Theoretical light-matter interaction on nanoscale

Plasmonic properties of noble metal structures

Skjølstrup, Enok Johannes Haahr

Publication date:
2020

Document Version
Publisher's PDF, also known as Version of record

THEORETICAL LIGHT-MATTER INTERACTION ON NANOSCALE – PLASMONIC PROPERTIES OF NOBLE METAL STRUCTURES

BY
ENOK JOHANNES HAAHR SKJØLSTRUP

DISSERTATION SUBMITTED 2019

AALBORG UNIVERSITY
DENMARK

Theoretical light-matter interaction on nanoscale - Plasmonic properties of noble metal structures

Ph.D. Dissertation
Enok Johannes Haahr Skjølstrup

Dissertation submitted December 23, 2019

Dissertation submitted: December 23, 2019

PhD supervisor: Associate Professor Thomas Møller Søndergaard
Aalborg University, Denmark

PhD co-supervisor: Professor Thomas Garm Pedersen
Aalborg University, Denmark

PhD committee: Professor Sergey Sorokin (chairman)
Aalborg University

Professor Jesper Mørk
Technical University of Denmark

Professor Femius Koenderink
Center for Nanophtotonics, AMOLF

PhD Series: Faculty of Engineering and Science, Aalborg University

Department: Department of Materials and Production

ISSN (online): 2446-1636
ISBN (online): 978-87-7210-572-7

Published by:
Aalborg University Press
Langagervej 2
DK – 9220 Aalborg Ø
Phone: +45 99407140
aauf@forlag.aau.dk
forlag.aau.dk

Printed in Denmark by Rosendahls, 2020

English abstract

This thesis deals theoretically with the interaction between light and noble metal structures on nanoscale. This interaction sometimes involve the excitation of a particular wave, called a surface plasmon polariton (SPP), which is a wave bound to and propagating along an interface between a metal and a dielectric. Such an SPP can be found in many different structures, where the following structures are considered in this thesis: An array of ultrasharp or rectangular grooves in gold, a dielectric gap located between two parallel metal walls, and a metal slab sandwiched between different dielectrics.

First a classical model neglecting quantum effects is applied to study the transition regarding the optics from one to multiple grooves in gold. It is found that the optical cross sections scale almost linearly with the number of grooves, thus not explaining why a single groove can be a good scatterer, while a groove array can be a good absorber with very low reflectance. Instead it is found that when the incident field is a narrow Gaussian beam focused entirely with an array of 20 grooves, the reflectance becomes the same as for an infinite array illuminated by a plane wave. Here, the reflectance from an ultrasharp and rectangular groove array becomes low in a broadband and narrowband wavelength interval, respectively.

Quantum effects are included afterwards, where the electron density is calculated using density-functional theory in the jellium model. Studying SPPs propagating in a dielectric gap between two parallel gold walls, the plasmon mode index converges to the refractive index of bulk gold in the limit of vanishing gap width. Thereby it restores the correct physical behaviour, and removes the unphysical divergence obtained using a classical model. Furthermore, the calculated reflectance from an ultrasharp groove array becomes in much better agreement with previous measurements compared to a classical model.

Next, a metal slab sandwiched between different dielectrics is studied with quantum effects taken into account. Here it is found that the plasmon mode index does not diverge in the limit of vanishing slab width, instead the plasmon mode ceases to exist for slabs below a cut-off thickness in the subatom range. For larger slabs, quantum effects imply that the imaginary

part of the mode index is significantly enhanced, and surprisingly, for wide slabs approaching bulk, the increase is 20 %. This is explained in terms of strong plasmonic absorption mostly taking place at narrow peaks a few Å outside the surface. In addition, anisotropy is found to modify the imaginary part of the mode index for both kinds of plasmon modes, but the relative increase is much larger for the long-range mode.

Finally, electron-energy loss spectroscopy (EELS) is studied in the non-retarded limit, where an electron is incident perpendicular or parallel to a thin metal film. SPPs are excited in the metal when the energy and momentum of the electron and the SPPs are the same, implying that the electron loses an amount of energy. Surprisingly, the energy loss calculated in a classical model is in good agreement with that obtained in the random-phase approximation taking quantum effects into account. This is explained in terms of the electron motion in the metal, which mainly takes place in the direction parallel to the film, where there is no confinement, thus reducing the quantum effects from the interfaces.

Danish abstract

I denne afhandling undersøges teoretisk hvordan lys interagerer med strukturer af ædelmetaller på nanoskala. Denne interaktion kan indbefatte excitationer af en bestemt bølgetype, kaldet overfladeplasmoner, som er bundne til en overflade mellem et metal og et dielektriaka, og som propagerer langs overfladen. Sådanne overfladeplasmoner findes i mange forskellige metalstrukturer, hvor der i afhandlingen fokuseres på følgende strukturer: Periodiske ultraskarpe eller rektangulære riller i guld, et dielektrika placeret mellem to parallelle metalvægge og til sidst en metalfilm omgivet af forskellige dielektrika.

Først anvendes en klassisk model uden kvanteeffekter til at studere overgangen fra én til mange riller i guld. Det findes at de optiske spredningstværsnit skalerer stort set lineært med antal riller. Det forklarer altså ikke, hvorfor en enkelt rille kan sprede lys meget, samtidig med at en uendelig række af den samme rille absorberer lyset meget, hvilket medfører en meget lav reflektans fra strukturen. Istedet anvendes et snævert Gaussisk beam som det indsendte felt, og det findes at når dette beam er fokuseret udelukkende indenfor en række af 20 riller, så opnås samme reflektans som ved en uendelig række af riller, når det indsendte felt er en planbølge. Grundet forskellen i geometri på de ultraskarpe og rektangulære riller, fås en reflektans fra førstnævnte som er lav for alle synlige bølgelængder, mens reflektansen fra sidstnævnte kun er lav for en specifik bølgelængde bestemt udfra dybden af rillen.

Dernæst tages kvanteeffekter med, hvor elektrontætheden beregnes udfra tæthedsfunktionalteori i jellium modellen. Undersøges her plasmoner, som propagerer i et snævert gab mellem to guldoverflader, findes det at mode indekset for plasmonen konvergerer mod brydningsindekset i rent guld i grænsen når gabets tykkelse går mod nul. Dette genopretter den korrekte fysiske opførsel, og fjerner den ufysiske divergens som findes i klassiske modeller uden kvanteeffekter. Derudover bliver den beregnede reflektans fra en række af ultraskarpe riller i guld i meget bedre overensstemmelse med tidligere målinger, end hvad tidligere fundet i en klassisk model.

Herefter undersøges den modsatte geometri, dvs. en metal film omringet af forskellige dielektrika. Her findes det at mode indekset ikke divergerer

når filmtykkelsen går mod nul, istedet ophører plasmonen med at eksistere når filmen bliver meget tyndere end et enkelt atomlag. For tykkere film, findes det at imaginærdelen af mode indekset stiger kraftigt som følge af kvanteeffekter, og at stigningen højst overraskende er hele 20 % for tykke film som opfører sig som bulk guld. Derudover er responsen generelt anisotrop da filmen kun er begrænset i én retning. Effekten af anisotropi er tilstede for begge typer af plasmoner i filmen, men den relative effekt er langt større for den langtrækkende plasmontype.

Til sidst studeres hvordan en elektron taber energi når den indsendes parallelt eller vinkelret på en tynd metal film. Elektronens tab af energy skyldes at den eksiterer plasmoner på overfladen. Overraskende nok er det beregnede tab i henholdsvis en klassisk model og en fuld kvantemekanisk model i god overensstemmelse. Dette forklares ved at elektronerne mest bevæger sig parallelt med filmen hvor den ikke er begrænset, hvilket reducerer kvanteeffekterne fra overfladerne.

Contents

Thesis Details

Thesis title:	Theoretical light-matter interaction on nanoscale - Plasmonic properties of noble metal structures
PhD student:	Enok Johannes Haahr Skjølstrup
Supervisor:	Associate Professor Thomas Møller Søndergaard
Co-supervisor:	Professor Thomas Garm Pedersen

The main body of this thesis consist of the following papers.

[I] **Enok J. H. Skjølstrup** and Thomas Søndergaard, Optics of multiple ultrasharp grooves in metal, J. Opt. Soc. Am. B **34**, 673 (2017).

[II] **Enok J. H. Skjølstrup**, Thomas Søndergaard, Kjeld Pedersen, and Thomas G. Pedersen, Optics of multiple grooves in metal: transition from high scattering to strong absorption, J. Nanophotonics, **11**(4), 046023 (2017).

[III] **Enok J. H. Skjølstrup**, Thomas Søndergaard, and Thomas G. Pedersen, Quantum spill-out in few-nanometer metal gaps: Effect on gap plasmons and reflectance from ultrasharp groove arrays, Phys. Rev. B **97**, 115429 (2018).

[IV] **Enok J. H. Skjølstrup**, Thomas Søndergaard, and Thomas G. Pedersen, Quantum spill-out in nanometer-thin gold slabs: Effect on the plasmon mode index and the plasmonic absorption, Phys. Rev. B **99**, 155427 (2019).

[V] A. R. Echarri, **Enok J. H. Skjølstrup**, Thomas G. Pedersen, and F. Javier Garcia de Abajo, Theory of EELS in atomically thin metallic films, submitted to Phys. Rev. Res. Available on ArXiv:1912.09414.

This thesis has been submitted for assessment in partial fulfillment of the PhD degree. The thesis is based on the published and submitted scientific papers listed above. Parts of the papers are used directly or indirectly in the summary in the thesis. As part of the assessment, co-author statements have been made available to the assessment committee and are also available at the Faculty.

Preface

This thesis summarizes the work that I have done as a PhD student at the Department of Materials and Production at Aalborg University in the period from January 2017 to December 2019. Many people have helped and supported me during this period, and I owe all of them a thanks.

First of all, I would like to thank my two supervisors, Thomas Søndergaard and Thomas Garm Pedersen, for providing me the opportunity of becoming a PhD student under their supervision. It has been exciting to work together with them and gain insight into their extensive knowledge within many physical subjects, and each in their own way, they have both learned me a lot. I would also like to thank Prof. F. Javier Garcia de Abajo for making my three months research stay at ICFO in Barcelona possible and pleasant. Also thanks to his PhD students for warmly welcoming me in their research group, especially to Alvaro Rodriguez Echarri whom I enjoyed working closely together with. Also thanks to my colleagues, in particular Jonas Have, Høgni Kamban, and Alireza Taghizadeh, for nice conversations and for making a good office environment.

Finally, I owe a big thanks to my family. Especially to my wife, Natascha Skjølstrup, who has always been of great help to me. This thesis would not have been possible without your love and support.

Enok Johannes Haahr Skjølstrup
Aalborg University, December 23, 2019

Chapter 1

Introduction

Light interacts with matter in many different ways. This results in a wide variety of optical phenomena, for example the refraction of light in a prism or a rainbow, and the reflection of light in the sea at sunset. To understand such optical phenomena, it is necessary to understand the optical properties of different materials. Metals are a class of materials, and in this thesis it is studied theoretically how light interacts with noble metals, which are partly characterized by their high reflectance and thereby shiny surfaces. The dielectric function of noble metals is to a good approximation described by the Drude model, where the electrons are modelled as being attracted to the atomic nucleus by a spring force, which causes them to oscillate out of phase with a driving electric field, with their motion being damped due to different scattering mechanisms. When noble metals interact with light, it sometimes involve the excitation of a particular kind of wave, called a surface plasmon polariton (SPP). This is an eigenmode of a system consisting of a noble metal and a dielectric, for example air, and behaves as a wave bound to and propagating along the interface [1].

Letting the x axis be the propagation direction, the wave number component k_x of the SPP is larger than the free space wave number k_0. This implies that the SPP eigenmode cannot simply be excited by incident light of any frequency from free space, as the momentum of the incident light will always be lower than the momentum of the plasmon. Hence in order to excite plasmons with incident light, the wave number along the interface has to be increased above its free space value, such that the momentum of the incident light can equal the momentum of the plasmon [1]. This can be obtained in the Otto [2] or Kretschmann [3] configurations, where light is incident on a glass prism, and generates an evanescent field at the backside of the prism. The evanescent field can then excite plasmons at a nearby interface between metal and air as illustrated in Fig. 1.1(a,b) from Ref. [1]. The excitation of

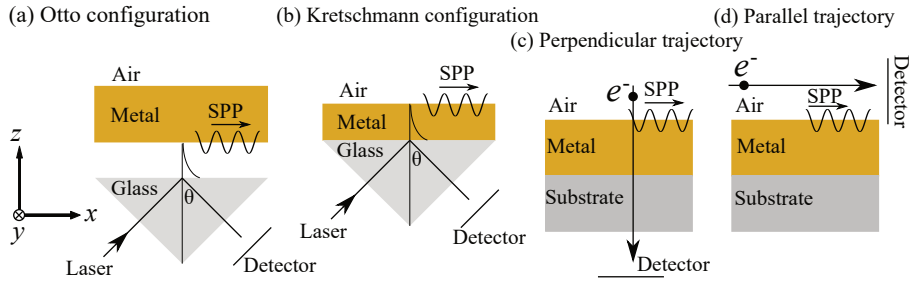

Fig. 1.1: Different methods to excite plasmons at an interface between metal and air. The excitation source is evanescent light in the Otto and Kretschmann configutations (a,b), while an electron incident perpendicular or parallel to the film serves as the excitation source in (c,d). The panels (a,b) are from Ref. [1].

SPPs can then be seen as minimum in the reflected light for a certain angle θ, and the results reported in Refs. [2,3] served as experimental verifications of SPPs propagating along metal-dielectric interfaces.

Another method for detecting plasmons was proposed by Ritchie in Ref. [4], whose idea was to shoot fast electrons onto a metal film, thus applying incident electrons instead of incident light as the excitation source. The moving electron serves as an evanescent source of radiation, as its point charge behaviour in real space implies a charge density that is completely delocalized in momentum space [5,6], thereby containing a parallel momentum necessary to excite a plasmon. The electron can be incident perpendicular or parallel to the metal film as illustrated in Fig. 1.1(c,d). In general, the electromagnetic field of an SPP propagating along a metal slab can be bound to both interfaces of the slab, which will be further elaborated upon below. Fig. 1.1 simply illustrates different methods to excite an SPP, which is shown as propagating along a single interface. The plasmons can be detected from the loss in kinetic energy of the electrons, a method which later became known as electron-energy loss spectroscopy (EELS) [5]. Peaks are found at certain frequencies in an EELS spectrum, and optical properties of a surface can thus be determined from the location of the peaks. The method was applied to detect plasmons in aluminium as reported in Ref. [7], while Refs. [8–10] reported the same detection in sodium, potassium, and silver. In gold, different damping mechanisms suppress certain peaks, which makes silver a good candidate for studying the optical properties of noble metal surfaces. Furthermore, the surface orientation plays a role on the optical properties, as can be revealed in e.g photoemission experiments, showing a difference between Ag(111) and Ag(100), as a projected gap is found in Ag(111) but not in Ag(100) [11]. Using other experimental techniques Refs. [12,13] measured the photoelectric yield of Ag(111), where the work function has been lowered

by placing the silver surface in an electrolyte, while Refs. [14, 15] measured two-photon photoemission (2PP) in Ag(111), in which the surface states (SS) due to the projected gap can be observed.

A single planar interface between metal and dielectric is the simplest physical system that supports plasmons [1]. More complex plasmonic structures are found in e.g dielectric gaps between parallel metal walls [16–19], a metal slab surrounded by different dielectrics [20–25], a nanometer size gap between spherical and triangular nanoparticles [26, 27], and in rectangular or tapered grooves [16, 19, 28, 29]. Here the metal slab supports two different kinds of plasmonic modes, which are classified as short- and long-range modes, respectively, with opposite symmetry on the electromagnetic fields [21–25]. Plasmons are characterized by their parallel wave number k_x, or correspondingly their dimensionless mode index, defined as $\beta = k_x/k_0$. Refs. [16–29] applied a classical model neglecting quantum effects, and found that β diverges in the limit of vanishing gap or slab width [16, 20–23, 28]. Especially for the gap structure, this cannot be correct from a physical point of view, from which it is expected that a structure consisting of an ultrathin dielectric gap between metals would behave almost as if there was no gap, i.e a structure consisting simply of pure metal. Furthermore, calculating the reflectance from an ultrasharp groove array using classical models [30, 31], it is necessary to assume a bottom width of the grooves of only 0.3 nm, (roughly the same as the diameter of a gold atom) in order to obtain a reflectance in reasonable agreement with the measured reflectance from Ref. [31]. The exact bottom width cannot be measured precisely, but is highly unlikely to be only 0.3 nm, and when assuming this bottom width, the calculated reflectance, in addition, contains oscillations that are not present in the measured spectrum. This illustrates the limitations of the classical models when pushing the size of the structure down to atomic scale, where it is now possible to fabricate structures [32].

Furthermore, the reflectance from a groove array (ultrasharp or rectangular) has been computed using two different methods within the classical model, where the incident light is p-polarized in both cases. The first method is based on the propagation of gap plasmons within the grooves using the stack matrix method [30], while the second method is based on a scattering problem, where the grooves are treated as scatterers, and the problem is solved using a full Greens function surface integral equation method (GF-SIEM) [33]. Importantly, the two methods give the same scattering cross sections from rectangular groove arrays [29] and the same reflection from an ultrasharp groove array in gold [28, 30]. The rectangular groove array is as a narrowband absorber, where the depth of the rectangular grooves determines which wavelengths that are absorbed [29], which is of particular interest in selective thermal emitters [34, 35], that can be advantageous in thermophotovoltaics [36, 37]. The ultrasharp groove array, on the other

hand, serves as a broadband absorber of light, giving a low reflectance in the entire visible spectrum, and is therefore called black gold [28,31]. A possible application for this is as polarizers for ultrashort laser pulses, where it is utilized that *p*- and *s*- polarized light will be efficiently absorbed and reflected, respectively [38]. However, the scattering cross section of a single groove (rectangular or ultrasharp) in gold may actually be larger than the physical dimension of the groove [29,39], which implies that a single groove is a good scatterer, while a groove array is a good absorber (narrowband or broadband). Hence the optical properties of a periodic array does not follow straightforwardly from that of a single groove, and it was stated in Ref. [39] that the low reflectance from the groove array could be due to destructive interference occurring between light scattered from the individual grooves, a statement that had not yet been tested.

A way to improve the accuracy of the models when shrinking to atomic scale is to take quantum effects into account. One of the consequences of quantum mechanics is the tunnelling effect, which states that an electron can tunnel through a potential barrier with a certain probability [40], a transition that is forbidden in classical mechanics. The tunnelling effect implies that the electron density of the free electrons in the conduction bands becomes smooth across an interface between metal and dielectric. The calculation of the position-dependent free electron density builds upon the pioneering work performed by Hohenberg, Kohn, and Sham back in the 1960s and 70s. In 1964, Hohenberg and Kohn proved the remarkable theorem, that the ground state energy of a system consisting of N electrons is simply a functional of the electron density $n(\vec{r})$ [41]. Hence the energy only depends on the 3-dimensional density, rather than the $3N$-dimensional wave function, thus significantly reducing the complexity of the problem. This implied that the Hohenberg-Kohn theorem laid the foundation of what became known as density-functional theory (DFT). A year later, Kohn and Sham applied the variational theorem on the energy functional and obtained the Kohn-Sham equations [42]. Here the Kohn-Sham orbitals are found as eigenvalues to a Hamiltonian whose potential consists of a Coulomb, an exchange, and a correlation term, all given as functionals of the electron density, implying that the equations have to be solved in a self-consistent way in order to obtain the density of the free electrons. Where the two first potential terms stems from operators obtained using rigid quantum mechanics, the correlation term describes what is missing to obtain the exact ground state energy, and of course it has no simple analytic expression. It is noticed that DFT is an *ab initio* method, meaning that it is, in principle, not necessary to have any prior knowledge of the system or to fit any value to experiments [43]. In addition, DFT can be further extended to account for time varying potentials, resulting in time-dependent DFT (TDDFT) [44].

However, the Hohenberg-Kohn theorem and the Kohn-Sham equations

do not give any prescription of the actual functionals, which are of course needed in practice when solving the equations. A method to obtain these functionals is to apply the jellium model, where it is assumed that the positive charge density of the ions is smeared out across the metal [45,46], while atomistic models include the discrete nature of the positive ions [47–51], which of course makes the models more complex. To obtain expressions for the exchange and correlation potentials, further simplifications have to be made. In the local density approximation (LDA), the system is approximated by a homogenous electron gas, which implies that the exchange potential can be calculated analytically as a functional of the density [44, 45]. The correlation energy is calculated using quantum Monte Carlo simulations [52,53] for some specific densities in a wide range, and the correlation potential as a functional of n is obtained by fitting an analytic function to these data. The most widely used correlation potentials within the LDA are the parametrizations of Perdew-Zunger [54] and Vosko-Wilk-Nusair [55]. A model beyond the LDA is the generalized gradient approximation (GGA), where density variations are also taken into account [56]. Within the GGA, the exchange and correlation functionals are given in Refs. [57, 58]. It is noticed that DFT calculations only give the density of the free electrons in the conduction bands, while the bound electrons in the lower lying energy bands have to be treated separately. In addition, the surface orientation is not considered in the jellium model, as information about the positions of the individual atoms is not taken into account.

DFT is a widely used method in many aspects of both physics, chemistry, and biology, and is contained in more than 15.000 papers every year [56], within topics spanning from interface phenomena [59, 60], condensed matter physics [61–63], and the modelling of enzyme reactions [64, 65] and chemical properties of molecules [66]. In closer relation to the work presented in this thesis, DFT has also been applied to calculate the optical cross sections of metal clusters and spheres [67, 68], metal nanowires [69–71], and the plasmon resonance of metal dimers [72,73] and semiconductor nanocrystals [74]. Shrinking the size of the physical system down to atomic scale, quantum effects are found to have a significant impact on the local field intensity in metal nanostructures [75] and the local density of states in metal gaps and slabs [76]. The latter being in agreement with experimental results from Ref. [77].

However, it takes a lot of time to calculate the electron density by solving the Kohn-Sham equations self-consistently, even when applying the LDA and the jellium model. This implies that DFT is a slow and very computationally demanding method, which limits its practical use for structures consisting of more than 100 atoms [78], and has motivated the development of faster methods. One of these methods is a semiclassical quantum corrected model (QCM), where a parameter is included to account for electron tunnelling [79–

81]. The quantum corrected dielectric function becomes similar to the Drude dielectric function, but with the damping term depending on the tunnelling parameter and neglecting the response from the bound electrons. The results obtained using QCM can be in good agreement with those obtained using TDDFT [79].

Other methods to describe the quantum dielectric function include nonlocality, meaning that the displacement field at a position $\vec{r}$ depends on the electric field at all other positions in space as $\vec{D}(\omega,\vec{r}) = \int \overleftrightarrow{\varepsilon}(\omega,\vec{r},\vec{r}\,') \cdot \vec{E}(\omega,\vec{r}\,')\,\mathrm{d}\vec{r}\,'$. In a local model the dielectric function $\overleftrightarrow{\varepsilon}(\omega,\vec{r},\vec{r}\,')$ reduces to $\overleftrightarrow{\varepsilon}(\omega,\vec{r})\delta(\vec{r}-\vec{r}\,')$, implying that $\vec{D}(\omega,\vec{r}) = \overleftrightarrow{\varepsilon}(\omega,\vec{r}) \cdot \vec{E}(\omega,\vec{r})$. The concept of nonlocality is, however, treated in two different ways. The hydrodynamical Drude model (HDM) considers the interacting electron gas as governed by fluid dynamics described by the Navier-Stokes equation [82], where the transport can be both convective and diffusive [83]. The nonlocality in the HDM enters in the hydrodynamical pressure described by the velocity $\beta = \sqrt{3/5}v_F$, where v_F is the Fermi velocity [83–85]. A differential equation is derived relating the electric field and the current density, where the classical Ohm's law is restored if $\beta = 0$. By solving the differential equation, it is found that nonlocality slightly blue-shifts the plasmon resonances in sodium and gold dimers [79, 80, 85, 86] and gold cylinders [84, 87], and that the effect of nonlocality is small compared to the effect of spill-out for plasmons propagating in narrow metal gaps and slabs [76]. Furthermore, the dielectric function is a tensor in the HDM, where β modifies the longitudinal component, while the transverse component is still given by the classical Drude model [79, 87]. Here, the response from the bound electrons is in Ref. [76] modelled as a step function changing abruptly at the jellium edge resulting in a dielectric function sufficiently far inside the metal in agreement with values obtained from experiments [88], while the bound electron response has been neglected in Refs. [79, 80, 83, 84].

A more accurate way to treat the dielectric response is by rigid quantum mechanical response theory, studying how materials respond when exposed to an external field. A widely used method is the random-phase approximation (RPA), where the exchange and correlation terms in the potential are neglected, meaning that the potential is solely described by Coulomb interactions [89]. Within the RPA, the dielectric response from the free electrons is obtained by perturbing them to first order by a plane wave external potential, giving rise to an induced charge proportional to the external potential. In a homogeneous medium, the dielectric function is isotropic and simplifies to $\overleftrightarrow{I}\,\varepsilon(\omega,\vec{r}-\vec{r}\,')$, and after a convolution one obtains $\vec{D}(Q,\omega) = \varepsilon(Q,\omega)\vec{E}(Q,\omega)$ in reciprocal space, which is related to the real space quantities through Fourier transforms. Here $\varepsilon(Q,\omega)$ can be computed as an integral over the Fermi sphere, and a closed form of the integral was first

obtained by the Danish professor Jens Lindhard in 1954 [90]. This expression for the dielectric function was applied by other pioneers in the field of RPA and the many-electron problem [91–93]. Later, the Lindhard function was generalized by Mermin in Ref. [94] to include damping while conserving the number of electrons, and furthermore contain the desired local limit as $\lim_{Q\to 0} \varepsilon^{\text{Mermin}}(Q,\omega) = \varepsilon(\omega)$, where the latter is the local Drude dielectric function. The original Lindhard (or Mermin) dielectric function is only valid in homogeneous media, but is extended in the specular reflection model (SRM) to account for non-homogeneous media by matching boundary conditions at the interfaces [78, 95]. In addition, Ref. [78] reported that nonlocality slightly blue-shifts the plasmon resonances in gold dimers, a result that was also obtained using the HDM in Refs. [79]. For 2-dimensional (2D) structures with the z axis perpendicular to the structure, the dielectric function is given as an inverse Fourier transform as $\overleftrightarrow{\varepsilon}(\vec{Q},\omega,z,z') = \int \overleftrightarrow{\varepsilon}(\omega,\vec{r},\vec{r}')e^{-i\vec{Q}\cdot(\vec{R}-\vec{R}')}\,\mathrm{d}^2\vec{R}\,\mathrm{d}^2\vec{R}'$, where $\vec{Q}$, $\vec{R}$, and $\vec{R}'$ are in the (x,y) plane, thus parallel to the structure. Calculating $\overleftrightarrow{\varepsilon}(\vec{Q},\omega,z,z')$ within the RPA, it consists of a local isotropic Drude term and an anisotropic nonlocal term [93], where a smooth electron density can be incorporated to account for quantum spill-out [96, 97]. Such an RPA model has previously been used to calculate the dynamical response in thin films [98, 99], the plasmonic properties of graphene [100] and thin metal films [101], and the loss probability for electrons incident on metal films [5, 102, 103]. Like in Ref. [76], the response from the bound electrons was in Refs. [5, 100–103] treated as a step function giving the measured dielectric constant [88] sufficiently far inside the metal regions.

1.1 This thesis

The main goal of this thesis is to obtain a theoretical understanding of plasmons propagating in noble metal structures at nanoscale. The structures of interest in this work consist of ultrasharp and rectangular groove arrays, dielectric gaps located between parallel metal walls, and metal slabs sandwiched between different dielectrics, where the four different structures are illustrated in Fig. 1.2. The papers I and II apply a classical dielectric function, while quantum effects are taken into account in the papers III, IV, and V, as it has turned out that classical models have their limitations when the structures become small enough.

Classical models

The papers I and II study the optics of a multiple (ultrasharp, rectangular, and tapered) groove array in metal within a classical model. Such structures can be considered as a reference structure, consisting of a plane interface be-

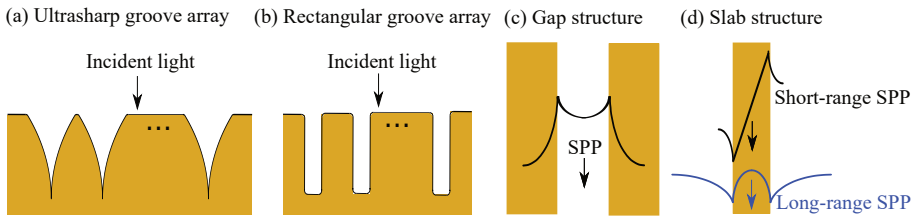

Fig. 1.2: Schematic of the different structures studied in this thesis.

tween metal and air, in which the groove array is placed as a scatterer. The structures are invariant in one spatial direction (Fig. 1.2(a,b)) and p-polarized light has its magnetic field in the same direction [29]. This implyies that the optical properties of the grooves can be calculated using the Greens function surface integral equation method, where the surface is two-dimensional (GFSIEM). From the wave equation, an integral equation is constructed relating the magnetic field and its normal derivative at the boundary of the scatterer, where the integration kernels consists of a scalar Greens function for the layered reference structure and its normal derivative. The Greens function is calculated using a plane wave expansion of the free space Greens function, where the reflected and transmitted parts are obtained using the Fresnel coefficients. By discretizing the structure, the integral equation is written on matrix form, and solving it gives the magnetic field and its normal derivative at the boundary of the scatterer [33]. These fields are afterwards used to calculate the scattered field, from which the scattering and extinction cross sections can be computed [29].

Recently, the GFSIEM has been applied to calculate the optical cross sections of a single ultrasharp groove in gold [39] and the reflectance from an infinite array of ultrasharp grooves [30] when the incident field illustrated by the arrow in Fig. 1.2(a,b) is a plane wave. For the infinite array, a periodic Greens function from Ref. [33] was applied. As mentioned above, a single ultrasharp groove is a good broadband scatterer [39], while an ultrasharp groove array is a good broadband absorber [30, 31], showing that the optical properties of an array does not follow straightforwardly from the case of a single groove. It was stated in Ref. [39] that the low reflectance from the periodic array could be due to destructive interference between light scattered from the individual grooves. This statement is tested in paper I, studying the transition regarding the optics of a single ultrasharp groove to a periodic, but finite, array of grooves. In addition, a Gaussian beam is applied as the incident field, and it is examined how the reflectance from a groove array depends on the width of beam and the number of grooves, the latter being in order to examine how many grooves that are necessary for the array to

behave as infinite. At last, it is investigated how the ultrasharp groove array reflects, transmits and scatters an incident plasmon. The same method is applied in paper II for a finite array of rectangular and tapered grooves acting as narrowband absorbers [29], and therefore behaving quite differently compared to the ultrasharp groove array.

Quantum models

While a classical model is applied in paper I and II, quantum effects are taken into account in the papers III, IV, and V. Here the electron density of the free electrons across the metal structures in Fig. 1.2(c,d) is calculated using DFT in the jellium model, and it becomes smooth across the interfaces between metal and dielectric due to the effect of quantum spill-out. The free electron density in the jellium model is an essential part of the thesis, and is applied to construct a quantum dielectric function, whose level of complexity is increased through paper III, IV, and V. As mentioned, rigid quantum mechanical response theory within the RPA implies that the quantum dielectric function corresponding to the free electrons consists of a local isotropic Drude term plus an anisotropic nonlocal term [96]. Here the Drude term depends on the spatially varying electron density, while the nonlocal term depends on the wave functions and energies of the discrete quantum states. All quantities which are obtained within the DFT model. Similarly to several papers [76, 100, 101, 103], the response from the bound electrons is modelled as a step function at the jellium edge giving the measured dielectric response sufficiently far inside the metal regions. Since nonlocality only slightly blueshifts the plasmon resonances in different nanostructures [78–80, 84, 87], and is a small modification compared to the effect of spill-out [76], nonlocality is simply neglected in paper III. The quantum dielectric function of the free electrons is therefore described entirely as a local isotropic Drude term with varying electron density. In this paper, the focus is on plasmons propagating in a nanometer thin gap of air located between two parallel walls of gold, where the magnetic field of the plasmon is illustrated by the black curve in Fig. 1.2(c). Here the plasmon mode index unphysically diverges in the limit of vanishing gap width when using a classical model neglecting spill-out [16, 17, 20, 28]. It is expected from a physical point of view, that the plasmon mode index of an ultrathin gap should be almost the same as the refractive index of bulk gold, as the structure is almost bulk gold in this limit, and it is examined in paper III whether this correct physical behaviour can be restored when quantum spill-out is taken into account. Furthermore, it is investigated whether the calculated plasmon mode index with spill-out included can be applied to obtain a reflectance from an ultrasharp groove array in better agreement with measurements [31] compared to classical models.

Paper IV studies plasmons propagating in a metal slab sandwiched be-

tween different dielectrics as illustrated in Fig. 1.2(d). Here the magnetic field of the two kinds of plasmonic modes are illustrated by the black and blue curves. The quantum dielectric function in paper IV includes the nonlocality in the direction perpendicular to the slab, while neglecting it in the parallel direction, thus applying $\overleftrightarrow{\varepsilon}(\vec{0},\omega,z,z')$ for the dielectric function. Furthermore, based on symmetry arguments, the nonlocal part of the dielectric function regarding z and z' is modelled by an averaging procedure resulting in a local anisotropic step function, which gives the correct net response from the slab for both kinds of modes. As for the gap structure, the short-range plasmon mode index in a classical model diverges in the limit of vanishing slab width [21–23], and like in paper III, it is examined whether this divergence is removed when quantum spill-out is taken into account. The electron density is only modified due to spill-out in a narrow region close to the metal slab, and it is tested how this effect modifies the plasmons propagating in wide slabs approaching bulk. Furthermore, the effect of anisotropy on the plasmon mode indices is examined.

Paper V studies EELS in atomically thin metal films, where the electron can be incident parallel or perpendicular to the film as illustrated in Fig. 1.1(c,d). The incident electron excites plasmons, implying that it loses an amount of energy described by the loss probability, which is computed in the nonretarded limit, where the response from the electron is instantaneous. The response of the film is calculated within the full nonlocal RPA, thus not simplifying the dielectric function as in paper III and IV. The potential across the films is described by the same DFT potential as in paper III and IV, as well as by an atomic layer potential (ALP) based on fitting the work function, the projected gap, the surface states and the Fermi energy to experimental data [101, 104]. It is studied how the loss probability depends on the number of metal monolayers, the kinetic energy of the incident electrons, as well as the surface orientation. The results obtained using DFT and ALP are compared to those obtained using classical models, and the plasmon mode indices are compared to those obtained using the simplified RPA response in paper IV.

Chapter 2

Theory and methods

This chapter presents the theory and methods that have been used to produce the results in the thesis. Where the theory has only been briefly explained in some of the papers, this chapter focuses on explaining it in more detail. The chapter begins with a short description of the applied method regarding the optics of multiple grooves in metal applied in paper I and II. The DFT method in the jellium model, which has been applied to calculate the electron densities in the papers III, IV, and V, is presented afterwards, along with a review of the physical concepts behind SPPs. The theory regarding EELS in thin metal films will, however, not be included here, as it is explained in detail in paper V.

2.1 Optics of multiple grooves in metal

Paper I and II study the optics of multiple grooves in metal within a classical model, where the grooves are ultrasharp and rectangular/tapered, respectively. This section contains a description of the method that has been applied in these two papers. First the theory behind the scattering problem is outlined based on Ref. [33], after which the theory for dealing with a periodic, but finite, array will be explained.

2.1.1 The scattering problem

Fig. 2.1(b) illustrates an array of N identical grooves illuminated by a Gaussian beam of waist radius w_0 centered in x_0. There is a separation d between the grooves, and the array has total length L, which is related to the beam waist through the ratio parameter γ as $w_0 = \gamma L/2$, where the figure shows the case when $\gamma = 1$. SPPs are generated within the grooves and propagate along the interface as illustrated. The structure can be considered as

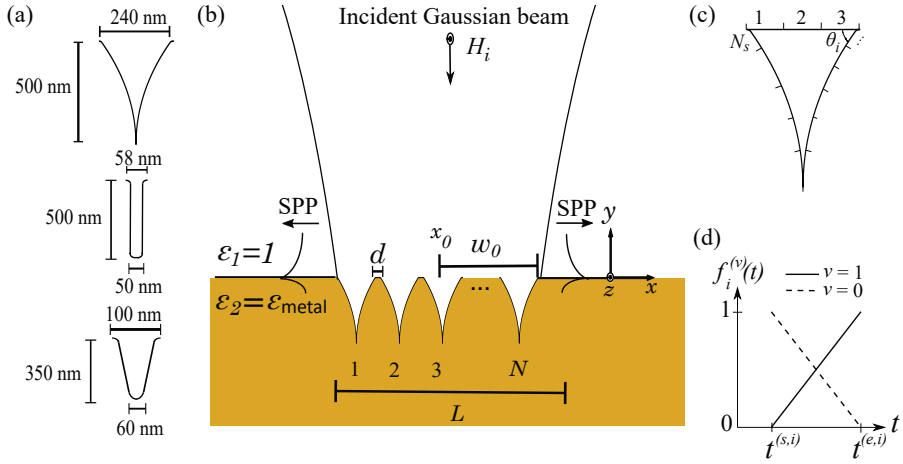

Fig. 2.1: (a) shows a schematic of the three different kinds of grooves (ultrasharp, rectangular, and tapered), while (b) illustrates the array of N grooves illuminated by a Gaussian beam of width w_0. A discretization of a groove is illustrated in (c), and the functions applied in the linearly varying fields are shown in (d).

a reference structure consisting of a plane interface between metal and air, in which the grooves are placed as scatterers of air in the metal region, and Fig. 2.1(a) shows a schematic of the ultrasharp, rectangular, and tapered grooves. The structure and the magnetic field are invariant in the z direction ($\vec{H}(\vec{r}) = \hat{z}H(\vec{r})$), implying that the Greens function surface integral equation method (GFSIEM) can be applied. Within that method, it can be shown that the magnetic field and its normal derivative satisfy the following integral equations [33]

$$H(\vec{r}) = \begin{cases} H_0(\vec{r}) - \oint_{C_1} \{g_1(\vec{r},\vec{r}')\varphi(\vec{r}') - H(\vec{r}')\hat{n}' \cdot \nabla' g_1(\vec{r},\vec{r}')\}\, \mathrm{d}l', & \vec{r} \in \Omega_1 \\ \oint_{C_2} \{g_2(\vec{r},\vec{r}')\varphi(\vec{r}') - H(\vec{r}')\hat{n}' \cdot \nabla' g_2(\vec{r},\vec{r}')\}\, \mathrm{d}l', & \vec{r} \in \Omega_2. \end{cases} \tag{2.1}$$

Here $H_0(\vec{r})$ is the incident field, while $\varphi(\vec{r}') = \hat{n}' \cdot \nabla' H(\vec{r}')$ is the normal derivative of the magnetic field, where the normal vector $\hat{n}'$ points out of the scatterer. The regions Ω_1 and Ω_2 are the reference structure and the scatterer, respectively, and the curves C_1 and C_2 lie just outside and inside the scatterer surface, respectively. The Greens function for the scatterer, i.e for $\vec{r} \in \Omega_2$, is given by

$$g_2(\vec{r},\vec{r}') = \frac{i}{4} H_0^{(1)}(k_0 n_1 |\vec{r} - \vec{r}'|), \tag{2.2}$$

where $H_0^{(1)}$ is the Hankel function of zero'th order and first kind, while $k_0 = \omega/c$ is the free space wave number, and $n_1 = 1$ is the refractive index of the scatterer (air). The corresponding Greens function for the reference structure, g_1, is not given given as such a simple analytic expression, but is instead constructed using a plane wave expansion of the Greens function in a homogeneous medium. For $y, y' < 0$, g_1 consists of a direct and an indirect part due to the reflection at the interface between metal and air, and is given by [33]

$$g_1(\vec{r},\vec{r}') = \frac{i}{2\pi}\int_{k_x=0}^{\infty} \frac{\cos(k_x(x-x'))\left(e^{ik_{y2}|y-y'|} + r_{21}(k_x)e^{-ik_{y2}(y+y')}\right)}{k_{y2}}\,\mathrm{d}k_x, \tag{2.3}$$

where r_{21} is the reflection coefficient for p-polarized light incident from metal to air, thus given by [1]

$$r_{21}(k_x) = \frac{\varepsilon_1 k_{y2} - \varepsilon_2 k_{y1}}{\varepsilon_1 k_{y2} + \varepsilon_2 k_{y1}}, \text{ where } k_{yj} = \sqrt{k_0^2\varepsilon_j - k_x^2}, \text{ with } \mathrm{Im}(k_{yj}) \geq 0. \tag{2.4}$$

The Greens functions in Eq. (2.2) and (2.3) are the same as in Ref. [29], and applies for a scatterer of aribitrary shape. A corresponding periodic Greens function is applied in Ref. [30] to calculate the scattering from the infinite array. The integration kernels in Eq. (2.1) are constructed from the Greens functions in Eq. (2.2) and (2.3), and the integral equation can thus be solved. The solution gives H and φ at the boundary of the scatterer, which are afterwards applied to calculate the scattered field and the reflectance from the groove array. Notice that the Greens functions are singular when $\vec{r} = \vec{r}'$, which must be taken care of when integrating the Greens functions, as will be shown below. Furthermore, the scatterer surface needs to be discretized in order to solve the integral equation. Here, each groove is divided into N_s elements as illustrated in Fig. 2.1(c), where the position on the scatterer, $\vec{s}$, is given by the parameter t moving clockwise around the groove, with the start and end of element i denoted as $t^{(s,i)}$ and $t^{(e,i)}$, respectively. It is assumed that H and φ are linearly varying within each element, which is accounted for by introducing the following functions, which are also illustrated in Fig. 2.1(d).

$$f_i^{(v)}(t) = \begin{cases} 1 - (t - t^{(s,i)})/\Delta t_i, & t^{(s,i)} < t < t^{(e,i)}, \quad v = 0 \\ (t - t^{(s,i)})/\Delta t_i, & t^{(s,i)} < t < t^{(e,i)}, \quad v = 1 \\ 0, & \text{otherwise.} \end{cases} \tag{2.5}$$

Here $\Delta t_i = t^{(e,i)} - t^{(s,i)}$ is the length of element i, and H and φ along one of the curves C_u, where $u = \{1, 2\}$, can be written as

$$H_u(\vec{s}(t)) \simeq \sum_{i=1}^{N_s} \sum_{v=0}^{1} H_{u,i,v} f_i^{(v)}(t) \quad \text{and} \quad \varphi_u(\vec{s}(t)) \simeq \sum_{i=1}^{N_s} \sum_{v=0}^{1} \varphi_{u,i,v} f_i^{(v)}(t). \tag{2.6}$$

In general, it is possible to let the fields within an element be varying according to a higher order polynomial in order to achieve better convergence [33], but only the linearly varying fields have been considered in this work. From the discretization, the integral equation in Eq. (2.1) can now be written on matrix form. Inserting Eq. (2.6) in Eq. (2.1) and taking the limit where $\vec{r}$ approaches the surface from both sides (Ω_1 and Ω_2) gives the discretized integral equations.

$$H_{1,i}^s = H_{0,i} - \sum_{j=1}^{N_s} \sum_{v=0}^{1} \left(A_{i,j}^{(1,v)} \varphi_{1,j,v} - B_{i,j}^{(1,v)} H_{1,j,v} \right) \tag{2.7}$$

$$H_{2,i}^s = \sum_{j=1}^{N_s} \sum_{v=0}^{1} \left(A_{i,j}^{(2,v)} \varphi_{2,j,v} - B_{i,j}^{(2,v)} H_{2,j,v} \right). \tag{2.8}$$

Here the superscript s indicates that it is the field at the start of element i, and the integration kernels are given by [33]

$$A_{i,j}^{(u,v)} = P \int_{t'} g_u \left(\vec{s}_i^{\,s}, \vec{s}(t') \right) f_j^{(v)}(t') \, \mathrm{d}t' \tag{2.9}$$

$$B_{i,j}^{(u,v)} = P \int_{t'} \hat{n}' \cdot \nabla' g_u \left(\vec{s}_i^{\,s}, \vec{s}(t') \right) f_j^{(v)}(t') \, \mathrm{d}t' + \frac{\delta_{ij}\delta_{0v}}{2\pi} (\delta_{u1}\theta_i - \delta_{u2}(2\pi - \theta_i)). \tag{2.10}$$

Here P denotes the principal integral value, meaning that the singular point when $\vec{s}(t') = \vec{s}_i^{\,s}$ is simply excluded from the integration. The Greens function itself is only logarithmic singular, implying that there is no contribution to $A_{i,j}^{(u,v)}$ from the singularity. However, the singularity from the normal derivative of the Greens functions is stronger, and the contribution from it is given explicitly in the last term of $B_{i,j}^{(u,v)}$. Here the δ_{ij}'s are the Kronecker delta, while θ_i is the inner angle between two elements as illustrated in Fig. 2.1(c). Eq. (2.7) and (2.8) are now written on matrix form as

$$\begin{bmatrix} \left(\bar{\bar{I}} - \bar{\bar{B}}^{(1,0)} - \bar{\bar{B}}^{(1,1)} \bar{\bar{D}}_H \right) & \left(\bar{\bar{A}}^{(1,0)} + \bar{\bar{A}}^{(1,1)} \bar{\bar{D}}_\varphi \right) \\ \left(\bar{\bar{I}} + \bar{\bar{B}}^{(2,0)} + \bar{\bar{B}}^{(2,1)} \bar{\bar{D}}_H \right) & -\frac{\varepsilon_1}{\varepsilon_2} \left(\bar{\bar{A}}^{(2,0)} + \bar{\bar{A}}^{(2,1)} \bar{\bar{D}}_\varphi \right) \end{bmatrix} \cdot \begin{bmatrix} \vec{H}^{(0)} \\ \vec{\varphi}_1^{(0)} \end{bmatrix} = \begin{bmatrix} \vec{H}_0 \\ \vec{0} \end{bmatrix}, \tag{2.11}$$

where the vectors have been introduced as $\vec{H}^{(v)} = [H_{1v}, H_{2v}, \cdots, H_{N_s v}]^T$, $\vec{H}_0 = [H_{01}, H_{02}, \cdots, H_{0N_s}]^T$, and $\vec{\varphi}_u^{(v)} = [\varphi_{u1v}, \varphi_{u2v}, \cdots, \varphi_{uN_s v}]^T$ with $H_{jv} := H_{1,j,v} = H_{2,j,v}$ as the magnetic field is conserved across the interfaces. Furthermore $\overline{\overline{A}}^{(u,v)}$ and $\overline{\overline{B}}^{(u,v)}$ are the matrices containing the elements $A_{i,j}^{(u,v)}$ and $B_{i,j}^{(u,v)}$, respectively. As the end of element i is the same as the start of element $i+1$, the $\overline{\overline{D}}$-matrices are given by

$$\overline{\overline{D}}_H = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 1 & 0 & 0 & \cdots & 0 \end{bmatrix}, \quad \overline{\overline{D}}_\varphi = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -1 & 0 & 0 & \cdots & 0 \end{bmatrix}, \tag{2.12}$$

and relates the fields in neighbouring elements as $\vec{H}^{(1)} = \overline{\overline{D}}_H \vec{H}^{(0)}$ and $\vec{\varphi}_u^{(1)} = \overline{\overline{D}}_\varphi \vec{\varphi}_u^{(0)}$. Here the two -1 in $\overline{\overline{D}}_\varphi$ are found at elements where $\theta_i = 0$ according to Fig. 2.1(c), which means that the normal vector changes sign [33].

2.1.2 From one to multiple grooves

While the theory outlined in the previous section focused on a scatterer of arbitrary shape, the properties of the finite array of grooves illustrated in Fig. 2.1(b) will be utilized here. To simplify the notation, the large matrix in Eq. (2.11) is denoted $\overline{\overline{M}} := \begin{bmatrix} \overline{\overline{M}}_{11} & \overline{\overline{M}}_{12} \\ \overline{\overline{M}}_{21} & \overline{\overline{M}}_{22} \end{bmatrix}$, where each block matrix has dimension $(NN_s) \times (NN_s)$, and Eq. (2.11) is written as $\overline{\overline{M}} \cdot \vec{x} = \vec{b}$, with $\vec{x} = [\vec{H}^{(0)}, \vec{\varphi}_1^{(0)}]^T$ and $\vec{b} = [\vec{H}_0, \vec{0}]^T$. Furthermore, each block matrix $\overline{\overline{M}}_{ab}$ where $\{a, b\} = \{1, 2\}$ can be further divided into submatrices $\overline{\overline{m}}_{ab}(i, j)$ of dimensions $N_s \times N_s$, describing all the interactions from one groove to another, where $i, j = 1, 2, \cdots, N$.

The Greens functions in Eq. (2.2) and (2.3) can be written as $g_i(x, x', y, y') = g_i(|x - x'|, y, y')$ where the y-dependence goes as $|y - y'|$ for the direct term and $y + y'$ for the indirect term. As the grooves are found along the x direction, this implies that the interaction between two grooves only depend on their relative position. However, when computing the integration kernels in Eq. (2.10), it is noticed that the observation point is chosen to be in the start of element i, thus at $\vec{s}_i^s$, which implies that there is a small difference between $A_{i,j}^{(u,v)}$ and $A_{j,i}^{(u,v)}$, and that the relative position between two grooves thereby needs to be calculated with a sign. The submatrices can then be simplified such that $\overline{\overline{m}}_{ab}(i, j) = \overline{\overline{m}}_{ab}(i - j)$, implying that each block matrix can

be written as

$$\overline{\overline{M}}_{ab} = \begin{bmatrix} \overline{\overline{m}}_{ab}(0) & \overline{\overline{m}}_{ab}(1) & \cdots & \overline{\overline{m}}_{ab}(N-1) \\ \overline{\overline{m}}_{ab}(-1) & \overline{\overline{m}}_{ab}(0) & \cdots & \overline{\overline{m}}_{ab}(N-2) \\ \vdots & \vdots & \vdots & \vdots \\ \overline{\overline{m}}_{ab}(-N+1) & \overline{\overline{m}}_{ab}(-N+2) & \cdots & \overline{\overline{m}}_{ab}(0) \end{bmatrix}, \quad \{a,b\} = \{1,2\}. \tag{2.13}$$

Such a block matrix contains N^2 submatrices, but importantly, it is seen that only $2N-1$ of them are different, which significantly reduces the problem of constructing $\overline{\overline{M}}$. Clearly a solution to $\overline{\overline{M}} \cdot \vec{x} = \vec{b}$ is simply $\vec{x} = \overline{\overline{M}}^{-1} \cdot \vec{b}$, which however needs the entire matrix $\overline{\overline{M}}$ and not only the $4(2N-1)$ *different* submatrices $\overline{\overline{m}}_{ab}(i-j)$. To handle large systems, the limiting parameter will often be the RAM rather than computational time [105], and a method to solve the integral equation without using a lot af RAM is to use an iterative method based on matrix-vector products rather than matrix inversion. These matrix-vector products are then computed solely from the $4(2N-1)$ different submatrices by simple row-column multiplication, implying that only the submatrices need to be stored, and not the entire matrix $\overline{\overline{M}}$.

An iterative method tends to minimize the residual $\vec{r} = \vec{b} - \overline{\overline{M}} \cdot \vec{x}$ where $\vec{x}_0$ is given as an initial guess for the solution. A search direction is chosen based on the residual, and is used to update the initial guess in every iteration. Hence the method generates a sequence of vectors, the corresponding residuals, and search directions until convergence is obtained when the norm of the residual is below some threshold. Several different iterative methods exist, where some of them depends on whether the matrix $\overline{\overline{M}}$ is symmetric, definite, or Hermitian. [106]. The following iterative methods have been tested in this work: The conjugate gradient (CG) algorithm, the Bi-conjugate gradient (BiCG), the Bi-conjugate gradient stabilized (BiCGSTAB), the generalized minimal residual (GMRES), the conjugate gradient squared method (CGS), and the quasi-minimal residual method (QMR). For a further description of the methods, see Ref. [106]. To solve the integral equation constructed for multiple grooves, it was found that the GMRES is the fastest. This is because more memory is built into that method, meaning that the search direction in one iteration is computed based on quantities for multiple previous iterations. When testing all the methods, the initial guess was chosen to be $\vec{x}_0 = \vec{b}$.

The ultrasharp groove illustrated in Fig. 2.1(a) is discretized into roughly 400 elements, where many elements are found near the bottom and corners. Hence for an array consisting of 40 grooves, the matrix $\overline{\overline{M}}$ has dimension 32.000×32.000, thus containing a billion (complex) numbers. However, for this structure it was still possible to solve the integral equation using matrix

inversion on a good stationary computer, and this was in fact much faster than solving it using GMRES.

Solving the integral equation gives the magnetic field and its normal derivative at the boundary of the scatterer. The scattered field and the cross sections, when the incident field is a plane wave ($\gamma = \infty$), are found using the method presented in appendix B in Ref. [29], while the method for computing the reflection from the groove array when the incident field is a Gaussian beam is found in the appendix in paper I. However, much of the physics regarding the optical properties of groove arrays can be explained in terms of gap plasmons propagating within the grooves, and calculating the reflectance within that formalism gives almost the same results as obtained using GFSIEM [28–30]. This will be utilized in paper III when calculating the reflectance from an ultrasharp groove array taking quantum effects into account.

2.2 Density-functional theory in the jellium model

Hohenberg and Kohn proved in 1964 the remarkable theorem that the ground state energy E of an N-electron system is simply a functional of the electron density, thus $E = E[n(\vec{r})]$ [41]. Thereby the problem is reduced from being $3N$-dimensional to become only 3-dimensional, as the density n only depends on the spatial coordinate $\vec{r}$. However, the Hohenberg-Kohn theorem does not give a prescription for calculating the density or the ground state energy, but these quantities can be determined when applying the variational theorem on the energy functional, resulting in the Kohn-Sham equations [42]

$$\left\{ -\frac{\hbar^2}{2m_e}\nabla^2 + V_C(\vec{r}) + V_{xc}(\vec{r}) \right\} \Phi_i(\vec{r}) = E_i \Phi_i(\vec{r}). \tag{2.14}$$

Here V_C and V_{xc} are the Coulomb and exchange-correlation potentials, respectively, and m_e is the electron mass. Eq. (2.14) is an eigenvalue problem, where the Φ_i's are the Kohn-Sham orbitals and E_i's the energies belonging to the Hamiltonian in the curly brackets. Having obtained the Φ_i's, the density is calculated as $n(\vec{r}) = 2\sum_i f_i |\Phi_i(\vec{r})|^2$ [42], where the factor of 2 is due to spin summation and f_i is the Fermi function at zero temperature. The principle in DFT is to solve the Kohn-Sham equations in Eq. (2.14) and afterwards calculate the electron density. In the next section, explicit expressions are obtained for the potential terms given as functionals of the density n.

The total potential is denoted $V(\vec{r}) = +V_C(\vec{r}) + V_{xc}(\vec{r})$, and must account for electron-electron interactions, which are important in metals due to the high electron density. However, the electrons in noble metals are highly delocalized with typical mean free paths of more than 50 times the lattice constant [1, 107], which implies that they do not experience the attractions from

the individual nuclei, but rather from a structure where the nuclear potential is smoothed out. Thus the ionic charge distribution is assumed to be uniform within the metal, which is accounted for in the jellium model [45, 46] as illustrated in Fig. 2.2. The positive charge density n^+ equals Zn_0, where Z is the number of valence electrons per atom, and n_0 is the atomic density. The latter can be written as $n_0 = 3/(4\pi r_s^3)$, where the Wigner-Seitz radius r_s is the radius of a sphere containing one atom [46]. Gold and silver both have one valence electron per atom, while r_s equals 3.01 Bohr for gold and 3.02 Bohr for silver [107].

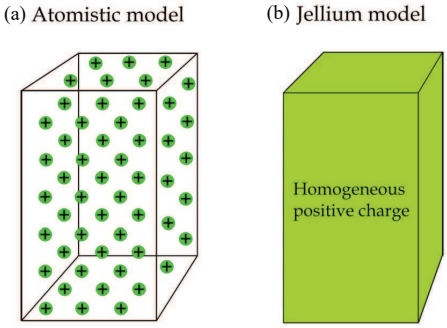

Fig. 2.2: (a) illustrates an atomistic model, where the positive charge is located at the discrete nuclei of the metal, while (b) illustrates the jellium model assuming that the charge is smoothed out across the metal.

The present work studies plasmons propagating in nanometer thin gaps and slabs as illustrated in Fig. 1.2(c,d). Choosing the z axis to be perpendicular to the structures, the total potential is only a function of z, thus $V(\vec{r}) = V(z)$. Thereby the Kohn-Sham orbitals in Eq. (2.14) become separable such that $\Phi_i(\vec{r}) = A^{-1/2}\varphi_j(z)e^{i\vec{k}_\parallel\cdot\vec{r}_\parallel}$ and $E_i = \varepsilon_j + \hbar^2 k_\parallel^2/(2m_e)$ [96]. Here A is a normalization area, while $i \in \{j, \vec{k}_\parallel\}$, $\vec{r}_\parallel = (x, y)$, and m is the electron mass. Furthermore, $\varphi_j(z)$ and ε_j satisfy the following Kohn-Sham equations, where the Hamiltonian and wave functions only depend on z

$$\hat{H}(z)\varphi_j(z) = \varepsilon_j\varphi_j(z), \text{ where } \hat{H}(z) = \frac{-\hbar^2}{2m_e}\frac{\mathrm{d}^2}{\mathrm{d}z^2} + V(z). \tag{2.15}$$

The electron density also only depends on z as

$$n(\vec{r}) = 2\sum_i f_i|\Phi_i(\vec{r})|^2 = \sum_{j\in occ}|\varphi_j(z)|^2\frac{m_e}{\pi\hbar^2}(E_F - \varepsilon_j) = n(z), \tag{2.16}$$

where the sum is over the occupied states, for which $\varepsilon_j < E_F$, where E_F is the Fermi energy. Hence, for the gap and slab structures considered in this

work, the principle in the DFT method is to solve the eigenvalue problem in Eq. (2.15) and calculate the electron density using Eq. (2.16).

The geometry is chosen such that the slab in paper IV is located from $z = -d/2$ to $z = d/2$, while the gap structure in paper III is constructed as two thick slabs, where the gap size is the distance between the slabs, and with the gap centered at $z = 0$. This implies that the ionic charge distributions introduced above become $n^+_{\text{slab}}(z) = n_0\theta(d/2 - |z|)$ and $n^+_{\text{gap}}(z) = n_0 - n^+_{\text{slab}}(z)$ for the slab and gap structure, respectively. Notice that the slab and gap geometries are just opposite.

Explicit expressions are needed for the different potential terms in order to solve the Kohn-Sham equations. Here the Coulomb potential is the simplest and is found by just solving the Poisson equation $\frac{\mathrm{d}^2}{\mathrm{d}z^2}V_C(z) = \rho(z)/\varepsilon_0$ [108], where $\rho(z) = n^+(z) - n(z)$ is the total charge computed as the positively charged jellium density $n^+(z)$ minus the negatively charged electron density $n(z)$. The sign has been reversed in the Poisson equation, and V_C is therefore the potential acting on the negatively charged electrons. Utilizing integration by parts, the solution is found to be

$$V_C[n(z)] = \frac{1}{\varepsilon_0}\int_{-\infty}^{z}(z - z')[n^+(z') - n(z')]\,\mathrm{d}z'. \tag{2.17}$$

In a symmetric structure, the Coulomb potential is symmetric in z, which will be utilized below when writing the Kohn-Sham equations on matrix form.

2.2.1 Local density approximation

While the expression for the Coulomb potential in Eq. (2.17) was simply obtained by solving the Poisson equation, a further simplification is needed in order to calculate the exchange-correlation potential. In quantum mechanics, the exchange operator is given by [43]

$$\hat{K}'_j\Phi(\vec{r}) = \int \Phi^*_j(\vec{r}\,')\Phi(\vec{r}\,')\frac{1}{|\vec{r} - \vec{r}\,'|}\,\mathrm{d}\vec{r}\,'\Phi_j(\vec{r}), \tag{2.18}$$

and is non-local in the sense that in order to evaluate $\hat{K}'_j\Phi(\vec{r})$ in some particular point $\vec{r}$, one must know $\Phi(\vec{r})$ at all other points $\vec{r}\,'$. In the local density approximation (LDA), it is assumed that the exchange-correlation energy of an inhomogeneous electron gas is the same as for a corresponding electron gas with constant density [44]. The exchange-correlation energy E_{xc} is then separated into individual exchange and correlation energies, such that $E_{xc}[n] = E_x[n] + E_c[n]$, and similar for the potential V_{xc}.

In the LDA, an analytical expression can be obtained for the exchange energy as a functional of n. Within the Hartree-Fock method [43], the exchange

energy per electron as function of wave number k can be shown to be

$$\varepsilon_x(k) = \frac{k_F}{2\pi^3}\left(1 + \frac{1-k^2/k_F^2}{2k/k_F}\ln\left|\frac{1+k/k_F}{1-k/k_F}\right|\right), \tag{2.19}$$

where $k_F = \sqrt{2m_eE_F}/\hbar$, and the integration over the angular part of $\vec{k}$ has been carried out. Integrating now over the radial part, the exchange energy per electron becomes

$$\varepsilon_x[n(\vec{r})] = \int_0^{k_F} k^2\varepsilon_x(k)\,\mathrm{d}k = -\frac{3}{4}\left(\frac{3n(\vec{r})}{\pi}\right)^{1/3}. \tag{2.20}$$

The mathematical details of the derivation can be found in Ref. [109]. The total exchange energy for all the electrons is $E_x[n(\vec{r})] = \int n(\vec{r})\varepsilon_x[n(\vec{r})]\,\mathrm{d}^3r$, from which the exchange potential is found as the functional derivative of $E_x[n]$, giving [42]

$$V_x[n(\vec{r})] = \frac{\partial E_x}{\partial[n(\vec{r})]} = -\left(\frac{3n(\vec{r})}{\pi}\right)^{1/3}. \tag{2.21}$$

It turns out that the Coulomb and exchange potentials are not sufficient to correctly describe the ground state energy of a quantum system. The remaining part is stored in a term called the correlation potential, which of course has no analytic expression [43]. However, the accurate energy of a homogeneous electron gas can be calculated using quantum Monte Carlo simulations [52], and the correlation energy can be fitted to the results. Different parametrizations exist for the correlation energy in the LDA, where the following Perdew-Zunger parametrization for the correlation energy per electron has been used in the present work [54]

$$\varepsilon_c[n(z)] = \begin{cases} A\ln(r_s(z)) + B + Cr_s(z)\ln(r_s(z)) + Dr_s(z), & r_s(z) \le 1 \\ \gamma/(1+\beta_1\sqrt{r_s(z)} + \beta_2 r_s(z)), & r_s(z) > 1, \end{cases} \tag{2.22}$$

where $r_s(z) = (3/(4\pi n(z)))^{1/3}$, and the coefficients are for a spin unpolarized electron gas as given in the table in Fig. 2.3(a). Here unpolarized means that the densities for electrons with spin up and spin down are the same [57]. Fig. 2.3(b) shows $\varepsilon_c[r_s]$ together with values obtained from Monte Carlo simulations for a homogeneous electron gas [53]. The data from Ref. [53] contains the total energy per electron, from which the kinetic and exchange energy has been subtracted in order to get the correlation energy. The exchange energy per electron is given by Eq. (2.20), while the kinetic energy per electron is $\varepsilon_{\text{kin}} = 3E_F/5 = 3^{5/3}\pi^{4/3}n^{2/3}/10$ [107]. The black dots in Fig. 2.3 thus denotes the correlation energy per electron obtained using Monte Carlo simulation, and the blue line obtained from the Perdew-Zunger parametrization

(a)

A	0.0311
B	-0.048
C	0.020
D	-0.0116
β_1	1.0529
β_2	0.03334
γ	-0.1423

(b)

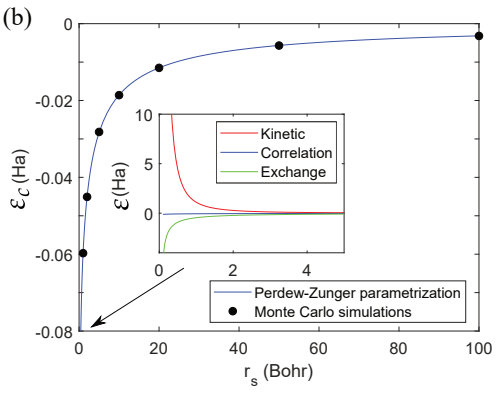

Fig. 2.3: (a) shows the coefficients used in the Perdew-Zunger parametrization in Eq. (2.22). The black dots in (b) show the correlation energy per electron obtained using Monte Carlo simulation from Ref. [53], while the blue line shows the corresponding Perdew-Zunger parametrization. The inset shows the kinetic, correlation, and exchange energy per electron for low r_s (high density).

in Eq. (2.22) is seen to perfectly fit the simulation. The inset in the figure shows the three different energy terms for small r_s, as illustrated by the arrow, where the correlation energy shown by the blue line is the same as in the outer figure. In the limit of low r_s (high density), the kinetic energy term is dominant, while the correlation term plays a minor role. The Coulomb potential vanish in the LDA as $n^+(z) = n(z)$, and is therefore not shown in Fig. 2.3(b).

The corresponding correlation potential is found as the functional derivative of $E_c[n(\vec{r})] = \int n(\vec{r})\varepsilon_c[n(\vec{r})]\,\mathrm{d}^3r$, yielding

$$V_c[n(z)] = \begin{cases} A\ln(r_s) + B - \frac{1}{3}A + \frac{2}{3}Cr_s\ln(r_s) + \left(\frac{2}{3}D - \frac{1}{3}C\right)r_s, & r_s \leq 1 \\ \gamma\dfrac{1 + \frac{7}{6}\beta_1\sqrt{r_s} + \frac{4}{3}\beta_2 r_s}{(1 + \beta_1\sqrt{r_s} + \beta_2 r_s)^2}, & r_s > 1, \end{cases} \tag{2.23}$$

where the argument of r_s has been omitted. The exchange and correlation potentials in Eq. (2.21) and (2.23) become symmetric in z, as they are given as functionals of a symmetric density $n(z)$. Hence the total potential is symmetric in z, which will be utilized in the next section, where the Kohn-Sham equations are written on matrix form.

2.2.2 The Kohn-Sham equations on matrix form

In this section, the Kohn-Sham equations in Eq. (2.15) are turned into a matrix eigenvalue problem. As the geometry is centered at $z = 0$, it is chosen

to expand the wave functions in the following orthonormal base

$$\varphi(z) = \sum_{j}^{N} c_j f_j(z), \text{ where } f_j(z) = \sqrt{\frac{2}{L}} \sin(k_j(z + L/2)). \tag{2.24}$$

Here $k_j = j\pi/L$, and the computational domain is $z \in [-L/2, L/2]$, where its length L being 2 nm larger than the slab width d is sufficient to model the quantum spill-out correctly. Multiplying Eq. (2.15) from the left by $\varphi(z)$ and integrating, the following matrix equation is obtained.

$$\overline{\overline{H}}\vec{c} = \varepsilon\vec{c} \tag{2.25}$$

where $\vec{c} = (c_1, c_2, \ldots, c_N)^T$ and the matrix elements in $\overline{\overline{H}}$ are given by

$$H_{ij} = \int_{-L/2}^{L/2} f_i(z)\hat{H}(z)f_j(z)\,\mathrm{d}z. \tag{2.26}$$

As the base $\{f_j\}$ is orthonormal, the kinetic energy part is described by the diagonal matrix $\overline{\overline{H}}_0 = \hbar^2/(2m_e)\mathrm{diag}(k_1^2, k_2^2, \ldots, k_N^2)$, while the matrix elements of the potential matrix $\overline{\overline{V}}$ are given by

$$V_{ij} = \begin{cases} 2\int_{-L/2}^{0} f_i(z)V(z)f_j(z)\,\mathrm{d}z, & \text{for } i+j \text{ even} \\ 0, & \text{for } i+j \text{ odd,} \end{cases} \tag{2.27}$$

where $V(z) = V_C(n(z)) + V_x(n(z)) + V_c(n(z))$ is the total potential, which is symmetric in z as utilized in Eq. (2.27). The matrix in Eq. (2.25) is then written as $\overline{\overline{H}} = \overline{\overline{H}}_0 + \overline{\overline{V}}$, and as this matrix is Hermitian, it is sufficient to calculate the integrals in Eq. (2.27) for $i \leq j$. The energy states ε and the vector $\vec{c}$ are thus found as the eigenvalues and eigenvectors of $\overline{\overline{H}}$, respectively. The total negative charge of the electrons must equal the total positive charge of the ions, thus $\int n(z)\,\mathrm{d}z = n_0 d$. Inserting the density from Eq. (2.16) into this equation, the following expression is obtained for the Fermi energy [101]

$$E_F = \frac{1}{M}\left(\frac{\pi\hbar^2 n_0 d}{m_e} + \sum_{j=1}^{M}\varepsilon_j\right), \tag{2.28}$$

where M is the highest occupied band. From the coefficients $\{c_j\}$, the wave functions are calculated from Eq. (2.24), and at last the density is calculated from Eq. (2.16) using the Fermi energy in Eq. (2.28).

However, the potential $V(z)$ appearing in Eq. (2.27) is only given as a functional of n. Thus, expressions involving the wave functions $\varphi(z)$ appear on both sides of Eq. (2.25), which implies that the equation has to be solved in a self-consistent way.

2.2.3 Self-consistency

The method for solving the Kohn-Sham equations self-consistently is first to apply an initial guess of the density, from which the total potential is computed. Then the eigenvalue problem in Eq. (2.25) is solved, and the output density is calculated from Eq. (2.16). Self-consistency is then obtained only if the input and output densities are equal [42]. Otherwise, an iterative procedure is started, where the output density can be used as the input density in the next iteration. However, this method fails as the effect of the exchange-correlation and Coulomb potential from the previous iteration is too strong. Instead, these effects can be reduced by applying a mixing of input and output densities in one iteration as the input density in the next iteration. The simplest form of mixing is just to use a linear combination of input and output densities in the following way [44]

$$n_{\text{in}}^{i+1}(z) = \alpha n_{\text{out}}^{i}(z) + (1-\alpha) n_{\text{in}}^{i}(z), \tag{2.29}$$

where $\alpha \in (0,1)$ is a mixing parameter, which must be below some threshold in order for the process to converge. A way to speed it up is to use information from earlier iterations when updating the density. Hence, the mixing parameter is not just constant, but is optimized in each iteration. The method is called Anderson mixing and does the following [44]

$$\begin{aligned} n_{\text{in}}^{i+1}(z) &= \alpha \hat{n}_{\text{out}}^{i}(z) + (1-\alpha)\hat{n}_{\text{in}}^{i}(z) \\ \hat{n}_{p}^{i}(z) &= \beta n_{p}^{i-1}(z) + (1-\beta) n_{p}^{i}(z), \text{ where } p = \{\text{in}, \text{out}\}. \end{aligned} \tag{2.30}$$

Here α is still a constant mixing parameter, but the parameter β is chosen such that $|n_{\text{in}}^{i} - n_{\text{out}}^{i}|^2$ is minimized, where $|h|^2$ means $\int h(z)^2 \, \mathrm{d}z$. After some algebra, it is found that

$$\beta = \frac{\langle \Delta n^i(z) | \Delta n^i(z) - \Delta n^{i-1}(z) \rangle}{|\Delta n^i(z) - \Delta n^{i-1}(z)|^2}, \tag{2.31}$$

where $\Delta n^i(z) = n_{\text{out}}^{i}(z) - n_{\text{in}}^{i}(z)$ and $\langle h | g \rangle = \int h(z) g(z) \, \mathrm{d}z$. In general, the threshold value for α depends strongly on the size of the system, where a larger system forces a smaller threshold. An example of how the threshold value depends on the slab thickness will be given in Sec. 3.2.

The iterative method for obtaining a self-consistent solution of the Kohn-Sham equations is illustrated by the flow chart in Fig. 2.4. Here, the mixing function f can be either the simple mixing from Eq. (2.29) or the Anderson mixing from Eq. (2.30). In practice, the check for self-consistency is performed based on the Fermi energy from Eq. (2.28) rather than the electron density. Self-consistency is said to be obtained when the difference in Fermi energy for two consecutive iterations is below 10^{-7} Ha.

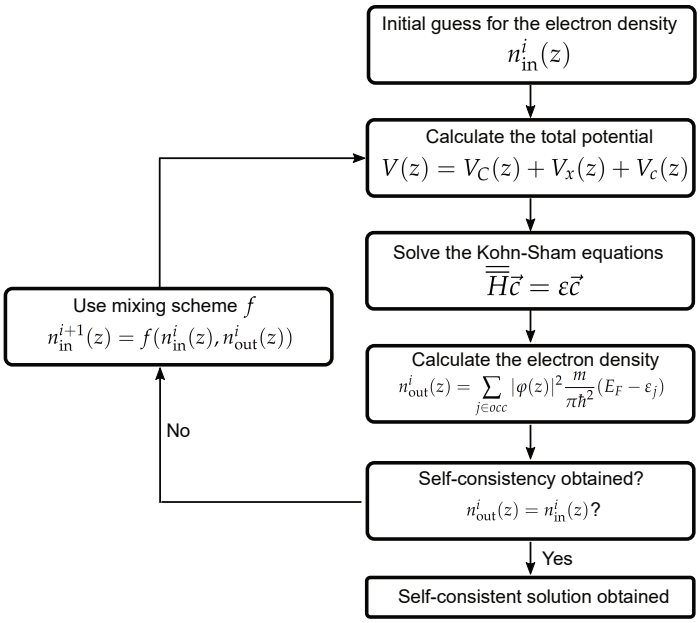

Fig. 2.4: Flowchart describing the iterative procedure for obtaining a self-consistent solution of the Kohn-Sham equations. The mixing function f can be the simple mixing from Eq. (2.29) or the Anderson mixing from Eq. (2.30). Figure inspired by Ref. [44].

2.3 Surface plasmon polaritons

In this section, the theory regarding surface plasmon polaritons (SPPs) is presented, which is an essential part of the thesis. First the basic concepts are explained based on the classical Drude model, while the model is extended afterwards to include quantum- and non-local effects.

2.3.1 Classical dielectric function

The dielectric function of noble metals is to a good approximation described by the Drude model $\varepsilon(\omega) = \varepsilon_b - \omega_p^2/(\omega^2 + i\Gamma\omega)$, where ε_b is the background dielectric function stemming from bound electrons in the lower lying d bands, ω_p is the plasma frequency, and Γ is a damping term [1]. For optical frequencies, the real part of ε is negative, while the imaginary part is positive, with $-\text{Re}(\varepsilon) > \text{Im}(\varepsilon)$, and the free electrons oscillate out of phase with the driving electric field. The quanta of these oscillations in charge density are called SPPs. The following subsections contain analyses of SPPs propagating in different structures: along a plane interface between a metal and a dielectric, in a gap between two metals, and in a metal slab sandwiched between dielectrics.

Plasmons at plane interfaces

The simplest physical system that supports SSPs is a plane interface between a metal and a dielectric [1], and this case is studied in detail in this subsection, as it lays the foundation for the gap and slab geometries to be considered below.

The reflection and transmission of light when incident on a plane interface is illustrated in Fig. 2.5(a,b) for p- and s-polarization, respectively, where the dielectric functions of the two media are ε_1 and $\varepsilon_2(\omega)$. Only the fields that

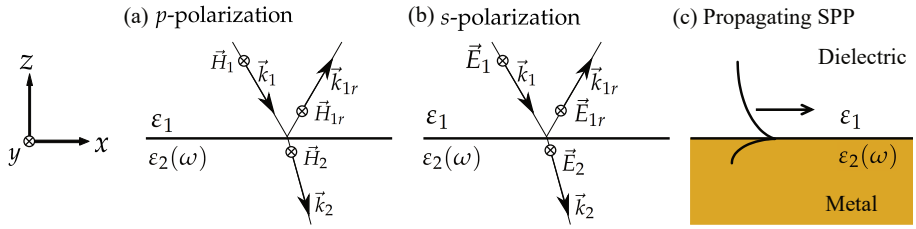

Fig. 2.5: (a,b) Reflection and transmission for p- and s-polarization. (c) The magnetic field of an SPP propagating along a plane interface between a metal and a dielectric. Figure inspired by Ref. [1].

are tangential to the interface are illustrated, which is $\vec{H}$ in (a) and $\vec{E}$ in (b). In both cases $\vec{E}$ is perpendicular to $\vec{H}$ and the propagation direction $\vec{k}$. For p-polarization in Fig. 2.5(a) the magnetic field can be written as $\vec{H}^p(\vec{r}) = \hat{y}H^p(x,z)$, where

$$H^p(x,z) = He^{ik_x x}\begin{cases} e^{-ik_{z,1}z} + re^{ik_{z,1}z}, & z > 0 \\ te^{-ik_{z,2}z}, & z \leq 0, \end{cases} \tag{2.32}$$

where $k_{z,j} = \sqrt{k_0^2\varepsilon_j - k_x^2}$, $j = 1,2$, and the corresponding electric field is found from Maxwells equations as $\vec{E}^p(x,z) = i\vec{\nabla} \times [\hat{y}H^p(x,z)]/(\omega\varepsilon_0\varepsilon(z))$. The coefficients r and t are found by imposing the boundary condition that the tangential component of the magnetic field and the normal component of the displacement field are conserved across the interface. The same can be done for s-polarization in Fig. 2.5(b), and the following Fresnel coefficients are obtained [1]

$$r_{12}^p(k_x) = \frac{\varepsilon_2 k_{z,1} - \varepsilon_1 k_{z,2}}{\varepsilon_2 k_{z,1} + \varepsilon_1 k_{z,2}}, \quad r_{12}^s(k_x) = \frac{k_{z,1} - k_{z,2}}{k_{z,1} + k_{z,2}}, \quad t_{12} = 1 + r_{12}. \tag{2.33}$$

An SPP is a purely bound mode, and for it to exist the fields must be exponentially decaying away from the interface. Choosing $k_{z,j} = i\delta_j$, where $\mathrm{Re}(\delta_j) > 0$, the field in Eq. (2.32) is clearly decaying for $z \leq 0$, while it is only

decaying for $z > 0$ if the reflected field is much larger than the incident field. This is obtained when the denominator in r is zero, thus an SPP is found as a pole in the reflection coefficient [33]. For the single interface in Fig. 2.5 this is obtained for p-polarization when $\varepsilon_2 k_{z,1} + \varepsilon_1 k_{z,2} = 0$, and by combining with $k_{x,j}^2 + k_{z,j}^2 = k_0^2 \varepsilon_j$, the following SPP dispersion relation is obtained relating the parallel wave number k_x and the frequency ω [1]

$$k_x = k_0 \sqrt{\frac{\varepsilon_1 \varepsilon_2(\omega)}{\varepsilon_1 + \varepsilon_2(\omega)}}. \tag{2.34}$$

The corresponding wave number in the z direction within the two media are given as $k_{z,j} = k_0 \sqrt{\varepsilon_j^2(\omega)/(\varepsilon_1 + \varepsilon_2(\omega)}$. For the SPP to be bound and propagate along the interface, k_x and $k_{z,j}$ have to be purely real and imaginary, respectively, which is fulfilled if both the sum and product of ε_1 and $\varepsilon_2(\omega)$ are negative. This condition is fulfilled if material 1 is a dielectric and material 2 is a metal at optical frequencies. Hence an SPP is found near the interface between a metal and a dielectric, where its magnetic field is illustrated in Fig. 2.5(c). For s-polarization, the denominator in the reflection coefficient in Eq. (2.33) can never be zero, which implies that SPPs only exist for p-polarization [33]. It is noticed that SPPs are eigenmodes of the system, implying that the associated electric field satisfies the homogeneous wave equation [1]

$$\vec{\nabla} \times \vec{\nabla} \times \vec{E}(\vec{r}, \omega) - k_0^2 \varepsilon(\vec{r}, \omega) \vec{E}(\vec{r}, \omega) = 0. \tag{2.35}$$

To get a deeper understanding of plasmons propagating along a plane interface, the dispersion relation from Eq. (2.34) is plotted in Fig. 2.6(a,b) using different models for $\varepsilon_2(\omega)$ while fixing $\varepsilon_1 = 1$. In gold the plasma frequency is $\hbar\omega_p = 9.06$ eV, while the damping term is $\hbar\gamma = 71$ meV [88]. For gold described by the Drude dielectric function with constant background of $\varepsilon_b = 9$, the dispersion relation is shown in Fig. 2.6(a), where the light line $\omega = ck_x$ is shown by the dotted line. The dispersion relation clearly shows two branches, called the low-energy and high-energy branch, respectively. The real part of the metal dielectric function becomes positive for $\omega > \omega_b = \omega_p/\sqrt{\varepsilon_b}$, and above this frequency the field propagates through the metal instead of being bound to the interface. Hence no true SPPs are found in this high-energy region. In the low-energy region, the metal dielectric function is negative and true SPPs exist. The dispersion curve converges for large k_x to $\omega_s = \omega_p/\sqrt{\varepsilon_b + \varepsilon_1}$, at which frequency the denominator in Eq. (2.34) becomes zero, after which it bends back into the high-energy branch. The limiting frequencies in gold are $\hbar\omega_b = 2.84$ eV and $\hbar\omega_s = 2.71$ eV, and these are called the bulk plasmon frequency and the surface plasmon frequency, respectively.

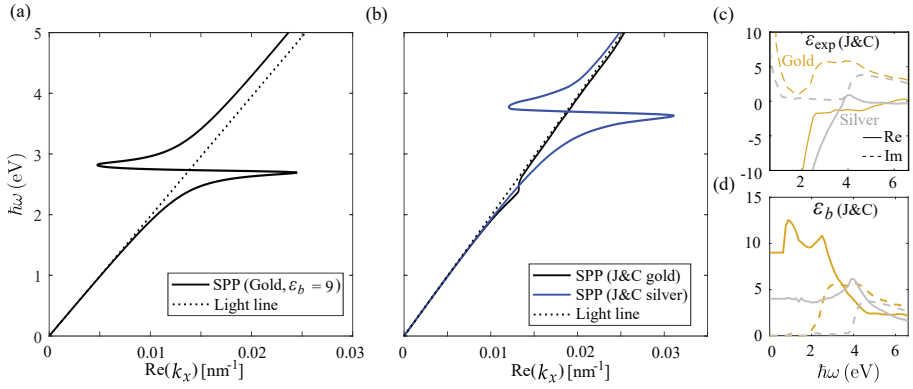

Fig. 2.6: (a,b) show the dispersion relation from Eq. (2.34) with $\varepsilon_1 = 1$ and the light line $\omega = ck_x$ depicted by the dotted line. In (a) the dielectric function of the metal is described by the Drude model for gold with $\hbar\omega_p = 9.06$ eV, $\hbar\gamma = 71$ meV, and ε_b=9. The experimental values of the metal dielectric functions from Johnson & Christy (J&C, Ref. 88) are shown in (c) and have been applied in (b). The corresponding background dielectric constants ε_b are shown in (d).

The experimental dielectric constants of gold and silver measured by Johnson & Christy (J&C, Ref. [88]) are shown in Fig. 2.6(c), where the solid and dashed lines denote the real and imaginary parts, respectively. The corresponding background dielectric constants ε_b are obtained by subtracting the Drude part from the experimental values, and Fig. 2.6(d) shows $\varepsilon_b(\omega)$, where the applied Drude parameters for silver are $\hbar\omega_p = 9.17$ eV and $\hbar\gamma = 21$ meV [88]. No measurements from J&C are found for $\hbar\omega < 0.6$ eV, and ε_b in this region is approximated by a constant value taken from $\hbar\omega = 0.6$ eV as seen in Fig. 2.6(d). Using the experimental value from J&C, the dispersion relation for gold is shown by the black line in Fig. 2.6(b). The reason that the dispersion does not bends back into the high-energy branch, as in Fig. 2.6(a), is that ε_b for gold has an imaginary part for $\hbar\omega > 2$ eV, as seen in Fig. 2.6(d), which implies that the bending is suppressed by damping. The blue curve in Fig. 2.6(b) shows the dispersion relation based on the experimental values for silver, which has the same behaviour as Fig. 2.6(a). This is due to the fact that the imaginary part of ε_b in silver is small near the limiting frequencies, which in silver takes the values $\hbar\omega_b = 3.78$ eV and $\hbar\omega_s = 3.63$ eV, respectively, as seen in the figure. Below the surface plasmon frequency, the dispersion curve is at the right of the light line, implying that SPPs cannot simply be excited by incident light of any frequency from free space. Instead the configurations illustrated in Fig. 1.1 can be applied to excite the SPPs.

Plasmons in three-layer structures

Plasmons propagating in layered structures is an essential part of this thesis, and the basic theory of the topic is explained in this section, where the structure consists of 3 layers.

Two different three-layer structures were mentioned in the introduction and illustrated in Fig. 1.2(c,d) as a gap and slab structure, respectively. For convenience, they are illustrated again in Fig. 2.7 with more details. As ex-

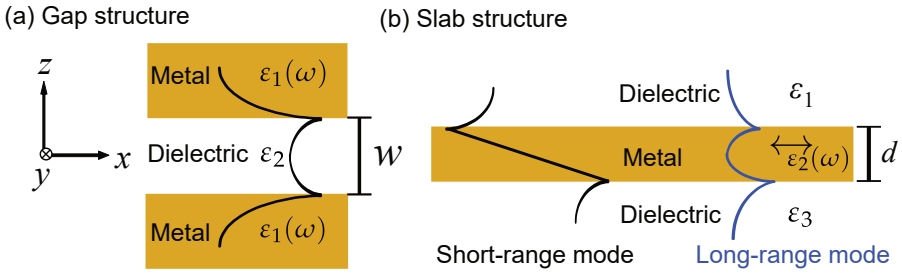

Fig. 2.7: Plasmons propagating in three-layer structures, where the curves denote the tangential magnetic fields. The slab structure in (b) supports a short-range and a long-range mode, as illustrated by the black and blue curve, respectively.

plained in the previous section, plasmons exist only for p-polarization where the magnetic field is tangential to the interfaces, and for a three-layer structure, the field in Eq. (2.32) is generalized to

$$H^p(x,z) = e^{ik_0\beta x}H(z). \tag{2.36}$$

Here $\beta = k_x/k_0$ has been introduced as a dimensionless mode index, and $H(z)$ is the transverse magnetic field illustrated by the curves in Fig. 2.7 to be further elaborated upon below. The structures are considered as invariant in the x- and y-directions, and they are also called metal-dielectric-metal (MDM) and dielectric-metal-dielectric (DMD) structures. As in the case with plane interface studied above, the dielectrics are described by constant dielectric functions, while the dielectric function of the metals depend on frequency. In addition, the metal dielectric function in the slab structure is in general anisotropic, as it only has boundaries in the z-direction, implying that the dielectric function becomes a tensor. The metal regions in the gap structure are semi-infinite, why their response is isotropic.

As for the case with a plane interface, the SPP solution is found as a pole in the reflection coefficient for p-polarization. A layered structure can be described by a structure matrix $\overline{\overline{S}}$, which relates the magnetic field to the left of the structure to that at the right of the structure. $\overline{\overline{S}}$ consists of interface

matrices describing reflection and transmission of light across the interfaces, and of propagation matrices describing the propagation of light within each layer. The two kind of matrices are given by [110]

$$\overline{\overline{T}}_{ij} = \frac{1}{t_{ij}} \begin{pmatrix} 1 & r_{ij} \\ r_{ij} & 1 \end{pmatrix}, \quad \overline{\overline{T}}_{j} = \begin{pmatrix} e^{-ik_{z,j}d_j} & 0 \\ 0 & e^{ik_{z,j}d_j} \end{pmatrix}, \tag{2.37}$$

where d_j is the thickness of layer j while r_{ij} and t_{ij} are the reflection and transmission coefficients for p-polarization in anisotropic materials given by [111]

$$r_{ij} = \frac{k_{z,i}\varepsilon_{\parallel,j} - k_{z,j}\varepsilon_{\parallel,i}}{k_{z,i}\varepsilon_{\parallel,j} + k_{z,j}\varepsilon_{\parallel,i}}, \quad t_{ij} = 1 + r_{ij}, \tag{2.38}$$

where $k_{z,i} = \sqrt{k_0^2\varepsilon\|, i - k_x^2\varepsilon_{\parallel,i}/\varepsilon_{\perp,i}}$ with $\mathrm{Im}(k_{z,i}) \geq 0$. It is seen that the original Fresnel coefficients in Eq. (2.33) are restored if the materials are isotropic. The total reflection coefficient across the structure is given as $r = S_{21}/S_{11}$. For three layers the structure matrix is given by $\overline{\overline{S}} = \overline{\overline{T}}_{12}\overline{\overline{T}}_{2}\overline{\overline{T}}_{23}$, and the reflection coefficient becomes

$$r = \frac{r_{12} + r_{23}e^{2ik_{z,2}d}}{1 + r_{12}r_{23}e^{2ik_{z,2}d}}, \tag{2.39}$$

where d is the thickness of the middle layer. Plasmons are found when the denominator in r is zero, and after some algebra the plasmon dispersion relation becomes

$$\left(1 + \frac{\varepsilon_{\parallel,2}k_{z,1}}{\varepsilon_{\parallel,1}k_{z,2}}\right)\left(1 + \frac{\varepsilon_{\parallel,2}k_{z,3}}{\varepsilon_{\parallel,1}k_{z,3}}\right) = \left(1 - \frac{\varepsilon_{\parallel,2}k_{z,1}}{\varepsilon_{\parallel,1}k_{z,2}}\right)\left(1 - \frac{\varepsilon_{\parallel,2}k_{z,3}}{\varepsilon_{\parallel,3}k_{z,2}}\right)e^{2ik_{z,2}d}. \tag{2.40}$$

It is noticed that this equation is valid for both the gap and slab geometry illustrated in Fig. 2.7. Unlike the case with a single planar interface, the equation can not be solved analytically, and is instead solved with the Newton-Raphson method. In the limit of $d \to \infty$, Eq. (2.40) simplifies to $\varepsilon_{\parallel,2}k_{z,1} + \varepsilon_1 k_{x,2} = 0$ and/or $\varepsilon_{\parallel,2}k_{z,3} + \varepsilon_3 k_{x,2} = 0$, which describe plasmons bound to the two individual interfaces, as these are decoupled in this limit. But for smaller d the plasmons bound to the individual interfaces overlap, and hybridize into two kinds of modes whose electric and magnetic fields have the opposite symmetry. If the structure is symmetric, $\varepsilon_1 = \varepsilon_3$, the two modes satisfy the following dispersion relations

$$\text{long-range} : \tanh(-ik_{z,2}d/2) = -\frac{\varepsilon_{\parallel,2}k_{z,1}}{\varepsilon_1 k_{z,2}}, \tag{2.41}$$

$$\text{short-range} : \coth(-ik_{z,2}d/2) = -\frac{\varepsilon_{\parallel,2}k_{z,1}}{\varepsilon_1 k_{z,2}}, \tag{2.42}$$

where the names will be explained below. It is clearly seen that the two modes coincide when $d \to \infty$ as $\tanh(d) = \coth(d)$=1 in this limit. But in the other limit, i.e when $d \to 0$, they behave very different. This is illustrated in Fig. 2.8(a) showing the solutions of Eq. (2.41) and (2.42) in the low-energy branch for the slab structure, where $\varepsilon_1 = \varepsilon_3 = 1$ and $\varepsilon_{\parallel,2}$ is the measured dielectric constant of silver from J&C. In the figure, the light line is shown

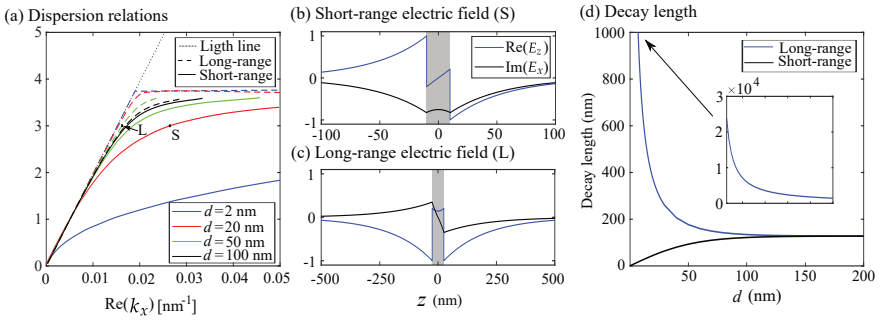

Fig. 2.8: (a) Dispersion relations of SPPs propagating in a silver slab of thickness d surrounded by air. (b,c) show the electric field components of the short- and long-range mode for slab widths of 20 and 50 nm, respectively. The fields are computed at the k_x and $\hbar\omega = 3$ eV illustrated by the dots in (a). The decay lengths are shown in (d) as a function of slab width at $\hbar\omega = 3$ eV.

by the dotted line while the long- and short-range modes are shown by the dashed and solid lines, respectively, which are almost the same when $d = 100$ nm. In this case, the two modes behave almost as two decoupled plasmons bound to the individual interfaces, and the dispersion relations are almost the same as the red line in the low-energy branch in Fig. 2.6(b), converging to $\hbar\omega_s$ before it bends back into the high-energy branch. For smaller d, one mode is close to the light line, implying that it has a long decay length into the surrounding air, why it is called a long-range mode. The other mode has a much larger k_x, implying that it decays rapidly into the air, why it is called a short-range mode. As mentioned, the high-energy branch describes propagation through the metal and not true SPPs bound to the surface, and for sufficiently thin films, this branch cease to exist. Instead the long-range mode converges from above to $\hbar\omega_s$ for large k_x, while the short-range mode converges to the same value from below.

The magnetic fields of the two modes were illustrated in Fig. 2.7(b), and have the opposite symmetry. The corresponding electric fields for slab widths of 20 and 50 nm are shown in Fig. 2.8(b,c) for the short- and long-range mode, respectively. The figures show the x- and z-components of the electric fields, which are tangential and normal to the interfaces, respectively. The tangential component is conserved across the interfaces, while the other component

jumps in order for the normal component of the displacement field to be conserved [1]. For both modes, the fields have been calculated at an energy of $\hbar\omega$ =3 eV, resulting in k_x of 0.0155 nm^{-1} and 0.027 nm^{-1}, as illustrated by the dots denoted L and S in Fig. 2.8(a). In the air regions surrounding the slab, the wave number in the z-direction is computed as $k_z = \sqrt{k_0^2\varepsilon_1 - k_x^2}$, from which the decay length is found as $1/\mathrm{Im}(k_z)$. For the two modes, the decay lengths are found to be 45 nm and 175 nm, as can be verified from the fields in Fig. 2.8(b,c). The decay lengths of the two modes as a function of slab width is shown in Fig. 2.8(d) at fixed $\hbar\omega = 3$ eV. For thin slabs the decay lengths are very different, and it becomes clear that they are indeed long-range and short-range, while they become practically the same for a 200 nm thick slab, as in this case the plasmons behave as bound to a single interface between silver and air.

For the structure consisting of a dielectric gap located between two parallel metal surfaces, only the mode whose magnetic field profile is illustrated in Fig. 2.7(a) will be considered. The behaviour and propagation of this mode is the main topic of paper III.

2.3.2 Quantum dielectric function

Having now explained the basic concepts of plasmons propagating in different structures based on a classical description, quantum effects are incorporated in this section.

In the classical Drude model presented in the previous section, the dielectric function of the metal is only a function of the frequency, thus with no spatial dependence. Hence the dielectric function across an interface between metal and dielectric becomes a step function, taking one value in each layer. When including quantum effects, the electron density becomes a smooth function across an interface between metal and dielectric (to be shown in Fig. 3.4(a)). Furthermore, the corresponding quantum dielectric function is non-local, and is for a structure that only varies in the z direction (like the gap and slab structures illustrated in Fig. 2.7), within the random-phase approximation (RPA) given by [96]

$$\overleftrightarrow{\varepsilon}(\vec{Q}, z, z', \omega) = \left(1 - \frac{e^2 n(z)}{m_e \varepsilon_0(\omega^2 + i\Gamma\omega)}\right)\delta(z - z')\overleftrightarrow{I} - \frac{1}{2\pi^2\varepsilon_0\hbar\omega^2}\sum_{n,m}\int \frac{f_m(|\vec{k}_\parallel - \vec{Q}|) - f_n(|\vec{k}_\parallel|)}{\omega + i\Gamma + \omega_m - \omega_n}\vec{j}_{nm}(\vec{k}_\parallel - \vec{Q}, z)\vec{j}_{mn}(\vec{k}_\parallel - \vec{Q}, z')\,\mathrm{d}^2k_\parallel. \tag{2.43}$$

The first line describes a local isotropic Drude term, where the spatially varying electron density $n(z)$ corresponds to the plasma frequency $\omega_p^2(z) =$

$e^2 n(z)/(m_e \varepsilon_0)$, where e is the electron charge. The second line is a nonlocal contribution, where n and m denote the quantum numbers. Here $f_n(|\vec{k}_\parallel|) = \theta(E_F - \hbar\omega_n - \hbar^2|\vec{k}_\parallel|^2/2m_e)$, is a Fermi factor at zero temperature, while $\vec{k}_\parallel$ and $\vec{Q}$ are parallel to the structure, thus in the (x, y) plane. Furthermore, the $\vec{j}_{nm}$'s are matrix elements of the transition current given by [96]

$$\vec{j}_{nm}(\vec{k}_\parallel, \vec{k}'_\parallel, z) = \frac{ie\hbar}{2m_e}\left(i(\vec{k}_\parallel + \vec{k}'_\parallel)A_{nm}(z) + \hat{z}B_{nm}(z)\right), \text{ where} \tag{2.44}$$

$$A_{nm}(z) = \varphi_n(z)\varphi^*_m(z), \quad B_{nm}(z) = \varphi^*_m(z)\frac{\partial\varphi_n(z)}{\partial z} - \varphi_n(z)\frac{\partial\varphi^*_m(z)}{\partial z}. \tag{2.45}$$

Here the $\varphi_n(z)$ are the eigenstates of the Hamiltonian $\hat{H}(z)$ from Eq. (2.15) with corresponding eigenvalues $\varepsilon_n = \hbar\omega_n$. Hence the quantum dielectric function depends on ε_n and φ_n which are calculated using DFT as explained in Sec. 2.2. Eq. (2.43) shows that rigid quantum mechanical response theory implies that the metal dielectric function consists of a local isotropic Drude term plus a nonlocal anisotropic term. It has been found in several papers [77–79,84,87,112], that for metal dimers and cylinders the local Drude term dominates the response, while the nonlocal term only slightly blueshift the plasmon resonances, implying that nonlocality is a relatively small modification to the dielectric response [76]. The nonlocal term is neglected in paper III, while paper IV neglects the parallel part of the nonlocality by setting $\vec{q}_\parallel = \vec{0}$, and furthermore use an averaging procedure for the perpendicular part, which implies that the nonlocal term becomes a local anisotropic step function. A full nonlocal model is applied in paper V.

In general, the dielectric tensor is given by $\overleftrightarrow{\varepsilon} = \text{diag}(\varepsilon_\parallel, \varepsilon_\parallel, \varepsilon_\perp)$, and becomes local in both paper III and IV. It is now modified to contain the effects of substrates and interband transitions. The latter is due to the bound electrons in the lower lying d-bands of the atom, which also contribute to the dielectric function [1]. However, in contrast to the free electrons described by the spatial varying density $n(z)$, it is assumed that the density of the bound electrons is constant within the jellium region and zero outside. Thus the contribution from the bound electrons is described by a step function, whose value is given as $\varepsilon_{\text{bound}}(\omega) = \varepsilon_{\text{metal}}(\omega) - \varepsilon_{\text{bulk}}(\omega)$, where $\varepsilon_{\text{metal}}(\omega)$ is the measured dielectric constant from J&C, and $\varepsilon_{\text{bulk}}(\omega) = 1 - e^2 n_0/(m_e\varepsilon_0(\omega^2 + i\Gamma\omega))$ is the Drude dielectric constant of bulk gold. The same treatment of the bound electrons have been applied in several papers [5, 76, 78, 100–103]. In paper V, the response from the bound electrons is included by a screening function to be shown explicitly below. The shape of the bound electron step function is opposite for the gap and slab structures due to their opposite geometries. Furthermore, different dielectrics can be found as substrates and superstrates in the slab structure, and this term

is described by a step function whose shape is the same as for the bound electron term in the gap structure. In total, the parallel part of the dielectric function for the gap and slab structure can thus be written as

$$\varepsilon_{\parallel}(\omega, z) = 1 - \frac{\omega_p^2(z)}{\omega^2 + i\Gamma\omega} + \varepsilon_{\text{bound}}(\omega)h(z) + g(z), \tag{2.46}$$

where

$$\begin{cases} h(z) = \theta(d/2 - |z|), & g(z) = (\varepsilon_s(z) - 1)\theta(|z| - d/2) \quad \text{for the slab structure} \\ h(z) = \theta(|z| - w/2), & g(z) = 0 \quad \text{for the gap structure.} \end{cases}$$

The gap and slab width w and d were illustrated in Fig. 2.7, and $\varepsilon_s(z)$ is the dielectric function of the substrate and superstrate. As mentioned in the previous section, the metal regions in the gap structure are modelled as semi-infinite and thereby isotropic, thus $\varepsilon_{\perp}^{\text{gap}}(\omega, z) = \varepsilon_{\parallel}^{\text{gap}}(\omega, z)$, while the response of the metal slab is anisotropic due to its confinement only in the z direction. The perpendicular part of the dielectric function for the slab is given by

$$\varepsilon_{\perp}^{\text{slab}}(\omega, z) = \varepsilon_{\parallel}^{\text{slab}}(\omega, z) + \varepsilon_{\text{ani}}(\omega, d)\theta(d/2 - |z|), \text{ with} \tag{2.47}$$

$$\varepsilon_{\text{ani}}(\omega, d) = \frac{e^2}{\pi d \varepsilon_0 \hbar^2 (\omega + i\Gamma)^2} \sum_{m,n} g_{mn} \frac{E_{mn}^2 (E_F - E_n)\theta(E_F - E_n)}{E_{mn}^2 - \hbar^2(\omega + i\Gamma)^2}, \tag{2.48}$$

where $g_{mn} = 2m_e |\langle \varphi_m | z | \varphi_n \rangle|^2 (E_m - E_n)/\hbar^2$ is the oscillator strength. A derivation of the expression for $\varepsilon_{\text{ani}}(\omega, d)$ can be found in Appendix B in paper IV.

Paper V includes the effect from the bound electrons as a screened Coulomb potential, which in free space (no screening) equals $V(\vec{r}, \vec{r}\,') = 1/|\vec{r} - \vec{r}\,'|$. Fourier transforming along the parallel direction gives $W(\vec{Q}, z, z') = \int V(\vec{r}, \vec{r}\,') e^{-i\vec{Q}\cdot(\vec{R} - \vec{R}\,')} \, \mathrm{d}\vec{R} \, \mathrm{d}\vec{R}\,'$ which becomes [5] $W(\vec{Q}, z, z') = 2\pi/Q e^{-Q|z - z'|}$, thus only depending of the norm of $\vec{Q}$ denoted Q. When including the bound electrons as a background dielectric function, the screening function W consists of a direct term and an indirect term describing the reflections within the background, thus $W = W^D + W^I$. For the slab geometry in Fig. 2.7, the direct term is given by

$$W^D(Q, z, z') = \frac{2\pi}{Q} e^{-Q|z - z'|} \begin{cases} 1/\varepsilon_1, & z, z' > d/2 \\ 1/\varepsilon_b, & -d/2 \leq z, z' \leq d/2 \\ 1/\varepsilon_3, & z, z' < -d/2 \\ 0, & \text{otherwise,} \end{cases} \tag{2.49}$$

where ε_b is the metal background dielectric function from Fig. 2.6(d). The

indirect term can be generalized from Ref. [101] and becomes

$$W^I(Q,z,z') = \frac{2\pi/Q}{(\varepsilon_1+\varepsilon_b)(\varepsilon_b+\varepsilon_3)-(\varepsilon_b-\varepsilon_1)(\varepsilon_b-\varepsilon_3)e^{-2Qd}} \times$$
$$\begin{cases} 1/\varepsilon_1 \left[(\varepsilon_1-\varepsilon_b)(\varepsilon_b+\varepsilon_3)e^{Qd}+(\varepsilon_1+\varepsilon_b)(\varepsilon_b-\varepsilon_3)e^{-Qd}\right] e^{-Q(z+z')}, & z,z' > d/2 \\ 2\left[(\varepsilon_b+\varepsilon_3)e^{-Q(z-z')}+(\varepsilon_b-\varepsilon_3)e^{-Q(z+z'+d)}\right], & -d/2 \le z' \le d/2 < z \\ 4\varepsilon_b e^{-Q(z-z')}, & z' < -d/2, z > d/2 \\ 2\left[(\varepsilon_b+\varepsilon_3)e^{Q(z-z')}+(\varepsilon_b-\varepsilon_3)e^{-Q(z+z'+d)}\right], & -d/2 \le z \le d/2 < z' \\ 1/\varepsilon_b\{(\varepsilon_b-\varepsilon_1)(\varepsilon_b-\varepsilon_3)\left[e^{-Q(z-z'+2d)}+e^{-Q(-z+z'+2d)}\right]+ \\ \quad (\varepsilon_1+\varepsilon_b)(\varepsilon_b-\varepsilon_3)e^{-Q(z+z'+d)}+(\varepsilon_b+\varepsilon_3)(\varepsilon_b-\varepsilon_1)e^{-Q(-z-z'+d)}\} & -d/2 \le z,z' \le d/2 \\ 2\left[(\varepsilon_1+\varepsilon_b)e^{-Q(z-z')}+(\varepsilon_b-\varepsilon_1)e^{Q(z+z'-d)}\right], & z' < -d/2 \le z \le d/2 \\ 4\varepsilon_b e^{Q(z-z')}, & z < -d/2, z' > d/2 \\ 2\left[(\varepsilon_1+\varepsilon_b)e^{Q(z-z')}+(\varepsilon_b-\varepsilon_1)e^{Q(z+z'-d)}\right], & z < -d/2 \le z' \le d/2 \\ 1/\varepsilon_3 \left[(\varepsilon_3-\varepsilon_b)(\varepsilon_1+\varepsilon_b)e^{Qd}+(\varepsilon_b+\varepsilon_3)(\varepsilon_b-\varepsilon_1)e^{-Qd}\right] e^{Q(z+z')}, & z,z' < -d/2. \end{cases} \tag{2.50}$$

It is constructed such that the potential and the normal component of the displacement field is conserved across the interfaces in $z = \pm d/2$ [1].

Calculating the plasmon modes

In order to calculate the plasmon modes of a structure described by a quantum dielectric function, the structure of interest is discretized along the z direction as illustrated in Fig. 2.9. The dielectric function is modelled as con-

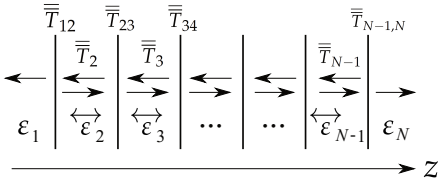

Fig. 2.9: Discretization of the dielectric function $\overleftrightarrow{\varepsilon}(z)$ into N layers along z. The matrices $\overline{\overline{T}}_j$ describe propagation within each layer, while the matrices $\overline{\overline{T}}_{ij}$ describe the reflection and transmission across each interface, and the arrows illustrate the direction of light propagation.

stant within each layer j as indicated by the dielectric constants $\overleftrightarrow{\varepsilon}_j$, where scalar values are found in the layers 1 and N, as these materials are modelled as semi-infinite and thereby isotropic. However, all the dielectric constants are modelled as isotropic for the gap structure. Within a layer, the electric and

magnetic fields consist of waves propagating left and right as illustrated by the arrows. As plasmons are bound to the structure, the fields are decaying away from the structure, why only out-going waves are found in layer 1 and N. The matrices $\overline{\overline{T}}_j$ describe propagation within each layer, while the matrices $\overline{\overline{T}}_{ij}$ describe the reflection and transmission across each interface, where expressions for the matrices were given in Eq. (2.37).

When using a classical dielectric function as in Sec. 2.3.1, the structure consists of only three layers, which implies that the reflection coefficient is obtained analytically as in Eq. (2.39). But for the quantum dielectric function, the structure typically needs to be discretized into more than 1000 layers, and the reflection coefficient is therefore calculated by a transfer matrix method. The structure matrix relating the magnetic field to the left of the structure to that at the right of the structure is given by

$$\overline{\overline{S}} = \prod_{j=1}^{N-2} (\overline{\overline{T}}_{j,j+1} \overline{\overline{T}}_{j+1}) \overline{\overline{T}}_{N-1,N} := \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}, \tag{2.51}$$

from which the reflection coefficient is found as $r = S_{21}/S_{11}$, and the plasmons are found as poles in r, thus when $S_{11} = 0$. The same condition is obtained when considering the fields which are only out-going left and right of the slab, where the magnetic field left of the structure is given by

$$\begin{pmatrix} 0 \\ H_L^- \end{pmatrix} = \overline{\overline{S}} \begin{pmatrix} H_R^+ \\ 0 \end{pmatrix}, \tag{2.52}$$

where + and - denote the direction in which the light propagates, while R and L denote right and left of structure, respectively. From the equation, it is clearly seen that $S_{11} = 0$. Furthermore, $H_L^- = S_{21} H_R^+$, and the symmetry of the mode can thus be determined from the sign of S_{21}, where positive and negative signs correspond to long- and short-range modes in the slab, respectively. S_{21} becomes exactly ± 1 in symmetric structures ($\varepsilon_1 = \varepsilon_3$) and the magnetic field of the modes becomes symmetric and antisymmetric, respectively, which will be illustrated in Sec. 3.3. It is noticed that for a structure consisting of three layers, the condition $S_{11} = 0$ is equivalent to Eq. (2.40).

The matrices $\overline{\overline{T}}_j$ and $\overline{\overline{T}}_{i,j}$ in Eq. (2.51) depend on k_x through Eq. (2.38), which implies that S_{11} is a function of k_x and thereby of $\beta = k_x/k_0$. By introducing the function $f : \mathbb{C} \to \mathbb{C}$ with $f(\beta) = S_{11}(\beta)$, the plasmon mode indices are found as the roots in f. The numerical procedure of derermining the roots is first to evaluate f for a range of β in the complex plane. When sign changes are observed for both the real and imaginary part of f, a region in $\mathbb{C}$ is identified where f is close to zero, and the Newton-Raphson method is afterwards applied to obtain the exact root.

An example of the numerical procedure is shown in Fig. 2.10 for a gold slab of width $d = 50$ Bohr surrounded by air at a wavelength of 775 nm ($\hbar\omega =$

1.6 eV), where the long- and short-range mode indices are shown in (a,b), respectively. Here the blue and red curves are contour curves illustrating the

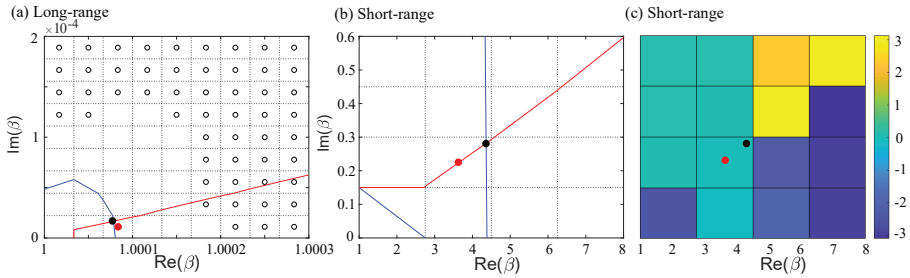

Fig. 2.10: The numerical procedure to obtain the plasmonic modes for a gold slab of width $d = 50$ Bohr surrounded by air at a wavelength of 775 nm. In (a,b) the blue and red curves are contour curves illustrating where the real and imaginary parts of f is zero. Starting the Newton-Raphson method in the points marked with asterisks lead to the solutions marked with a square. If the Newton-Raphson method is instead started at some of the points marked with a circle in (a), it converges to the solution in (b). A colormap of the phase of f is shown in (c), where the colors correspond to the phase value in the bottom left corner of the rectangles.

roots in the real and imaginary parts of f, respectively, and the proper roots are thus found where the two curves intersect. The dotted lines divide the complex plane into smaller rectangles, where f is evaluated in each corner, and thus changes sign in a specific rectangle only if both the blue and red contour curves go through it. A root of f thus exists in a certain rectangle only when this condition is fulfilled, and the Newton-Raphson method is started where the initial guess is in the center of this rectangle. The initial guess is illustrated by the red circles in the figures, and it converges to the exact root illustrated by the black filled circles in 4 iterations.

It is noticed that a very fine grid in the complex plane is necessary to find the solution of the long-range mode in Fig. 2.10(a), and that this solution is very close to 1. In fact, if the initial guess in the Newton-Raphson method is instead in the center of some other rectangle, the iterative method will instead converge to the short-range solution in Fig. 2.10(b), which is located relatively far from the long-range solution. The rectangles where this is the case are marked with black hollow circles, illustrating that it is necesary to sample f in a very narrow and exact region of the complex plane in order to find the solution for the long-range mode. The short-range mode is on the other hand much easier to find, as a rough grid is sufficient to obtain the sign changes and correct convergence as illustrated in Fig. 2.10(b). In fact, starting the Newton-Raphson in the center of any rectangle in Fig. 2.10(b) leads to the correct short-range solution.

Another numerical method to calculate the mode indices is based on the

phase of the complex function f, where the phase is chosen to be in the interval $[-\pi, \pi]$. Again the function is first evaluated in a region of the complex plane, but here the phase of f is considered instead of the sign changes in its real and imaginary parts. The phase of a complex function jumps by a factor of 2π when going in a certain direction near its root. This implies that a solution exists in a rectangle only if the magnitude of the phase change is larger than π in precisely one side of the rectangle [33]. Again the Newton-Raphson method is apploed to obtain the exact root. The method is illustrated in Fig. 2.10(c) for the short-rage mode, where the phase of f is shown by the colormap in the same region as in Fig. 2.10(b). Here, the color of a rectangle corresponds to the phase value in the bottom left corner of the rectangle. Again, the initial guess for the Newton-Raphson method is shown by the red circle, while the solution is shown by the black filled circle. It is observed that the phase is $\simeq \pm\pi/2$ and $\simeq \pi$ for values close to the blue and red contour curves in Fig. 2.10(b), respectively. The rough resolution applied in the colormap is sufficient to obtain the correct root of f.

The two numerical methods perform equally, and the method considering the sign changes has been applied to obtain the results presented in the thesis.

Chapter 3

Summary of results

This chapter contains selected results from the published papers. Sec. 3.1 contains results from paper I and II regarding the optics of multiple (ultra-sharp and rectangular) grooves in metal, while Sec. 3.2 shows quantities calculated using DFT in the jellium model as applied in paper III, IV, and V. Selected results regarding plasmons propagating in the gap and slab structures are found in Sec. 3.3, and Sec. 3.4 contains a collection of the results from paper V regarding electron-energy loss spectroscopy (EELS).

3.1 Optics of multiple grooves in metal

This section contains results regarding the optics of multiple grooves in gold, where the structure and incident field were illustrated in Fig. 2.1(b). The dielectric constant of gold in Fig. 2.6(c) (from Ref. [88]) has been applied in all the calculations in this section.

3.1.1 Plane wave as incident field

By applying the theory outlined in Sec. 2.1 when the incident field is a plane wave ($\gamma = \infty$), the optical cross sections are calculated and shown in Fig. 3.1(a). Here, extinction (EXT) refers to the amount of power removed from the reflected beam due to scattering and absorption, while OUP is the power scattered out-of-plane, and SPP (not shown here) is the power scattered into surface plasmons propagating away from the grooves along the interface. By normalizing the scattered and extinction powers with the power per unit area of the incident plane wave, the corresponding cross sections are obtained [29]. The absorption (ABS) cross section is calculated as the extinction cross sections minus OUP and SPP cross sections.

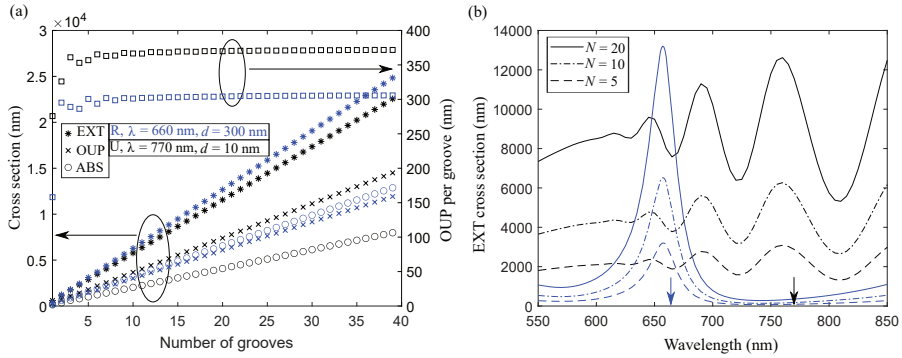

Fig. 3.1: The left y axis in (a) shows the EXT, OUP, and ABS cross sections as a function of the number of grooves, while the squares on the right y axis denote the OUP per groove. The black and blue quantities refer to ultrasharp (U) and rectangular (R) grooves, respectively, where λ and d are specified in the legend. (b) shows EXT spectra, where the U and R grooves scatter broadband and narrowband, respectively. The arrows indicate the fixed wavelengths applied in (a). The figures are combined from paper I and II and are reprinted with permission from ©The Optical Society.

The EXT, OUP, and ABS cross sections are shown on the left y- axis as a function of the number of grooves, N, and are seen to scale almost linearly with N. Here the black quantities are for an ultrasharp groove array (U) at the wavelength λ = 770 nm with $d = 10$ nm between the grooves (see Fig. 2.1(b)), while the blue quantities are for a rectangular groove array (R) at λ = 660 nm and $d = 300$ nm. This distance has been chosen as it maximizes the OUP and EXT as the period for this d is roughly half a wavelength. This will be further elaborated upon in paper II. It is highly surprising that the cross sections keep increasing linearly with N, as $N = 40$ corresponds to a total groove array length of $L \simeq 10$ μm which equals more than 13 wavelengths (for $\lambda = 770$ nm). The OUP cross section per groove is shown by the squares on the right y axis in Fig. 3.1(a), and is larger for multiple grooves than for a single groove, which will be explained below. For the ultrasharp grooves, the OUP per groove converges to roughly 1.5 times the top width of a groove. For a single ultrasharp groove, the OUP cross sections can exceed its physical dimension [39], but this was not expected for an array of many wavelengths, especially because an infinite array of ultrasharp grooves illuminated by a plane wave has a low reflectance [31]. It was stated in Ref. [39] that the low reflectance could be due to destructive interference between scattered light from the individual grooves, which would imply that the OUP per groove would become small when many grooves are found in the array. Fig. 3.1(a) clearly shows that this statement was wrong, and instead the OUP cross sections is significantly larger than the physical dimension of the grooves,

which is the first main result of paper I and II.

While Fig. 3.1(a) showed the cross sections for fixed wavelengths, a spectrum of the EXT cross section is shown in Fig. 3.1(b), where again the black and blue lines refer to an ultrasharp and rectangular groove array, respectively. For both arrays, the EXT scales linearly with N for all wavelengths, and the same is found for OUP and ABS (not shown). The main difference between the two types of groove arrays, is that the ultrasharp has cross sections that oscillate with the wavelength but is large in the entire spectrum, while the cross sections for the rectangular is only large in a narrow band of wavelengths. The black and blue arrows in the figure indicate the fixed wavelengths that was applied in Fig. 3.1(a). Hence the cross sections keep increasing the N, thus not explaining the low reflectance from the infinite array, which was found both theoretically and experimentally in Ref. [30] and [31], respectively.

3.1.2 Gaussian beam as incident field

The optical properties of the groove structures are instead investigated when the incident field is a Gaussian beam with finite beam radius as illustrated in Fig. 2.1(b). The reflectance from an array is calculated as the total reflected power divided by the incident power (see the appendix in paper I for details), and the reflectance from an array of 20 grooves is shown in Fig. 3.2(a) as a function of γ on a log scale. The colors have the same meaning as in Fig.

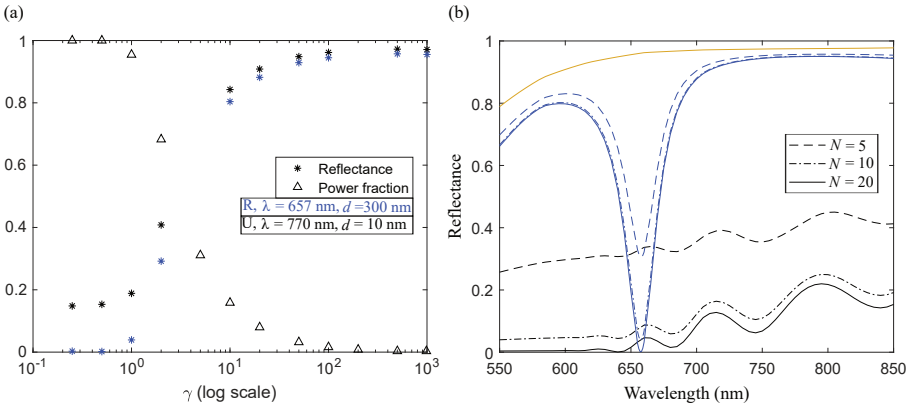

Fig. 3.2: (a) shows the reflectance from an array of 20 grooves as function of γ with corresponding power fraction that hits within the grooves, where the meaning of the colors are specified in the legend. (b) shows the reflectance spectra from an array of 5, 10, and 20 grooves for fixed beam radius corresponding to $\gamma = 0.5$ for 20 grooves. The reflectance from a planar gold surface is shown by the gold curve. The figures are combined from paper I and II and are reprinted with permission from ©The Optical Society.

3.1(a), but the wavelength for the rectangular groove array has been slightly modified to 657 nm, as it is the exact maximum for the blue curves in Fig. 3.1(b). The fraction of the incident power that actually hits the grooves is called PF and equals

$$PF(\gamma) = \frac{\int_0^L e^{-2(x-x_0)^2/w_0^2}\,\mathrm{d}x}{\int_{-\infty}^{\infty} e^{-2(x-x_0)^2/w_0^2}\,\mathrm{d}x} = \frac{2\int_0^1 e^{-2\tilde{x}^2/\gamma^2}\,\mathrm{d}\tilde{x}}{\gamma\sqrt{\pi/2}}, \tag{3.1}$$

where the transformation $\tilde{x} = 2(x - x_0)/L$ has been applied along with $w_0 = \gamma L/2$ and $\int_{\infty}^{\infty} e^{-ax^2}\,\mathrm{d}x = \sqrt{\pi/a}$. The power fraction is shown by the triangles in Fig. 3.2(a) and converges to 1 when $\gamma \leq 0.5$, as the Gaussian beam in this limit is so narrow that it is entirely focused within the grooves, and the reflectance is thus completely due to the grooves, and not the surrounding planar gold surface. In the other limit, when $\gamma \geq 500$ the power fraction is almost zero, as the beam is so wide such that it mainly hits the planar gold surface surrounding the grooves. Thereby the reflectance converges to that of a planar interface between air and gold, as the contribution from the grooves become negligible in this limit. Notice that the reflectance and power fraction in Fig. 3.2 converge in both limits at the same value of γ.

The reflectance as a function of wavelength is shown in Fig. 3.2(b) for the two types of arrays, where the beam radius is fixed corresponding to γ = 0.5 for 20 grooves, as this ensures that the planar surface does not contribute to the reflectance. For both types of arrays, the reflectance converges when 20 grooves are present, where the reflectance from ultrasharp and rectangular grooves is low in a broadband and narrowband spectrum, respectively. Importantly, the reflectance from the ultrasharp groove array is the same as obtained theoretically in Ref. [28], where an infinite groove array were illuminated by a plane wave. Furthermore, it was shown in Ref. [28] that the same reflectance can be obtained when using a stack matrix method, where each layer is modelled by a classical plasmon mode index. Hence it is found that an array of 20 grooves illuminated by a Gaussian beam that hits entirely within the grooves behave similarly as an infinite array of grooves illuminated by a plane wave, which is the second main result obtained in paper I and II.

The energy transportation has been investigated based on the time-averaged Poynting vector $\langle\vec{S}\rangle = 1/2\mathrm{Re}(\vec{E} \times \vec{H}^*)$, and it points in the direction of the power flow with its magnitude describing the power flow per area [1]. The magnitude of $\langle\vec{S}\rangle$ is shown on the colormap in Fig. 3.3(a,b,c,d), where the structure consists of 20 rectangular grooves with $d = 300$ nm. Here the length of the array is illustrated by the black horizontal lines in the figures, the wavelength is 657 nm, and the incident field is a Gaussian beam with ratios γ of 0.5, 1, 2 and ∞ (plane wave). For $\gamma = 0.5$, the magnitude of $\langle\vec{S}\rangle$ is only non-zero for positions right above the grooves, while more light is

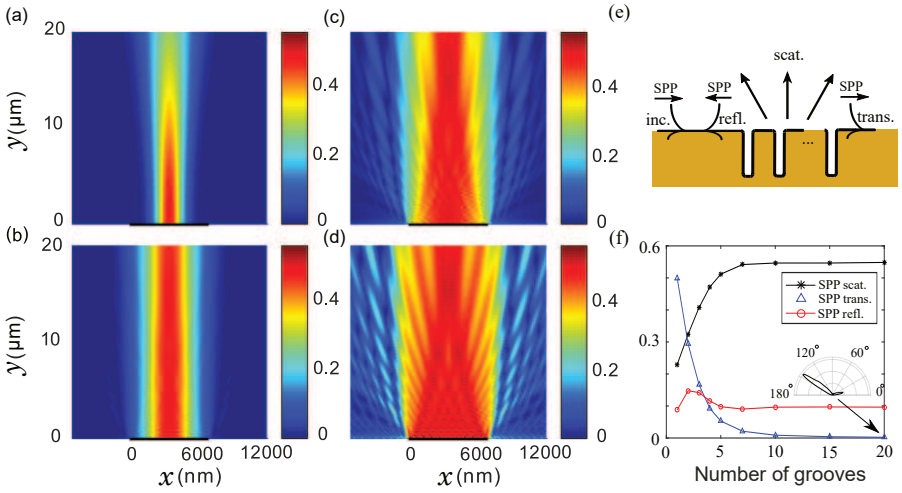

Fig. 3.3: Magnitude of the time-averaged Poynting vector for a Gaussian beam incident on 20 rectangular grooves with $d = 300$ nm. The ratio γ is 0.5, 1, 2, and ∞ (plane) wave in (a)-(d), respectively. (e) shows a schematic of a plasmon incident from the left on a groove array, which can be reflected, transmitted, and scattered out of the plane. (f) shows these quantities as a function of the number of grooves, where the inset shows the differential scattered power for a structure of 20 grooves. The wavelength is 657 nm. The figures are from paper II and are reprinted with permission.

scattered out of the plane when the ratio is increased.

At last, it is studied how the groove array reflects, transmits, scatters, and absorbs light incident as a plasmon from the left, as illustrated in Fig. 3.3(e). The quantities are shown in Fig. 3.3(f) at the wavelength 657 nm, and refers to the scattered, transmitted, and reflected powers divided by the power of the incident plasmon. Most of the incident light is transmitted across a single groove, while an array of 20 grooves scatters most of the light and transmits only a vanishing part. Notice that the absorption level is also increased when more grooves are present. The inset shows the differential scattered power on a polarplot, illustrating that most of the light is actually scattered backwards at an angle of roughly 150°. The fact that an incident plasmon can be scattered out of the plane explains why the OUP per groove for multiple groove is larger than for a single groove in Fig. 3.1(a).

3.2 Density-functional theory in the jellium model

This section shows selected results obtained using the DFT method in the jellium model presented in Sec. 2.2. This is an important part of the thesis,

as the electron densities presented here are fundamental to the papers III, IV, and V.

3.2.1 Slab structure

The electron density across a gold slab of width $d = 50$ Bohr is shown in Fig. 3.4(a) in units of the bulk gold density n_0. The ion charge is illustrated

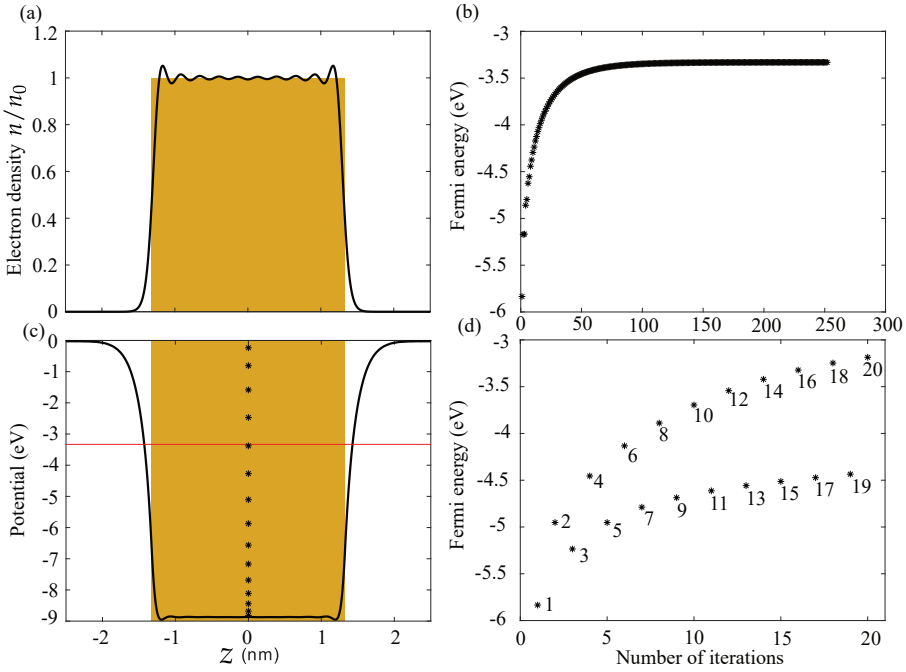

Fig. 3.4: The electron density in units of the bulk gold density n_0 is shown in (a), while the corresponding potential is shown in (c), where the red horizontal line and black dots depict the Fermi energy and the discrete quantum states, respectively. The Fermi energy as a function of number of iterations in the iterative method is shown in (b,d), where $\alpha = 0.03$ is small enough for the method to converge in (b), while $\alpha = 0.04$ is too high and the Fermi energy will never converge in (d).

by the colored area, and electron spill-out is clearly seen to occur as the electron density contains an exponential tail stretching $\simeq 0.3$ nm into the vacuum region. In addition, Friedel oscillations are found in the metal region close to the boundaries, in agreement with Refs. 45, 46. The corresponding potential V is shown in Fig. 3.4(c), where the black dots depict the energies of the discrete quantum states, while the red horizontal line depicts the Fermi energy. It is noticed that the oscillations in potential near the boundaries are smaller than the corresponding Friedel oscillations in Fig. 3.4(a). Practically

the same electron density and potential are found across a silver film, as the free electron density in bulk gold and bulk silver is almost the same [107]. The density is calculated in the sine base (Eq. (2.24)), and it is found that 15 basis functions per nm of computational domain are sufficient in order for the density to converge with respect to the number of basis functions.

As mentioned in Sec. 2.2.3, the mixing parameter α must be below some threshold value in order for the iterative method to converge. For the slab of width $d = 50$ Bohr, the threshold value is found to be 0.03, and Fig. 3.4(b) shows the Fermi energy as a function of number of iterations for this slab width. The Fermi energy converges after 252 iterations to $E_F = -3.33$ eV, as illustrated by the red line in Fig. 3.4(c). An example of the iterative method for a mixing parameter that is too high is shown in Fig. 3.4(d), where $\alpha = 0.04$ has been applied. The Fermi energy becomes oscillating into two different branches: one branch when the number of iterations is even, and another branch when it is odd, as illustrated by the numbers in Fig. 3.4(d). Clearly, this Fermi energy will never converge.

The threshold value for the mixing parameter decreases strongly with the slab width, as illustrated by the blue dots on the left y axis in Fig. 3.5(a). No

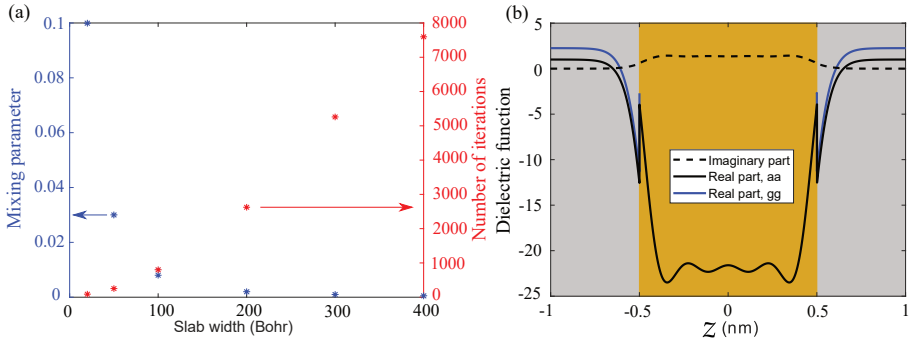

Fig. 3.5: (a) shows the mixing parameter (left y axis) and number of iterations (right y axis) as a function of slab width. (b) shows real (solid lines) and imaginary (dashed lines) parts of the quantum dielectric functions for a slab structure of 1 nm. (b) is from paper IV and is reprinted with permission.

analytic expression exists for the threshold value, which is instead found by trial and error. First the iterative method is started for some value of the mixing parameter. After a sufficient number of iterations, it is tested whether the Fermi energy evolves into two branches like in 3.4(d). If this is the case, the calculation is stopped and a lower mixing parameter is examined. Choosing a very low mixing parameter will ensure convergence, but according to Eq. (2.30) this implies that the input density in one iteration is almost the same as for the previous iteration, thus not utilizing much of the output density.

This implies that extremely many iterations are needed. Thus, it is desired to choose the mixing parameter as high as possible, but still below the threshold value. The corresponding number of iterations is shown by the red dots on the right y axis in Fig. 3.5(a). Due to the low mixing parameter for wide slabs, the number of iterations increases significantly with the slab width. Furthermore, both the number of basis functions and discretization points along the z axis scale linearly with d. This implies that the complexity of constructing the Hamiltonian using Eq. (2.26) and (2.27), which is the most demanding part of the flowchart in Fig. 2.4, is $\mathcal{O}(d^3)$. As the number of iterations in Fig. 3.5(a) scales faster than linearly, this implies that the complexity in the DFT method is at least $\mathcal{O}(d^4)$, illustrating that DFT is not a practical method when the systems become too large [78]. To compare, the complexity of the tight binding (TB) method is $\mathcal{O}(N^3)$ where N is the total number of orbitals in the model, while models based on the Dirac equation (DE) are scale invariant [113]. However, in e.g. graphene and carbon nanotubes the accuracy of TB and DE is also lower than in DFT [114].

The parallel part of the quantum dielectric function is computed from Eq. (2.46), and is shown across a slab of width 1 nm in Fig. 3.5(b), where the real and imaginary part are shown by the solid and dashed lines, respectively. Quantum spill-out implies that the dielectric function from the free electrons becomes smooth across the interfaces, while the interband part $\varepsilon_{\text{bound}}$ is seen as the step function, and the effect of a substrate is illustrated by the blue lines.

3.2.2 Gap structure

A gap structure has the opposite geometry compared to the slab structure (see Fig. 1.2), and it consists of a narrow gap of width w located between two semi-infinite metal surfaces. However, the surfaces are modelled as two thick slabs of width d separated by the gap width w. It is found that slabs of widths $d = 50$ Bohr are sufficient to avoid artefacts from finite-size effects. Notice that the density across such a slab was seen in Fig. 3.4(a). The density across three different gaps in units of n_0 is seen in Fig. 3.6(a,b,c), where the gap width is 1 nm, 0.5 nm, and 0.3 nm, respectively, and again the ion charge is illustrated by the colored areas. For the two smallest gaps, there is no true vacuum region between the metal surfaces, as the exponential tails from each surface overlap in the middle of the gap. Remember that the range of electron spill-out is $\simeq 0.3$ nm from each slab. This implies that the density in the middle of the gap of width 0.3 nm only decreases to roughly 9%, while it decreases to roughly 1% when the gap width is 0.5 nm, as seen in Fig. 3.6(b,c). Only when the gap width is larger than 0.6 nm, there is a true vacuum region between the metal surfaces, as clearly seen in Fig. 3.6(a) where the gap width is 1 nm.

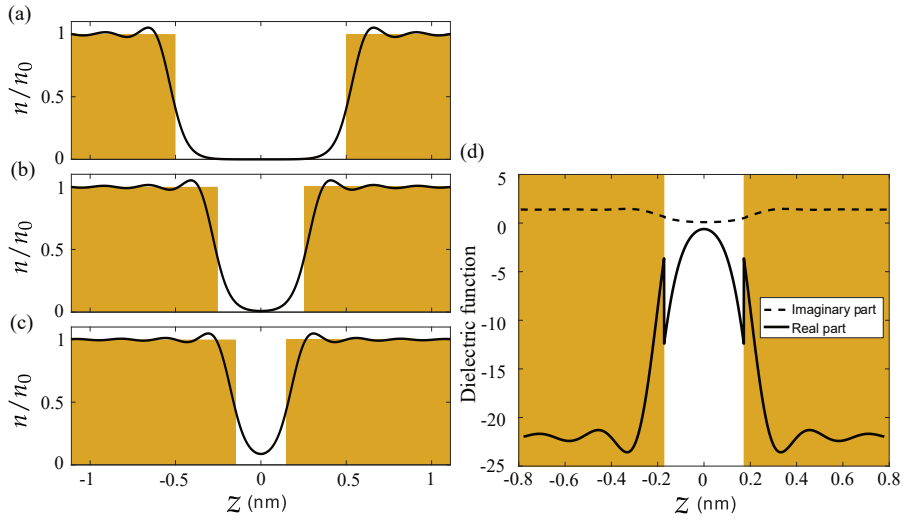

Fig. 3.6: (a,b,c) show electron densities across three gap structures in units of n_0, where the gap width is 1 nm, 0.5 nm, and 0.3 nm, respectively. The corresponding dielectric function across a gap of width 0.35 nm is shown in (d), which is from paper III and is reprinted with permission.

The corresponding parallel part of the quantum dielectric function is shown in Fig. 3.6(d) for a gap width of 0.35 nm. In order for the plasmonic modes to converge, it is found sufficient to include four periods of Friedel oscillations near the boundary and to apply the bulk value beyond this range as illustrated.

3.3 Surface plasmon polaritons

This section contains results dealing with the propagation of plasmons in gap and slab structure of gold. The dielectric function is calculated from Eq. (2.46), and was illustrated in Fig. 3.5(b) and 3.6(d).

3.3.1 Plasmon mode index

The plasmon mode indices have been calculated using the transfer matrix method presented in Sec. 2.3.2 and are shown for the gap structure in Fig. 3.7(a), where the wavelength is 775 nm. The real and imaginary parts without spill-out (w/o SO), i.e in a classical model, are shown as the blue solid and dashed lines, respectively, and diverge in the limit of vanishing gap width. When including spill-out, the exponential tails from the gold regions overlap and there is no true vacuum region for sufficiently thin gaps (Fig. 3.6). This

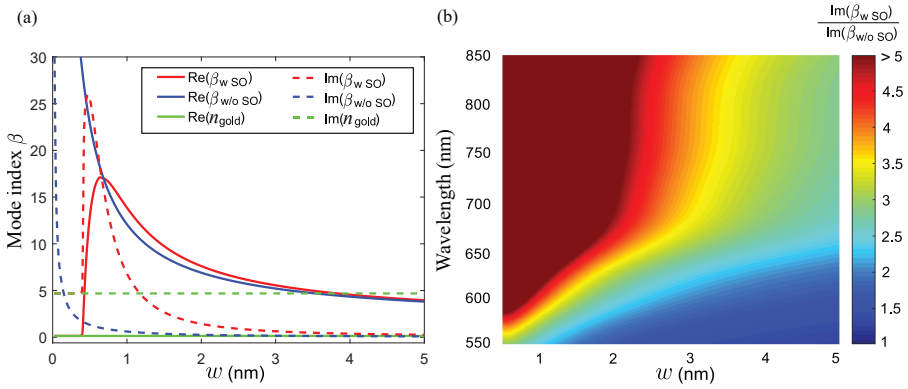

Fig. 3.7: (a) shows the real (solid line) and imaginary parts (dashed line) of the mode index for a gap structure at a wavelength of 775 nm. The horizontal green lines show the refractive index of bulk gold. (b) shows the ratio between imaginary parts of the mode index with and without spill-out (SO) on a colorplot. The figures are from paper III and are reprinted with permission.

implies that the plasmon mode index (w SO, red lines) for thin gaps converges to the refractive index of bulk gold, n_{gold}, as shown by the green lines. The mode index converges when the slab width is 0.35 nm, for which the dielectric function was shown in Fig. 3.6(d). Hence even though this dielectric function is quite different from that of bulk gold, it behaves almost as bulk gold when it comes to the plasmon mode index. Thus, quantum spill-out implies that the mode index does not unphysically diverge in the limit of vanishing slabs, instead it converges to the refractive index of bulk gold, thus restoring correct physical behaviour. This is the first the main result in paper III.

For gap widths above 1 nm, the real part of the mode index with and without spill-out is almost the same, while there is a large difference for the imaginary part. This is further examined in Fig. 3.7(b) showing the ratio between the imaginary parts with and without spill-out on a colorplot. It is seen that for gap widths up to 5 nm, there is a significant increase in the imaginary part of the mode index, even though the spill-out range is only 0.3 nm. The reason to the increased imaginary part due to spill-out will be given below.

The corresponding short-range mode indices for a gold slab are shown in Fig. 3.8(a) at the wavelength 775 nm, where the blue and red lines are for gold surrounded by the glass (gg) and air (aa), respectively. These mode indices have been calculated using an isotropic response, thus setting $\varepsilon_{\text{ani}} = 0$ in Eq. (2.47). In general, anisotropy modifies the mode index for both kinds of modes, while the relative change is much larger for the long-range mode, but

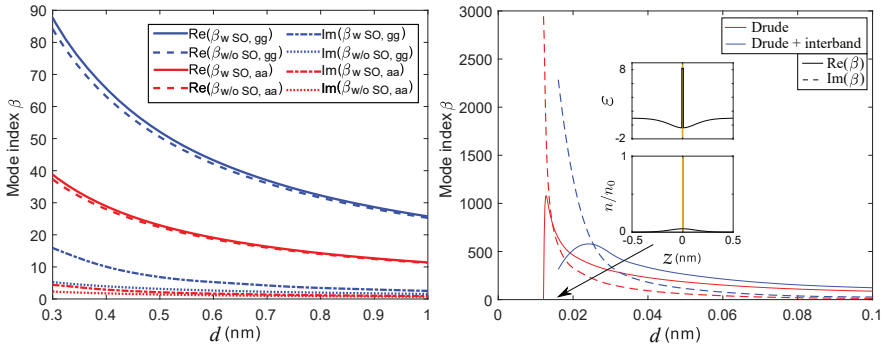

Fig. 3.8: (a) shows the mode index of a gold slab surrounded by glass (gg) and air (aa). (b) shows the mode index when including spill-out for a slab surrounded by air at subatom slab thickness. The insets show the electron density and dielectric function across a slab of width $d = 0.015$ nm. The red curves are calculated where the response is pure Drude, while the interband term has been included in the calculations shown by the blue curves. The wavelength is 775 nm and the response is isotropic. (a) is from paper IV and is reprinted with permission.

all the results in the following are for an isotropic response, while anisotropy has been included in some of the results in paper IV below. Again, the mode indices in Fig. 3.8(a) diverges in a classical model neglecting spill-out. To study whether this divergence is again an artefact, the mode index when including spill-out is calculated in thinner slabs as shown in Fig. 3.8(b) to examine how it behaves in the limit of vanishing slab width. The diameter of a gold atom is roughly 0.3 nm [31], why it is totally unphysical to consider slab widths below this limit, and Fig. 3.8(b) is therefore highly theoretical. But as spill-out is included by applying the jellium model, it is possible to calculate a theoretical electron density for slabs in the subatom range, which does not simply not exist in an atomistic model.

The results shown in Fig. 3.8(b) are for an isotropic response for a gold slab surrounded by air at a wavelength of 775 nm. The red curves have been calculated where only the free electron term is included in the quantum dielectric function, while the interband term (see Eq. (2.46)) has been included in the results shown by the blue curves. Again the real and imaginary parts are shown by the solid and dashed lines, respectively. The mode index increases for decreasing slab widths, but for some sufficiently thin slab, the plasmon mode ceases to exist. At this cuf-off, the electron density has become so delocalized that the dielectric function is everywhere positive. Recall from Sec. 2.3.1 that a plamon only exists then the metal dielectric function is negative. The electron density and dielectric function for a slab width of $d = 0.015$ nm ($\simeq$0.3 Bohr) are illustrated in the inset, where the interband

part of the dielectric function gives rise to the step function at the interfaces, such that its value within the slab is roughly 8, clearly illustrating that the system is unphysical. When applying only the Drude term, the step function is omitted, and the plasmonic modes exists for slightly thinner slabs as illustrated by the red lines. The theoretical outcome of this figure is that the mode index does not diverge in the limit of vanishing slab thickness when including spill-out, instead the plasmonic modes cease to exist for slabs in the subatom range. From now on, however, only slab widths larger than 0.3 nm will be considered.

The ratio between the imaginary part of the mode index with and without spill-out was shown for the gap structure on the colorplot in Fig. 3.7(b). The corresponding ratio for the slab structure surrounded by glass (gg) is shown in Fig. 3.9(a,b) for the short- and long-range mode, respectively, where the ratios between real parts are shown by the solid lines and is practically unaffected by spill-out. The ratio between the imaginary parts converges to 1.2

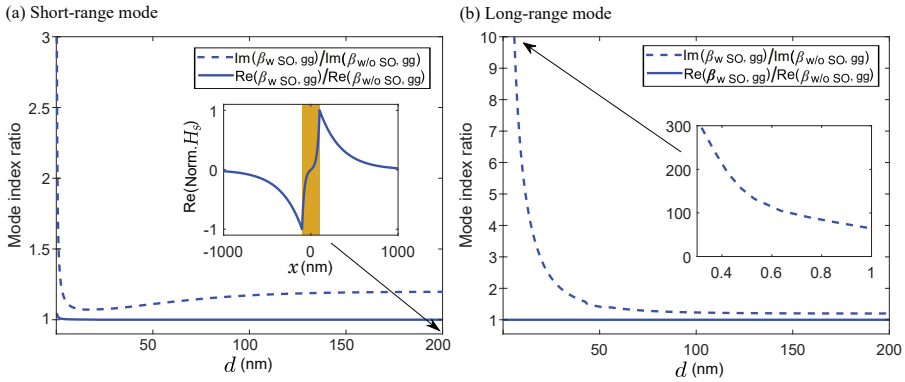

Fig. 3.9: Ratio between mode indices with and without spill-out at a wavelength of 775 nm for a gold slab surrounded by glass (gg). The modes are short- and long-range in (a) and (b), respectively, and the response is isotropic. The figure is from paper IV and is reprinted with permission.

when the slab width is 200 nm for both kinds of modes. Thus, spill-out gives rise to an increase in the imaginary part of 20% for relatively thick slabs that can be fabricated. This is highly surprising as the electron density is only modified in regions close to the surface. The significant increase in the imaginary part of the mode index due to spill-out is the first main result of paper IV. The increased imaginary part will manifest itself in decreased propagation lengths in fabricated plasmonic structures as shown in Refs. [115–118]. The inset in Fig. 3.9(a) shows the real part of the magnetic field for a slab width of 200 nm, and it behaves as two decoupled plasmons bound to the

individual interfaces for such a wide slab.

3.3.2 Plasmonic absorption

The physical explanation of the enhanced imaginary part of the mode index is found in the absorption density. The electric field of the plasmons is found from the magnetic field in Eq. (2.36) as

$$\vec{E}(x,z) = \frac{i\overleftrightarrow{\varepsilon}^{-1}(z)}{\omega\varepsilon_0}\nabla \times [\hat{y}H^p(x,z)], \tag{3.2}$$

from which the absorption density is calculated as

$$A(x,z) = \mathrm{Im}(\vec{E}^*(x,z) \cdot \overleftrightarrow{\varepsilon}(z) \cdot \vec{E}(x,z)). \tag{3.3}$$

The free electron part of the part of the quantum dielectric function is smooth across the structure, implying that its real part becomes zero at certain positions close to the jellium edges (see Fig. 3.5(b)). As the electric field and absorption density scale as $\overleftrightarrow{\varepsilon}^{-1}(z)$, they contain narrow peaks at these positions. Notice that the imaginary part of the dielectric function is small but non-zero at these positions, which ensures that the peaks in absorption density are finite. The absorption density near a gold slab is shown in Fig. 3.10(a), where the response is isotropic and the kind of substrate and superstrate slightly shifts the peak positions. In a classical model neglecting spill-

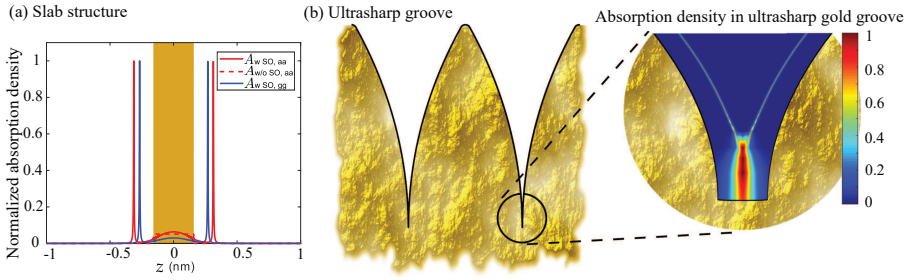

Fig. 3.10: Absorption density near a gold slab of width 0.3 nm in shown in (a), while it is illustrated in the bottom of an ultrasharp groove in (b). The wavelength is 775 nm and the response is isotropic. (a) is from paper IV and is reprinted with permission.

out, absorption can only take place in the gold, and the enhanced absorption is therefore a consequence of spill-out. The enhanced imaginary part of the mode index due to spill-out is thus explained in terms of strong plasmonic absorption taking place at narrow peaks close to the surface, which is the second main result of paper IV.

The free electron part of the quantum dielectric function in a gap structure is also smooth and its real part for gap widths above a certain threshold is zero at positions close to the jellium edges. For smaller gaps, the overlap in electron densities from the gold surfaces implies that the dielectric function is everywhere negative, see Fig. 3.6(d). A continuous range of gaps between metal surfaces is found in ultrasharp grooves as illustrated in Fig. 3.10(b), where the colormap shows the absorption density near the bottom of such a groove. It shows that absorption mostly takes place near the bottom, but also in narrow regions close to the jellium edge. As in Fig. 3.10(a), the real part of the dielectric function is zero at these positions.

The increased absorption density in the ultrasharp groove implies that the reflectance from such a groove array is lowered. The reflectance from an array of 20 grooves was calculated using GFSIEM in a classical model with a narrow Gaussian beam as the incident field as shown in Fig. 3.2(b). Importantly, as also stated in Sec. 3.1.2, this reflectance is the same as for an infinite array illuminated by a plane wave. The latter can be calculated using the GFSIEM with a periodic Greens function, or with a stack matrix method based on the classical gap plasmon mode index giving the same result [30]. To calculate the reflectance from a groove array when spill-out is taken into account, the same stack matrix method is applied as in Ref. [30], but with the gap plasmon mode index calculated with spill-out, see Fig. 3.7(a). Here the ultrasharp groove is modelled as a staircase as illustrated in Fig. 3.11(a), where each level i is modelled as having the refractive index equal to the mode index for a gap of width w_i. Hence, the ultrasharp groove is modelled

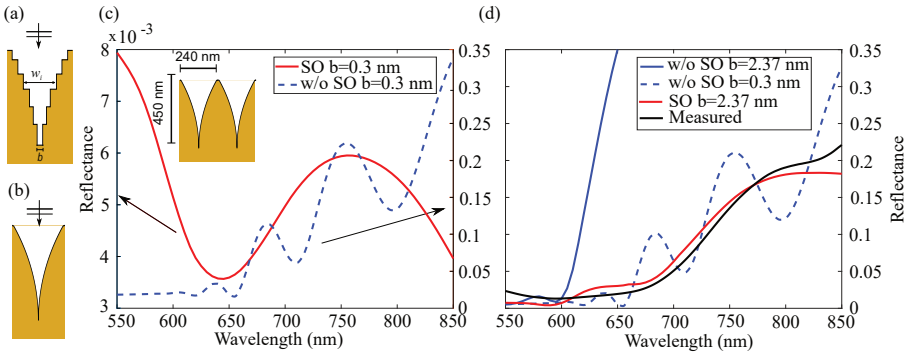

Fig. 3.11: (a) and (b) illustrate how an ultrasharp groove is modelled as a staircase consisting of 6 and 500 steps, respectively. The calculated reflectance of an ultrasharp groove array with a bottom width of b = 0.3 nm is shown in (c), while a comparison with the measured reflectance from Ref. [31] is shown in (d). (c) and (d) are from paper III and are reprinted with permission.

as a multilayer structure, and the reflectance is calculated by a simple transfer

matrix method [110]. It is found that by applying 500 steps in the staircase, the structure behaves as the smooth groove illustrated in Fig. 3.11(b).

Fig. 3.11(c) shows the reflectance from an ultrasharp groove array with and without spill-out, where the dimensions of the grooves are shown in the inset, and the applied bottom width is b = 0.3 nm. By noticing the very different scaling on the two y-axis, it is clearly seen that spill-out significantly lowers the reflectance. The calculated reflectance without spill-out for b = 0.3 nm does to some extent agree with the measured reflectance from the groove array from Ref. [31] shown by the black curve in Fig. 3.11(d). However, the calculated reflectance contains oscillations which are not present in the measured spectrum. Furthermore, the exact bottom width is impossible to measure precisely, but is highly unlikely only the 0.3 nm as assumed in the calculation shown the blue dashed line. Notice that 0.3 nm is roughly the diameter of a gold atom [31]. Calculating instead the reflectance when spill-out is included, and minimizing the root mean square error between measured and calculated reflectance, it is found that a bottom width of b = 2.37 nm gives the best reflectance. The reflectance for this bottom width is shown by the red curve in Fig. 3.11(d) and is close to the measured reflectance shown in black. The corresponding reflectance without spill-out for b = 2.37 nm is shown by the blue solid line and is much higher than the measured reflectance. Hence, when including spill-out the calculated reflectance becomes in much better agreement with measurements compared to classical models neglecting spill-out, and in addition, the bottom width becomes 2.37 nm, which is much more realistic than 0.3 nm as assumed in the classical model. This is the second main result of paper III.

3.4 Electron-energy loss spectroscopy

The theory regarding electron-energy loss spectroscopy (EELS) in the non-retarded limit is presented in detail in paper V, and a summary of the results is found here. Retardation effects being ignored means that the response from the electron is instantaneous, which is a reasonable approximation as the considered electron velocities are much smaller than the velocity of light, and the film thicknesses are much smaller than the optical wavelength [5].

The loss probability Γ^{EELS} can be written in terms of the imaginary part of the reflection coefficient r_p [102]. Within the full nonlocal RPA, r_p is calculated using Eq. (18) in paper V, and its imaginary part is shown on the colorplot in Fig. 3.12(a,b,c) for a metal film of 10 monolayers (ML) of facet 111 [107], where the film is surrounded by air and modelled using DFT in the jellium model (denoted JEL in the following). Here the colorbar is logarithmic, the ML thickness is $a = 0.236$ nm for Ag(111) and $a = 0.235$ nm for Au(111) [104], and Q has the same meaning as k_x in Fig. 2.6(a,b) denoting the wave

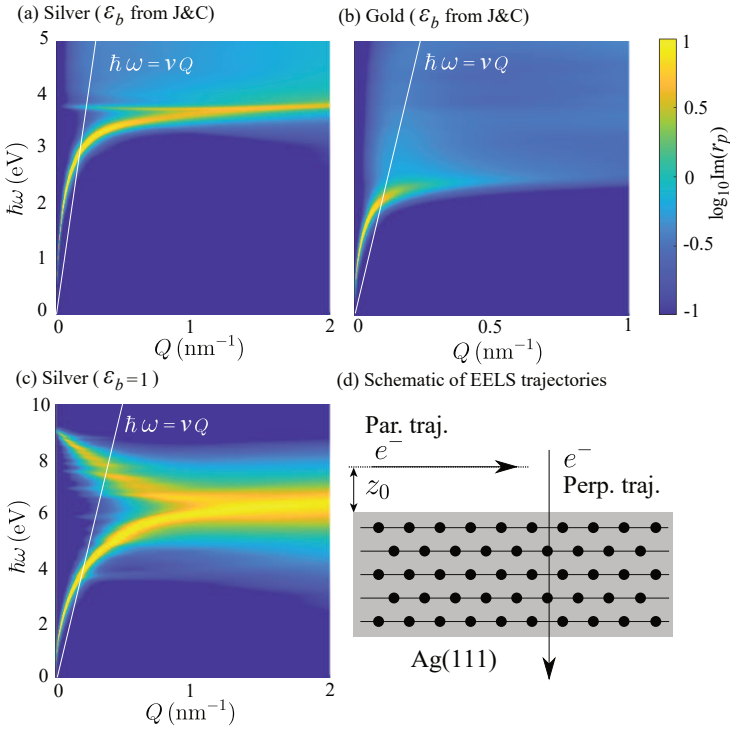

Fig. 3.12: Colorplots of the imaginary part of the reflection coefficient for a metal film of 10 ML of facet 111, where the white lines show $\hbar\omega = vQ$ for $v = c/10$. In (a) and (b), the background dielectric constants ε_b are from J&C (Ref. [88]) for silver and gold, respectively, while (c) is for silver with $\varepsilon_b = 1$. A schematic of the parallel and perpendicular trajectories in EELS is shown in (d). The figures are from paper V and are reprinted with permission.

number component parallel to the film. The dispersion relation of an incident electron travelling with velocity v parallel to the film is $\omega = vQ$, and the white lines in Fig. 3.12(a,b,c) depict $\hbar\omega = vQ$ for $v = c/10$. The background dielectric function ε_b applied in Fig. 3.12(a,b) is from J&C for silver and gold, respectively. Due to the non-zero imaginary part of ε_b in gold for $\hbar\omega > 2$ eV, the reflection coefficient for such frequencies is suppressed due to damping, which was also observed in Fig. 2.6(b), where a single interface between gold and air was studied. The imaginary part of ε_b is smaller in silver, implying that the maximum in r_p is more pronounced as seen in Fig. 3.12(a), where the back bending is also visible near the bulk plasmon frequency $\hbar\omega_b \simeq 3.78$ eV, around which $\mathrm{Im}(\varepsilon_b)$ starts to become non-zero, see Fig. 2.6(d). Setting $\varepsilon_b = 1$ gives the reflection coefficient shown in Fig. 3.12(c), where the bulk and surface plasmon frequencies are clearly seen at ω_p and $\omega_p/\sqrt{2}$, respectively.

The applied damping rate $\hbar\gamma$ is 21 meV in silver and 71 meV in gold [88]. Several EELS experiments have been performed on silver films in order to characterize their optical properties [9,10,12,13], and the loss functions shown below are therefore calculated for silver, while they are shown for gold in Fig. 8 in paper V.

As mentioned, plasmons are found as poles in the reflection coefficient [33], where their parallel wave number Q (or k_x) becomes complex when the metal dielectric constant is complex. Both components of the plasmon mode index $\beta = Q/k_0$ are considered in paper III and IV, and were shown in Fig. 3.7(a) and 3.8(a) above. When only considering the real part of Q, as in e.g Fig. 3.12 and Refs. [76, 119], the plasmons are found as maxima, not poles, in the reflection coefficient. Hence the plasmon dispersion relations are seen by the yellow curves in Fig. 3.12 denoting the maxima of $\mathrm{Im}(r_p)$. The curves are similar if considering $|r_p|$ instead. Extending Fig. 3.8(a) to larger d, the mode index for a slab of 10 ML Au(111) ($d = 2.35$ nm) is found to be $4.89 + 0.32i$ at a wavelength of 775 nm. This corresponds to $Q = 0.04$ nm^{-1} at $\hbar\omega = 1.6$ eV, which is found to be in good agreement with Fig. 3.12(b). Nonlocal effects have been neglected and included in Fig. 3.8(a) and Fig. 3.12(a), respectively, and the fact that the calculated plasmon mode indices within the two models are in good agreement illustrates that nonlocality is a relatively small correction when plasmons propagate in thin slabs, as was also found in Ref. [76]. This will be further elaborated upon below.

In EELS an electron can be incident parallel or perpendicular to the film, as illustrated in Fig. 3.12(d). Its velocity v is determined from the kinetic energy of the electron beam in the relativistic form as [120]

$$E_0 = m_e c^2 \left(\frac{1}{\sqrt{1 - v^2/c^2}} - 1 \right), \tag{3.4}$$

where c is the speed of light in vacuum. An electron is a moving point charge in real space, implying that its charge density becomes completely delocalized in reciprocal space [5, 6]. The incident electron can excite plasmons when their parallel momentum and energy are the same, as first discovered by Ritchie [4]. He proposed that fast electrons could excite plasmons, which is an alternative to the Otto or Kretschmann configurations, where an evanescent field from incident light is applied as excitation sources [2, 3]. When an electron excites a plasmon it loses an amount of energy as described by the loss function Γ^{EELS}. An electron moving parallel to the film has parallel wave number $Q = \omega/v$ as illustrated by the white lines in Fig. 3.12, which implies that only the modes right to this line can be excited by a parallel trajectory. The loss function for this trajectory is given by [102] (Eq. (17) in paper V)

$$\Gamma_{\parallel}^{\mathrm{EELS}}(\omega) = \frac{2e^2 L}{\pi\hbar v^2} \int_{\omega/v}^{\infty} \frac{\mathrm{d}Q}{\sqrt{Q^2 - \omega^2/v^2}} e^{-2Qz_0} \mathrm{Im}(r_p(Q,\omega)), \tag{3.5}$$

where L is the length of the parallel trajectory. The loss function contains a peak at the frequency where the parallel momentum of the electron and the plasmon is the same, which is where the white lines cross the yellow curves in Fig. 3.12(a,b,c) [102]. Another peak is found close to the surface plasmon frequency, at which the plasmon dispersion has a low slope, thus giving a large contribution from $\mathrm{Im}(r_p)$ when integrating along Q. These peaks are observed in Fig. 3.13 showing the loss functions for a film of 10 ML Ag(111), where ε_b is specified in each panel, and with the loss functions for parallel and perpendicular trajectories shown in black on the left y-axis and in blue on the right y-axis, respectively. The parallel loss has been computed using $z_0 = 0.5$ nm. The loss for the perpendicular trajectory when $\varepsilon_b = 1$

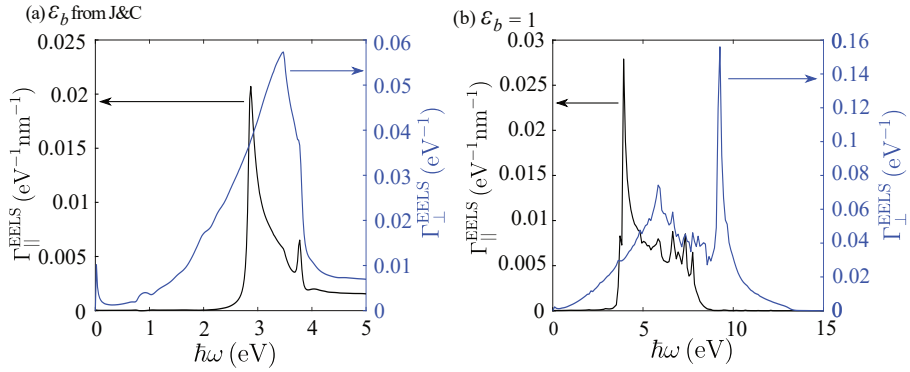

Fig. 3.13: Loss functions for a film of 10 ML Ag(111) modelled using DFT, where ε_b is from J&C in (a) and is set to 1 in (b). In both figures, the loss functions for the parallel and perpendicular trajectories are shown by the black curve on the left y axis and the blue curve on the right y axis, respectively.

consists also of two main peaks, located at the bulk and surface plasmon frequencies, respectively. But when using the measured dielectric function from J&C in Fig. 3.13(a), the loss function consists of a single peak near the surface plasmons frequency, as the other peak has been suppressed due to damping caused by the imaginary part of ε_b.

An example of the integration kernel for the perpendicular trajectory, $\Gamma_{\perp}^{\mathrm{EELS}}(Q,\omega)$ (Eq. 13 in paper V) is shown in Fig. 3.14(a) for $N = 10$ ML of Ag(111) at a fixed $\hbar\omega = 3$ eV and with ε_b from J&C. The curve consists of a relatively narrow peak around $Q = 0.17$ nm^{-1}, as a plasmon is excited for this particular Q, as seen in Fig. 3.12(a), whereby the electron loses the energy described by the kernel $\Gamma_{\perp}^{\mathrm{EELS}}(Q,\omega)$. In order to calculate the correct energy loss, it is necessary to use a resolution along Q that is fine enough to capture the narrow peak in Fig. 3.14(a), where the location of the peak depends on the frequency. In practice, the integration along Q is performed

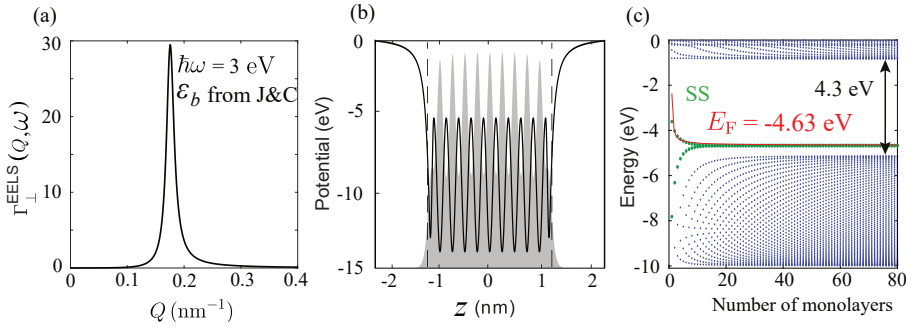

Fig. 3.14: (a) shows $\Gamma_\perp^{\mathrm{EELS}}(Q,\omega)$ (Eq. (13) in paper V) at fixed $\hbar\omega = 3$ eV for a film of 10 ML Ag(111) modelled using DFT. The peak corresponds to the maxima in r_p from Fig. 3.12(a). (b) shows the potential across the same film modelled using ALP, while (c) shows the discrete energy levels calculated using ALP. Here, the red line and green dots denote the Fermi energy and surface states (SS), respectively. (b) and (c) are from paper V and are reprinted with permission.

by first calculating the kernel using a rough resolution in Q, from which the location of the peak can be identified. Afterwards, a fine resolution is applied for Q close to the peak and trapezoidal integration is applied to obtain the loss function. Same method is applied to calculate the loss function for the parallel trajectory, where the kernel is the integrand in Eq. (3.5) (Eq. (17) in paper V).

In order to account for the surface orientation (facet 111) of the silver film, another model is applied for the potential across the film. It is called an atomic layer potential (ALP) and is described as an analytic function with parameters based on fitting the work function, the fermi energy, the surface states, and the projected gap to experimental data [104]. An explicit expression for the potential as well as the applied parameters for Ag(111) can be found in appendix F of Ref. [101]. The potential is shown across a film of 10 ML Ag(111) as the solid line in Fig. 3.14(b), where the corresponding electron density is illustrated by the shaded area, while the vertical dashed lines denote the jellium edges. The potential clearly contains stronger oscillations than the DFT potential in Fig. 3.4(c), and they give rise to a projected energy gap between the highest occupied state and the lowest unoccupied state. Furthermore, the ALP contains another type of states called surface states (SS) [101, 104], and their wave functions are concentrated near the surface. Fig. 3.14(c) shows the discrete energy levels in ALP as a function of the ML in the film, where the red line and green dots denote the Fermi energy E_F and surface states (SS), respectively. However, here the effective mass of the occupied energy bands have been varied such that $m_j^*/m_e = a\hbar\varepsilon_j + b$, where the blue dots in Fig. 3.14(c) denote $\hbar\varepsilon_j$, and the value a and b are fitted to

experiments [104,121–124]. The effective mass of the SS and the empty bands are set to a constant and the electron mass, respectively. The applied values are shown in Table 1 in paper V. Using a constant effective mass of all the bands corresponds to setting $a = 0$ and $b = 1$. It is noticed that ε_j in paper V, and in the following, denotes a frequency, with $\hbar\varepsilon_j$ corresponding to the energy ε_j as applied in the dielectric function in paper III and IV. The varying effective mass model leads to a Fermi energy given in Eq. (10) in paper V, where Eq. (2.28) above is restored if $m_j^* = m_e$ for all j.

The energy bands of a film of 10 ML Ag(111) calculated using ALP are shown in Fig. 3.15(a,b) when the effective mass is constant and varying, respectively. In the former, all the bands have the same curvature, which implies that certain resonances are enhanced in the optical response, thus giving rise to spurious features. This is clearly seen in Fig. 3.15(c) showing the imaginary part of the reflection coefficient on a colorplot, where the bands split up for energies above 3 eV. When the effective mass of the bands is varied, all the bands have different curvature, as seen in Fig. 3.15(b), which gives rise to the reflection coefficient shown in Fig. 3.15(d). Here, the splitting of the bands has been washed out as no resonances are artificially enhanced. Remarkably, this reflection coefficient is not so different from the one calculated using the JEL (DFT) potential in Fig. 3.12(a), despite the fact that JEL does not consider the surface orientation, and therefore contains no projected energy gap. In addition, a constant effective mass has been applied in JEL. Fig. 3.15(e,f) show the loss functions for the perpendicular and parallel trajectories with $z_0 = 0.5$ nm, respectively, where the line types refer to the model used for the effective mass. Here it is seen that splitting of bands in Fig. 3.15(c) imply a splitting of peaks in the loss functions. Hence, in order to avoid artificial splitting phenomena when using the ALP, it is necessary to account for the varying effective mass of the energy bands. This is one of the main results of paper V.

Having now established the JEL and ALP models, loss functions calculated using these two models are compared in Fig. 3.16(a,b) and (c,d) for the perpendicular and parallel trajectories with $z_0 = 0.5$ nm, respectively. Here the number of layers is fixed and the electron energy varied in (a,c), while the energy is fixed in (b,d) varying the number of layers. In the perpendicular case, the classical model refers to the expression derived in Refs. [4, 6, 125], while it in the parallel case refers to applying Eq. (3.5) with the classical reflection coefficient fromEq. (2.39). Remarkably, the results obtained using the classical model are in good agreement with those obtained using the full nonlocal quantum dielectric response within the RPA. Furthermore, the former is almost independent of the actual shape of the potential, as ALP and JEL give roughly the same results. The same was found for the reflection coefficient above.

The small effect of nonlocality is assigned to the fact that electron motion

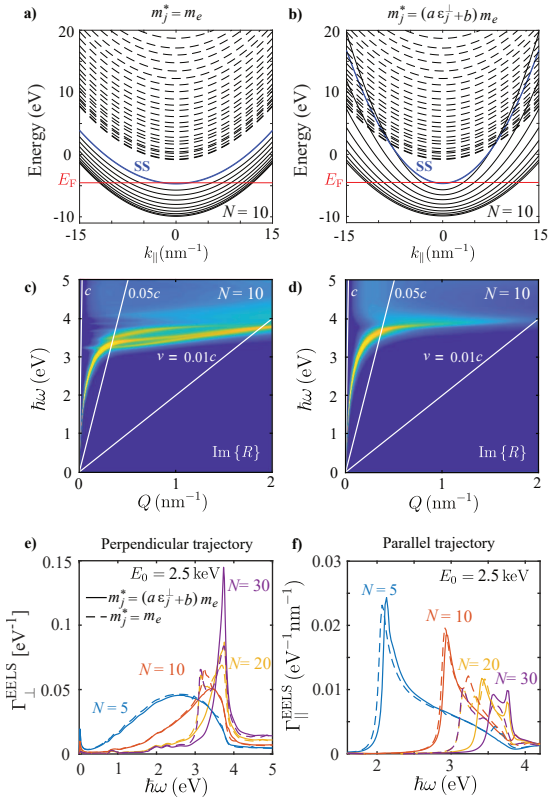

Fig. 3.15: The energy bands for 10 ML Ag(111) are shown in (a,b) where the effective mass is specified on top of the panels, and with the corresponding reflection coefficient shown in (c,d). The loss functions for perpendicular and parallel trajectories ($z_0 = 0.5$ nm) are shown in (e,f), where the line types refer to the effective mass model. The figure is from paper V and is reprinted with permission.

in the film mostly takes place in the direction parallel to the film, as the film is strongly confined in the perpendicular direction. Here the motion parallel to the film is described using a free electron model, where the effective mass is, however, modified in ALP. Hence in thin metal films, there are two directions in which the electrons behave as (almost) free, implying that the effect from the strong nonlocality in the third (perpendicular) direction is relatively small. This is in agreement with Ref. [76] studying plasmons propagating in thin metal gaps and slabs, finding that nonlocality is a relatively small correction. In contrast to nano structures like dimers, clusters, and spheres, which are confined in all spatial directions, and where nonlocality is more important and blue-shifts the plasmon resonances [78–80, 85, 86]. The small

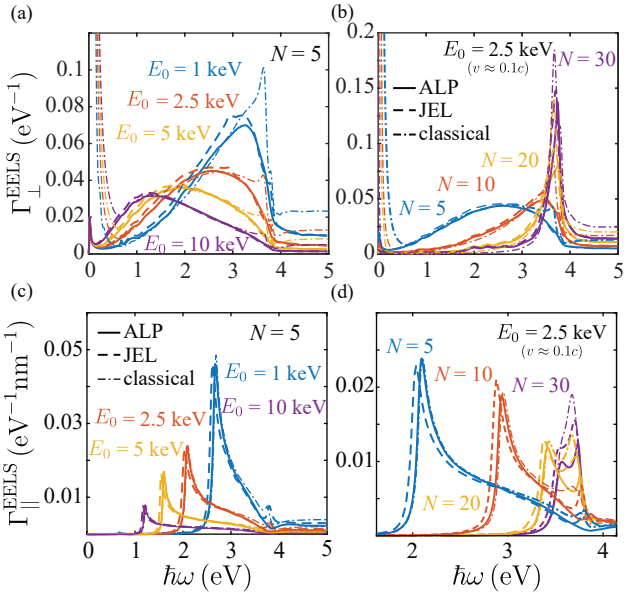

Fig. 3.16: Loss functions for the perpendicular and parallel trajectory ($z_0 = 0.5$ nm) in upper and bottom row, respectively, where $N = 5$ ML is fixed in (a,c) and $E_0 = 2.5$ keV is fixed in (b,d). The figure is from paper V and is reprinted with permission.

effect from nonlocality for plasmons propagating in thin metal slabs, along with the physical explanation of the phenomena, is another main result of paper V, and it also justifies the simplified treatment of nonlocality in paper III and IV.

Chapter 4

Conclusions

The main goal of this thesis has been to obtain a theoretical understanding of plasmons propagating in different noble metal structures at nanoscale. The structures have consisted of arrays of ultrasharp and rectangular grooves in gold, a dielectric gap between two parallel metal walls, and a metal slab sandwiched between different dielectrics.

First, a classical model was applied to study the transition from one to multiple grooves in gold, calculating the optical cross sections and the reflectance from the groove array. Here arrays of ultrasharp and rectangular grooves are good absorbers in a broadband and narrowband wavelength interval, respectively. Surprisingly, it was found that the scattering and extinction cross sections increase almost linearly with the number of grooves for all wavelengths. In addition, the out-of-plane scattering cross sections per groove is actually larger than the physical dimension of a groove, even for a large structure of more than 13 wavelengths. Hence, the optical cross sections can not explain why the reflectance from an infinite array can be very low. Instead a Gaussian beam has been applied as the incident field, and the reflectance computed as the total reflected power divided by the incident power. It is found that a narrow Gaussian beam focused entirely within an array of 20 grooves gives rise to the same reflectance as an infinite array illuminated by a plane wave. However, in order for the calculated reflectance from an ultrasharp groove array to be in reasonable agreement with previous measurements, it was necessary to assume a bottom width of the groove of only 0.3 nm, which is roughly the diameter of a gold atom.

Next, quantum effects were taken into account as classical models have limitations when the size of the structure is shrinked to atomic scale. The electron density was calculated using density-functional theory in the jellium model (JEL), and becomes smooth across an interface between metal and air with a spill-out range of roughly 0.3 nm. Neglecting nonlocal effects in the

dielectric function, but applying the smooth electron density, the mode index for plasmons propagating in a thin gap between two parallel gold walls, was found to converge to the refractive index of bulk gold in the limit of vanishing gap. Thus restoring the correct physical behaviour, and removing the unphysical divergence obtained using classical models. Furthermore, the calculated reflectance from an ultrasharp groove array when quantum spill-out is taken into account, is in much better agreement with measurements compared to classical models. In addition, the bottom width of the groove is fitted to 2.37 nm, which is much more realistic than the applied 0.3 nm in the classical model.

Furthermore, the opposite geometry, i.e a metal slab sandwiched between different dielectrics, has been studied taking spill-out into account. Here, only the parallel part of the nonlocal contribution was neglected, while the perpendicular part was modelled by an averaging procedure, resulting in a local, but anisotropic, dielectric function. Anisotropy was found to modify both kinds of plasmonic modes, but the relative change was much larger for the long-range mode. Surprisingly, the effect of spill-out significantly increases the imaginary part of the mode index, and for wide slabs approaching bulk, the increase is 20 %. This is explained in terms of strong plasmonic absorption taking place at narrow regions close to the surface, as the real part of the quantum dielectric function is zero at these positions, showing that the phenomena is a consequence of quantum spill-out. In addition, the plasmon mode index does not diverge in the limit of vanishing slab width, instead the plasmonic mode ceases to exist below a cut-off in the subatom range.

Finally, electron-energy loss spectroscopy (EELS) was studied in thin metal films within the random-phase approximation, thus giving rise to a full nonlocal response. Electrons can be incident perpendicular or parallel to the film, and they excite plasmons when their energy and momentum are the same, thereby losing an amount of energy. The energy loss calculated within a classical model is in good agreement with the loss obtained using JEL and ALP, showing that nonlocality is a relatively small effect for plasmons propagating in thin metal films. Importantly, in order to avoid spurious features in the reflection coefficient and loss functions, it is necessary to account for a varying effective mass of the occupied states within ALP. The small impact from nonlocality is explained in terms of the electron motion within the film, which mostly occurs in the direction parallel to the film, where the electrons behave as almost free. This is in stark contrast to e.g. metal clusters and spheres, which are confined in all three spatial directions, and where the effect of nonlocality is larger.

References

[1] L. Novotny and B. Hecht. *Principles of Nano-Optics*. (Cambridge University Press, 2012) 2nd edition.

[2] Andreas Otto. *Z. Physik*, 216, 398-410, (1968).

[3] Erwin Kretschmann. *Z. Physik*, 241, 313-324, (1971).

[4] R. Ritchie. *Phys. Rev.*, 106, 874-881, (1957).

[5] F. J. Garcia de Abajo. *Rev. Mod. Phys.*, 82, 209-275, (2010).

[6] E. Kröger. *Z. Phys.*, 216, 115-135, (1968).

[7] C. Powell and J. Swan. *Phys. Rev.*, 115, 869-875, (1959).

[8] K. Tsuei, E. Plummer, and P. Feibelman. *Phys. Rev. Lett.*, 63, 2256-2259, (1989).

[9] J. Daniels. *Z. Physik*, 203, 235-249, (1967).

[10] M. Rocca, L. Yibing, F. de Mongeot, and U. Valbusa. *Phys. Rev. B*, 52, 947-953, (1995).

[11] A. Goldman, V. Dose, and G. Borstel. *Phys. Rev. B*, 32, 1971-1980, (1985).

[12] J. Sass, H. Laucht, and K. Kliewer. *Phys. Rev. Lett.*, 35, 1461-1464, (1975).

[13] T. López-Ríos and G. Hincelin. *Phys. Rev. B*, 38, 3561-3563, (1988).

[14] M. Reutzel, A. Li, B. Gumhalter, and H. Petek. *Phys. Rev. Lett.*, 123, 017404, (2019).

[15] K. Giesen, F. Hage, F. Himpsel, H. Riess, and W. Steinmann. *Phys. Rev. Lett.*, 55, 300-303, (1985).

[16] S. I. Bozhevolnyi and J. Jung. *Opt. Express*, 16, 2676-2684, (2008).

[17] S. I. Bozhevolnyi. *Opt. Express*, 14, 9467-9476, (2006).

[18] T. Søndergaard and S. I. Bozhevolnyi. *New. J. Phys.*, 15, 013034, (2013).

[19] F. J. García-Vidal, L. Martin-Moreno, T. Ebbesen, and L. Kuipers. *Rev. Mod. Phys.*, 82, 729, (2010).

[20] C. L. C. Smith, N. Stenger, A. Kristensen, N. A. Mortensen, and S. I. Bozhevolnyi. *Nanoscale*, 7, 9355-9368, (2015).

[21] T. Søndergaard and S. I. Bozhevolnyi. *Phys. stat. sol. (b)*, 245, 9-19, (2008).

[22] P. Berini. *Adv. Opt. Photon.*, 1, 484-588, (2009).

[23] D. Sarid. *Phys. Rev. Lett.*, 47, 1927-1930, (1981).

[24] E. N. Economou. *Phys. Rev.*, 182, 539-554, (1969).

[25] A. Yanai and U. Levy. *Opt. Express*, 17, 14270, (2009).

[26] D. C. Marinica, M. Zapata, P. Nordlander, A. K. Kazansky, P. M. Echenique, J. Aizpurua, and A. G. Borisov. *Sci. Adv.*, 1, 1501095, (2015).

[27] G. Haberfehlner, F.-P. Schmidt, G. Schaffernak, A. Hörl, A. Trügler, A. Hohenau, F. Hofer, J. Krenn, U. Hohenester, and G. Kothleitner. *Nano Lett.*, 17, 11, 6773-6777, (2017).

[28] T. Søndergaard and S. I. Bozhevolnyi. *New J. Phys.*, 15, 013034, (2013).

[29] A. Roberts, T. Søndergaard, M. Chirumamilla, A. Pors, J. Beermann, K. Pedersen, and S. Bozhevolnyi. *Phys. Rev. B*, 93, 075413, (2016).

[30] M. Odgaard, M. G. Laursen, and T. Søndergaard. *J. Opt. Soc. Am. B*, 31, 1853-1860, (2014).

[31] T. Søndergaard, S. Novikov, T. Holmgaard, R. Eriksen, J. Beermann, Z. Han, K. Pedersen, and S. I. Bozhevolnyi. *Nat. Commun.*, 3, 969, (2012).

[32] R. Maniyara, D. Rodrigo, R. Yu, J. Canet-Ferrer, D. Ghosh, R. Yongsunthon, D. Baker, A. Rezikyan, F. J. García de Abajo, and V. Pruneri. *Nat. Photon.*, 13, 328, (2019).

[33] T. Søndergaard. *Green's Function Integral Equation Methods in Nano-Optics*. (CRC Press, Boca Raton, 2019) 1st edition.

[34] J. Greffet, R. Carminati, K. Joulain, J. Mulet, S. Mainguy, and Y. Chen. *Nature*, 416, 61-64, (2002).

[35] H. Miyazaki, K. Ikeda, T. Kasaya, K. Yamamoto, Y. Inoue, K. Fujimura, T. Kanakugi, M. Okada, K. Hatade, and S. Kitagawa. *Appl. Phys. Lett.*, 92, 141114, (2008).

[36] H. Sai and H. Yugami. *Appl. Phys. Lett.*, 85, 3399-3401, (2004).

[37] Thomas Bauer. *Thermophotovoltaics – Basic Principles and Critical Aspects of System Design*. (Springer Verlag, Berlin, 2011) 1st edition.

[38] E. Skovsen, T. Søndergaard, C. Lemke, T. Holmgaard, T. Leißner, R. Eriksen, J. Beermann, M. Bauer, K. Pedersen, and S. Bozhevolnyi. *Appl. Phys. Lett.*, 103, 211102, (2013).

[39] T. Søndergaard and S. I. Bozhevolnyi. *Opt. Lett.*, 41, 2903-2906, (2016).

[40] A. C. Phillips. *Introduction to quantum mechanics*. (Wiley, 2003) 1st edition.

[41] P. Hohenberg and W. Kohn. *Phys. Rev.*, 136, 864-871, (1964).

[42] W. Kohn and L. J Sham. *Phys. Rev.*, 140, 1133-1138, (1965).

[43] Jens P. Dahl. *Introduction to the quantum world of atoms and molecules*. (World Scientific, London, 2001) 1st edition.

[44] J. Kohanoff. *Electronic structure calculations for solids and molecules - Theory and computational methods*. (Cambridge University Press, 2006) 1st edition.

[45] N. D. Lang and W. Kohn. *Phys. Rev. B*, 1, 4555-4568, (1970).

[46] N. D. Lang and W. Kohn. *Phys. Rev. B*, 3, 1215-1223, (1971).

[47] L. Zhigilei, A. Volkov, and A. Dongare. *Computational Study of Nanomaterials: From Large-Scale Atomistic Simulations to Mesoscopic Modeling*. (Springer, Netherlands, 2012) pp. 470-480.

[48] D. Frenkel and B. Smit. *Understanding Molecular Simulation: From Algorithms to Applications*. (Academic, San Diego, 1996).

[49] X. Chen, P. Liu, and L. Jensen. *J. Phys. D: Appl. Phys.*, 52, 363002, (2019).

[50] M. Barbry, P. Koval, F. Marchesin, R. Esteban, A. Borisov, J. Aizpurua, and D. Sánchez-Portal. *Nano Lett.*, 15, 5, 3410-3419, (2015).

[51] M. Kuisma, A. Sakko, T. Rossi, A. Larsen, J. Enkovaara, L. Lehtovaara, and T. Rantala. *Phys. Rev. B*, 91, 115431, (2015).

[52] D. Ceperley. *Phys. Rev. B*, 18, 3126-3138, (1978).

[53] D. Ceperley and B. Adler. *Phys. Rev. Lett.*, 45, 566-569, (1980).

[54] J. Perdew and A. Zunger. *Phys. Rev. B*, 23, 5048-5079, (1981).

[55] S. Vosko, L. Wilk, and M. Nusair. *Can. J. Phys*, 58, 1200, (1980).

[56] Y. Foucaud, M. Badawi, L. Filippov, I. Filippova, and S. Lebegue. *Min. Eng.*, 143, 106020, (2019).

[57] J. Perdew and Y. Wang. *Phys. Rev. B*, 45, 23, (1992).

[58] J. Perdew, K. Burke, and M. Ernzerhof. *Phys. Rev. Lett.*, 77, 18, (1996).

[59] M. Chebbi, S. Chibani, J. Paul, L. Cantrel, and M. Badawi. *Micropor. Mesopor. Mater.*, 239, 111-122, (2017).

[60] E. Hessou, H. Jabraoui, M. Hounguè, J. Mensah, M. Pastore, and M. Badawi. *Crystal Mater.*, 234, 469-482, (2019).

[61] G. Csonka, J. Perdew, A. Ruzsinszky, P. Philipsen, S. Lebègue, J. Paier, O. Vydrov, and J. Ángyán. *Phys. Rev. B*, 79, 155107, (2009).

[62] S. Lebègue and O. Eriksson. *Phys. Rev. B*, 79, 115409, (2009).

[63] T. Bučko, S. Lebègue, J. Hafner, and J. Ángyán. *J. Chem. Theory Comput.*, 9, 4293-4299, (2013).

[64] M. van der Kamp, J. Zurek, F. Manby amd J. Harvey, and A. Mulholland. *J. Phys. Chem. B*, 114(34), 11303-11314, (2010).

[65] M. van der Kamp and A. Mulholland. *Biochemistry*, 52(16), 2708-2728, (2013).

[66] J. Liang, Y. Xu, R. Liu, and X. Zhu. *Sci.*, 6, 213, (2019).

[67] O. M. Bakr, V. Amendola, C. M. Aikens, W. Wenseleers, R. Li, L. D. Negro, G. C. Schatz, and F. Stellacci. *Angew. Chem.*, 48, 5921-5926, (2009).

[68] M. Zhu, C. M. Aikens, F. J. Hollander, G. C. Schatz, and R. Jin. *J. Am. Chem. Soc.*, 130, 5883-5885, (2008).

[69] W. Yan, M. Wubs, and N. A Mortensen. *Phys. Rev. Lett.*, 115, 137403, (2015).

[70] T. V. Teperik, P. Nordlander, J. Aizpurua, and A. G. Borisov. *Phys. Rev. Lett.*, 110, 263901, (2013).

[71] L. Stella, P. Zhang, F. J. García-Vidal, A. Rubio, and P. García-Gonzalez. *J. Phys. Chem. C*, 117, 8941-8949, (2013).

[72] G. Maroulis. *J. Chem. Phys.*, 121, 10519-10524, (2004).

[73] J. Zuloaga, E. Prodan, and P. Nordlander. *Nano Lett.*, 9, 887, (2009).

[74] H. Zhang, V. Kulkarni, E. Prodan, P. Nordlander, and A. O. Govorov. *J. Phys. Chem. C*, 118, 16035-16042, (2014).

[75] M. Nezami G. Hajisalem and R. Gordon. *Nano Lett.*, 14, 6651, (2014).

[76] C. David and F. J. Garcia de Abajo. *ACS Nano*, 8, 9558-9566, (2014).

[77] K. J. Savage, M. M. Hawkeye, R. Esteban, A. G. Borisov, J. Aizpurua, and J. J. Baumberg. *Nature (London)*, 491, 574, (2012).

[78] F. J. Garcia de Abajo. *J. Phys. Chem. C*, 112, 17983, (2008).

[79] W. Zhu, R. Esteban, A. G. Borisov, J. J. Baumberg, P. Nordlander, H. J. Lezec1, J. Aizpurua, and K. B. Crozier. *Nat. Commun.*, 7, 11495, (2016).

[80] R. Esteban, A. G. Borisov, P. Nordlander, and J. Aizpurua. *Nat. Commun.*, 3, 825, (2012).

[81] R. Esteban, A. Zugarramurdi, P. Zhang, P. Nordlander F. García-Vidal, A. Borisov, and J. Aizpurua. *Faraday Discuss.*, 178, 151-183, (2015).

[82] A. Boardman. *Electromagnetic surface modes: Hydrodynamic theory of plasmon-polaritons on plane interfaces*. (John Wiley and Sons, Chichester, 1982).

[83] S. Bozhevolnyi, L. Martin-Moreno, and F. Garcia-Vidal. *Quantum plasmonics*. (Springer, Switzerland, 2017) 1st edition.

[84] S. Raza, G. Toscano, A. Jauho, M. Wubs, and N. A. Mortensen. *Phys. Rev. B*, 84, 121412(R), (2011).

[85] S. Raza, S. Bozhevolnyi, M. Wubs, and N. A. Mortensen. *J. Phys. Condens. Matter*, 27, 183204, (2015).

[86] T. Christensen, W. Yan, S. Raza, A. Jauho, N. A. Mortensen, and M. Wubs. *ACS Nano*, 8, 1745-1758, (2014).

[87] W. Yan, N. A. Mortensen, and M. Wubs. *Phys. Rev. B*, 88, 155414, (2013).

[88] P. B. Johnson and R. W. Christy. *Phys. Rev. B*, 6, 4370, (1972).

[89] L. Hedin and S. Lundqvist. *Solid State Physics*. (Academic Press, 1970).

[90] J. Lindhard. *Dan. Vidensk. Selsk. Mat. Fys. Medd.*, 28, 8, (1954).

[91] H. Ehrenreich and M. Cohen. *Phys. Rev.*, 115, 786-790, (1959).

[92] N. Wiser. *Phys. Rev. B*, 129, 62-69, (1963).

[93] S. L. Adler. *Phys. Rev. B*, 126, 413-420, (1962).

[94] N. Mermin. *RPB*, 1, 2362-2363, (1970).

[95] R. Ritchie and L. Marusak. *Surf. Sci.*, 4, 234-240, (1966).

[96] O. Keller. *Phys. Rev. B*, 33, 990-1009, (1986).

[97] P. J. Feibelman. *Phys. Rev. B*, 12, 1319-1336, (1975).

[98] L. Marusic and M. Sunjic. *Phys. Script.*, 63, 336-341, (2001).

[99] Z. Penzar and M. Sunjic. *Phys. Script.*, 30, 431-443, (1984).

[100] S. de Vega and F. J. Garcia de Abajo. *ACS Photon.*, 4, 2367-2375, (2017).

[101] A. Rodríguez Echarri, J. D. Cox, and F. J. Garcia de Abajo. *Optica*, 6, 630-641, (2019).

[102] F. J. Garcia de Abajo. *ACS Nano*, 7, 11409-11419, (2013).

[103] V. Nazarov, V. Silkin, and E. Krasovskii. *Phys. Rev. B*, 93, 035403, (2016).

[104] E. Chulkov, V. Silkin, and P. Echenique. *Surf. Sci.*, 437, 330-352, (1999).

[105] A. V. Lavrinenko, J. Lægsgaard, N. Gregersen, F. Schmidt, and T. Søndergaard. *Numerical methods in photonics*. (CRC Press, Boca Raton, 2015) 1st edition.

[106] R. Barrett, M. Berry, T. Chan, J. Demmel, J. Donato, J. Dongarra, V. Eijkhout, R. Pozo, C. Romine, and H. Vorst. *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods*. (SIAM, 1994) 2nd edition.

[107] Charles Kittel. *Introduction to solid state physics*. (John Wiley & Sons, 2005) 8th edition.

[108] J. Reitz, F. Milford, and R. Christy. *Foundations of electromagnetic theory*. (Pearson, New York, 2009) 4th edition.

[109] Middle East technical university Department of physics. *Lecture XII: Exchange and correlation*. http://www.physics.metu.edu.tr/ hande/teaching/741-lectures/lecture-12.pdf. Last visited December 23, 2019.

[110] M. V. Klein and T. E. Furtak. *Optics*. (John Wiley & Sons, New York, 1986) 2nd edition.

[111] R. Petersen, T. G. Pedersen, M. N. Gjerding, and K. S. Thygesen. *Phys. Rev. B*, 035128, (2016).

[112] N. A. Mortensen, S. Raza, M. Wubs, T. Søndergaard, and S. I. Bozhevolnyi. *Nat. Commun.*, 5, 3809, (2014).

[113] M. R. Thomsen, S. J. Bruun, and T. G. Pedersen. *J. Phys. Condens. Matter*, 26, 335301, (2014).

[114] F. Bonabi, S. J. Bruun, and T. G. Pedersen. *Phys. Rev. B*, 96, 155419, (2017).

[115] S. Boriskina, T. Cooper, L. Zeng andG. Ni, J. Tong, Y. Tsurimaki, Y. Huang, L. Merouoeh, G. Mahan, and G. Chen. *Adv. Opt. Phot.*, 9,775-827, (2017).

[116] J. Khurgin. *Nat. Nanotech.*, 10, 2-6, (2015).

[117] R. Zia, M. Selker, and M. Brongersma. *Phys. Rev. B*, 71, 165431, (2005).

[118] B. Lambrecht, J. Krenn, G. Schider, H. Ditlbacher, M. Salerno, N. Felidj, A. Leitner, F. Ausseness, and J. Weeber. *Appl. Phys. Lett.*, 79, 51, (2001).

[119] X. Li, A. Teng, M. Özer, J. Shen, H. Weitering, and Z. Zhang. *New J. Phys.*, 16, 065014, (2014).

[120] P. Schwarz and J. Schwarz. *Special relativity - from Einstein to strings*. (Cambridge University Press, 2004) 1st edition.

[121] V. Silkin, J. Pitarke, E. Chulkov, and P. Echenique. *Phys. Rev. B*, 721 115435, (2005).

[122] F. Reinert, G. Nicolay, S. Schmidt, D. Ehm, and S. Hüfner. *Phys. Rev. B*, 63, 115415, (2001).

[123] R. Paniago, R. Matzdorf, G. Meister, and A. Goldmann. *Surf. Sci.*, 336, 113, (1995).

[124] A. Garcia-Lekue, J. Pitarke, E. Chulkov, A. Liebsch, and P. Echenique. *Phys. Rev. B*, 68, 045103, (2003).

[125] F. J. Garcia de Abajo, A. Rivacoba, N. Zabala, and N. Yamamoto. *Phys. Rev. B*, 69, 155420, (2004).

Publications

List of publications

[I] **Enok J. H. Skjølstrup** and Thomas Søndergaard, Optics of multiple ultrasharp grooves in metal, J. Opt. Soc. Am. B **34**, 673 (2017).

[II] **Enok J. H. Skjølstrup**, Thomas Søndergaard, Kjeld Pedersen, and Thomas G. Pedersen, Optics of multiple grooves in metal: transition from high scattering to strong absorption, J. Nanophotonics, **11**(4), 046023 (2017).

[III] **Enok J. H. Skjølstrup**, Thomas Søndergaard, and Thomas G. Pedersen, Quantum spill-out in few-nanometer metal gaps: Effect on gap plasmons and reflectance from ultrasharp groove arrays, Phys. Rev. B **97**, 115429 (2018).

[IV] **Enok J. H. Skjølstrup**, Thomas Søndergaard, and Thomas G. Pedersen, Quantum spill-out in nanometer-thin gold slabs: Effect on the plasmon mode index and the plasmonic absorption, Phys. Rev. B **99**, 155427 (2019).

[V] A. R. Echarri, **Enok J. H. Skjølstrup**, Thomas G. Pedersen, and F. Javier Garcia de Abajo, Theory of EELS in atomically thin metallic films, submitted to Phys. Rev. Res. Available on ArXiv:1912.09414.

Paper I

Optics of multiple ultrasharp grooves in metal

Enok J. H. Skjølstrup and Thomas Søndergaard

The paper has been published in
Journal of the optical society of America B **34**, 673-680 (2017).

The layout has been revised.

Journal of the Optical Society of America B
OPTICAL PHYSICS

Optics of multiple ultrasharp grooves in metal

ENOK J. H. SKJØLSTRUP* AND THOMAS SØNDERGAARD

Department of Physics and Nanotechnology, Aalborg University, Skjernvej 4A, DK-9220 Aalborg East, Denmark
**Corresponding author: ejs@nano.aau.dk*

Received 13 January 2017; revised 2 February 2017; accepted 6 February 2017; posted 7 February 2017 (Doc. ID 284438); published 27 February 2017

The optics of multiple ultrasharp sub-wavelength grooves in metal is studied theoretically. Focus is on the transition from a single groove, where the scattering cross section is significant and can exceed the groove width, to infinitely many grooves in a periodic array with very low reflectance. When the multiple-groove array is illuminated by a plane wave the out-of-plane scattering is found to be extraordinarily large compared with the expected maximum from a geometric-optics estimate even for array widths of many wavelengths. The out-of-plane scattering is even higher per groove compared to the single-groove case. This is explained as an effect of surface plasmon polaritons (SPPs) generated at one groove being scattered out of the plane by other grooves. This is supported by studies of the transmittance, reflectance, and out-of-plane scattering, when an SPP is incident on multiple grooves. When illuminating instead with a Gaussian beam, and observing the limit where the beam is confined well within the multiple-groove array, the total reflectance is very low and practically no scattering occurs.

OCIS codes: (050.0050) Diffraction and gratings; (240.6680) Surface plasmons; (290.5825) Scattering theory; (260.3910) Metal optics.

https://doi.org/10.1364/JOSAB.34.000673

1. INTRODUCTION

Recently, it has been found that a single sub-wavelength ultrasharp groove in metal may have a significant out-of-plane scattering cross section [1]. On the other hand a periodic array of the same grooves can transform a shiny and highly reflecting metal into a black material with broadband absorption and thus very low reflectance of light [2,3]. The optics of the periodic array of grooves thus does not follow straightforwardly from the optics of a single groove. It has been suggested in [1] that the lack of reflection for the periodic array can be explained as an effect of mutual (destructive) interference between scattered waves originating from each groove in the array. In this paper we test this hypothesis and explore the transition from one to infinitely many grooves. We will show among other things that in the case of plane-wave illumination of multiple grooves the out-of-plane scattering actually does not cancel out as previously hypothesized, while it holds true in the case of illumination with a Gaussian beam having a width being smaller than the multiple-groove array.

While we have chosen in this paper to consider ultrasharp grooves the phenomenon of near-unity absorption of periodic arrays of grooves in metal also occurs in the case of rectangular grooves, and for other tapered grooves that are not sharp [4–6]. In those cases, however, the near-unity absorption occurs only for a narrow band of wavelengths where the groove acts as an optical resonator. Similar to our case the single groove of this type in metal may also have a large scattering cross section (at resonance) [4], and again the optics of the periodic array does not follow straightforwardly from the optics of a single groove.

One possible application of arrays of ultrasharp grooves is as polarizers for ultrashort laser pulses exploiting that one polarization of light will be efficiently absorbed, while the other polarization will be efficiently reflected, and with negligible dispersion in the reflection process [7]. The rectangular grooves or other grooves that are not sharp may on the other hand be used for selective thermal emitters that predominantly emit light with narrow bands of wavelengths [8,9], which is of interest for thermophotovoltaics [10,11]. Broadband omnidirectional absorbers and angularly selective emitters can also be constructed using arrays of tapered grooves in metal [12]. It is also possible to make broadband absorbers using crossed-groove arrays that are not tapered much if grooves are made deep enough [13], while other crossed-groove arrays with selective absorption can be used for all-metal structural color printing [14].

The structure of interest in this paper is illustrated in Fig. 1. Here, an incident p-polarized Gaussian beam with beam waist radius w_0 is centered at $x = x_0$ being in the middle of an array of N identical ultrasharp grooves, where each groove is characterized by a top width of 240 nm, a depth of 500 nm,

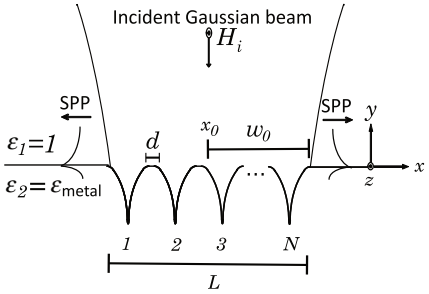

Fig. 1. Illustration of N identical ultrasharp grooves in metal separated by the distance d. The grooves are illuminated by a normally incident Gaussian beam with beam waist radius w_0 centered in the middle of the groove array ($x = x_0$).

a bottom groove width of 0.3 nm, and parallel groove walls near the groove bottom. The bottom of the groove is rounded by a small circle of radius 0.15 nm. The period of the finite array of grooves is 250 nm such that edges of neighbor grooves are separated by a distance $d = 10$ nm (Fig. 1). Except for a marginal 10 nm difference in the top groove width the grooves are identical to those considered in [1]. The total length of the finite groove array is denoted L. The beam waist radius can be related to the array length as $w_0 = \gamma L/2$, where γ is a ratio parameter of beam waist radius to half the length of the grooves. Figure 1 is illustrated with $\gamma = 1$. It will be assumed that the structure and fields are invariant along the z-axis (2D calculations). The magnetic field thus only has a z-component ($\mathbf{H}(\mathbf{r}) = \hat{z}H(\mathbf{r}) = \hat{z}H(x, y)$). The incident light can be either reflected or scattered upward, absorbed in the metal, or scattered into surface plasmon polaritons (SPPs), which are electromagnetic waves bounded to and propagating along the metal surface.

All calculations presented throughout the paper apply the dielectric constant of gold from [15] and have been made using the Greens function surface integral equation method (GFSIEM) outlined in Appendix B of [4]. The case of large N was handled by using the iterative solution method GMRES (see, e.g., [16,17] for a further description). The periodic nature of the structure was taken advantage of when constructing the matrix equation by using that the coupling between two grooves in the array only depends on their relative position. This implies setting up the matrix scales as $2N$ instead of N^2.

The paper is organized in the following way. In Section 2 the incident field is a plane wave ($w_0 = \infty$) and the extinction, scattering, and absorption cross sections are presented for a structure consisting of a varying number of grooves. In Section 3 the case of a finite beam waist of the incident Gaussian beam is considered, and the transition to the case of the beam waist diameter being smaller than the groove array length L is observed. Here, we instead calculate angular reflection spectra and total out-of-plane reflected power. In Section 4 the incident field is a plasmon and the reflectance, transmittance, and out-of-plane scattering is studied as a function of the number of grooves in the array. We offer our conclusions in Section 5. Appendix A outlines the theory applied in the paper.

2. PLANE WAVE AS INCIDENT FIELD

In this section we consider the situation in Fig. 1 in the limit of infinite beam waist ($w_0 \to \infty$), such that the incident field is a plane wave. Due to scattering by the multiple grooves some light will be coupled into the SPPs propagating away from the grooves along the metal surface, some light will be scattered out of the plane, and some additional light will be absorbed. The amount of power removed from the reflected beam due to scattering and absorption by the grooves will be referred to as extinction (EXT). By normalizing the power scattered into the SPPs, the power scattered out of the plane, and the extinction power, with the power per unit area of the incident plane wave, we obtain the corresponding EXT, out-of-plane scattering (OUP), and SPP optical cross sections. In addition we will need the differential out-of-plane scattering cross section being the out-of-plane scattered power per unit angle again normalized with the incident power per unit area. The total OUP cross section is obtained by integrating the differential cross section over all relevant angles. In addition, the absorption cross section (ABS) can be defined as the extinction cross section minus out-of-plane and SPP scattering cross sections.

We consider first the EXT, OUP, and ABS optical cross sections for a fixed wavelength of 770 nm in Fig. 2(a). These cross sections are almost linear functions of the number of grooves. In addition, the scattering cross section is larger than the absorption cross section for this wavelength, while the opposite is the case for a wavelength of, e.g., 550 nm. For the case of $N = 40$ grooves the total length of the grooves is $L \approx 10$ μm

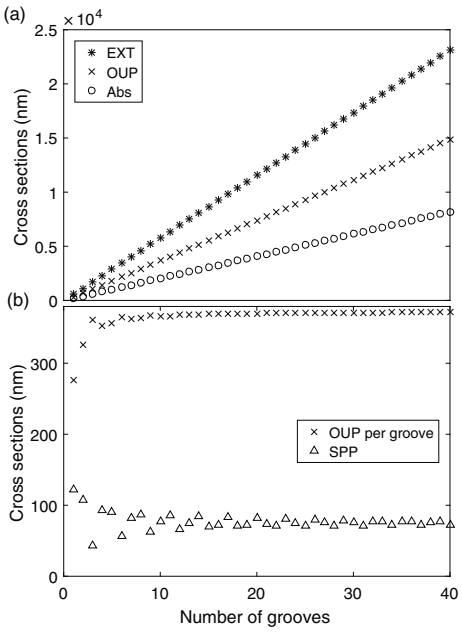

Fig. 2. (a) EXT, OUP, and ABS cross sections. (b) SPP and OUP cross section per groove. The wavelength is 770 nm in both (a) and (b).

but the OUP cross section is seen to be approximately 15 μm, and thus exceeds the length L by a factor of approximately 1.5. This can also be seen from Fig. 2(b) where the OUP cross section is normalized with the number of grooves. The OUP cross section per groove converges to approximately 370 nm, which is thus approximately 1.5 times the period between grooves. It should be noted that in the case of infinitely many grooves arranged in a periodic array, and being illuminated by a plane wave of the same wavelength, the reflectance will be less than 16% [2], and finding such a large scattering cross section is thus a surprising and remarkable result. It is not unusual that a subwavelength plasmonic scatterer can have optical cross sections that greatly exceed the physical width of the scatterer [1]. However, here the physical length L is more than 13 wavelengths (for $\lambda = 770$ nm), and for a structure of this size it is normally assumed that geometric optics is applicable. The observed scattering for the multiple grooves with large L (or N) is thus extraordinarily large.

The SPP cross section as a function of number of grooves can be seen by the triangles in Fig. 2(b) and is a damped oscillatory function. More grooves do not generally imply a larger SPP cross section, which can be due to two different phenomena: the plasmons excited in different grooves can interfere destructively, and a plasmon excited at one groove can be coupled out of the plane by another groove as studied in [18–20]. Note that the OUP cross section per groove is increasing for a small number of grooves and converges for approximately 12 grooves at a value that is higher than for a single groove. This can be explained from SPPs excited at one groove being coupled out of the plane at other grooves, which leads to an overall increase in the out-of-plane scattering per groove for the case of multiple grooves. This out-coupling of SPPs is studied in more detail in Section 4.

Note that if the purpose of the array of grooves would have been to efficiently excite SPPs with normally incident light this would require instead a period close to the plasmon wavelength of interest [21,22]. Furthermore, the dimensions of the individual grooves in an array can be optimized as examined in [23,24] where it was found that most of the incident light can be excited into SPPs for optimal groove parameters.

While Fig. 2(a) only considered a wavelength of 770 nm, the EXT cross sections as functions of the wavelength is seen in Fig. 3(a) for 1, 5, 10, and 20 grooves. The cross section is found to scale approximately linear with the number of grooves for all wavelengths. The same applies for the OUP and ABS cross sections (not shown). Furthermore, the peaks and valleys in spectra do not shift much when more grooves are present. The spectra are thus highly similar to that of a single groove scaled up by a factor N. The SPP cross section spectra for different numbers of grooves are seen in Fig. 3(b), and also here peaks and valleys do not shift much with N. But in general no scaling occurs as for the other cross sections, as was also observed in Fig. 2(b). For some other groove dimensions than considered in this paper the cross sections of a single groove are studied in detail in [1], establishing several fundamental results. Based on the linear scaling of cross sections versus number of grooves observed in Figs. 2(a) and 3(a) for our specific groove dimensions, we have reason to believe that the same linear scaling is valid for other groove dimensions.

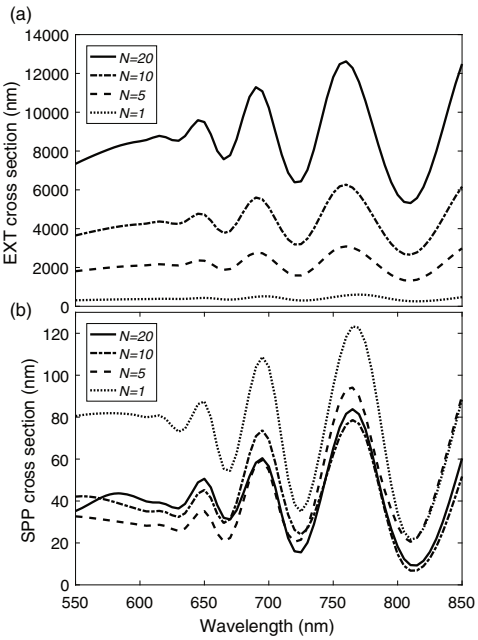

Fig. 3. (a) EXT cross sections for 1, 5, 10, and 20 grooves. (b) SPP cross sections for 1, 5, 10, and 20 grooves.

In Fig. 2(a) it was observed that the OUP cross section increases almost linearly with the number of grooves. To further study the scattering from multiple grooves, the differential OUP cross section is shown in Fig. 4. The differential cross sections for different N have been normalized such that they all have a maximum of 1. It is clearly seen that the angular distribution becomes more narrow when more grooves are present.

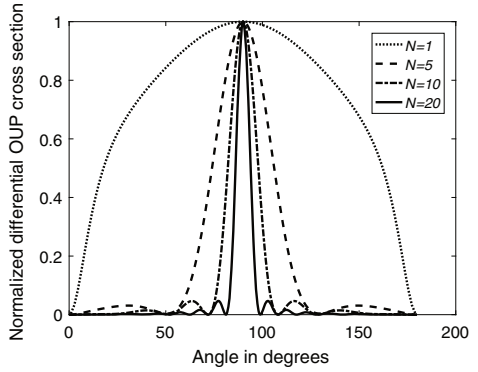

Fig. 4. Normalized differential OUP cross sections for 1, 5, 10, and 20 grooves for $\lambda = 770$ nm.

3. GAUSSIAN BEAM AS INCIDENT FIELD

In this section we consider again the situation in Fig. 1 but now with a finite beam waist radius $w_0 = \gamma L/2$. We will consider the effect of varying the ratio parameter γ and observe the effect of using beams that are both wider and more narrow than the structure. Here, we shall no longer use optical cross sections since the incident field is not a plane wave. Instead the incident field has a well-defined power, and the same goes for the power coupled into SPPs, and the total power reflected into the upper half-plane. Total reflected power refers to the total power reflected or scattered into any out-of-plane direction θ in the upper half-plane.

The reflected power per angle [the differential of Eq. (A10) in Appendix A] is seen in black in Fig. 5 as a function of angle for a structure consisting of 20 grooves when γ is 0.5, 2, and 100. A relatively wide structure with 20 grooves was chosen such that these values of γ correspond to paraxial beams. The spectra have been normalized by the maximum incident power per angle. This is also equivalent to the maximum of the reflected power versus angle when the incident field is reflected by a perfect mirror. The incident differential power is seen by the bright lines (blue online) in the same figure for $\gamma = 0.5$ and 2. When $\gamma = 100$ the beam is so wide that most of the incident field will be hitting the flat metal surface outside the grooves and the grooves only marginally change the total reflection. In this case the incident power is almost the same as the reflected power shown as the black solid line, except that it peaks at the value 1, while the black solid line has a smaller peak value due to the reflectance from gold being slightly below 100%.

The reflectance being the ratio of total reflected and incident power [Eq. (A12) in Appendix A] can be found as the area under the black curve divided by the area under the corresponding blue curve, and is found to be 0.153, 0.408, and 0.96 when $\gamma = 0.5$, 2, and 100, respectively. The incident field behaves almost like a plane wave for $\gamma = 100$, which is why the reflected power is very narrow in angular distribution around 90°. When γ gets smaller the angular power distributions become broader. When $\gamma = 0.5$ the Gaussian beam is entirely focused within the grooves, and the calculated reflectance is therefore entirely due to the grooves and not the planar interface outside the grooves as is partly the case for a wider Gaussian beam.

Importantly notice the difference between the spectra in Figs. 4 and 5. In Fig. 4 the incident field is a plane wave and only the squared amplitude of the scattered field is shown, while the incident field is a Gaussian beam in Fig. 5 and the squared amplitude of the total reflected field is shown, which is the sum of the scattered field and the reflected field due to the reference structure [Eq. (A9) in Appendix A].

The reflectance from 20 grooves at a wavelength of 770 nm as a function of γ is seen by the asterisks in Fig. 6 and is seen to converge for both small and large ratio. When the ratio gets small the reflectance converges to the proper reflectance that is only due to the scattering structure itself, and not the surrounding planar interface, and this is found to be the case when $\gamma \leq 0.5$. On the other hand, when the ratio gets large the reflectance converges to that of a planar interface. Notice that the γ-axis in the figure is on a log scale, and when the ratio is 500 the reflectance is almost the same as for a planar interface shown by the horizontal line, which was also partly observed for $\gamma = 100$ in Fig. 5.

The geometric fraction of the incident power that hits the scatterer depends strongly on γ as seen by the triangles in the same figure. For $\gamma \leq 0.5$ the Gaussian beam is entirely focused within the groove array and the geometric fraction is practically 1, which implies that the reflectance is only due to the groove array. On the other hand, for $\gamma \geq 500$ the Gaussian beam is so wide that the groove array becomes negligible and the geometric fraction is practically 0, which implies that the reflectance becomes that of a flat interface. Notice that the reflectance and power fraction converge for the same γ.

A. Reflectance as a Function of Number of Grooves

This subsection considers the question of how the total reflectance varies with the number of grooves for a fixed beam waist radius w_0, which we shall set to 1250 nm. The reflectance as a

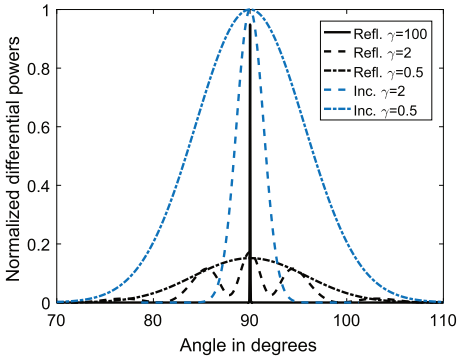

Fig. 5. Normalized differential incident power $dP_i/d\theta$ for varying γ is seen by the bright lines (blue online) while normalized differential reflected power $dP_r/d\theta$ is in black. The scattering structure consists of 20 grooves and the wavelength is 770 nm.

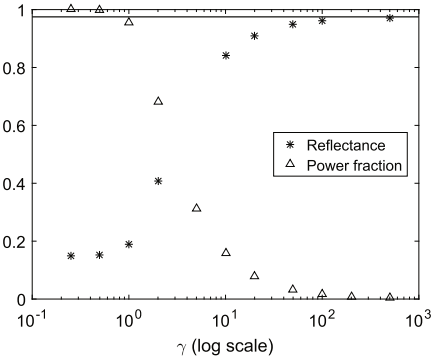

Fig. 6. Reflectance as a function of γ for 20 grooves at a wavelength of 770 nm, and corresponding geometric fraction of incident power that hits the grooves. The horizontal line is the reflectance of a planar gold interface.

function of number of grooves for $\lambda = 770$ nm is seen in Fig. 7(a) and is seen to have converged when 20 grooves are present. Here, 20 grooves is equivalent to $\gamma = 0.5$. Also shown is the geometric fraction of the incident light that hits the scattering structure, and this fraction has converged to 1 when 20 grooves are present. Hence, the reflectance converges when all the incident light hits the grooves and not partly the planar interface, which was also observed in Fig. 6. Thus, 20 grooves are sufficient to obtain the same total reflectance as for infinitely many grooves.

While Fig. 7(a) only considered the reflectance at $\lambda = 770$ nm, the reflectance as a function of wavelength is seen in Fig. 7(b) for a structure consisting of 5, 10, and 20 grooves, still with a fixed beam waist radius of 1250 nm. The reflectance has converged in the case of 20 grooves, as the same reflectance is obtained using 30 grooves (not shown). In the case of 20 grooves we find the same reflectance as has been previously obtained for a plane wave incident on infinitely many grooves arranged in a periodic array [2,3]. It should also be noted that as γ is made large the scattered fields and SPP fields become identical to those obtained with plane-wave incidence. The reflectance for several other groove dimensions is also considered in [2,3], and we have reason to believe that the same phenomena apply for these dimensions, namely that the same reflectance is found when using a Gaussian beam entirely focused within a finite number of grooves.

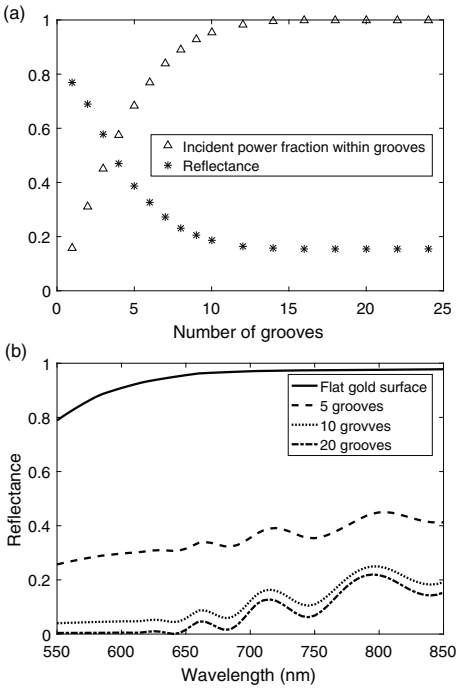

Fig. 7. (a) The reflectance as a function of number of grooves and the geometrical power fraction within the grooves for a wavelength of 770 nm. (b) Reflectance as a function of wavelength for a scatterer consisting of 5, 10, and 20 grooves. The beam waist radius is fixed at 1250 nm in both (a) and (b).

B. Energy Transportation

In this section we consider the time-averaged Poynting vector $\langle \mathbf{S} \rangle = 1/2\,\mathrm{Re}(\mathbf{E} \times \mathbf{H}^*)$ defined in terms of its complex electric field $\mathbf{E}$ and magnetic field $\mathbf{H}$ of a Gaussian beam of different beam waist radius being incident on 20 grooves. The wavelength is fixed at 770 nm. The Poynting vector flux through a (closed) surface is equivalent to the power flow through the surface, and the Poynting vector is commonly interpreted as a vector pointing in the direction of power flow and with a magnitude describing the power flowing per unit area [25].

The magnitude of the time-averaged Poynting vector is shown in Fig. 8 for a structure consisting of 20 grooves when the incident field is a Gaussian beam with ratios 0.5, 1, 2, and ∞ (plane wave). The length of the groove array is illustrated by the black horizontal lines. Notice that the calculated fields consist of both the reference fields and the scattered fields. When $\gamma = 0.5$ the incident power fraction within the grooves is practically 1 (Fig. 6), which implies that only a small part of the incident light is scattered out of the plane. This is clearly seen in Fig. 8(a) where the magnitude of the Poynting vector is only non-zero for positions right above the middle of the grooves. As γ increases more light is scattered out of the plane, which is clearly seen in Figs. 8(b)–8(d). For an incident plane wave as in Fig. 8(d) the region with non-zero Poynting vector is not so different from the case with $\gamma = 2$ in Fig 8(c).

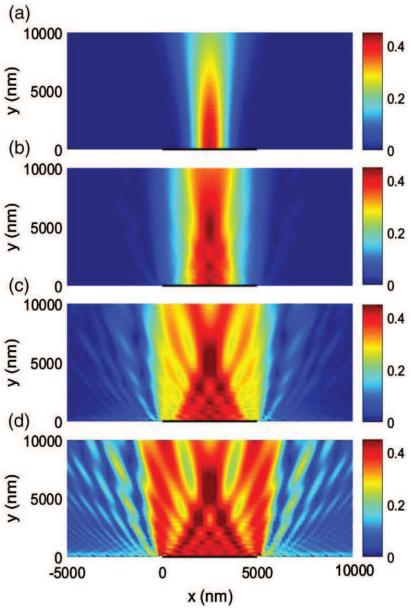

Fig. 8. Magnitude of time-averaged Poynting vector when a Gaussian beam is incident on a structure consisting of 20 grooves at a wavelength of 770 nm. The length of the groove array is illustrated by the black horizontal lines. The ratio is (a) $\gamma = 0.5$, (b) $\gamma = 1$, (c) $\gamma = 2$, and (d) $\gamma = \infty$ (plane wave).

This is because the part of a plane wave hitting the planar surface far outside the grooves will experience almost 100% reflectance implying a vanishing net power flow there. The angular distribution of the scattered field for $\gamma = \infty$ is broader compared with, e.g., $\gamma = 2$. The same phenomenon was partly observed in Fig. 5 as oscillations in the differential power for angles deviating from 90°.

4. PLASMON AS INCIDENT FIELD

In this section we consider an SPP being incident on multiple grooves. A schematic of this situation is shown in Fig. 9(a). Part of the SPP power will be reflected, transmitted, and scattered out of the plane. The reflectance, transmittance, and out-of-plane scattering of an SPP incident on multiple rectangular grooves has previously been considered in [18,19], while the case of a single groove or ridge of different shapes was considered in [20], and the case of a rectangular hole in [26]. The case

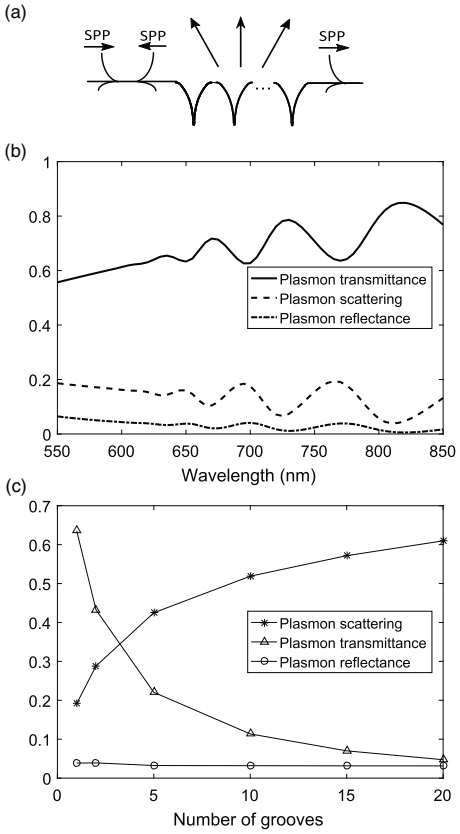

Fig. 9. (a) Schematic of an SPP wave incident from left on multiple grooves. The incident plasmon can be reflected, transmitted, scattered out of the plane, or absorbed in the metal. (b) Plasmon reflectance, transmittance, and scattering in the case of one groove as a function of wavelength. (c) Plasmon reflectance, transmittance, and scattering as a function of number of grooves for $\lambda = 770$ nm.

of short- and long-range SPPs of thin-metal films being incident on a rectangular metal nanostrip forming a gap-plasmon resonator together with the metal film has also been considered [27].

Figure 9(b) shows the reflectance, transmittance, and out-of-plane scattering for a plasmon being incident on a single groove as a function of wavelength. The reflectance is below 7% for all wavelengths; thus only a small part of a plasmon is reflected. The transmittance is much higher implying that most of the SPP power is transmitted across one groove, and below 20% is scattered out of the plane. The fact that a part of the incident plasmon is scattered out of the plane serves as a qualitative explanation of why the out-of-plane scattering per groove is higher for multiple grooves compared to the single groove case [Fig. 2(b)]. The groove dimensions are not optimized as in [23,24]; thus it is likely that another structure will result in a lower scattering out of the plane and a higher absorption level. Figure 9(c) shows the reflectance, transmittance, and out-of-plane scattering as a function of number of grooves at $\lambda = 770$ nm. The transmittance through one groove is 64% but decreases as more grooves are present and is only 5% for 20 grooves, while the out-of-plane scattering has increased to 61%. The reflectance is almost the same for any number of grooves. However, the out-of-plane scattering and absorption increase as more grooves are present.

5. CONCLUSION

In conclusion, the optics of multiple ultrasharp grooves in metal have been studied theoretically in order to investigate the transition from the case of a single groove to the case of infinitely many grooves arranged in a periodic array. It is found that when the incident field is a plane wave the OUP scattering cross section is almost a linear function of the number of grooves. Furthermore, for the considered example the OUP cross section per groove is approximately 1.5 times the period between grooves, and is thus higher than what could have been expected from a geometric-optics estimate of the maximum OUP cross section. Such a result is not unusual for a single sub-wavelength scatterer but here it is unusual and thus extraordinarily high because the total scattering structure has a width of many wavelengths.

The low reflectance of infinitely many grooves arranged in a periodic array and being illuminated with a plane wave can thus not be straightforwardly explained as an effect of mutual interference of waves originating from each groove. Instead, as the number of grooves is increased, an increasing scattered power is found within a decreasing angular interval. In addition, the OUP cross section per groove is higher for multiple grooves than for a single groove. This is explained as an effect of SPPs generated at one groove being coupled out of the plane by other grooves, which is supported by studies of the reflectance, transmittance, and scattering when an SPP is incident on multiple grooves.

When using instead a Gaussian beam as the incident field the reflectance was found to strongly depend on the width of the beam. In the case when the incident beam is sufficiently narrow to be located entirely within a groove array with many grooves, e.g., 20 grooves, there is practically no longer any

scattering, and low total reflectance is found being practically identical to the case of a plane wave being incident on the infinitely many grooves arranged in a periodic array.

APPENDIX A

In this appendix we will briefly present the theory applied for calculating the reflectance and optical cross sections.

We consider incident Gaussian beams with a beam profile $H_{0,i}(x, y = 0) = A\exp(-(x - x_0)^2/w_0^2)$ at $y = 0$, where A is the amplitude of the beam. By assuming a large enough beam waist radius w_0 that the paraxial approximation can be applied ($k_0 w_0 \gg 1$, where $k_0 = 2\pi/\lambda$ is the free-space wavenumber), and by using the angular spectrum representation [25] the incident magnetic field is found as

$$H_{0,i}(x,y) = Aw_0\sqrt{\frac{1}{w_0^2 - \frac{2iy}{k_0 n_1}}}e^{-ik_0 n_1 y}\exp\left(\frac{-(x-x_0)^2}{w_0^2 - \frac{2iy}{k_0 n_1}}\right), \quad y > 0. \tag{A1}$$

Here $n_1 = 1$ is the refractive index of air. When no grooves are present the corresponding reflected field will be

$$H_{0,r}(x,y) = Ar_{12}w_0\sqrt{\frac{1}{w_0^2 + \frac{2iy}{k_0 n_1}}}e^{ik_0 n_1 y}\exp\left(\frac{-(x-x_0)^2}{w_0^2 + \frac{2iy}{k_0 n_1}}\right), \quad y > 0, \tag{A2}$$

where $r_{12} = (n_2 - n_1)/(n_2 + n_1)$ is the reflection coefficient for p-polarized light incident normal to a planar interface between air and gold and n_2 is the refractive index of gold. The corresponding transmitted field will be

$$H_t(x,y) = At_{12}w_0\sqrt{\frac{1}{w_0^2 - \frac{2iy}{k_0 n_2}}}e^{-ik_0 n_2 y}\exp\left(\frac{-(x-x_0)^2}{w_0^2 - \frac{2iy}{k_0 n_2}}\right), \tag{A3}$$

where $t_{12} = 1 + r_{12}$ is the transmission coefficient from air into gold.

Hence, the reference field (no grooves) is given by

$$H_0(x,y) = \begin{cases} H_{0,i}(x,y) + H_{0,r}(x,y), & \text{for } y > 0 \\ H_t(x,y), & \text{for } y < 0 \end{cases}. \tag{A4}$$

This reference field is applied in the GFSIEM as presented in Appendix B of [4] for modeling of the grooves, and applying the GFSIEM gives first the magnetic field (H) and its normal derivative ($\hat{\mathbf{n}} \cdot \nabla H$) at the boundary of all the grooves, which is later used to calculate the scattered field. The field outside the grooves will be given on the form

$$H(x,y) = H_0(x,y) + H_{\text{scat}}(x,y), \tag{A5}$$

where the scattered part of the field is obtained as

$$H_{\text{scat}}(\mathbf{r}) = -\oint_{C_2}\{g(\mathbf{r},\mathbf{r}')\hat{\mathbf{n}}' \cdot \nabla' H(\mathbf{r}') - H(\mathbf{r}')\hat{\mathbf{n}}' \cdot \nabla' g(\mathbf{r},\mathbf{r}')\}\mathrm{d}l', \tag{A6}$$

where $\hat{n}$ is the outward surface normal vector and the Greens function is

$$g(\mathbf{r},\mathbf{r}') = \frac{i}{2\pi}\int_{k_x=0}^{\infty}\frac{t_{21}(k_x)\cos(kx(x-x'))e^{-ik_{y2}y' + ik_{y1}y}}{k_{y2}}\mathrm{d}k_x, \tag{A7}$$

with $k_{yi} = \sqrt{k_0^2\varepsilon_i - k_x^2}$, and $t_{21}(k_x) = 2\varepsilon_1 k_{y2}/(\varepsilon_1 k_{y2} + \varepsilon_2 k_{y1})$. The integration path C_2 in Eq. (A6) is on the outer surface of the grooves buried in the gold layer in the limit where the upper part of the grooves tends to air gold interface. For a further description see Appendix B in [4].

A far-field approximation ($k_0 r \gg 1$) of the scattered field valid at large distances above the grooves is available in [4]. The corresponding far-field approximation for the reference reflected field is found to be

$$H_{0,r}^{ff}(r,\theta) \approx Ar_{12}w_0\sqrt{\frac{k_0 n_1}{2r}}e^{-i\pi/4}e^{ik_0 n_1 r}e^{-i\delta} \times \exp\left(-\frac{(\pi/2-\theta)^2}{4}(k_0 n_1 w_0)^2\right), \tag{A8}$$

where $\theta \in [0;\pi]$ is the polar angle with $x = r\cos\theta$, $y = r\sin\theta$, and $\delta = k_0 n_1 x_0 \cos\theta$ is a phase accounting for the fact that the beam is centered at $x = x_0$. The last exponential in Eq. (A8) implies that the field is largest for $\theta = \pi/2$, which is in the vertical direction in Fig. 1, and decreases fast for other angles. The total reflected field in the far field is thus given by

$$H_{\text{tot},r}^{ff}(r,\theta) = H_{0,r}^{ff}(r,\theta) + H_{\text{scat}}^{ff}(r,\theta). \tag{A9}$$

The total out-of-plane reflected power is calculated using the flux of the time average of the Poynting vector $\langle\mathbf{S}\rangle = 1/2\,\mathrm{Re}(\mathbf{E}_{\text{tot},r} \times \mathbf{H}_{\text{tot},r}^*)$ through a semi-circle in the far field, where $*$ denotes complex conjugation [25], which yields

$$P_r = \frac{1}{2n_1}\sqrt{\frac{\mu_0}{\varepsilon_0}}\int_0^{\pi}|H_{\text{tot},r}^{ff}(r,\theta)|^2 r\mathrm{d}\theta. \tag{A10}$$

The incident power is found in the same way:

$$P_i = \frac{1}{2n_1}\sqrt{\frac{\mu_0}{\varepsilon_0}}|A|^2 w_0\sqrt{\frac{\pi}{2}}. \tag{A11}$$

The total reflectance R of the structure is defined as the ratio between the total reflected power and the incident power; thus

$$R = \frac{P_r}{P_i}. \tag{A12}$$

SPPs are excited by the grooves and carry away a power P_{SPP}, which can be found using the method in Appendix B in [4]. We have checked that there is energy conservation when absorption losses in the metal are absent ($R + P_{\text{SPP}}/P_i = 1$), and when absorption is present $R + P_{\text{SPP}}/P_i < 1$.

The case of a plane wave being incident is obtained in the limit $w_0 \to \infty$. Here, we instead calculate the out-of-plane scattered power P_{OUP} as

$$P_{\text{OUP}} = \frac{1}{2n_1}\sqrt{\frac{\mu_0}{\varepsilon_0}}\int_0^{\pi}|H_{\text{scat}}^{ff}(r,\theta)|^2 r\mathrm{d}\theta. \tag{A13}$$

In the case of plane-wave incidence the calculation of the power removed from the (now very wide) reflected beam P_{EXT} due to scattering and absorption caused by the grooves is also given in Appendix B of [4]. The absorbed power is

$P_{ABS} = P_{EXT} - P_{OUP} - P_{SPP}$, and by normalizing the powers P_{SPP}, P_{OUP}, P_{ABS}, and P_{EXT} by the power per unit area of the incident plane wave we obtain the corresponding optical cross sections.

Funding. Danish Council for Independent Research (DFF) (1335-00104).

Acknowledgment. Support was received from The Danish Council for Independent Research—The FTP project: PlasTPV, Contract 1335-00104. S. I. Bozhevolnyi is acknowledged for helpful discussions.

REFERENCES

1. T. Søndergaard and S. I. Bozhevolnyi, "Optics of a single ultrasharp groove in metal," Opt. Lett. **41**, 2903–2906 (2016).
2. T. Søndergaard, S. Novikov, T. Holmgaard, R. Eriksen, J. Beermann, Z. Han, K. Pedersen, and S. Bozhevolnyi, "Plasmonic black gold by adiabatic nanofocusing and absorption of light in ultra-sharp convex grooves," Nat. Commun. **3**, 969 (2012).
3. T. Søndergaard and S. I. Bozhevolnyi, "Theoretical analysis of plasmonic black gold: periodic arrays of ultra-sharp grooves," New J. Phys. **15**, 013034 (2013).
4. A. Roberts, T. Søndergaard, M. Chirumamilla, A. Pors, J. Beermann, K. Pedersen, and S. Bozhevolnyi, "Light extinction and scattering from individual and arrayed high-aspect-ratio trenches in metals," Phys. Rev. B **93**, 075413 (2016).
5. J. L. Perchec, P. Quémerais, A. Barbara, and T. López-Ríos, "Why metallic surfaces with grooves a few nanometers deep and wide may strongly absorb visible light," Phys. Rev. Lett. **100**, 066408 (2008).
6. F. Pardo, P. Bouchon, R. Haidar, and J. Pelouard, "Light funneling mechanism explained by magnetoelectric interference," Phys. Rev. Lett. **107**, 093902 (2011).
7. E. Skovsen, T. Søndergaard, C. Lemke, T. Holmgaard, T. Leißner, R. Eriksen, J. Beermann, M. Bauer, K. Pedersen, and S. Bozhevolnyi, "Plasmonic black gold based broadband polarizers for ultra-short laser pulses," Appl. Phys. Lett. **103**, 211102 (2013).
8. J. Greffet, R. Carminati, K. Joulain, J. Mulet, S. Mainguy, and Y. Chen, "Coherent emission of light by thermal sources," Nature **416**, 61–64 (2002).
9. H. Miyazaki, K. Ikeda, T. Kasaya, K. Yamamoto, Y. Inoue, K. Fujimura, T. Kanakugi, M. Okada, K. Hatade, and S. Kitagawa, "Thermal emission of two-color polarized infrared waves from integrated plasmon cavities," Appl. Phys. Lett. **92**, 141114 (2008).
10. T. Bauer, *Thermophotovoltaics–Basic Principles and Critical Aspects of System Design* (Springer Verlag, 2011).
11. H. Sai and H. Yugami, "Thermophotovoltaic generation with selective radiators based on tungsten surface gratings," Appl. Phys. Lett. **85**, 3399–3401 (2004).
12. C. Argyropoulos, K. Le, N. Mattiucci, G. Daguanno, and A. Alù, "Broadband absorbers and selective emitters based on plasmonic Brewster metasurfaces," Phys. Rev. B **87**, 205112 (2013).
13. M. Bora, E. Behymer, D. Dehlinger, J. Britten, C. Larson, A. Chang, K. Munechika, H. Nguyen, and T. Bond, "Plasmonic black metals in resonant nanocavities," Appl. Phys. Lett. **102**, 251105 (2013).
14. Z. Li, W. Wang, D. Rosenmann, D. Czaplewski, X. Yang, and J. Gao, "All-metal structural color printing based on aluminum plasmonic metasurfaces," Opt. Express **24**, 20472–20480 (2016).
15. P. Johnson and R. Christy, "Optical constants of the noble metals," Phys. Rev. B **6**, 4370–4379 (1972).
16. R. Barrett, M. Berry, T. Chan, J. Demmel, J. Donato, J. Dongarra, V. Eijkhout, R. Pozo, C. Romine, and H. Vorst, *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods*, 2nd ed. (SIAM, 1994).
17. Mathworks, gmres, https://se.mathworks.com/help/matlab/ref/gmres.html.
18. F. López-Tejeira, F. J. García-Vidal, and L. Martín-Moreno, "Scattering of surface plasmons by one-dimensional periodic nanoindented surfaces," Phys. Rev. B **72**, 161405(R) (2005).
19. G. Brucoli and L. Martín-Moreno, "Effect of depth on surface plasmon scattering by subwavelength surface defects," Phys. Rev. B **83**, 075433 (2011).
20. A. Y. Nikitin, F. López-Tejeira, and L. Martín-Moreno, "Scattering of surface plasmon polaritons by one-dimensional inhomogeneities," Phys. Rev. B **75**, 035129 (2007).
21. C. Ropers, C. C. Neacsu, T. Elsaesser, M. Albrecht, M. B. Raschke, and C. Lienau, "Grating-coupling of surface plasmons onto metallic tips: a nanoconfined light source," Nano Lett. **7**, 2784–2788 (2007).
22. I. P. Radko, S. I. Bozhevolnyi, G. Brucoli, L. Martin-Moreno, F. Garcia-Vidal, and A. Boltasseva, "Efficiency of local surface plasmon polariton excitation on ridges," Phys. Rev. B **78**, 115115 (2008).
23. P. Lalanne, J. Hugonin, H. Liu, and B. Wang, "A microscopic view of the electromagnetic properties of sub-λ metallic surfaces," Surf. Sci. Rep. **64**, 453–469 (2009).
24. B. Wang and P. Lalanne, "Surface plasmon polaritons locally excited on the ridges of metallic gratings," J. Opt. Soc. Am. A **27**, 1432–1441 (2010).
25. L. Novotny and B. Hecht, *Principles of Nano-Optics*, 2nd ed. (Cambridge, 2012).
26. H. Liu and P. Lalanne, "Microscopic theory of the extraordinary optical transmission," Nature **452**, 728–731 (2008).
27. T. Søndergaard, V. Siahpoush, and J. Jung, "Coupling light into and out from the surface plasmon polaritons of a nanometer-thin metal film with a metal nanostrip," Phys. Rev. B **86**, 085455 (2012).

Paper II

Optics of multiple grooves in metal: transition from high scattering to strong absorption

Enok J. H. Skjølstrup, Thomas Søndergaard, Kjeld Pedersen, and Thomas G. Pedersen

The paper has been published in
Journal of Nanophotonics **11**(4), 046023 (2017).

The layout has been revised.

Optics of multiple grooves in metal: transition from high scattering to strong absorption

Enok J. H. Skjølstrup,* Thomas Søndergaard, Kjeld Pedersen, and Thomas G. Pedersen

Aalborg University, Department of Materials and Production, Aalborg East, Denmark

Abstract. This paper theoretically studies how the optics of multiple grooves in a metal change as the number of grooves gradually increased from a single groove to infinitely many arranged in a periodic array. In the case of a single groove, the out-of-plane scattering (OUP) cross section at resonance can significantly exceed the groove width. On the other hand, a periodic array of identical grooves behaves radically different and is a near-perfect absorber at the same wavelength. When illuminating multiple grooves with a plane wave, the OUP cross section is found to scale roughly linearly with the number of grooves and is comparable with the physical array width even for widths of many wavelengths. The normalized OUP cross section per groove even exceeds that of a single groove, which is explained as a consequence of surface plasmon polaritons generated at one groove being scattered out of the plane by other grooves. In the case of illuminating instead with a Gaussian beam and observing the limit as the incident beam narrows and is confined within the multiple-groove array, it is found that the total reflectance becomes very low and that there is practically no OUP. The well-known result for periodic arrays is thus recovered. All calculations were carried out using Green's function surface integral equation methods taking advantage of the periodic nature of the structures. Both rectangular and tapered grooves are considered. © *2017 Society of Photo-Optical Instrumentation Engineers (SPIE)* [DOI: 10.1117/1.JNP.11.046023]

Keywords: surface plasmons; diffraction and gratings; scattering theory; metal optics.

Paper 17136P received Sep. 6, 2017; accepted for publication Nov. 30, 2017; published online Dec. 28, 2017.

1 Introduction

Optics of grooves in metal has attracted attention due to their interesting scattering and absorption (ABS) properties. The optical cross sections of a single subwavelength ultrasharp or tapered groove in metal have been theoretically studied in detail in Ref. 1 establishing several fundamental results describing how the cross sections depend on the groove dimension. Here, it was found that the out-of-plane scattering (OUP) cross section can exceed the physical width of the groove in a broad wavelength interval. A periodic array of ultrasharp grooves is on the other hand found to give rise to broadband ABS, thus turning a shiny, highly reflecting surface into a black surface.[2,3] Thus, a single groove and a periodic array of grooves behave quite different, and in Ref. 1, it was suggested that the very low reflectance of the periodic array was due to mutual destructive interference between the scattered fields from the individual grooves. This hypothesis was recently tested in Ref. 4, where the optics of multiple ultrasharp grooves in metal was studied as the transition from one to infinitely many grooves. Surprisingly, it was found that the hypothesis was not correct when the incident field is a plane wave. The OUP cross section was found to scale approximately linear with the number of grooves and to be ∼1.5 times larger than the physical width of the grooves even for widths of many wavelengths. Instead, it was found that when illuminating 20 grooves with a Gaussian beam entirely focused within the grooves, the reflectance is the same as for a periodic array illuminated by a plane wave.

*Address all correspondence to: Enok J. H. Skjølstrup, E-mail: ejs@mp.aau.dk

In this paper, we explore the same transition as in Ref. 4 but for rectangular and tapered (not ultrasharp) grooves and show that the same principles apply for these types of grooves when they are combined in an array of multiple grooves. For these types of grooves, the cross sections are found to be significantly large only for a narrow band of wavelengths,[5–8] and thus, the optical cross-section spectra are quite different from previous work.

An application for arrays of tapered grooves in metal is to use them in constructing broadband omnidirectional absorbers and angularly selective emitters.[9] As only p-polarized light will be efficiently absorbed in the grooves while s-polarized light will be almost perfectly reflected, an array of ultrasharp grooves can be applied as polarizers for ultrashort laser pulses.[10] As the rectangular and tapered grooves only absorb light for wavelengths close to the resonance, an application for those grooves is in selective thermal emitters,[11,12] which can be advantageous in thermophotovoltaics.[13,14]

The structure of interest in this paper is shown in Fig. 1(c), where the grooves can be either rectangular [Fig. 1(a)] or tapered [Fig. 1(b)]. The rectangular grooves have a depth of 500 nm and a width of 50 nm, whereas the tapered grooves have a depth of 350 nm, a top width of 100 nm, and a bottom width of 60 nm, where all the corners are rounded by a circle with a radius of 4 nm as in Ref. 5, implying that the top width of the rectangular grooves is 58 nm. The incident field in Fig. 1(c) is a normal Gaussian beam with beam waist radius w_0 centered at $x = x_0$ in the middle of the array of N identical grooves. There is a distance d between the grooves, and the total length of the groove array is denoted L. The beam waist radius is related to the array length by $w_0 = \gamma L/2$, where γ is a ratio parameter determining the width of the Gaussian beam. The schematic in Fig. 1(c) corresponds to $\gamma = 1$. The incident light can be either reflected, scattered upward, absorbed in the metal, or scattered into surface plasmon polaritons (SPPs), which are electromagnetic waves bounded to and propagating along the metal surface. As in Ref. 4 the magnetic field only has a z-component [$\mathbf{H}(r) = \hat{z}H(r) = \hat{z}H(x, y)$], and the structure is considered invariant in the z-direction, which implies that 2-D-calculations are performed. Gold is applied as the metal, and the dielectric constant of gold is from Ref. 15. The calculations are performed using the Green's function surface integral equation method (GFSIEM) as presented in Appendix B in Ref. 5. See Ref. 4 for a further description of how the matrix equation is constructed and solved using the iterative method GMRES.[16,17]

The paper is organized in the following way: Sec. 2 contains the case with a plane wave ($w_0 = \infty$) as the incident field, and here, extinction (EXT), scattering, and ABS cross sections are calculated for a structure of varying N, where the grooves are rectangular with the aforementioned dimensions. A Gaussian beam is used as the incident field in Sec. 3, where the grooves are still rectangular. Here, the beam waist radius is varied, and angular reflection spectra and total out-of-plane reflected power are calculated. In Sec. 4, the incident field is a plasmon and the reflectance, transmittance, and OUP is studied as a function of the number of grooves. In Sec. 5 the grooves are tapered, and here, both optical cross sections and reflectance as a function of wavelength are presented.

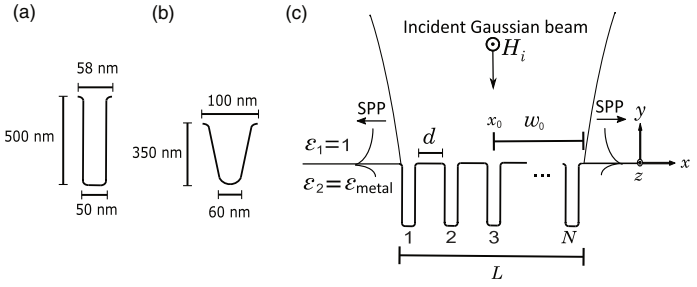

Fig. 1 (a) A rectangular groove, (b) a tapered groove, (c) schematic of N identical rectangular grooves in metal separated by the distance d. The incident field is a normal Gaussian beam with beam waist radius w_0 centered in the middle of the groove array ($x = x_0$).

2 Plane Wave as Incident Field

In this section, the incident field is a plane wave, which means that the beam waist radius w_0 in Fig. 1(c) tends to infinity. When light is incident on the multiple grooves, scattering occurs and some light is coupled into SPPs propagating along the metal surface away from the grooves, some light is scattered out of the plane, and some light is absorbed in the metal. EXT refers to the amount of power removed from the reflected beam due to scattering and ABS. The corresponding EXT, OUP, and SPP cross sections are obtained by normalizing the respective powers by the power per unit area of the incident light. The ABS cross section is given by the EXT cross-section minus the OUP and SPP cross sections. See Appendix B in Ref. 5 for a description of how the cross sections are calculated using the GFSIEM.

The rectangular shape of the grooves is found to give rise to a narrowband resonant behavior in all the cross sections.[5] For the particular dimensions of the grooves considered here, the cross sections of a single groove are found to be resonant at a wavelength of 660 nm as will later be shown in Fig. 5(a). Before multiple grooves are considered, it is chosen to study cross sections for a structure of only two grooves depending on the distance between them. It is chosen to fix the wavelength at $\lambda_0 = 660$ nm, and the EXT, OUP, and SPP cross sections are seen in Fig. 2 as a function of distance between the two grooves. All the cross sections are oscillating with certain extreme values to be explained. For the SPP cross section, the first minimum is found at $d = 260$ nm. With a top width of a single groove at 58 nm [see Fig. 1(a)], the period of the structure for this d is 318 nm, which equals a half plasmon wavelength, where $\lambda_{\mathrm{SPP}} = \sqrt{(\varepsilon_1 + \varepsilon_2')/(\varepsilon_1 \varepsilon_2')}\lambda_0 = 635$ nm, and ε_2' is the real part of the metal dielectric constant ε_2.[18] Hence at $d = 260$ nm, the SPPs generated at the different grooves interfere destructively implying that the SPP cross section is practically 0. When this happens almost all of the scattered light is coupled out of the plane, as seen by the fact that the OUP cross section is a maximum at almost the same d for which the SPP cross section is minimized. When d increases toward a plasmon wavelength, plasmons generated at the different grooves interfere constructively implying that scattering into SPPs has a maximum, and at approximately the same d, both the OUP and EXT cross sections are minimized. At $d = 300$ nm, the EXT cross section has a maximum; at $d = 600$ nm, it has a minimum; therefore, it is chosen to consider these distances in the following. The inset in Fig. 2 shows the differential OUP cross section for $d = 600$ nm. Here, interference similar to a double slit predicts that destructive interference occurs at 60 deg and 120 deg, which is verified in the inset.[19]

For multiple grooves, the EXT, OUP, and ABS cross sections are seen at a wavelength of 660 nm in Fig. 3 when the distance d is 300 nm in (a) and 600 nm in (b). Especially in Fig. 3(a), the cross sections are almost linear functions of the number of grooves, while in Fig. 3(b) the linear behavior first begins after ~10 grooves. It is clearly seen that the cross sections in Fig. 3(a) are much larger than those in Fig. 3(b), and this large difference is not entirely caused by the fact that the EXT and OUP cross sections are smaller for $d = 600$ nm than for $d = 300$ nm according to Fig. 2. As will later be shown in Fig. 5(b) for $d = 600$ nm, the resonance wavelength is

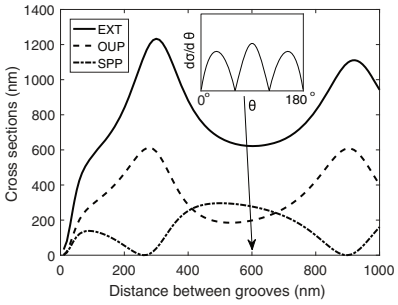

Fig. 2 EXT, OUP, and SPP cross sections for a structure of two grooves with varying distance d between the grooves. The inset shows the differential OUP cross section at $d = 600$ nm. The wavelength is 660 nm.

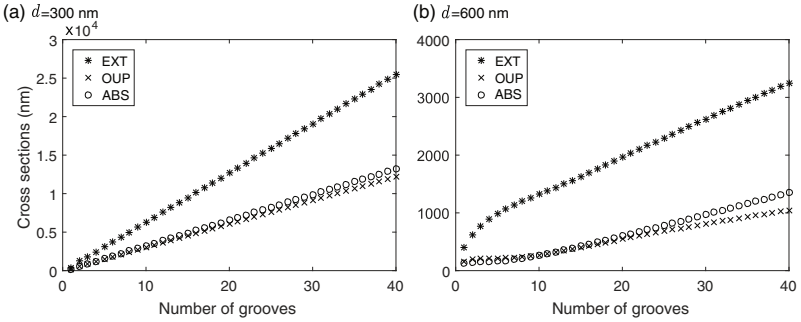

Fig. 3 EXT, OUP, and ABS cross sections at $\lambda = 660$ nm, where (a) $d = 300$ nm and (b) $d = 600$ nm.

blue-shifted from the 660 nm being the resonance wavelength of a single groove and furthermore a smaller peak occurs around 700 nm. Hence, when the wavelength is 660 nm, as is the case in Fig. 3, there is no resonance for $d = 600$ nm, which implies that the cross sections are much smaller compared with Fig. 3(a). For $N = 40$ grooves the OUP cross section is ~ 12 μm for $d = 300$ nm, and here, the total length of the groove array is $L \approx 14$ μm. However, the grooves themselves only occupy approximately one sixth of the length, as the distance d between the grooves is much larger than the groove width. The OUP cross section per groove is seen in Fig. 4, where again $d = 300$ nm in (a) and $d = 600$ nm in (b). In Fig. 4(a), the OUP cross section per groove converges to $\sim$300 nm, which is $\sim$0.85 times the groove period, but more than five times larger than a single groove width. For $d = 600$ nm, the OUP cross section for 40 grooves is ~ 1 μm as seen in Fig. 3(b), and the OUP per groove thus converges to $\sim$25 nm as seen in Fig. 4(b). This much smaller cross section per groove is again due to the fact that for $d = 600$ nm the wavelength at 660 nm is not resonant.

The linear behavior of the EXT, OUP, and ABS cross section as a function of number of grooves was recently found for ultrasharp grooves in Ref. 4, where the OUP per groove was found to be $\sim$1.5 times the groove period. In this study, the grooves had a wide opening at 240 nm in the top and a bottom width of only 0.3 nm with only $d = 10$ nm between the grooves. Furthermore, the reflectance of an infinite array of the same grooves illuminated by a plane wave was found in Ref. 2 to be 16% for this particular wavelength at 770 nm, and the extraordinary large OUP cross section for a structure consisting of 40 grooves was, therefore, a surprising and remarkable result. For the rectangular grooves considered in this paper, the OUP cross section per groove is thus comparable with Ref. 4, but with the difference that it is smaller than

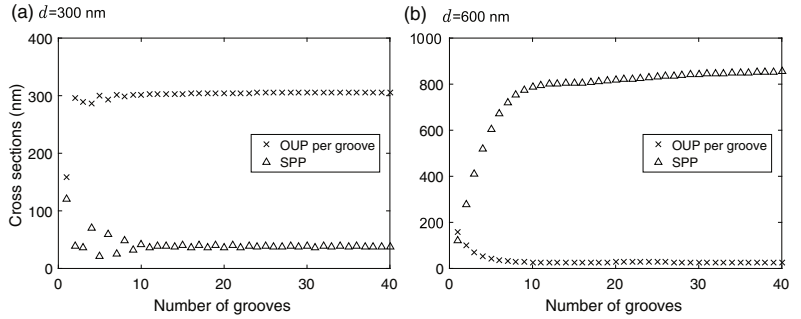

Fig. 4 SPP and OUP cross sections per groove at $\lambda = 660$ nm, where (a) $d = 300$ nm and (b) $d = 600$ nm.

the groove period (by the factor ≈0.85) but much larger than a single groove width (by the factor > 5).

The SPP cross section is shown by the triangles in Fig. 4 being a damped oscillatory function in Fig. 4(a) and an increasing function in Fig. 4(b). According to Fig. 2, the distance $d = 300$ nm is close to the minimum where the plasmons generated at different grooves interfere destructively resulting in a small SPP cross section. Another phenomenon that implies a small SPP cross section is the fact that a plasmon generated at one groove can be coupled out of the plane by another groove as examined in Refs. 20–22. This effect will be studied in detail in Sec. 4. In Fig. 4(b), the distance d is close to the maximum of SPP cross section according to Fig. 2. Here, plasmons generated at different grooves interfere constructively resulting in an overall increase in the total SPP cross section. Hence, if the purpose of the structure is to efficiently excite plasmons the groove period should be close to the plasmon wavelength as studied in Refs. 23 and 24. In addition, it was found in Refs. 25 and 26 that the dimensions of the individual grooves can be optimized in such a way that most of the incident light is excited into SPPs. Furthermore, the efficiency of SPP excitation from a single and multiple (but finite) rectangular grooves is studied in Ref. 27, showing that the SPPs can be excited with a significantly higher efficiency when more grooves are present.

While only a wavelength of 660 nm has been considered so far, the EXT cross section is seen as a function of wavelength for a structure consisting of 1, 5, 10, and 20 grooves in Fig. 5 for $d = 300$ nm in (a) and $d = 600$ nm in (b). Especially in Fig. 5(a), the EXT cross-section scales approximately linear with N for all considered wavelengths, and the same is found for the OUP and ABS cross sections (not shown). Here, the resonance wavelength has slightly changed from 660 nm for one groove into 657 nm for 20 grooves. This change in resonance wavelength is more pronounced in Fig. 5(b) where it is blue-shifted by ∼30 nm compared with the spectrum for a single groove, but as N increases, the resonance wavelength slightly red-shifts again. As in Fig. 5(a) the EXT cross-section scales roughly linear with N at most wavelengths. However, in Fig. 5(b) the groove period is comparable with the wavelength, which implies that Rayleigh–Wood anomalies split the resonance wavelength into two peaks instead of one.[28–30] This splitting is more pronounced for $N = 20$ implying that the linear scaling of EXT with N fails for wavelengths around 700 nm for $N < 20$.

The same linear scaling of EXT cross section for many wavelengths was recently found in Ref. 4 for ultrasharp grooves, where the cross sections were large in a much broader wavelength interval. Based on this linear scaling for a specific groove dimension, and the study of the cross sections of a single groove for many different groove dimensions in Ref. 1, it was postulated in Ref. 4 that the linear scaling will also be valid for other groove dimensions, and Fig. 5(a) confirms this for a rectangular groove being a narrowband resonator.

As seen in Fig. 3, the OUP cross-section scales approximately linear with the number of grooves. To further study the out-of-plane scattering from multiple grooves, the differential OUP cross section is shown in Fig. 6 for a structure consisting of 1, 5, 10, and 20 grooves

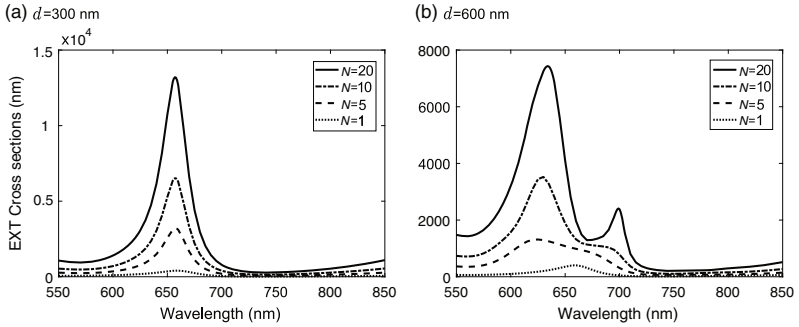

Fig. 5 EXT cross sections for 1, 5, 10, and 20 grooves as a function of wavelength, where (a) $d = 300$ nm and (b) $d = 600$ nm.

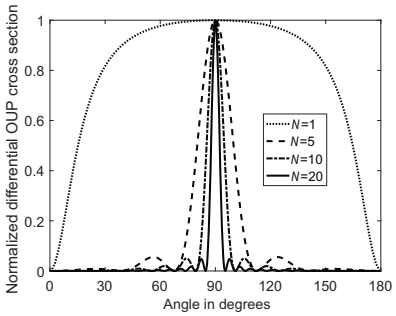

Fig. 6 Normalized differential OUP cross sections for 1, 5, 10, and 20 grooves for $\lambda = 660$ nm and $d = 300$ nm.

with $d = 300$ nm and $\lambda = 660$ nm. The differential cross sections have all been normalized such that they have a maximum of 1. When many grooves are present the angular distribution clearly becomes much more narrow. Notice that the OUP cross section is found by integrating the differential OUP cross section from 0 deg to 180 deg.

3 Gaussian Beam as Incident Field

In this section, the incident field is a Gaussian beam with beam waist radius $w_0 = \gamma L/2$ as shown in Fig. 1(c) for $\gamma = 1$. The same type of calculations as in Ref. 4 is performed, see for example, the appendix in Ref. 4 for a further explanation of the calculation of relevant terms.

3.1 *Reflectance as a Function of Beam Waist Radius*

First, it is studied how the incident field and reflected field depend on the ratio parameter γ, which determines the beam waist radius of the Gaussian beam. A structure consisting of 20 grooves with $d = 300$ nm between the grooves is considered, and the wavelength is chosen to be 660 nm. The incident power per angle is seen by the bright lines (blue online) in Fig. 7(a) for $\gamma = 0.5$ and 2, and even for such small γ the beam is still paraxial (meaning that $2\pi w_0/\lambda \gg 1$) as the structure has been chosen to be sufficiently wide such that this is achieved for $\gamma \geq 0.5$. It is clearly seen that a smaller γ implies that the angular distribution

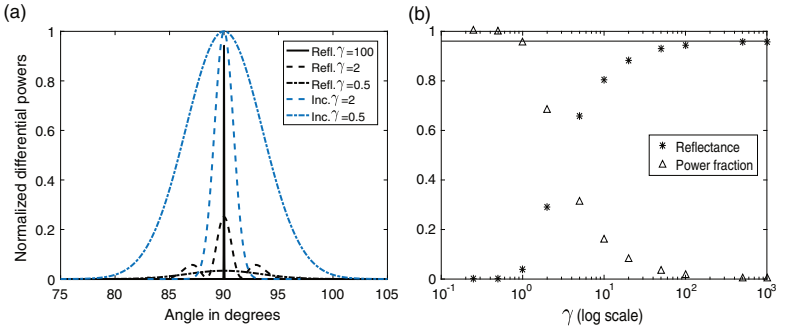

Fig. 7 (a) Normalized differential incident power for varying γ is seen by the bright lines (blue online) and the normalized differential reflected power in black at a wavelength of $\lambda = 660$ nm. (b) Reflectance and power fraction as a function of γ for 20 grooves at a wavelength of $\lambda = 657$ nm. The distance $d = 300$ nm in both (a) and (b).

is broader. The reflected power per angle is seen in black in the same figure when $\gamma = 0.5$, 2, and 100. For $\gamma = 100$, the incident field is so wide that it mostly hits the planar surface surrounding the grooves and the grooves themselves only marginally contribute. Therefore, the incident and reflected beams are almost the same, why only the reflected field is shown in Fig. 7(a), and it is very narrow in angular distribution as it behaves almost, such as a plane wave. The reflectance is found as the area under the black curve divided by the area under the corresponding blue curve and is found to be 0.03, 0.31, and 0.95 when $\gamma = 0.5$, 2, and 100, respectively. Hence, the reflectance strongly depends on γ since a high γ implies that the Gaussian beam also hits the planar surface and not entirely the grooves. This is found not to be the case for $\gamma \leq 0.5$, why in this case the calculated reflectance is entirely due to the grooves and not the surrounding planar surface.

The reflectance is thus calculated as a function of γ and is shown by the asterisks as in Fig. 7(b) for the same structure consisting of 20 grooves with $d = 300$ nm. Notice that the γ-axis in the figure is on a log scale.

As mentioned in Sec. 2, the resonance wavelength for the structure consisting of 20 grooves has slightly changed into 657 nm, why this wavelength has been used in Fig. 7(b) while 660 nm was the wavelength in Fig. 7(a). The reflectance is seen to converge for both small and large γ, where it converges to the proper reflectance, which is entirely due to the grooves when $\gamma \leq 0.5$ and converges to that of a flat gold surface when γ is large, as illustrated by the black horizontal line in the figure. The power fraction illustrated by the triangles in the same figure shows the geometric fraction of the incident power that actually hits the grooves. This fraction is practically 1 for $\gamma \leq 0.5$ and is practically 0 for $\gamma \geq 500$. Importantly, notice that the reflectance and power fraction converge for the same γ.

3.2 *Reflectance as a Function of Number of Grooves*

In this subsection, it is studied how the reflectance depends on the number of grooves for a fixed beam waist radius w_0, which is set to 1715 nm corresponding to $\gamma = 0.5$ for 20 grooves. In Fig. 8, the reflectance as a function of number of grooves is seen for $\lambda = 657$ nm and converges to practically 0 when 20 grooves are present. Thus, it is found that 20 grooves are sufficient to obtain the same reflectance as a structure consisting of infinitely many grooves. As in Fig. 7(b), the triangles show the power fraction and it converges to 1 when 20 grooves are present as was also observed in Fig. 7(b). Hence, Fig. 8 shows that the reflectance converges to the proper reflectance of the grooves when all the incident light hits the groove array. The same result was obtained in Ref. 4 for ultrasharp grooves.

Whereas Fig. 8 only considered $\lambda = 657$ nm the reflectance as a function of wavelength is seen in Fig. 9 for 5 to 20 grooves, where $d = 300$ nm and $w_0 = 1715$ nm. The reflectance of the structure consisting of 20 grooves is found to be the same as for a periodic array of the same grooves illuminated by a plane wave as studied in Ref. 5, and the same result was found in Ref. 4 for ultrasharp grooves. The reflectance is shown for 5, 7, 10, and 20 grooves and is low, close to the resonance wavelength and comparable with that of pure gold for longer wavelengths. Here, more grooves imply that the reflectance at resonance becomes lower as was also shown in Fig. 8,

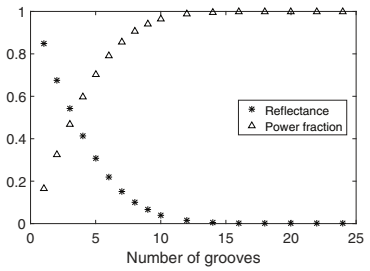

Fig. 8 Reflectance and power fraction as a function of number of grooves for a fixed beam waist $w_0 = 1715$ nm. The wavelength is 657 nm and $d = 300$ nm.

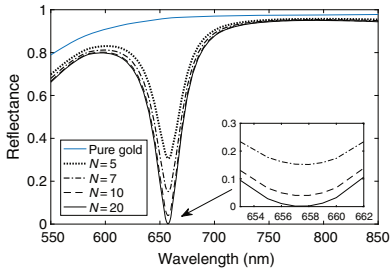

Fig. 9 Reflectance as a function of wavelength for 5 to 20 grooves, with $d = 300$ nm and $w_0 = 1715$. Inset shows the reflectance in a smaller wavelength interval.

but off resonance the reflectance is almost independent of the number of grooves. The inset in Fig. 9 shows the reflectance for a smaller wavelength interval close to the resonance wavelength, and here it is possible to see the difference in reflectance between 10 and 20 grooves. It is remarkable that the reflectance of 20 grooves when illuminating with a Gaussian beam becomes practically 0, when the same groove structure illuminated with a plane wave has a very large OUP cross section as observed in Sec. 2. The same kind of result was found in Ref. 4 for ultra-sharp grooves, and there, it was postulated to be valid for other groove dimensions as well, and Fig. 9 confirms this for rectangular grooves.

3.3 *Energy Transportation*

As in Ref. 4, the energy transportation is investigated based on the time-averaged Poynting vector $\langle \mathbf{S} \rangle = 1/2\,\mathrm{Re}(\mathbf{E} \times \mathbf{H}^*)$, where $\mathbf{E}$ and $\mathbf{H}$ are the complex electric and magnetic field, respectively, and where * denotes complex conjugation. The interpretation of the Poynting vector is that it points in the direction of the power flow with a magnitude describing the power flow per unit area.[18] Figure 10 shows the magnitude of the time-averaged Poynting vector when a Gaussian beam with ratios of 0.5, 1, 2, and ∞ (plane wave) is incident on a structure consisting of 20 grooves with $d = 300$ nm between the grooves and with a wavelength of 660 nm. The black horizontal lines in the figure illustrate the length of the groove array. When $\gamma = 0.5$ the

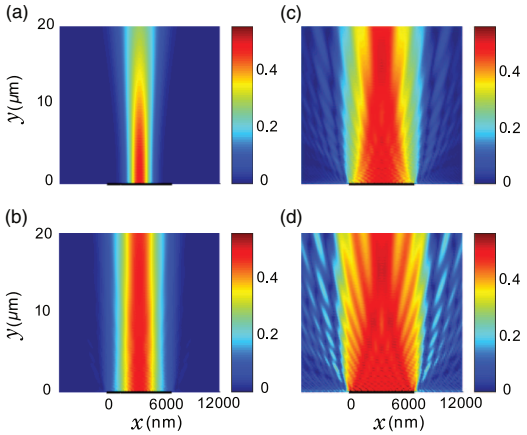

Fig. 10 (Color online) Magnitude of time-averaged Poynting vector when a Gaussian beam is incident on a structure consisting of 20 grooves at a wavelength of 660 nm with $d = 300$ nm. The horizontal black lines denote the groove array and the ratios are (a) $\gamma = 0.5$, (b) $\gamma = 1$, (c) $\gamma = 2$, and (d) $\gamma = \infty$ (plane wave).

incident field is entirely focused within the grooves [see Fig. 7(b)], which implies that there is almost no light scattered out of the plane. This is clearly observed in Fig. 10(a), where the magnitude of the Poynting vector is zero for all other positions than right above the groove array. When γ increases, more light is scattered out of the plane as shown in Figs. 10(b)–10(d). When the incident field is a plane wave, as shown in Fig. 10(d), most of the incident light hits the planar surface surrounding the grooves where they experience almost total reflectance as pure gold is almost a perfect mirror at this wavelength. Hence, far outside the grooves, the net power flow is very low. The angular distribution of the scattered field is broader for $\gamma = \infty$ than for $\gamma = 2$ [Fig. 10(c)], which was also partly shown in Fig. 7(a) as the oscillations in differential powers for angles deviating from 90 deg.

4 Plasmon as Incident Field

In this section, we consider an SPP being incident on multiple rectangular grooves with $d = 300$ nm between the grooves. A schematic of this situation is shown in Fig. 11(a), where the SPP is incident from left, thus it can be reflected, transmitted, scattered out of the plane, or absorbed. The reflectance, transmittance, and OUP of an SPP incident on multiple rectangular grooves have previously been considered in Refs. 20 and 21 whereas the case of a single groove or ridge of different shapes was considered in Ref. 22 and the case of a rectangular hole in Ref. 31. The case of short- and long-range SPPs of thin-metal films being incident on a rectangular metal nanostrip forming a gap-plasmon resonator together with the metal film has also been considered.[32]

Figure 11(b) shows the reflectance, transmittance, and OUP for a plasmon being incident on a single groove as a function of wavelength. Like in Figs. 5(a) and 9(a), clear resonant behavior is observed with a dip in transmittance relatively close to the resonant wavelength at 657 nm. The fact that a part of the incident plasmon is scattered out of the plane at resonance serves as a qualitative explanation of why the OUP per groove is higher for multiple grooves compared with the single groove case as was observed in Fig. 4(a). Figure 11(c) shows the reflectance, transmittance, and OUP, as a function of number of grooves at $\lambda = 657$ nm. The plasmon transmittance through one groove is almost 50% but decreases as more grooves are present while the scattering and ABS level increase. When 10 grooves are present less than 1% of the incident light is transmitted, and therefore, all three curves remain constant hereafter. For ultrasharp grooves, it was found in Ref. 4 that still 5% of the incident light was transmitted through a structure consisting of 20 grooves (at $\lambda = 770$ nm). Hence, the array of rectangular grooves blocks the light of an incident plasmon better compared with an array of ultrasharp grooves. Notice that the plasmon reflectance is highest when two grooves are present, which is due to the fact that the distance between the grooves is roughly a half wavelength, thus the reflected light from the two

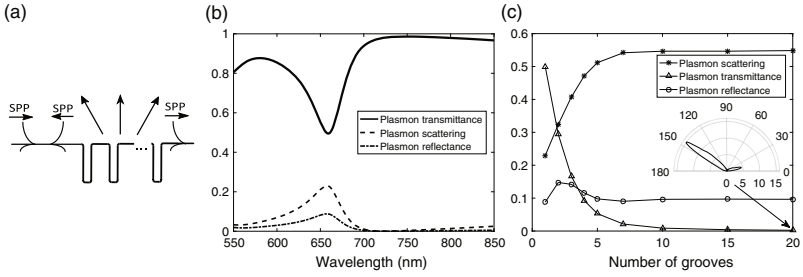

Fig. 11 (a) Schematic of an SPP wave incident from left on multiple rectangular grooves. The incident plasmon can be reflected, transmitted, or scattered out of the plane. (b) Plasmon reflectance, transmittance, and scattering in the case of one groove as a function of wavelength. (c) Plasmon reflectance, transmittance, and scattering as a function of number of grooves for $\lambda = 657$ nm and $d = 300$ nm. Inset shows normalized differential power for a structure consisting of 20 grooves.

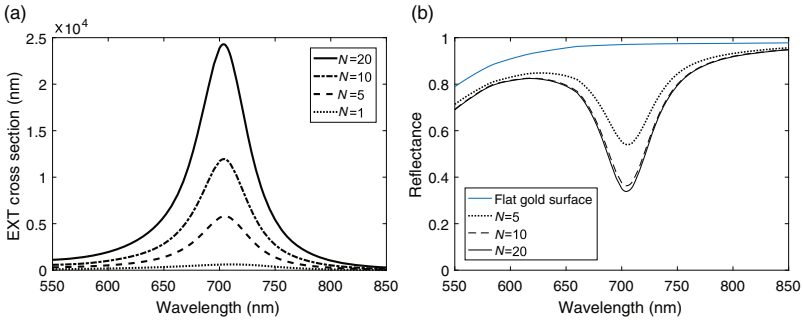

Fig. 12 (a) EXT cross section as a function of wavelength for 1, 5, 10, and 20 tapered grooves. (b) Reflectance as a function of wavelength for 5, 10, and 20 tapered grooves with $w_0 = 1925$ nm. The distance d between the grooves is 300 nm in both (a) and (b).

grooves interferes constructively resulting in a higher plasmon reflectance. When more grooves are present, multiple reflections within the groove array imply that the plasmon reflectance decreases. The inset in Fig. 11(c) shows a polarplot of the normalized differential power when 20 grooves are present showing that most of the scattered light is actually scattered obliquely backward at an angle of approximately 150 deg. The plot has been normalized such that integrating the differential power from 0 deg to 180 deg gives 0.55, which is the plasmon scattering for 20 grooves.

5 Tapered Grooves

Until now rectangular grooves have been examined as shown in Fig. 1(a). In this section, tapered grooves are considered with a depth of 350 nm, a top width of 100 nm, and a bottom width of 60 nm as seen in Fig. 1(b). In Ref. 5, tapered grooves are found to give rise to broader resonances compared with rectangular grooves, which can be understood in terms of the resonator formalism presented therein. The tapered grooves are studied here following the same procedure as for the rectangular grooves studied in Secs. 2 and 3. Thus first two grooves are studied, and the cross sections found as a function of distance between the grooves. The result is found to be very similar to Fig. 2 and with approximately the same d giving rise to the extrema. The EXT cross section as a function of wavelength is seen in Fig. 12(a) for a structure consisting of 1, 5, 10, and 20 grooves with $d = 300$ nm between the grooves, where the resonance wavelength is 704 nm. The spectra are quite similar to those of the rectangular grooves in Fig. 5(a) but with a broader resonance. Again the EXT cross section scales almost linear with the number of grooves for the considered wavelengths. The reflectance of the same structure when illuminating with a Gaussian beam is seen in Fig. 12(b) and follows the same principles as the rectangular grooves in Fig. 9. Here, the beam waist radius w_0 has been fixed at 1925 nm, which here corresponds to $\gamma = 0.5$ for a structure consisting of 20 grooves. Again the reflectance of the structure consisting of 20 grooves is the same as for a periodic array of the same grooves illuminated by a plane wave. Here, the reflectance at resonance is 0.34 even when 20 grooves are present, illustrating that the grooves have to be sufficiently narrow in order for the reflectance at resonance to be very low. Hence for the tapered grooves considered here, the resonance is broader, and it is not possible to achieve perfect ABS as for rectangular grooves.

6 Conclusion

The optics of multiple rectangular and tapered grooves in metal has been studied theoretically to examine the transition from a single groove to infinitely many grooves arranged in a periodic array. When the incident field is a plane wave the OUP cross section depends approximately

linearly on the number of grooves, a result that was also recently found for multiple ultrasharp grooves. The OUP cross section per groove is comparable with the groove period even though the scattering structure has a width of many wavelengths. Furthermore, the OUP cross section per groove is higher than for a single groove, which is explained by the fact that an SPP generated in one groove can be scattered out of the plane by other grooves, and this is supported by studying the reflectance, transmittance, and OUP when an SPP is incident on multiple grooves. A structure consisting of infinitely many grooves in a periodic array illuminated by a plane wave has a very low reflectance at resonance, but this is not due to destructive interference occurring between the scattered fields of different grooves. Instead, a narrow Gaussian beam focused entirely within the grooves has to be used as the incident field in order for the reflectance of multiple grooves to be the same as for an infinite array of grooves illuminated by a plane wave.

When the distance between the rectangular grooves is 300 nm a structure consisting of 20 grooves is found to be a near-perfect absorber for wavelengths close to 657 nm while it, for longer wavelengths, is a nearly perfect mirror. When the distance between the grooves increases to 600 nm the reflectance is found to be close to that of pure gold independent on the number of grooves in the structure. For tapered grooves, the resonance is broader, but the minimal reflectance is found to be 0.34 even when 20 grooves are present, and the tapered grooves considered here can, therefore, not be used as a perfect absorber.

Acknowledgments

This work was supported by Villum Kann Rasmussen (VKR) center of excellence QUSCOPE. This is an extended and revised paper based on publication in the SPIE Proceeding Optics + Photonics 2017, San Diego, Vol. 10346.

References

1. T. Søndergaard and S. I. Bozhevolnyi, "Optics of a single ultrasharp groove in metal," *Opt. Lett.* **41**, 2903–2906 (2016).
2. T. Søndergaard et al., "Plasmonic black gold by adiabatic nanofocusing and absorption of light in ultra-sharp convex grooves," *Nat. Commun.* **3**, 969 (2012).
3. T. Søndergaard and S. I. Bozhevolnyi, "Theoretical analysis of plasmonic black gold: periodic arrays of ultra-sharp grooves," *New J. Phys.* **15**, 013034 (2013).
4. E. J. Skjølstrup and T. Søndergaard, "Optics of multiple ultrasharp grooves in metal," *J. Opt. Soc. Am. B* **34**, 673–680 (2017).
5. A. Roberts et al., "Light extinction and scattering from individual and arrayed high-aspect-ratio trenches in metals," *Phys. Rev. B* **93**, 075413 (2016).
6. J. L. Perchec et al., "Why metallic surfaces with grooves a few nanometers deep and wide may strongly absorb visible light," *Phys. Rev. Lett.* **100**, 066408 (2008).
7. F. Pardo et al., "Light funneling mechanism explained by magnetoelectric interference," *Phys. Rev. Lett.* **107**, 093902 (2011).
8. J. Guo, Z. Li, and H. Guo, "Near perfect light trapping in a 2D gold nanotrench grating at oblique angles of incidence and its application for sensing," *Opt. Express* **24**, 17259–17271 (2016).
9. C. Argyropoulos et al., "Broadband absorbers and selective emitters based on plasmonic brewster metasurfaces," *Phys. Rev. B* **87**, 205112 (2013).
10. E. Skovsen et al., "Plasmonic black gold based broadband polarizers for ultra-short laser pulses," *Appl. Phys. Lett.* **103**, 211102 (2013).
11. J. Greffet et al., "Coherent emission of light by thermal sources," *Nature* **416**, 61–64 (2002).
12. H. Miyazaki et al., "Thermal emission of two-color polarized infrared waves from integrated plasmon cavities," *Appl. Phys. Lett.* **92**, 141114 (2008).
13. T. Bauer, *Thermophotovoltaics—Basic Principles and Critical Aspects of System Design*, 1st ed., Springer Verlag, Berlin (2011).
14. H. Sai and H. Yugami, "Thermophotovoltaic generation with selective radiators based on tungsten surface gratings," *Appl. Phys. Lett.* **85**, 3399–3401 (2004).

15. P. Johnson and R. Christy, "Optical constants of the noble metals," *Phys. Rev. B* **6**, 4370–4379 (1972).
16. R. Barrett et al., *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods*, 2nd ed., SIAM, Philadelphia (1994).
17. Mathworks, "GMRES," https://se.mathworks.com/help/matlab/ref/gmres.html (29 June 2017).
18. L. Novotny and B. Hecht, *Principles of Nano-Optics*, 2nd ed., Cambridge University Press, Cambridge (2012).
19. H. D. Young, R. A. Freedman, and A. L. Ford, *University Physics with Modern Physics*, 13th ed., Pearson, San Francisco (2012).
20. F. López-Tejeira, F. J. García-Vidal, and L. Martín-Moreno, "Scattering of surface plasmons by one-dimensional periodic nanoindented surfaces," *Phys. Rev. B* **72**, 161405 (2005).
21. G. Brucoli and L. Martín-Moreno, "Effect of depth on surface plasmon scattering by sub-wavelength surface defects," *Phys. Rev. B* **83**, 075433 (2011).
22. A. Y. Nikitin, F. López-Tejeira, and L. Martín-Moreno, "Scattering of surface plasmon polaritons by one-dimensional inhomogeneities," *Phys. Rev. B* **75**, 035129 (2007).
23. C. Ropers et al., "Grating-coupling of surface plasmons onto metallic tips: a nanoconfined light source," *Nano Lett.* **7**, 2784–2788 (2007).
24. I. P. Radko et al., "Efficiency of local surface plasmon polariton excitation on ridges," *Phys. Rev. B* **78**, 115115 (2008).
25. P. Lalanne et al., "A microscopic view of the electromagnetic properties of sub-λ metallic surfaces," *Surf. Sci. Rep.* **64**, 453–469 (2009).
26. B. Wang and P. Lalanne, "Surface plasmon polaritons locally excited on the ridges of metallic gratings," *J. Opt. Soc. Am. A* **27**, 1432–1441 (2010).
27. S. de la Cruz et al., "Compact surface structures for the efficient excitation of surface plasmon-polaritons," *Phys. Status Solidi B* **249**, 1178–1187 (2012).
28. T. Søndergaard et al., "Extraordinary optical transmission with tapered slits: effect of higher diffraction and slit resonance orders," *J. Opt. Soc. Am. B* **29**, 130–137 (2012).
29. A. A. Maradudin et al., "Rayleigh and wood anomalies in the diffraction of light from a perfectly conducting reflection grating," *J. Opt.* **18**, 024004 (2016).
30. U. Fano, "The theory of anomalous diffraction gratings," *J. Opt. Soc. Am.* **31**, 213–222 (1941).
31. H. Liu and P. Lalanne, "Microscopic theory of the extraordinary optical transmission," *Nature* **452**, 728–731 (2008).
32. T. Søndergaard, V. Siahpoush, and J. Jung, "Coupling light into and out from the surface plasmon polaritons of a nanometer-thin metal film with a metal nanostrip," *Phys. Rev. B* **86**, 085455 (2012).

Biographies for the authors are not available.

Paper III

Quantum spill-out in few-nanometer metal gaps: Effect on gap plasmons and reflectance from ultrasharp groove arrays

Enok J. H. Skjølstrup, Thomas Søndergaard, and Thomas G. Pedersen

The paper has been published in
Physical Review B **97**, 115429 (2018).

The layout has been revised.

Quantum spill-out in few-nanometer metal gaps: Effect on gap plasmons and reflectance from ultrasharp groove arrays

Enok J. H. Skjølstrup,* Thomas Søndergaard, and Thomas G. Pedersen
Department of Materials and Production, Aalborg University, Skjernvej 4A, DK-9220 Aalborg East, Denmark

(Received 15 December 2017; revised manuscript received 7 March 2018; published 19 March 2018)

Plasmons in ultranarrow metal gaps are highly sensitive to the electron density profile at the metal surfaces. Using a quantum mechanical approach and assuming local response, we study the effects of electron spill-out on gap plasmons and reflectance from ultrasharp metal grooves. We demonstrate that the mode index of ultranarrow gap plasmons converges to the bulk refractive index in the limit of vanishing gap and, thereby, rectify the unphysical divergence found in classical models. Surprisingly, spill-out also significantly increases the plasmonic absorption for few-nanometer gaps and lowers the reflectance from arrays of ultrasharp metal grooves. These findings are explained in terms of enhanced gap plasmon absorption taking place inside the gap 1–2 Å from the walls and delocalization near the groove bottom. Reflectance calculations taking spill-out into account are shown to be in much better agreement with measurements compared with classical models.

DOI: 10.1103/PhysRevB.97.115429

I. INTRODUCTION

In the past decade, plasmonic structures have attracted attention, in part due to their efficiency in absorbing incident light [1,2] and their ability to squeeze light below the diffraction limit [3–6]. Furthermore, plasmonic structures can be applied in, e.g., solar cells, lasers, and biosensors [7–9]. Metal surfaces support surface plasmon polaritons (SPPs) that are electromagnetic waves bound to and propagating along the surface, while deep-subwavelength gaps between metal surfaces may support gap plasmons, i.e., waves confined to and propagating along the gap. Such gap plasmons localized in gaps of nanometer size between spherical and triangular nanoparticles have been studied in Refs. [3,10]. Furthermore, the propagation of gap plasmons in wider gaps between two parallel metal surfaces has been studied in Refs. [11–14], and when propagating in rectangular or tapered grooves in Refs. [15–17]. In all these papers, quantum spill-out is neglected, such that the dielectric function takes one value in the gap region and another value in the metal, thus changing abruptly at the interfaces.

In this paper, we focus on gaps of a few nanometers in metals, which can be found in ultrasharp groove arrays [12,18–21]. In such grooves, nearly parallel metal surfaces are separated by an ultranarrow gap near the bottom. These structures are broadband absorbers of light [19,20], and most of the absorption takes place in the bottom part of the grooves [22]. In addition, most of the physics can be explained in terms of gap plasmons propagating back and forth in the grooves. So far, the modeling of ultrasharp grooves has not taken quantum spill-out effects into account, and in the extreme limit of vanishing gap width, the resulting gap plasmon mode index diverges [12–16], which is clearly unphysical. Importantly, only minor oscillations are observed in the measured reflectance spectra from Ref. [20], which does not match present theories neglecting spill-out.

We show in this paper that by assuming local response taking quantum spill-out into account leads to a drastically improved agreement with the measured reflectance from arrays of ultrasharp grooves in gold films. Furthermore, the mode index of gap plasmons converges to the refractive index of bulk gold for vanishing gaps, thus restoring physically correct behavior. The range of electron spill-out is only about 0.3 nm, implying that when the gap width is below 0.6 nm the electron distributions from the two gold surfaces overlap, and electrons can tunnel across the gap, while the surfaces do not couple electronically for wider gaps. Surprisingly, we demonstrate in the following that the effect of spill-out also significantly increases the absorption when the gap width is a few nanometers, thus far outside the tunnel regime.

II. QUANTUM DIELECTRIC FUNCTION

In a quantum mechanical description of metal surfaces, the electron density has an exponential tail stretching into the vacuum region due to tunneling through the surface barrier. A highly efficient model of such spill-out effects is provided by density-functional theory (DFT) in the jellium approximation treating the positive ions as a constant charge density inside the metal [23,24]. The distribution of free (s,p band) electrons in this positive background produces an inhomogeneous electron density with a characteristic spill-out into the vacuum region.

The density of free electrons in the vicinity of a gap between metal surfaces is found by self-consistently solving the Kohn-Sham equations [23] within the jellium model. The exchange and correlation potentials that appear in these equations are calculated applying the local density approximation (LDA) [25], using the Perdew-Zunger parametrization in the correlation term [26]. The applied Wigner-Seitz radius for gold is $r_s = 3.01$ Bohr [24]. The electron density is calculated for a structure consisting of two parallel gold slabs of width d separated by the gap width w. The width of the slabs must be thick enough such that artificial finite-size effects are negligible, and we find

*ejs@mp.aau.dk

that $d = 2.6$ nm is sufficient. The density is calculated using a standing-wave basis of the form $\sin(n\pi(x/L + 1/2))$, where $n = 1,2,\ldots,200$ and $L = 8$ nm, and it has been checked that the density has converged with respect to the number of basis functions. A variation in Fermi energy between two iterations below 10^{-7} Ha is used as the convergence criterion in the optimization process, and an Anderson mixing scheme [27] with mixing parameter 0.005 is applied.

Such a DFT model has in other cases previously been applied to calculate the optical cross sections of metal nanowires [28–31], metal clusters and spheres [28,32–34], and the plasmon resonance of metal dimers [35,36] and of semiconductor nanocrystals [37]. In addition, it was shown in Ref. [38] that electron spill-out has a significant impact on the local density of states for gap plasmons propagating between two gold surfaces. In a similar manner, quantum tunneling was found to have a significant impact on the local field intensity in metal nanostructures [39]. In Ref. [40] different quantum mechanical effects in plasmonic structures with subnanometer gaps were studied, and it was found that spill-out plays a significant role in the optical cross sections and near-field enhancement factors of a sodium dimer, which is in agreement with the experimental results from Ref. [41]. Furthermore, nonlocality was studied in Refs. [40,42] for sodium and gold dimers, respectively, and it was found that nonlocal effects slightly blue-shift the plasmon resonances. A similar result was found in Refs. [43,44] for gold cylinders, where an analytic estimate of the blue-shift is found in Ref. [44]. Furthermore, it was found in Ref. [38] that for plasmons propagating in narrow metal gaps, the spill-out effect generally dominates compared to nonlocal effects. Reference [29] studied the optical response of metal nanowires using time-dependent DFT (TDDFT), as well as classical local and nonlocal response. It was found that by including screening as in Refs. [45–47], results for both the local and nonlocal response become in excellent agreement with results obtained using TDDFT.

The separation of the optical response into local and nonlocal effects follows rigorously from quantum mechanical response theory [48] within the random phase approximation [25]. Hence, considering surfaces normal to the x axis the following full dielectric response can be derived from the electric conductivity in Ref. [48],

$$\varepsilon(x,x') = \left(1 - \frac{e^2 n(x)}{m_e \varepsilon_0 \omega^2}\right)\delta(x - x') + \frac{i}{2\pi^2 \varepsilon_0 \hbar\omega} \int f_{nm} \frac{j_{nm}(x) j_{mn}(x')}{\omega + i\Gamma - \omega_m + \omega_n} d^2k. \quad (1)$$

Here, $n(x)$ denotes the spatially varying electron density corresponding to a position-dependent plasma frequency $\omega_p^2(x) = e^2 n(x)/(m_e \varepsilon_0)$ [49], where e and m_e denote the electron charge and mass, respectively. In the second line of Eq. (1) j_{nm} is a matrix element of the paramagnetic transition current between states n and m with energies $\hbar\omega_n$ and $\hbar\omega_m$, respectively, and the integral is over the two-dimensional electron momentum $\vec{k}$. Also, f_{nm} denotes the difference between Fermi factors for the states and Γ is a damping term.

In Eq. (1), the last term is the nonlocal one responsible for the blue-shifts in plasmon resonances observed in Refs. [40–44]. In this paper, similar to Refs. [31–33,37], we neglect such nonlocal effects and thereby treat the dielectric function as a local response. As a consequence, we can approximate $\varepsilon(x,x') = \varepsilon(x)\delta(x - x')$ with the local response $\varepsilon(x) \approx 1 - \omega_p^2(x)/\omega^2$, i.e., the Drude dielectric function. This simple expression is modified below to contain damping and interband effects. It is shown in the next two sections how this local treatment rectifies the unphysical divergence of the mode index found in classical models, and leads to a calculated reflectance from ultrasharp groove arrays in excellent agreement with measured reflectance spectra. Hence, most of the physics regarding plasmons in narrow metal gaps can be explained using a local model, and it seems sufficient only to include the quantum effect of electron spill-out and not the nonlocal effects in order to get reliable results.

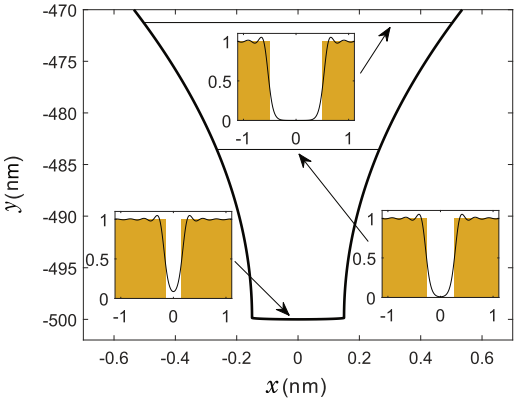

FIG. 1. Schematic of the bottom 30 nm of an ultrasharp groove with a bottom width of 0.3 nm. The three insets show the electron density n/n_0 at the bottom of the groove and for gap widths of 0.5 nm and 1 nm. The colored areas in the insets show the position of the gold surfaces.

A continuous range of widths between metal surfaces can be found in ultrasharp grooves [12,19,20]. Figure 1 shows a schematic of the bottom 30 nm of a groove with bottom width of 0.3 nm in accordance with the geometry considered in Refs. [19,20]. The three insets show the calculated electron density $n(x,w)$ in units of the bulk gold electron density n_0 at three different cross sections of the ultrasharp groove corresponding to gap widths $w = 0.3$, 0.5, and 1 nm, respectively. The colored areas mark the position of the gold surfaces. Due to spill-out, the electron density extends a small distance into the gap and, especially for the smallest gap width of 0.3 nm, the density only decreases to roughly 9% of the bulk gold density in the center of the gap, while it decreases to about 1% when the gap width is 0.5 nm. Hence, there is no true vacuum region between the gold surfaces for these gap widths. For the larger gap of 1 nm, the electron density drops practically to zero 0.3 nm from the gold surfaces, such that the two gold surfaces do not couple electronically. The range of spill-out is therefore 0.3 nm, and electrons can tunnel from one gold surface to the other only when the gap width is below 0.6 nm. In addition, the electron density inside the metal is also affected near the surface and shows Friedel oscillations, in agreement

with previous studies of the electron density across a single boundary between metal and air [23,24,50].

It is noticed that all the calculated densities only include spill-out from parallel gold slabs, and these densities are merged together in a region, where the curvature of the groove walls is small in order to form the density across the two-dimensional ultrasharp groove. Electron spill-out also occurs from the bottom of the groove and affects the density on a length scale comparable to the spill-out range. However, in this paper we ignore spill-out from the bottom, an assumption that will be explained in Sec. IV.

Next, the electron density is applied to compute the dielectric function ε across the structure as described by the local Drude model modified to include the inhomogeneous density as mentioned above. In addition to the free electrons, bound electrons are found in lower lying d bands, and they contribute to the interband part of the dielectric function [49]. In contrast to the free electrons, we assume no spill-out of bound electrons into the gap region, and they are therefore entirely located in the bulk region. In Refs. [45–47], a thin surface layer with ineffective screening has been applied to account for the interband part of the dielectric function, such that this contribution is a step function changing abruptly a few Å from the surface at the metal side of the interface. Applying this model, plasmon resonances in nanometer-size clusters of gold, silver, and copper are calculated and are in good agreement with measurements. In this paper, however, similar to Ref. [38], we ignore this thin surface layer and assume thereby that the interband part of the dielectric function is a step function changing abruptly at the interfaces between air and gold, thus at the same position as the jellium edge. It is shown in Sec. IV that this simple description of the bound electrons leads to good agreement with measurements of the reflectance of an ultrasharp groove array.

In the bulk, the electron density n_0 implies a bulk plasma frequency of $\omega_{p,\mathrm{bulk}} = \sqrt{n_0 e^2/(m_e \varepsilon_0)}$ and an accompanying Drude response $\varepsilon_{p,\mathrm{bulk}}(\omega) = 1 - \omega_{p,\mathrm{bulk}}^2/(\omega^2 + i\omega\Gamma)$ [49]. Similar to Ref. [38] we include the interband contribution by requiring the bulk response to equal the measured dielectric function $\varepsilon_{\mathrm{gold}}(\omega)$ from Ref. [51] such that the final dielectric function is given by

$$\varepsilon(\omega,x,w) = 1 - \frac{\omega_p^2(x,w)}{\omega^2 + i\Gamma\omega} + [\varepsilon_{\mathrm{gold}}(\omega) - \varepsilon_{p,\mathrm{bulk}}(\omega)]\theta(|x| - w/2). \quad (2)$$

This is a modification of the dielectric function introduced above, as it also contains both damping and interband effects. Here, the first term describes the local Drude response of free electrons with plasma frequency $\omega_p(x,w) = \sqrt{n(x,w)e^2/(m_e\varepsilon_0)}$ determined by the electron density $n(x,w)$ calculated using DFT. In addition, the damping term in gold is $\hbar\Gamma = 65.8$ meV [49] and the step function models the abrupt behavior assumed for the bound electron interband term.

An example of the dielectric function is seen in Fig. 2(a) for $w = 0.35$ nm at a wavelength of 775 nm, where the colored

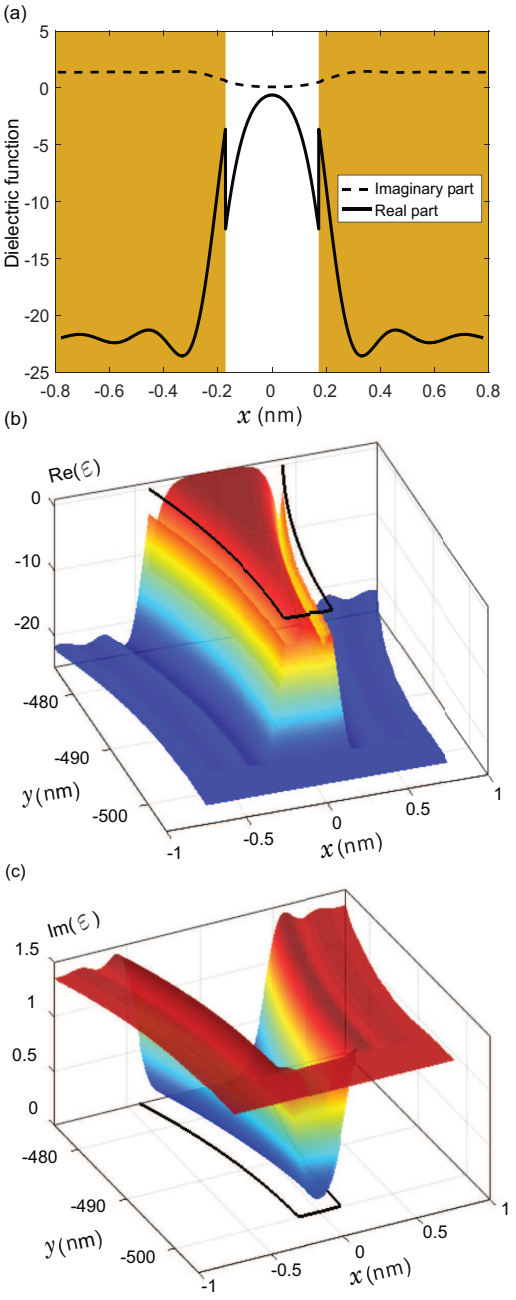

FIG. 2. (a) Real and imaginary parts of $\varepsilon(x,w)$ for $w = 0.35$ nm, where the colored areas show the position of the gold surfaces. (b) and (c) 3D plots of real and imaginary parts of $\varepsilon(x,y)$ in the bottom 25 nm of an ultrasharp groove. The wavelength is 775 nm.

areas mark the position of the gold surfaces. The real part is clearly seen to jump according to the step function in Eq. (2) and the dielectric function equals the bulk value in the gold regions that are sufficiently far from the air-gold interfaces. The imaginary part in Fig. 2(a) also jumps, but it is difficult to see in the figure. The Friedel oscillations in the electron density (Fig. 1) result in corresponding oscillations in the dielectric function [Fig. 2(a)].

Figure 2(b) shows a 3D plot of the real part of the dielectric function $\varepsilon(x,y)$ in the bottom 25 nm of the ultrasharp groove at wavelength 775 nm. The black curve in the xy plane shows the structure of the groove. A 3D plot of the imaginary part of $\varepsilon(x,y)$ is seen in Fig. 2(c). The imaginary part is small but nonzero in the entire shown region implying that absorption of light takes place even in the middle of the gap.

As spill-out from the bottom of the groove is neglected there is pure gold at positions below $y=-500$ nm as illustrated by the blue color in Fig. 2(b) and the red color in Fig. 2(c). As y increases and the groove width increases one should think that the effect of electron spill-out would become negligible as it only occurs very close to the groove walls as observed in Fig. 1. However, most surprisingly, when it comes to the mode index and the absorption density, spill-out also has a great influence for gap widths of a few nanometers, even though they far exceed the tunnel regime.

III. MODE INDEX AND ABSORPTION DENSITY OF A PROPAGATING GAP PLASMON

A gap between two metal surfaces supports gap plasmons propagating in the y direction. Gap plasmons are p-polarized electromagnetic waves, implying that the corresponding magnetic field $\vec{H}(\vec{r}\,)=\hat{z}H(x,y)$ for a constant w is given by [18]

$$H(x,y)=e^{ik_0\beta y}H(x), \tag{3}$$

where $H(x)$ is the transverse field distribution, β is the complex mode index, and $k_0=2\pi/\lambda$ is the free space wave number. Both the transverse field distribution and the mode index depend strongly on the gap width w.

A gap can also support ordinary waveguide modes, but in order for the associated wave to be propagating, the gap has to be larger than half a wavelength, and such wide gaps are not considered in this paper. For smaller gaps, waveguide modes can also exist, but then the imaginary part of the mode index is much higher than its real part, implying that the waves are exponentially damped in the waveguide [52]. They therefore play a negligible role compared to the much longer propagating plasmonic modes.

The mode index is calculated for a fixed width w by applying a transfer-matrix method [53]. This is done by dividing the x axis into N sufficiently thin layers, each modeled as having a constant dielectric function. For the mode index to converge it is found that thicknesses of 2.7×10^{-4} nm are sufficient. A structure matrix $\mathcal{S}$ is constructed, which relates the magnetic field to the left of the structure to that at the right of the structure,

$$\mathcal{S}=\mathcal{T}_{g1}\prod_{i=1}^{N-1}(\mathcal{T}_i\mathcal{T}_{i,i+1})\mathcal{T}_N\mathcal{T}_{N,g}:=\begin{pmatrix}\mathcal{S}_{11}&\mathcal{S}_{12}\\ \mathcal{S}_{21}&\mathcal{S}_{22}\end{pmatrix}. \tag{4}$$

Here, a matrix with a single index denotes propagation in that particular layer, a matrix with two indices denotes an interface matrix, and the subscript g denotes bulk gold. Expressions for the propagation and interface matrices can be found in Ref. [53]. The field to the left of the structure is then given by

$$\begin{pmatrix}0\\ H_L^-\end{pmatrix}=\mathcal{S}\begin{pmatrix}H_R^+\\ 0\end{pmatrix}. \tag{5}$$

Here, R and L denote right and left of the structure, respectively, and $+$ and $-$ denote the direction, in which light propagates. Left and right of the structure there is only light propagating in the negative and positive direction, respectively. From Eq. (5) it is found that the matrix element $\mathcal{S}_{11}$ must be zero. The matrices in Eq. (4) depend on the mode index β, implying that $\mathcal{S}_{11}$ is a function of β, and to find the roots $\mathcal{S}_{11}$ is evaluated for a range of complex β values. By observing sign changes in the real and imaginary parts of $\mathcal{S}_{11}$ a region of β values is identified where $\mathcal{S}_{11}$ is close to zero. The Newton-Raphson method is then applied to obtain the exact root. Four periods of Friedel oscillations are found sufficient in the modeling of the dielectric function near the gold surface in order for the mode index to converge, c.f. Fig. 2(a), and the bulk value is applied beyond this range.

Applying this method the mode index of a gap plasmon has been calculated using the dielectric function from the previous section. The mode index as a function of w is shown in Fig. 3(a) for $\lambda=600$ nm and in Fig. 3(b) for $\lambda=775$ nm. In Ref. [38], David and de Abajo studied gap plasmons propagating between two gold surfaces including spill-out. In that paper, in particular, the local density of states is calculated, from which it is possible to obtain a dispersion relation. Hence, a quantitative comparison with the present work can be made. From Figs. 3(a) and 3(b) in this paper, it is found that for a gap width of 1 nm, the real part of the mode index, when including spill-out, is 20.2 when $\lambda=600$ nm and 13.9 when $\lambda=775$ nm. These values correspond to parallel wave numbers (identical to $k_0\beta$) of 0.21 nm^{-1} and 0.11 nm^{-1}, respectively, which agree well with values estimated from Fig. 3(f) in Ref. [38]. The same accordance is found when neglecting spill-out and also for a gap width of 0.5 nm, which can be seen by comparing the values obtained from Figs. 3(a) and 3(b) in this paper with values estimated from Figs. 3(a), 3(b) and 3(e) in Ref. [38]. Hence, the real part of the mode index calculated in this paper is in quantitative agreement with Ref. [38]. In the present work, however, we calculate both components of the complex mode index exactly and, in contrast to Ref. [38], we show in Figs. 3(a) and 3(b) how they are explicitly dependent on w.

The red solid and dashed lines are the real and imaginary parts of the mode index obtained when including electron spill-out. The horizontal green lines represent the real and imaginary part of the refractive index n_{gold} from Ref. [51], and the mode index is seen to converge to this value for sufficiently small w. It is difficult to see in the figure, but the real part of n_{gold} shown by the solid green line is positive but very small. The blue lines show the mode index obtained when neglecting electron spill-out similar to previous studies [12–14]. Here, the mode index diverges for $w\to0$, which cannot be correct from a physical point of view. The mode index must converge to the refractive index of bulk gold for small gaps, since in this case the structure is simply bulk gold. For distances w below 0.35 nm, the mode

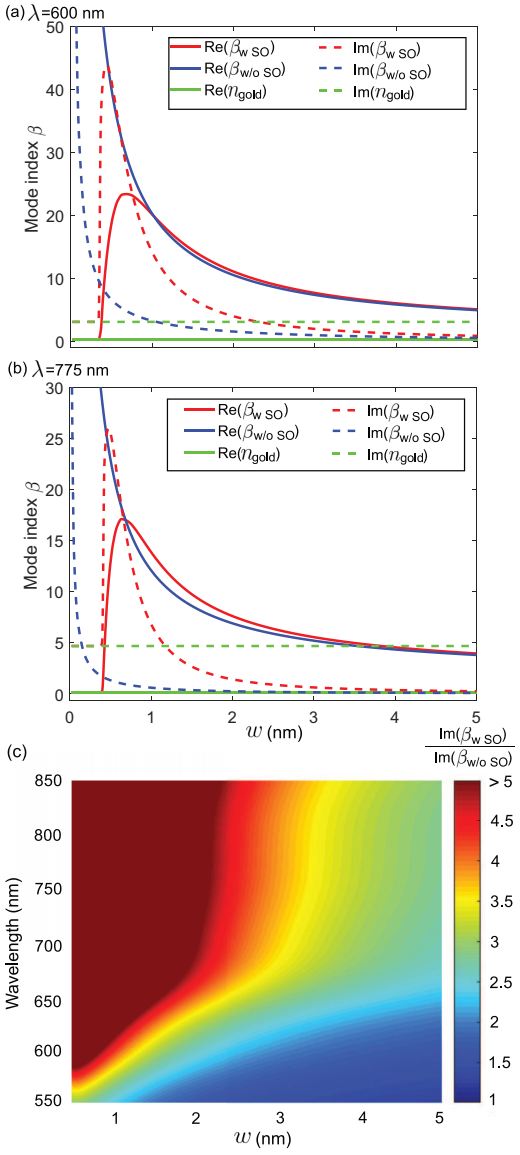

FIG. 3. (a) and (b) Real part (solid red line) and imaginary part (dashed red line) of the mode index of a gap plasmon as a function of the gap width w when taking electron spill-out into account. The horizontal green lines show the refractive index of bulk gold. The blue lines show the corresponding mode index when neglecting spill-out. The wavelength is 600 nm in (a) and 775 nm in (b). (c) Ratio between imaginary parts of the mode index with and without spill-out (SO).

index of the gap plasmon is almost the same as the refractive index of pure gold. Hence, even though the dielectric function for $w = 0.35$ nm is quite different from that of pure gold [see

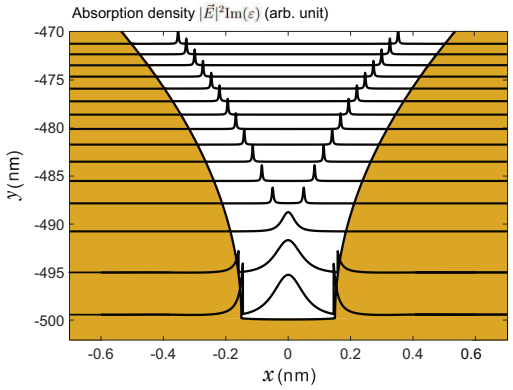

FIG. 4. Absorption density in the bottom 30 nm of an ultrasharp groove. In the bottom 10-nm absorption takes place across the entire gap, but as y increases and the groove gets broader, absorption mostly takes place 0.15 nm from the interfaces. The wavelength is 775 nm.

Fig. 2(a)], the system nevertheless behaves almost as pure gold when it comes to the mode index of a propagating gap plasmon. For large w, the mode index both with and without spill-out converges to $\sqrt{\varepsilon_{\mathrm{gold}}/(\varepsilon_{\mathrm{gold}}+1)}$ [49], as in this case the wave behaves as an SPP bound to a single interface (not shown).

On the other hand, when the gap width is a few nanometers, thus far outside the tunnel regime, the real part of the mode index is almost the same with and without spill-out, as seen by comparing the red and blue solid lines in Figs. 3(a) and 3(b). However, there is a large difference between the imaginary parts of the mode index, as seen by comparing the corresponding dashed lines. This is further illustrated in Fig. 3(c) showing the ratio of the imaginary parts of the mode index with and without spill-out (SO). For gaps of a few nanometers, the imaginary part of the mode index is seen to be much higher when including spill-out, especially for long wavelengths. In the figure, the color is dark red for every value above 5, but the maximal value is more than 20 which is found for a gap width of 0.5 nm at a wavelength of 850 nm. Hence for few-nanometer gaps the effect of spill-out on the imaginary part of the mode index is significant.

To elucidate the physics behind the increased imaginary part, the absorption density,

$$A(x,y) = |\vec{E}(x,y)|^2 \mathrm{Im}(\varepsilon(x,y)), \tag{6}$$

is calculated as shown in Fig. 4 in the bottom 30 nm of an ultrasharp groove at a wavelength of 775 nm, where the colored areas show the position of the gold. Here, the electric field has been calculated from the magnetic field in Eq. (3) as [49]

$$\vec{E} = \frac{i}{\omega\varepsilon_0\varepsilon}\vec{\nabla}\times\hat{z}H. \tag{7}$$

In the bottom of the groove, where the gap width is 0.3 nm, the absorption density clearly jumps across the boundaries in order for the normal part of the displacement field to be continuous across the interfaces [49]. It is noticed that the dielectric function on the gap side of the interface has a numerically higher value than on the gold side of the interface, which is due

to the abrupt jump in the bound electron term [see Fig. 2(a)]. In order for the displacement field to be continuous, this implies that the electric field magnitude is correspondingly lower on the gap side than on the gold side, which explains why the absorption density in the bottom of Fig. 4 increases across the interface. If spill-out is neglected, the dielectric function has a numerically higher value on the gold side of the interface, which implies that in this case the electric field magnitude drops across the interface. As y increases and the groove gets broader, the absorption density mostly consists of two peaks located about 0.15 nm from the groove walls. At these positions, the real part of the dielectric function is zero (at the wavelength 775 nm), while the imaginary part is small but nonzero, [see Figs. 2(b) and 2(c)], which ensures that the peaks in absorption density are finite and not diverging.

When neglecting spill-out the dielectric function has zero imaginary part in the gap and absorption can only take place in the metal. It is highly surprising that the effect of spill-out significantly increases the absorption density also for few-nanometer gaps that far exceed the tunnel regime. How this affects the reflectance from an ultrasharp groove array is studied in the next section.

IV. REFLECTANCE FROM AN ULTRASHARP GROOVE ARRAY

Periodic arrays of ultrasharp grooves absorb light almost perfectly in a broad wavelength interval, which is utilized in plasmonic black gold, where the grooves turn a shiny gold surface into a broadband absorber [19,20]. Rectangular or tapered (not ultrasharp) grooves may on the other hand absorb efficiently in a narrow band of wavelengths [17,54–56], which is advantageous in thermophotovoltaics [57,58]. Hence, the reflectance spectrum of an array of grooves strongly depends on the groove shape. Previously the reflectance of a groove array has only been calculated when neglecting spill-out, but in this section it is calculated when taking spill-out into account. This is done by applying the stack matrix method (SMM) of Ref. [18], where the grooves are divided into layers with a refractive index corresponding to the gap plasmon mode index. In Ref. [18], the SMM was applied to calculate reflectance from ultrasharp groove arrays when neglecting spill-out, and the results were practically identical to results obtained using a full Greens function surface integral equation method.

As mentioned in the previous section, the mode index of the propagating gap plasmon is almost the same as in pure gold for gap widths below 0.35 nm [see Figs. 3(a) and 3(b)] even though the dielectric function is quite different from that of pure gold [see Fig. 2(a)]. By looking at the groove structure in Fig. 1, it is found that the distance between the groove walls is below 0.35 nm at the bottom 8 nm of the groove, thus the gap plasmon behaves almost as pure gold in this region. As mentioned in Sec. II electron spill-out also occurs from the bottom and affects the density in a short range from the bottom. But, as the range of spill-out is much shorter than the 8 nm, it only influences the dielectric function at positions where the mode index already behaves almost as pure gold and where minor variations in dielectric function implies no change in mode index. This explains why we have reasons to ignore the spill-out from the bottom of the groove.

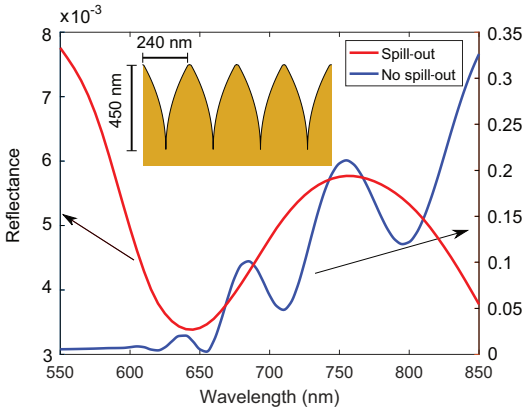

FIG. 5. Reflectance from an ultrasharp groove array in gold where the grooves have a top width of 240 nm, a bottom width of 0.3 nm, and a groove height of 450 nm. The grooves are illuminated by normally incident light. Spill-out is included in the calculated reflectance shown by the red line on the left y axis, and neglected in the corresponding reflectance shown by the blue line on the right y axis. The inset shows a schematic of the groove geometry.

The array of ultrasharp grooves in gold is illustrated in the inset in Fig. 5. Here the reflectance from an ultrasharp groove array illuminated by normally incident light is shown for the wavelength interval 550–850 nm, where the groove height is 450 nm, the top width is 240 nm, and the bottom width is 0.3 nm. Including spill-out leads to the reflectance shown by the red line on the left y axis in Fig. 5. The same groove dimensions were applied in Ref. [20] where electron spill-out was neglected, which gives the reflectance shown by the blue line on the right y axis in the same figure. By noticing the very different scale on the left and right axes in the figure, it is clearly seen that the effect of spill-out significantly lowers the reflectance from an ultrasharp groove array in gold. This is expected since the imaginary part of the gap plasmon mode index is higher when including spill-out [see Fig. 3(c)].

The mode index and, thus, the degree of absorption depends on how the field profile is distributed between the gap and the metal regions. It is therefore investigated how spill-out affects the field profile for a small gap. The electric field of the gap plasmon in Eq. (7) has both an x and a y component, where the x component jumps across the interface as observed in the bottom of Fig. 4. The corresponding magnetic field in Eq. (3) only has a z component, which is continuous across the interface [49], and makes the magnetic field preferable for illustrating the penetration of the field into the metal.

The gap plasmon transverse field magnitude $|H(x)|$ is shown in Fig. 6(a) for a gap width of $w = 0.45$ nm both with and without spill-out at a wavelength of 775 nm. The field distributions have been normalized by their maximum value. The blue line shows the field when neglecting spill-out. In this case the mode index β is high [see blue lines in Fig. 3(b)], which implies that the imaginary part of $k_x = k_0\sqrt{n_{\text{gold}}^2 - \beta^2}$ is also high. The field is therefore partly localized in the air

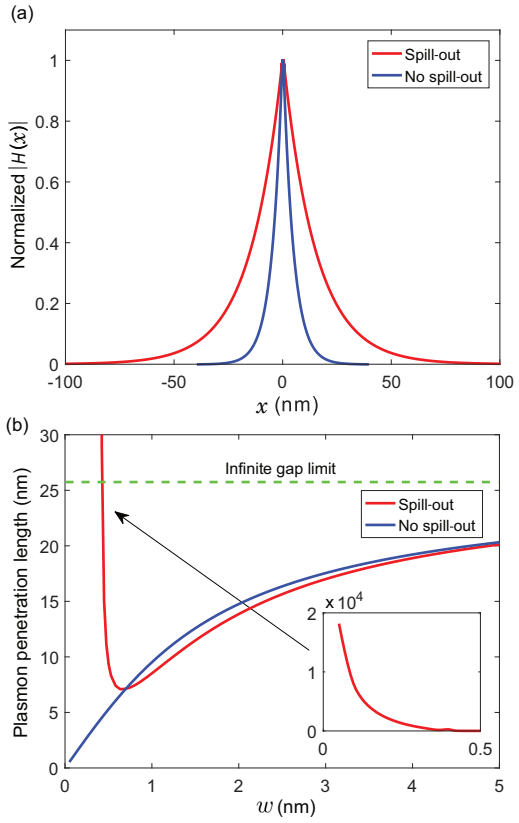

FIG. 6. (a) Normalized $|H(x)|$ for a gap plasmon propagating in a gap of width $w = 0.45$ nm. Spill-out is included and neglected in the calculations of the fields shown by the red and blue lines, respectively. (b) Corresponding plasmon penetration length as a function of w with and without spill-out. Horizontal green line shows the penetration length of an SPP bound to a single gold surface, i.e., in the limit of an infinite gap. The wavelength is 775 nm in both (a) and (b).

gap region and the penetration length calculated as $1/\mathrm{Im}(k_x)$ is only 4.8 nm.

When including spill-out for the same w the mode index is closer to that of pure gold [see red lines in Fig. 3(b)], and k_x is therefore smaller, which yields a more delocalized field as shown by the red line in Fig. 6(a). Here the penetration length is 15 nm, and most of the field is therefore located in the pure gold regions. But as the absorption density for this w consists of two peaks, as in the upper part of Fig. 4, absorption mostly takes place 0.15 nm from the interfaces and not in the pure gold regions, even though the magnetic field in Fig. 6(a) is mostly located there. The figure shows that when including spill-out the field profile becomes approximately three times broader. This effect becomes more pronounced for smaller w, and for $w = 0.35$ nm the field is almost 100 times broader, which is due to the fact that the mode index is almost equal to the refractive index of bulk gold when including spill-out.

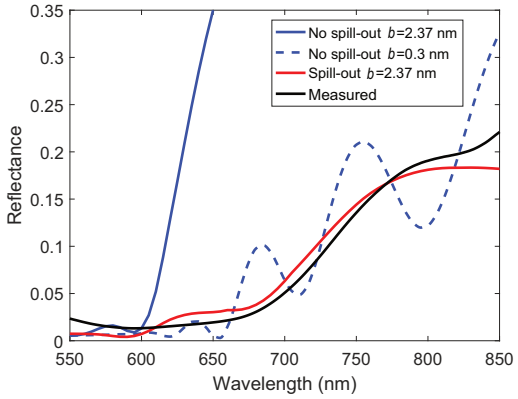

FIG. 7. Comparison between the calculated reflectance with and without spill-out and the measured reflectance from Ref. [20] shown in black. Spill-out is included in the calculated reflectance shown by the red line and neglected in the calculated reflectance shown by the blue lines. The groove height is 450 nm, the top width is 240 nm, and b denotes the bottom width.

The penetration length as a function of w is shown in Fig. 6(b) with and without spill-out at a wavelength of 775 nm. The blue line shows the case when neglecting spill-out, and for small w the field is highly localized implying a very short penetration length, which was also observed in Fig. 6(a). When including spill-out the penetration length shown by the red line for small w becomes very high and diverges as shown in the inset. Thus, the field behaves almost like a plane wave in the limit $w \to 0$. On the other hand, when w increases the penetration length becomes almost the same with and without spill-out. Then, at a first glance, the effect of spill-out seems to be negligible for gaps of a few nanometers, but as was found in the previous section significant absorption takes place in the gap region 0.15 nm from the interfaces. It is astonishing that even though the field penetration into the gold surfaces is almost the same with and without spill-out for few-nanometer gaps the absorption is still much higher when including spill-out. The horizontal green line in Fig. 6(b) shows the penetration length of an SPP bound to a single interface between gold and air [49]. When the gap becomes sufficiently wide the gap plasmon behaves almost as a single SPP bound to a gold surface, and the penetration lengths shown by the red and blue lines are found to converge to the horizontal green line for large w (not shown).

The bottom width b of an ultrasharp groove is impossible to measure precisely. In the calculations in Refs. [19,20], where spill-out was neglected, the bottom width was set to 0.3 nm, which is close to the gold atom diameter. This was necessary in order to obtain a reflectance comparable to measured values. However, it was found in Fig. 5 that the reflectance from an ultrasharp groove array with this bottom width is significantly lower when spill-out is included. It is therefore investigated in Fig. 7, which impact the bottom width has on the calculated reflectance of a 450-nm deep ultrasharp groove array with

top width 240 nm. For bottom widths below 1 nm the reflectance is always below 1.3% (not shown), being thus still significantly lower than both measured and calculated values in Refs. [19,20]. By adjusting b to minimize the root mean square error between calculated and measured reflectance, it is found that the bottom width $b = 2.37$ nm gives the best reflectance. The calculated reflectance when including spill-out for this b is shown by the red line in Fig. 7, and is in excellent agreement with the measured reflectance from Ref. [20] shown in black in the same figure.

The blue dashed and solid lines in Fig. 7 show the calculated reflectance when neglecting spill-out for bottom widths of 0.3 nm and 2.37 nm, respectively. For $b = 2.37$ nm the reflectance is now much higher than the measured reflectance. The result for 0.3 nm, also shown in Fig. 5, is similar in magnitude to the measured reflectance shown in black but the oscillations in the calculated reflectance spectrum are clearly not present in the measured reflectance. On the other hand such oscillations are not present in the calculation that includes spill-out, in which case a much better agreement with the measured reflectance is obtained.

In Ref. [20], the reflectance has been measured for several different fabricated arrays of ultrasharp grooves. In the theoretical calculations performed in this paper and in Refs. [19,20] it has been assumed that all the grooves in the periodic array are identical. This is extremely hard to guarantee in practice when fabricating the grooves, and SEM and optical microscope images of the arrays of grooves in Ref. [20] also show that there are minor variations. The fact that the fabricated groove arrays are not perfectly periodic may explain some of the deviation between the calculated reflectance including spill-out and the measured reflectance.

V. CONCLUSION

Using a quantum mechanical approach and assuming local response, the properties of gap plasmons in ultranarrow metal gaps have been investigated. Electron spill-out is found to play a crucial role for both plasmon propagation and reflectance from ultrasharp groove arrays. In these geometries, a classical approach based on bulk optical properties leads to unphysically diverging mode indices in the limit of vanishing gap width. We demonstrate, however, that divergencies are avoided when spill-out is taken into account. Importantly, spill-out also has a great impact on gaps of a few nanometers, since power is strongly absorbed in the gap region 1–2 Å from the interfaces. As a consequence, calculated reflectance spectra are in excellent agreement with measured reflectance spectra for ultrasharp groove arrays.

ACKNOWLEDGMENT

This work is supported by Villum Kann Rasmusen (VKR) center of excellence QUSCOPE.

[1] P. Lalanne, J. Hugonin, H. Liu, and B. Wang, Surf. Sci. Rep. **64**, 453 (2009).
[2] B. Wang and P. Lalanne, J. Opt. Soc. Am. A **27**, 1432 (2010).
[3] D. C. Marinica, M. Zapata, P. Nordlander, A. K. Kazansky, P. M. Echenique, J. Aizpurua, and A. G. Borisov, Sci. Adv. **1**, 1501095 (2015).
[4] J. A. Schuller, E. S. Barnard, W. Cai, Y. C. Jun, J. S. White, and M. L. Brongersma, Nat. Mater. **9**, 193 (2010).
[5] D. K. Gramotnev and S. I. Bozhevolnyi, Nat. Photon. **4**, 83 (2010).
[6] M. Kuttge, F. de Abajo, and A. Polman, Opt. Express **17**, 10385 (2009).
[7] H. A. Atwater and A. Polman, Nat. Mater. **9**, 205 (2010).
[8] P. Berini and I. D. Leon, Nat. Photon. **6**, 16 (2012).
[9] J. Feng, V. Siu, A. Roelke, V. Mehta, S. Rhieu, G. Tayhas, R. Palmore, and D. Pacifici, Nano Lett. **12**, 602 (2011).
[10] G. Haberfehlner, F.-P. Schmidt, G. Schaffernak, A. Hörl, A. Trügler, A. Hohenau, F. Hofer, J. Krenn, U. Hohenester, and G. Kothleitner, Nano Lett. **17**, 6773 (2017).
[11] F. J. García-Vidal, L. Martin-Moreno, T. Ebbesen, and L. Kuipers, Rev. Mod. Phys. **82**, 729 (2010).
[12] T. Søndergaard and S. I. Bozhevolnyi, New J. Phys. **15**, 013034 (2013).
[13] S. I. Bozhevolnyi and J. Jung, Opt. Express **16**, 2676 (2008).
[14] C. L. C. Smith, N. Stenger, A. Kristensen, N. A. Mortensen, and S. I. Bozhevolnyi, Nanoscale **7**, 9355 (2015).
[15] S. I. Bozhevolnyi, Opt. Express **14**, 9467 (2006).
[16] Z. Han and S. I. Bozhevolnyi, Rep. Prog. Phys. **76**, 016402 (2013).
[17] A. S. Roberts, T. Søndergaard, M. Chirumamilla, A. Pors, J. Beermann, K. Pedersen, and S. I. Bozhevolnyi, Phys. Rev. B **93**, 075413 (2016).
[18] M. Odgaard, M. G. Laursen, and T. Søndergaard, J. Opt. Soc. Am. B **31**, 1853 (2014).
[19] E. J. H. Skjølstrup and T. Søndergaard, J. Opt. Soc. Am. B **34**, 673 (2017).
[20] T. Søndergaard, S. Novikov, T. Holmgaard, R. Eriksen, J. Beermann, Z. Han, K. Pedersen, and S. I. Bozhevolnyi, Nat. Commun. **3**, 969 (2012).
[21] S. Raza, N. Stenger, A. Pors, T. Holmgaard, S. Kadkhodazadeh, J. Wagner, K. Pedersen, M. Wubs, S. I. Bozhevolnyi, and N. A. Mortensen, Nat. Commun. **5**, 1 (2014).
[22] T. Søndergaard and S. I. Bozhevolnyi, Opt. Lett. **41**, 2903 (2016).
[23] N. D. Lang and W. Kohn, Phys. Rev. B **1**, 4555 (1970).
[24] N. D. Lang and W. Kohn, Phys. Rev. B **3**, 1215 (1971).
[25] J. Kohanoff, *Electronic Structure Calculations for Solids and Molecules—Theory and Computational Methods*, 1st ed. (Cambridge University Press, Cambridge, 2006).
[26] J. P. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1981).
[27] V. Eyert, J. Comput. Phys. **124**, 271 (1996).
[28] W. Yan, M. Wubs, and N. A. Mortensen, Phys. Rev. Lett. **115**, 137403 (2015).
[29] T. V. Teperik, P. Nordlander, J. Aizpurua, and A. G. Borisov, Phys. Rev. Lett. **110**, 263901 (2013).

[30] L. Stella, P. Zhang, F. J. García-Vidal, A. Rubio, and P. García-Gonzalez, J. Phys. Chem. C **117**, 8941 (2013).
[31] D. C. Marinica, A. K. Kazansky, P. Nordlander, J. Aizpurua, and A. G. Borisov, Nano Lett. **12**, 1333 (2012).
[32] O. M. Bakr, V. Amendola, C. M. Aikens, W. Wenseleers, R. Li, L. D. Negro, G. C. Schatz, and F. Stellacci, Angew. Chem. **48**, 5921 (2009).
[33] M. Zhu, C. M. Aikens, F. J. Hollander, G. C. Schatz, and R. Jin, J. Am. Chem. Soc. **130**, 5883 (2008).
[34] R. Zhang, L. Bursi, J. Cox, Y. Cui, C. Krauter, A. Alabastri, A. Manjavacas, A. Calzolari, S. Corni, E. Molinari, E. Carter, F. de Abajo, H. Zhang, and P. Nordlander, ACS Nano **11**, 7321 (2017).
[35] G. Maroulis, J. Chem. Phys. **121**, 10519 (2004).
[36] J. Zuloaga, E. Prodan, and P. Nordlander, Nano Lett. **9**, 887 (2009).
[37] H. Zhang, V. Kulkarni, E. Prodan, P. Nordlander, and A. O. Govorov, J. Phys. Chem. C **118**, 16035 (2014).
[38] C. David and F. de Abajo, ACS Nano **8**, 9558 (2014).
[39] G. Hajisalem, M. S. Nezami, and R. Gordon, Nano Lett. **14**, 6651 (2014).
[40] W. Zhu, R. Esteban, A. G. Borisov, J. J. Baumberg, P. Nordlander, H. J. Lezec, J. Aizpurua, and K. B. Crozier, Nat. Commun. **7**, 11495 (2016).
[41] K. J. Savage, M. M. Hawkeye, R. Esteban, A. G. Borisov, J. Aizpurua, and J. J. Baumberg, Nature (London) **491**, 574 (2012).
[42] F. de Abajo, J. Phys. Chem. C **112**, 17983 (2008).
[43] S. Raza, G. Toscano, A. Jauho, M. Wubs, and N. A. Mortensen, Phys. Rev. B **84**, 121412(R) (2011).
[44] W. Yan, N. A. Mortensen, and M. Wubs, Phys. Rev. B **88**, 155414 (2013).
[45] E. Cottancin, G. Celep, J. Lermé, M. Pellarin, J. R. Huntzinger, J. L. Vialle, and M. Broyer, Theor. Chem. Acc. **116**, 514 (2006).
[46] V. V. Kresin, Phys. Rev. B **51**, 1844 (1995).
[47] A. Liebsch, Phys. Rev. B **48**, 11317 (1993).
[48] P. J. Feibelman, Phys. Rev. B **12**, 1319 (1975).
[49] L. Novotny and B. Hecht, *Principles of Nano-Optics*, 2nd ed. (Cambridge University Press, Cambridge, 2012).
[50] W. Yan, Phys. Rev. B **91**, 115416 (2015).
[51] P. B. Johnson and R. W. Christy, Phys. Rev. B **6**, 4370 (1972).
[52] J. A. Kong, *Electromagnetic Wave Theory*, 2nd ed. (John Wiley & Sons, New York, 1990).
[53] M. V. Klein and T. E. Furtak, *Optics*, 2nd ed. (John Wiley & Sons, New York, 1986).
[54] H. Miyazaki, K. Ikeda, T. Kasaya, K. Yamamoto, Y. Inoue, K. Fujimura, T. Kanakugi, M. Okada, K. Hatade, and S. Kitagawa, Appl. Phys. Lett. **92**, 141114 (2008).
[55] J. Greffet, R. Carminati, K. Joulain, J. Mulet, S. Mainguy, and Y. Chen, Nature (London) **416**, 61 (2002).
[56] E. J. H. Skjølstrup, T. Søndergaard, K. Pedersen, and T. G. Pedersen, J. Nanophoton. **11**, 046023 (2017).
[57] T. Bauer, *Thermophotovoltaics—Basic Principles and Critical Aspects of System Design*, 1st ed. (Springer Verlag, Berlin, 2011).
[58] H. Sai and H. Yugami, Appl. Phys. Lett. **85**, 3399 (2004).

Paper III.

Paper IV

Quantum spill-out in nanometer-thin gold slabs: Effect on the plasmon mode index and the plasmonic absorption

Enok J. H. Skjølstrup, Thomas Søndergaard, and Thomas G. Pedersen

The paper has been published in
Physical Review B **99**, 155427 (2019).

The layout has been revised.

Quantum spill-out in nanometer-thin gold slabs: Effect on the plasmon mode index and the plasmonic absorption

Enok J. H. Skjølstrup,* Thomas Søndergaard, and Thomas G. Pedersen
Department of Materials and Production, Aalborg University, Skjernvej 4A, DK-9220 Aalborg East, Denmark

(Received 12 October 2018; published 26 April 2019; corrected 26 April 2019)

A quantum mechanical approach and local response theory are applied to study plasmons propagating in nanometer-thin gold slabs sandwiched between different dielectrics. The metal slab supports two different kinds of modes, classified as long- and short-range plasmons. Quantum spill-out is found to significantly increase the imaginary part of their mode indices, and, surprisingly, even for slabs wide enough to approach bulk the increase is 20%. This is explained in terms of enhanced plasmonic absorption, which mainly takes place in narrow regions located near the slab surface.

DOI: 10.1103/PhysRevB.99.155427

I. INTRODUCTION

Recently, it was found that the effect of quantum spill-out in nanometer-thin gaps in gold has a significant impact on the propagation of surface plasmon polaritons (SPPs) in such structures. In the limit of vanishing gap, the SPP mode index was found to converge to the refractive index of bulk gold [1], while classical models neglecting spill-out find a diverging mode index [2,3]. In addition, it was discovered in Ref. [1] that spill-out significantly increases the plasmonic absorption in these gaps. Furthermore, the predicted reflectance from an ultrasharp groove array is in much better agreement with measurements [4] than the classical model [5,6].

In this paper, we study the opposite geometry, i.e., a nanometer-thin gold slab surrounded by different dielectrics. Such a structure supports long- and short-range SPPs, which are *p*-polarized electromagnetic waves bound to and propagating along the slab [7–11]. For a nanometer-thin slab, the short-range mode is strongly bound, meaning that a large part of the field profile is located in the slab region, while the long-range mode is weakly bound with most of its field profile located in the dielectric regions. The magnetic fields of the modes are symmetric and antisymmetric, respectively, if the metal slab is sandwiched between identical dielectrics, while the symmetry is broken when sandwiched between different dielectrics [9]. Applications of such SPPs are found in, e.g., plasmonic lenses for biosensors and as mode couplers into dielectric or plasmonic waveguides [12,13]. In addition, plasmonic structures find applications within, e.g., solar cells [14] and, furthermore, they can be applied to squeeze light below the diffraction limit [15,16], and can be utilized in lasers [17].

Plasmons supported by metal slabs have previously been studied using both classical and quantum models [7–11,18–21]. Thus, Refs. [7–11] applied a classical model neglecting quantum spill-out, such that the dielectric function takes one value in the metal region and another value in the dielectric region, thus changing abruptly at the interfaces, while quantum effects have been included in, e.g., Refs. [18–21]. Furthermore, gold films with thicknesses down to 1 nm have recently been fabricated [22].

Here, we examine the effect of quantum spill-out on plasmons propagating in nanometer-thin gold slabs. Local response theory is applied to calculate the mode indices and associated electromagnetic fields. We show that spill-out significantly increases the imaginary part of the mode index, even for slabs wide enough to approach bulk. This is explained in terms of strong plasmonic absorption mainly taking place a few Å from the slab surface, a phenomenon not found in classical models.

II. QUANTUM DIELECTRIC FUNCTION

In the vicinity of the gold slab, the electron density and the effective potential arising from the free electrons (in the *s*, *p* band) are significantly modified due to electron tunneling through the surface barrier. To capture this effect, we calculate the electron density using density-functional theory (DFT) in the jellium model [23,24] (see Appendix A for further description). The optical cross sections of metal nanowires [25–28], metal clusters, and spheres [25,29–31] have previously been calculated by applying such a DFT model in the jellium approximation. Likewise the plasmon resonances of metal dimers and semiconductor nanocrystals have been calculated in Refs. [32–34], while Ref. [20] studied the plasmonic properties of ultrathin metal films. In addition, Ref. [35] examined the role of electron spill-out and nonlocal effects on the plasmon dispersion relation for gap plasmons propagating between two gold surfaces as well as plasmons propagating in gold slabs surrounded by air. It was found in that paper that spill-out has a significant impact while the influence from nonlocal effects was minor. In Refs. [20,35], only the real part of the parallel wave number (analogous to mode index) was considered, with no studies of the dependence of the slab (or gap) width on plasmon propagation. In this paper, in contrast, we compute both real and imaginary parts of the mode index

*ejs@mp.aau.dk

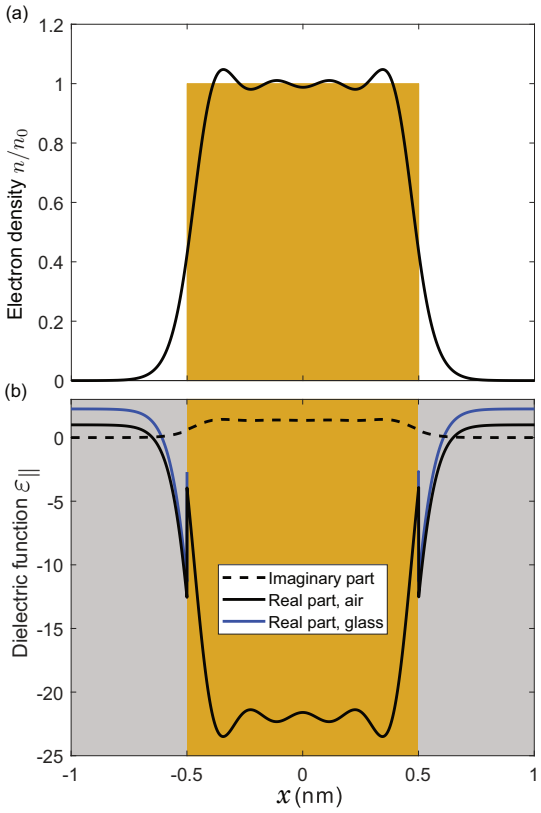

FIG. 1. (a) Electron density in units of the bulk gold density n_0 across a gold slab of width 1 nm. (b) Real and imaginary parts of the dielectric function ε across the same gold slab at a wavelength of 775 nm, with the shaded areas indicating the surrounding dielectrics. The solid black (blue) curve is for a slab surrounded by air (glass) on both sides. The imaginary part is unaffected by the kind of dielectric. In both figures, the colored areas represent the ion charge.

and, furthermore, investigate in detail how they depend on the slab width.

The electron density n across a gold slab of width $d = 1$ nm is shown in Fig. 1(a) in units of the bulk gold density n_0, where the geometry is chosen such that the x axis is perpendicular to the slab, while the plasmons are propagating in the y direction. The colored area in the figure shows the position of the ion charge in the jellium model, and spill-out is clearly seen to occur as the electron density contains an exponential tail that stretches $\sim$0.3 nm into the dielectric region. In addition, charge is conserved ($\int n(x)dx = n_0 d$), and the density inside the slab is thus also affected by spill-out. In Fig. 1(a) the integrated value of the spilled-out charge is $0.05 n_0 d$. As the slab gets wider, the electron density near the slab boundary contains Friedel oscillations in agreement with Refs. [23,24,36] where the electron density at a single interface between gold and air was studied.

The dielectric response function has been studied in several papers [37–43] using a fully quantum mechanical approach within the random phase approximation [44]. As the metal slab is a two-dimensional (2D) material, it only has boundaries in the direction perpendicular to the slab, implying that the response is anisotropic. The tensorial dielectric function $\overleftrightarrow{\varepsilon}(x, x')$ can be obtained from the electric conductivity in Ref. [43] as

$$\overleftrightarrow{\varepsilon}(x, x') = \left(1 - \frac{e^2 n(x)}{m_e \varepsilon_0(\omega^2 + i\Gamma\omega)}\right)\delta(x - x')\overleftrightarrow{I} - \frac{1}{2\pi^2 \varepsilon_0 \hbar\omega^2}\sum_{n,m}\int f_{mn}(\vec{k}_\parallel) \times \frac{\vec{j}_{nm}(\vec{k}_\parallel, x)\vec{j}_{mn}(\vec{k}_\parallel, x')}{\omega + i\Gamma + \omega_m - \omega_n} d^2 k_\parallel. \quad (1)$$

Here, the first line is a local isotropic Drude term, where the spatially varying electron density $n(x)$ corresponds to a position-dependent plasma frequency $\omega_p^2(x) = e^2 n(x)/(m_e \varepsilon_0)$, where e and m_e denote the electron charge and mass, respectively. In the second line of Eq. (1), $f_{nm}(\vec{k}_\parallel)$ denotes the difference between Fermi factors of states with energies $\hbar\omega_n$ and $\hbar\omega_m$, respectively, where $\vec{k}_\parallel$ is in the (y, z) plane, thus parallel to the slab. Furthermore, $\vec{j}_{nm}$ is a matrix element of the transition current (see Appendix B for an explicit expression) and Γ is a damping term. The second line of Eq. (1) is denoted $\overleftrightarrow{\varepsilon}_{NL}(x, x')$. Equation (1) shows, thus, that rigid quantum mechanical response theory implies a dielectric response function that is separable into a local isotropic and a nonlocal anisotropic term. The nonlocality has been studied in several papers [18,21,45–48], and for metal dimers and cylinders it is found to only slightly blue-shift the plasmon resonances, illustrating, as in Ref. [35], that nonlocality is a relatively small modification to the dielectric response.

For the short-range mode, the x and y components of the electric field are odd and even, respectively [7,8]. From the expressions of $\vec{j}_{nm}$ in Appendix B this implies that the nonlocal part of the displacement field $\vec{D}_{NL}(x) = \varepsilon_0 \int \overleftrightarrow{\varepsilon}_{NL}(x, x') \cdot \vec{E}(x')dx'$ is zero on average across the slab, where the contribution from the x component of the field is zero because the field is odd, while the contribution from the y component is zero due to orthogonality of the wave functions applied in $\vec{j}_{nm}$. For the y component the same applies for the long-range mode. Based on symmetry arguments, this explains the physics behind the small impact from the nonlocal effects on the short-range mode. The same phenomenon has previously been discussed for other cases in Refs. [18,21,35,45–48].

For the long-range mode, the x component of the field is almost constant across the slab [7,8], which implies that the contribution to $\vec{D}_{NL}$, which is nonzero on average, is reduced to an integral of the $\hat{x}\hat{x}$ component of $\overleftrightarrow{\varepsilon}_{NL}$ with respect to x'. In this paper, the nonlocal part of the dielectric function is taken into account by an averaging procedure, which leads to an anisotropic step function that gives the correct net response from the slab for both kinds of modes. It is modeled as a piecewise constant function taking the value $\iint \overleftrightarrow{\varepsilon}_{NL} dx\, dx'/d$ inside the slab and zero outside (see Appendix B for details). The effect of spill-out is included

in the first line of Eq. (1), while the effects of interband transitions and dielectric substrates will be incorporated by modifying the local dielectric function as shown below.

In general, the dielectric tensor is given by $\overleftrightarrow{\varepsilon} = \mathrm{diag}(\varepsilon_\perp, \varepsilon_\parallel, \varepsilon_\parallel)$ [49], where the parallel part $\varepsilon_\parallel$ is similar to the one in Ref. [1]. In the bulk, the electron density n_0 implies a bulk plasma frequency of $\omega_{p,\mathrm{bulk}} = \sqrt{n_0 e^2/(m_e \varepsilon_0)}$, which gives rise to a Drude response $\varepsilon_{p,\mathrm{bulk}}(\omega) = 1 - \omega^2_{p,\mathrm{bulk}}/(\omega^2 + i\Gamma\omega)$ [50]. Bound electrons in the lower-lying d bands also contribute to the dielectric function [50], but in contrast to the free electrons, we assume, as in Ref. [35], that they are entirely located in the jellium region, thus not tunneling through the potential barrier. The response from the bound electrons is calculated from the experimental response of bulk gold $\varepsilon_{\mathrm{gold}}(\omega)$ from Ref. [51] as $\varepsilon_{\mathrm{bound}}(\omega) = \varepsilon_{\mathrm{gold}}(\omega) - \varepsilon_{p,\mathrm{bulk}}(\omega)$. The parallel part of the final dielectric tensor in the vicinity of a gold slab with a jellium region spanning from $x = -d/2$ to $d/2$ is therefore given by

$$\varepsilon_\parallel(\omega, x) = 1 - \frac{\omega_p^2(x)}{\omega^2 + i\Gamma\omega} + [\varepsilon_s(x) - 1]\theta(|x| - d/2) + \varepsilon_{\mathrm{bound}}(\omega)\theta(d/2 - |x|). \quad (2)$$

Here, the first term describes the local Drude response of free electrons with position-dependent plasma frequency $\omega_p(x) = \sqrt{n(x)e^2/(m_e\varepsilon_0)}$ determined by the electron density $n(x)$ calculated using DFT. Also, $\hbar\Gamma = 65.8$ meV has been applied for the damping term [50]. The dielectric substrate and superstrate, which in general can be different, are described by $\varepsilon_s(x)$, and the Heaviside step function θ in the first line makes sure that the dielectric function sufficiently far from the slab equals the correct values in the substrate and superstrate. Hence, it has been assumed that the electron density across the slab does not depend on the kind of substrate and superstrate it is surrounded by. Lastly, the abrupt behavior assumed for the bound electron term is modeled with the same step function θ as used to describe the anisotropic response discussed above. The perpendicular part of the dielectric tensor is written as

$$\varepsilon_\perp(\omega, x) = \varepsilon_\parallel(\omega, x) + \varepsilon_{\mathrm{ani}}(\omega, d)\theta(d/2 - |x|), \quad (3)$$

where the expression for the anisotropic term $\varepsilon_{\mathrm{ani}}(\omega, d)$ is given in Appendix B. Anisotropy has previously been included in Refs. [21,49] but neglected in several other papers [1,19,20,28–30].

An example of the parallel part of the dielectric function is seen in Fig. 1(b) for a slab width of 1 nm at a wavelength of 775 nm. For a gold slab placed on a glass substrate with air as superstrate, the blue curve to the left and the black curve to the right of the slab describe the dielectric function in the glass and air, respectively. The real part of the dielectric function is clearly seen to jump at the slab boundary due to the step function in the second line of Eq. (2). Although it is difficult to see in the figure, the imaginary part of the dielectric function also jumps across the interfaces. Since the substrate and superstrate are assumed lossless, the imaginary part of the dielectric function is unaffected by these materials.

III. MODE INDEX OF PROPAGATING PLASMONS

The magnetic field of the SPPs only has a z component and is given by [6]

$$\vec{H}_m(\vec{r}) = \hat{z}H_m(x, y) = \hat{z}\,\exp(ik_0\beta_m y)H_m(x), \quad (4)$$

where the subscripts $m = \{l, s\}$ indicate that the field and associated complex mode index β can be either long range or short range, respectively, in agreement with Ref. [9]. In Eq. (4), $k_0 = 2\pi/\lambda$ is the free-space wave number, and $H_m(x)$ is the transverse magnetic field distribution. Both the mode index and the transverse magnetic field depend strongly on d, especially for the short-range mode, as will be shown below.

The plasmon mode indices are found as poles in the reflection coefficient [50,52], which is calculated by a transfer matrix method relating the magnetic fields to the left and right of the structure [1] (see Appendix B for the incorporation of anisotropy and Appendix C for classification of modes).

Figures 2 and 3 show, respectively, the short- and long-range mode indices as a function of d at a wavelength of 775 nm, where the response is isotropic in Fig. 2. For the blue and red curves in Fig. 2(a), the geometric structure is symmetric, while it is asymmetric for the green curves in Fig. 2(b), as indicated in the text above each subfigure. This implies that the magnetic fields associated with the blue and red curves are antisymmetric, while the symmetry of the associated fields is broken for the corresponding green curves, which will be demonstrated in the next section.

It is found that the relative effect of anisotropy on both components of the short-range mode indices is small, why Fig. 2 only shows the mode indices for an isotropic response. Mode indices calculated with an anisotropic response are instead shown for the long-range mode in Fig. 3. The results in Fig. 2 are in agreement with previous studies [7–9] when spill-out is neglected. The real parts are almost unaffected by spill-out, while it plays a significant role for the imaginary parts, as will be elaborated upon below. A similar calculation of the mode index with and without spill-out for gap plasmons propagating in narrow gaps in gold showed that the mode index when including spill-out converges to the refractive index of bulk gold in the limit of vanishing gap width [1], while neglecting spill-out leads to an unphysically diverging mode index [2,3]. For plasmons bound to the slab, the mode index when neglecting spill-out also diverges unphysically in the limit of vanishing slab thickness [7–9]. This is not the case with spill-out included, as plasmonic modes only exist when the real part of the metal dielectric constant is negative in some region along the direction normal to the slab [50]. It is found that for a slab of subatom thickness (∼0.3 Å), the electron density becomes so delocalized that the real part of the dielectric constant is everywhere positive. Hence, with spill-out included, the mode index does not diverge in the limit of vanishing slab thickness. Instead, plasmonic modes cease to exist for slabs below a cutoff thickness in the subatom range. However, since a gold atom has a diameter of roughly 0.3 nm [4], we only consider slab widths larger than this value.

For an asymmetric structure with glass as substrate and air as superstrate, the short-range mode index is shown in Fig. 2(b). Here, it is difficult to see the difference in mode

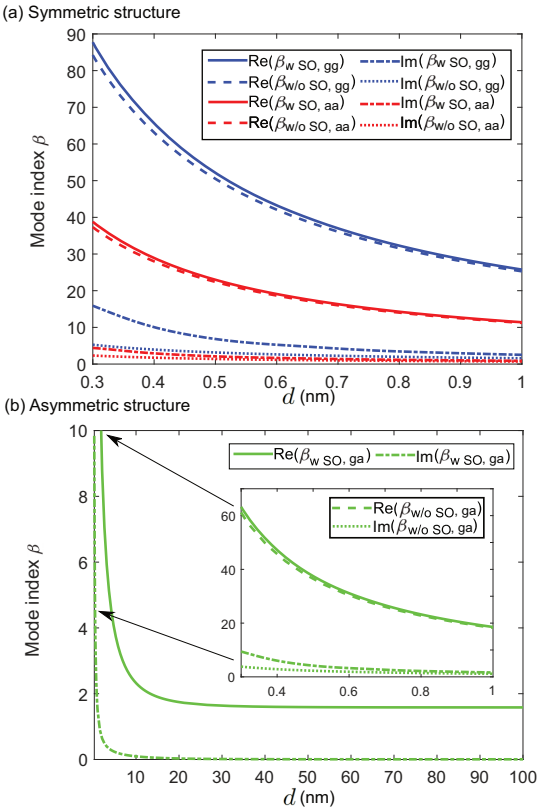

FIG. 2. Real (solid and dashed lines) and imaginary (dotted and dashed-dotted lines) parts of the short-range mode index vs slab width d at a wavelength of 775 nm. Results are shown for spill-out (SO) included (solid and dashed-dotted lines) and neglected (dashed and dotted lines). In (a) the structure is symmetric, and the blue and red lines are for slabs surrounded by glass (gg) and air (aa), respectively. In (b) the structure is asymmetric, where the slab is surrounded by glass and air (ga), and the inset shows a zoom for d below 1 nm. In both (a) and (b), the response is isotropic.

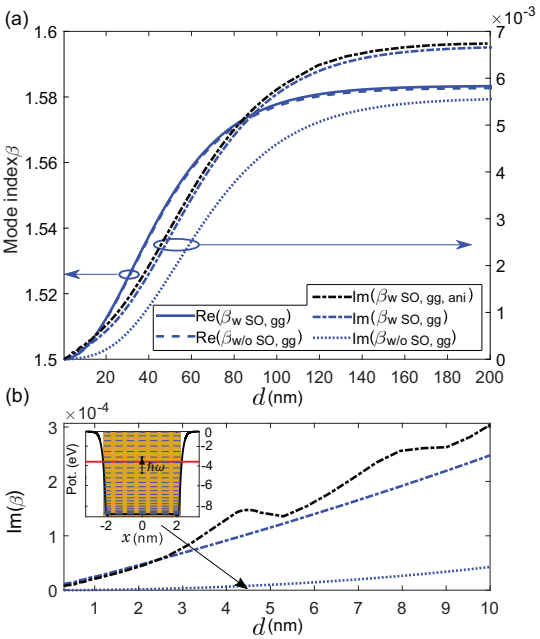

FIG. 3. Long-range mode index for a slab surrounded by glass (gg) at a wavelength of 775 nm. In (a), the real and imaginary parts are shown on the left and right y axes, respectively, as indicated by the arrows. The black dashed-dotted line shows the imaginary part of the mode index when the response is anisotropic. (b) Shows a zoom of the imaginary part in (a) for d below 10 nm, where the line types have the same meaning as in (a). Inset shows the potential barrier for a slab of width 4.5 nm, as indicated by the arrow, where the Fermi energy is shown by the red horizontal line, and all the discrete quantum energy levels as the dashed blue horizontal lines. A transition between an occupied and an empty state with energy difference $\hbar\omega$ is illustrated by the vertical arrow, and is near resonance at the wavelength of 775 nm.

index with and without spill-out when the slab width exceeds 1 nm. Therefore, the mode index when neglecting spill-out is only included in the inset showing results for d below 1 nm, where β has the same behavior as for the symmetric structure in Fig. 2(a). Although it is hard to see in the figure, the imaginary part is small but nonzero for all slab widths. The mode index has converged when $d = 100$ nm, and for such a wide slab the plasmon behaves as if bound to a single interface between glass and gold [50].

The asymmetric structure also supports long-range modes, but only for slab thicknesses above a certain threshold [8]. The long-range mode is mainly bound to the air-gold interface, with a mode index that is lower than the refractive index of the glass substrate. This implies that the normal component of the wave vector becomes real (with a very small imaginary part due to loss in the gold) on the glass side of the structure, leading to a wave propagating in the substrate, thus not a truly bound mode [10]. Hence, the wave will leak out into the substrate, where conservation of momentum determines the leakage angle [52]. However, if the dielectric constants of the substrate and superstrate are not too different, it is possible to obtain a long-range mode that is truly bound to both interfaces (see, e.g., Fig. 3 in Ref. [9]). The phenomenon of leaky modes can be examined using leakage radiation microscopy (see, e.g., Refs. [10,53,54]).

Figure 3(a) shows the mode index of the corresponding long-range mode for a gold slab surrounded by glass, where the response is isotropic for the blue lines and anisotropic for the black line. Again, the results are in agreement with previous studies [7–9] when spill-out is neglected. In this case, the mode index for an ultrathin slab is very close to the refractive index of the substrate, which implies that the mode is weakly bound. The mode is therefore long range with most of its field profile located in the dielectric regions, which will be illustrated in the next section. As the slab width

increases to 200 nm, the mode index without spill-out has converged to $\sqrt{\varepsilon_{\text{gold}}\varepsilon_{\text{glass}}/(\varepsilon_{\text{gold}}+\varepsilon_{\text{glass}})}$, which is the mode index of a plasmon bound to a single interface between gold and glass [50]. In addition, the corresponding short-range mode index when neglecting spill-out in Fig. 2(a) converges to the same value for $d = 200$ nm (not shown) in agreement with Refs. [8,9,12]. With spill-out included, the long-range mode index for small slabs in Fig. 3(a) is also close to the refractive index of the substrate, and the imaginary part is very low. As for the short-range mode in Figs. 2(a) and 2(b), especially the real part of the mode index is almost the same with and without spill-out, as seen by comparing the solid and dashed lines. But, importantly, spill-out significantly increases the imaginary part of the mode index, even for slab widths up to 200 nm, as seen by comparing the dotted and dashed-dotted lines in Fig. 3(a). When including anisotropy in the model, it is found that the real part of the mode index is almost unchanged (not shown), while the imaginary part is modified. For wide slabs the effect of anisotropy is small, in agreement with Ref. [49], as can be seen by comparing the blue and black dashed-dotted lines. The same small effect of anisotropy for wide slabs is found for the short-range mode in Fig. 2(a) (not shown). In the other limit when the slab width is only a few nm, the imaginary part of the long-range mode index is modified due to anisotropy, as seen in Fig. 3(b) showing a zoom of the imaginary parts in Fig. 3(a) for d below 10 nm. As these values are on the order of 10^{-4}, this mode is relatively much more sensitive to anisotropy than the corresponding short-range mode in Fig. 2, even though the absolute change due to anisotropy is comparable for the two modes. The anisotropic part contains peaks corresponding to certain resonances in electronic transitions (see Appendix B for details). For a wavelength of 775 nm, a resonance is found at a slab width of 4.5 nm, for which the potential barrier is shown in the inset, where the zero in potential is chosen to be sufficiently far from the slab. A transition between an occupied and an empty state with energy difference $\hbar\omega$ is illustrated by the vertical arrow. For wider slabs, the individual energy levels are closer, and become continuous in the bulk limit, implying that the response tends to become isotropic in this limit.

Having now briefly accounted for the relative small impact of anisotropy, we turn to further illustrate the effect of spill-out in a symmetric structure. Hence, in the remaining part of the paper, the shown results are all obtained using an isotropic response. The ratio between mode indices with and without spill-out is shown in Figs. 4(a) and 4(b) for the short- and long-range modes, respectively. For both kinds of modes, the real part of the mode index is almost unaffected by spill-out, as the ratios shown by the solid lines in Fig. 4 have converged to 1.0004 when $d = 50$ nm. For the short-range mode the ratio between the imaginary parts is approximately 3.0 for a slab width of 0.3 nm, while it converges to ∼1.2 for $d = 200$ nm. The real part of the normalized magnetic field profile when spill-out is included is shown in the inset of Fig. 4(a) for a slab width of 200 nm, as indicated by the arrow. It behaves as two decoupled plasmons bound to the interfaces between glass and gold, as the field profiles bound to the individual interfaces do not interact for such a wide slab.

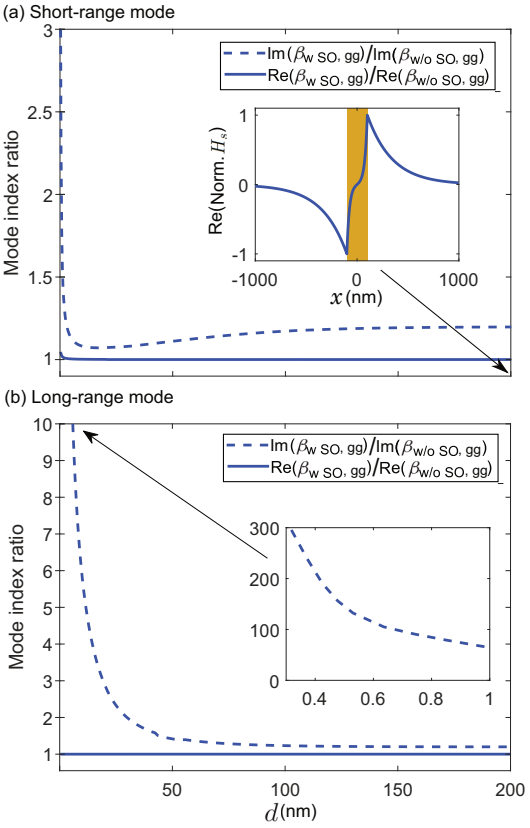

FIG. 4. Ratio between mode indices with and without spill-out at a wavelength of 775 nm for a slab surrounded by glass (gg). In (a), the mode is short range, and the inset shows the real part of the transverse magnetic field for $d = 200$ nm, as indicated by the arrow, where the colored area represents the ion charge. In (b), the mode is long range and the inset shows a zoom for d below 1 nm. The response is isotropic.

For the long-range mode, the corresponding ratio between the imaginary parts is extremely high for small d as seen in the inset in Fig. 4(b). However, as d increases, the ratio decreases monotonically and converges to ∼1.2 when $d = 200$ nm. This is an important result showing that quantum spill-out increases the imaginary part of the mode index by 20%, even for relatively thick slabs that can readily be fabricated [12,53] and approach bulk gold. It is highly surprising that spill-out plays such a significant role for wide slabs, as it only modifies the electron density in a region very close to the ion charge. In addition, it is noticed that the ratios between the imaginary parts of the two modes converge to the same value when the slab is wide enough, as in this case the field profiles bound to the individual interfaces are decoupled, similarly to classical models [8,9,12]. Furthermore, the short-range mode indices with and without spill-out in the asymmetric structure in Fig. 2(b) converge to the same values as for

the long-range mode in Fig. 3(a), as both modes behave as bound to a single interface between gold and glass. Hence, spill-out also increases the imaginary part of the mode index by 20% in an asymmetric structure. The physical meaning of the enhanced imaginary part is that the plasmon propagation length will be significantly shorter, while the unaltered real part implies that the propagation velocity will be unchanged.

Reference [20] applied an isotropic dielectric function with components analogous to Eq. (2) to study the effect of spill-out on plasmons propagating in a magnesium slab ($r_s = 2.66$ bohrs) surrounded by silicon and air. With the present method, the real part of the calculated mode index agrees well with values estimated from Fig. 5 in Ref. [20], showing quantitative agreement between that paper and the method presented here. Likewise, the mode index calculated in this paper agrees well with values estimated from Figs. 6 and S9 in Ref. [35] regarding plasmons propagating in gold slabs surrounded by air.

IV. FIELD PROFILE AND PLASMONIC ABSORPTION

Once the mode indices have been computed, the magnetic field from Eq. (4) is calculated using the same transfer matrix method as described in Refs. [1,55]. In the last part of Appendix B it is shown how the method is slightly modified to describe the anisotropy. Applying the same phase convention as in Ref. [7], the normalized real part of the short-range transverse magnetic field $H_s(x)$ across a gold slab of 0.3 nm is shown in Fig. 5(a) at a wavelength of 775 nm. The associated imaginary parts of the fields are not shown as they are small compared to the real parts, similarly to classical models [7].

When neglecting spill-out, the slope of the magnetic field, corresponding to the normal component of the electric field, becomes discontinuous across the slab surfaces in agreement with Refs. [7–9]. With spill-out included, the slope is still discontinuous due to the abrupt change in the bound electron term in Eq. (2), although it is difficilt to see in Fig. 5(a). But, in the vicinity of the slab surface, the field profiles behave more smoothly, and their maximum positions are slightly shifted into the dielectric region. We have checked that the appropriate boundary conditions regarding electromagnetic fields across an interface [50] are satisfied. Further away from the slab, the field profiles with and without spill-out become almost identical. Consequently, the decay lengths into the dielectrics, calculated as $1/\mathrm{Im}(k_x)$, where $k_x = k_0\sqrt{\varepsilon_s - \beta^2}$ is the wave number in the x direction, are very similar and both are on the order of a few nm. This illustrates that the short-range mode is strongly bound to the slab, as it decays very rapidly into the dielectrics [9], and thereby has a large part of its field profile located in the slab region. Notice that as the real part of the mode index is much higher than its imaginary part, the decay length mostly depends on the real part of the mode index. Figure 5(a) demonstrates that the short-range magnetic field is antisymmetric for the two symmetric structures shown by the blue and red curves, while this is no longer the case for an asymmetric structure, as shown by the green curves. When the slab width increases, the field profile broadens, as shown in the inset of Fig. 4(a) for a 200-nm-wide slab. For such a wide slab, the field profiles with

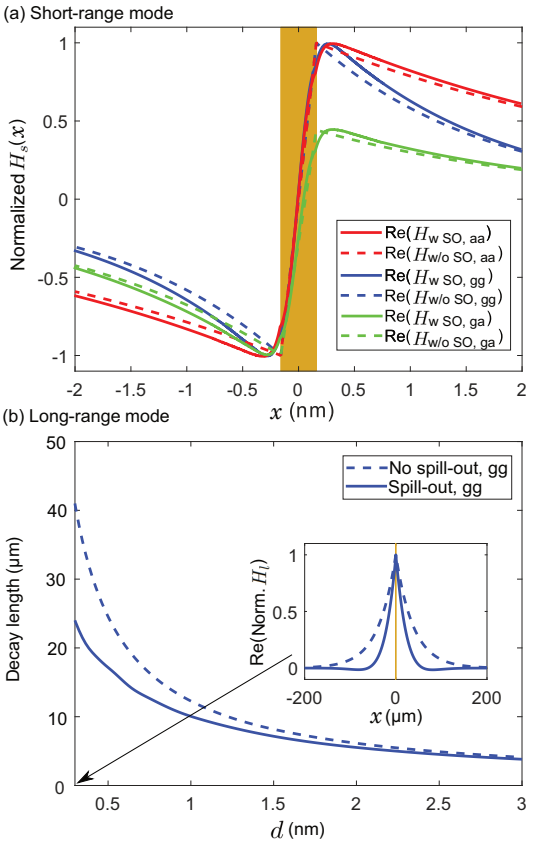

FIG. 5. (a) Normalized real part of $H_s(x)$ across a slab of width 0.3 nm. (b) Shows the decay length of the long-range mode for a gold slab surrounded by glass (gg) with and without spill-out. The inset shows the real part of the magnetic field profiles for a slab of width 0.3 nm as indicated by the arrow. In both (a) and (b), the colored areas represent the ion charge, the wavelength is 775 nm, and the response is isotropic.

and without spill-out are almost identical, and both behave as two decoupled plasmons bound to the interfaces between glass and gold. When including anisotropy, the short-range mode indices were almost unchanged. This also applies for the magnetic fields, why Fig. 5(a) only shows the fields for an isotropic response.

As mentioned above, the long-range mode is weakly bound. Consequently, the electromagnetic fields for a few-nm slab have decay lengths of several micrometers, as shown for a gold slab surrounded by glass in Fig. 5(b). As the decay length mostly depends on the real part of the mode index, the decay length in Fig. 5(b) is in practice unchanged when the response is anisotropic. The long decay length implies that most of the field profiles are located in the dielectric regions. The field profiles are broader when spill-out is neglected, as also seen in the inset showing the real part of the magnetic fields across a slab of width 0.3 nm, i.e., the same slab as in

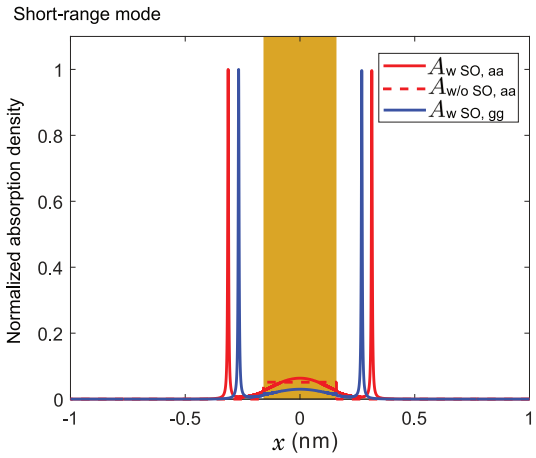

FIG. 6. Normalized absorption density for the short-range mode across a gold slab of width 0.3 nm at a wavelength of 775 nm. The solid and dashed lines show the absorption density when spill-out is included and neglected, respectively, where the red and blue lines are for slabs surrounded by air (aa) and glass (gg), respectively. The response is isotropic.

Fig. 5(a). Including spill-out effectively implies a broader slab [see Fig. 1(a)], which means that the fields become slightly more localized with a shorter decay length. For slabs wider than 3 nm, the decay lengths with and without spill-out are very similar, and both converge to the decay length of a plasmon bound to a single interface between gold and glass (not shown).

As argued above, spill-out plays almost no role for the decay length for slabs of a few nm. On the other hand, it significantly increases the imaginary part of the mode index as shown in Fig. 4. This leads us to investigate how spill-out affects the electric field and plasmonic absorption across the slab. First, the electric field is calculated from the magnetic field in Eq. (4) as [50]

$$\vec{E}_m(x,y) = \frac{i\overleftrightarrow{\varepsilon}^{-1}(x,y)}{\omega\varepsilon_0}\vec{\nabla}\times[\hat{z}H_m(x,y)]. \tag{5}$$

The electric field is subsequently used to calculate the absorption density defined as

$$A_m(x,y) = \mathrm{Im}(\vec{E}_m^*(x,y)\cdot\overleftrightarrow{\varepsilon}(x,y)\cdot\vec{E}_m(x,y)). \tag{6}$$

By considering the time average of the Poynting vector, $\langle\vec{S}\rangle = 1/2\,\mathrm{Re}(\vec{E}\times\vec{H}^*)$ [50], it can be shown that conservation of energy implies that the plasmonic absorption and the imaginary part of the mode index are related in the following way:

$$\mathrm{Im}(\beta) = \frac{c\varepsilon_0\int A_m(x,y)dx}{2\int \mathrm{Re}(\vec{E}_m(x,y)\times\vec{H}_m^*(x,y))\cdot\hat{y}\,dx}. \tag{7}$$

We have checked that this relation is satisfied for both kinds of modes with and without spill-out. The normalized absorption density is shown in Fig. 6 for the short-range mode across the slab of width 0.3 nm at a wavelength of 775 nm. If spill-out is neglected, absorption can only take place in the gold as the surrounding dielectrics are assumed lossless. In this case, the absorption density is almost unaffected by the kind of surrounding dielectric, why Fig. 6 only shows it for a slab surrounded by air. But, with spill-out included, strong plasmonic absorption occurs, and the absorption density mostly consists of two narrow peaks located in the dielectric regions close to the interfaces. At these positions, similarly to Refs. [1,56], the real part of the dielectric function is zero (at the wavelength 775 nm), while its imaginary part is small but nonzero, which ensures that the peaks in the absorption density are finite. The narrow peaks are found a few Å outside the ion charge, and the same is found for the long-range mode (not shown). The contribution from these peaks leads to enhanced plasmonic absorption, as they are a consequence of electron spill-out, and therefore not found in classical models. The peaks correspond to the increase in imaginary part of the mode index observed in Sec. III. For a slab surrounded by glass, the peaks occur slightly closer to the slab, as the real part of the dielectric function has its zero shifted slightly compared to the case with a slab surrounded by air [see Fig. 1(b)]. The same phenomenon was found in Ref. [20] for a magnesium slab surrounded by silicon and air. The peaks in absorption density due to spill-out were recently discussed in Ref. [1], where they were found to significantly reduce the reflectance from an ultrasharp groove array in much better agreement with measurements [4] compared to classical models [5,6]. Like for the magnetic field, the absorption density is almost unaffected by anisotropy, why Fig. 6 only shows results for an isotropic response.

The increased absorption loss due to spill-out will manifest itself as decreased propagation lengths in fabricated plasmonic structures. Losses in such structures have been studied in, e.g., Refs. [57–60], where it was found that the measured propagation length of plasmons propagating in a 70-nm silver film deposited on glass is significantly shorter than the one calculated using classical models [59,60]. In addition, attenuated total reflectance (ATR) has been studied for a silver film on glass in the Kretschmann configuration. It is found that the measured reflectance at the Kretschmann angle [50] is significantly lower than the one calculated using classical models [61,62], illustrating that the losses are higher than expected from a classical point of view. Hence, these works together with Ref. [1] also support the finding that a classical model is not sufficient to correctly describe losses occurring in plasmonic waveguides.

V. CONCLUSION

In conclusion, we have applied a quantum mechanical approach and local response theory to study the propagation of plasmons in nanometer-thin gold slabs surrounded by different dielectrics. The effect of spill-out is found to be small on the real part of the mode indices but remarkably increases the corresponding imaginary part, and even for slabs wide enough to approach bulk the increase is 20%. This is explained in terms of enhanced plasmonic absorption mainly taking place in narrow regions located a few Å outside the ion charge. It is highly surprising that spill-out plays such a significant role for wide slabs, as it only modifies the electron density in a region

very close to the ion charge. For slab widths above a few nanometers, the decay length of the fields into the dielectrics is almost unaffected by spill-out, as it mostly depends on the real part of the mode index. By taking anisotropy into account, a comparable change is found in the imaginary part of the mode index for both kinds of modes, while the relative change, however, is much larger for the long-range mode. When the slab width increases, the effect of anisotropy becomes less pronounced, and for a 200-nm-wide slab the response is almost isotropic. Furthermore, in contrast to classical models, the short-range mode index does not diverge in the limit of vanishing slab thickness when spill-out is included. Instead, plasmonic modes cease to exist for slab widths below a cutoff thickness in the subatom region.

ACKNOWLEDGMENT

This work was supported by the QUSCOPE center sponsored by the Villum Foundation.

APPENDIX A: CALCULATION OF ELECTRON DENSITY

In this Appendix, we discuss in more detail how the electron density is calculated. Within the jellium model it is assumed that the charge of the gold ions is smeared out, such that their charge density is constant within the slab [23,24]. The characteristic spill-out, as seen in Fig. 1(a), stems from the distribution of free electrons in the vicinity of this positive background. The Kohn-Sham equations [23] are solved self-consistently within the local density approximation (LDA) [44], applying the Perdew-Zunger parametrization [63] for the correlation term. The applied Wigner-Seitz radius for gold is $r_s = 3.01$ bohrs [24].

It is found that 2500 basis functions on the form $\sin(m\pi(x/L + 1/2))$ are sufficient to describe the density for slab widths up to 200 nm. The length L is 1 nm larger than the slab width d, and the slab is centered at $x = 0$. As in Ref. [1], the density is said to converge when a variation in Fermi energy between two iterations below 10^{-7} Ha is achieved. Furthermore, in the Anderson mixing scheme [64], the mixing parameter α must be below a certain threshold which strongly decreases with d. It is found that $\alpha \leqslant 5 \times 10^{-4}$ is necessary for slab thicknesses up to 20 nm. The potentials for wider slabs can afterward be constructed from the potential of the 20-nm slab, as the oscillations in potential near its center are negligible, meaning that the effective potential near the center can be seen as constant. This constant potential is added in the central region of wider slabs.

APPENDIX B: ANISOTROPY

In this Appendix, we discuss how the anisotropic term $\varepsilon_{\text{ani}}(\omega, d)$ in Eq. (3) is calculated, and how the transfer matrix method is slightly modified to describe an anisotropic response. As mentioned in Sec. II, we approximate the nonlocal part of the dielectric tensor by a piecewise constant function taking the value $\iint \overleftrightarrow{\varepsilon}_{NL}(x, x')dx\,dx'/d$ inside the slab and zero outside. From Ref. [43], the transition matrix elements used in Eq. (1) are given by

$$\vec{j}_{nm}(\vec{k}_\parallel, x) = -\frac{e\hbar}{2m_e i}(2\vec{i}k_\parallel A_{nm}(x) + \hat{x}B_{nm}(x)), \tag{B1}$$

where

$$A_{nm}(x) = \varphi_n(x)\varphi_m^*(x),$$
$$B_{nm}(x) = \varphi_m^*(x)\frac{\partial\varphi_n(x)}{\partial x} - \varphi_n(x)\frac{\partial\varphi_m^*(x)}{\partial x}. \tag{B2}$$

Here, the φ_n's are the wave functions corresponding to the energies $E_n = \hbar\omega_n$ in Eq. (1). From Eqs. (B1) and (B2), the parallel part of the step function is found to be zero due to the orthogonality of the wave functions. Using the commutator relation $\langle\varphi_m|\hat{p}_x|\varphi_n\rangle = im_e E_{mn}\langle\varphi_m|x|\varphi_n\rangle/\hbar$, where $E_{mn} = E_m - E_n$, and the oscillator strength $g_{mn} = 2m_e|\langle\varphi_m|x|\varphi_n\rangle|^2 E_{mn}/\hbar^2$, the perpendicular part describing the anisotropy can be shown to be

$$\varepsilon_{\text{ani}}(\omega, d) = \frac{e^2}{\pi d\varepsilon_0\hbar^2(\omega + i\Gamma)^2} \times \sum_{m,n} g_{mn}\frac{E_{mn}^2(E_F - E_n)\theta(E_F - E_n)}{E_{mn}^2 - \hbar^2(\omega + i\Gamma)^2}. \tag{B3}$$

Here, the energies and number of bands depend strongly on d. A resonance is found when E_{mn} for a certain transition is close to the photon energy $\hbar\omega$, as illustrated by the vertical arrow in the inset of Fig. 3(b). When the slab width increases to 200 nm, the anisotropic term tends to zero, implying that the response becomes isotropic in this limit.

When describing anisotropy, the reflection and transmission coefficients in the transfer matrix method are slightly modified into [65]

$$r_{ij} = \frac{k_{x,i}\varepsilon_{\parallel,j} - k_{x,j}\varepsilon_{\parallel,i}}{k_{x,i}\varepsilon_{\parallel,j} + k_{x,j}\varepsilon_{\parallel,i}}, \quad t_{ij} = 1 + r_{ij}, \tag{B4}$$

where $k_{x,i} = k_0\sqrt{\varepsilon_{\parallel,i} - \beta^2\varepsilon_{\parallel,i}/\varepsilon_{\perp,i}}$. It is seen that if the material is isotropic, $\varepsilon_{\parallel,i} = \varepsilon_{\perp,i}$, the original coefficients from Ref. [55] are restored.

APPENDIX C: CLASSIFICATION OF PLASMONIC MODES

In this short Appendix, we discuss how the plasmonic modes are classified. The mode index is calculated by the same type of transfer matrix method as presented in detail in Ref. [1]. However, only one kind of plasmonic mode was studied in that paper, while the classification of modes was not presented there. A structure matrix $\mathcal{S}$ is constructed, which relates the magnetic fields to the left and right of the structure, and a mode index is found when the matrix element $\mathcal{S}_{11}$ is zero, as this condition yields a pole in the reflection coefficient. The mode is classified by the sign of $\mathcal{S}_{21}$, where positive and negative signs correspond to long- and short-range modes, respectively, and $\mathcal{S}_{21}$ is exactly ± 1 for symmetric structures. Expressions for the matrix elements can be found in Refs. [1,55]. In addition, the x axis is divided into tiny segments, each modeled as having a constant dielectric function. Similarly to Ref. [1], we find that segments of 2.7×10^{-4} nm are sufficient to avoid discretization errors.

[1] E. J. H. Skjølstrup, T. Søndergaard, and T. G. Pedersen, Phys. Rev. B **97**, 115429 (2018).
[2] F. J. García-Vidal, L. Martin-Moreno, T. Ebbesen, and L. Kuipers, Rev. Mod. Phys. **82**, 729 (2010).
[3] C. L. C. Smith, N. Stenger, A. Kristensen, N. A. Mortensen, and S. I. Bozhevolnyi, Nanoscale **7**, 9355 (2015).
[4] T. Søndergaard, S. Novikov, T. Holmgaard, R. Eriksen, J. Beermann, Z. Han, K. Pedersen, and S. I. Bozhevolnyi, Nat. Commun. **3**, 969 (2012).
[5] E. J. H Skjølstrup and T. Søndergaard, J. Opt. Soc. Am B **34**, 673 (2017).
[6] M. Odgaard, M. G. Laursen, and T. Søndergaard, J. Opt. Soc. Am. B **31**, 1853 (2014).
[7] T. Søndergaard and S. I. Bozhevolnyi, Phys. Status Solidi B **245**, 9 (2008).
[8] P. Berini, Adv. Opt. Photonics **1**, 484 (2009).
[9] D. Sarid, Phys. Rev. Lett. **47**, 1927 (1981).
[10] A. Drezet, A. Hohenau, D. Koller, A. Stepanov, H. Ditlbacher, B. Steinberger, F. R. Aussenegg, A. Leitner, and J. R. Krenn, Mater. Sci. Eng. B **149**, 220 (2008).
[11] A. Davoyan, I. Shadrivov, S. Bozhevolnyi, and Y. Kivshar, J. Nanophotonics **4**, 043509 (2010).
[12] F. Liu, Y. Rao, Y. Huang, W. Zhang, and J. Peng, Appl. Phys. Lett. **90**, 141101 (2007).
[13] A. Yanai and U. Levy, Opt. Express **17**, 14270 (2009).
[14] H. A. Atwater and A. Polman, Nat. Mater. **9**, 205 (2010).
[15] D. K. Gramotnev and S. I. Bozhevolnyi, Nat. Mater. **9**, 193 (2010).
[16] M. Kuttge, F. de Abajo, and A. Polman, Opt. Express **17**, 10385 (2009).
[17] P. Berini and I. D. Leon, Nat. Photonics **6**, 16 (2012).
[18] N. A. Mortensen, S. Raza, M. Wubs, T. Søndergaard, and S. I. Bozhevolnyi, Nat. Commun. **5**, 3809 (2014).
[19] H. Qian, Y. Xiao, and Z. Liu, Nat. Commun. **7**, 13153 (2016).
[20] X. Li, A. Teng, M. Özer, J. Shen, H. Weitering, and Z. Zhang, New J. Phys. **16**, 065014 (2014).
[21] W. Yan, N. A. Mortensen, and M. Wubs, Phys. Rev. B **88**, 155414 (2013).
[22] R. Maniyara, D. Rodrigo, R. Yu, J. Ferrer, D. Ghosh, R. Yongsunthon, D. Baker, A. Rezikyan, F. de Abajo, and V. Pruneri, Nature Photonics **13**, 328 (2019).
[23] N. D. Lang and W. Kohn, Phys. Rev. B **1**, 4555 (1970).
[24] N. D. Lang and W. Kohn, Phys. Rev. B **3**, 1215 (1971).
[25] W. Yan, M. Wubs, and N. A. Mortensen, Phys. Rev. Lett. **115**, 137403 (2015).
[26] T. V. Teperik, P. Nordlander, J. Aizpurua, and A. G. Borisov, Phys. Rev. Lett. **110**, 263901 (2013).
[27] L. Stella, P. Zhang, F. J. García-Vidal, A. Rubio, and P. García-Gonzalez, J. Phys. Chem. C **117**, 8941 (2013).
[28] D. C. Marinica, A. K. Kazansky, P. Nordlander, J. Aizpurua, and A. G. Borisov, Nano Lett. **12**, 1333 (2012).
[29] O. M. Bakr, V. Amendola, C. M. Aikens, W. Wenseleers, R. Li, L. D. Negro, G. C. Schatz, and F. Stellacci, Angew. Chem. **48**, 5921 (2009).
[30] M. Zhu, C. M. Aikens, F. J. Hollander, G. C. Schatz, and R. Jin, J. Am. Chem. Soc. **130**, 5883 (2008).
[31] R. Zhang, L. Bursi, J. Cox, Y. Cui, C. Krauter, A. Alabastri, A. Manjavacas, A. Calzolari, S. Corni, E. Molinari, E. Carter, F. de Abajo, H. Zhang, and P. Nordlander, ACS Nano **11**, 7321 (2017).
[32] G. Maroulis, J. Chem. Phys. **121**, 10519 (2004).
[33] J. Zuloaga, E. Prodan, and P. Nordlander, Nano Lett. **9**, 887 (2009).
[34] H. Zhang, V. Kulkarni, E. Prodan, P. Nordlander, and A. O. Govorov, J. Phys. Chem. C **118**, 16035 (2014).
[35] C. David and F. de Abajo, ACS Nano **8**, 9558 (2014).
[36] W. Yan, Phys. Rev. B **91**, 115416 (2015).
[37] S. L. Adler, Phys. Rev. **126**, 413 (1962).
[38] N. Wiser, Phys. Rev. **129**, 62 (1963).
[39] J. K. Lindhard, Dan. Vidensk. Selsk. Mat. Fys. Medd. **28**, 8 (1954).
[40] N. D. Mermin, Phys. Rev. B **1**, 2362 (1970).
[41] P. J. Feibelman, Phys. Rev. B **12**, 1319 (1975).
[42] P. J. Feibelman, Prog. Surf. Sci. **12**, 287 (1982).
[43] O. Keller, Phys. Rev. B **33**, 990 (1986).
[44] J. Kohanoff, *Electronic Structure Calculations for Solids and Molecules: Theory and Computational Methods*, 1st ed. (Cambridge University Press, Cambridge, 2006).
[45] W. Zhu, R. Esteban, A. G. Borisov, J. J. Baumberg, P. Nordlander, H. J. Lezec1, J. Aizpurua, and K. B. Crozier, Nat. Commun. **7**, 11495 (2016).
[46] F. de Abajo, J. Phys. Chem. C **112**, 17983 (2008).
[47] S. Raza, G. Toscano, A. Jauho, M. Wubs, and N. A. Mortensen, Phys. Rev. B **84**, 121412(R) (2011).
[48] K. J. Savage, M. M. Hawkeye, R. Esteban, A. G. Borisov, J. Aizpurua, and J. J. Baumberg, Nature (London) **491**, 574 (2012).
[49] S. Laref, J. Cao, A. Asaduzzaman, K. Runge, P. Deymier, R. Ziolkowski, M. Miyawaki, and K. Muralidharan, Opt. Express **21**, 11827 (2013).
[50] L. Novotny and B. Hecht, *Principles of Nano-Optics*, 2nd ed. (Cambridge University Press, Cambridge, 2012).
[51] P. B. Johnson and R. W. Christy, Phys. Rev. B **6**, 4370 (1972).
[52] T. Søndergaard, *Green's Function Integral Equation Methods in Nano-Optics*, 1st ed. (CRC Press, Boca Raton, FL, 2019).
[53] P. Simesen, T. Søndergaard, E. Skovsen, J. Fiutowski, H. Rubahn, S. I. Bozhevolnyi, and K. Pedersen, Opt. Express **23**, 16356 (2015).
[54] S. Ramos, H. Farooq, H. Alghasam, A. Bernussi, and L. Peralta, Opt. Express **26**, 19718 (2018).
[55] M. V. Klein and T. E. Furtak, *Optics*, 2nd ed. (Wiley, New York, 1986).
[56] Z. F. Öztürk *et al.*, J. Nanophotonics **5**, 051602 (2011).
[57] S. Boriskina, T. Cooper, L. Zeng, G. Ni, J. Tong, Y. Tsurimaki, Y. Huang, L. Meroueh, G. Mahan, and G. Chen, Adv. Opt. Photonics **9**, 775 (2017).
[58] J. Khurgin, Nat. Nanotechnol. **10**, 2 (2015).
[59] R. Zia, M. D. Selker, and M. L. Brongersma, Phys. Rev. B **71**, 165431 (2005).
[60] B. Lambrecht, J. Krenn, G. Schider, H. Ditlbacher, M. Salerno, N. Felidj, A. Leitner, F. Ausseness, and J. Weeber, Appl. Phys. Lett. **79**, 51 (2001).
[61] J. W. Petefish and A. C. Hillier, Anal. Chem. **86**, 2610 (2014).
[62] W. Knoll, Annu. Rev. Phys. Chem. **49**, 569 (1998).
[63] J. P. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1981).
[64] V. Eyert, J. Comput. Phys. **124**, 271 (1996).
[65] R. Petersen, T. G. Pedersen, M. N. Gjerding, and K. S. Thygesen, Phys. Rev. B **94**, 035128 (2016).

Correction: The article identification number was incorrect in Ref. [1] and has been fixed, enabling access to the online article.

Paper V

Theory of EELS in atomically thin metallic films

A. R. Echarri, Enok J. H. Skjølstrup, Thomas G. Pedersen, and
F. Javier Garcia de Abajo

The paper has been submitted to Physical Review Research and is available
on
ArXiv:1912.09414

The layout has been revised.

Theory of EELS in atomically thin metallic films

A. Rodríguez Echarri,[1] Enok Johannes Haahr Skjølstrup,[2] Thomas G. Pedersen,[2] and F. Javier García de Abajo[1,3,*]

[1] *ICFO-Institut de Ciencies Fotoniques, The Barcelona Institute of Science and Technology, 08860 Castelldefels (Barcelona), Spain*

[2] *Department of Materials and Production, Aalborg University, Skjernvej 4A, DK-9220 Aalborg East, Denmark*

[3] *ICREA-Institució Catalana de Recerca i Estudis Avançats, Passeig Lluís Companys 23, 08010 Barcelona, Spain*

(Dated: December 20, 2019)

We study strongly confined plasmons in ultrathin gold and silver films by simulating electron energy-loss spectroscopy (EELS). Plasmon dispersion relations are directly retrieved from the energy- and momentum-resolved loss probability under normal incidence conditions, whereas they can also be inferred for aloof parallel beam trajectories from the evolution of the plasmon features in the resulting loss spectra as we vary the impinging electron energy. We find good agreement between nonlocal quantum-mechanical simulations based on the random-phase approximation and a local classical dielectric description for silver films of different thicknesses down to a few atomic layers. We further observe only a minor dependence of quantum simulations for these films on the confining out-of-plane electron potential when comparing density-functional theory within the jellium model with a phenomenological experimentally-fitted potential incorporating atomic layer periodicity and in-plane parabolic bands of energy-dependent effective mass. The latter shows also a small dependence on the crystallographic orientation of silver films, while the unphysical assumption of energy-independent electron mass leads to spurious features in the predicted spectra. Interestingly, we find electron band effects to be more relevant in gold films, giving rise to blue shifts when compared to classical or jellium model simulations. In contrast to the strong nonlocal effects found in few-nanometer metal nanoparticles, our study reveals that a local classical description provides excellent quantitative results in both plasmon strength and dispersion when compared to quantum-mechanical simulations down to silver films consisting of only a few atomic layers, thus emphasizing the in-plane nearly-free conduction-electron motion associated with plasmons in these structures.

Physics Subject Headings: EELS; Surface plasmons; Thin films; Quantum-well states; Nanophotonics; Nonlocal effects.

I. INTRODUCTION

Surface plasmons –the collective electron oscillations at material surfaces and interfaces– provide the means to concentrate and amplify the intensity of externally applied light down to nanoscale regions [1, 2], where they interact strongly with molecules and nanostructures, thus becoming a powerful asset in novel applications [3] such as biosensing [2, 4, 5], photocatalysis [6, 7], energy harvesting [8, 9], and nonlinear optics [10–13].

Surface plasmons were first identified using electron energy-loss spectroscopy (EELS), starting with the prediction [14] and subsequent measurement of associated loss features in electrons scattered under grazing incidence from Al [15], Na and K [16, 17], and Ag [17, 18] surfaces. The main characteristics of surface plasmons in noble and simple metals were successfully explained using time-dependent density-functional theory (TD-DFT) [19] within the jellium model [20, 21], while inclusion of electron band effects were required for other metals [22]. Interestingly, multipole surface plasmons were predicted as additional resonances originating in the smooth electron density profile across metal-dielectric interfaces [22–24], and subsequently found in experiments performed on simple metals such as K and Na [25], but concluded to be too weak to be observed in Al [25] and Ag [26]. These studies focused on the relatively high-energy plasmons supported by planar surfaces in the short-wavelength regime. However, plasmons can hybridize with light forming surface-plasmon polaritons (SPPs) in planar surfaces, which become light-like modes at low energies, thus loosing confinement, as they are characterized by in-plane wavelengths slightly smaller than those of light and long-range penetration into the dielectric material or empty space outside the metal [27–29].

Highly confined plasmons can also be achieved in sharp metallic tips and closely spaced metal surfaces [30], where strong redshifts are produced due to the attractive Coulomb interaction between neighboring non-coplanar interfaces. This effect, which depends dramatically on surface morphology, can also be observed in planar systems such as ultrathin noble metal films [31, 32] and narrow metal-dielectric-metal waveguides [12, 31, 33]. More precisely, hybridization takes place in metal films between the plasmons supported by their two interfaces, giving rise to bonding and antibonding dispersion branches that were first revealed also through EELS is self-standing aluminum foils [34]; in ultrathin films of only a few atomic layers in thickness, the antibonding plasmon dispersion is pushed close to the light line, whereas the bonding plasmon becomes strongly confined

*Corresponding author:javier.garciadeabajo@nanophotonics.es

(reaching the quasistatic limit [31, 32]), as experimentally corroborated through angle-resolved low-energy EELS in few-monolayer Ag films [35] and monolayer (ML) $DySi_2$ [36], as well as in laterally confined wires formed by In [37] and silicide [38], and even in monoatomic Au chains grown on Si(557) surfaces [39]. Additionally, graphene has been shown to support long-lived mid-infrared and terahertz plasmons [40] that can be tuned electrically [41, 42] and confined vertically down to few nanometers when placed in close proximity to a planar metal surface [43, 44]. While most graphene plasmon studies have been performed using far- and near-field optics setups [45–47], low-energy EELS has also revealed their dispersion relation in extended films [48, 49]. Here, we focus instead on visible and near-infrared plasmons supported by atomically thin metal films, which have been recently demonstrated in crystalline Ag layers [32], where they also experience strong spatial confinement.

In this paper, we investigate plasmons in atomically thin noble metal films by theoretically studying EELS for electron beams either traversing them or moving parallel outside their surface. We provide quantum-mechanical simulations based on the random-phase approximation (RPA), which are found to be in excellent agreement with classical dielectric theory based on the use of frequency-dependent dielectric functions for both Ag and Au films of small thickness down to a few atomic layers. This result is in stark contrast to the strong nonlocal effects observed in metal nanoparticles of similar or even larger diameter [50, 51], a result that we attribute to the predominance of in-plane electron motion associated with the low-energy plasmons of thin films, unlike the combination of in- and out-of-plane motion in higher energy SPPs.

II. THEORETICAL FORMALISM

We present the elements needed to calculate EELS probabilities in the nonretarded approximation using the linear response susceptibility to represent the metallic thin film. The latter is obtained in the RPA, starting from the one-electron wave functions of the system, which are organized as vertical quantum-well (QW) states, discretized by confinement along the out-of-plane direction and exhibiting quasi-free motion along the plane of the film. We further specify the EELS probability for electron trajectories either parallel or perpendicular with respect to the metal surfaces.

A. Calculation of EELS probabilities from the susceptibility in the nonretarded limit

The loss probability $\Gamma^{\rm EELS}(\omega)$ measured through EELS in electron microscopes must be normalized in such a way that $\int_0^\infty d\omega\, \hbar\omega\, \Gamma^{\rm EELS}(\omega)$ gives the average energy loss experienced by the electrons. Taking the latter to follow a straight-line trajectory with constant velocity vector $\mathbf{v}$ parallel to the z axis and impact parameter $\mathbf{R}_0 = (x_0, y_0)$, we can write [52]

$$\Gamma^{\rm EELS}(\omega) = \frac{e}{\pi\hbar\omega} \int dz\, {\rm Re}\left\{ E_z^{\rm ind}(\mathbf{R}_0, z, \omega)\, {\rm e}^{-{\rm i}\omega z/v} \right\} \quad (1)$$

as the integral along the electron trajectory of the frequency-resolved self-induced field $E_z^{\rm ind}(\mathbf{r},\omega) = \int dt\, E_z(\mathbf{r},t){\rm e}^{{\rm i}\omega t}$, which can be in turn calculated by solving the classical Maxwell equations with the electron point charge acting as an external source in the presence of the sample. This equation is rigorously valid within the approximations of linear response and nonrecoil (i.e., small energy loss $\hbar\omega$ compared with the electron kinetic energy E_0).

In the present study, we consider relatively small electron velocities $v \ll c$ and films of small thickness compared with the involved optical wavelengths. This allows us to work in the quasistatic limit and write the field $E_z^{\rm ind}(\mathbf{r},\omega) = -\partial_z \phi^{\rm ind}(\mathbf{r},\omega)$ as the gradient of a scalar potential, so Eq. (1) can be integrated by parts to yield

$$\Gamma^{\rm EELS}(\omega) = \frac{e}{\pi\hbar v} \int dz\, {\rm Im}\left\{ \phi^{\rm ind}(\mathbf{R}_0, z, \omega)\, {\rm e}^{-{\rm i}\omega z/v} \right\}. \quad (2)$$

We can now express the induced potential in terms of the induced charge as

$$\phi^{\rm ind}(\mathbf{r},\omega) = \int d^3\mathbf{r}'\, \nu(\mathbf{r},\mathbf{r}')\, \rho^{\rm ind}(\mathbf{r}',\omega), \quad (3)$$

where $\nu(\mathbf{r},\mathbf{r}')$ is the Coulomb interaction between point charges located at positions $\mathbf{r}$ and $\mathbf{r}'$. Likewise, we write the induced charge as $\rho^{\rm ind}(\mathbf{r},\omega) = \int d^3\mathbf{r}'\, \chi(\mathbf{r},\mathbf{r}',\omega)\phi^{\rm ext}(\mathbf{r}',\omega)$, where $\chi(\mathbf{r},\mathbf{r}',\omega)$ is the linear susceptibility, $\phi^{\rm ext}(\mathbf{r}',\omega) = \int d^3\mathbf{r}'\, \nu(\mathbf{r},\mathbf{r}')\, \rho^{\rm ext}(\mathbf{r}',\omega)$ is the external electric potential generated by the electron charge density $\rho^{\rm ext}(\mathbf{r},\omega) = -e\int dt\, \delta(\mathbf{r} - \mathbf{R}_0 - \mathbf{v}t)\, {\rm e}^{{\rm i}\omega t} = (-e/v)\delta(\mathbf{R} - \mathbf{R}_0)\, {\rm e}^{{\rm i}\omega z/v}$, and we use the notation $\mathbf{r} = (\mathbf{R}, z)$ with $\mathbf{R} = (x, y)$.

In free space one has $\nu(\mathbf{r},\mathbf{r}') = \nu_0(\mathbf{r} - \mathbf{r}') = 1/|\mathbf{r} - \mathbf{r}'|$, but we are interested in retaining a general spatial dependence of $\nu(\mathbf{r},\mathbf{r}')$ in order to describe the polarization background produced in the film by interaction with everything else other than conduction electrons (see below). Combining these elements with Eq. (2), we find the loss probability

$$\Gamma^{\rm EELS}(\omega) = \frac{e^2}{\pi\hbar v^2} \int d^3\mathbf{r} \int d^3\mathbf{r}'\, w^*(\mathbf{r})\, w(\mathbf{r}') \times {\rm Im}\left\{ -\chi(\mathbf{r},\mathbf{r}',\omega) \right\}, \quad (4)$$

where

$$w(\mathbf{r}) = \int dz' \nu(\mathbf{r}, \mathbf{R}_0, z')\, {\rm e}^{{\rm i}\omega z'/v} \quad (5)$$

is the *external* potential created by the electron and we have made use of the reciprocity property $\chi(\mathbf{r},\mathbf{r}',\omega) =$

$\chi(\mathbf{r}',\mathbf{r},\omega)$ to extract the complex factors w outside the imaginary part. Next, we apply this expression to calculate EELS probabilities from the RPA susceptibility. But first, for completeness, we note that the integral in Eq. (5) can be performed analytically for the bare Coulomb interaction [53] yielding

$$w(\mathbf{r}) = 2K_0(\omega|\mathbf{R}-\mathbf{R}_0|/v)\,\mathrm{e}^{\mathrm{i}\omega z/v},$$

where K_0 is a modified Bessel function [53], thus allowing us to write

$$\begin{aligned}\Gamma^{\mathrm{EELS}}(\omega) =& \frac{4e^2}{\pi\hbar v^2}\int d^3\mathbf{r}\int d^3\mathbf{r}'\,\cos\left[\frac{\omega}{v}(z'-z)\right]\\ &\times K_0\left(\frac{\omega}{v}|\mathbf{R}-\mathbf{R}_0|\right)K_0\left(\frac{\omega}{v}|\mathbf{R}'-\mathbf{R}_0|\right)\\ &\times \mathrm{Im}\left\{-\chi(\mathbf{r},\mathbf{r}',\omega)\right\}\end{aligned}$$

for the loss probability, which we can directly apply to systems in which any background polarization is already contained in χ, or when ν is well described by the bare Coulomb interaction (e.g., in simple metals).

B. RPA susceptibility of thin metal films

We follow the same formalism as in Ref. [33], which is extended here to account for an energy-dependence of the in-plane electron effective mass. One starts by writing $\chi(\mathbf{r},\mathbf{r}',\omega)$ in terms of the non-interacting susceptibility $\chi^0(\mathbf{r},\mathbf{r}',\omega)$ through $\chi = \chi^0\cdot(\mathcal{I}-\nu\cdot\chi^0)^{-1}$, where we use matrix notation with spatial coordinates $\mathbf{r}$ and $\mathbf{r}'$ acting as matrix indices, so that matrix multiplication involves integration over $\mathbf{r}$, and $\mathcal{I}(\mathbf{r},\mathbf{r}') = \delta(\mathbf{r}-\mathbf{r}')$. We further adopt the RPA by calculating χ^0 as [33, 54]

$$\chi^0(\mathbf{r},\mathbf{r}',\omega) = \frac{2e^2}{\hbar}\sum_{ii'}(f_{i'}-f_i)\,\frac{\psi_i(\mathbf{r})\psi_i^*(\mathbf{r}')\psi_{i'}^*(\mathbf{r})\psi_{i'}(\mathbf{r}')}{\omega+\mathrm{i}\gamma-(\varepsilon_i-\varepsilon_{i'})} \tag{6}$$

from the one-electron wave functions ψ_i of energies $\hbar\varepsilon_i$ and Fermi-Dirac occupation numbers f_i. Here, the factor of 2 accounts for spin degeneracy and γ is a phenomenological damping rate.

We describe metal films assuming translational invariance along the in-plane directions and parabolic electron dispersion with different effective mass m_j^* for each vertical QW band j. This allows us to write the electron wave functions as [55] $\psi_i(\mathbf{r}) = \varphi_j(z)\mathrm{e}^{\mathrm{i}\mathbf{k}_\parallel\cdot\mathbf{R}}/\sqrt{A}$, where $\mathbf{k}_\parallel$ is the 2D in-plane wave vector, A is the quantization area, and the state index is multiplexed as $i\to(j,\mathbf{k}_\parallel)$. Likewise, the electron energy can be separated as $\hbar\varepsilon_{j,k_\parallel} = \hbar\varepsilon_j^\perp + \hbar^2k_\parallel^2/2m_j^*$, where $\hbar\varepsilon_j^\perp$ is the out-of-plane energy that signals the QW band bottom. Inserting these expressions into Eq. (6) and making the customary substitution $\sum_i \to A\sum_j\int d^2\mathbf{k}_\parallel/(2\pi)^2$ for the state sums, we find [56]

$$\chi(\mathbf{r},\mathbf{r}',\omega) = \int\frac{d^2\mathbf{Q}}{(2\pi)^2}\,\chi(Q,z,z',\omega)\,\mathrm{e}^{\mathrm{i}\mathbf{Q}\cdot(\mathbf{R}-\mathbf{R}')}, \tag{7}$$

which directly reflects the in-plane homogeneity of the film. We can now work in $\mathbf{Q}$ space, where Eq. (6) reduces, using the above assumptions for the wave functions, to

$$\begin{aligned}&\chi^0(Q,z,z',\omega) \\ &= \frac{2e^2}{\hbar}\sum_{jj'}\chi_{jj'}(Q,\omega)\varphi_j(z)\varphi_j^*(z')\varphi_{j'}^*(z)\varphi_{j'}(z')\end{aligned} \tag{8}$$

where

$$\begin{aligned}\chi_{jj'}(Q,\omega) =& \int\frac{d^2\mathbf{k}_\parallel}{(2\pi)^2}\left(f_{j',|\mathbf{k}_\parallel-\mathbf{Q}|}-f_{j,k_\parallel}\right)\\ &\times\frac{1}{\omega+\mathrm{i}\gamma-\left[\varepsilon_j^\perp-\varepsilon_{j'}^\perp+\frac{\hbar}{2}\left(k_\parallel^2/m_j^*-|\mathbf{k}_\parallel-\mathbf{Q}|^2/m_{j'}^*\right)\right]},\end{aligned} \tag{9}$$

which only depends on the modulus of $\mathbf{Q}$ due to the in-plane band isotropy. We evaluate the integral in Eq. (9) assuming zero temperature [i.e., $f_{j,k_\parallel} = \theta(E_\mathrm{F}-\hbar\varepsilon_{j,k_\parallel})$, where E_F is the Fermi energy] and taking $\mathbf{Q} = (Q,0)$ without loss of generality.

Incidentally, simple manipulations of the above expressions reveal a dependence on frequency and damping through $(\omega+\mathrm{i}\gamma)^2$ that is maintained in the local limit ($Q\to 0$), in contrast to $\omega(\omega+\mathrm{i}\gamma)$ in the Drude model. The RPA formalism thus produces spectral features with roughly twice the width of the Drude model in the local limit. This problem (along with a more involved issue related to local conservation of electron number for finite attenuation) can be solved through a phenomenological prescription proposed by Mermin [57], which unfortunately becomes rather involved when applied to the present systems. As a practical and reasonably accurate solution, we proceed instead by setting $\gamma = \gamma^{\mathrm{exp}}/2$ in the above expressions (i.e., half the experimental damping rate, see Appendix A).

We obtain the out-of-plane wave functions $\varphi_j(z)$ as the eigenstates of the 1D Hamiltonian $-(\hbar^2/2m_\mathrm{e})\partial_{zz}+V(z)$, using the free-electron mass m_e for the transversal kinetic term and two different models for the confining potential $V(z)$: (*i*) the self-consistent solution in the jellium (JEL) approximation within density-functional theory (DFT) [20, 21]; and (*ii*) a phenomenological atomic-layer potential (ALP) that incorporates out-of-plane bulk atomic-layer corrugation and a surface density profile with parameters fitted to reproduce relevant experimental band structure features, such as affinity, surface state energy, and projected bulk band gap, which depend on material and crystal orientation as compiled in Ref. [62].

The JEL model corresponds to the self-consistent DFT solution for a thin slab of background potential and energy-independent effective mass $m_j^* = m_\mathrm{e}$ [20, 21], computed here through an implementation discussed elsewhere [63].

In the ALP model we fit m_j^* to experimental data (see Table I) and consider an effective electron density n_eff. Upon integration over the density of states of the parabolic QW bands, we can then write the Fermi energy

Material	$a(\mathrm{eV}^{-1})$	b	$m^*(\mathrm{SS})/m_\mathrm{e}$	m_0/m_e	n_eff/n_0	E_F (eV)
Ag(111)	-0.1549	-0.5446	0.40 [58]	0.25 [59]	0.8381	-4.63 [60]
Ag(100)	-0.0817	0.2116	-	0.40 [61]	0.8710	-4.43 [62]
Au(111)	-0.1660	-0.8937	0.26 [58]	0.26 [59]	0.9443	-5.50 [60]

TABLE I: Parameters used to describe the parabolic dispersion of quantum wells (QWs) in Ag(111), Ag(100), and Au(111) films. We take the effective mass of each QW j to linearly vary as $m_j^*/m_\mathrm{e} = a\hbar\varepsilon_j^\perp + b$ with band-bottom energy $\hbar\varepsilon_j^\perp$, where the parameters a and b are taken to match m_0 at the highest occupied QW (below the SS) in the semi-infinite surface and $m^* = m_\mathrm{e}$ at the bottom of the conduction band. The effective electron density n_eff, given here relative the bulk conduction electron density n_0, is required to fit the experimentally observed Fermi energy E_F and SS energy.

of a N-layer film as

$$E_\mathrm{F} = \left(\sum_{j=1}^{M} m_j^*\right)^{-1} \left(n_\mathrm{eff} a_s N\hbar^2\pi + \sum_{j=1}^{M} m_j^*\hbar\varepsilon_j^\perp\right), \quad (10)$$

where $j = M$ is the highest partially populated QW band (i.e., $\varepsilon_M^\perp < E_\mathrm{F}/\hbar < \varepsilon_{M+1}^\perp$) and a_s is the atomic interlayer spacing (i.e., the film thickness is $d = Na_s$, with $a_s = 0.236\,\mathrm{nm}$ for Ag(111) and Au(111), and $a_s = 0.205\,\mathrm{nm}$ for Ag(100)). This expression reduces to a similar one in Ref. [33] when m_j^* is independent of j. We adjust n_eff for each type of metal surface in such a way that Eq. (10) gives the experimental bulk values of E_F listed in Table I. Incidentally, although the effective mass of surface states also varies with energy [64, 65], we take it as constant because of the lack of data for ultrathin Au and Ag films; this should be a reasonable approximation for films consisting of $N \geq 5$ layers, where the surface state energy is already close to the semi-infinite surface level.

Conduction electrons interact through the bare Coulomb potential in simple metals, which in $\mathbf{Q}$ space reduces to $\nu(Q,z,z') = (2\pi/Q)\mathrm{e}^{-Q|z-z'|}$. However, polarization of inner electronic bands plays a major role in the dielectric response of Ag and Au. We describe this effect by modifying $\nu(Q,z,z')$ in order to account for the interaction between point charges in the presence of a dielectric slab of local background permittivity fitted to experimental data [66] after subtracting a Drude term representing conduction electrons (see Appendix A). We thus adopt the local response approximation for this contribution originating in localized inner electron states, whereas conduction electrons are treated nonlocally through the above RPA formalism. Similar to Eq. (7), translational symmetry in the film allows us to write

$$\nu(\mathbf{r},\mathbf{r}') = \int \frac{d^2\mathbf{Q}}{(2\pi)^2}\, \nu(Q,z,z')\, \mathrm{e}^{\mathrm{i}\mathbf{Q}\cdot(\mathbf{R}-\mathbf{R}')}, \quad (11)$$

where $\nu(Q,z,z')$ is reproduced for convenience from Ref. [33] in Appendix A. We note that Eq. (11) neglects the effect of lateral atomic corrugation in this interaction (i.e., the background permittivity is taken to be homogeneous inside the film).

Finally, we calculate $\chi(Q,z,z',\omega)$ from the noninteracting susceptibility [Eq. (8)] and the screened interaction by discretizing both of them in real space coordinates (z,z') and numerically performing the linear matrix algebra explained above. We obtain converged results with respect to the number of discretization points and also compared with an expansion in harmonic functions [33].

C. EELS probability under normal incidence

Direct insertion of Eqs. (7) and (11) into Eqs. (4) and (5) leads to the result

$$\Gamma_\perp^\mathrm{EELS}(\omega) = \int_0^\infty dQ\, \Gamma_\perp^\mathrm{EELS}(Q,\omega) \quad (12)$$

with

$$\Gamma_\perp^\mathrm{EELS}(Q,\omega) = \frac{e^2\,Q}{2\pi^2\hbar v^2} \times \int dz \int dz'\, I_\perp^*(Q,z) I_\perp(Q,z')\, \mathrm{Im}\left\{-\chi(Q,z,z',\omega)\right\}, \quad (13)$$

where

$$I_\perp(Q,z) = \int dz'\, \nu(Q,z,z')\, \mathrm{e}^{\mathrm{i}\omega z'/v} \quad (14)$$

contains the external electron potential. For completeness, we note that when $\nu(Q,z,z')$ is the bare Coulomb interaction $(2\pi/Q)\mathrm{e}^{-Q|z-z'|}$, Eq. (14) becomes $I_\perp(Q,z) = 4\pi\mathrm{e}^{\mathrm{i}\omega z/v}/(Q^2+\omega^2/v^2)$, so Eq. (13) reduces to

$$\Gamma_\perp^\mathrm{EELS}(Q,\omega) = \frac{8e^2}{\hbar v^2}\, \frac{Q}{(Q^2+\omega^2/v^2)^2} \times \int dz \int dz'\, \cos\left[\omega(z-z')/v\right]\, \mathrm{Im}\left\{-\chi(Q,z,z',\omega)\right\}, \quad (15)$$

where we have used reciprocity again [i.e., $\chi(Q,z,z',\omega) = \chi(Q,z',z,\omega)$].

In the simulations that we present below, we compare the RPA approach just presented with classical electromagnetic calculations based on the use of a local frequency-dependent dielectric function for the metal. This configuration has been theoretically studied for a long time [67], and in particular, we use the analytical expressions derived in a previous publication for an electron normally incident on a dielectric slab [68] with the bulk contribution integrated up to a cutoff wave vector $Q = 5\,\mathrm{nm}^{-1}$.

D. EELS probability in the aloof configuration

For an electron moving parallel to the film at a distance z_0 from the metal surface, it is convenient to make the substitutions $z \to x$, $\mathbf{R} \to (y, z)$, and $\mathbf{R}_0 \to (0, z_0)$ in Eqs. (4) and (5), so combining them with Eqs. (7) and (11), and retaining $\mathbf{R} = (x, y)$ in the latter, we readily obtain

$$\Gamma_{\parallel}^{\rm EELS}(\omega) = \frac{e^2 L}{\pi^2 \hbar v^2} \int_0^\infty dQ_y \times \int dz \int dz' \nu^*(Q, z, z_0) \nu(Q, z', z_0) \, {\rm Im}\left\{-\chi(Q, z, z', \omega)\right\}, \tag{16}$$

where $Q = \sqrt{\omega^2/v^2 + Q_y^2}$ and L is the electron path length. Again for completeness, when $\nu(Q, z, z')$ is the bare Coulomb interaction, Eq. (16) reduces to

$$\Gamma_{\parallel}^{\rm EELS}(\omega) = \frac{4e^2 L}{\hbar v^2} \int_0^\infty \frac{dQ_y}{Q^2} \times \int dz \int dz' {\rm e}^{-Q(|z-z_0|+|z'-z_0|)} \, {\rm Im}\left\{-\chi(Q, z, z', \omega)\right\}.$$

The above expressions can be applied to electron impact parameters z_0 both inside or outside the metal, but they can be simplified when the beam is not overlapping the conduction electron charge [see Fig. 3(a)], so that $z_0 > z, z'$ in the region inside the above integrals in which $\chi(Q, z, z', \omega)$ is nonzero, and therefore, changing the variable of integration from Q_y to Q, we can write

$$\Gamma_{\parallel}^{\rm EELS}(\omega) = \frac{2e^2 L}{\pi \hbar v^2} \int_{\omega/v}^\infty dQ \, \frac{{\rm e}^{-2Qz_0}}{\sqrt{Q^2 - \omega^2/v^2}} \, {\rm Im}\{r_{\rm p}(Q, \omega)\}, \tag{17}$$

where

$$r_{\rm p}(Q, \omega) = -\frac{Q}{2\pi} \int dz \int dz' \, \nu^*(Q, z, z_0) \nu(Q, z', z_0) \times {\rm e}^{2Qz_0} \, \chi(Q, z, z', \omega) \tag{18}$$

is the Fresnel reflection coefficient of the film for p polarization in the quasistatic limit. Incidentally, Eq. (18) is independent of the source location z_0 when it does not overlap the metal because $\nu(Q, z, z_0)$ then depends on z_0 only through a factor ${\rm e}^{-Qz_0}$ (see Appendix A). Equation (17), which agrees with previous derivations from classical dielectric theory [69], reveals ${\rm Im}\{r_{\rm p}(Q, \omega)\}$ as a loss function, which is used below to visualize the surface plasmon dispersion. We also provide results from a local dielectric description based on the textbook solution of the Poisson equation for the reflection coefficient [33]

$$r_{\rm p}^{\rm classical} = \frac{(\epsilon^2 - 1)\left(1 - {\rm e}^{-2Qd}\right)}{(\epsilon + 1)^2 - (\epsilon - 1)^2 {\rm e}^{-2Qd}} \tag{19}$$

for a metal film of thickness d and permittivity ϵ.

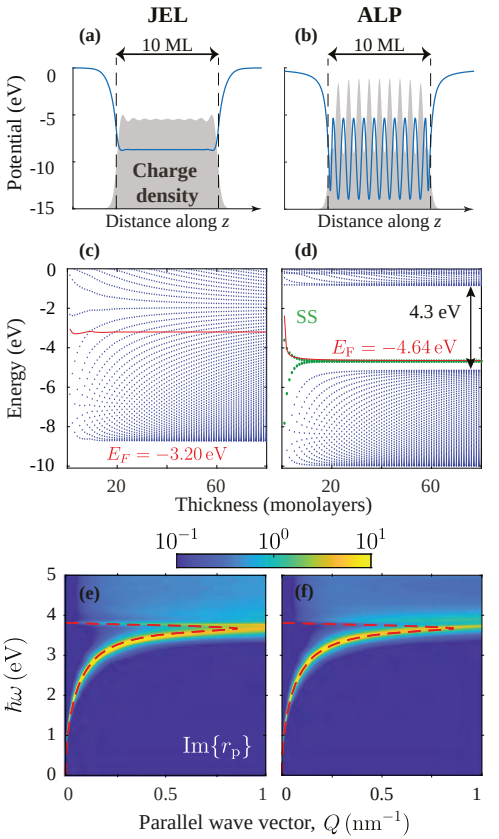

FIG. 1: RPA description of plasmons in atomically thin Ag(111) films. (a,b) Effective confining potential for conduction electrons across a 10 ML film. The conduction charge density is shows as shaded areas. (d,c) Electronic energies as a function of film thickness expressed as the number of (111) atomic layers (blue dots). Red curves and green dots represent the Fermi energy and the surface states (SSs). (e,f) Loss function ${\rm Im}\{r_{\rm p}\}$ calculated in the RPA [color plot, Eq. (18)], compared with the plasmon dispersion relation in the local Drude dielectric model (red curves). Left (right) panels are calculated in the jellium (ALP) model.

III. RESULTS AND DISCUSSION

We show examples of the two types of confining electron potentials used in our RPA calculations for Ag films in Fig. 1(a,b), along with the resulting conduction electron charge densities. The JEL potential is smooth at the surface and describes electron spill-out and Friedel oscillations [70]. The phenomenological ALP potential further incorporates corrugations due to the atomic planes in the bulk, which result in strong oscillations of the den-

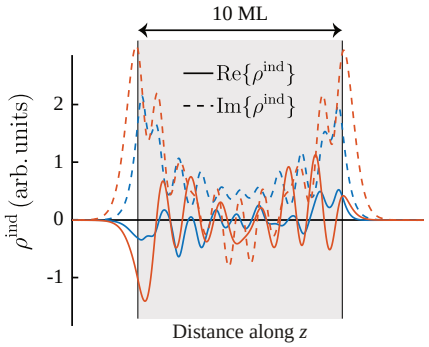

FIG. 2: Plasmon charge density across a thin 10 ML Ag(111) film. We plot the real (solid curves) and imaginary (dashed curves) parts of the induced charge density $\rho^{\rm ind}$ as calculated in the RPA for excitation by a source placed to the left of the film at the plasmon energies $\hbar\omega = 3.54\,$eV and $\hbar\omega = 3.47\,$eV corresponding to a parallel wave vector $Q = 0.5\ \text{nm}^{-1}$ in the ALP (blue) and JEL (orange) models, respectively.

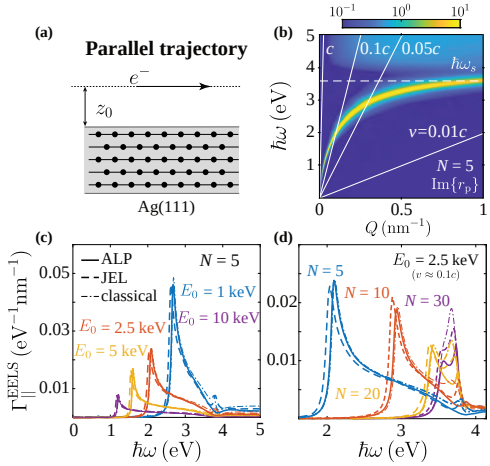

FIG. 3: Aloof EELS in thin Ag(111) films. (a) Scheme showing an electron moving parallel to a $N = 5$ ML Ag(111) metal film at a distance z_0 from its upper surface. (b) Dispersion diagram showing $\text{Im}\{r_{\rm p}\}$ calculated in the ALP model for the film shown in (a). White solid lines correspond to $\omega = vQ$ for different velocities v, while the dashed horizontal line shows the classical high-Q asymptotic surface-plasmon energy $\hbar\omega_{\rm s} \simeq 3.7\,$eV. (c,d) EELS probability per unit of path length for $z_0 = 0.5\,$nm calculated using different models [see legend in (c)] for (c) different electron kinetic energies E_0 with fixed $N = 5$ and (d) different N's with $E_0 = 2.5\,$keV.

sity. The computed electron energies $\hbar\varepsilon_j$ (see Sec. II B), which correspond to the bottom points of the QW bands (i.e., for vanishing in-plane momentum), are distributed with N of them below the Fermi level in a Ag(111) film consisting of N monolayers [Fig. 1(c,d)]. The band structure quickly evolves toward the semi-infinite surface for a few tens of MLs in both models. Additionally, the ALP potential hosts surface states and a projected bulk gap of energies fitted to experiment [62]. We note that this gap depends on surface orientation: it is present in Ag(111) but absent in Ag(100) at the Fermi level, as revealed by photoemission measurements [**?**] see also Fig. 9(a) in Appendix B]. Remarkably, despite the important differences in the details of the potentials and electron bands, both models predict a similar plasmon dispersion [Fig. 1(e,f), density plots, obtained from Eq. (18)], which is in excellent agreement with classical theory [Fig. 1(e,f), red curves, obtained from the poles of Eq. (19)]. Incidentally, we observe the response to also converge toward the semi-infinite surface limit for a few tens of atomic layers [see Fig. 10 in Appendix B. Similar good agreement is found in the reflection coefficients of Ag films computed for different thickness with either of these potentials, with a square-barrier potential, or with a model potential constructed by glueing on either film side a jellium DFT potential tabulated for semi-infinite surfaces [20] [see Fig. 11 in Appendix **??**.

The transversal distribution of change densities associated with thin film plasmons show a clear resemblance when calculated using the ALP or JEL model potentials, although one can still observe substantial discrepancies between the two of them [see for example Fig. 2, where the ALP model charge appears to be smaller in magnitude]. However, this different behavior hardly reflects in the dispersion relation and plasmon strength [Fig. 1]. Interestingly, the z-integrated charge is nonzero, revealing that plasmons involves net charge oscillations along the in-plane directions for finite wave vector Q.

We conclude from these results that it is the effective number of valence electrons participating in the plasmons what determines their main characteristics, irrespective of the details of the electron wave functions and induced charge densities.

The loss function $\text{Im}\{r_{\rm p}\}$ provides a convenient way to represent the plasmon dispersion relation, as plasmons produce sharp features in the Fresnel reflection coefficient for p polarization. A weighted integral of this quantity over in-plane wave vectors gives the EELS probability under parallel aloof interaction [Fig. 3(a)] according to Eq. (17). However, the integration limit has a threshold at $\omega = Qv$ and the weighting factor multiplying the loss function in the integrand diverges precisely at that point. The cutoff condition $\omega = Qv$ is represented in Fig. 3(b) for different electron velocities (white lines) along with the loss function (density plot). As expected, the points of intersection with the plasmon band produce a dominant contribution that pops up as sharp peaks in the resulting EELS spectra [Fig. 3(c,d)]. An increase in electron velocity (i.e., in the slope of the threshold line) results in a redshift of the spectral peak [Fig. 3(c)],

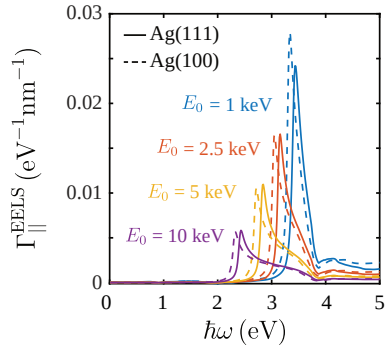

FIG. 4: Plasmon dependence on crystallographic surface orientation: Ag(111) and Ag(100) films. We compare EELS spectra calculated in the ALP model under the same conditions as in Fig. 3(c,d) for $N = 13$ ML Ag(111) and $N = 15$ ML Ag(100) films (thickness ratio differing by $< 0.1\%$).

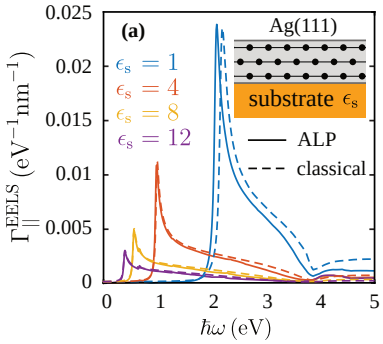

FIG. 5: Substrate-induced plasmon shift. We show EELS spectra for 2.5 keV electrons calculated in either the ALP model or the local classical description under the same conditions as in Fig. 3 for a Ag(111) film consisting of $N = 5$ MLs supported on a planar dielectric substrate of permittivity ϵ_s as indicated by labels.

and likewise, thinner films show plasmons moving farther away from the $\omega = Qc$ light line, thus producing shifts toward higher plasmon energies in the EELS spectra for fixed electron energy. We remark that RPA and classical calculations lead to quantitatively similar results for this configuration, and the former are roughly independent of the choice of confining electron potential.

The ALP model incorporates experimental information on electronic bands, which depend on crystallographic orientation (see Table I). We explore the effects of this dependence by comparing aloof EELS spectra obtained from Ag(111) and Ag(100) films in Fig. 4. In order to eliminate discrepancies arising from differences in thickness, we consider films consisting of $N = 13$ and $N = 15$ MLs, respectively, so that the thickness ratio is $(2/\sqrt{3}) \times (13/15) \approx 1.001$. We remind that Ag(111) displays a projected bulk gap in the electronic bands, in contrast to Ag(100) [see Fig. 9(a) in Appendix B]; as a consequence the former supports electronic surface states unlike the latter [62]. Despite these remarkable differences in electronic structure, the resulting spectra look rather similar, except for a small redshift of Ag(100) plasmon peaks relative to Ag(111), comparable in magnitude to those observed in semi-infinite Ag(111) and Ag(110) crystal surfaces through angle-resolved low-energy EELS [72], although the actual magnitude of the shift might be also influenced by electron confinement in our ultrathin films.

The presence of a dielectric substrate of permittivity ϵ_s is known to redshift the plasmon frequency of thin films by a factor $\sim 1/\sqrt{1+\epsilon_s}$ due to the attractive image interaction [73]. This effect is observed in our calculated aloof EELS spectra, for which we obtain the combined film-substrate reflection coefficient by using a Fabry-Perot approach, as discussed elsewhere [33]. We find again excellent agreement between RPA simulations using the ALP potential and classical calculations [Fig. 5], and in fact, the resemblance between the spectral profiles obtained with both methods increases with ϵ_s.

In Fig. 6 we examine the way lateral dispersion of QW states affects the plasmonic properties of ultrathin Ag films when using the ALP potential. Comparison of the band structures calculated with [Fig. 6(b)] and without [Fig. 6(a)] inclusion of an energy dependence in the in-plane effective mass anticipates a clear difference between the two of them: the latter shows the same energy jumps between different bands irrespective of the electron parallel wave vector $k_{\parallel}$; those energy jumps will therefore be favored in the optical response, giving rise to spurious spectral features. In contrast, differences in lateral dispersion associated with the energy dependence of the effective mass (described here by fitting existing angle-resolved photoemission data [58, 59, 61, 74, 75]) should at least partially wash out those spectral features. This is clearly observed in the resulting dispersion diagrams [Fig. 6(c,d)] and aloof EELS spectra [Fig. 6(e,f)]. In particular, the dispersion relation for constant m_j^* [Fig. 6(c)] reveals a complex mixture of resonances at energies above 3 eV, which we find to be strongly affected by the HOMO-LUMU gap energy (not shown); these resonances cause fine structure in the EELS spectra that disappears when a realistic energy dependence is introduced in the lateral effective mass [Fig. 6(e)].

We also analyse EELS spectra for normally impinging electron beams [Fig. 7]. The momentum- and energy-resolved EELS probability given by Eq. (13) reveals the plasmon dispersion in analogy to the loss function [cf. Figs. 3(b) and 7(b)]. But now, this quantity is directly accessible under normal incidence by recording angle- and energy-dependent electron transmission intensities, as already done in pioneering experiments for thicker Al films showing both bonding and antibonding plasmon disper-

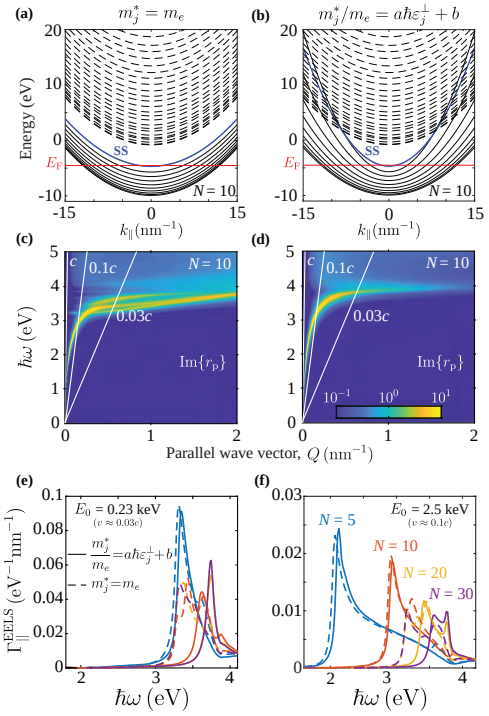

FIG. 6: The role of the electron effective mass. (a,b) In-plane parabolic QW bands of a $N = 10$ ML Ag(111) film in the ALP model with (a) constant and (b) energy-dependent effective mass ($m_j^* = m_e$ and $m_j^* = (a\hbar\varepsilon_j^\perp + b)m_e$, respectively, see Table I). The surface state bands (blue curves) have a mass $0.4\,m_e$. Solid (dashed) curves represent bands that are occupied (unoccupied) at $k_\parallel = 0$. The Fermi level is shown as a horizontal red line. (c,d) Loss function $\mathrm{Im}\{r_p\}$ under the conditions of (a,b), respectively. (e,f) EELS probability under parallel aloof interaction at a distance $z_0 = 0.5$ nm for two different electron energies corresponding to the $\omega = Qv$ lines shown in (c,d) and different film thicknesses (see labels) calculated in the ALP model with constant (dashed curves) and energy-dependent (solid curves) electron effective mass.

sions [34]. In contrast to the aloof configuration, the transmission EELS spectra exhibit broader plasmon features [Fig. 7(c,d)], which in the thin film limit [69] are the result of weighting the loss function with a profile $Q^2/(Q^2 + \omega^2/v^2)^2$ [see also Eq. (15), where an extra factor of Q emerges from χ in the small Q limit], represented in Fig. 7(b) for 2.5 keV electrons and different energies $\hbar\omega$ (colored curves); these spectra reveal indeed a broad spectral overlap with the plasmon band. Again, we observe very similar results from RPA and classical descriptions, and just a minor dependence on electron potential in the former.

We conclude by showing EELS calculations for

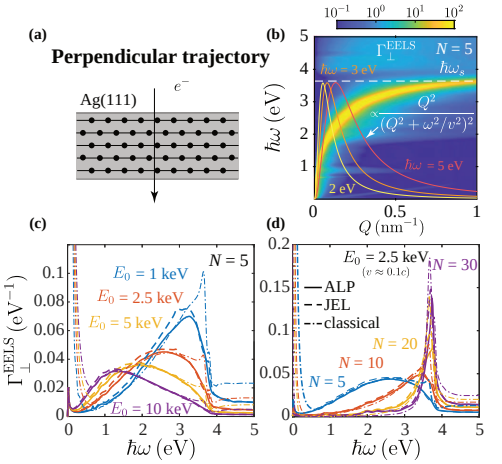

FIG. 7: EELS in thin Ag(111) films under normal incidence. (a) Scheme showing an electron normally traversing a $N = 5$ ML Ag(111) metal film. (b) Momentum- and energy-resolved EELS probability $\Gamma_\perp^{\mathrm{EELS}}(Q, \omega)$ [Eq. (13)] calculated for $E_0 = 2.5$ keV electrons ($v/c \approx 0.1$) in the ALP model for the film shown in (a). Colored solid curves show $Q^2/(Q^2 + \omega^2/v^2)^2$ profiles as a function of Q for different energy losses $\hbar\omega = 2$, 3, and 5 eV, while the dashed horizontal line indicates $\hbar\omega_s$. (c,d) EELS probability calculated using different models [see legend in (d)] for (c) different electron kinetic energies E_0 with fixed $N = 5$ and (d) different N's with $E_0 = 2.5$ keV.

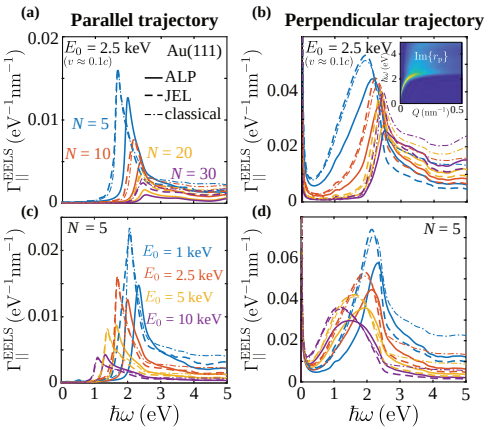

FIG. 8: EELS spectra for gold Au(111) films. We consider for (a,c) aloof and (b,d) normal trajectories for either (a,b) fixed electron energy ($E_0 = 2.5$ keV) and varying film thickness ($N = 5$-30 MLs) or (c,d) fixed $N = 5$ and varying electron energy. Calculations for the same models as in Fig. 3 are presented. The plasmon dispersion is shown for $N = 10$ MLs using the ALP model in the inset of (b).

Au(111) films in Fig. 8. This noble metal has a similar conduction electron density as Ag, but the Au d-band is closer to the Fermi energy, therefore producing large screening ($\epsilon_\mathrm{b} \sim 9$ in the plasmonic region) compared with Ag ($\epsilon_\mathrm{b} \sim 4$, see Fig. S4 in SI). This causes a shift of the high-Q surface plasmon asymptote down to $\hbar\omega_\mathrm{s} \simeq 2.5\,\mathrm{eV}$. Additionally, damping is also stronger (more than three times larger than in Ag, see Appendix A), which results in broader spectral features [cf. Fig. 8 for Au and Figs. 3(c,d) and 7(c,d)]. Interestingly, we observe significant blue shifts in the plasmon spectral features when using the ALP potential as compared with both jellium DFT and classical models. This effect could originate in a more substantial role played by the electronic band structure in Au(111) because the projected bulk gap extends further below the Fermi level, and additionally, the surface state band is also more deeply bound [see Fig. 9(b) in Appendix B]. This is consistent with the general dependence of the optical surface conductivity on Fermi momentum k_F and velocity v_F: in the Drude model for graphene and the two-dimensional electron gas, this quantity is proportional to $k_\mathrm{F} v_\mathrm{F}$ and the surface plasmon frequency scales as $\propto \sqrt{k_\mathrm{F} v_\mathrm{F}}$; the situation is more complicated in our thin films because they have multiple 2D bands crossing the Fermi level, but the presence of a deeper gap in Au(111) indicates that the effective band-averaged value of $k_\mathrm{F} v_\mathrm{F}$ (i.e., with k_F defined by the crossing of each QW at the Fermi level and v_F as the slope of the parabolic dispersion at that energy) is larger than in Ag surfaces, characterized by the presence of shallower bands near E_F; we thus expect an increase in Drude weight, and consequently, a plasmon blue shift, in Au(111) relative to Ag; this argument is reinforced by the small effective mass of surface states in Au(111) compared with Ag(111), which also pushes up their associated v_F. In summary, the plasmon blue shifts observed in Au(111) when using the realistic ALP potential seem to have a physical origin, although more sophisticated first principles simulations might be needed to conclusively support this finding.

IV. CONCLUSION

In summary, we have shown that a local classical dielectric model predicts reasonably well the intensities and dispersion relations of plasmons in ultrathin silver films when compared to quantum-mechanical simulations based on the RPA with different potentials used to simulate the conduction one-electron wave functions. We attribute the small effect of nonlocality in the plasmonic response of these films to the fact that their associated electron motion takes place along in-plane directions, in contrast to metal nanoparticles with a similar size as the film thickness here considered (i.e., electron surface scattering is unavoidable in such particles, thus introducing important nonlocal effects). We confirm this agreement between classical and quantum simulations in Ag films down to a few atomic layers in thickness [33, 63], consistent with previous smooth-interface hydrodynamic theory [76]. Additionally, our quantum RPA simulations are relatively insensitive to the details of the confining electron potential, so similar results are obtained when using either a smooth jellium DFT model or a phenomenological potential that incorporates atomic-layer corrugation to fit relevant elements of the electronic band structure. In particular, the latter produces results that are rather independent of the crystallographic orientation of the film. Nonetheless, it is important to introduce the correct energy dependence of the out-of-plane effective mass in the phenomenological potential model, as otherwise spurious features show up in the calculated plasmon spectra. Although these potentials lead to substantially different plasmon charge distributions, spatial integration gives rise to similar plasmon dispersion relations. Interestingly, band effects described in the ALP potential model are more significant in Au, where they produce plasmon blue shifts relative to the predictions of classical and jellium DFT simulations; we attribute this different behavior in Au(111) relative to Ag(111) and Ag(100) to the fact that the former surface exhibits a projected bulk gap that extends further below the Fermi level, and additionally, this gives rise to more bound surface states. We remark that EELS provides the means to access the dispersion relations of strongly confined plasmons in ultrathin metal films, which are too far from the light line to be measured by means of optical techniques.

Acknowledgments

This work has been supported in part by the ERC (Advanced Grant 789104-eNANO), the Spanish MINECO (MAT2017-88492-R and SEV2015-0522), the Catalan CERCA Program, the Fundació Privada Cellex, and the Quscope center sponsored by the Villum Foundation.

Appendix A: Background screened interaction

We introduce the effect of interband polarization in the plasmonic spectral region of noble metals through a dielectric slab of permittivity $\epsilon_\mathrm{b}(\omega) = \epsilon(\omega) + \omega_\mathrm{p}^2/\omega(\omega + \mathrm{i}\gamma^\mathrm{exp})$, that is, the local dielectric function of the bulk metal $\epsilon(\omega)$ from which we subtract a classical bulk Drude term representing the contribution of conduction electrons. In practice, we take $\epsilon(\omega)$ from measured optical data [66] and use parameters $\hbar\omega_\mathrm{p} = 9.17\,\mathrm{eV}$ and $\hbar\gamma^\mathrm{exp} = 21\,\mathrm{meV}$ for Ag, and $\hbar\omega_\mathrm{p} = 9.06\,\mathrm{eV}$ and $\hbar\gamma^\mathrm{exp} = 71\,\mathrm{meV}$ for Au. The resulting $\epsilon_\mathrm{b}(\omega)$ is plotted in Fig. 12 in Appendix B. Incidentally, as we explain in Sec. II B, we set the damping parameter to $\gamma = \gamma^\mathrm{exp}/2$ in the RPA formalism in order to fit the experimental plasmon width. Following previous work [19], we take the background dielectric slab to have a thickness $d = Na_s$, where N is the number of atomic layers and a_s is the interlayer spacing, so that it extends symmetrically a distance $a_s/2$ outside

the outer atomic plane on each side of the film.

We reproduce for convenience a previously reported expression [33] for the screened interaction, used here to account for background polarization in the a self-standing metal film of thickness d and background permittivity $\epsilon_{\rm b}$ contained in the $0 < z < d$ region:

$$\nu(Q,z,z') = \nu^{\rm dir}(Q,z,z') + \nu^{\rm ref}(Q,z,z'),$$

where

$$\nu^{\rm dir}(Q,z,z') = \frac{2\pi}{Q}\,{\rm e}^{-Q|z-z'|} \times \begin{cases} 1, & z,z' \le 0 \text{ or } z,z' > d \\ 1, & 0 < z,z' \le d \\ 0, & \text{otherwise} \end{cases}$$

and

$$\nu^{\rm ref}(Q,z,z') = \frac{(2\pi/Q)}{(\epsilon_{\rm b}+1)^2 - (\epsilon_{\rm b}-1)^2{\rm e}^{-2Qd}} \times \begin{cases} (1-\epsilon_{\rm b}^2)\left({\rm e}^{2Qd}-1\right){\rm e}^{-Q(z+z')}, & d < z,z' \\ 2\left[(\epsilon_{\rm b}+1){\rm e}^{-Q(z-z')} + (\epsilon_{\rm b}-1){\rm e}^{-Q(z+z')}\right], & 0 < z' \le d < z \\ 4\epsilon_{\rm b}\,{\rm e}^{-Q(z-z')}, & z' \le 0 \text{ and } d < z \\ 2\left[(\epsilon_{\rm b}+1){\rm e}^{Q(z-z')} + (\epsilon_{\rm b}-1){\rm e}^{-Q(z+z')}\right], & 0 < z \le d < z' \\ (1/\epsilon_{\rm b})\big\{(\epsilon_{\rm b}^2-1)\left[{\rm e}^{-Q(z+z')} + {\rm e}^{-Q(2d-z-z')}\right] & \\ \quad +(\epsilon_{\rm b}-1)^2\left[{\rm e}^{-Q(2d+z-z')} + {\rm e}^{-Q(2d-z+z')}\right]\big\}, & 0 < z,z' \le d \\ 2\left[(\epsilon_{\rm b}+1){\rm e}^{-Q(z-z')} + (\epsilon_{\rm b}-1){\rm e}^{-Q(2d-z-z')}\right], & z' \le 0 < z \le d \\ 4\epsilon_{\rm b}\,{\rm e}^{Q(z-z')}, & z \le 0 \text{ and } d < z' \\ 2\left[(\epsilon_{\rm b}+1){\rm e}^{Q(z-z')} + (\epsilon_{\rm b}-1){\rm e}^{-Q(2d-z-z')}\right], & z \le 0 < z' \le d \\ (1-\epsilon_{\rm b}^2)\left(1-{\rm e}^{-2Qd}\right){\rm e}^{Q(z+z')}. & z,z' \le 0 \end{cases}$$

For completeness, we illustrate the dramatic effects of interband processes in Fig. 13 in Appendix B by comparing calculations obtained for Ag films using either screened or bare Coulomb interactions.

Appendix B: Additional figures

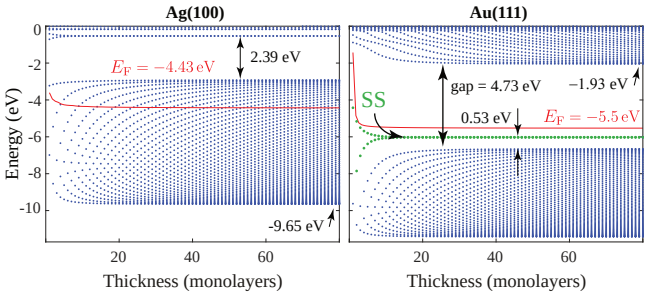

FIG. 9: ALP model calculations similar to those of Fig. 1(d), but for Ag(100) and Au(111) films.

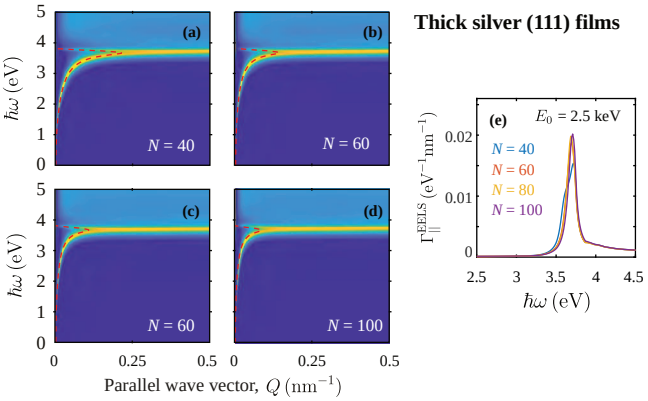

FIG. 10: ALP model calculations for (a-d) the loss function $\mathrm{Im}\{r_p\}$ of Ag(111) films of different thickness N and (b) the resulting EELS spectra of normally incident 2.5 keV electrons.

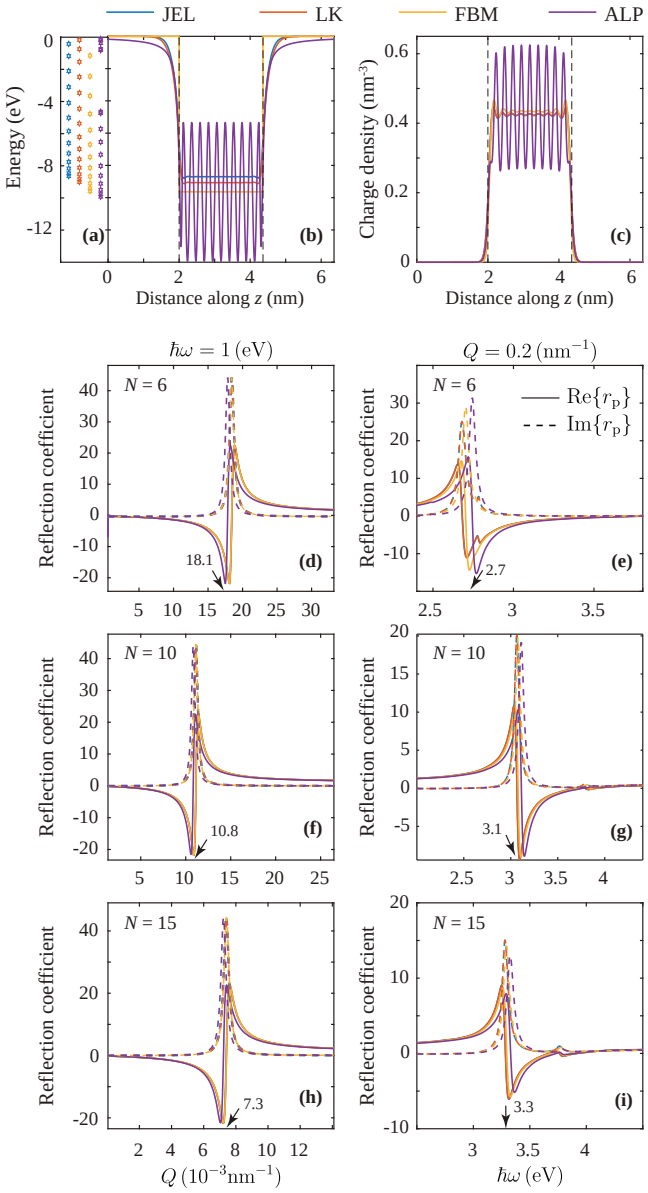

FIG. 11: Dependence of the RPA response on model potential. We show (a) the binding conduction electron energies, (b) the confining potential, and (c) the conduction electron density for a $N = 10$ ML Ag(111) film, as well as (d-i) the reflection coefficient r_p of Ag(111) films of different thickness N for either (d,f,h) fixed photon energy $\hbar\omega$ as a function parallel wave vector Q or (e,g,i) fixed Q as a function of $\hbar\omega$. We calculate r_p in the RPA and consider different confining electron potentials, as indicated by the upper labels: JEL and ALP, defined in the main text; LK, a superposition of the parametrized jellium DFT potential for semi-infinite surfaces taken from Lang and Kohn [Phys. Rev. B **1**, 4555 (1970)] for a one-electron radius $r_s = 3$ a.u., adopted for each of the film surfaces and glued by hand at the film center; and FBM, a square-well finite-barrier model potential. Only the ALP incorporates an energy dependence on the lateral effective mass, while the rest of the models assume $m_j^* = m_\mathrm{e}$.

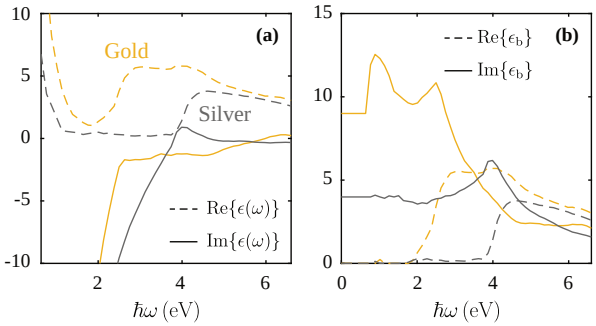

FIG. 12: Dielectric function $\epsilon(\omega)$ of Ag and Au taken from Jonhson and Christy [Phys. Rev. B **6**, 4370 (1972)] and background permittivity $\epsilon_{\rm b}(\omega) = \epsilon(\omega) + \omega_{\rm p}^2/\omega(\omega + i\gamma^{\rm exp})$ obtained by subtracting a Drude term with parameters $\hbar\omega_{\rm p} = 9.17\,$eV and $\hbar\gamma^{\rm exp} = 21\,$meV for Ag, and $\hbar\omega_{\rm p} = 9.06\,$eV and $\hbar\gamma^{\rm exp} = 71\,$meV for Au. We take $\epsilon_{\rm b} = 4$ and $\epsilon_{\rm b} = 9.5$ for Ag and Au in the $\hbar\omega < 0.6\,$eV region, which is not covered in the above reference.

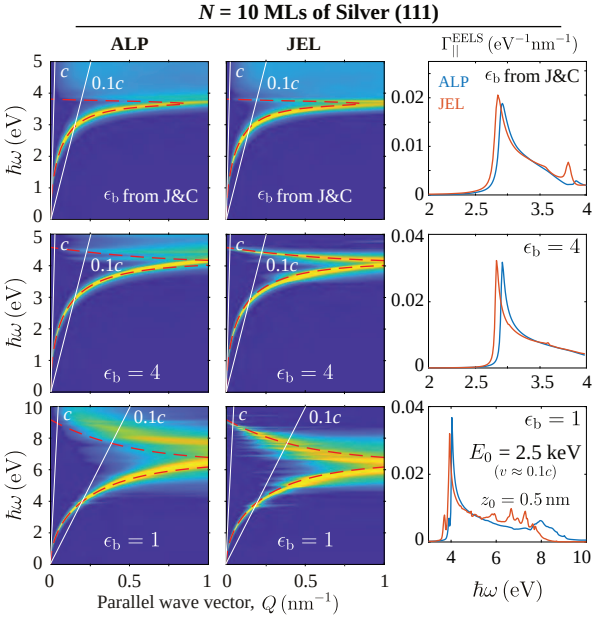

FIG. 13: Effect of background screening. We show dispersion diagrams (color plots) showing the loss function $\mathrm{Im}\{r_{\rm p}\}$ of $N = 10\,$ML Ag(111) films as calculated in the RPA for ALP (left) and JEL (center) model potentials when the background permittivity $\epsilon_{\rm b}$ is obtained from optical data (upper plots, see Fig. 12) or set to a constant value $\epsilon_{\rm b} = 4$ (middle) or $\epsilon_{\rm b} = 1$ (bottom). We find $\epsilon_{\rm b} = 4$ to represent approximately well background screening in silver over the plasmonic spectral region, whereas $\epsilon_{\rm b} = 1$ gives rise to unrealistic blue shifts. These conclusions are maintained when examining aloof EELS spectra (right plots, calculated for 2.5 keV electrons passing at a distance of 0.5 nm from the metal surface).

[1] D. J. Bergman and M. I. Stockman, Phys. Rev. Lett. **90**, 027402 (2003).
[2] H. Xu, E. J. Bjerneld, M. Käll, and L. Börjesson, Phys. Rev. Lett. **83**, 4357 (1999).
[3] A. Polman, Science **322**, 868 (2008).
[4] J. N. Anker, W. P. Hall, O. Lyandres, N. C. Shah, J. Zhao, and R. P. Van Duyne, Nat. Mater. **7**, 442 (2008).
[5] D. Rodrigo, O. Limaj, D. Janner, D. Etezadi, F. J. García de Abajo, V. Pruneri, and H. Altug, Science **349**, 165 (2015).
[6] Z. W. Seh, S. Liu, M. Low, S.-Y. Zhang, Z. Liu, A. Mlayah, and M.-Y. Han, Adv. Mater. **24**, 2310 (2012).
[7] C. Clavero, Nat. Photon. **8**, 95 (2014).
[8] H. A. Atwater and A. Polman, Nat. Mater. **9**, 205 (2010).
[9] C. F. Guo, T. Sun, F. Cao, Q. Liu, and Z. Ren, Light Sci. Appl. **3**, e161 (2014).
[10] M. Danckwerts and L. Novotny, Phys. Rev. Lett. **98**, 026104 (2007).
[11] R. W. Boyd, *Nonlinear optics* (Academic Press, Amsterdam, 2008), 3rd ed.
[12] A. R. Davoyan, I. V. Shadrivov, and Y. S. Kivshar, Opt. Express **16**, 21209 (2008).
[13] S. Palomba and L. Novotny, Phys. Rev. Lett. **101**, 056802 (2008).
[14] R. H. Ritchie, Phys. Rev. **106**, 874 (1957).
[15] C. J. Powell and J. B. Swan, Phys. Rev. **115**, 869 (1959).
[16] K. D. Tsuei, E. W. Plummer, and P. J. Feibelman, Phys. Rev. Lett. **63**, 2256 (1989).
[17] M. Rocca, L. Yibing, F. B. de Mongeot, and U. Valbusa, Phys. Rev. B **52**, 14947 (1995).
[18] J. Daniels, Z. Physik **203**, 235 (1967).
[19] A. Liebsch, Phys. Rev. B **48**, 11317 (1993).
[20] N. D. Lang and W. Kohn, Phys. Rev. B **1**, 4555 (1970).
[21] N. D. Lang and W. Kohn, Phys. Rev. B **7**, 3541 (1973).
[22] J. M. Pitarke, V. M. Silkin, E. V. Chulkov, and P. M. Echenique, Rep. Prog. Phys. **70**, 1 (2007).
[23] J. F. Dobson and G. H. Harris, J. Phys. C **21**, L729 (1988).
[24] A. Eguiluz, S. C. Ying, and J. J. Quinn, Phys. Rev. B **11**, 2118 (1975).
[25] K. D. Tsuei, E. W. Plummer, A. Liebsch, K. Kempa, and P. Bakshi, Phys. Rev. Lett. **64**, 44 (1990).
[26] A. Liebsch and W. L. Schaich, Phys. Rev. B **52**, 14219 (1995).
[27] H. Raether, *Surface Plasmons on Smooth and Rough Surfaces and on Gratings*, vol. 111 of *Springer Tracks in Modern Physics* (Springer-Verlag, Berlin, 1988).
[28] S. A. Maier, *Plasmonics: Fundamentals and Applications* (Springer, New York, 2007).
[29] W. L. Barnes, A. Dereux, and T. W. Ebbesen, Nature **424**, 824 (2003).
[30] R. A. Álvarez-Puebla, L. M. Liz-Marzán, and F. J. García de Abajo, J. Phys. Chem. Lett. **1**, 2428 (2010).
[31] E. N. Economou, Phys. Rev. **182**, 539 (1969).
[32] Z. M. Abd El-Fattah, V. Mkhitaryan, J. Brede, L. Fernández, C. Li, Q. Guo, A. Ghosh, A. R. Echarri, D. Naveh, F. Xia, et al., ACS Nano **13**, 7771 (2019).
[33] A. R. Echarri, J. D. Cox, and F. J. García de Abajo, Optica **6**, 630 (2019).
[34] R. Vincent and J. Silcox, Phys. Rev. Lett. **31**, 1487 (1973).
[35] F. Moresco, M. Rocca, T. Hildebrandt, and M. Henzler, Phys. Rev. Lett. **83**, 2238 (1999).
[36] E. P. Rugeramigabo, T. Nagao, and H. Pfnür, Phys. Rev. B **78**, 155402 (2008).
[37] H. V. Chung, C. J. Kubber, G. Han, S. Rigamonti, D. Sánchez-Portal, D. Enders, A. Pucci, and T. Nagao, Appl. Phys. Lett. **96**, 243101 (2010).
[38] E. P. Rugeramigabo, C. Tegenkamp, H. Pfnür, T. Inaoka, and T. Nagao, Phys. Rev. B **81**, 165407 (2010).
[39] T. Nagao, S. Yaginuma, T. Inaoka, and T. Sakurai, Phys. Rev. Lett. **97**, 116802 (2006).
[40] G. X. Ni, A. S. McLeod, Z. Sun, L. Wang, L. Xiong, K. W. Post, S. S. Sunku, B.-Y. Jiang, J. Hone, C. R. Dean, et al., Nature **557**, 530 (2018).
[41] Z. Fei, A. S. Rodin, G. O. Andreev, W. Bao, A. S. McLeod, M. Wagner, L. M. Zhang, Z. Zhao, M. Thiemens, G. Dominguez, et al., Nature **487**, 82 (2012).
[42] J. Chen, M. Badioli, P. Alonso-González, S. Thongrattanasiri, F. Huth, J. Osmond, M. Spasenović, A. Centeno, A. Pesquera, P. Godignon, et al., Nature **487**, 77 (2012).
[43] M. B. Lundeberg, Y. Gao, R. Asgari, C. Tan, B. V. Duppen, M. Autore, P. Alonso-González, A. Woessner, K. Watanabe, T. Taniguchi, et al., Science **89**, 035004 (2017).
[44] D. Alcaraz Iranzo, S. Nanot, E. J. C. Dias, I. Epstein, C. Peng, D. K. Efetov, M. B. Lundeberg, R. Parret, J. Osmond, J.-Y. Hong, et al., Science **360**, 291 (2018).
[45] L. Ju, B. Geng, J. Horng, C. Girit, M. Martin, Z. Hao, H. A. Bechtel, X. Liang, A. Zettl, Y. R. Shen, et al., Nat. Nanotech. **6**, 630 (2011).
[46] A. N. Grigorenko, M. Polini, and K. S. Novoselov, Nat. Photon. **6**, 749 (2012).
[47] D. N. Basov, M. M. Fogler, and F. J. García de Abajo, Science **354**, aag1992 (2016).
[48] Y. Liu, R. F. Willis, K. V. Emtsev, and T. Seyller, Phys. Rev. B **78**, 201403 (2008).
[49] Y. Liu and R. F. Willis, Phys. Rev. B **81**, 081406(R) (2010).
[50] U. Kreibig and M. Vollmer, *Optical Properties of Metal Clusters* (Springer-Verlag, Berlin, 1995).
[51] J. A. Scholl, A. L. Koh, and J. A. Dionne, Nature **483**, 421 (2012).
[52] F. J. García de Abajo, Rev. Mod. Phys. **82**, 209 (2010).
[53] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products* (Academic Press, London, 2007).
[54] L. Hedin and S. Lundqvist, in *Solid State Physics*, edited by D. T. Frederick Seitz and H. Ehrenreich (Academic Press, 1970), vol. 23 of *Solid State Physics*, pp. 1–181.
[55] O. Keller, Phys. Rev. B **33**, 990 (1986).
[56] L. Marušić and M. Šunjić, Phys. Scripta **63**, 336 (2001).
[57] N. D. Mermin, Phys. Rev. B **1**, 2362 (1970).
[58] F. Reinert, G. Nicolay, S. Schmidt, D. Ehm, and S. Hüfner, Phys. Rev. B **63**, 115415 (2001).
[59] V. M. Silkin, J. M. Pitarke, E. V. Chulkov, and P. M. Echenique, Phys. Rev. B **72**, 115435 (2005).
[60] R. Paniago, R. Matzdorf, G. Meister, and A. Goldmann, Surf. Sci. **336**, 113 (1995).
[61] A. Garcia-Lekue, J. M. Pitarke, E. V. Chulkov, A. Liebsch, and P. M. Echenique, Phys. Rev. B **68**, 045103

(2003).
[62] E. Chulkov, V. Silkin, and P. Echenique, Surf. Sci. **437**, 330 (1999).
[63] E. J. H. Skjølstrup, T. Søndergaard, and T. G. Pedersen, Phys. Rev. B **99**, 155427 (2019).
[64] D. Popović, F. Reinert, S. Hüfner, V. G. Grigoryan, M. Springborg, H. Cercellier, Y. Fagot-Revurat, B. Kierren, and D. Malterre, Phys. Rev. B **72**, 045419 (2005).
[65] M. Kerker, *The scattering of light and other electromagnetic radiation: physical chemistry: a series of monographs*, vol. 16 (Academic Press, 2013).
[66] P. B. Johnson and R. W. Christy, Phys. Rev. B **6**, 4370 (1972).
[67] E. Kröger, Z. Phys. **216**, 115 (1968).
[68] F. J. García de Abajo, A. Rivacoba, N. Zabala, and N. Yamamoto, Phys. Rev. B **69**, 155420 (2004).
[69] F. J. García de Abajo, ACS Nano **7**, 11409 (2013).
[70] N. W. Ashcroft and N. D. Mermin, *Solid State Physics* (Harcourt College Publishers, Philadelphia, 1976).
[71] A. Goldman, V. Dose, and G. Borstel, Phys. Rev. B **32**, 1971 (1985).
[72] S. Suto, K.-D. Tsuei, E. W. Plummer, and E. Burstein, Phys. Rev. Lett. **63**, 2590 (1989).
[73] F. J. García de Abajo, ACS Photon. **1**, 135 (2014).
[74] M. A. Mueller and T. M. T.-C. Chiang, Phys. Rev. B **41**, 5214 (1990).
[75] I. Matsuda, T. Tanikawa, S. Hasegawa, H. W. Yeom, K. Tono, and T. Ohta, e-J. Surf. Sci. Nanotechnol. **2**, 169 (2004).
[76] C. David and F. J. García de Abajo, ACS Nano **8**, 9558 (2014).

ISSN (online): 2446-1636
ISBN (online): 978-87-7210-572-7 AALBORG UNIVERSITY PRESS