Digital plasmonics: from concept to microscopy

Gjonaj, B.

Citation for published version (APA):
CHAPTER 1

INTRODUCTION

Controlling short wavelength plasmon waves via wavefront shaping techniques towards a plasmonic microscope is the goal of this thesis. In this chapter we will introduce the basic concepts underlying the work in this PhD thesis. We will introduce (1) Surface Plasmon Polaritons quasi-particles, the state of art in the field of plasmonics, as well as advantages and disadvantages these surface waves have to offer for high confinement of electromagnetic fields. We will also describe (2) wavefront shaping techniques used to achieve flexible and active control of the plasmonic fields. Finally, (3) microscopy techniques, which can benefit in resolution from a combination with plasmonics, will be introduced.
INTRODUCTION

1.1 General Introduction

Since antiquity, controlling light has attracted the attention of the most brilliant minds of the time, from Archimedes [1] to Newton [2]. Even nowadays, controlling light at small length scales is one of the major research directions in physics and beyond [3, 4]. Due to the wave nature, the ability to control light is intrinsically limited by the diffraction limit which introduces a length scale limitation of the order of the wavelength of light. Below this length scale the spatial variations of the light field are not controllable with conventional optics such as mirrors, prisms, and lenses. Light is a powerful investigative tool in many areas of physics, biology, chemistry, medicine and so forth; but limited control of light fields inhibits investigation below the diffraction length scale. Controlling light beyond the diffraction limit is the essence of nanophotonics and an active area of research with large potential for applications.

With the development of Focused Ion Beam and Electron Beam Lithography, samples with features as small as the electron (or ion) wavelength became available. Typically the electron wavelength is orders of magnitude shorter than that of light, thus samples fabricated via electron lithography can have nanometer features: These nanostructured samples have features smaller than the diffraction limit of light. The availability of these nano-shaped samples gave rise to the field of ”nanophotonics”. The main goal of this expanding field is to control light beyond the diffraction limit using nano structured samples.

The field of nanophotonics can be divided in three main research areas very related to each other: (1) photonic crystals, (2) metamaterials, and (3) plasmonics. Spectacular results have been shown for photonic crystals [5, 6] as means of controlling the life times of molecules [7] and quantum dots [8, 9], or as optical-waveguide [10] circuitry with potential for lab-on-chip [11] and optical computation [12]. The field of metamaterials [13] introduced new fundamental concepts, such as perfect lenses [14, 15] using negative refraction [16, 17, 18, 19]. The potentiality of an invisibility cloak [20] using transformation optics [21, 22] and metamaterials [23] is being developed both theoretically and experimentally.

1.1.1 The field of plasmonics

One of the most active fields in nanophotonics is the control of light via Surface Plasmons [24, 25, 26]. Surface plasmons [27] are optically excited
oscillations of the free electrons at the surface of a metal. Because the electromagnetic field confinement achieved with these surface oscillations is tighter than the optical confinement, plasmonics offer control [28] of the electromagnetic radiation beyond the optical diffraction limit. The booming field of plasmonics allows for the excitation and the manipulation of surface plasmons via complex metallic nanostructures [29, 30]. With plasmonics, optical excitations are confined in nanoscale volumes which can be orders of magnitudes smaller than the standard optical confinement [31]. The effective light-matter interaction in these plasmonically confined volumes is strongly enhanced with profound implications in imaging, microscopy, sensing, nonlinear optics and lithography.

High confinement via plasmonics is achieved in two principal approaches: using localized surface plasmons, and using propagating surface plasmons.

Large research in plasmonics is directed towards the study of localized surface plasmons [32] in individual metal nanoparticles and particle assemblies [33]. The plasmonic harvesting of the optical energy within arrays of metal nanoparticles can lead to the formation of nanoscale hot spots [34] in which the intensity of light from an incident beam can be orders of magnitude more concentrated than by diffraction limited optics. This tight optical confinement has lead to a large improvement in techniques that use optical fields, such as surface enhanced Raman spectroscopy [35] where the Raman signal varies as the fourth power of the electric field. This improvement in Raman spectroscopy holds potential for applications in medical diagnostics. Similarly, metallic nanoparticles designed to selectively bond only to tumorous cells offer new hopes for cancer treatment [36]. When sick mice or human bodies infiltrated with these nanoparticles are exposed to light, the light energy will be strongly confined into the nanoparticles. The highly confined energy in the nanoparticles heats up the particles which will than burn the neighboring tumorous cells. Nonlinear processes, such as extreme-ultraviolet light from pulsed-laser high harmonic generation [37], also largely benefit from the field confinement offered by plasmonic nanoparticle arrays. The use of plasmonic nanoparticle arrays as efficient directional antennas for harvesting light, is giving rise to a new generation of thin film solar cells [38].

In the context of this thesis, propagating surface plasmons [39, 40] will be mainly discussed. When light couples to the electron oscillations on the surface of metallic films, propagating polariton waves are excited. These waves are the Surface Plasmon Polaritons (SPP). SPP waves propagating along metallic samples of precisely controlled material and geometry allow
for nanoscale wave guiding properties that cannot be realized with regular dielectrics. To be specific, extremely short wavelengths of the propagating plasmons can be achieved at optical frequencies. For example, it has been shown that for metal-insulator-metal waveguides [41], the wavelength of SPPs excited from visible light was reduced below 100 nm [42, 43, 44]. Thus for a fixed light frequency, the plasmonic wavelength can be up to an order of magnitude shorter than the optical wavelength. Provided a high degree of control of these SPP waves, their short wavelengths hold large potential for biomedical imaging [45], lithography [46] and optical data storage [47]. The short wavelengths of SPPs also enables new miniaturization prospectives in optical computation, for example by using plasmonic waveguides and nanowires for creating optical circuits [30] or the more complex optical transistor [48, 49]. Surface plasmons coupled to single photon sources provide new developing tools for quantum optics [50] and related applications. One of the latest proposed entries in propagating plasmonics is the substitution of the metal by graphene [51].

In the following sections we will provide a formalism for understanding propagating Surface Plasmon Polaritons on metal-dielectric interfaces. These propagating waves will be derived starting from the Maxwell’s equation for a few interesting geometries, like single interface plasmons or double interface plasmons.

1.2 Single interface Surface Plasmon Polaritons

In this section we introduce surface plasmon polaritons propagating in a single metal-dielectric interface. The SPP wave propagates along the interface and its amplitude decays exponentially with the distance form the interface. We calculate here the propagation and decay constants of these waves for gold and silver films on two typical dielectric materials (air and glass).

1.2.1 Electromagnetic waves from Maxwell’s equations

Surface plasmon polaritons are electromagnetic waves thus we will start with describing electromagnetic waves in the formalism of Maxwell’s equations
and in absence of sources and currents.

\[
\nabla \cdot \mathbf{B} = 0, \quad \text{(1.1a)} \\
\nabla \cdot \mathbf{E} = 0, \quad \text{(1.1b)} \\
\n\nabla \times \mathbf{B} = \varepsilon(\omega)\varepsilon_0\mu_0 \frac{\partial \mathbf{E}}{\partial t}; \quad \varepsilon_0\mu_0 \equiv \frac{1}{c^2}, \quad \text{(1.1c)} \\
\n\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}. \quad \text{(1.1d)}
\]

The speed of light is \(c\) while \(\varepsilon(\omega)\) is the dielectric constant of the material under consideration relative to the vacuum. For the frequencies and structures we are interested in, the relative magnetic permeability is equal to unity and thus has been removed from the equations.

The above Maxwell’s equations are partial differential equations in space and time. We can separate the variables and find the wave equation. The time dependency of the electric and magnetic fields is found to be \(e^{i\omega t}\). We will thus look only to the spatial dependency of the fields

\[
\nabla^2 \mathbf{B}(\mathbf{r}) = \varepsilon(\omega)k_0^2 \mathbf{B}(\mathbf{r}), \quad \text{(1.2a)} \\
\mathbf{E}(\mathbf{r}) = \frac{-c^2}{i\omega\varepsilon(\omega)} \nabla \times \mathbf{B}(\mathbf{r}), \quad \text{(1.2b)}
\]

with: \(k_0 \equiv \omega/c\), \(\text{(1.2c)}\)

and: \(\varepsilon(\omega) \equiv \varepsilon'(\omega) + i\varepsilon''(\omega) \equiv n^2(\omega)\). \(\text{(1.2d)}\)

These equation are in their most general form with no boundary conditions imposed. We have imposed no constrictions on the dielectric constant or the refractive index \(n\). To find the SPP fields we have to impose the structure of the dielectric constant and impose boundary conditions.

1.2.2 Piecewise dielectric function for the single interface

Surface Plasmon Polaritons are Transverse Magnetic (TM) waves at a metal-dielectric interface. In order to describe them we need to add this geometry to the wave equation described by Maxwell’s equations. We divide the three-dimensional space into two regions separated by the plane of the interface, \(z = 0\). For this single interface we consider a piecewise homogeneous dielectric function

\[
\varepsilon(\omega) = \begin{cases} 
\varepsilon_1(\omega), & z \geq 0 \\
\varepsilon_2(\omega), & z < 0.
\end{cases} \quad \text{(1.3)}
\]
INTRODUCTION

We can solve the wave equation separately for the two regions and later impose the continuity conditions for the electric and magnetic fields as well as momentum conservation in both regions.

1.2.3 SPPs as Transverse Magnetic waves

Here we will describe the solution for transverse magnetic (TM) waves which are bound to the interface \( z = 0 \) and exponentially decay with the distance from the interface in both regions. We assume the direction of propagation to be along the \( x \) axis. For TM waves this assumption implies that the magnetic field has a non-zero component only along the \( y \) axis. Because we are looking for waves confined to the plane of the interface, we require this magnetic field to decay exponentially with the distance from the interface in both regions. We can write

\[
B(r) = \begin{cases} 
(0, 1, 0)N_1 \exp (ik_{x,1}x - k_{z,1}z), & z \geq 0 \\
(0, 1, 0)N_2 \exp (ik_{x,2}x + k_{z,2}z), & z < 0.
\end{cases}
\]  

(1.4)

Inserting this expression for the magnetic field into Maxwell’s equation 1.2b yields the electric field

\[
E(r) = \begin{cases} 
(k_{z,1}, 0, ik_{x,1}) - \frac{c^2 N_1}{i\omega \varepsilon_1} \exp (ik_{x,1}x - k_{z,1}z), & z \geq 0 \\
(-k_{z,2}, 0, ik_{x,2}) - \frac{c^2 N_2}{i\omega \varepsilon_2} \exp (ik_{x,2}x + k_{z,2}z), & z < 0.
\end{cases}
\]  

(1.5)

Thus the electric field has two components: \( E_x \) in the direction of the propagation and \( E_z \) in the decaying direction. The first component, \( E_x \), is tangential to the interface and, apart from non-zero constants, we have \( E_x \propto \varepsilon(\omega)^{-1} \partial B_y/\partial z \).

To determine the two normalization constants \((N_1 \text{ and } N_2)\), as well as the k-vectors along \( x \) and \( z \), we use the continuity of the magnetic and electric fields along the interface \((B_y \text{ and } E_x)\). Regarding the normalization constants: the electric and magnetic fields are always determined except for a global normalization constant \( N \). In specific cases it is possible to determine this global normalization constant, for example in experiments where a known light flux is incident (normalizing to the incident flux). In general the normalization problem can be very complicated, especially for continuous spectra [52].
1.2. Single interface Surface Plasmon Polaritons

From the equality of the tangential component of the magnetic field at the interface, \( B_y(z = 0) \), we have

\[
k_{x,1} = k_{x,2} \equiv K_S \equiv k_S + i/2L_S.
\]  

In the above equation we define the complex-valued SPP momentum \( K_S \) as well as its real part and imaginary parts. The real part of the SPP k-vector describes the plasmonic wave propagation and is related to the plasmon wavelength \( \lambda_S \) by \( k_S = 2\pi/\lambda_S \). The imaginary part of the SPP k-vector describes losses in the metal as the wave propagates along the interface. The propagation length along \( z \) is exactly given by \( L_S \).

Similarly to the magnetic field, we require also the equality of the tangential components of the electric field (\( E_x \)) in the two regions as they approach the plane of the interface. We find

\[
\begin{align*}
N_1 &= N_2 \\
-N_1 \frac{k_{z,1}}{\varepsilon_1} &= N_2 \frac{k_{z,2}}{\varepsilon_2}.
\end{align*}
\]  

Or in matrix form

\[
\begin{bmatrix}
1 & -1 \\
\frac{k_{z,1}}{\varepsilon_1} & \frac{k_{z,2}}{\varepsilon_2}
\end{bmatrix}
\begin{bmatrix}
N_1 \\
N_2
\end{bmatrix} =
\begin{bmatrix}
0 \\
0
\end{bmatrix}.
\]  

This system of homogeneous equations has always the trivial solution \( N_i = 0; i = 1, 2 \). Only when the determinant of the matrix is zero the system has a non trivial solution which corresponds to surface waves propagating along the interface

\[
\frac{k_{z,1}}{\varepsilon_1} = -\frac{k_{z,2}}{\varepsilon_2}.
\]  

We interpret this equation based on the possible values of the two dielectric functions. To maintain realistic physical conditions, it is necessary to impose that \( k_{z,i} \) is positive (real part), so as the wave decays exponentially in both regions. This condition is only satisfied in eq. (1.9) when the real part of dielectric constants for the two media have opposite signs. Such condition is satisfied for a metal-dielectric interface.
1.2.4 Dispersion relation for Single interface SPPs

To fully describe the surface plasmon propagation for a metal dielectric interface we have to determine the SPP wave vector $K_S$. We define the region $z > 0$, $z < 0$ as the metal and the dielectric, respectively, and substitute in the equations the index $m$, $d$ for the upper region and the lower region, respectively. Finally we use the momentum conservation in both media to find the SPP k-vector.

\[ \varepsilon_m k_0^2 = K_S^2 - k_{z,m}^2, \]  \hspace{1cm} (1.10a)

\[ \varepsilon_d k_0^2 = K_S^2 - k_{z,d}^2. \]  \hspace{1cm} (1.10b)

The $k$-vector along $y$ was taken to be zero, and the minus sign is due to the evanescent definition of the $k$-vector of $k_z$ ($i^2 = -1$). By combining the above equations for the momentum conservation with eq. (1.9), we find the SPP dispersion relation

\[ K_S(\omega) = k_0(\omega)\sqrt{\frac{\varepsilon_m(\omega)\varepsilon_d(\omega)}{\varepsilon_m(\omega) + \varepsilon_d(\omega)}}. \]  \hspace{1cm} (1.11)

The dielectric constants of gold and silver, two metals that are commonly used in plasmonics, have been tabulated for different photon energies. We have used these tabulated values [53] to calculate the SPP dispersion relation for interfaces between gold and silver as metal combined with air ($n_a = 1$) and glass ($n_g = 1.5$) as dielectric. Results for these these two metal-dielectric interfaces are shown in Fig. 1.1.

Because we are interested in plasmonic microscopy, we also present the SPP dispersion relations as a function of the free space wavelength $\lambda_0 = 2\pi \cdot c/\omega$. These results, as well as the SPP propagation lengths, are shown in Fig. 1.2. From this figure, the beneficial properties of the plasmons for high field confinement can be quantitatively determined as the ratio between the SPP and the light wavelength (in vacuum and in the dielectric). The wavelength of silver-glass SPPs excited from blue light is as small as 200 nm: thus twice as small as the vacuum wavelength of light. Notice that as SPP wavelength decreases, so does their propagation length due to losses in the metal. A propagation length smaller than the SPP wavelength is unpractical for most applications.
1.2. Single interface Surface Plasmon Polaritons

Figure 1.1: SPP dispersion relations. (a) Dispersion relation for Au. (b) Dispersion relation for Ag. The dashed lines are the photon light lines, \( \omega = n \cdot c k \). Refractive indices of Au and Ag taken from reference [53].

Figure 1.2: Alternative presentation of the \( \omega - k \) dispersion relations of Fig.1.1. The SPP wavelength plotted as function of the vacuum wavelength \( \lambda_0 \). (a)-(b) Dispersion relation of Au (a) and Ag (b). (c) and (d) Propagation length of SPPs for the two metals. Material properties for the metals taken from reference [53].
1.2.5 Lack of Transverse Electric SPP waves

We have defined SPPs as transverse magnetic (TM) polarized waves bound to metal-dielectric interfaces. Here we prove the definition correct by demonstrating that no bound transverse electric (TE) waves can be excited. The calculation is very similar to the TM case, but the electric and magnetic field exchanged. The dielectric function in the two regions is given by Eq. (1.3), as for the TM case.

For a TE polarized wave that propagates in the interface, we suppose an in-plane electric field along the $y$ axis. This electric field decays exponentially in both regions as a function of the distance from the interface.

$$E(r) = \begin{cases} 
(0, 1, 0)N_1 \exp(ik_{x,1}x - k_{z,1}z), & z \geq 0 \\
(0, 1, 0)N_2 \exp(ik_{x,2}x + k_{z,2}z), & z < 0 
\end{cases} \tag{1.12}$$

Similarly as for the TM waves, the magnetic field is determined using Maxwell’s equations.

$$B(r) = \begin{cases} 
(k_{z,1}, 0, ik_{x,1}) \frac{cN_1}{i\omega} \exp(ik_{x,1}x - k_{z,1}z), & z \geq 0 \\
(-k_{z,2}, 0, ik_{x,2}) \frac{cN_2}{i\omega} \exp(ik_{x,2}x + k_{z,2}z), & z < 0 
\end{cases} \tag{1.13}$$

The crucial difference between the TM and TE polarized waves is the absence of the dielectric constant in equation eq. 1.13 (compared to eq. 1.5). This absence has major consequences when we apply the continuity condition to the electric and magnetic field and require realistic solutions (exponentially decaying with the distance from the interface).

The continuity of the magnetic field is equivalent to the continuity of $\partial E_y/\partial z$. From the continuity of the tangential components of the electric and magnetic fields, a set of homogeneous equations in the normalization constants ($N_i$) is derived.

$$\begin{bmatrix} 1 & -1 \\ k_{z,1} & k_{z,2} \end{bmatrix} \begin{bmatrix} N_1 \\ N_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{1.14}$$

As for the TM case, apart from the trivial solution of no fields $N_i = 0$, other solutions are possible when the determinant of the above matrix is zero. For the TE case this yields

$$k_{z,1} = -k_{z,2}. \tag{1.15}$$
1.3 Double interface Surface Plasmon Polaritons

By a direct comparison with the TM case of eq. (1.9) we notice that the dielectric functions in the two regions are not part of the TE equation. The TE solution of eq. (1.15) is physically unacceptable because the fields must decay exponentially with the distance from the interface, which dictates that both $k_{z,1}$ and $k_{z,2}$ must be positive; a condition that cannot be satisfied simultaneously with eq. (1.15). Thus the only acceptable solution is the trivial solution $N_i = 0$. Hence no bound TE SPP waves do exist.

1.3 Double interface Surface Plasmon Polaritons

In this section we show that a geometry with two metal-dielectric interfaces can achieve higher field confinement (shorter wavelength) than the single metal-dielectric interface. The dispersion relations are described for a general three layer structure. We will focus our attention on the Metal-Insulator-Metal (MIM) structures. Our starting point is the piecewise dielectric constant of the structure, with $d$ the thickness of the insulator

$$\varepsilon(\omega) = \begin{cases} 
\varepsilon_1(\omega), & z \geq d \\
\varepsilon_2(\omega), & 0 < z < d \\
\varepsilon_3(\omega), & z \leq 0.
\end{cases} \quad (1.16)$$

As for the single interface we are looking for evanescent (TM) solutions to the wave equation propagating along $x$. The magnetic field in the three regions is given by

$$B(r) = (0,1,0)e^{iK_Sx} \begin{cases} 
N_1 \exp[-k_{z,1} \cdot (z - d)], & z \geq d \\
N_2 \exp(-k_{z,2}z) + N_3 \exp[k_{z,2} \cdot (z - d)], & 0 < z < d \\
N_4 \exp(k_{z,3}z), & z \leq 0.
\end{cases} \quad (1.17)$$

The k-vectors are defined as for the single interface and we have assumed the propagation of a SPP wave with k-vector $K_S = k_{x,i}$. We impose the continuity of the tangential components of the magnetic ($H_y$) and electric ($E_x$) fields at both interfaces ($z = 0$ and $z = d$). From the single interface we noticed that the continuity of the tangential electric filed was equivalent to the continuity of $\varepsilon^{-1}\partial H_y(z)/\partial z$. These conditions result in a system of four equations for the four unknown $N_i$’s
The system has a non trivial solution \( N_i \neq 0 \) only if the determinant of the associated matrix is zero. For simplicity we denoted \( e^{-k_z d} \) with \( A \)

\[
\begin{vmatrix}
-1 & A & 1 & 0 \\
k_{z,1} & -k_{z,2} & k_{z,2} & 0 \\
\frac{1}{\varepsilon_1} & \frac{1}{\varepsilon_2} & A & -1 \\
0 & -k_{z,2} & k_{z,2} & \frac{k_{z,3}}{\varepsilon_3} & -k_{z,2} & \varepsilon_2 \end{vmatrix} = 0. \tag{1.19}
\]

By calculating the determinant above we find the condition for the dual interface SPP waves. The value of the SPP k-vector \( K_S \) is included in the expressions for all the other k-vectors due to the momentum conservation \( (k_{z,i}^2 = K_S^2 - \varepsilon_i k_0^2) \) in the three different spatial regions.

\[
(k_{z,1} + k_{z,2}) (k_{z,3} + k_{z,2}) = (k_{z,1} - k_{z,2}) (k_{z,3} - k_{z,2}) e^{-2k_z d}. \tag{1.20}
\]

Eq. 1.20 is the dispersion relation for dual interface SPPs. The value of the SPP k-vector \( K_S \) is included in the expressions for all of the other k-vectors due to the momentum conservation \( (k_{z,i}^2 = K_S^2 - \varepsilon_i k_0^2) \) in the three different spatial regions. In contrast with the single interface case, the dual interface condition of eq. (1.20) cannot be solved analytically. Nevertheless, we immediately notice that for \( d \) tending to infinity the right part of eq. (1.20) becomes zero and we find two solutions given by the single interface dispersion relation. As a result, for large thicknesses there will be two SPP modes, one for each interface. When the thickness decreases these two modes will start to couple to each other and will hybridize into new SPP modes (MIM SPPs) of much shorter wavelength.

### 1.3.1 SPP dispersion relation for a MIM waveguide

We calculate the dispersion relations of surface plasmons propagating in a metal-insulator-metal (MIM) waveguide by solving eq. 1.20 graphically. We
use the same metal in the two periphery regions, and thus we substitute the index \(m\) in both regions. Similarly we use the index \(d\) for the dielectric core in the central region. We rewrite condition (1.20) for the MIM structure

\[
\left( \frac{k_{z,m}}{\varepsilon_m} + \frac{k_{z,d}}{\varepsilon_d} \right) = \pm \left( \frac{k_{z,m}}{\varepsilon_m} - \frac{k_{z,d}}{\varepsilon_d} \right) e^{-k_{z,d}d}.
\]

The \(+(-)\) solution due to the square root represent a symmetric (antisymmetric) magnetic field distribution relative to the plane \(z = d/2\). With a bit of algebra and remembering that \(\tanh(x/2) = \coth(x/2)^{-1} = \frac{e^x-1}{e^x+1}\) we write

\[
\frac{k_{z,m}}{\varepsilon_m} \cdot \frac{k_{z,d}}{\varepsilon_d} = -\tanh \left( -\frac{1}{2}k_{z,d}d \right), \quad (1.22a)
\]

\[
\frac{k_{z,m}}{\varepsilon_m} \cdot \frac{k_{z,d}}{\varepsilon_d} = -\coth \left( -\frac{1}{2}k_{z,d}d \right). \quad (1.22b)
\]

The equations can be solved graphically as a function of the propagating SPP \(k\)-vector \(K_S\). Again, the propagating SPP \(k\)-vector is included implicitly in eq. (1.22a) via the momentum conservation in the various regions, \(\varepsilon_i k_{0i}^2 = K_S^2 - k_{i}^2\), where the index \(i\) indicates either metal \((m)\) or dielectric \((d)\).

For the MIM waveguide plasmons, only eq. (1.22a) has solutions. These solutions are shown in Fig. 1.3 for gold and silver at a double interface with a planar silicon nitrate dielectric core \((n = 2)\). Alternatively, in Fig. 1.4, the same solutions are shown as a function of the SPP and vacuum wavelengths. The MIM waveguide provides shorter wavelengths (higher confinement) of the SPP waves as the dielectric thickness \(d\) decreases.

Compared to the single interface plasmons (infinite thickness data of Fig. 1.4), shorter SPP wavelengths can be achieved. For a fixed photon frequency and for reasonable propagation lengths of the plasmons, the MIM SPP wavelength is up to a factor three shorter than the wavelength of light in the dielectric (given by the light line in the dielectric). Therefore, MIM plasmons with a given dielectric middle layer could resolve microscopic objects with nearly triple the resolution of a solid immersion lens composed of the same dielectric.
Figure 1.3: MIM SPP dispersion relations. (a) Dispersion relation for Au. (b) Dispersion relation for Ag. Dashed lines are the photon light lines in $\text{Si}_3\text{N}_4$. Refractive indexes of Au and Ag taken from reference [53].

Figure 1.4: Alternative presentation of the $\omega - k$ dispersion relations of Fig. 1.3. The SPPs wavelength is plotted as function of the vacuum wavelength $\lambda_0$. (a)-(b) Dispersion relation for a Au-SiN-Au (a) and Ag-SiN-Ag (b). (c) and (d) Propagation length of SPPs. Material properties for the metals taken from reference [53].
1.4 Microscopy techniques benefiting from plasmonics

We have shown in the previous section the potential of plasmonics for high field confinement. The confinement is due to a much shorter SPP wavelength compared to either the free space wavelength or the wavelength in the dielectric. This short wavelength offers high potential for microscopy because the resolution of a microscope is inversely proportional to the wavelength. Here we briefly describe two microscopy techniques, widely used in biology, that could perform better with plasmonic illumination than with light.

1.4.1 Total internal reflection microscopy (TIRM)

For a variety of biological systems it is important to distinguish the surface information from the core. For living cells, for example, most of the interaction with the external environment is happening at the surface. The cell’s dynamics are based on the information it receives from the surface. Total internal reflection microscopy is a widely used technique for surface-only investigation.

Total internal reflection microscopy (TIRM) is based on evanescent waves. These waves have a purely imaginary component of the k-vector, thus they decay exponentially along that direction. The standard method to generate these waves is via total internal reflection. When light impinges on the interface between two dielectrics with an angle above the critical angle, all light is reflected back (total reflection); except for an evanescent tail which decays exponentially with distance from the interface. If a scatterer or a fluorescent molecule is placed in the vicinity of the interface, then the scatterer will be exited only via the evanescent tail of the electric field. When the scatterer to interface distance is much larger than the decay constant of the evanescent wave, no electric field is present in proximity of the scatterer which therefore cannot be excited. Only scatterers very close to the surface will feel the presence of the evanescent field. This fact allows for surface investigation. The resolution along the direction of the evanescent decay (axial resolution) is provided not by the diffraction limit (wavelength dependent), but by the evanescent tail. Only a thin slice of the biological sample within the decay constant is imaged, as shown in Fig. 1.5a. Nevertheless, the lateral resolution along the surface is still determined by the diffraction limit.
INTRODUCTION

Figure 1.5: Principle of total internal reflection microscopy (TIRM). (a) Optical TIRM. Light above the critical angle is totally reflected except for an evanescent wave. This evanescent wave is scattered only by particles next to the surface. The scattered light is collected by the objective. Because of the detection diffraction limit two nearby particles (red circles) are unresolved. (b) Plasmonic TIRM is achieved via focused plasmonic wave excitation. The in-plane resolution is determined by plasmonic diffraction limit of the excitation (shown in blue). Two nearby scatterers are resolved.

1.4.2 Plasmonic contribution to TIRM

Plasmonic waves are surface waves that decay exponentially with the distance from a metal-dielectric (or more complicated multilayer) interfaces. On one hand, plasmonics offers the same evanescent axial resolution for TIRM as evanescent light waves (thus they can achieve surface only imaging). On the other hand, SPPs TIRM offers an additional advantage: an increased lateral resolution due to shorter SPP wavelengths compared to light in the dielectric.

The principle of plasmonic TIRM is shown in Fig. 1.5b where the focused evanescent plasmonic field exciting the scatterer is shown in blue. Compared to optical TIRM (Fig. 1.5a) in which two optically excited scatterers (in red) are indistinguishable because of the optical diffraction limit, plasmonic fields focused below optical diffraction can distinguishably excite each scatterer separately. The resolution is determined by the excitation (plasmonic diffraction limit) and not by the detection optics. Nevertheless, to achieve in-plane resolution enhancement, the plasmonic excitation must be tightly focused by using an SPP lens. The creation of such a plasmonic
lens will be one of the goals of this thesis.

1.4.3 Structured Illumination Microscopy (SIM)

We introduce the principles of structured illumination microscopy (SIM), a technique that achieves a resolution two times better than that given by the optical diffraction limit [54, 55]. The core ingredient of structured illumination microscopy is the structured illumination: instead of a wide field illumination a sinusoidal pattern is used. In k-vector space, this sinusoidal pattern will carry additional information. Let the sample spatial distribution be \( s(r) \), with \( r \) a two-dimensional vector on the surface of the sample. In SIM such a sample is illuminated with a periodic intensity pattern \( f(r) \) along \( k_f \) created by two counter propagating waves, of which the relative phase \( \phi \) determines the positions of the maxima of intensity

\[
f(r) \equiv 1 + \cos(k_f \cdot r + \phi) = 1 + \frac{1}{2} \left[ e^{i(k_f \cdot r + \phi)} + e^{-i(k_f \cdot r + \phi)} \right]. \tag{1.23}
\]

When an image of the sample is taken with a lens of which point-spread function (psf) is \( h(r) \), the measured spatial intensity \( I(r) \) and its Fourier transform \( \tilde{I}(k) \) are

\[
I(r) = [f(r)s(r)] \otimes h(r),
\]

\[
\tilde{I}(k) = F \left[ f(r)s(r) \right] \cdot \tilde{h}(k) = \tilde{s}(k) + \frac{e^{-i\phi}}{2} \tilde{s}(k+k_f) + \frac{e^{i\phi}}{2} \tilde{s}(k-k_f) \cdot \tilde{h}(k). \tag{1.24c}
\]

The Fourier transform \( F \) has been calculated using eq. (1.23) and by noting with \( \tilde{s} \) and \( \tilde{h} \) the transforms of \( s \) and \( h \) respectively. The psf \( (h) \) is assumed as a gaussian with FWHM equal to the diffraction limit resolution. Thus its Fourier transform \( \tilde{h} \) is also a gaussian with FWHM \( (k_h) \) which value is inversely proportional to the diffraction limited resolution. For simplicity we further assume that \( \tilde{h} \) is a circle with diameter \( k_h \).

In the intensity spectrum of eq. (1.24c) we notice three contributions: a central one and two contributions shifted by \( \pm k_f \). The central one \( [\tilde{s}(k)\tilde{h}(k)] \) is the sample spectra truncated to \( k_h \) because of the limited resolution. The two shifted contributions are the the sample spectrum shifted by \( \pm k_f \). Because of these shifts due to the sinusoidal illumination, part of the sample spectra above the limiting \( k_h \) fall now into the observable detection limit.
Figure 1.6: Effect of the structured illumination in $k$-vector space. The rectangle represents the sample $k$-vectors. The sample separation in three areas is simply illustrative. (a) The small circles are the Fourier transform of the point spread function. Because of the sinusoidal illumination, the Fourier transform of the recorded intensity is the sum of three different components. Together with the wide-field information (vertical stripes), information from two more areas of the sample (diagonal stripes), is mixed in the recorded image. (b) If the three information components recorded in the image are properly separated, they can be used to increase the resolution.

In principle, if one combines the information contained in these three contributions, an imaging resolution can be obtained that is better than the diffraction limit. Nevertheless the three contributions are overlapped with each other in the measured image as shown in Fig. 1.6a. As a result a single image is not sufficient to separate the three contributions: three images are required. The location of the sinusoidal illumination (not the direction) can be changed by varying the phase $\phi$. Taking pictures for three different phases of the structured illumination $\phi_1, \phi_2, \phi_3$ we obtain

$$
\begin{bmatrix}
\tilde{I}_1(k) \\
\tilde{I}_2(k) \\
\tilde{I}_3(k)
\end{bmatrix} = \frac{\tilde{h}(k)}{2} \begin{bmatrix}
2 & e^{-i\phi_1} & e^{i\phi_1} \\
2 & e^{-i\phi_2} & e^{i\phi_2} \\
2 & e^{-i\phi_3} & e^{i\phi_3}
\end{bmatrix} \cdot \begin{bmatrix}
\tilde{s}(k) \\
\tilde{s}(k+k_f) \\
\tilde{s}(k-k_f)
\end{bmatrix}.
$$

By inverting the system and solving it we can separate the three contributions. Once the three contributions are separated, the two shifted contri-
butions have to be shifted back to their original position at $\pm k_f$, respectively. To obtain an isotropically improved resolution, the above procedure is repeated for few other directions, for example by keeping the module of $k_f$ constant but changing its direction. In general 12 images suffice: four directions, with three different phases for each direction. The improvement of the point-spread function in k-vector space is shown in Fig. 1.7a-b. The psf achieved via SIM, shown as the black circle in Fig. 1.7b, is larger than the original psf of Fig. 1.7a. In the real space ($r$) this larger SIM psf yields improved resolution.

At best, the resolution can improve by a factor two: because the structured illumination is created by the same optics, the magnitude of $k_f$ cannot be larger than $k_h/2$. In other words, the momentum of the fringes has to be within the circle given by the diffraction limit. The best configuration is achieved when the structured illumination pattern is created with the maximally allowed resolution by the optics we have that $k_f = k_h/2$. The new point spread function has now a diameter of $2k_h$, as shown in Fig. 1.7b. Thus by back Fourier transforming in real space, the resolution is two times better than the diffraction limit.

1.4.4 Plasmonic contribution to SIM

When the sinusoidal pattern is optically created, each fringe has to be resolved from the others (the minima yields zero intensity) for SIM to work. This criteria implies that the fringe periodicity cannot be made smaller than twice the diffraction limit. Alternatively in momentum space $k_f \leq k_h/2$ which for the limiting case the SIM resolution is twice as better as the original diffraction limit.

When the sinusoidal pattern is created from surface plasmons this limit can be bypassed. The plasmon fringes do not have to be created optically. The SPP fringes are created by two counter-propagating SPP waves on the surface of a metal-dielectric interface. The fringe periodicity is dependent on the SPP wavelength ($\lambda_S = 2\pi/k_S$) and not on the diffraction limit. Thus, for SPP structured illumination the fringe momentum is $k_f \equiv 2k_S \geq k_h/2$. Such fringes cannot be resolved optically but this fact has no relevance for SIM.

Because the fringe momentum of SPP is not limited by the optical diffraction limit but by the plasmonic one, the resolution enhancement of plasmonic SIM can be higher than the enhancement of standard optical SIM. Such enhancement is shown in Fig. 1.7. Compared to the optical diffraction limit.
Figure 1.7: Effect of the structured illumination. The black circles are the point spread functions in momentum space. The larger the circle the better the resolution. (a) Original point spread function of the lens without SIM. (b) Reconstructed point spread function via optical SIM for four directions (the arrows). Twelve images are acquired (three for each direction) to reproduce a high resolution one. The resolution can only be two times better than the diffraction limit because the illumination pattern has to be optically resolved (momentum of the fringes is within the white circle). (c) Reconstructed point spread function via Plasmonic SIM. The illumination fringes are created from counter propagating SPPs. These fringes are not optically resolved. For a continuous point spread function, plasmonic SIM achieves triple better imaging resolution as the diffraction limit.

(Fig. 1.7a), optical SIM achieves double resolution (Fig. 1.7b), while plasmonic SIM achieves triple resolution (Fig. 1.7c). The Fourier Transform of the point spread function has to be a continuous function in momentum space (so as to provide information in real space over continuous length scales). To satisfy this continuity given by \( k_f = k_h \), plasmonic SIM can only achieve triple resolution. Nevertheless, if a two step approach is used (first optical SIM with \( k_f, 1 = \pm k_h /2 \), followed by SPP SIM with \( k_f, 2 = \pm 3k_h /2 \)), then the imaging resolution can be four times better than the diffraction limit and the reconstructed point spread function is continuous.

An alternative way of using plasmonics for SIM microscopy is to use a nanohole grating. SPP waves propagating on the grating will become plasmonic Bloch modes. Because of different Bloch modes of the grating, multiple fringes will be created. Fringes from the counter-propagation of the \( m^{th} \) plasmonic Bloch mode have momentum \( k_f = 2(k_S \pm mk_g) \). The principal grating momentum, \( k_g \), is inversely proportional to the hole periodicity of the array.

The consequences of these multiple fringe illumination for SIM microscopy
are easily understandable. Related to SIM microscopy, using multiple fringes is identical to using multiple values of $k_f$ in parallel and not limited by the diffraction limit. In fact these values cover the entire inverse lattice of the grating. A similar technique has already been applied to optical SIM [56] by using nonlinearities: multiple $k_f$ are created by the principal harmonic, the second one, and so on. Using plasmonic Bloch modes involves no nonlinearities.

1.5 Wavefront Shaping

We have seen in the previous section the potential of plasmonics for high resolution microscopy. To develop this potential a high degree of control of the propagating SPP waves is an absolute requirement. For example, to take advantage of the short wavelengths of these surface waves, we should be able to focus and redirect them at will and thus to create a plasmonic lens.

In this section we introduce an optical method that achieves high control over the optical wavefronts with the prospective of a possible extension of such over plasmonic wavefronts.

1.5.1 Principle of Wavefront Shaping (WS)

Wavefront Shaping was originally developed for multiple scattering media like: paper, paint, milk, sugar, biological tissue or the human brain. Light passing through this kind of media is scattered multiple times and the phase information is typically lost: we cannot see through paper. An even clearer example is glass: we can see through a glass window because the phase information is preserved, but we cannot see through glass powder because of the scattering from all the grains. Wavefront Shaping restores the phase information or at least part of it.

The principle of Wavefront Shaping is shown in Fig. 1.8. A plane wave from a laser source is sent to the multiply scattering sample. Such a sample can be described as an ensemble of optical channels or modes, with each channel introducing a specific path length. The spatial-intensity distribution in transmission is the interference pattern produced by the sum of all these channels. Because of the large distribution of path lengths, the phases of light transmitted via these channels are completely uncorrelated resulting in an interference pattern which is completely random. Such random interference pattern is the well-known laser speckle, as shown in Fig. 1.8a. The crucial idea behind Wavefront Shaping is to control the phase of sample channels to
Figure 1.8: Principle of Wavefront Shaping. (a) A plane wave incident on the multiply scattering sample is fully distorted giving rise to a speckle pattern in transmission. (b) When the incident wavefront is phase structured it becomes possible to spatially focus the transmitted light. The interference of all light channels at that particular point is constructive due to the structured wavefront. Picture taken from reference [57]

change the speckle pattern. By shaping the phase of the incident wavefront it is possible to associate a different incident phase to each channel. When these incident phases compensate for the path length of the channel, then light from all channels will interfere constructively. This concept is shown in Fig. 1.8b. Due to the shaped wavefront, the constructive interference of a large number of channels at a given spot creates a focus in transmission.

An intuitive illustration of Wavefront Shaping is shown in Fig. 1.9. Before Wavefront Shaping (incident plane wave) the resultant electric field at a given spot is the sum of the electric fields from all the channels. These electric fields have random phases (directions of the arrows) thus the resultant electric field is also random. When Wavefront Shaping is applied, the electric fields from all channels have the same phase (aligned arrows) and the resultant electric field is the maximal.

1.5.2 Finding the right wavefront

Since its introduction in 2007, there have been many extensions to Wavefront Shaping. The major ones are focusing light inside random media [58], projecting a predesigned image through random media [59], focusing light
1.5. Wavefront Shaping

**Figure 1.9:** Wavefront Shaping as constructive interference. For a non optimized wavefront, the electric fields (arrows) of different channels have random phases. The resultant (sum) electric field for an unstructured wavefront is also randomly oriented and has random amplitude. When the wavefront is optimized, the electric fields from different channels have the same phase. The amplitude of the resultant electric field is maximal for the optimized wavefront.

Pulses in space and time through random media [60, 61, 62], and use Wavefront Shaping for high resolution imaging [63, 64, 65]. The main problem that is solved in these interesting cases is: given a desired effect, how to determine the required wavefront?

The wavefront is phase and/or amplitude shaped using a pixelated spatial light modulator (SLM). The phase of each pixel can be independently controlled between 0 and $2\pi$. In experiments typically 1000 to 10000 independent pixels are used, while the number of sample channels has to be larger to avoid redundancy. Finding the required wavefront for focusing (for example) among all the possible phase combinations of all pixels is certainly not a trivial problem. The problem of finding the fastest and most efficient algorithm to determine the wavefront is an active research topic. Various approaches have been proposed and demonstrated experimentally. We will introduce two widely used methods to determine the wavefront: (1) using a feedback approach, and (2) using a time-reversal approach. In most systems (samples) of interest these two approaches yield equivalent solutions.
1.5.3 Finding the wavefront via feedback loops

In the first investigations on Wavefront Shaping, the required phase front was determined via feedback loops. In this approach the sample is considered a black box with a large number of inputs and outputs. In transmission, among all possible outputs only one is chosen, for example creating a very bright spot (a focus). Such an output is very different from the normal speckle pattern because it shows a dramatic intensity increase at one point, the focus. To achieve the focused spot, the algorithm used optimizes the input (the incident wavefront) stepwise till achieving the desired output (focusing at a target spot). The input is optimized using as feedback parameter the intensity at the target. The feedback criteria is to maximize the intensity at the target.

Each input channel (pixel) is optimized independently from the others. The optimization principle is shown in Fig. 1.10. To determine the optimal phase of a pixel, the target intensity (the feedback parameter) is recorded as a function of the pixel’s phase. This variation is provided by the interference with the constant background electric field. This background field is created by the light from all the other pixels, the phases of which is held fixed during the optimization step. A sinusoidal intensity at the target is observed as a function of the pixel phase. The phase for which the intensity is maximal is the optimal phase of the pixel, while the amplitude of the variation is related to the amplitude of the pixel’s field.

This process is repeated for all pixels. When the phases of all pixels are determined the are all loaded together to create the desired focus. The full

Figure 1.10: Finding the optimal phases for two different SLM pixels. The intensity at the target as a function of the single pixel phase is a sinusoidal wave. The optimal phase of the pixel is the one that corresponds to the maxima of intensity. For a different pixel, an different interference pattern is observed, and thus a different optimal phase.
cycle of the algorithm is repeated starting from the created focus till it converges (the focus does not get any brighter). The convergence is guaranteed by the linearity of the algorithm.

Variations to this algorithm and geometry have been shown. For example one may focus at many target points at the same time. Focusing at multiple targets is the first step towards creating an image in transmission. In other experiments, the same algorithm was used to focus inside scattering media using fluorescent particles as target points. These last experiments have implications for deep tissue imaging and medical laser therapy.

1.5.4 Finding the wavefront via optical Time Reversal

Another very efficient way of determining the required wavefront for focusing and beyond is the application of time reversal. This technique was originally developed for ultrasound waves, but was rapidly generalized to other waves like microwaves, seismic waves and finally light waves. The principle of Time Reversal (TR) is simple: the sample is described as matrix $T$ with the element $T_{ji}$ describing the complex-valued transfer between input channel $i$ and output channel $j$. If this matrix is known, then any required wavefront for any desired output is fully determined by the time reversed matrix $T^{-1}$.

Let $V$ be the vectorial space of all possible inputs and $W$ be the vectorial space of all possible outputs. For simplicity we assume that both vectorial spaces have the same dimension $N$. The $V$ space comprises all the possible linear combinations of an orthonormal set $\{v_i\}$, for example the independent pixels of the SLM. Similarly, $W$ is formed by all the possible linear combinations of an orthonormal set $\{w_j\}$, for example the pixels of a CCD detector after the sample. We write

$$V \equiv \left\{ v = \sum_{i=1:N} a_i v_i \mid a_i \in \mathbb{C} \right\},$$

(1.26a)

$$W \equiv \left\{ w = \sum_j b_j w_j \mid b_j \in \mathbb{C} \right\},$$

(1.26b)

$$w = T \cdot v.$$  

(1.26c)

The transfer function of the sample, $T$, has to be determined experimentally. It is sufficient to determine it only for basis vectors, $v_i$. As for feedback algorithms, to determine the element $T_{ji}$ an interferometric measurement is needed. The transmitted electric field from the single SLM pixel ($v = v_i$)
interferes with a constant background field. The interference pattern at the CCD fully determines the amplitude and phase of the transmitted electric field from this SLM pixel to all the CCD pixels.

Once the matrix is determined, it is in principle straightforward to find the required incident wavefront $v_d$ for any desired output $w_d$. In fact

\[ w_d = T v_d \Rightarrow v_d = T^{-1} w_d. \]  

The last step involves the inversion of the sample matrix to determine the desired wavefront. Matrix inversion is an exponential algorithm and thus the inversion of a 10000 by 10000 matrix is not realistic. Fortunately, when the sample is not absorbing, the matrix is unitarian and the matrix inversion becomes much simpler. The inverse of an unitarian matrix is its conjugated transpose: $T^{-1}_{ji} = T^{*}_{ij}$. Such "flipping" of the matrix elements can be quickly processed. For these unitarian cases Time Reversal experiments have elegantly demonstrated the ability to create a variety of outputs, like focusing or even the creation of a programmed image.

1.5.5 Comparison between feedback algorithms and Time Reversal

We have previously described two main ways to retrieve the derided optical wavefront: feedback Wavefront Shaping and Time Reversal mirror. Now we compare advantages and disadvantages of these two techniques. Wavefront Shaping and optical Time Reversal yield the same results for the largest majorities of the cases, so these algorithms share more similarities than differences. The few differences are the convergence speed, signal quality and range of outputs.

Time Reversal is a full matrix approach while Wavefront Shaping is a projective approach. With TR the sample matrix has to be measured only once and the desired wavefront for any given output is than quickly calculated from the sample. With Wavefront Shaping algorithms, for any desired output, the retrieval of the wavefront is a new experimental measurement. Thus, with TR, switching between different outputs is faster than with WS. Nevertheless, because Wavefront Shaping algorithms are based on output optimization, the output quality is higher. Applying the feedback algorithm twice, generally improves the quality of the desired output by up to a factor three. For example, theory predicts that Wavefront Shaping of 256 independent segments yields a maximal focus enhancement of 201 times: optimization algorithms [57] report a focus enhancement of nearly 150 times, while
1.5. Wavefront Shaping

Time Reversal algorithms [66] report an enhancement of 54 times. Whenever choosing one algorithm or the other depends on the desired application. For fixed samples (no internal motion) and with no absorption centers, TR might be preferred. For dynamic samples or samples with absorbing centers, WS feedback algorithms perform better.

1.5.6 Wavefront Shaping of plasmonic channels

Wavefront Shaping tunes the constructive interference of different sample channels. For multiply scattering samples, these channels are typically transmission or reflection channels. For metallic nanostructures, these channels could very well be plasmonic channels. Thus using Wavefront Shaping it might be possible to control plasmonic channels at will, for example to create a plasmonic lens. The shape of the lens is deformable because it’s shape is determined by the incident optical wavefront. The creation and the application of this deformable plasmonic lens for microscopy will be the central theme of this thesis work.