We review an on-chip approach for spontaneous Raman spectroscopy and surface-enhanced Raman spectroscopy based on evanescent excitation of the analyte as well as evanescent collection of the Raman signal using complementary metal oxide semiconductor (CMOS)-compatible single mode waveguides. The signal is either directly collected from the analyte molecules or via plasmonic nanoantennas integrated on top of the waveguides. Flexibility in the design of the geometry of the waveguide, and/or the geometry of the antennas, enables optimization of the collection efficiency. Furthermore, the sensor can be integrated with additional functionality (sources, detectors, spectrometers) on the same chip. In this paper, the basic theoretical concepts are introduced to identify the key design parameters, and some proof-of-concept experimental results are reviewed.

1. Introduction

Raman scattering is a process where incident photons are scattered inelastically by vibrations characteristic to the scattering molecules [1–3]. The frequency of the scattered photon is shifted by an amount corresponding to vibrational energies of the molecules, and the intensity of the scattered light is proportional to the number of scattering molecules. Hence, the Raman spectrum constitutes a specific pattern that allows one to identify and quantify the number of molecules. Thanks to its non-invasive and label-free nature, Raman spectroscopy has been established as a powerful spectroscopic technique for basic research and a myriad of applications ranging from biology, chemistry to material sciences [2,3]. Scientists have applied Raman spectroscopy techniques for a wide variety of remarkable applications such as the study of viruses [4], the classification of tumour cells [5] and a single molecule detection [6].

Despite the enormous progress made over the past decades in demonstrating and identifying several applications, Raman spectroscopy has not yet found a widespread use as a point-of-need tool outside of the dedicated laboratories. The major drawback of Raman spectroscopy is the extremely small cross section of the spontaneous Raman scattering process. It is of the order of $10^{-30}$ cm$^2$, which is 15 orders of magnitude smaller compared with fluorescence [1]. The extremely weak nature of the Raman scattering process necessitates advanced techniques to enhance and detect the signal. Coherent Raman scattering [7], stimulated Raman scattering and surface-enhanced Raman scattering (SERS) are only a few examples of techniques devised to enhance the Raman signal. However, in conventional set-ups, these techniques still require a confocal microscope combined with advanced laser sources, detectors and a large monochromator, hence, limiting the use of the Raman spectroscopic techniques to laboratory environments.

The development of optoelectronic and photonic technologies such as diode lasers and charge-coupled device (CCD) detectors has enormously advanced the use of Raman spectroscopy for numerous applications [3]. Recent advances...
in integrated photonics technology have opened up unique ways towards low-cost and compact point-of-need Raman analysis tools by integrating the essential components of the spectroscopic system on a chip [10,11]. As for many other optical systems, a Raman spectroscopy system may be miniaturized and integrated on a chip. Indeed, all the essential components such as lasers [12], detectors, planar concave gratings [13], arrayed waveguide gratings [11,14], Mach–Zehnder interferometer filters [15,16], ring cavity filters [17] and Bragg gratings [18] have been demonstrated using chip-scale nanophotonics. However, effective transduction methods to generate Raman signals in an on-chip integrated Raman spectroscopic system have remained largely unexplored. A large enhancement is needed to eliminate the necessity of cooled detectors and expensive filters which are typically used in Raman spectroscopy. A small étendue is essential to keep the size of the integrated spectrometer as small as possible for a given resolution. These requirements have neither been identified nor addressed in an integrated set-up.

Nevertheless, Kanger et al. [19] have demonstrated the use of multilayer planar waveguides for the evanescent excitation of adsorbed thin layers for Raman spectroscopy. Although this technique is not readily integrable with an optimal on-chip spectrometer, and still requires a conventional microscope to collect the signal, this work took the idea of confining light and extending the interaction volume to boost the Raman signal using waveguides proposed by Rabolt et al. [20]. We refer elsewhere [21] for a review of spectroscopic techniques based on planar waveguides. Photonic crystal fibres filled with gaseous analyte [22] and hollow core fibres filled with liquid analyte [23] are key examples of the techniques that recognize waveguiding as a mechanism to enhance Raman signal, but fail to be directly integrable on a chip.

We have recently shown that single mode waveguides can be used to evanescently excite and collect Raman signals and have intrinsic assets to implement Raman spectroscopy in a laboratory-on-a-chip framework [24]. Waveguides combine a large detection volume with the field enhancement near a high index of adsorbed thin layers for Raman spectroscopy. Although this technique is not readily integrable with an optimal on-chip spectrometer, it still requires a conventional microscope to collect the signal, this work took the idea of confining light and extending the interaction volume to boost the Raman signal using waveguides proposed by Rabolt et al. [20]. We refer elsewhere [21] for a review of spectroscopic techniques based on planar waveguides. Photonic crystal fibres filled with gaseous analyte [22] and hollow core fibres filled with liquid analyte [23] are key examples of the techniques that recognize waveguiding as a mechanism to enhance Raman signal, but fail to be directly integrable on a chip.

Figure 1. (a) Schematic of the beam geometry for a confocal microscope. Only signal from confocal volume contributes significantly to the signal. (b) Schematic of the waveguide-based evanescent Raman sensors. A long interaction length leads to a very large detection volume.

2. Collection efficiency of diffraction-limited optics

In case of confocal microscopes employing diffraction-limited excitation and collection, the collected signal essentially originates from particles located in the confocal volume, i.e. within the depth-of-focus 2zR, where zR is the Rayleigh range of the beam. Further, only the emission inside a solid angle Ω corresponding to a half-acceptance angle θ (figure 1) of the collection optics is collected. We assume that the collection optics has the same acceptance width (slit width of the spectrometer) as the 1/e² width of the exciting beam, such that the half-acceptance angle is the same as that of the divergence of the input beam. In terms of differential scattering cross section σ and the density of the particles ρ, the overall efficiency of excitation and collection can then be calculated as [28]

\[
P_{\text{eff}} \cdot P_{\text{pump}} = 2zR \Omega \sigma \rho.
\]

Assuming a small beam divergence, \(\theta = w_0/z_R = \lambda_0/(n \pi w_0)\), where \(w_0\) is the beam waist at the focus, \(\lambda_0\) is the pump/Stokes wavelength and \(n\) is the refractive index of the medium.
medium, the solid angle can be expressed in terms of the parameters of the Gaussian beam

$$\Omega = 4 \pi \sin^2 \left( \frac{\theta}{2} \right) \approx \pi \left( \frac{w_0}{2R} \right)^2 = \frac{\lambda_0}{n R}. \quad (2.2)$$

Hence,

$$P_{\text{pump}} = 2 \left( \frac{\lambda_0}{n} \right) \Omega. \quad (2.3)$$

Note that the SI unit for solid angle steradian (sr) is implicit in the prefactor of equation (2.3). The approximation carried out in equation (2.2) overestimates the solid angle, as in general, $w_0/R > \sin (\theta)$. Hence, equation (2.3) sets the upper limit for the confocal microscopic systems. Microscopes with higher transmission losses and having larger divergence than diffraction-limited systems behave worse. Equation (2.3) is independent of any system-specific parameters and reveals the interplay between the beam divergence and the depth of focus, which restricts the detection volume and field intensity, hence, the collected power [29]. In the following sections, we describe the use of dielectric waveguides to overcome the limit imposed by diffraction by means of waveguiding and electromagnetic confinement. Section 3 briefly introduces dielectric waveguides and the relevant technology platform for the fabrication of such waveguides.

3. Silicon nitride-integrated waveguides for Raman spectroscopy

Dielectric waveguides constitute a high refractive index ($n_{\text{core}}$) material called core surrounded by, one or more, lower refractive index materials called cladding (with lowest refractive index $n_{\text{clad}}$), so that the light is guided in the core by total internal reflection [30]. Optical fibres are a well-known example of dielectric waveguides with a very low index contrast $\Delta n$, defined as $\Delta n = n_{\text{core}} - n_{\text{clad}}$. In the realm of integrated photonics, however, high index silicon (Si) core (refractive index $n_{\text{Si}} = 3.45$) material is defined on top of a silica ($n_{\text{SiO}}$) undercladding (refractive index $n_{\text{ox}} = 1.45$), by selectively etching the unwanted Si region [31,32]. This platform for defining Si waveguides on SiO$_2$ is popularly called the silicon-on-insulator (SOI) technology platform and it is the epitome of high index contrast waveguide integration platforms in a CMOS-compatible fabrication technology. High index contrast waveguide is the enhancement of the field intensity in the vicinity of the waveguides, thus increasing the light–matter interaction which we shall explore in §4. To understand the enhancement effects inherent to the waveguides and the role of refractive index contrast, we investigate a common planar waveguide with a silica undercladding, an air uppercladding ($n_i = 1$) and variable core index $n_c$ and core thickness $d$. An emitting molecule near a waveguide can be modelled as a dipole oscillating in a certain direction at a given frequency [36,37]. The total field density at the location of the molecule, which determines the emission, can be calculated by expanding the dipole

shorter wavelengths, remains manageable. A near-infrared wavelength of 785 nm is a popular choice in Raman spectroscopy for biological applications because of the low water absorption, low fluorescence originating from biological molecules, but still fairly short wavelength for achieving a reasonably high scattering cross section. Availability of high-quality and low-cost sources and detectors in the 700–1000 nm wavelength region is also an important drive for the choice of 785 nm wavelength. Unfortunately, the popular SOI platform is not suitable for 785 nm wavelength as silicon absorbs heavily for wavelengths less than 1 $\mu$m, so an alternative integration platform is required.

Silicon nitride ($\text{Si}_3\text{N}_4$) is another common material used in CMOS fabrication technology, the backbone of modern electronics. It has a small absorption in the 500–2500 nm wavelength region, a moderately high refractive index $n_{\text{SiN}} = 1.8–2.2$, exhibits low fluorescence and is a very robust material capable of handling large optical powers [33,34]. Thus, Si$_3$N$_4$ waveguides [35,36] defined on top of a silica undercladding constitute a good trade-off for Raman spectroscopy at visible and near-infrared wavelengths, satisfying several technological constraints.

Si$_3$N$_4$ waveguide circuits discussed in this article (except for the hybrid nanoplasmonic waveguides) are fabricated on 200 mm diameter silicon wafer with 200 mm diameter and thickness of 700 $\mu$m. A stack of 220 nm thick Si$_3$N$_4$ on top of 2.4 $\mu$m thick SiO$_2$ is deposited on top of the Si wafer. The Si$_3$N$_4$ as well as the SiO$_2$ are deposited by plasma-enhanced chemical vapour deposition [35]. The waveguide structures are defined with 195 nm optical lithography and subsequently etched by fluorine-based inductive coupled plasma-reactive ion-etch process to get the final structure. Figure 1b shows the structure of a typical Si$_3$N$_4$ waveguide.

4. Enhancement and collection efficiency with waveguides

In this section, we introduce waveguide enhancement with the help of planar waveguides and show the importance of index contrast for the waveguide enhancement. Next, we calculate the overall evanescent collection efficiency for single mode channel waveguides, which are generally used in the context of integrated photonics.

Unless otherwise noted, the calculations are done for a wavelength $\lambda_0$ of 785 nm in case of Si$_3$N$_4$ and $\lambda_0 = 1550$ nm in case of SOI waveguides. The refractive indices of Si, Si$_3$N$_4$, SiO$_2$ and isopropyl alcohol (IPA; analyte used as uppercladding) are respectively taken to be $n_{\text{Si}} = 3.45$, $n_{\text{SiN}} = 1.89$, $n_{\text{SiO}} = 1.45$ and $n_{\text{IPA}} = 1.37$. The waveguide height $h$ is 220 nm.

4.1. Enhancement of emission near dielectric waveguides

To understand the enhancement effects inherent to the waveguides and the role of refractive index contrast, we investigate a common planar waveguide with a silica undercladding, an air uppercladding ($n_i = 1$) and variable core index $n_c$ and core thickness $d$. An emitting molecule near a waveguide can be modelled as a dipole oscillating in a certain direction at a given frequency [36,37]. The total field density at the location of the molecule, which determines the emission, can be calculated by expanding the dipole
field into plane wave components, and applying the electromagnetic boundary conditions [37,38]. We investigate two common CMOS-compatible core materials, Si ($n_0 = 3.45$) and Si$_3$N$_4$ ($n_0 = 1.89$), for the two orthogonal dipole orientations with respect to the slab waveguide surface. Figure 2a shows the power $P$ coupled to the fundamental quasi transverse electric (TE) and quasi transverse magnetic (TM) modes of a slab waveguide from a dipole located at the surface of the waveguide, normalized to the total free-space emission $P_0$ of the dipole. Depending on the orientation of the dipole and the polarization of the mode, a strong coupling of the dipole emission (even exceeding $P_0$) can be observed. The coupled power scales with the overlap of the dipole field and the modal field at the dipole location and is strongly dependent on the core thickness and core index. The same conclusions hold for the case of excitation from the waveguide modes [25,39]. Hence, in the absence of non-radiative transitions, the overall efficiency of excitation and collection by the waveguide is a quadratic function of the field intensity. These results from slab waveguides underscore the role of waveguide geometry and index contrast for efficient excitation and collection [39]. Interestingly, for both material systems at their respective wavelengths ($\text{Si}_3\text{N}_4$: 785 nm and Si: 1550 nm), the optimal slab thicknesses are near 120 and 240 nm, respectively, for TE and TM excitation.

4.2. Conversion efficiency of channel waveguides

The excitation efficiency of a particle located at a position $r_0$ by a channel waveguide mode, and the subsequent collection by the same mode can be described by a parameter called integrated luminosity $\Lambda_{wg}$ of the waveguide. It is defined as the ratio of collected power ($P_{wg}$) to pump power ($P_{\text{pump}}$) for a unit scattering cross section ($\sigma$)

$$\Lambda_{wg}(r_0) = \frac{P_{wg}(r_0)}{\sigma P_{\text{pump}}}, \quad (4.1)$$

For a random ensemble of incoherently scattering particles with a uniform volume density $\rho$, distributed in a volume $V$, the total scattering efficiency can then be calculated as

$$\frac{P_{\text{tot}}}{P_{\text{pump}}} = \sigma \rho \int_V \Lambda_{wg}(r_0) \, dr. \quad (4.2)$$

In case of arbitrarily shaped channel waveguides, having a translational symmetry in the longitudinal direction transverse to the surface $S$, we define a quantity called specific conversion efficiency $\eta_0$ as

$$\eta_0 = \frac{d}{dz} \int_V \Lambda_{wg}(r_0) \, dr = \int_S \Lambda_{wg}(r_0) \, dr, \quad (4.3)$$

such that the power $P_{\text{tot}}$ collected by a lossless waveguide section of length $l$ is given by

$$\frac{P_{\text{tot}}}{P_{\text{pump}}} = \eta_0 \rho atf. \quad (4.4)$$

Neglecting the Stokes shift, $\Lambda_{wg}$ can be written as [25],

$$\Lambda_{wg}(r_0) = \frac{1}{n(\omega)} \left( \frac{n_g(\omega)\epsilon_0 |\mathbf{e}_a(r_0, \omega)|^2}{\int |\mathbf{e}_a(r, \omega)|^2 \, dr} \right)^2, \quad (4.5)$$

where $n_g$ is the group index of the mode. Thus, the conversion efficiency is given by

$$\eta_0 = \frac{1}{n(\omega)} \int_S \left( \frac{n_g(\omega)\epsilon_0 |\mathbf{e}_a(r_0, \omega)|^2}{\int |\mathbf{e}_a(r, \omega)|^2 \, dr} \right)^2. \quad (4.6)$$

The conversion efficiency $\eta_0$ evidently depends on the dielectric function that defines the waveguide field distribution and can be calculated using standard mode solvers. The conversion efficiency $\eta_0$ has the unit of solid angle (sr), reminiscent of the collection efficiency of a diffraction-limited optical microscope described by equation (2.1).

In Figure 2b, we show the theoretical conversion efficiency $\eta_0$ calculated for Si$_3$N$_4$ and SOI strip waveguides for normalized waveguide widths $k_0 w$ and fixed height $h = 220$ nm corresponding to single mode operation of the waveguides. As the waveguide width is increased, $\eta_0$ increases to a maximum corresponding to an optimal mode confinement in the core and the cladding. Once this condition is reached, further increase in the width leads to a decrease in $\eta_0$ owing to more
confinement of the modal field in the core. This reduces the interacting field in the uppercladding region of the analyte. The discontinuity of the electric field at the core–top cladding interface of the TM modes leads to a higher evanescent field at the position of the analyte. Thus, higher index contrast waveguides with TM polarization generally perform better than TE-polarized modes. The optimal widths for Si₃N₄ waveguides at 785 nm are near 260 nm and 300 nm for TE and TM excitation and collection respectively, whereas for SOI waveguides, it is near 150 and 240 nm, respectively.

4.3. Slotted channel waveguides

As depicted in figure 3a, slotted waveguides have a narrow slot of width s in between the rails of the core material of width r. Owing to the boundary conditions of the electromagnetic field, the existence of a narrow slot leads to a huge enhancement of the field in the slotted region, and hence also to the value of η₀. For a fixed rail width r = 275 nm and thickness h = 220 nm, figure 3a shows η₀ calculated for Si₃N₄ and Si slot waveguides for variable slot widths using equation (4.6). A significant increase in η₀ is observed as s is reduced for TE modes owing to the enhancement of the TE-field in the slot region as s is decreased. In case of SOI waveguides, where the index contrast is higher and much higher field enhancement can be expected, the value of η₀ reaches up to 40 sr for s = 20 nm.

Figure 3b shows η₀ calculated for Si₃N₄ and SOI for a fixed slot width s = 150 nm and for various total widths w, inclusive of the slot width s (note: r = (w − s)/2). As observed in the case of strip waveguides, η₀ decreases when the waveguide width is increased as a result of the confinement of the mode in the core and consequent reduction of the interaction volume in the analyte region. Owing to the continuity of the electric field, the field enhancement in the slot is much less in the case of TM polarization, leading to relatively insignificant enhancement in case of TM polarization.

4.4. Equivalent length and enhancement

Comparing equations (4.4) and (2.3), we can define an equivalent length $l_{eq}$ which corresponds to the length of the waveguide that gives the same signal as that of diffraction-limited confocal microscopes

$$l_{eq} = \frac{2}{\eta_0} \left( \frac{\lambda_0}{n} \right).$$

Depending on the index contrast, polarization and design of the waveguides, η₀ may typically vary from 0.1 to 50 sr. Thus, a waveguide length less than 20 $\lambda_0$ wavelengths is enough to provide an efficiency equivalent to that of a diffraction-limited confocal microscope. A single mode waveguide loss of 1 dB cm⁻¹ is typical for the considered integrated photonics platform, allowing more than a centimetre of waveguide without significant loss of the pump or collected signal. This leads to three to five orders of magnitude enhancement of the signal compared with diffraction-limited systems.

4.5. Lossy waveguides

So far, we have neglected the waveguide losses. Waveguide losses can be incorporated in our model by defining a function $\xi(l)$ such that (24)

$$\xi(l) = \frac{P_{col}}{P_{tx}} = \frac{\eta_0 \rho \sigma}{2} \left( e^{\Delta \alpha l} - 1 \right),$$

where $P_{tx}$ is the transmitted pump power at the detector position and $\Delta \alpha$ is the difference in waveguide losses at the pump and Stokes wavelength.

5. Hybrid nanophotonic–plasmonic waveguides

In §4.4, we showed that nanophotonic waveguide-based evanescent Raman spectroscopy provides a promising improvement in the sensitivity to standard confocal microscopy for bulk sensing. We showed that the most significant component of the enhancement is due to an increase in the interaction volume, and only a small component comes from the enhancement of the field itself. As a result, it becomes difficult to detect a small number of analyte molecules or nanoparticles in solution. Specifically designed metallic nanostructures allow a considerable enhancement of the electric field near the metal surface [40]. Such structures can hence locally boost the signal in order to allow detection of molecules with an extremely weak Raman cross section or ultra-low concentrations (i.e. SERS). SERS signals are, however, mainly excited and collected by bulky and expensive microscopy systems. The development of nanophotonic circuits functionalized with dedicated nanoplasmic antennas would, hence, present a significant step towards the realization of dense SERS probes, allowing multiplexed detection of extremely weak Raman signals. Up till now, these integration efforts
The quantity $\text{FOM}(N)$ contains all design parameters to assess the SERS signal strength for a given waveguide geometry and is therefore the proper figure of merit in comparing different antenna geometries.

In figure 5, we compare $\eta_A$ and $\text{FOM}(N)$ for four different bowtie antennas with fixed $\alpha = 60^\circ$, $\Delta = 10$ nm, but varying length $L$ (70, 90, 110, 130 nm) and coated with a 1 nm 4-nitrothiophenol (NTP) layer.

For a given pump and Stokes wavelength, $\text{FOM}(N)$ is maximal for an optimum antenna number

$$N_{\text{opt}} = \frac{\log(\log(e_s)/\log(e_p))}{\log(e_s/e_p)}.$$

For the selected Stokes wavelength (877 nm) from an NTP monolayer, a bowtie with length $L = 90$ nm generates the highest SERS signal as the $\text{FOM}(N_{\text{opt}})$ is maximal for this configuration. For this optimal bowtie geometry, a single antenna will produce about 4 fW of Stokes power for each 1 W of guided pump power as can be derived from the middle part of figure 5. Organizing these bowties in an array with $N_{\text{opt}}$ antennas will generate about 30 fW of Stokes power at the output for 1 W guided power. Similar calculations can also be repeated for other geometries. In the experimental section, we show that the absolute Raman power predicted by this analytical model fits very well with the experimentally obtained Raman power.

6. Experiments

In this section, we describe the experimental set-up and the experimental results obtained for the silicon nitride waveguides with and without integrated nanoantennas, and compare with the theoretical models described in §§4 and 5.

6.1. A brief description of the experimental set-up

The details of the set-up have been discussed elsewhere [24,27], but for convenience of the readers, the set-up is reproduced in figure 6, and briefly described below. A tunable Ti : sapphire laser is set to a pump wavelength of $\lambda_p = 785$ nm (red) after which the polarized beam passes through a half-wave plate ($\lambda_p/2$) in order to rotate the polarization to a TE-polarized beam. The pump beam then passes through a laser line filter for side-band suppression before it is coupled into the chip by an aspheric lens (ASPH). The output beam is collected with an objective (OBJ) and passes through a polarizer $P$ (set to TE) before it is filtered by a dichroic mirror into the chip by an aspheric lens (ASPH). The output beam is collected with an objective (OBJ) and passes through a polarizer $P$ (set to TE) before it is filtered by a dichroic mirror which reflects the pump beam and transmits all Stokes wavelengths (green). The Stokes light is collected into a fibre using a parabolic mirror collimator (PMC) after which the fibre is split by a fibre splitter of which 1% goes to a power meter and 99% to a commercial spectrometer from ANDOR (Shamrock 303i spectrometer and iDus 416 cooled CCD camera). The 1% fibre tap is used during alignment and to measure the transmitted power.

6.2. Experiments with waveguides without nanoantennas

The analyte, which is pure IPA here, is drop-cast on the chip containing the waveguides and covers the relevant waveguide region. The Raman signal of the analyte is collected via the same waveguide and measured with a commercial

![Figure 4. Si₃N₄ waveguide functionalized with an array of N bowtie antennas. The top inset depicts the geometrical parameters of a bowtie antenna (gap Δ, length L and apex angle α). The bottom inset shows an SEM image of an integrated bowtie antenna (adapted from [27]).](http://rsfs.royalsocietypublishing.org/Downloaded from http://rsfs.royalsocietypublishing.org/)
length 416, spectrometer and cooled CCD detector from ANDOR; BS, beamsplitter; ¼ (EFL 0.5); S, sample stage; DM, dichroic mirror; PMC, parabolic mirror collimator chosen. Both the waveguides have a height width l of IPA obtained using such a procedure for a strip waveguide guide before application of IPA and the Raman spectrum of waveguides are 1 cm in length and are coiled as spirals (typical footprint: 800 220 nm). The antennas were coated with Si3N4, this guarantees that any Raman signal has an SERS nature. The fundamental TE mode of a Si3N4 not bind to Si3N4, this guarantees that any Raman signal probe molecule to detect SERS events. Because NTP does not bind to Si3N4, this guarantees that any Raman signal has an SERS nature. The fundamental TE mode of a Si3N4 absorbs at 819 cm−1, which is consistent with the red peak in figure 7b. We observe strong peaks that match well with the known Raman spectrum of α-D glucose solution.

One key highlight of figure 7a is that about an eightfold increase in the signal is observed for slotted waveguides compared with the strip waveguides, in accordance with the theory. To verify the theory further, in figure 8, the experimentally determined values of η0 for various waveguide geometries and polarizations are compared with the theoretically values obtained using equation (4.3) and the signal corresponding to the 819 cm−1 vibration of IPA. The theoretical curves of η0 as a function of waveguide width w are found to be within the error margin of the experimental data for both polarizations and the waveguide types studied. These experimental results validate the theoretical model discussed in §4.

The reported experiments and simulations are performed for IPA as the analyte in the uppercladding. However, the experimental and theoretical assessments have also been performed using other liquid and solid analytes, such as, acetone, dimethyl sulfoxide and spin-coated photoresists. The confinement properties of the waveguide mode changes if the refractive index of the uppercladding varies; therefore, according to equation (4.3), the numerical values of the conversion efficiency will slightly differ. Nonetheless, the results discussed in this article are pertinent for analytes of general interest such as a glucose solution.

Next, we discuss experimental results relating to the integration of plasmonic nanoantennas on top of the waveguides.

6.3. Experiments with hybrid nanophotonic–plasmonic waveguides

The nanoantennas were patterned on top of Si3N4 waveguides as described in [27]. The antennas were coated with NTP, which is assumed to form a self-assembled monolayer on the gold through an Au–S bond and which acts as a probe molecule to detect SERS events. Because NTP does not bind to Si3N4, this guarantees that any Raman signal has a SERS nature. The fundamental TE mode of a Si3N4 spectrometer as described in §6.1. For strip waveguides, a width w of 700 nm is chosen, whereas for slot waveguides, a slot width s = 150 nm. We can clearly distinguish all the major peaks of IPA obtained using both waveguide types. As an alternative analyte, we also apply a 1 M glucose solution on top of the slot waveguide (w = 700 nm, s = 150 nm). In figure 7c, the spectra obtained from the glucose solution before and after background correction are shown. We observe strong peaks that match well with the known Raman spectrum of α-D glucose solution.

Figure 7b shows typical Raman spectra of IPA obtained using such a procedure for a strip waveguide (w = 700 nm) and a slotted waveguide (w = 700 nm and s = 150 nm). The red and cyan line mark the respective pump and Stokes wavelength (adapted from [26,27]).

Figure 5. Comparison of the antenna extinction E (dB), η0 and FOM(ν) for four different bowtie antennas with fixed α = 60° and Δ = 10 nm but varying length l (70, 90, 110, 130 nm). The red and cyan line mark the respective pump and Stokes wavelength (adapted from [26,27]).
rib waveguide excites the SERS signal, which is subsequently collected back by the same waveguide mode.

Figure 9 depicts the Raman spectra, before and after coating with NTP, of waveguides functionalized with varying \( N = 10, 20, 30, 40 \) (for a fixed bowtie geometry with \( \alpha = 60^\circ, \Delta = 48 \pm 13 \text{ nm}, L = 106 \pm 8 \text{ nm} \)). The spectrum of a reference waveguide \(( N = 0)\) is also shown. Each Raman spectrum is averaged 10 times. The spectral regions where an NTP Stokes peak is expected \((1080, 1110, 1340, 1575 \text{ cm}^{-1})\) are highlighted by the cyan-shaded areas. The peaks at 1250 and 1518 cm\(^{-1}\) (marked by the black-dashed lines) are already visible before coating, so they are not due to the NTP, but are attributed to interference effects of the Au array [26,27]. After coating, additional peaks, which perfectly coincide with the expected NTP Stokes peaks, appear. While the non-functionalized reference waveguide does not generate any NTP SERS signal, it does generate a relatively large Raman background originating from the Si3N4 waveguide core (with a length of 1 cm). The shot noise associated with this background will limit the detection of the smallest NTP Raman features. This is evidenced by the fact that the smallest 1110 cm\(^{-1}\) peak only appears for the \( N = 40 \) waveguide, which has a reduced background when compared with the \( N = 10 \) case owing to the attenuation caused by the larger number of antennas. However, the absolute signal strength of the \( N = 40 \) peaks...
is lower than the \( N = 10 \) peaks as shown in figure 10 where the signal strength and signal-to-noise ratio (SNR) of the 1340 cm\(^{-1}\) peak is plotted for different \( N \). The solid lines represent a fit of the experimental data to our earlier described on-chip SERS model. The shaded areas represent a distribution of possible signal counts and SNR values that arise owing to fabrication-induced variations among all antennas in the array. Additional details can be found in [27].

Furthermore, one can see that the signal is maximized for \( N = N_{\text{opt}} \), whereas the SNR is maximal for \( N = N^{\text{SNR}}_{\text{opt}} > N_{\text{opt}} \). It is expected that background mitigation through proper chip designs, and not through the antennas itself, can significantly enhance the SNR, because one maintains the possibility of optimizing the signal while reducing the background simultaneously. From figure 10, it moreover follows that a minimum number of antennas \( N_{\text{min}} \) is required to achieve an SNR > 1. Proper background mitigation is also expected to push towards the \( N_{\text{min}} = 1 \) limit such that signals from a single antenna can be detected.

For the data shown in figure 9, we obtained a fitted value of \( \eta_{\lambda} = 2.6 \pm 0.77 \times 10^{-15} \) which was in excellent correspondence with the theoretically predicted value \( \eta_{\lambda} = 2.2 \times 10^{-15} \) as detailed elsewhere [27]. This clearly establishes the validity of our model and its ability to provide quantitative predictions of the absolute Raman power coupled into a single mode waveguide. For this particular case, a single antenna, coated with an NTP monolayer, will hence produce 2.6 \pm 0.77 fW for 1 W of guided power.

7. Comparative performance analysis of waveguides with and without nanoantennas

For the case of evanescent waveguide sensing without nanoantennas, the signal enhancement is mostly due to the large interaction volume distributed over the length of the waveguide. Hence, these sensors are suitable for sensing relatively larger analyte volumes. The enhancement of these sensors cannot be increased arbitrarily and a very high enhancement, needed for, e.g. single molecule detection, is not possible using standard designs.

However, for the waveguides with integrated nanoantennas, the enhancement is the result of a plasmonic resonance effect at discrete ‘hot spots’, i.e. regions of extreme localization of the electric field. Hence, waveguides integrated with nanoantennas are suitable for studying a very small number of localized molecules. We have theoretically shown that for integrated nanoantennas, an enhancement as high as 10\(^{10}\) is possible [46]. Hence, single molecule detection in an integrated framework is possible with such hybrid nanophotonic–plasmonic waveguides.

Depending on the dimensions and the geometry of the antennas and the waveguide, we estimated that each integrated antenna discussed in this paper has an equivalent Si\(_3\)N\(_4\) rib waveguide length of 250–2500 nm. This equivalent length is the physical length of a bare Si\(_3\)N\(_4\) waveguide (without any gold) that is required to generate the same signal as one integrated nanoplasmonic antenna. As a rough rule of thumb, 1 \( \mu \)m of a Si\(_3\)N\(_4\) rib waveguide produces the same Raman signal as 1 plasmonic nanoantenna integrated on top of the waveguide. However, for the nanoantennas, the loss associated with gold absorption, elastic scattering, surface roughness, etc., is about \( \epsilon_{\text{p}} \sim \epsilon_{\text{s}} \sim 0.5 \) dB per antenna [27]. The loss per antenna is quite high compared with the \( 10^{-4} \) dB \( \mu \)m\(^{-1}\) for waveguides without the antenna. For bulk sensing, the integration of nanoplasmonic antennas is hence completely superfluous, because the signal generated by bare waveguides will be much larger owing to the large waveguide length and small waveguide loss. On-chip SERS will, however, outperform the bare waveguide sensing case when particular analytes are adsorbed on the metal surface. In that case, one can make use of the extreme local Raman enhancement of the plasmonic hot spots, which can be orders of magnitude higher than the field enhancement of dielectric rib or slot waveguides.

As discussed in §4, depending on the waveguide geometry, more than three to five orders of magnitude higher signal is expected from a waveguide (bare) compared with diffraction-limited systems. Indeed, for a realistic silicon nitride waveguide (\( h = 220 \) nm, \( s = 150 \) nm, \( w = 700 \) nm, \( l = 1 \) cm), about three orders of magnitude is obtained experimentally [29, 47]. As a consequence of this enhancement, a higher SNR is obtained compared with the detector dark-noise-limited system, despite the background noise from the waveguide core [47]. Still higher performance is expected for a very thin layer of molecules functionalized on top of the chips. It results from the fact that the signal from microscopes is limited by the thickness \( t \) of the layer, so that \( z_0 \) in equation (2.1) is replaced by \( t \). Whereas for waveguide systems, the signal essentially remains the same, as the strongest contribution for \( \eta_\lambda \) in equation (4.4) comes from the molecules near the waveguides in either case. More than four orders of magnitude signal enhancement is expected for a very thin layer using a centimetre of waveguide length.

Our on-chip SERS platform improves on existing hybrid nanophotonic circuits where waveguides are used to probe SERS signals from external metallic nanoparticles, in the sense that it allows a quantitative prediction of the Raman enhancement and coupling with the underlying waveguide. The complete integration moreover allows, by fabrication, a robust and optimized alignment and can potentially improve the coupling efficiency between the excitation beam and the nanoplasmonic antennas when compared with free-space excitation owing to the near-field interaction of the guided mode and the plasmonic mode [48]. Currently, we do not gain in terms of maximum Raman enhancement factor (about 10\(^{10}\)) when compared with existing technologies. It is, however, expected that by decreasing the gap size of the bowtie antennas, the maximum Raman enhancement can be boosted by another two to three orders of magnitude.
8. Conclusion

We reviewed an integrated waveguide approach for on-chip spontaneous and surface-plasmon-enhanced Raman spectroscopy using the evanescent field of the dielectric waveguide for excitation and collection. The effects of the waveguide parameters and polarization of the modes on the overall evanescent excitation and collection efficiency of spontaneous Raman scattering were discussed and validated experimentally. Thanks to the use of low-loss, high index contrast waveguides, more than two orders of magnitude improvement in the overall excitation and collection efficiency is estimated for a centimetre of waveguide when compared with the free-space excitation and collection. Moreover, we developed an analytical model for on-chip SERS to identify the key design parameters and figure of merit for a hybrid nanophotonic–plasmonic platform. The experimentally obtained SERS signals were in excellent correspondence with our theoretical analysis. Depending on the number of analyte molecules to be probed, the integrated waveguide approach provides alternative routes for Raman spectroscopy. We conclude that ‘bare’ waveguides perform better when a large numbers of molecules have to be probed. Integrated nanoantennas are particularly useful when analyte adsorbs to the metal surface. In that case the Raman signal can be boosted by several orders of magnitude owing to the plasmonic enhancement. Moreover, on-chip SERS could be recommended for probing extremely weak Raman scatterers or ultra-low concentrations provided the analyte molecules find their way to the plasmonic hot spots. This enables a comprehensive design freedom and quantitative control on the Raman signal acquisition depending on various applications. In combination with other on-chip devices, such as sources, filters and arrayed waveguide gratings, the presented on-chip Raman platform allows multiplexed detection of extremely weak Raman signals on a highly dense integrated platform.

Competing interests. We declare we have no competing interests.

Funding. ERC Advanced Grant, Inspectra, Fonds Wetenschappelijk Onderzoek (FWO) and Bijzonder Onderzoeksfonds (BOF-UGent).

Acknowledgements. The authors acknowledge imec, Leuven for fabrication of the waveguides.

References

3. Laserna JJ. 1996.


