Research Article | Vol. 26, No. 16 | 6 Aug 2018 | OPTICS EXPRESS 19885

Optics EXPRESS

Surface sensing with integrated optical waveguides: a design guideline

Johannes Milvich,[1,2,3,6] Daria Kohler,[1,4,6] Wolfgang Freude[1] AND Christian Koos[1,5]

[1]*Institute of Photonics and Quantum Electronics (IPQ), Karlsruhe Institute of Technology (KIT), Engesserstr. 5, 76131 Karlsruhe, Germany*
[2]*Robert Bosch GmbH, Robert-Bosch-Campus 1, 71272 Renningen, Germany*
[3]*johannes.milvich@bosch.com*
[4]*daria.kohler@kit.edu*
[5]*christian.koos@kit.edu*
[6]*These authors contributed equally to this work*

Abstract: Waveguide-based biochemical sensors exploit detection of target molecules that bind specifically to a functionalized waveguide surface. For optimum sensitivity, the waveguide should be designed to mediate maximum influence of the surface layer on the effective refractive index of the guided mode. In this paper, we define a surface sensitivity metric which quantifies this impact and which allows to broadly compare different waveguide types and integration platforms. Focusing on silicon nitride and silicon-on-insulator (SOI) as the most common material systems, we systematically analyze and optimize a variety of waveguide types, comprising simple strips, slot and double slot structures, as well as sub-wavelength gratings (SWG). Comparing the highest achievable surface sensitivities, we provide universal design guidelines and physically interpret the observed trends and limitations. Our findings allow to select the appropriate WG platform and to optimize sensitivity for a given measurement task.

OCIS codes: (130.6010) Sensors; (130.0130) Integrated optics; (280.1415) Biological sensing and sensors; (230.7370) Waveguides.

References and links

1. S. Schmidt, J. Flueckiger, W. Wu, S. M. Grist, S. T. Fard, V. Donzella, P. Khumwan, E. R. Thompson, Q. Wang, P. Kulik, J. Kirk, K. C. Cheung, L. Chrostowski, and D. Ratner, "Improving the performance of silicon photonic rings, disks, and bragg gratings for use in label-free biosensing," Proc. SPIE **9166**, 91660M (2014).
2. V. M. N. Passaro, B. Troia, M. La Notte, and F. De Leonardis, "Chemical sensors based on photonic structures" in *Advances Chemical Sensors*, W. Wang, ed. (InTech, 2012).
3. S. Janz, A. Densmore, D.-X. Xu, P. Waldron, J. Lapointe, J. H. Schmid, T. Mischki, G. Lopinski, A. Delâge, R. McKinnon, P. Cheben, and B. Lamontagne, "Silicon Photonic Wire Waveguide Sensors" in *Advanced photonic structrues for biological and chemical detection*, X. Fan, ed. (Springer, 2009).
4. P. Kozma, F. Kehl, E. Ehrentreich-Förster, C. Stamm, and F. F. Bier, "Integrated planar optical waveguide interferometer biosensors: A comparative review," Biosens. Bioelectron. **58**, 287–307 (2014).
5. A. Densmore, D.-X. Xu, S. Janz, P. Waldron, J. Lapointe, T. Mischki, G. Lopinski, A. Delâge, J. Schmid, and P. Cheben, "Sensitive label free Sensitive label-free biomolecular detection using thin silicon waveguides," Adv. Opt. Technol. **2008**, 725967 (2008).
6. L. Chrostowski, S. Grist, J. Flueckiger, W. Shi, X. Wang, E. Ouellet, H. Yun, M. Webb, B. Nie, Z. Liang, K. Cheung, S. Schmidt, D. Ratner, and N. Jaeger, "Silicon photonic resonator sensors and devices," Proc. SPIE **8236**, 823620 (2012).
7. X. Wang, J. Flueckiger, S. Schmidt, S. Grist, S. T. Fard, J. Kirk, M. Doerfler, K. C. Cheung, D. M. Ratner, and L. Chrostowski, "A silicon photonic biosensor using phase-shifted Bragg gratings in slot waveguide," J. Biophotonics **6**(10), 821–828 (2013).
8. J. Flueckiger, S. Schmidt, V. Donzella, A. Sherwali, D. M. Ratner, L. Chrostowski, and K. C. Cheung, "Sub-wavelength grating for enhanced ring resonator biosensor," Opt. Express **24**(14), 15672–15686 (2016).
9. H. Yan, L. Huang, X. Xu, S. Chakravarty, N. Tang, H. Tian, and R. T. Chen, "Unique surface sensing property and enhanced sensitivity in microring resonator biosensors based on subwavelength grating waveguides," Opt. Express **24**(26), 29724–29733 (2016).

#326159 https://doi.org/10.1364/OE.26.019885
Journal © 2018 Received 14 Mar 2018; revised 10 May 2018; accepted 13 May 2018; published 24 Jul 2018

Optics EXPRESS

10. B. Sepúlveda, J. S. Del Rio, M. Moreno, F. J. Blanco, K. Mayora, C. Domínguez, and L. M. Lechuga, "Optical biosensor microsystems based on the integration of highly sensitive Mach-Zehnder interferometer devices," J. Opt. A, Pure Appl. Opt. **8**(7), 561–566 (2006).
11. A. Densmore, D.-X. Xu, P. Waldron, S. Janz, P. Cheben, J. Lapointe, A. Delâge, B. Lamontagne, J. Schmid, and E. Post, "A silicon-on-insulator photonic wire based evanescent field sensor," IEEE Photonics Technol. Lett. **18** (23), 2520–2522 (2006).
12. S. T. Fard, V. Donzella, S. A. Schmidt, J. Flueckiger, S. M. Grist, P. Talebi Fard, Y. Wu, R. J. Bojko, E. Kwok, N. A. Jaeger, D. M. Ratner, and L. Chrostowski, "Performance of ultra-thin SOI-based resonators for sensing applications," Opt. Express **22**(12), 14166–14179 (2014).
13. A. Kargar and C.-Y. Chao, "Design and optimization of waveguide sensitivity in slot microring sensors," J. Opt. Soc. Am. A **28**(4), 596–603 (2011).
14. L. Huang, H. Yan, X. Xu, S. Chakravarty, N. Tang, H. Tian, and R. T. Chen, "Improving the detection limit for on-chip photonic sensors based on subwavelength grating racetrack resonators," Opt. Express **25**(9), 10527–10535 (2017).
15. H. Sun, A. Chen, and L. Dalton, "Enhanced evanescent confinement in multiple-slot waveguides and its application in biochemical sensing," IEEE Photon. J., **A1**(1), 48–57 (2009).
16. I. Khodadad, N. Clarke, M. Khorasaninejad, D. Henneke, and S. S. Saini, "Optimization of multiple-slot waveguides for biochemical sensing," Appl. Opt. **53**(23), 5169–5178 (2014).
17. C. A. Barrios, "Analysis and modeling of a silicon nitride slot-waveguide microring resonator biochemical sensor," Proc. SPIE **7356**, 735605 (2009).
18. T. Claes, W. Bogaerts, and P. Bienstman, "Experimental characterization of a silicon photonic biosensor consisting of two cascaded ring resonators based on the Vernier-effect and introduction of a curve fitting method for an improved detection limit," Opt. Express **18**(22), 22747–22761 (2010).
19. F. Dell'Olio and V. M. N. Passaro, "Optical sensing by optimized silicon slot waveguides," Opt. Express **15**(8), 4977–4993 (2007).
20. J. Gonzalo Wangüemert-Pérez, P. Cheben, A. Ortega-Moñux, C. Alonso-Ramos, D. Pérez-Galacho, R. Halir, I. Molina-Fernández, D.-X. Xu, and J. H. Schmid, "Evanescent field waveguide sensing with subwavelength grating structures in silicon-on-insulator," Opt. Lett. **39**(15), 4442–4445 (2014).
21. A. Dhakal, A. Raza, F. Peyskens, A. Z. Subramanian, S. Clemmen, N. Le Thomas, and R. Baets, "Efficiency of evanescent excitation and collection of spontaneous Raman scattering near high index contrast channel waveguides," Opt. Express **23**(21), 27391–27404 (2015).
22. D.-X. Xu, J. Schmid, G. Reed, G. Mashanovich, D. Thomson, M. Nedeljkovic, X. Chen, D. Van Thourhout, S. Keyvaninia, and S. Selvaraja, "Silicon photonic integration platform-Have we found the sweet spot?" IEEE J. Sel. Topics Quantum Electron., **A20**(4), 8100217 (2014).
23. M. G. Scullion, T. F. Krauss, and A. Di Falco, "Slotted photonic crystal sensors," Sensors (Basel) **13**(3), 3675–3710 (2013).
24. J. Hu and D. Dai, "Cascaded-ring optical sensor with enhanced sensitivity by using suspended Si-nanowires," IEEE Photonics Technol. Lett. **23**(13), 842–844 (2011).
25. U. Manual, Version 2016, CST-Computer Simulation Technology GmbH, Darmstadt, Germany, 2017, *http://www.cst.com*.
26. R. Halir, P. J. Bock, P. Cheben, A. Ortega-Moñux, C. Alonso-Ramos, J. H. Schmid, J. Lapointe, D.-X. Xu, J. G. Wangüemert-Pérez, Í. Molina-Fernández, and S. Janz, "Waveguide sub-wavelength structures: a review of principles and applications," Laser Photonics Rev. **9**(1), 25–49 (2015).
27. S. Hughes, L. Ramunno, J. F. Young, and J. E. Sipe, "Extrinsic optical scattering loss in photonic crystal waveguides: role of fabrication disorder and photon group velocity," Phys. Rev. Lett. **94**(3), 033903 (2005).
28. W. Bogaerts, P. De Heyn, T. Van Vaerenbergh, K. De Vos, S. Kumar Selvaraja, T. Claes, P. Dumon, P. Bienstman, D. Van Thourhout, and R. Baets, "Silicon microring resonators," Laser Photonics Rev. **6**(1), 47–73 (2012).
29. H. Kogelnik, "Theory of optical waveguides" in *Guided-wave Optoelectronics*, T. Tamir, ed. (Springer, 1988).
30. C. Koos, *Nanophotonic Devices for Linear and Nonlinear Optical Signal Processing* (Univ.-Verlag Karlsruhe, 2007).
31. J.-M. Brosi, *Slow Light Photonic Crystal Devices for High Speed Optical Signal Processing* (Univ.-Verlag Karlsruhe, 2009).
32. J. D. Joannopoulos, R. D. Meade, and J. N. Winn, *Photonic Crystals: Molding the Flow of Light* (Princeton University Press, 1995).

1 Introduction

Waveguide-based optical sensors are used in a variety of applications such as label-free detection of chemical or biological analytes that specifically bind to functionalized waveguide (WG) surfaces [1–10]. Such sensors exhibit large potential for miniaturization and cost-efficient mass production, utilizing established photonic integration platforms such as silicon or silicon nitride. Sensor schemes are most commonly based on interferometers, e.g., in Mach-Zehnder and Young configuration [2,4,5,10,11], or on resonant devices, such as ring,

Optics EXPRESS

disk and Bragg resonators [1,2,5–9,12–17], which can be further enhanced by exploiting the Vernier effect [18]. Enabling large effective interaction lengths with the analyte, these sensor structures combine high sensitivity with small device footprint and lend themselves to high-density integration into massively parallel arrays.

The sensor principle relies on an optical WG, guiding a mode which significantly extends into the cladding medium that surrounds the WG core. The interaction between the optical mode field and the varying surface layer properties alters the effective refractive index by Δn_e and thus the optical phase shift accumulated during propagation. The strength of this effect is expressed by the so-called surface sensitivity, which, in combination with the effective-index sensitivity from the phase measurement of the underlying resonator or interferometer, determines the overall sensitivity. Proper optimization of the WG towards high surface sensitivities is hence key for realizing high-performance sensors.

Over the last years, various approaches for optimizing special types of WG were published, both for surface sensing [5,9,10,15–20] and for detection of bulk refractive index changes in the WG cladding (homogeneous sensing) [1,2,6–9,11–22]. However, these investigations are often limited to specific WG types and geometries on certain material platforms, such as silicon [1,2,5–9,11–14,16,19,22], silicon nitride (Si_3N_4) [5,10,11,17,21] and polymers [11,19]. It is hence impossible to broadly compare the highest achievable surface sensitivities across different WG types and integration platforms. Moreover, most sensitivity analyses consider only a specific type of surface layer with prescribed refractive index.

In this paper, we define a universal surface sensitivity which is broadly applicable to layers of different thicknesses and refractive indices. Focusing on Si_3N_4 and Si as the most common integration platforms, we systematically analyze a wide variety of WG types, comprising simple strip WG, slot and double slot WG, as well as sub-wavelength grating (SWG) WG. For each of these WG types, we identify the optimum geometry for both TE and TM polarization, and we compare the highest achievable surface sensitivities, taking into account implementation limitations that are associated with state-of-the art fabrication technologies. The focus of our analysis is on WG types that can be reliably mass-produced by optical lithography and single-etch structuring of WG on solid substrate layers. Note that even higher sensitivities can be achieved by more sophisticated WG concepts comprising ultra-small features [23] or suspended WG sections [24]. These schemes, however, require dedicated fabrication processes which are not yet accessible through scalable foundry processes. Exploiting the scalability of Maxwell's equations with respect to refractive index and geometry, we derive and physically explain general trends and design rules to corroborate the numerical results. Our findings can be used as design guidelines to select the appropriate WG platform and to optimize sensitivity for a given measurement task.

2 Scope

2.1 Basic sensing principle: Wave propagation and effective refractive index

For illustrating the basic sensing principle, we regard homogeneous sensing with a waveguide core embedded in an infinitely extended cladding medium. First, we define a few quantities: The propagation of monochromatic plane waves with vacuum wavelength λ in a homogenous medium is determined by the propagation constant $\beta = nk_0$ (refractive index n, vacuum propagation constant $k_0 = 2\pi/\lambda$). Dielectric WG consist of a high-refractive index core (n_{core}) and a low-refractive index cladding medium (n_{M}). For integrated optical WG made from silicon or silicon nitride, the core is usually supported by a buried oxide layer (BOX, bottom cladding) with refractive index n_{BOX}. The evanescent parts of the WG mode, which are essential for the sensing process, extend into the cladding region. The actual field

Optics EXPRESS

distribution in the various materials determines the WG propagation constant β, which can be expressed by an effective refractive index $n_e = \beta/k_0$. If the refractive index of the WG cladding changes, β and hence n_e are modified, which impacts the phase shift $\varphi = -\beta L = -n_e k_0 L$ accumulated over a propagation length L. Due to the large optical frequencies, a change of n_e is measured with high accuracy. For a given WG length L, the measured phase shift can be referred to a change Δn_e of the effective index, which finally allows to sense a change of the cladding index. The larger Δn_e becomes for a certain cladding index change, the more sensitive the device becomes.

2.2 Parameters for WG-based sensing

Numerous parameters determine how sensitive Δn_e reacts on a cladding index change Δn_M. Figure 1 summarizes the essential design elements for a WG: The material platform, the WG type, the polarization, and the WG geometry.

Platform Common integration platforms for optical sensors rely on a layer stack of a silicon (Si) or silicon nitride (Si_3N_4) device layer on a several micrometer thick BOX (SiO_2) as a bottom cladding, mechanically supported by a Si substrate. WG are structured in the device layer, and the BOX thickness is chosen to avoid leakage into the high-refractive index silicon substrate as well as to optimize grating coupler efficiency.

While Si WG are operated in the near infrared (NIR), Si_3N_4 WG are suitable for operation across the whole visible (VIS) and NIR spectrum. As the target media for biosensors are usually provided in the form of aqueous solutions, sensor operation at VIS wavelengths is much less impaired by water absorption than in the NIR. However, a large wavelength allows relaxed WG fabrication accuracies, and reduces scattering loss due to WG roughness. In this paper, we thus consider Si WG operated at NIR telecom wavelengths around 1550 nm, where low-cost laser sources are readily available, as well as Si_3N_4 WG operated at 600 nm as an example of low-wavelength sensors, Fig. 1(a).

Waveguide type For both integration platforms, we study four typical WG types, denoted as strip, slot, double slot and subwavelength grating (SWG) WG, Fig. 1(b). The last three types require significantly smaller feature sizes than the strip WG and thus lead to more challenging fabrication processes. We concentrate on single-mode WG, combining stable operation with high surface sensitivity.

Polarization We study the fundamental mode of both (quasi-)TE and (quasi-)TM polarization, Fig. 1(c). The term (quasi-)TE refers to the case where the dominant transverse electric field component is oriented parallel to the substrate. Similarly, (quasi-)TM denotes a configuration in which the dominant transverse magnetic field component is parallel to the substrate. For the TE polarization, an enhancement of the electric field E_x exists at the WG

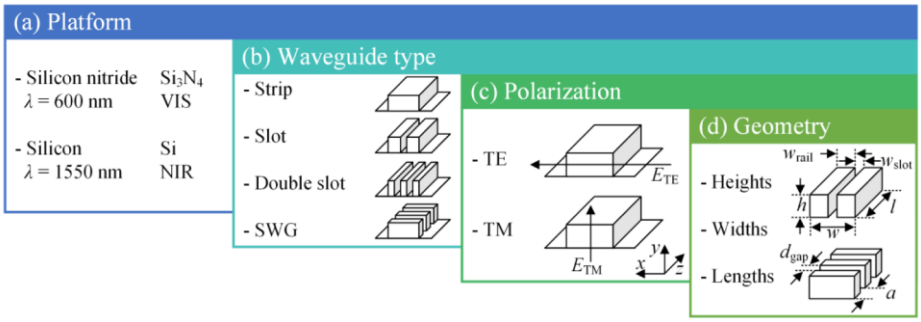

Fig. 1. Essential design elements for maximizing surface sensitivity. For each combination of (a) material platform, (b) WG type and (c) polarization, the optimum (d) geometry can be determined, observing reasonable technological constraints.

Optics EXPRESS

sidewalls because the normal component $D_x = \varepsilon_0 n^2(x) E_x(x)$ of the displacement must be continuous, while for TM polarization this field enhancement is to be seen at the top and at the bottom WG surfaces.

Geometry For all combinations of platform, WG type and polarization, we study the impact of the WG geometry in terms of height, width, and length (SWG only) of WG features, Fig. 1(d). Three standard device layer heights h of 220 nm, 250 nm and 340 nm are considered for Si WG structured on silicon-on-insulator (SOI) wafers. Wafers for structuring Si_3N_4 WG are available with a maximum Si_3N_4 thickness of several hundred nanometers, if stoichiometric growth is important, and we therefore consider typical WG heights h of 200 nm, 300 nm and 400 nm. For all WG types, we vary the overall width w in steps of 10 nm. For slot and double-slot WG, we additionally vary the slot width w_{slot} in four steps. For SWG WG, we vary the period a and the spacing d_{gap} between the WG elements, which can also be quantified by the fill factor $\text{FF} = (a - d_{\text{gap}})/a$. We chose a minimum feature size of $80\,\text{nm}$ to meet commercial technological conditions.

In the following, we define the surface sensitivity as a quantitative metric and maximize it by varying the various WG parameters. Given a certain sensor application, this data set allows to find the optimum design for a WG-based sensor.

3 Methods

3.1 Sensitivity of waveguide surfaces with respect to attached molecules

In WG-based sensing, two basic approaches are commonly used, which are compared in Fig. 2. For so-called homogeneous sensing, a bare WG core is exposed to a typically aqueous homogeneous target medium with refractive index n_{M}, Fig. 2(a). For surface sensing, a WG core is functionalized such that target molecules from an aqueous solution can bind to the core forming a surface layer with effective thickness t_{SL} and refractive index n_{SL}, Fig. 2(b).

For homogeneous sensing, a change in n_{M} causes a change Δn_{e} of the effective index. This change is the stronger the more the mode optical field extends into the target medium. Homogeneous sensing is usually unspecific, i.e., Δn_{e} cannot be traced back to a specific substance in the target medium if it is unknown which constituent is actually changing.

For surface sensing, the refractive index n_{M} of the aqueous solution remains essentially fixed, while the molecules, bound to the surface layer, influence the optical mode. By functionalizing the WG surface with dedicated capture agents, surface sensing can be used for

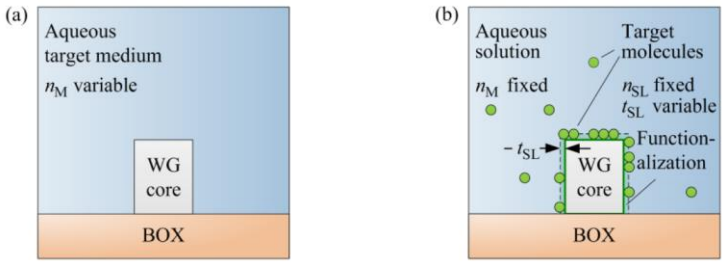

Fig. 2. Plain and functionalized strip WG on a buried silicon oxide (BOX) layer. (a) Homogeneous sensing. The aqueous target medium with refractive index n_{M} forms the homogeneous cladding of the WG core. A variation of n_{M} leads to a change in the effective refractive index of a guided mode. (b) Surface sensing. The refractive index n_{M} of the aqueous cladding solution remains constant. Target objects such as molecules, cells, vesicles or other corpuscules attach to the WG core, often mediated by a specific surface functionalization. The effect on the wave propagation is modelled with a surface layer (SL) having a refractive index n_{SL} and an effective layer thickness t_{SL}. The effective layer thickness takes into account the size of the target objects as well as the ratio of occupied binding sites.

Optics EXPRESS

specific detection of certain target analytes. The change Δn_e is the stronger, the more the optical field is concentrated within the surface layer. The definition of an effective layer thickness t_{SL} accounts for a possibly inhomogeneous distribution of target molecules within the surface layer. With biological samples this effective thickness is in the nanometer range, and n_{SL} is typically around 1.5.

For the remainder of this paper, we concentrate on surface sensing, offering a wide variety of applications. We are hence interested in the detailed influence of the surface layer on the electric and magnetic field distribution.

Figures 3 and 4 show the simulated electric field magnitudes of the fundamental quasi-TE mode of strip, slot and double slot WG, and for subwavelength grating WG, respectively. Details on the simulation parameters can be found in Appendix A. White contours mark the surface layer where molecules bind, leading to a change of the local refractive index.

A higher field concentration in these regions increases the surface sensitivity. While in a typical strip WG, Fig. 3(a), a large portion of the power is located inside the WG core, slot and double slot WG concentrate the power between the rails, Fig. 3(b) and Fig. 3(c). This means that the surface layer experiences higher field strengths in slotted WG, leading to a larger Δn_e compared to a strip WG.

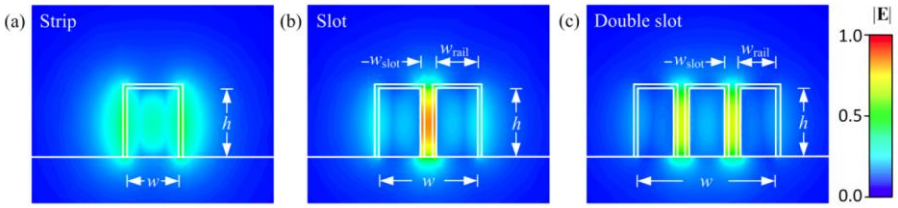

Fig. 3. Simulated electric field magnitudes of the fundamental quasi-TE mode in different WG geometries with height h and total width w. White contours mark the surface layer on the WG core. The larger the field strength in this region is, the larger the surface sensitivity becomes. The surface layer is disregarded for the field calculation. (a) Strip WG. The surface layer experiences only moderate field strengths. (b) Slot WG. A large field strength is located in the surface layers of the slot. (c) Double slot WG. The field strength in each of the two slots is smaller than for a single slot, but the relevant surface layer area has doubled.

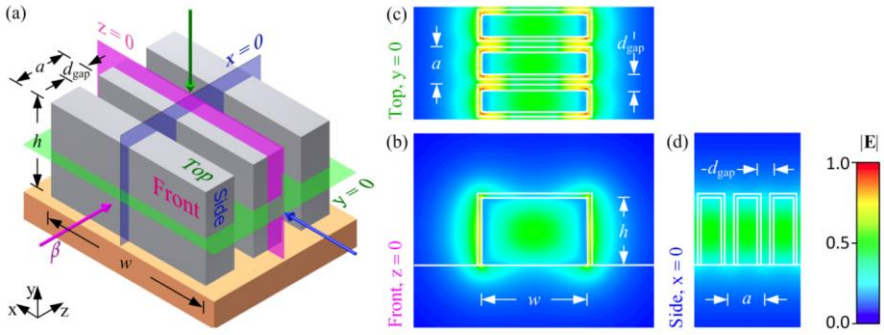

Fig. 4. Section of a subwavelength grating (SWG) WG and simulated electric field magnitudes of the fundamental quasi-TE mode. Compared to a strip WG, the surface layer area per unit length of the WG is increased. (a) Schematic of a SWG WG with width w, height h, period a and gap size d_{gap}. The pink arrow marks the direction of propagation. Specific cross-sections $z = 0$, $y = 0$, $x = 0$ are indicated with colored planes. Front, Top and Side mark the associated views. (b)-(d) Electric field magnitudes. White contours mark the surface layers on the WG core, which are disregarded for the field calculation. The larger the field strength in this region and the larger the surface, the bigger the surface sensitivity becomes. (b) Front view at $z = 0$. The surface layer experiences only moderate field strengths. (c) Top view at $y = 0$. Large field strengths (red areas) occur at the vertical edges of the blocks. (d) Side view at $x = 0$. Moderate field strengths are found at surfaces $z = \text{const}$.

Optics EXPRESS

SWG WG are composed of a multitude of individual WG elements, thus increasing the surface layer area per unit length of the WG. Figure 4(a) shows a section of a SWG WG with three periods along the propagation axis (pink arrow). For determining the surface sensitivity, a 3D elementary cell has to be simulated using periodic boundary conditions in the axial direction [25]. The three coloured planes mark the cross sections where the field distributions of Fig. 4(b)- 4(d) are plotted. In Fig. 4(c), high field strengths (red regions) are located at the vertical edges of the blocks.

3.2 Surface sensitivity and field perturbation approach

In the following, we define the surface sensitivity $S^{(\mathrm{surf})}$ as a quality metric for quantifying which WG design leads to the potentially highest overall sensor sensitivity. In order to obtain $S^{(\mathrm{surf})}$ for a specific WG with a specific surface layer, a full simulation of the WG with and without the surface layer would be required. This would include various combinations of surface layer properties like layer refractive index n_{SL} and layer thickness t_{SL} and thus requires a multitude of high-resolution simulations to resolve the surface layer with a thickness in the range of a few nanometers. Exploring the whole parameter space as discussed in Section 2 and Fig. 2 would hence be a time-consuming and probably unrealistic endeavour.

To overcome this problem we use a perturbation approach. To this end, we perform a single finite-element method (FEM) simulation of a bare WG for each WG geometry, store the resulting fields and compute the influence of an additional surface layer using a field interaction factor. This technique is only valid for small perturbations, i.e., the modal field does not change significantly with the surface layer, and hence the effective refractive index n_{e} changes only slightly, too. In our case, binding events of molecules change the refractive index of a surface layer of thickness t_{SL} from n_{M} to n_{SL}, where n_{SL} is the refractive index of the attached molecules and n_{M} denotes the background refractive index of the solution. The conditions for the perturbation approach hold, if either the thickness is small – then the refractive index change n_{SL}-n_{M} can be larger – or if n_{SL}-n_{M} is small, in which case t_{SL} can be larger.

We will first consider the general case of a z-variant SWG WG. According to Eq. (14) and Eq. (15) in Appendix C, the local perturbation in the surface layer can be translated into a change $\Delta n_e^{(\mathrm{surf})}$ of the effective refractive index via the field interaction factor $\Gamma_{\mathrm{SL}}(t_{\mathrm{SL}})$,

$$\Delta n_{\mathrm{e}}^{(\mathrm{surf})} = \left(n_{\mathrm{SL}} - n_{\mathrm{M}}\right)\Gamma_{\mathrm{SL}}, \qquad \Gamma_{\mathrm{SL}}(t_{\mathrm{SL}}) = \frac{1}{\mathrm{Z}_0}\frac{\frac{1}{\mathrm{a}} n_{\mathrm{M}} \iiint_{V_{\mathrm{SL}}} |\mathbf{E}|^2 \,\mathrm{d}V}{\iint \mathrm{Re}(\mathbf{E}\times\mathbf{H}^*)\cdot\mathbf{e}_z \mathrm{d}A} = \frac{c}{n_{\mathrm{g,SL}}}\frac{W_{\mathrm{SL}}}{P}. \quad (1)$$

Note that the field interaction factor $\Gamma_{\mathrm{SL}}(t_{\mathrm{SL}})$ in Eq. (1) is different from the intensity-related field confinement factor $\Gamma_i^{(\mathrm{conf})} = W_{\mathrm{SL}}/W$, because it describes also slow-light propagation, see Eq. (10). For computing $\Gamma_{\mathrm{SL}}(t_{\mathrm{SL}})$, we simulate the fields of a unit cell for a bare SWG WG without surface layers, and integrate $|\mathbf{E}|^2$ over the fictitious surface layer volume V_{SL} within that unit cell of length a. The result is proportional to the electric energy W_{SL} in the surface layer per unit cell, Eq. (7) and Eq. (10) for $i = \mathrm{SL}$, normalized to the cross-sectional power P of Eq. (5). The integral in the denominator extends over the entire (x,y)-plane.

For z-invariant WG such as strip, slot, or double slot structures, Eq. (1) can be simplified by exploiting the fact that the integral over $|\mathbf{E}|^2$ is invariant along z. Simplifying the volume integral in the numerator, we obtain

$$\Gamma_{\mathrm{SL}}(t_{\mathrm{SL}}) = \frac{1}{\mathrm{Z}_0}\frac{n_{\mathrm{M}} \iint_{A_{\mathrm{SL}}} |\mathbf{E}|^2 \mathrm{d}A}{\iint \mathrm{Re}(\mathbf{E}\times\mathbf{H}^*)\cdot\mathbf{e}_z \mathrm{d}A} \quad (2)$$

Optics EXPRESS

for the special case of *z*-invariant WG. We use Eq. (2) to analyze $\Gamma_{\mathrm{SL}}(t_{\mathrm{SL}})$ for a set of widely different silicon strip WG geometries, operated in quasi-TE polarization, see in Fig. 5(a) "Perturbation". As a comparison, we extract the numerically exact values $\Gamma_{\mathrm{SL}}^{(\mathrm{sim})}(t_{\mathrm{SL}}) = \Delta n_{\mathrm{e}}^{(\mathrm{surf,\,sim})} / \left(n_{\mathrm{SL}} - n_{\mathrm{M}}\right)$ from $\Delta n_{\mathrm{e}}^{(\mathrm{surf,\,sim})}$ obtained by FEM simulations of quasi-TE fields that include the actual surface layers, and plot them in Fig. 5(a) with a dashed line "Simulation". The agreement is very good except for large surface layer thicknesses t_{SL} on narrow and high WG. In this case the majority of the field is contained within the surface layers, violating the assumption of a small field perturbation. The electric field outside the core decays approximately exponentially. For a penetration depth much larger than t_{SL}, the decay of the field magnitude and of the power within the surface layer can be approximated by a linear function. As a consequence, Γ_{SL} in Fig. 5(a) exhibits a region where it depends linearly on t_{SL}.

We now want to define a sensing sensitivity which is – in the framework of the perturbation approach – independent of the surface layer thickness. This can be achieved by looking at the derivative $\partial\left(\Delta n_{\mathrm{e}}^{(\mathrm{surf})}\right)/\partial t_{\mathrm{SL}}$ at $t_{\mathrm{SL}} = 0$. Because n_{SL} influences the result, we define the surface sensitivity as the derivative of the field interaction factor with respect to the surface layer thickness,

$$\Delta n_{\mathrm{e}}^{(\mathrm{surf})} = \left(n_{\mathrm{SL}} - n_{\mathrm{M}}\right) S^{(\mathrm{surf})} \Delta t_{\mathrm{SL}} = \left(n_{\mathrm{SL}} - n_{\mathrm{M}}\right) S^{(\mathrm{surf})} t_{\mathrm{SL}}, \qquad S^{(\mathrm{surf})} = \left.\frac{\partial \Gamma_{\mathrm{SL}}}{\partial t_{\mathrm{SL}}}\right|_{t_{\mathrm{SL}}=0}. \quad (3)$$

Within the framework of the perturbation approach, the surface sensitivity $S^{(\mathrm{surf})}$ is independent of both the thickness and the refractive index of the surface layer, thus providing a universal guideline for a sensor design. To confirm this, we calculate the derivatives $\partial\Gamma_{\mathrm{SL}}/\partial t_{\mathrm{SL}}$ for the same WG types as used in Fig. 5(a), and plot them semi-logarithmically in Fig. 5(b) in a range $t_{\mathrm{SL}} = (1\ldots350)\,\mathrm{nm}$. We see that $\partial\Gamma_{\mathrm{SL}}/\partial t_{\mathrm{SL}}$ is constant for very small $\Delta\varepsilon(\mathbf{r}) = \varepsilon_0 \cdot \Delta\varepsilon_r(\mathbf{r})$, owing to the approximately linear dependence of Γ_{SL} on t_{SL} in this region. We extrapolate the curve to $t_{\mathrm{SL}} = 0$, where the computation fails due to the finite spatial discretization. In Fig. 5(b) these extrapolated values are marked with filled circles and denoted by $S^{(\mathrm{surf})}$, see Eq. (2) and Eq. (3). Note that a WG design for best homogeneous sensitivity $S^{(\mathrm{hom})} = \Gamma_{\mathrm{M}} \approx \lim_{t_{\mathrm{SL}}\to\infty} \Gamma_{\mathrm{SL}}$ is not necessarily optimum when it comes to surface

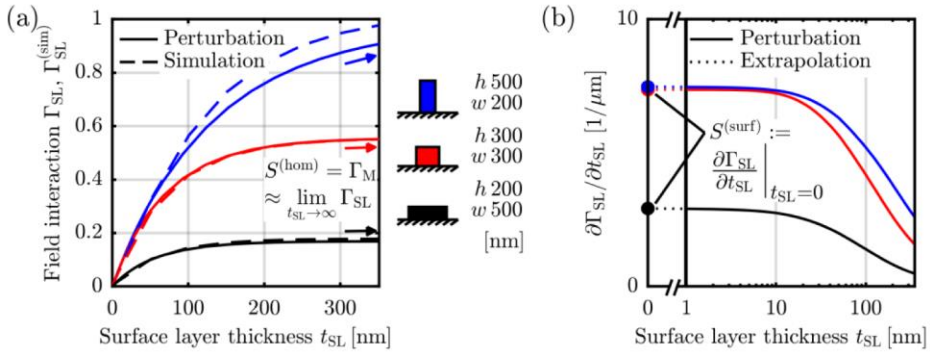

Fig. 5. Validation of perturbation model and definition of surface sensitivity $S^{(\mathrm{surf})}$ for a set of widely different silicon strip WG geometries propagating quasi-TE fields, see legend. (a) Field interaction factor Γ_{SL} in surface layer of thickness t_{SL}. For each geometry, we compare Γ_{SL} computed with a perturbation approach (based on a single FEM simulation without a surface layer) with $\Gamma_{\mathrm{SL}}^{(\mathrm{sim})}$ obtained from numerically exact calculations (FEM simulations with different surface layers). For small t_{SL} the agreement is very good. For large t_{SL} (marked by $\rightarrow$) and $n_{\mathrm{SL}} \approx n_{\mathrm{M}}$, the field interaction factor Γ_{SL} approaches Γ_{M} and thus the homogeneous sensitivity $S^{(\mathrm{hom})}$. (b) The surface sensitivity $S^{(\mathrm{surf})} = \partial\Gamma_{\mathrm{SL}}/\partial t_{\mathrm{SL}}\big|_{t_{\mathrm{SL}}\to\infty}$ allows an easy comparison of different WG, irrespective of the actual surface layer properties.

Optics EXPRESS

sensing, see Eq. (16) in Appendix C for rigorous definition of $S^{(\mathrm{hom})}$. This can be inferred by comparing $S^{(\mathrm{hom})}$ of the two top strip WG in Fig. 5(a) (blue and red arrows) to the corresponding $S^{(\mathrm{surf})}$ of the same WG in Fig. 5(b) (blue and red dots at $t_{\mathrm{SL}} = 0$): A larger homogeneous sensitivity does not lead to a difference in surface sensitivity. In sensing applications, the desired measurement quantity is the effective surface layer thickness t_{SL}. From a measurement of Δn_{e} and for known bulk refractive indices n_{SL} and $\Delta n(\mathbf{r}) \ll n(\mathbf{r})$, t_{SL} can be inferred from Eq. (3). A proper choice of the WG then maximizes $S^{(\mathrm{surf})}$ and therefore the measurement sensitivity for t_{SL}.

4 Results

We extract the surface sensitivity $S^{(\mathrm{surf})}$ according to Section 3.2 for all kinds of WG outlined in Section 2.2. For each TE- or TM-operated WG core in Si_3N_4 or alternatively in Si, a few typical values of heights h, slot widths w_{slot}, periods a and fill factors FF are considered. For each combination of these parameters, the total WG width w (see Fig. 3) is then optimized to obtain optimum surface sensitivity $S_{\mathrm{opt}}^{(\mathrm{surf})}$. To this end, the WG width is swept with a step size of 10 nm. The range of w was chosen to maintain single-mode operation and to avoid substrate leakage for a BOX thickness of $2\,\mu\mathrm{m}$. The following WG surface sensitivity analysis relates to Si_3N_4 cores operated in the VIS (Section 4.1) as well as to Si cores operated in the NIR (Section 4.2).

4.1 Silicon nitride – VIS

For Si_3N_4 strip WG operated at a vacuum wavelength of 600 nm, Fig. 6(a) displays the surface sensitivity $S^{(\mathrm{surf})}$ as a function of the WG width w for three different WG heights h and for both polarizations. Surface sensitivity generally benefits from large electric field strengths in the region of the surface layer as indicated in Fig. 3. Within their single-mode range, each WG shows a maximum $S_{\mathrm{opt}}^{(\mathrm{surf})}$ at a distinct width w_{opt}, marked by ● for TE and by ▲ for TM. Any deviation from this optimum width w_{opt} reduces the interaction factor of the mode field with the surface layers, e.g., by concentrating the modal field to the WG core in the case of large w, or by spreading the mode field over a wider cladding region for small w. In the second case, a large portion of the modal field propagates in the BOX, where it cannot contribute to sensing. In Fig. 6(b), the extracted w_{opt} (green, left vertical axis) and the corresponding $S_{\mathrm{opt}}^{(\mathrm{surf})}$ (magenta, right vertical axis) are plotted as a function of the WG height h. The largest $S_{\mathrm{opt}}^{(\mathrm{surf})}$ are found for large h and small w for both polarizations, with TE operation being more sensitive than TM operation. In the TE case, the two sides of the WG exploit the field enhancement, and only a small portion of the mode propagates in the BOX, see Section 5.2 for a more detailed discussion.

We apply the same procedure to TE-operated Si_3N_4 slot and double slot WG and plot the results in Fig. 7. As for the strip WG, the sensitivity increases for higher WG. Decreasing WG width first leads to increasing sensitivity, which drops again as the width becomes too small and the mode extends far into the cladding, Fig. 7(a). Smaller slot widths increase the field strength in the slot, Fig. 7(b). This, together with the growth of the total core surface as compared to a strip WG, increases $S^{(\mathrm{surf})}$. The slot WG behaves like two strip WG if $w_{\mathrm{slot}} \gg w_{\mathrm{rail}}$. In this case, each of the strips carries half the power of the slot WG and both the optimum rail width and the optimum sensitivity converge to those of a single optimum strip WG, see top light green line (∞) and light magenta line (∞) in Fig. 7(b).The double slot WG behaves like three separate strip WG if $w_{\mathrm{slot}} \gg w_{\mathrm{rail}}$. In this case, each strip carries one third of the power of the slot WG. As before, both the optimum rail width and the optimum sensitivity converge to those of a single optimum strip WG, see top light green line (∞) and light magenta line (∞) in Fig. 7(c) and in Fig. 7(d).

Optics EXPRESS

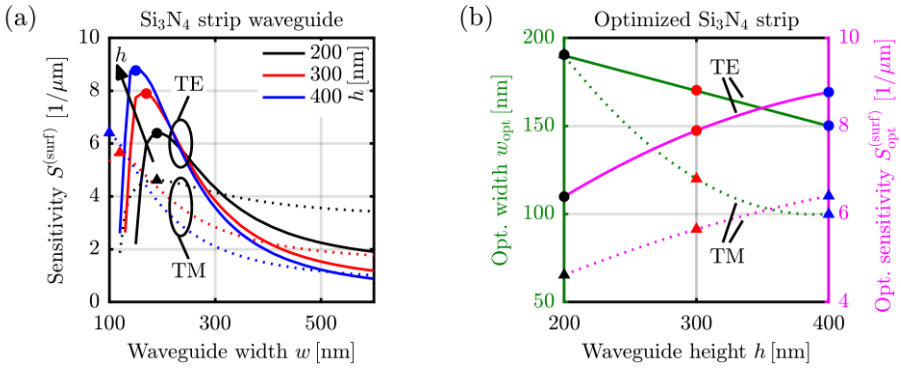

Fig. 6. Optimization of Si_3N_4 strip WG. (a) Surface layer sensitivity $S^{(surf)}$ of Si_3N_4 strip WG for three standard heights h as a function of WG width w in TE and TM operation. For decreasing WG width, the sensitivity first increases to its maximum value and then drops. The drop is caused by the fact that the mode extends far into the cladding or is even lost to the substrate for very small WG widths. The optimum sensitivities $S_{opt}^{(surf)}$ at the corresponding optimum widths w_{opt} are marked by dots and triangles. (b) Comparison of optimum WG width (green, left axis) and optimized surface sensitivity (magenta, right axis) as a function of WG height. High and narrow WG (blue markers) are most sensitive.

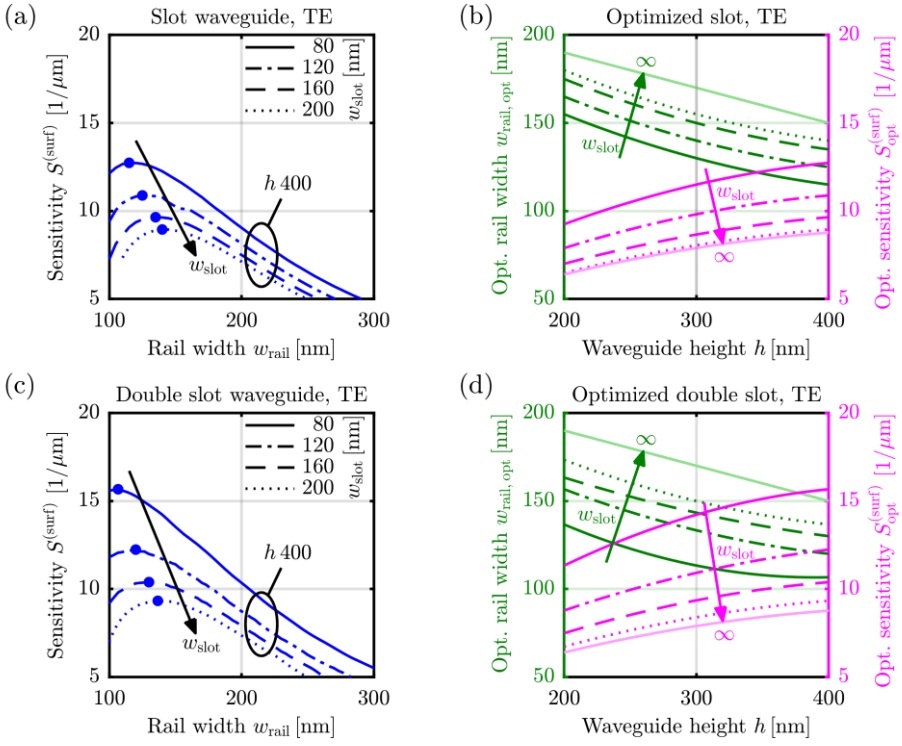

Fig. 7. Optimization of Si_3N_4 TE-operated slot and double slot WG. (a,c) Surface layer sensitivity $\mathbf{H}(t,\mathbf{r}) = \mathbf{H}(\mathbf{r})e^{i\omega t}$ of Si_3N_4 slot and double slot WG for a height of h = 400 nm as a function of rail width w for different slot widths. For decreasing rail width, the sensitivity first increases to its maximum value and then drops. The drop is caused by the fact that the mode extends far into the cladding or is even lost to the substrate for very small rail widths. The optimum sensitivities $S_{opt}^{(surf)}$ at the corresponding optimum widths w_{opt} are marked by dots. (b,d) Optimized geometries for slot and double slot Si_3N_4 WG. Sensitivities are larger for small slot widths and higher WG, the fabrication of which is limited by technological constraints. The light solid lines indicate the limits for $w_{slot} \to \infty$ (light green and light magenta).

Optics EXPRESS

4.2 Silicon – NIR

For silicon WG operated in the NIR at $\lambda = 1550\,\mathrm{nm}$, we apply the same technique as described in Section 4.1 for extracting the optimized widths $\mathbf{E}(t,\mathbf{r}) = \mathbf{E}(\mathbf{r})e^{j\omega t}$ and sensitivities $S_{\mathrm{opt}}^{(\mathrm{surf})}$. In Fig. 8(a) we display the surface sensitivity $S^{(\mathrm{surf})}$ of a strip WG with different heights h and polarizations as a function of the strip width w. As for Si_3N_4 strip WG, Fig. 6(a), we see a sensitivity maximum within the range of single-mode operation. The TE sensitivity exhibits a pronounced maximum when varying the WG width, whereas the maximum of the TM sensitivity is much less pronounced, Fig. 8(a).

We extract the optimum sensitivity and find that, similar to Si_3N_4, the TE sensitivity can be optimized by choosing a high WG with a narrow width, while the optimum TM sensitivity depends only weakly on the WG height, Fig. 8(b). This is in contrast to the findings for the Si_3N_4 WG, Fig. 6(a) and Fig. 6(b). We attribute this to the fact that the refractive index of Si is significantly larger than that of Si_3N_4 and hence the asymmetry introduced by the underlying BOX is less significant for the case of Si WG. Consequently, for the WG heights h under consideration and for small w, the TM modal field does not extend into the BOX as strongly as for the Si_3N_4 WG. Regarding Si slot and double slot WG, Fig. 8(c) and Fig. 8(d),

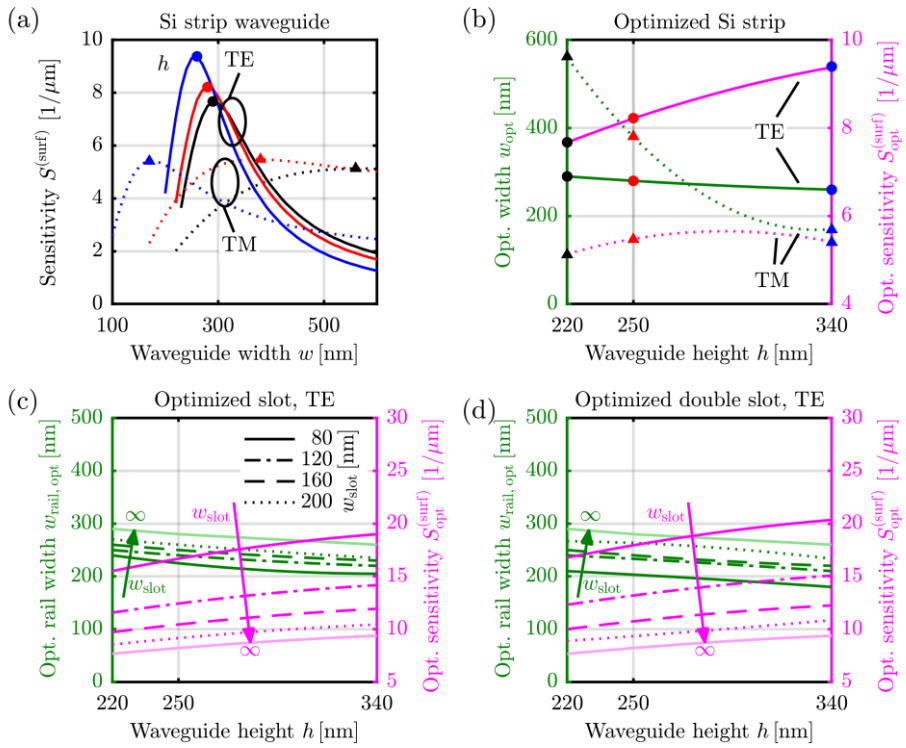

Fig. 8. Optimization of Si strip, slot and double slot WG. (a) Surface layer sensitivity $S^{(\mathrm{surf})}$ for Si strip WG with three standard heights h as a function of WG width w in TE and TM operation. For decreasing WG width, the sensitivity first increases to its maximum value and then drops. The drop is caused by the fact that the mode extends far into the cladding or is even lost to the substrate for very small WG widths. The optimum sensitivities $S_{\mathrm{opt}}^{(\mathrm{surf})}$ at the corresponding optimum widths w_{opt} are marked by dots and triangles in (a) and compared in (b). For TE polarization, high and narrow WG yield the best results. For TM, no pronounced sensitivity gain is observed for higher WG. (c),(d) Optimized geometries for slot and double slot Si WG. The sensitivity increases for smaller slot widths and higher WG, the fabrication of which is limited by technological constraints. The light solid lines indicate the limits for $w_{\mathrm{slot}} \to \infty$ (light green and light magenta), which correspond to the case of individual strip WG as shown in (b). Compared to Si_3N_4, operated in the VIS, the Si rails, operated in the NIR, are typically 100 nm wider while the sensitivity is slightly larger.

Optics EXPRESS

the trends for optimum WG parameters are similar to those of Si_3N_4 WG, Fig. 7, and thus the same conclusions as in Section 4.1 can be drawn.

For achieving better sensitivities, completely different WG structures in form of sub-wavelength gratings (SWG) were proposed [8,14,20], see Fig. 4. For a systematic comparison to strip and slot WG, we calculate and optimize the surface sensitivities of SWG WG, see Fig. 9. Figure 9(a), left y-axis, shows the results obtained from two exemplary families of TE-operated SWG, which share the same period a, but differ in height h, and feature three different fill factors (50%, 60%, 70%, differing in opacity). For the smaller WG height ($h = 220\,\text{nm}$, black lines) the sensitivity does not depend strongly on the WG width. Strip and slotted WG show a fast decrease of sensitivity for widths larger than w_{opt}, since the fields become smaller at the sensitive side walls and upper surfaces and are more confined to the core. This effect is less pronounced in the SWG WG, since the sensitive surface in between the individual blocks are still experiencing high field strengths. The maximum sensitivity shifts to smaller w, when the FF increases. This is to be expected: If the FF becomes larger, the effective refractive index increases and the modal field is stronger confined to the SWG "core". For a sufficient influence of the analyte, the field must then extend into the SWG "cladding", which is achieved by decreasing the WG width.

Interestingly, for higher SWG WG ($h = 340\,\text{nm}$, blue lines), no optimum sensitivity within the observed range of w is found. Instead, the sensitivity increases strongly for larger FF and wider WG. This increase is caused by a decrease of the group velocity: Larger FF and increased w lead to an increase of the Bragg wavelength associated with the SWG structure. Once the Bragg wavelength comes close to the operation wavelength of the sensor, the device enters the so-called slow-light regime [26], which is characterized by a greatly reduced group velocity and a greatly increased effective group index n_{eg}.

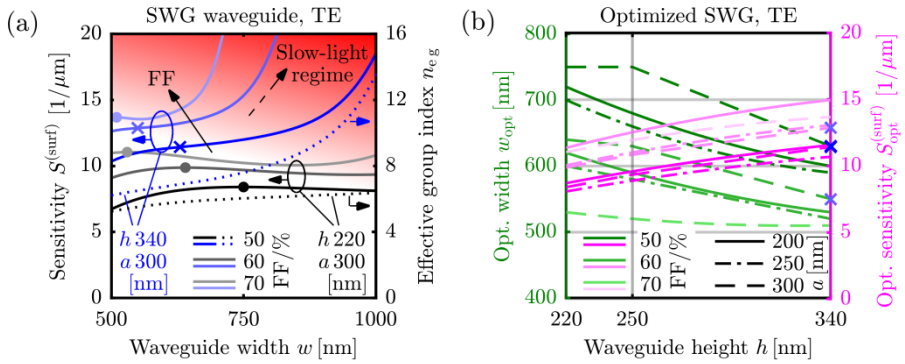

Fig. 9. Optimization of Si sub-wavelength grating (SWG) WG. (a) Sensitivity of two SWG WG families with different heights, but common period $a = 300\,\text{nm}$ as a function of WG width w for varying fill factors FF. Black lines with various shadings stand for a WG height of $h = 220\,\text{nm}$. In this case, the sensitivity does not change significantly with w. For higher SWG WG ($h = 300\,\text{nm}$, blue lines), the sensitivity traces do not exhibit an optimum before entering into the so-called slow-light regime, where the sensitivity is dominated by a largely increased effective group index n_{eg} (dotted lines for two different heights for $\text{FF} = 50\%$). For these traces, we choose the inflection points (blue crosses) to define reasonable sensor designs that are not subject to the impairments associated with slow-light operation. Note that this does not represent an optimum in a strict mathematical sense. (b) Overview of optimized SWG WG sensitivities and geometries outside the slow-light regime for different heights, periods, and fill factors. As a trend, the sensitivity increases for high WG and small gap size $d_{gap} = a(1-\text{FF})$. Sensitivities obtained from inflection points at the transition to slow-light operation are again marked by blue crosses. Note that the traces for FF = 70% and a = 200 nm or a = 250 nm do not appear in the plot since the associated gap sizes are below the minimum feature size of 80 nm.

Optics EXPRESS

This leads to strong interaction of the guided mode with the surface layer and hence to a sharp increase of the sensitivity even though the field confinement $\Gamma_i^{(\mathrm{conf})}$ of the surface layer does not significantly increase, see Appendix C, Eq. (7). Note, however, that operation in the slow-light regime strongly increases the detrimental impact of environmental influences and fabrication tolerances [27] on mode propagation. Amongst others, the sensitivity and the optical propagation loss can vary significantly, such that reproducible device properties and hence a reliable sensor read-out are difficult to obtain in practice. This susceptibility to tiny changes can be seen exemplarily for the sensitivity of WG operating in the slow-light regime, Fig. 9(a), which is why we exclude them from the following discussion. For sensitivity traces entering into the slow-light regime without showing a pronounced maximum in Fig. 9(a), we choose the inflection points (blue crosses) to define reasonable sensor designs that are not subject to the impairments associated with slow-light operation without being optimum in a strict mathematical sense. Note that there is no distinct transition from "normal" SWG operation to the slow-light regime and that a strict classification is therefore difficult. The choice of the inflection point is motivated by the fact that it marks the width at which the influence of the increased effective group index n_{eg} starts dominating over the impact of the field confinement factor $\Gamma^{(\mathrm{conf})}$, Eq. (10) in Appendix C, and dotted lines for two different heights for $\mathrm{FF} = 50\,\%$, see Fig. 9(a).

The optimal sensitivities of the SWG WG and those obtained for the transition to the slow-light regime are summarized Fig. 9(b), indicating again a general sensitivity increase with WG height. Sensitivities obtained from inflection points at the transition to slow-light operation are marked by blue crosses. As a trend, $\mathbf{E}(\mathbf{r})$ increases for small $d_{\mathrm{gap}} = a(1-\mathrm{FF})$.

4.3 Comparison of optimized sensitivities

As a summary of the last two subsections, we compare the sensitivities for the Si_3N_4 platform in the VIS (λ = 600 nm) and for the Si platform in the NIR (λ = 1550 nm). In Fig. 10, we display the optimized surface layer sensitivities $S_{\mathrm{opt}}^{(\mathrm{surf})}$ of four WG types (strip, slot, double slot, SWG) with optimized widths w_{opt} and three different heights h for each platform. The left (right) panel shows the results for TE (TM) polarization. Three SWG positions are empty because either the gap size is below the minimum feature size of 80 nm, or because the structures do not support a well-guided WG mode ("poor waveguiding").

Generally, WG operated in TE polarization tend to be more sensitive than their TM-operated counterparts. Taking into account technological constraints such as a minimum feature size, the sensitivity of sophisticated WG designs can be increased by approximately a factor of 2 compared to a simple optimized strip WG. Moreover, slotted WG exhibit higher surface sensitivities than strip and SWG WG. Generally, the silicon platform exhibits higher sensitivities than the silicon nitride platform, which can be attributed to the fact that a higher index contrast leads to a smaller penetration depth into the WG cladding and hence to a stronger interaction with a thin surface layer. Larger WG heights tend to be generally better. For the silicon photonic platform, a sensitivity improvement of a factor of 4 can be achieved by using optimum double slot WG operating in TE polarization instead of non-optimum standard strip WG featuring, e.g., WG dimensions of $w \times h = 450 \times 220\,\mathrm{nm}^2$ for TM operation. A detailed physical explanation of these trends is given in Section 5.2.

5 Discussion

For a better understanding of the trends summarized in Fig. 10, we first discuss in Sect. 5.1 the influence of the cladding asymmetry and we introduce universal scaling laws of Maxwell's equations that allow to generalize the findings to other wavelengths and material platforms. Subsequently, Section 5.2 explains in detail the trends when varying platform, operating wavelength, polarization, and WG geometry. Section 5.3 discusses our results in

Optics EXPRESS

view of an overall sensor performance, including the impact of mode loss and analyte delivery as well as a practical instruction on finding global sensitivity optima.

5.1 Physical interpretation — Cladding asymmetry and scaling laws

Cladding asymmetry We consider WG cores that are supported by a SiO_2 substrate (BOX) and are surrounded by an aqueous medium. Since the refractive index of SiO_2 is larger than that of water, the cladding of the WG is asymmetric. It is hence impossible to increase the interaction of the guided light with the functionalized surface of the WG core by simply indefinitely reducing the core dimensions. Any attempt to do so will predominantly increase the fraction of the mode fields in the BOX, which do not contribute to sensing, and therefore decrease the fraction in the surface layer, leading to a decrease of surface sensitivity.

Scaling laws According to Appendix D, Table 1, line (1), scaling the geometry of a WG by σ_g while simultaneously scaling the operating frequency by $\sigma_\omega = 1/\sigma_g$ leads to a simple geometrical scaling of the eigenfunctions of the electric and magnetic fields by a factor of σ_g. This effect has no impact on the homogeneous sensitivity, since the relative portion of the mode fields in the cladding medium remains the same. However, the surface sensitivity Eq. (3) as defined by a derivative with respect to the layer thickness t_{SL} scales by $1/\sigma_g$. *The optimum surface sensitivity of a WG with fixed RI profile hence scales proportionally to frequency and requires an inverse scaling of the geometry*.

Furthermore, scaling the refractive index profile of a WG by σ_n while simultaneously scaling the operating frequency by $\sigma_\omega = 1/\sigma_n$ does not change the distributions of the electric and the magnetic field. Note, however, that the ratio of the electric-field eigenfunction $\mathbf{E}(\mathbf{r})$ and the magnetic field eigenfunction $\mathbf{H}(\mathbf{r})$ must be scaled by a factor of $1/\sigma_n$, see Eq. (19) and the corresponding discussion. In total, this leaves the sensitivity unchanged, see Appendix D, Table 1, line (2). *The optimum surface sensitivity of a WG with fixed geometry does not change for a scaling of the RI profile along with an inverse scaling of the frequency*.

5.2 Trends for increased surface sensitivity

High index-contrast platform The surface sensitivity is maximized by an increased field concentration at the core surface which is in contact with the analyte-delivering aqueous medium. For a given wavelength, the surface sensitivity tends to increase if the refractive index ratio $n_{\mathrm{core}}/n_{\mathrm{M}}$ between core and cladding is increased. This is caused by the fact that the enhancement of surface-normal electric-field components is proportional to the square $(n_{\mathrm{core}}/n_{\mathrm{M}})^2$ of the index ratio. Similarly, the surface sensitivity increases when the refractive

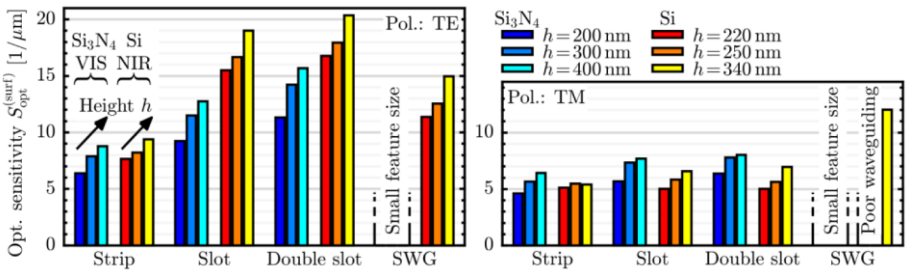

Fig. 10. Comparison of the optimized sensitivity for the Si_3N_4 platform in the VIS (λ = 600 nm) and for the Si platform in the NIR (λ = 1550 nm). We consider four WG types with three standard heights, operated in TE and TM polarization. Sensitivity can be increased by higher WG cores, by using Si rather than Si_3N_4, and by TE-operation instead of TM. The three more advanced WG types (slot, double slot and SWG) offer an enhancement of up to a factor of 2 compared to TE strip WG, coming at the cost of a more complex fabrication process. Some SWG WG are excluded because either the gap size is below 80 nm ("small feature size"), or because the structures do not support a well-guided WG mode ("poor waveguiding"). Typical slot, double slot and SWG WG do not benefit from TM operation.

Optics EXPRESS

index profile is scaled up by a factor of $\sigma_n > 1$ while the geometry is scaled down by $\sigma_g = 1/\sigma_n$ to maintain the same operating frequency, see Appendix D, Table 1, line (4).

Short operating wavelength For a given platform, higher surface sensitivities can be achieved with shorter wavelengths. The geometry of the WG has to be chosen guided by the scaling law in Appendix D, Table 1, line (1). A Si_3N_4 WG operated in the VIS will achieve a higher optimized surface sensitivity than an up-scaled Si_3N_4 WG operated in the NIR. When comparing optimized Si_3N_4 WG operated in the VIS with Si WG in the NIR, we see similar sensitivities, see Fig. 10. This is due to a combination of scaling the operating frequency along with WG geometry and an increase of the index contrast, which have contrary effects on the surface sensitivity that partially cancel each other. However, optimized WG in the high index-contrast Si platform will be as good as or better than those in the Si_3N_4 platform, as the limitation caused by the cladding asymmetry is more severe for lower core refractive indices.

TE polarization The TE-mode sensitivity generally outperforms that of the TM-mode. For high index-contrast WG as considered in this work, the surface sensitivity is dominated by regions exhibiting a large field enhancement due to field discontinuities at the core-cladding interface. In the case of strip WG and of vertical-slot WG, TE mode operation exploits more of these surfaces showing a field enhancement compared to TM mode operation, where the interface to the bottom oxide cladding (BOX) does not contribute to the sensitivity.

It has to be noted that in typical telecom applications, WG with large widths and small heights are used (e.g. $w \times h = 450 \times 220\,\mathrm{nm}^2$). These established WG geometries are often directly transferred to sensing applications. For these WG and for improper choices of WG widths exceeding 400 nm, the TM-polarized mode may exhibit higher sensitivity than the TE, see Fig. 8(a). However, under the assumption that the width w is chosen properly, TE-polarized modes exhibit optimized sensitivities. If an application demands Si WG operated with TM polarized light, SWG WG are best, see Fig. 10.

Enlarged surface Larger core surfaces can generally lead to higher sensitivities. This is exploited by slot, double slot and SWG WG, which introduce additional sensor surfaces compared to strip WG. For slot and double slot WG operated in TE polarization, these additional surfaces lead to an enhancement of the dominant transvers electric-field component and can hence further increase surface sensitivity, see Fig. 3. For SWG WG, the enhancement at the additional surfaces between the blocks affects the weaker longitudinal electric-field component, Fig. 4. The sensitivity gain of SWG WG is hence smaller than that of slot and double slot WG and depends only weakly on the polarization.

Narrow waveguides Optimized sensing WG have a narrower core than typical routing WG designed for low-loss light transport. Narrow WG increase the sensitivity for both TE and TM polarization. When narrowing the WG, the sensitive area of the surface layer remains essentially constant for TE polarization, but the surface field strength initially increases. For very small WG widths, the modal field expands into the cladding and the field interaction with the surface layer decreases, hence reducing surface sensitivity. For TM polarization, an additional effect comes into play: When reducing the WG width, the top surface, containing high field strengths due to the field discontinuity, becomes smaller, while the overall field strengths at the core surface become more dominant as the mode expands into the cladding. This interplay leads to a less pronounced optimum with respect to the WG width than for TE polarization, see Fig. 6(a) and Fig. 8(a).

High waveguides Higher WG cores outperform thinner WG cores. Higher WG cores reduce the relative portion of the fields located in the BOX (cladding asymmetry). Since this area does not contribute to sensing, less field strength in the BOX leads to larger sensitivities. In addition, and especially for the TE mode, higher WG cores enlarge the sensitive sidewall regions that are subject to field enhancement.

Optics EXPRESS

Small slot widths and gap sizes Smaller slot widths and gap sizes generally increase the sensitivity. For slot WG, the sensitivity gain for TE is more pronounced than for TM polarization.

5.3 Overall sensor system performance and the impact of mode loss

It is important to note that, when implementing these WG into functional sensors, the overall system performance will also depend on additional aspects that are outside the scope of our current analysis. One of the most important aspects is the propagation loss of the optical mode: Adsorption of target molecules to the WG surface is usually measured by recording the phase shift accumulated over a certain propagation length L. In technical implementations of sensors, the phase measurement accuracy depends on the precision with which amplitudes can be measured in an interference setup. This precision and ultimately the detection limit of the sensor system decreases with increasing propagation losses, which, like the sensitivity, depends on WG platform, type, polarization and geometry. A rigorous system optimization would ideally have to take into account all these effects to account for mutual trade-offs and to find a global optimum of the WG design. This, however, would require a quantitative relationship between the WG geometry and the associated power propagation loss $\exp(\alpha L)$ characterized by the loss constant α, which is impossible to state in a general and reliable manner across different integration platforms.

For finding optima for the overall system sensitivity S_{sys} for a specific sensor implementation, we need to know three dependencies: The surface sensitivity $S^{(\mathrm{surf})}$ as a function of the WG geometry, provided in this paper, the WG loss constant α as a function of the WG geometry, which must be experimentally determined for the specific WG type and technology, and the influence of this loss on the effective-index sensitivity $\mathbf{H}(\mathbf{r}'/\sigma_g), \mathbf{E}(\mathbf{r}'/\sigma_g)$, which translates $\Delta n_{\mathrm{e}}^{(\mathrm{surf})}$ into a measurable output signal, e.g., a current ([1,4,28]). If the linewidth of the light source is neglected, the overall system sensitivity can be expressed by the product of the surface sensitivity and the effective-index sensitivity

$$S_{\mathrm{sys}} = S^{(\mathrm{surf})} S_{\mathrm{e}}(\alpha). \tag{4}$$

The results presented in this paper can hence serve both as guidelines for identifying and selecting promising WG designs and as quantitative measures to determine the overall system sensitivity S_{sys} once the technology-dependent loss constant α and the architecture-specific relationship $S_{\mathrm{e}}(\alpha)$ are known.

Another aspect that may influence the overall performance of practical sensors is the analyte delivery to the sensor surface. In this context, narrow gaps or other high aspect-ratio voids tend to have less exchange with the surrounding liquid or might even not be accessible to large target objects such as cells or cellular vesicles. In these cases, simple strip WG geometries might exhibit advantages in comparison to more complex concepts featuring narrow strips and slots.

6 Summary

We have performed a comparative study of different WG types for application in label-free detection of chemical or biological analytes that specifically bind to functionalized WG surfaces. To this end, we have introduced the so-called surface sensitivity $S^{(\mathrm{surf})}$ as a general quality metric that is broadly applicable to surface layers of different thicknesses and refractive indices. This metric allows to optimize the WG design and to compare the fundamentally achievable sensor performances across different WG types and integration platforms. We specify optimized WG designs along with the corresponding surface sensitivities for the most common WG parameter combinations. We also introduce and

Optics EXPRESS

explain a computationally efficient recipe for applying the methodology to additional WG types that are not covered by our current study.

The following key findings result from our study: First, sensitivity trends can be explained by the interplay of a cladding asymmetry, the scaling laws of Maxwell's equations and the effect of field enhancement. Second, WG with optimum surface sensitivities are typically high and narrow, are fabricated on high-index contrast platforms and operated in TE polarization. Third, the surface sensitivity of highly sophisticated WG such as slot, double slot or sub-wavelength grating exceeds that of optimized strip WG by a factor of 2, and that of standard telecom strip WG by a factor of 4 if realistic feature size constraints are observed.

Appendices

A. Simulation parameters and mesh considerations

In the following, we describe the relevant parameters used for the simulation, which are performed with CST (Computer Simulation Technology GmbH) Microwave Studio. Modal fields of the WG are calculated in the frequency domain. Simulation boundaries in the cross section are perfectly absorbing. The computational mesh is tetrahedral, and an ultra-fine mesh is required around the WG core to obtain accurate field data within the surface layers. The field interaction factor Γ_{SL} is well-described by a linear approximation for t_{SL} smaller than the penetration depth, see Fig. 5(a). The surface sensitivity $S^{(surf)}$ can be reliably extracted from the derivative $\partial\Gamma_{SL}/\partial t_{SL}$ at $t_{SL}=0$ by a linear extrapolation of data within the first 10 nm, see Fig. 5(b). We have found that a mesh size of < 3 nm around the WG core region is sufficient. The choice of the mesh size in the cladding region was left to the program. The simulation area had a total size of 4 µm × 4 µm, where the bottom half of the simulation region was occupied by the 2 µm thick BOX. For WG significantly narrower than w_{opt}, the WG mode is not well-confined to the core anymore, and much larger simulation windows are required. These results were omitted from the discussion, but lie at non-practical WG geometries far away from any sensitivity optimum (Fig. 11). The refractive indices were assumed to be $n_{Si}=3.48$, $n_{H_2O}=1.33$, $n_{SiO_2}=1.44$ for NIR light and $n_{Si_3N_4}=2.01$, $n_{H_2O}=1.33$, $n_{SiO_2}=1.46$ for VIS light.

B. TM simulations WG geometries

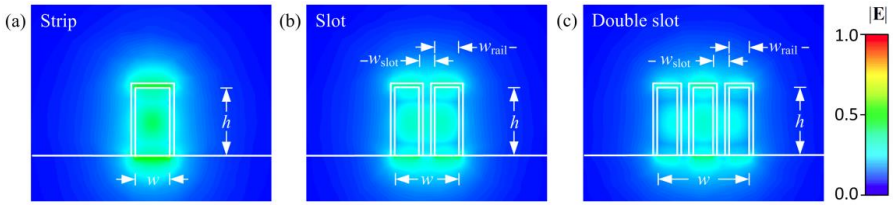

Fig. 11. Simulated electric field magnitudes of the fundamental quasi-TM mode in different WG types with height h and total width w. White contours mark the surface layer on the WG core, which is disregarded for the field calculation. Surface sensitivity generally benefits from large electric field strengths in the region of the surface layer. (a) Strip WG. Large portion of the fields is lost to the non-sensitive interface between core and BOX. (b),(c) Slot WG and double slot WG. In contrast to the TE-Mode (Fig. 3) there is no enhancement of the electrical field in the slot. The surface layer at the top of the core and in the slot experiences only moderate field strengths.

C. Propagation and sensitivity in waveguides

In the following discussion we assume a positive time dependence $\exp(j\omega t)$. The complex vectorial electric mode field $\mathbf{E}(\omega,\mathbf{r})$ and magnetic mode field $\mathbf{H}(\omega,\mathbf{r})$ depend on angular frequency ω and position vector $\mathbf{r}$. We assume dispersive dielectric and non-magnetic materials which could be periodic along z with a period a. The propagation constant is β

Optics EXPRESS

and the effective refractive index is $n_e = \beta/k_0$, $k_0 = \omega/c$. Within each region, all refractive indices and all other parameters of the material are assumed to be constant. The various regions are denoted by a subscript i.

Cross-sectional power, stored energy, and field confinement factor The cross-sectional power P associated with a guided mode is constant along the propagation direction z (unit vector $\mathbf{e}_z$) if losses are neglected. This is true even if the WG geometry varies periodically in the direction of propagation, as in the case of SWG WG. The cross-sectional power is expressed by the real part of the time-averaged complex Poynting vector in the direction of propagation, integrated over the WG cross-section A with $\mathrm{d}A = \mathrm{d}x\,\mathrm{d}y$,

$$P = \tfrac{1}{2}\iint_A \mathrm{Re}\left(\mathbf{E}\times\mathbf{H}^*\right)\cdot\mathbf{e}_z \mathrm{d}A\ . \tag{5}$$

The time-averaged stored energy per unit length is obtained by an integration of the modal electric and magnetic energy densities over a unit cell volume V [29]. The expression can be simplified with the help of the space-dependent material group index $n_g = n + \omega\partial n/\partial\omega$ and with the identity $\iiint \varepsilon_0\varepsilon_r |\mathbf{E}|^2\,\mathrm{d}V = \iiint \mu_0\mu_r |\mathbf{H}|^2\,\mathrm{d}V$ [29], where, for non-magnetic materials, $\mu_r = 1$. This leads to

$$W = \frac{1}{a}\iiint_V \frac{1}{4}\left(\frac{\partial}{\partial\omega}\left(\omega\varepsilon_0\varepsilon_r(\omega)\right)|\mathbf{E}|^2 + \frac{\partial}{\partial\omega}\left(\omega\mu_0\mu_r(\omega)\right)|\mathbf{H}|^2\right)\mathrm{d}V = \frac{1}{a}\iiint_V \frac{1}{2}\varepsilon_0 n n_g |\mathbf{E}|^2\,\mathrm{d}V. \tag{6}$$

Note that in this relation, W is a length-related energy density with unit J/m and denotes the ratio of stored energy within a WG section, e.g., a unit cell, and the length of the section, e.g., the unit cell length a. For z-invariant WG, the volume element is simply $\mathrm{d}V = a\,\mathrm{d}A$.

The field confinement factor $\Gamma_i^{(\mathrm{conf})}$ represents the ratio of the mode energy in a partial volume V_i related to the energy in the total volume V of a unit cell,

$$\Gamma_i^{(\mathrm{conf})} = \frac{\iiint_{V_i} n_i n_{g,i} |\mathbf{E}|^2\,\mathrm{d}V}{\iiint_V n n_g |\mathbf{E}|^2\,\mathrm{d}V} = \frac{W_i}{W}, \qquad \sum_i \Gamma_i^{(\mathrm{conf})} = 1. \tag{7}$$

Inside any partial volume V_i the refractive indices n_i and the associated group refractive index $n_{g,i}$ are constant. Summing the partial field confinement factors over all partial volumes results in 1 because $V = \sum_i V_i$

Variation theorem, effective group index, and field interaction factor The influence of small perturbations on the propagation constant $\Delta\beta$ can be examined by extending a variation theorem [29, Eq. (2.2.73)] for dielectric WG,

$$\Delta\beta = \frac{1}{4aP}\iiint_V \left(\Delta(\omega\varepsilon_0\varepsilon_r)|\mathbf{E}|^2 + \Delta(\omega\mu_0)|\mathbf{H}|^2\right)\mathrm{d}V, \tag{8}$$

where $\Delta(\omega\varepsilon_0\varepsilon_r)$ and $\Delta(\omega\mu_0)$ denote the perturbations. If we introduce only a frequency perturbation $\Delta\omega$ at a fixed dielectric profile, we find that the effective group index $n_{eg} = c\,\mathrm{d}\beta/\mathrm{d}\omega = n_e + \omega\,\mathrm{d}n_e/\mathrm{d}\omega$ is proportional to the ratio of the total energy W per unit cell length, and the power P,

$$n_{eg} = c\frac{\Delta\beta}{\Delta\omega} = \frac{1}{Z_0}\frac{\frac{1}{a}\iiint_V n n_g |\mathbf{E}|^2\,\mathrm{d}V}{\iint_A \mathrm{Re}\left(\mathbf{E}\times\mathbf{H}^*\right)\cdot\mathrm{e}_z\,\mathrm{d}A} = c\frac{W}{P}. \tag{9}$$

Optics EXPRESS

Here, $Z_0 = 1/(\varepsilon_0 c)$ is the free-space wave impedance. Combining Eq. (7) and Eq. (9) we define the field interaction factor $\Gamma_i = n_{\mathrm{eg}}/n_{\mathrm{g},i} \cdot \Gamma_i^{(\mathrm{conf})}$ of a certain WG region denoted by subscript *i*,

$$\Gamma_i = \frac{n_{\mathrm{eg}}}{n_{\mathrm{g},i}} \Gamma_i^{(\mathrm{conf})} = \frac{1}{Z_0} \frac{\frac{1}{a} n_i \iiint_{V_i} |\mathbf{E}|^2 \,\mathrm{d}V}{\iint_A \mathrm{Re}(\mathbf{E} \times \mathbf{H}^*) \cdot \mathrm{e}_z \,\mathrm{d}A}, \qquad n_{\mathrm{eg}} = \sum_i \Gamma_i n_{\mathrm{g},i}. \tag{10}$$

Note that the definition of $\Gamma_i = n_{\mathrm{eg}}/n_{\mathrm{g},i} \cdot \Gamma_i^{(\mathrm{conf})}$ is equivalent to the definition of $\Gamma_i = \Delta n_{\mathrm{e}}/\Delta n_i$ as the ratio of a local refractive index change Δn_i in region *i* and the associated change Δn_e of the effective modal index n_{e}, see next section. The effective group index can be expressed by a sum of the partial field interaction factors Γ_i multiplied with the respective material group indices. The field interaction factor is determined by the relative mode energy $\Gamma_i^{(\mathrm{conf})}$ in region i and by the time the mode dwells in that region, expressed by the effective modal group index n_{eg}, which can be larger than $n_{g,i}$. This is exploited in slow-light applications. For this case the field interaction factor can become larger than one.

Dielectric profile perturbation and definition of sensitivities If the 3D refractive index profile $n(\mathbf{r})$ with $\mathbf{r} = (x, y, z)$ of a WG is changed by a small amount $\Delta n(\mathbf{r}) \ll n(\mathbf{r})$, the associated change of the propagation constant $\Delta\beta$ can be calculated with the perturbation approach, Eq. (8). A small refractive index change $\Delta n(\mathbf{r})$ corresponds to a change $\Delta\varepsilon(\mathbf{r}) = \varepsilon_0 \cdot \Delta\varepsilon_r(\mathbf{r})$ in electric permeability, where [30]:

$$\Delta\varepsilon_{\mathrm{r}}(\mathbf{r}) = \big(n(\mathbf{r}) + \Delta n(\mathbf{r})\big)^2 - n^2(\mathbf{r}) \approx 2n(\mathbf{r})\Delta n(\mathbf{r}) \tag{11}$$

Typical index differences are in the order of $\Delta n = 0.1 \ldots 0.2$, which justifies the approximation in Eq. (11). For homogeneous sensing, only the refractive index of the aqueous cladding medium (partial volume $i = \mathrm{M}$) in the volume V_{M} changes by Δn_{M}, see Fig. 2(a),

$$\Delta n^{(\mathrm{hom})} = \begin{cases} \Delta n_{\mathrm{M}} & \text{inside } V_{\mathrm{M}} \\ 0 & \text{else} \end{cases}, \quad \Delta\varepsilon_{\mathrm{r}}^{(\mathrm{hom})} = \begin{cases} 2n_{\mathrm{M}}\Delta n_{\mathrm{M}} & \text{inside } V_{\mathrm{M}} \\ 0 & \text{else} \end{cases}. \tag{12}$$

For surface sensing, we consider a thin surface layer (partial volume $i = \mathrm{SL}$) of refractive index $\mathbf{H}(\mathbf{r}'/\sigma)$ around the WG core, which locally replaces the aqueous medium with refractive index n_{M}. The refractive index change is confined to and constant within the surface layer volume V_{SL} of thickness t_{SL}, see Fig. 2(b),

$$\Delta n^{(\mathrm{surf})} = \begin{cases} n_{\mathrm{SL}} - n_{\mathrm{M}} & \text{inside } V_{\mathrm{SL}} \\ 0 & \text{else} \end{cases}, \quad \Delta\varepsilon_{\mathrm{r}}^{(\mathrm{surf})} = \begin{cases} 2n_{\mathrm{M}}\left(n_{\mathrm{SL}} - n_{\mathrm{M}}\right) & \text{inside } V_{\mathrm{SL}} \\ 0 & \text{else} \end{cases}. \tag{13}$$

We introduce the perturbations defined by Eq. (11)-(13) at a fixed frequency into the general variation theorem of Eq. (8). Since the permittivity perturbation is limited to V_{SL} in the case of surface sensing, the change of the propagation constant $\Delta\beta^{(\mathrm{surf})}$ can be expressed as

$$\Delta\beta^{(\mathrm{surf})} = \frac{\omega}{4aP} \iiint_V \varepsilon_0 \Delta\varepsilon_{\mathrm{r}}^{(\mathrm{surf})}(\mathbf{r}) |\mathbf{E}|^2 \,\mathrm{d}V = k_0 \frac{1}{Z_0} \frac{\frac{1}{a} n_{\mathrm{M}} \iiint_{V_{\mathrm{SL}}} |\mathbf{E}|^2 \,\mathrm{d}V}{\iint_A \mathrm{Re}(\mathbf{E} \times \mathbf{H}^*) \cdot \mathbf{e}_z \,\mathrm{d}A} \Delta n^{(\mathrm{surf})} = k_0 \Gamma_{\mathrm{SL}} \Delta n^{(\mathrm{surf})}. \tag{14}$$

Similarly, for homogeneous sensing, the permittivity perturbation is limited to V_{M}, the change of the propagation constant $\Delta\beta^{(\mathrm{hom})}$ is obtained from Eq. (14) by replacing the superscript "(surf)" by "(hom)" and by integrating over V_{M} instead of V_{SL}. A direct link

Optics EXPRESS

between $\Delta\beta$ and the field interaction factor Γ Eq. (10) is observed. For a sensor it is important how the effective modal index n_e changes with respect to the local refractive index perturbation Δn. For surface (homogeneous) sensing, we have

$$\Delta n_{\mathrm{e}}^{(\mathrm{surf})} = \Gamma_{\mathrm{SL}}\,\Delta n^{(\mathrm{surf})},\ \Delta n_{\mathrm{e}}^{(\mathrm{hom})} = \Gamma_{\mathrm{M}}\,\Delta n^{(\mathrm{hom})}. \tag{15}$$

For our sensitivity analyses, we calculate the field interaction factors for a WG by numerically calculated mode fields. We directly define the homogeneous sensitivity $S^{(\mathrm{hom})}$ to be identical with the corresponding field interaction factor Γ_{M}, measuring the impact of the refractive index perturbation Δn_{M} within $|\mathbf{E}'(\mathbf{r})|/|\mathbf{H}'(\mathbf{r})| = 1/\sigma_{\mathrm{n}}\cdot|\mathbf{E}(\mathbf{r})|/|\mathbf{H}(\mathbf{r})|$ on the effective modal index n_{e}. We further define the surface sensitivity $S^{(\mathrm{surf})}$ as the derivative of the field interaction factor Γ_{SL} (surface layer volume V_{SL}) with respect to the surface layer thickness t_{SL},

$$\begin{aligned} S^{(\mathrm{hom})} &= \Gamma_{\mathrm{M}} = \frac{\Delta n_{\mathrm{e}}^{(\mathrm{hom})}}{\Delta n_{\mathrm{M}}}, \\ S^{(\mathrm{surf})} &= \left.\frac{\partial \Gamma_{\mathrm{SL}}}{\partial t_{\mathrm{SL}}}\right|_{t_{\mathrm{SL}}=0} = \frac{1}{(n_{\mathrm{SL}} - n_{\mathrm{M}})}\frac{\partial}{\partial t_{\mathrm{SL}}}\Delta n_{\mathrm{e}}^{(\mathrm{surf})}\Big|_{t_{\mathrm{SL}}=0}. \end{aligned} \tag{16}$$

D. Scaling laws of Maxwell's equations

We consider dielectric media, which are assumed to be lossless, isotropic, linear and non–magnetic at the (angular) frequencies $\omega = 2\pi f$ of interest. The vacuum speed of light is denoted by c. The (real) relative permittivity $\varepsilon_{\mathrm{r}}(\mathbf{r})$ is linked to the refractive index n by $\varepsilon_{\mathrm{r}} = n^2$. Reshaping Maxwell's equations for harmonic solutions of the form $\mathbf{E}(t,\mathbf{r}) = \mathbf{E}(\mathbf{r})e^{\mathrm{j}\omega t}$ and $\mathbf{H}(t,\mathbf{r}) = \mathbf{H}(\mathbf{r})e^{\mathrm{j}\omega t}$, we find the wave equations for the magnetic and electric fields [31],

$$\left(\mathrm{curl}\,\frac{1}{\varepsilon_{\mathrm{r}}(\mathbf{r})}\,\mathrm{curl}\right)\mathbf{H}(\mathbf{r}) = \frac{\omega^2}{c^2}\mathbf{H}(\mathbf{r}),\ \left(\frac{1}{\varepsilon_{\mathrm{r}}(\mathbf{r})}\,\mathrm{curl}\,\mathrm{curl}\right)\mathbf{E}(\mathbf{r}) = \frac{\omega^2}{c^2}\mathbf{E}(\mathbf{r}). \tag{17}$$

Equation (17), together with boundary conditions, defines an eigenvalue problem, where ω is the angular eigenfrequency and $\mathbf{H}(\mathbf{r})$ and $\mathbf{E}(\mathbf{r})$ are the corresponding eigenfunctions of the wave equations.

Scaling the geometry. If the geometry of the WG is scaled by a factor $\sigma_g > 0$, i.e., $\mathbf{r}' = \mathbf{r}\sigma_g$, $\varepsilon_{\mathrm{r}}'(\mathbf{r}) = \varepsilon_{\mathrm{r}}(\mathbf{r}/\sigma_g)$ and $\mathrm{curl}' = \mathrm{curl}/\sigma_g$, the magnetic field equation Eq. (17) can be expressed as

$$\left(\mathrm{curl}'\,\frac{1}{\varepsilon_{\mathrm{r}}'(\mathbf{r}')}\,\mathrm{curl}'\right)\mathbf{H}\left(\frac{\mathbf{r}'}{\sigma_{\mathrm{g}}}\right) = \frac{\omega'^2}{c^2}\mathbf{H}\left(\frac{\mathbf{r}'}{\sigma_{\mathrm{g}}}\right),\ \omega' = \frac{\omega}{\sigma_{\mathrm{g}}}. \tag{18}$$

An equivalent equation holds for the electric field. For the scaled WG, we find the same eigenfunctions $\mathbf{H}(\mathbf{r}'/\sigma_g), \mathbf{E}(\mathbf{r}'/\sigma_g)$ as in Eq. (17); we only have to scale the argument $\mathbf{r} = \mathbf{r}'/\sigma_g$ together with the associated angular eigenfrequency $\omega' = \omega\sigma_\omega$ with $\sigma_\omega = 1/\sigma_{\mathrm{g}}$ [32].

Scaling the refractive index. If we know the solutions for a WG with dielectric structure $\varepsilon_{\mathrm{r}}(\mathbf{r}) = n^2(\mathbf{r})$, and we look for the results of a WG with $\varepsilon_{\mathrm{r}}'(\mathbf{r}) = n'^2(\mathbf{r})$, where the refractive index is scaled everywhere with a real constant σ_{n} according to $n'(\mathbf{r}) = n(\mathbf{r})\sigma_{\mathrm{n}}$ and $\varepsilon_{\mathrm{r}}'(\mathbf{r}) = \varepsilon(\mathbf{r})\sigma_{\mathrm{n}}^2$, we find for the magnetic field equation

$$\left(\mathrm{curl}\,\frac{1}{\varepsilon_{\mathrm{r}}'(\mathbf{r})}\,\mathrm{curl}\right)\mathbf{H}(\mathbf{r}) = \frac{\omega'^2}{c^2}\mathbf{H}(\mathbf{r}),\ \omega' = \frac{\omega}{\sigma_{\mathrm{n}}}. \tag{19}$$

Optics EXPRESS

The eigenfunctions $\mathbf{H}(\mathbf{r})$ remain unchanged, but the associated angular eigenfrequencies are scaled to $\omega' = \omega\sigma_\omega$ with $\sigma_\omega = 1/\sigma_n$ [32]. The electric field equation can be formulated equivalently. Note that the relative ratio of the electric-field eigenfunction $\mathbf{E}(\mathbf{r})$ and the magnetic field eigenfunction $\mathbf{H}(\mathbf{r})$ must be scaled by a factor of $1/\sigma_n$ to still satisfy the Maxwell's curl equations that link the electric to the magnetic field and vice versa, $|\mathbf{E}'(\mathbf{r})|/|\mathbf{H}'(\mathbf{r})| = 1/\sigma_n \cdot |\mathbf{E}(\mathbf{r})|/|\mathbf{H}(\mathbf{r})|$.

Scaling the geometry and the refractive indices. If we scale both the geometry with σ_g and the refractive indices with σ_n, e.g. $\mathbf{r}' = \mathbf{r}\sigma_g$, $\varepsilon_r'(\mathbf{r}) = \varepsilon_r(\mathbf{r}/\sigma_g)\cdot\sigma_n^2$ and $\mathrm{curl}' = \mathrm{curl}/\sigma_g$, we find for the magnetic field equation

$$\left(\mathrm{curl}'\frac{1}{\varepsilon_r'(\mathbf{r}')}\mathrm{curl}'\right)\mathbf{H}\left(\frac{\mathbf{r}'}{\sigma}\right) = \frac{\omega'^2}{c^2}\mathbf{H}\left(\frac{\mathbf{r}'}{\sigma}\right), \omega' = \frac{\omega}{\sigma_n\sigma_g}. \tag{20}$$

The eigenfunction $\mathbf{H}(\mathbf{r}'/\sigma)$ remains unchanged with scaled arguments, and the associated angular eigenfrequencies are scaled to $\omega' = \omega\sigma_\omega$ with $\sigma_\omega = 1/(\sigma_n\sigma_g)$. The electric field equation can be written equivalently. Note that the relative ratio of the electric-field eigenfunction $\mathbf{E}(\mathbf{r})$ and the magnetic field eigenfunction $\mathbf{H}(\mathbf{r})$ must be scaled by a factor of $1/\sigma_n$ to still satisfy the Maxwell's curl equations that link the electric to the magnetic field and vice versa, $|\mathbf{E}'(\mathbf{r})|/|\mathbf{H}'(\mathbf{r})| = 1/\sigma_n \cdot |\mathbf{E}(\mathbf{r})|/|\mathbf{H}(\mathbf{r})|$.

If the frequency ω remains unchanged the geometry and the refractive indices must scale inversely with $\sigma_g = 1/\sigma_n$, so that the eigenfunction $\mathbf{H}(\mathbf{r}'/\sigma)$ remains unchanged. The relative ratio of the electric-field eigenfunction $\mathbf{E}(\mathbf{r})$ and the magnetic field eigenfunction $\mathbf{H}(\mathbf{r})$ must again be scaled by a factor of $1/\sigma_n$ to $|\mathbf{E}'(\mathbf{r})|/|\mathbf{H}'(\mathbf{r})| = 1/\sigma_n \cdot |\mathbf{E}(\mathbf{r})|/|\mathbf{H}(\mathbf{r})|$.

Impact on homogeneous and surface sensing. For comparison of Si_3N_4 WG in the VIS and Si WG in the NIR, we are especially interested in the impact of simultaneously scaling the refractive indices, the geometry and the frequency on the homogeneous sensitivity $S^{(\mathrm{hom})} = \Gamma_M$ and the surface sensitivity $S^{(\mathrm{surf})} = \partial\Gamma_{SL}/\partial t_{SL}$. The different scaling operations and results are summarized in Table 1.

In the case of scaling the geometry and accordingly the frequency with remaining refractive indices, the numerical value of the ratio of the integrals in Γ_{SL} remains constant, see Eq. (10) and therefore the homogeneous sensitivity remains unchanged, $S^{(\mathrm{hom})'} = S^{(\mathrm{hom})}$. However, due to the scaled surface layer thickness $t_{SL}' = \sigma_g t_{SL}$, the surface sensitivity scales according to $S^{(\mathrm{surf})'} = \partial\Gamma_{SL}/\partial t_{SL}' = S^{(\mathrm{surf})}/\sigma_g = S^{(\mathrm{surf})}\sigma_\omega$, see Table 1, line (1).

In the case of scaling the refractive indices and accordingly the frequency with remaining geometry, the numerical value of the ratio of the integrals in Γ_{SL} remains constant, as the ratio of the eigenfunctions scales with $|\mathbf{E}'(\mathbf{r})|/|\mathbf{H}'(\mathbf{r})| = 1/\sigma_n \cdot |\mathbf{E}(\mathbf{r})|/|\mathbf{H}(\mathbf{r})|$. Therefore the homogeneous and surface sensitivity do not change, see Table 1, line (2).

If the geometry and the refractive index of the WG is scaled, the frequency must be scaled accordingly, with $\sigma_\omega = 1/(\sigma_n\sigma_g)$. Under the premises of the scaled ratio of the eigenfunctions, the numerical value of the ratio of the integrals in Γ_{SL} remains constant, and therefore $S^{(\mathrm{surf})}$ scales with $1/\sigma_g$ due to the scaled surface layer thickness $t_{SL}' = \sigma_g t_{SL}$, see Table 1, line (3).

In the case of scaling the refractive indices with constant frequency, the geometry has to be scaled according to $\sigma_g = 1/\sigma_\omega$ in order to keep the eigenfunctions but with scaled ratio. The homogeneous sensitivity is again not changing, while the surface sensitivity scales with $\sigma_n\sigma_\omega = 1/\sigma_g$, see Table 1, line (4).

Optics EXPRESS

Table 1. Impact of scaling the frequency, geometry and refractive index on the homogeneous and surface sensitivities.

Geometry $\mathbf{r}'$	Refr. index n'	Frequency ω'	Scaling law	Hom. sensitivity $S^{(\mathrm{hom})\prime}$	Surf. sensitivity $S^{(\mathrm{surf})\prime}$	
$\mathbf{r}\sigma_{\mathrm{g}}$	n	$\omega\sigma_{\omega}$	$\sigma_{\omega}=1/\sigma_{\mathrm{g}}$	$S^{(\mathrm{hom})}$	$S^{(\mathrm{surf})}/\sigma_{\mathrm{g}}$	(1)
$\mathbf{r}$	$n\sigma_{\mathrm{n}}$	$\omega\sigma_{\omega}$	$\sigma_{\omega}=1/\sigma_{\mathrm{n}}$	$S^{(\mathrm{hom})}$	$S^{(\mathrm{surf})}$	(2)
$\mathbf{r}\sigma_{\mathrm{g}}$	$n\sigma_{\mathrm{n}}$	$\omega\sigma_{\omega}$	$\sigma_{\omega}=1/(\sigma_{\mathrm{n}}\sigma_{\mathrm{g}})$	$S^{(\mathrm{hom})}$	$S^{(\mathrm{surf})}/\sigma_{\mathrm{g}}$	(3)
$\mathbf{r}\sigma_{\mathrm{g}}$	$n\sigma_{\mathrm{n}}$	ω	$\sigma_{\mathrm{g}}=1/\sigma_{\mathrm{n}}$	$S^{(\mathrm{hom})}$	$S^{(\mathrm{surf})}/\sigma_{\mathrm{g}}$	(4)

Funding

Alfried Krupp von Bohlen und Halbach Foundation; European Research Council (ERC Starting Grant 'EnTeraPIC', number 280145); Robert Bosch GmbH (Renningen, Germany); Karlsruhe School of Optics and Photonics (KSOP).

Acknowledgments

We acknowledge support by the Karlsruhe Nano Micro Facility (KNMF), a Helmholtz Research Infrastructure at Karlsruhe Institute of Technology.