Principles for Sensitive and Robust Biomolecular Interaction Analysis: The Limits of Detection and Resolution of Diffraction-Limited Focal Molography

Andreas Frutiger,[1,†] Yves Blickenstorfer,[1,†] Silvio Bischof,[1] Csaba Forró,[1] Matthias Lauer,[2] Volker Gatterdam,[1] Christof Fattinger,[2,*] and János Vörös[1]

[1] *Laboratory of Biosensors and Bioelectronics, Institute of Biomedical Engineering, ETH Zürich, 8092 Zürich, Switzerland*

[2] *Roche Pharma Research and Early Development, Roche Innovation Center Basel, 4070 Basel, Switzerland*

Label-free biosensors enable the monitoring of biomolecular interactions in real time, which is key to the analysis of the binding characteristics of biomolecules. While refractometric optical biosensors such as surface plasmon resonance (SPR) are sensitive and well-established, they are susceptible to any change of the refractive index in the sensing volume caused by minute variations in composition of the sample buffer, temperature drifts, and most importantly nonspecific binding to the sensor surface in complex fluids such as blood. The limitations arise because refractometric sensors measure the refractive index of the entire sensing volume. Conversely, diffractometric biosensors–for example, focal molography–only detect the diffracted light from a coherent assembly of analyte molecules. Thus any refractive index distribution that is noncoherent with respect to this molecular assembly does not add to the coherent signal. This makes diffractometric biosensors inherently robust and enables sensitive measurements without reference channels or temperature stabilization. The coherent assembly is generated by selective binding of the analyte molecules to a synthetic binding pattern–the mologram. Focal molography has been introduced theoretically [C. Fattinger, Phys. Rev. X **4**, 031024 (2014)] and verified experimentally [V. Gatterdam, A. Frutiger, K.-P. Stengele, D. Heindl, T. Lübbes, J. Vörös, and C. Fattinger, Nat. Nanotechnol. **12**, 1089 (2017)] in previous papers. However, further understanding of the underlying physics and a diffraction-limited readout is needed to unveil its full potential. This paper introduces refined theoretical models, which can accurately quantify the amount of biological matter bound to the mologram from the diffracted intensity. In addition, it presents measurements of diffraction-limited molographic foci, i.e., Airy discs. These improvements enable us to demonstrate a resolution in real-time binding experiments comparable to the best SPR sensors without the need for temperature stabilization or drift correction and to detect low-molecular-weight compounds label free in an endpoint format. The presented experiments exemplify the robustness and sensitivity of the diffractometric sensor principle.

I. INTRODUCTION

Diffractive lenses or focusing holograms proposed by Augustin-Jean Fresnel have been known to humanity for 200 years and have experienced applicability in various fields such as photography [1], telescopes [2], spectroscopy [3], optical tweezers [4], and x-ray lenses [5]. Yet, nature discovered this principle much earlier. In certain organisms, biomolecules are assembled to create a focusing hologram for image formation in the eye [6]. Recently, thanks to advances in photolithography [7] and nonfouling photoactivatable surface chemistries in particular [8], it has become possible to apply the holographic principle to highly sensitive molecular detection. These molecular holograms can be used for real-time label-free detection of molecules by molecular recognition in complex samples [7]. This enables the monitoring of biomolecular interactions, which is key to the analysis of binding characteristics of biomolecules in a broad range of applications [9]. To date, the biosensing field has been dominated by refractometric optical sensors and most prominently, thanks to their high surface sensitivity, by techniques based on evanescent waves, such as surface plasmons or dielectric waveguide modes [10,11]. These analytical tools are well established to perform label-free binding assays with high sensitivity and low limits of detection (0.1–1 pg/mm^2) [12,13]. Refractometric biosensors [e.g.,

---

*christof.fattinger@roche.com

†A. Frutiger and Y. Blickenstorfer contributed equally to this work.

surface plasmon resonance (SPR)] measure the refractive index change upon receptor-ligand binding in the vicinity of the sensor surface. However, they are susceptible to any change in refractive index within the evanescent field caused by fluctuations in temperature, buffer composition, and most importantly, nonspecific binding to the sensor surface. This inherent feature of refractometric sensors often manifests as drift and causes jumps in the sensor signal—e.g., during sample exchange [14]. Therefore, these sensors typically operate continuously and in well-defined buffers because measurements in serum or plasma exhibit artifacts and stability problems.

The mentioned limitations arise from the fundamental inability of a refractometric sensor to distinguish between the molecules of the target analyte and all other influences that affect the refractive index of the sensing volume. Even for evanescent field sensors, the sensing volume is still enormous compared to the small volume of the target molecules. This makes it virtually impossible to compensate for these influences—even with a differential measurement [14]. There is, however, a physical phenomenon that only measures the refractive index difference between the target molecules and the refractive index of their displacement volume, namely the scattering of light. This rejects most of the influences from temperature and buffer changes by measuring only the refractive index contrast in the nanoenvironment of the binding events.

It is a common belief in the biosensing community that the most sensitive detection methods for label-free biomolecular interaction analysis in real time are based on the refractometric sensing principle. In this context, it is sometimes not believed that biomolecular interactions can be detected with high sensitivity by the scattering of light [7,15,16]. The single-molecule detection method interferometric scattering microscopy (iSCAT) [15] is based on interferometric detection of scattering. It demonstrates exquisite sensitivity for the analysis of biomolecular interactions through scattering. The single-molecular sensitivity of iSCAT allows the analysis of the heterogeneity in an ensemble of a molecular species. Yet, in other applications, it is sufficient to determine an averaged quantity of the ensemble. The accurate quantification of a biomarker concentration falls in this category. In such a measurement, single-molecular sensitivity can be beneficial, but is neither required nor should its importance be overestimated. In the case of iSCAT, the single-molecular sensitivity comes at the cost of a relatively complicated setup, since the noise has to be sufficiently low to detect every single protein independently. In addition, iSCAT cannot distinguish between different types of similarly sized proteins. Therefore, for measuring in complex fluids, iSCAT is currently limited by nonspecific binding similar to refractometric sensors. In both cases, the specificity is mostly determined by the choice of surface chemistry [15]. A protein-repellent (nonfouling) surface

chemistry is not enough to measure in complex samples since there is always a significant number of defects in the adlayer, and therefore of nonspecific binding to the sensor [17,18].

Conversely, molecular holograms are diffractometric sensors, which offer an additional mechanism to reduce the effect of nonspecific binding. This is achieved by constraining the specific binding to a coherent scattering system—i.e., a molecular hologram. The blueprint of this hologram is encoded into the surface adlayer. Namely, the recognition sites compose a coherent binding pattern—i.e., a mologram. The constructive interference of the scattered fields relates all bound analyte molecules and yields a quadratic scaling of the measured intensity with respect to the analyte number. On the other hand, the scattered field of randomly bound background molecules interferes with equal probability either constructively or destructively. Therefore, only its variance affects the coherent signal and thus it scales linearly with particle number. This implies that the nonspecific binding is efficiently suppressed for a sufficiently large ensemble of analyte molecules. In addition, other random scattering (noise) sources experience the same repression with respect to the signal. Therefore, it is considerably simpler to coherently detect an ensemble of molecules rather than to count them individually. In other words, a diffractometric sensor is inherently self-referencing on the submicron length scale of regions of constructive and destructive interference.

A pure diffractometric biosensor consists of coherently arranged binding sites without any diffractive power, or in other words, a massless affinity modulation [16]. A sensor with these properties is extremely robust and only produces a signal in the presence of the analyte [7]. The conception of an affinity modulation that has no optical modulation is reasonable to physicists. However, advanced molecular engineering capabilities are required to achieve this experimentally, since most nanolithographic techniques, such as imprinting or lift-off techniques, produce an inherent optical modulation due to coherent defects in the affinity keys [19,20]. Such an optical modulation reduces the robustness of the sensor. Due to the excluded volume, diffraction at such defects depends on the refractive index of the solution. Even worse, these coherent defects will give rise to an affinity modulation for background molecules, severely compromising the rejection of nonspecific binding. This fact makes the implementation of sensitive diffractometric biosensors interdisciplinary and demanding.

Focal molography is the first diffractometric sensor that may exhibit resolutions in direct binding assays that are comparable to the best refractometric sensors [7,16]. Briefly, in focal molography, the mologram is situated on a high-refractive-index slab waveguide and illuminated by the fundamental TE mode (Fig. 1). When the affinity modulation is exposed to a biological sample, the analyte
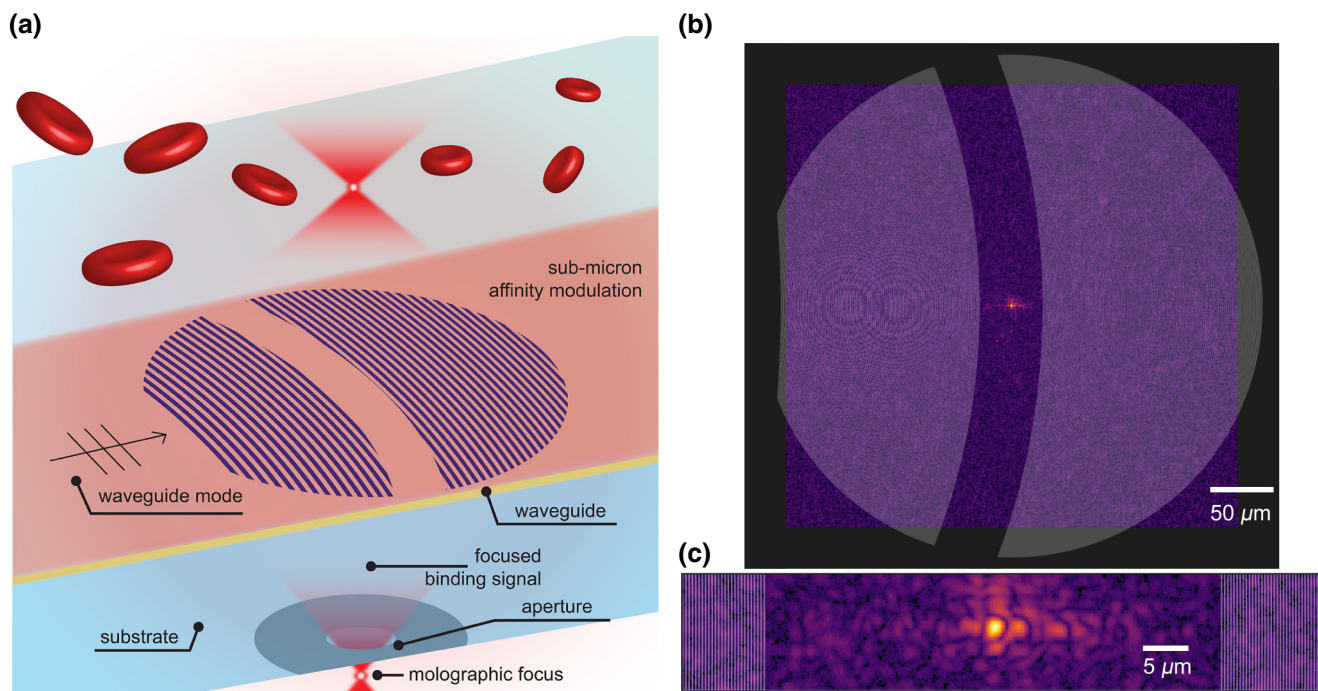
FIG. 1.   Focal molography incorporates the four essentials of a highly sensitive diffractometric biosensor (a) A submicrometer affinity modulation formed by specific binders is exposed to a biological sample (e.g., blood). The mode of a high-refractive-index waveguide provides perfect dark-field illumination of the molecules in the vicinity of the sensor surface and enhances the light intensity. The shape of the pattern acts as a diffractive lens, which concentrates the diffracted signal into a focal spot, whereas the background intensity is diluted over the entire solid angle. For efficient spatial filtering, the aperture of the optical system is matched to that of the mologram. (b) The molographic pattern with a recorded focal spot superimposed (bottom view) and enlarged in (c). The Airy disk of the mologram sticks out from the speckled background and monitors the binding activity of billions of recognition sites on an area that is nearly five orders of magnitude larger than the tiny focal spot. The image is nonlinearly scaled ($\gamma = 0.25$) to better visualize the speckles.

binds to the mologram. This induces an optical grating that diffracts light from the guided TE mode into a diffraction-limited focal spot. From the diffraction efficiency, the presence of molecules at the interaction sites is quantified. We will now outline the four pillars that need to be fulfilled for diffractometric biosensors to be highly sensitive and robust. Previously reported diffractometric concepts for biomolecular interaction analysis [21–25] do not incorporate all four pillars. This is the reason why they are limited in sensitivity or robustness. The concept—focal molography—was introduced with all the essentials necessary for highly sensitive and robust diffractometric biomolecular interaction analysis [16] (Fig. 1). (i) The first essential is a submicrometer affinity modulation on a nonfouling monolithic surface layer for efficient rejection of nonspecific binding. In the first demonstration of focal molography, this was achieved by the reactive immersion lithography (RIL) process, which produces an affinity modulation consisting of active regions (ridges) and passive regions (grooves) on a nonfouling brushed copolymer adlayer [7]. Ideally, this affinity modulation should be massless, as explained below as part of the second essential. (ii) In focal molography, the mologram is situated on an asymmetric high-refractive-index slab waveguide, which provides the

second essential—a proper dark-field illumination of the coherent affinity modulation. The two-dimensional light sheet of the guided TE mode only illuminates the first 100 nm of the sample solution close to the surface. This avoids any background scattering from particles in the sample solution that are further away. In this context, it becomes clear why the affinity modulation should be massless in the ideal case. Any optical modulation results in intrinsic diffraction and thus deteriorates the dark-field detection. (iii) The high-refractive-index waveguide also provides the third essential, namely an increase in the field intensity at the scatterer location. In other words, guiding to free-space mode coupling is more efficient than free-space to free-space coupling for a given amount of coherent biological matter [26]. (iv) The fourth and last essential is the observation of the diffracted signal in the far field of the mologram defined in terms of Fraunhofer distance. This near- to far-field transformation increases the SNR due to the directed character of the diffracted signal compared to the distributed background from random scatterers. However, the Fraunhofer distance of a linear diffraction grating with a length of 400 μm is roughly 50 cm for visible wavelengths. By using a lens, the far field can be observed much closer to the sensor surface. In our case, the mologram

itself performs the near- to far-field transformation by focusing the intensity holographically onto an Airy disk only a few hundred microns away from the sensor surface. The binding information of billions of recognition sites on an area five orders of magnitude larger is, therefore, contained in the tiny Airy disk [Figs. 1(b) and 1(c)]. This enables compact technical realizations of diffractometric biosensors.

From another viewpoint, focal molography can also be seen as a chemical radio, at least in the eyes of a physical chemist [27]. The transmission of radio signals is based on the modulation of an rf carrier signal and the subsequent demodulation at the receiver. Molography applies this principle at optical frequencies to the transmission of chemical signals. Molecules recognize the affinity modulation in the mologram and interact with it. The molecular interaction renders a coherent molecular pattern in the form of a diffractive lens. This diffractive lens modulates the momentum of the guided mode with the spatial frequency of the mologram. The demodulation in k-space is performed by Fourier optics and the molographic signal is separated from the carrier wave in the focal plane of the lens.

Recently, the first experimental measurements with focal molography were performed using the nondiffraction-limited ZeptoReader (Zeptosens AG), substantially compromising the fourth pillar [7]. The emphasis of that publication was the demonstration of the robust operation of focal molography and its insensitivity to nonspecific binding in complex samples rather than achieving high sensitivity. Nevertheless, a real-time detection limit of 5 pg/mm^2 was achieved and we made the projection that noise levels can be reduced by at least two orders of magnitude by observing the molographic signal in a reader capable of resolving the diffraction-limited focus.

The aim of this contribution is, therefore, to characterize the molographic signal in the proper far field and thus to explore the resolution limits of diffraction-limited focal molography for massless affinity modulations experimentally as well as to refine some of the theoretical concepts. First, we introduce a measurement setup that allows static and real-time measurements of diffraction-limited molographic focal spots. Second, we present a semi-analytical framework with which the field distribution in the focal spot can be accurately computed by summation of the scattered fields of individual molecules (dipole scatterers) on the waveguide surface. Third, we demonstrate that the simulated and experimental field distributions are in excellent agreement with each other and that the Airy disk dimensions of the mologram are consistent with the Airy disk of a diffraction-limited lens. Fourth, we show that our synthetic holograms produce diffraction-limited focal spots at least up to mologram diameters of 400 μm on high-refractive-index slab waveguides. Furthermore, it is verified that the analytical predictions for the intensity

of the focal spot through coupled mode theory made by Fattinger [16] coincide with Rayleigh scattering and can accurately describe the experimentally measured intensities for a given amount of coherent biological matter. Fifth, we address the relevance of different background sources that can scatter intensity into the focal plane and produce an inhomogeneous speckle pattern that limits the resolution and accuracy of the molographic measurement. Based on this discussion, a figure of merit for molography is formulated that allows direct comparison of different molographic arrangements with different waveguides and mologram sizes. Next, we apply our theoretical insights to calculate limits of detection for molography on Ta$_2$O$_5$-slab waveguides for endpoint and real-time detection. These predictions are then verified experimentally. In particular, the low-molecular-weight ($<$300 Da) molecule vitamin B7, commonly known as biotin, is detected label-free by molography in an endpoint measurement without any calibration of the sensor. These biotin molograms are most likely the faintest man-made holograms that have ever been measured. In addition, we demonstrate that it is possible to fabricate an affinity modulation without a detectable optical modulation and use it to acquire real-time binding curves with 500 pM streptavidin (SAv) in buffer that exhibit baseline noise levels below 100 fg/mm^2 over 20 min, which are comparable to those in the best commercially available label-free detection method [28]. However, while the commercial system is temperature stabilized to 0.01 °C, we achieve this stability without any temperature control, demonstrating the potential of focal molography for extremely sensitive and robust, real-time, label-free molecular interaction analysis.

II. DIFFRACTION-LIMITED MOLOGRAPHY

A. Measurement of foci formed by diffraction-limited molograms

The realization and quantification of diffraction-limited molographic experiments incorporates the design of a microscope, waveguide coupler, and fluidics as well as the development of appropriate algorithms for evaluation of the acquired images. The setup (MoloReader), which we developed for this purpose, is displayed in Fig. 2(d) as well as in Fig. 15 and a functional schematic is shown in Fig. 2(a). The setup allows the coupling of a TE-polarized He-Ne laser beam (632.8-nm wavelength) via a grating coupler (coupling angle −10.6°, period 318 nm, length 500 μm) into a thin-film optical waveguide (145-nm-thick Ta$_2$O$_5$) on a glass substrate (D263 Schott, 700 μm). Molecules located on the waveguide are illuminated by the evanescent field of the fundamental guided TE mode ($N = 1.814$, penetration depth is 82 nm) in a dark-field manner [Fig. 2(b)]. For most of the experiments presented in this paper, the molograms are composed of binding sites for the protein molecule SAv. The molograms on the
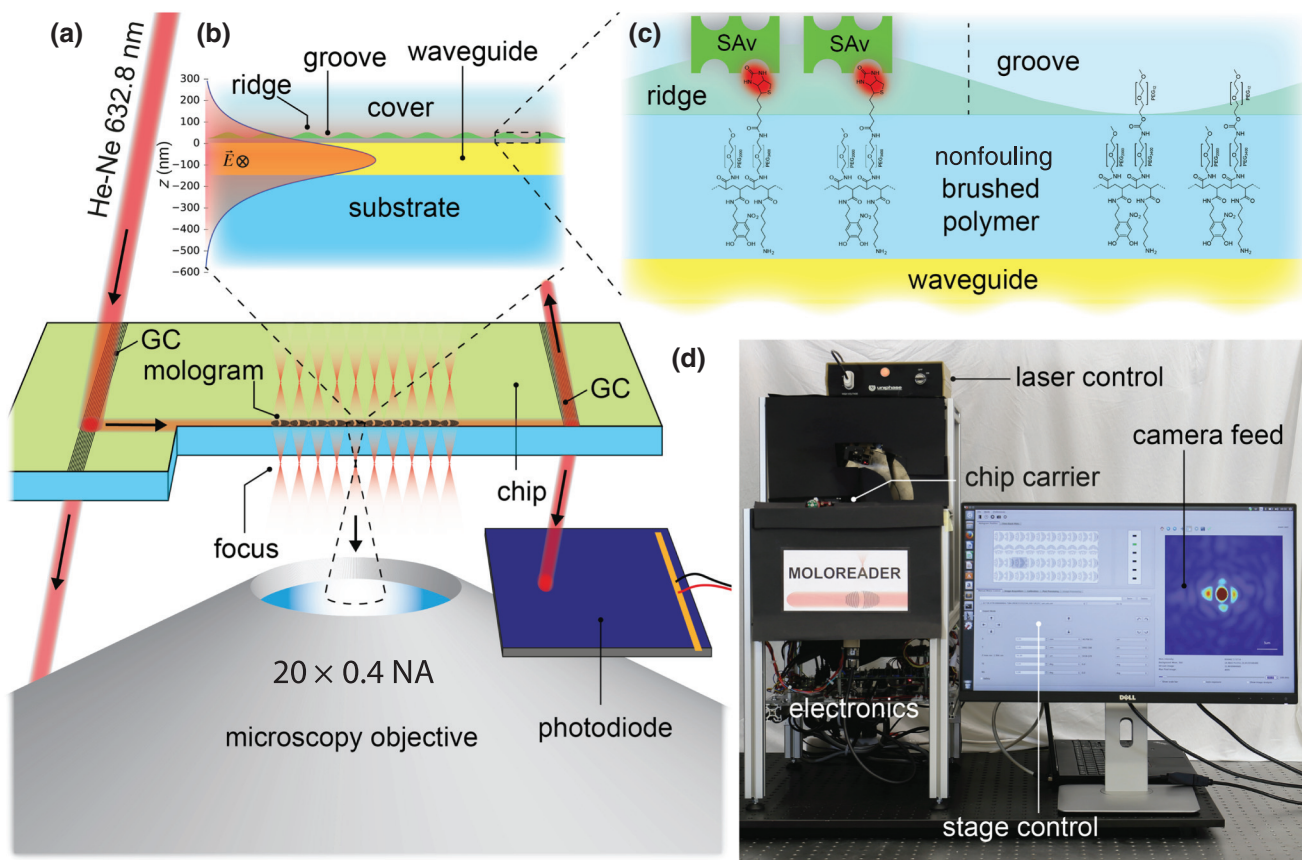
FIG. 2.   Detection of diffraction-limited molographic spots. (a) Schematic representation of the setup for the experimental demonstration of diffraction-limited molography: Light of a He-Ne laser is coupled into the fundamental TE mode of a high-refractive-index slab waveguide via a grating coupler (GC). The light propagates along the waveguide and is scattered at the molecules of interest (SAv) that form a focusing hologram. These molecules are captured from solution by binding to a coherent affinity modulation on top of a nonfouling polymer layer that is fabricated by reactive immersion lithography [7]. The molographic signal (the intensity of the focal point) of one of ten molograms in a row is collected by a microscope objective and captured by a camera. The total power in the waveguide is monitored with a photodiode, which measures the light diffracted by a physical out-coupling grating which is etched into the waveguide. (b) Enlarged view of the waveguide with the field profile of the TE mode. (c) In active regions (ridges), immobilized receptors (biotin) capture the protein of interest (SAv) and form a coherent assembly, whereas inert regions backfilled with polyethylene glycol (grooves) do not recognize the protein. (d) The experimental setup (MoloReader) in operation. The microscopy objective is focused on the focal plane of the mologram.

waveguide consist of alternating ridges, where SAv binds to immobilized biotin (molecular weight 227 g/mol); and grooves, which are backfilled with an inert polyethylene glycol (PEG) molecule (MeO-dPEG$_{12}$, molecular weight 570 g/mol). The PEG backfilling is performed to obtain a massless affinity modulation, as well as for blocking of free amine groups. As a side note, despite the higher molecular mass of the PEG, these molograms exhibit a nondetectable mass modulation (Video 3), most likely because PEG is the more flexible molecule and has a smaller refractive-index increment than biotin (0.12 compared to 0.16 ml/g). The SAv-bound mologram [Fig. 2(c)] is denoted as [NH-biotin/SAv|NH-PEG] and is fabricated by reactive immersion lithography as introduced previously [7]. We develop a new version of the illumination setup that achieves

higher peak-to-peak mass modulations of 540 pg/mm^2 (27%) compared to the previously published 283 pg/mm^2 (14%) [7] thanks to the higher spectral and spatial coherence of the laser source used (405 nm) (Figs. 13 and 14). This value is determined from a stimulated emission depletion microscopy (STED) measurement on a Leica SP8 STED as described in our previous publication [7]. When impinging on the mologram, a small portion of the light is coupled out into two converging beams that form two diffraction-limited foci above and below the waveguide. The diffracted light of the lower beam is collected by a $\times 20$, 0.4 numerical aperture (NA) microscopy objective and visualized by a CMOS camera. The molographic pattern has a diameter of 400 μm, a numerical aperture of 0.33, a focal length of 900 μm in glass (the substrate

refractive index $n_s = 1.521$), and a sickle-shaped central recess area (Bragg recess area) of 50 μm width to avoid second-order Bragg reflections [16]. The recess area is formed by two concentric circles with 1040- and 1140-μm diameters, respectively.

B. From protein molecules to molographic signals—simulations of molographic foci

The qualitative intensity distribution in and around the diffraction-limited molographic spot can be described by the coherent superposition of individual Rayleigh scatterers or by a mean-field approach through Fourier optics with both yielding the same results [35]. Here, we chose the first method to investigate the expected intensity distribution in the focal plane by summation of the scattered electric field of individual dipoles (molecules) located on the mologram [Fig. 3(a)]. We wrote a graphical processing unit (GPU)-based Python framework that can simulate the intensity distribution of a large amount (few 100 million) of scatterers on a plane with typically 150×150 pixels resolution within roughly an hour. This semi-analytical approach has the great advantage of calculating the field only where it is to be determined [compared to finite-difference time-domain method (FDTD) or FEM approaches]. The exact procedure is outlined in Appendix D and shall only be summarized here. First, the eigenmode equation of the dielectric waveguide is solved according to Marcuse [29] for the fundamental TE mode in order to calculate the excitation field at the position of the scatterers, which are placed on the molographic pattern [Fig. 3(b)]. Proteins can be modeled as Rayleigh scatterers due to their small size of only a few nm [36]. The dipole strength of a protein molecule depends only on its molecular mass and the immersion medium [Fig. 3(c)]. This is because the radius and the refractive index of the resulting sphere are related to the molecular mass and can be calculated as described in Appendix D3. For most practical purposes, the exact composition of the protein is negligible for its scattering properties. To account for reflections at the optical interfaces, we use the dipole potential approach outlined by Novotny and Hecht [30]. Furthermore, multibody interactions are disregarded because the scattering cross section of a typical protein is only of the order of 10^{-24} m^2 and the total diffracted power is typically less than 1%.

C. Comparison between analytical, numerical, and experimental results

1. Shape of the molographic Airy disk

Figure 4 illustrates the diffraction-limited focus obtained by simulations [Figs. 4(a) and 4(b)] and experiments [Figs. 4(c) and 4(d)] as an axial and radial slice through the focal point of the mologram. The experimental molographic spot is acquired from a SAv^{555} (Alexa Fluor^{TM} 555
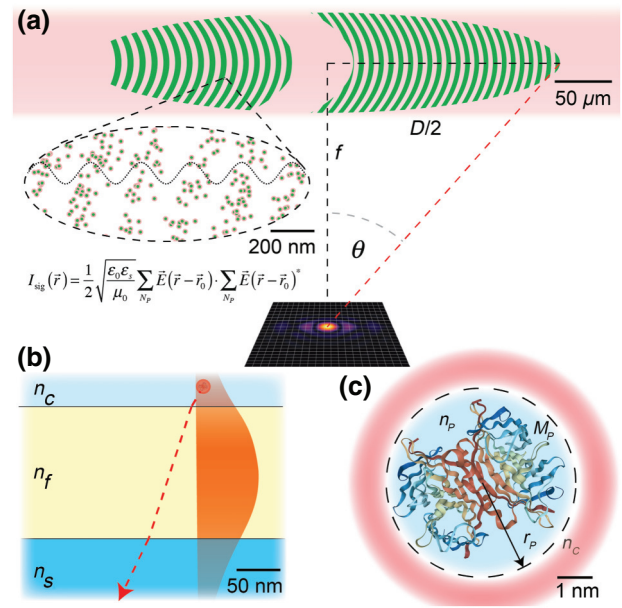

FIG. 3. Simulation of molographic foci. (a) The molographic signal emerges from the superposition of the scattered electric fields of many individual protein molecules on the surface of the waveguide [proteins are not drawn to scale but their number density corresponds to the 2.6 pg/mm^2 at the detection limit (Fig. 8)]. This field is computed for every pixel on a specified screen in the focal plane of the mologram. (b) The scattered field is calculated by modeling the proteins as Rayleigh scatterers excited by the evanescent field of the waveguide mode, which is obtained by solving the eigenvalue problem of the slab waveguide [29]. n_c, n_f, and n_s are the refractive indices of cover, film, and substrate, respectively. The dielectric interfaces can be accounted for by computing the dipole potentials of the layered interface as described by Novotny and Hecht [30–32]. (c) The optical properties necessary to determine the polarizability of the protein dipole, i.e., refractive index and radius, can be calculated from its molecular mass and the refractive-index increment for proteins in water [16,33,34].

labeled, Thermo Fisher Scientific) mologram in water. The fluorophore is only used for quality control and its scattering cross section is negligible compared to the one of SAv. Therefore, we will only refer to SAv for the rest of the paper. The chip is fabricated by reactive immersion lithography in dimethyl sulfoxide (DMSO) [7] followed by a 15 min incubation of 1 μM SAv in phosphate buffered saline with Tween 20 (PBS-T) buffer (pH 7.4; 0.05% Tween20). This yields a 540 pg/mm^2 peak-to-peak sinusoidal surface mass modulation. The experimental focal spot is acquired under water immersion by pipetting 10 μl of deionized (DI) water on the chip and performing a z stack with the MoloReader (vertical resolution of 1.3 μm). The computed focal spot is obtained from a simulation of 4.75 million SAv molecules (7.9 pg/mm^2 peak-to-peak modulation) sinusoidally distributed on the ridges of the mologram with water as the cover medium. This number of SAv molecules

is sufficient to demonstrate the excellent agreement of the numerical results with the measured experimental intensities. The simulated and experimental images only differ by the speckle pattern caused by scattering of the guided wave at noncoherent dipoles, which are not taken into account in the simulations.

The Airy disk radius for a diffraction-limited lens is determined by the wavelength and the NA of the mologram $\Delta x = 0.61(\lambda/\text{NA})$, which leads to 1.17 μm for our molograms. The Airy disk radii found in the measured (solid blue) and the simulated curves (dashed green) in Figs. 4(e) and 4(f) are 1.09 and 1.07 μm in the x and 1.34 and 1.22 μm in the y directions, respectively. The Airy disk is slightly elongated in the y direction in both experiment and simulation due to symmetry breaking of the central Bragg recess area of the mologram [Fig. 3(a)]. Without considering the central recess area, the Airy disk is perfectly symmetrical and has the size of a focal spot of a diffraction-limited lens (Fig. 17)[37]. In the experiment, there is additional broadening by scattering of the guided mode at waveguide imperfections into other guided modes with a small y component in the propagation vector.
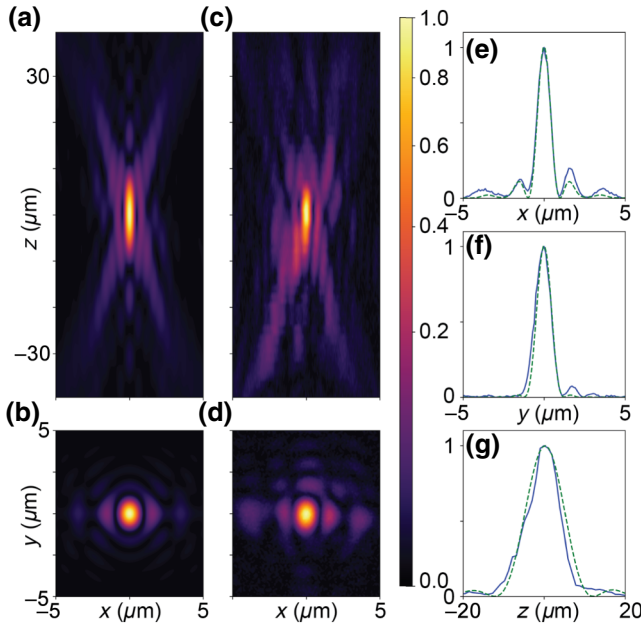

FIG. 4. The normalized intensity distribution in the vicinity of the focal spot. The contour plots show the vertical and the horizontal focal planes of the normalized intensity signal obtained by simulations (a),(b) and experiments (c),(d). The line plots (e)–(g) show the cross sections evaluated through the focal point along each axis. Both the simulated (dashed green) and the experimentally obtained (solid blue) curves in the focal plane exhibit the shape of an Airy pattern as expected from ideal lenses. The Airy disk is slightly enlarged in the y direction due to the sickle-shaped recess area in the middle of the mologram [16].

In the extreme case of a contaminated (strongly scattering) waveguide, the molographic spot attains a sickle shape (the so-called m line [38]). The depth of field of the mologram also follows the equation for a diffraction-limited lens: $\Delta z = 2n_s(\lambda/\text{NA}^2)$ (Chap. 4 in Ref. [30]). The depth of field depends on the refractive index of the medium in which it is observed. Since the thickness of our chip is smaller than the focal length of the mologram, we observe the molographic focus in air (Fig. 16). Yet, the simulation is carried out in an infinitely thick glass slide, therefore, the depth of field has to be compressed by a factor of 1/1.521 to match the experiment. After this adjustment, the experimental and the simulated depth of field amount to 11.23 and 12.61 μm, respectively [Fig. 4(g)]. These are in close agreement with the expected value of 11.62 μm for a diffraction-limited lens.

2. Quantitative intensity in the molographic focus by analytical predictions, simulations, and experiments

The quantitative intensity within the Airy disk is amenable through coupled mode theory (CMT) or summation of Rayleigh scatterers (RS) without considering multiple reflections of the scattered light at the interfaces of the waveguide. Here, we briefly show that the analytical expressions for the two approaches are equivalent and verify them by numerical simulations and experiments.

Fattinger [16] used CMT to obtain an expression for the ratio of power diffracted from the molecular assembly to the power guided by the waveguide (both powers are expressed in power per unit length) [39]. Here, we adapt this expression to yield the intuitive transfer function between the intensity on the waveguide surface and the average intensity in the Airy disk (derivation in Appendix E) which reads

$$I_{\text{avg,CMT}} = 5.59\text{NA}^2\left(\frac{dn}{dc}\right)^2\frac{D^2}{\lambda^4}\eta_{\text{mod}[A]}{}^2\Delta_\Gamma{}^2 I_0. \quad (1)$$

The subscript CMT stands for coupled mode theory. NA is the numerical aperture of the mologram, dn/dc is the refractive index increment for proteins in water [40], D is the diameter of the mologram, and λ is the wavelength. Here, we have adapted and generalized the canonical surface mass modulation $\Delta_{\Gamma_{\text{can}}}$ introduced by Fattinger [16] with the concept of the analyte efficiency of the modulation $\eta_{\text{mod}[A]}$. The analyte efficiency of the modulation is analogous to the diffraction efficiency of gratings with different grating functions [41]. The surface mass density modulation can be computed from $\Delta_\Gamma = m_{\text{mod}}/A_+$, where m_{mod} is the mass of the modulation and A_+ is the area of the ridges (see Appendix C). For a sinusoidal surface mass density modulation (obtained to a first approximation from phase mask lithography), this is equal to the peak-to-peak value. Here, we note that Fattinger defined the canonical

surface mass density modulation differently. His definition would correspond to the molographic surface mass density (see Appendix C) and is, therefore, a factor of two smaller. The prefactor in Eq. (1) arises from various considerations, such as taking into account the relative power incident on the Airy disk and its size. The intensity on the waveguide surface is given by

$$I_0 = 2 \frac{n_c \left(n_f^2 - N^2 \right)}{N t_{\text{eff}} \left(n_f^2 - n_c^2 \right)} P_{\text{WG}}, \tag{2}$$

where P_{WG} is the power per unit line (W/m) in the waveguide, t_{eff} is the effective thickness of the waveguide, N is the effective refractive index of the fundamental TE mode, and n_f and n_c are the refractive indices of the waveguide film and the cover medium.

The expression for Rayleigh scattering (neglecting the optical interfaces) is stated by Fattinger [16]. It can be written in the following form (see Appendix E):

$$I_{\text{avg,RS}} = 1.268\pi^2 \text{NA}^2 n_c^2 \frac{\left(n_P^2 - n_c^2 \right)^2}{\left(n_P^2 + 2n_c^2 \right)^2} \frac{D^2}{\lambda^4} \frac{\eta_{\text{mod}[A]}^2 \Delta_\Gamma^2}{\rho_P^2} I_0. \tag{3}$$

If one inserts the definition of the refractive index increment for proteins in dilute solutions [$dn/dc = 0.182$ ml/g (water)] [40,42],

$$\frac{dn}{dc} = \frac{3}{2} \frac{1}{\rho_P} n_c \frac{n_P^2 - n_c^2}{n_P^2 + 2n_c^2}, \tag{4}$$

one can easily verify that they yield the same result. ρ_P is the dry mass density of the protein calculated according to [43] and n_P is the refractive index of the dry protein sphere. For SAv, we compute the following values: $\rho_P = 1.412$ g/cm^3 and $n_P = 1.598$, which yield an equivalent $dn/dc = 0.36$ ml/g (air) that has to be used in the CMT model for comparison purposes.

To compare the analytical expressions [Eqs. (1) and (3)] with experiments, molograms of different diameters and numerical apertures are fabricated on a chip. Ten molograms of decreasing diameters (400, 343, 296, 255, 221, 193, 168, 148, 131, and 117 μm) and NAs of (0.33, 0.29, 0.25, 0.21, 0.19, 0.16, 0.14, 0.13, 0.11, and 0.1) at a constant focal length of 900 μm are designed in the first row of the phase mask. The same 10 molograms are arranged in the opposite order in the second row of the phase mask. The diameters are chosen such that the area differs by a factor of 1.4 from one mologram to the next (compensated for the Bragg recess area). To use the analytical expression for molograms with the Bragg recess area, the area has to be corrected by a factor of $2A_+/A_{\text{mologram}}$, where A_+ is the area of the ridges and A_{mologram} is the
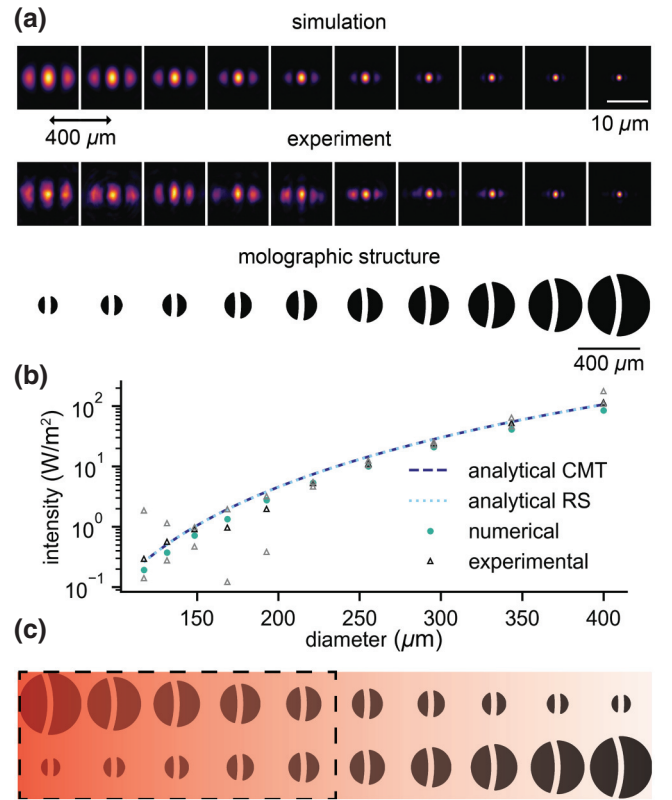

FIG. 5. Comparison between analytical models (CMT and RS), numerical simulations, and experiments for SAv mologram in air. (a) First row: Simulation of the intensity distribution in the focal plane for molograms with constant focal length (900 μm), constant sinusoidal surface mass density modulation (540 pg/mm^2, peak-to-peak), and varying diameters with air as the cover medium. Second row: Corresponding experimental measurement of the intensity distribution in the focal plane. Third row: Schematic of the underlying molograms. (b) Absolute average intensity values over the Airy disk for different molograms derived from measurements (median is in black, other measurements in gray), simulations, and the two analytical models (Eqs. (1) and (3)). The diameters of the molograms in the analytical formula are adjusted in order to account for the missing binding sites in the central Bragg recess area. The small differences between measured and calculated intensity for some molograms can be explained by alterations of the wave-front of the guided mode due to preceding molograms. (c) Chip geometry with the positions of the molograms taken for the analysis.

area of the molographic footprint (ridges + grooves + Bragg recess area). The experimental design of two rows is chosen because each mologram alters the mode shape slightly, and therefore the foci of the last molograms in a row are increasingly distorted in the y direction. Thus, the first five molograms of either row are used on three different measurement fields on the same chip (see Fig. 5(c) and Ref. [7] for chip geometry). The investigated molograms are the same SAv molograms as described in the last section.

The simulation is performed by placing exactly the amount of SAv molecules sinusoidally on the ridges of the mologram that exhibits the same diffraction efficiency as a peak-to-peak surface mass modulation of 540 pg/mm^2 (for the 400-μm mologram these are 231 million individual dipole scatterers, see Appendix D for the necessary conversions). The proteins are placed directly on the waveguide (the field at $z = 0$ is used to calculate the dipole moment). The scattered intensity is computed on a 150×150 grid around the focus, whereas individual grid points are spaced 110 nm apart, which is equivalent to the pixel size of the camera used in the experiment. Figure 5(a) shows the intensity distribution in the focal spot of the 10 molograms with varying diameters in air obtained from simulations and experiments, while the last row shows the underlying mologram. It can be seen that the intensity distributions of simulation and experiment are in perfect agreement over the entire range of diameters investigated. It has to be noted that it is nontrivial to achieve diffraction-limited focusing for molograms up to a diameter of 400 μm on high-refractive-index waveguides, since small gradients in the thickness (and therefore also in the effective refractive index) can already cause a significant accumulated phase shift between the guided mode and the synthetic hologram (designed for constant effective refractive index). Therefore, after a certain propagation distance, which we call the dephasing length, light scattered at the first and the last lines of the mologram can interfere destructively (see Appendix F and Fig. 12).

Figure 5(b) compares the average intensity in the focal point analytically (RS and CMT), numerically, and experimentally. The numerical intensity values are obtained by averaging the intensity on the screen over the Airy disk of the diffraction-limited lens. Experimentally, the mean intensity is calculated by subtracting the average background and then averaging on a circle of the size of the Airy disk centered at the maximum intensity. As can be readily seen, the CMT model and the RS model show nearly perfect agreement with the experimental results. However, there are a few effects which are not accounted for in these simple analytical models. These include free-space attenuation (all the dipoles are assumed to be at the center of the mologram), the angle dependence of Rayleigh scattering [36], the symmetry breaking of the central Bragg recess area, the observation in a half space with a denser medium, and reflections at the interfaces [44]. The numerical simulations incorporate them (see Appendix D), which result in a slightly lower intensity than the analytical models (factor of 1.33 in air). The fact that the experiment agrees closer with the simple analytical models can be explained by uncertainties in the measurement. These arise, for example, from the determination of the surface mass modulation on the mologram. Due to the nature of the quantification procedure (quantitative fluorescence on the submicron scale using STED), we expect to

have an uncertainty of roughly 10% in this measurement. Other possible sources of error, and most likely the prominent ones, are the estimation of the guided power P_{WG} at the mologram location (Appendix G). One can also see in Fig. 5 that the molograms closer to the in-coupling grating have higher intensities and their median values show less deviation to the curves expected from the analytical and numerical models. This can readily be explained by the alteration of the wave-front (guided-guided mode coupling) at every preceding mologram and at waveguide imperfections.

In summary, we have shown that the RS and CMT models are equivalent and show excellent agreement with numerical simulations and experiments for molograms with different diameters and numerical apertures in air. The analytical models are, therefore, a valid tool to make predictions of the limit of detection and to determine the surface mass modulation from the measured intensity in the molographic focal spot. Furthermore, we demonstrated the manufacturing of diffraction-limited molograms with diameters up to 400 μm on a high-refractive-index waveguide. Using much larger molograms (>1 mm in diameter) is not reasonable, since expressed proteins are valuable and often limited in biological experiments [45]. In addition, larger molograms compromise the multiplexing capabilities, because fewer measurements per area can be conducted.

III. BACKGROUND AND NOISE ANALYSIS FOR MOLOGRAPHY WITH MASSLESS AFFINITY MODULATION

Besides the intensity that originates from coherently arranged molecules on the waveguide, various background sources scatter intensity into the area of the focal spot. This can either obscure the coherent signal or limit its accuracy due to the stochastic variation of the background. We analyze the limit of resolution for the important case of molograms with massless affinity modulations. For such molograms, the signal of the empty mologram is hidden in the speckle background. Fattinger [16] provided a first estimation of the limit of detection by comparing the power diffracted by the mologram to the background power incident on the Airy disk. The background power was estimated by distributing the waveguide radiation loss uniformly over the solid angle (4π). While this serves as a good first approximation, we now refine the approach. First, we distinguish between background and noise. Whereas we describe the background as the mean intensity in the focal plane, we consider its spatial and/or temporal fluctuations as noise. Although the noise determines the limit of detection, it is worthwhile to investigate the background because the amplitude of the noise is in a fixed ratio to the mean background intensity. This is due

to the nature of the speckle pattern [46], which will be explained in more detail below.

The propagation loss (or attenuation) of a dielectric optical waveguide is an important quantity for its characterization. We evaluate the background with the help of the radiation loss as it has been performed in Ref. [16] for a first estimation of the limit of detection. However, two issues arise when approximating the background from the propagation loss. First, the attenuation constant is a sum of absorption and scattering loss $\alpha = \alpha_{abs} + \alpha_{sca}$, where we define the propagation loss as $P_{WG}(x) = P_{WG}(0) e^{-\alpha x}$. The scattering loss provides additional background photons to the area of the focal spot, whereas the absorption loss does not contribute any additional light. Therefore, determining the background intensity with the propagation loss is only possible when the absorption is small compared to the scattering. The second issue when estimating the background from the attenuation arises from the anisotropy of the scattering. The out-coupled power is not distributed isotropically over the solid angle. In order to determine the intensity in the focal plane, an additional parameter is needed to account for the anisotropy. This anisotropy parameter a_{ani} is explained in more detail in Fig. 6. The average intensity of the background in the focal plane I_{BG} can be conveniently written in terms of the scattering loss and the anisotropy parameter (detailed derivation in Appendix I):

$$I_{BG} = \frac{NA^2}{4} a_{ani} \alpha_{sca} P_{WG}. \tag{5}$$

We call the product $a_{ani}\alpha_{sca}$ the *scattering leakage*, since it refers to the light leaking into the focal plane of the mologram. As a side note, if the absorption is not negligible and the attenuation constant is used in Eq. (5), an experimentally determined a_{ani} will also correct for the absorption.

Next, we aim to show which scattering mechanisms exist, how they contribute to the background, and what needs to be considered when designing a waveguide for molographic sensing. We assume these mechanisms to be noncorrelated, which allows us to add the individual contributions such that

$$a_{ani}\alpha_{sca} = \sum_i a_{ani,i}\alpha_{sca,i}, \tag{6}$$

where the total scattering leakage $a_{ani}\alpha_{sca}$ corresponds to the experimentally measured value (see Appendix J). Fig. 7 shows six possible scattering sources for background photons: (a)–(c) are scattering processes inherent to the biosensing experiment and (d)–(f) depend on the waveguide manufacturing. We will now investigate the importance of each mechanism qualitatively. If possible, we will treat the relevant sources quantitatively. Before
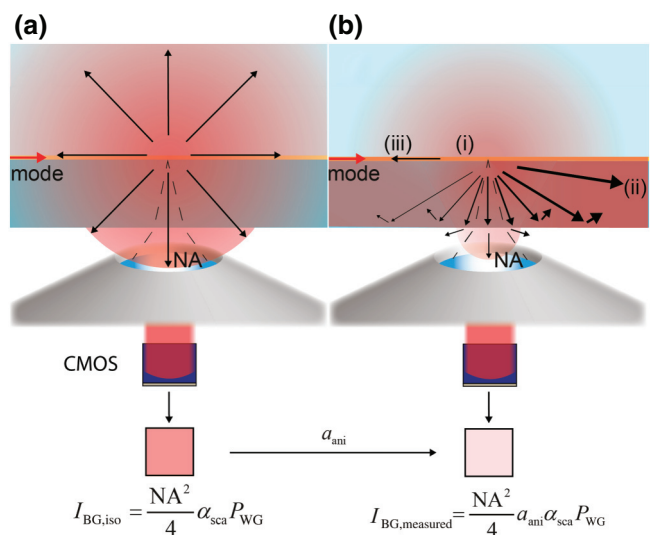

FIG. 6. Illustration of the anisotropy parameter a_{ani}. (a) The scattered power is distributed isotropically in all directions. Only angles that can be collected by the numerical aperture (dashed lines) of the objective contribute to the background. This results in the expression for $I_{BG,iso}$. The intensity is then multiplied with a_{ani} to match the average measured intensity in the focal plane (b) (see Fig. 19). In reality, scattering is an anisotropic process. There are three effects that contribute to anisotropy. (i) The asymmetry of the waveguide leads to a stronger scattering into the substrate due to its higher optical density [30]. (ii) Forward scattering is usually dominant over backward scattering [47]. (iii) Scattering into guided modes of the waveguide is more efficient than into freely propagating modes [30].

analyzing each scattering process individually, the relation between background and noise shall be explained. Any scattering is caused by an underlying stochastic refractive-index distribution within the angles of the NA of the optical system. This stochastic process is transformed by the coherent illumination to spatial intensity fluctuations in the focal plane—a speckle pattern. It is of utmost importance to distinguish between dynamic and static scattering processes. The speckle pattern of a dynamic scattering process exhibits a timescale much shorter than the required bandwidth of the sensor. It will be averaged to a homogeneous background. This background can be subtracted, which renders all dynamic scattering processes negligible. Conversely, static scattering backgrounds lead to speckle patterns that are relatively stable over the time course of the measurement and generally unknown *a priori* to the measurement. This generates an uncertainty when we determine the mass density on the mologram because the relative contribution of the static background to the intensity of the molographic focus is unknown. As a side note, if the NAs of the objective and the mologram match, background speckles and the mologram focus have the same size. Fortunately, the statistics of speckle patterns are well known and speckles exhibit a negative exponential distribution of the intensity [46]. The 99.7% percentile
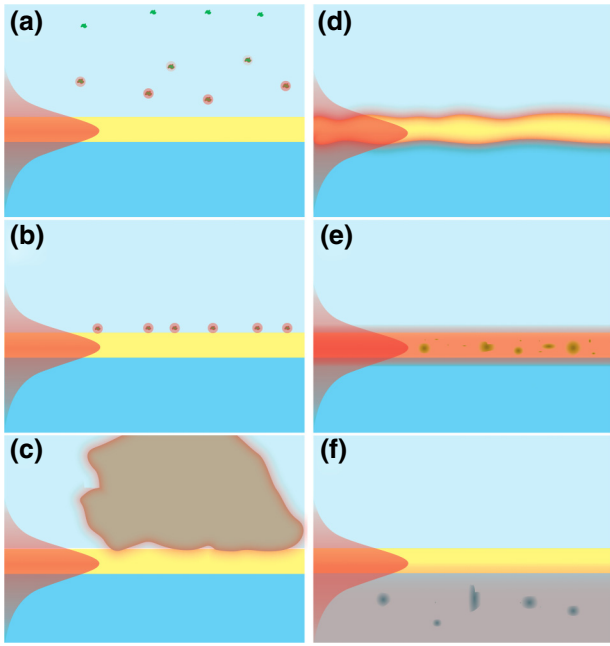
FIG. 7. Possible sources of background intensity due to scattering: (a) molecules in the solution; (b) nonspecifically bound molecules on the waveguide surface; (c) large particles such as dust or cell bodies on the waveguide surface; (d) waveguide surface roughness; (e) refractive index inhomogeneities inside the waveguide; (f) refractive index inhomogeneities inside the substrate.

[definition of the limit of detection (LOD), generally stated as $\mu + 3\sigma$] of the exponential distribution is always in a fixed ratio to the mean, and therefore knowing the mean background allows the estimation of the noise and the limit of detection.

Figure 7(a) illustrates the scattering at molecules in solution in the evanescent field above the waveguide. If the distance between two molecules changes by roughly half a wavelength, the interference for a given speckle in the focal plane can switch from completely constructive to completely destructive. When comparing this short distance to the diffusivity of proteins [48], it is apparent that the measurement time (approximately 1 s) is several orders of magnitude longer than the diffusion time over these length scales. Therefore, the scattering of any molecule in solution is a dynamic process and does not contribute to the noise in the background.

The scattering of randomly adsorbed proteins [Fig. 7(b)] on the waveguide surface has static and dynamic components. Most of these molecules adsorb reversibly to the low-energy nonfouling surface, and therefore have affinities in the mM range. Interactions with such affinities exhibit short lifetimes (μs to ms) [49]. Therefore, the rate at which new proteins adsorb and desorb on the surface is much faster than the acquisition of a single data point. On the other hand, a minority (below 10 pg/mm^2) [17] binds quasi-irreversibly to incoherent surface defects present

on any monolithic surface adlayer [18]. These are the static components of nonspecific binding. However, the contribution is extremely weak compared to the coherent signal. This can easily be understood if one recalls that the coherent signal scales with the number of adsorbed particles squared, whereas the nonspecific background scales linearly with this number [16]. To give an example, the 10 pg/mm^2 of irreversibly bound molecules upon serum exposure correspond to roughly 100 million molecules per mm^2. The same signal intensity can be achieved with a coherent arrangement of the square root of this number, which is about 10 000 molecules per mm^2. This corresponds to only 1 fg/mm^2 of coherent matter or four orders of magnitude less than the molecular mass from irreversible nonspecific binding. This consideration exemplifies again the insensitivity of focal molography to nonspecific binding [7]. The value is so low that in most applications, other background sources will be limiting.

The scattering originating from large particles [see Fig. 7(c)] will be static or dynamic depending of the flow conditions. The analytical treatment of this mechanism is difficult because such particles exceed the Rayleigh regime and the evanescent field of the waveguide. However, one can at least state that the large size of these scatterers causes strongly anisotropic forward scattering (Mie scattering regime [50]). Besides the strong anisotropy, the influence of this scattering process can be reduced by controlling the number of adsorbed particles through careful handling of the chips and filtering of the samples prior to analysis. We will, therefore, not cover this scattering process in our theory since, for most experiments, it can be minimized to a negligible level (see Fig. 24).

The static scattering process at the surface roughness of the two waveguide sidewalls [Fig. 7(d)] has been studied extensively and is not negligible [29,51–55]. In order to quantify the scattering, we have adapted the analytical formulas derived by Lacey and Payne [52,56] for the more general case of an asymmetric waveguide (see Appendix I2). This model has the advantage of providing an analytical solution for the attenuation constant [Eq. (I25)]. However, more relevant for our analysis is the expression for the scattering leakage caused by the roughness of the waveguide:

$$a_{\mathrm{ani},r}\alpha_{\mathrm{sca},r}$$

$$= \frac{\pi n_s}{\mathrm{NA}} \frac{\pi \left(n_f^2 - N^2\right)}{t_{\mathrm{eff}}\left(n_f^2 - n_c^2\right)} \frac{\sigma^2}{\lambda^2 N n_f}$$

$$\times \left[\left(n_f^2 - n_c^2\right)^2 + J^2\left(n_f^2 - n_s^2\right)^2\right]$$

$$\times L_c\beta \int_{\pi/2-\mathrm{NA}/n_s}^{\pi/2+\mathrm{NA}/n_s} \frac{1}{1 + \left[(\beta - kn_s\cos\theta)L_c\right]^2} d\theta,$$

$$(7)$$

where J accounts for the waveguide asymmetry,

$$
J = \cos\left[t_f \beta \frac{\left(n_f^2 - N^2\right)^{1/2}}{N} \right]
$$
$$
+ \frac{\left(N^2 - n_c^2\right)^{1/2}}{\left(n_f^2 - N^2\right)^{1/2}} \sin\left[t_f \beta \frac{\left(n_f^2 - N^2\right)^{1/2}}{N} \right]. \quad (8)
$$

σ is the rms roughness and L_c is the correlation length of the roughness. $\beta = 2\pi N/\lambda$ corresponds to the momentum of the mode in the propagation direction and t_f is the thickness of the waveguide. Eq. (7) can be used to estimate the background for different parameters. Yet, since it is difficult to determine a single correlation length from an AFM measurement, the model should be applied carefully. It can be improved by taking into account the full information of the power spectral density instead of assuming an exponential decay of the autocorrelation. Yet, this model is less intuitive. In addition, the assumption of no correlation between the two sidewall roughnesses and the two-dimensionality of the model can result in some inaccuracy. Despite these simplifications, the described model is a valuable tool for the estimation of the background intensity from the roughness properties of the waveguide surface (see Appendix J5). Furthermore, it is important to notice that the scattering leakage caused by roughness $a_{\mathrm{ani},r}\alpha_{\mathrm{sca},r}$ is fairly constant with respect to NA. Once characterized, this enables a straightforward comparison between experiments of molographic systems with different numerical apertures.

The volume scattering due to refractive index inhomogeneities in the waveguiding film [Fig. 7(e)] is a static process and amenable by similar theoretical investigation as the surface scattering [54]. However, the applicability of these models are limited since the characterization of the detailed distribution of those inhomogeneities by an orthogonal method is not trivial. For our waveguides, the volume scattering is much smaller than the scattering from surface roughness of the waveguide sidewalls (see Appendix J2) and can therefore be neglected. Yet, this only holds true for thin waveguides with a high refractive index.

Finally, the scattering from substrate inhomogeneities can be neglected because it is extremely low for a well-chosen material [Fig. 7(f)].

Based on the qualitative analysis, we conclude that for a general waveguide, one has to consider the volume and surface scattering of the waveguide for the background analysis $a_{\mathrm{ani}}\alpha_{\mathrm{sca}} = a_{\mathrm{ani},r}\alpha_{\mathrm{sca},r} + a_{\mathrm{ani},v}\alpha_{\mathrm{sca},v}$. Which of these processes is prevailing is determined by the configuration of the waveguide. In general, the relative importance of volume scattering increases with waveguide thickness (see

Appendix J6). On the other hand, the higher the refractive index of the film, the larger is the index contrast and the light intensity at the two waveguide sidewalls, which increases the contribution of surface scattering. Therefore, surface scattering is likely to be dominant for thinner waveguides with a high-refractive-index contrast (which is the case for the Ta_2O_5 waveguide discussed in this publication, see Appendix J6), whereas volume scattering will be limiting for thick waveguides with a low-refractive index (high fraction of power in the waveguide and hardly any index contrast at the waveguide sidewalls).

Yet, the signal also scales with the waveguide properties. Therefore, instead of minimizing the background, one needs to maximize the signal to background ratio. We define a figure of merit for a waveguide in order to easily identify the relevant parameters for this optimization. More generally, a figure of merit of a molographic biosensor can be formulated. The figure of merit for focal molography is the ratio of signal to background intensity, which stands for the ratio of mass sensitivity to dark-field illumination quality:

$$
\mathrm{FOM_{FM}} = \frac{D^2}{\lambda^4} \frac{n_c\left(n_f^2 - N^2\right)}{N t_{\mathrm{eff}}\left(n_f^2 - n_c^2\right) a_{\mathrm{ani}}\alpha_{\mathrm{sca}}}. \quad (9)
$$

This expression is obtained by dividing Eq. (3) by Eq. (5). We omitted the protein-related parameters and the analyte efficiency since, for most applications, these are fixed. By further excluding the diameter of the mologram (geometrical design parameter), one obtains the figure of merit of the waveguide as

$$
\mathrm{FOM_{WG}} = \frac{n_c\left(n_f^2 - N^2\right)}{\lambda^4 N t_{\mathrm{eff}}\left(n_f^2 - n_c^2\right) a_{\mathrm{ani}}\alpha_{\mathrm{sca}}}. \quad (10)
$$

The dependency on wavelength to the fourth power can be misleading. The choice of wavelength heavily affects the scattering leakage $a_{\mathrm{ani}}\alpha_{\mathrm{sca}}$. Other parameters such as N and t_{eff} also depend on the wavelength. Therefore, one should be careful when making predictions from Eq. (10). Experimentally, $a_{\mathrm{ani}}\alpha_{\mathrm{sca}}$ can be determined by measuring the intensity in the focal plane and the power in the waveguide and applying Eq. (5). Alternatively, $a_{\mathrm{ani}}\alpha_{\mathrm{sca}}$ can be estimated from the measured roughness properties with the help of Eq. (7) if volume scattering is negligible compared to surface roughness scattering (see Appendix J5). The relative importance of volume to surface scattering can be investigated by measuring the ratio of the scattered intensity with two different cover media, because volume scattering will hardly be affected by the change in cover medium. We perform these characterizations for our waveguide and find a scattering leakage of

$a_{\mathrm{ani},r}\alpha_{\mathrm{sca},r} = 3.12/\mathrm{m}$ in air and surface roughness scattering to be dominant over volume scattering (see Appendix J5). From this, we compute a figure of merit of our waveguide of $\mathrm{FOM_{WG}} = 1.27 \times 10^{30}$ /m^4. The figure of merit (FOM)] for molography for a mologram of diameter 400 μm on this waveguide is $\mathrm{FOM_{FM}} = 2 \times 10^{23}$ /m^2.

In summary, the current high-refractive-index Ta$_2$O$_5$ waveguide is already a good choice for molography, mainly thanks to the high field on the surface, which leads to a strong signal and also to its negligible volume scattering. This compensates for the fairly large surface scattering due to the substantial index contrast. Still, since surface scattering is dominant, we expect that the waveguide could be further optimized. The options to reduce the surface scattering include diminishing the rms of the surface roughness, which is determined with AFM to be 0.6 nm for our waveguide (see Appendix J5), or to adapt the waveguide parameters using Eqs. (7) and (10). However, this has to be performed with care, since Eq. (7) only considers surface roughness scattering. For other waveguides, such as thick low-refractive-index waveguides, volume scattering is most likely the dominant scattering source and the decrease in sensitivity for low-refractive-index waveguides is substantial. Nevertheless, once the scattering leakage is assessed experimentally, Eq. (10) allows a straightforward comparison of different waveguides.

IV. LIMITS OF DETECTION AND RESOLUTION

After having described the possible sources of background light and formulated the FOMs, we will use this knowledge to analyze the limits of detection and resolution of focal molography. The proper limit of detection for a specific assay is elaborate [57,58]. Yet, it is not practically feasible to compare different sensing platforms at the assay level since this would require standard assays to be performed with each of them. To establish the detection limit in endpoint measurements for molography, we define it as a confidence level for false positives, namely the 99.7th percentile (for the normal distribution equal to $\mu + 3\sigma$) or, if not possible due to a lack of experimental data points, the 99th or 99.5th percentile. The resolution is a common benchmark parameter for sensors in general and for real-time label-free biomolecular interaction analysis in particular [59]. It is defined as the temporal rms noise of the baseline after drift correction for the duration of a typical biosensing experiment. Next, we need to find a suitable unit to compare these two quantities among molography and other biosensors. This is not straightforward, since most biosensors measure a change in adlayer density, yet focal molography measures a change in adlayer density modulation. In order to enable comparison, we propose the molographic surface mass density Γ to be this quantity for focal molography. It is defined as the entire mass in the mass modulation uniformly spread over the mologram (see Appendix C). This quantity is calculated from the molographic signal intensity I_{sig} normalized by a reference intensity. A suitable intensity reference for massless affinity modulations is the mean of the speckle background I_{BG}, since it is affected in the same manner as the molographic focus by the majority of physical processes that cause noise or drift (see Appendix K1 for a discussion on intensity reference). It is not possible or is simply too expensive to obtain completely massless affinity modulations for all bioanalytical questions (i.e., a small ligand interacting with a large immobilized protein). In this case, a reference hologram will be required in order to calibrate the molographic signal in samples with varying bulk refractive index. This is not to be confused with the inherent self-referencing character of diffractometric sensors, which makes them more robust than referenced refractometric sensors. Contrary to these, diffractometric sensors only measure the refractive-index difference between ridges and grooves. Therefore, they do not need to compensate for refractive-index drifts in the entire volume of the evanescent field. Finally, we need to specify what the molographic signal I_{sig} refers to exactly. So far, we have discussed the average intensity in the Airy disk, which is related to the maximum by $I_{\mathrm{max}} = 4.378 I_{\mathrm{avg}}$, as the potential molographic signal. Equation (11) and the remainder of this paper use a different algorithm for the computation of the molographic signal, which amounts to $I_{\mathrm{sig}} = 2.012 I_{\mathrm{avg}}$. I_{sig} is the intensity on a single pixel of the focal plane image after convolution with a normalized Bessel kernel of the size of the expected Airy disk (see Appendix K2). The molographic surface mass density Γ can directly be calculated from the ratio of the molographic signal and the mean intensity of the background as

$$\Gamma = \Gamma_0 \sqrt{\frac{I_{\mathrm{sig}}}{I_{\mathrm{BG}}}} \qquad (11)$$

where Γ_0 (see Appendix K2 for derivation),

$$\Gamma_0 = 0.1056 \frac{A_+}{A_{\mathrm{mologram}}} \frac{1}{\eta_{\mathrm{mod}[A]}} \frac{1}{\sqrt{\mathrm{FOM_{FM}}\frac{dn}{dc}}}, \qquad (12)$$

can be seen as an equivalent molographic mass density (calculated from the intensity in the focal plane by using the scattering strength of biological matter). It is due to the stochastic variation of the significantly larger equivalent incoherent mass density of all the incoherent scatterers. The equivalent molographic mass density would contribute the same amount of intensity as the scattering leakage to the focal plane of the mologram (average background intensity in focal plane). For the remainder of this paper, we assume a sinusoidal mass modulation ($\eta_{\mathrm{mod}[A]} = 0.5$).

The limit of detection in terms of molographic surface mass density for a FOM of a given molographic system can be obtained by replacing I_{sig} with $I_{sig,LOD}$ and Γ with Γ_{LOD} in Eq. (11). The minimal detectable normalized intensity increase $\Delta I_{sig}/I_{BG}$ is determined by the readout scheme. We will now distinguish between two readout schemes: Endpoint detection and real-time measurements. We will derive the limit of detection and the limit of resolution for the two schemes, respectively. We also provide experimental evidence for our statements as well as the theoretical projection of the optimization potential.

A. The limit of detection of endpoint measurements is determined by the statistics of the speckle background

In an endpoint measurement, the operator has usually no *a priori* knowledge of the intensity distribution of the speckles and the location of the focal spot (except the focal distance to the surface). Therefore, the Airy disk needs a certain brightness compared to the background intensity to be detectable in a sufficiently large field of view. The detection limit is determined by the variation over many images of the ratio of the maximum pixel value to the image mean. In order to experimentally determine this value, 180 images of size 280×210 μm^2 with 110-nm pixel size on three different ZeptoMark (Zeptosens AG, Switzerland) chips are acquired and convoluted as described in Appendix K. The 99.5% percentile of this ratio equals to 13.8 (see Fig. 23). Hence, if the maximum pixel after convolution is 13.8 times brighter than the mean of the speckle background, it is likely to stem from a coherent binding signal. By inserting this into Eq. (11), one obtains a detection limit in terms of molographic surface mass density of 2.6 pg/mm^2 for a 400-μm mologram with a 0.4 NA on our waveguide, which has a figure of merit of 1.27×10^{30}/m^4 [Fig. 8(a)]. Therefore, only 336 fg of matter yield a signal that is clearly assignable to a coherent assembly of molecules. Figure 8(c) displays a contour plot of the two parameters that affect the detection limit of a molographic system—the diameter of the mologram and the FOM of the waveguide. An increase of either of these two parameters decreases the detection limit. It should be mentioned that the endpoint measurements with no *a priori* knowledge of the speckle intensities represents an upper bound of the achievable detection limit of diffraction-limited molography. Any readout scheme of greater sophistication will achieve lower detection limits.

1. Endpoint detection of vitamin B7 (biotin)

In this section, we demonstrate experimentally that molography can visualize a label-free mass modulation caused by a low-molecular-weight compound in an endpoint measurement. Molograms with the sole difference between grooves and ridges being the tiny molecule vitamin B7 (molecular weight 227 g/mol) are fabricated

[NH-biotin|NH$_2$] [Fig. 8(d)]. A chip is illuminated with a dose of 2000 mJ/cm^2, incubated with 1 mM N-hydroxysulfosuccinimide (sNHS) biotin in HEPES buffer with Tween 20 (HBS-T) buffer (10 mM HEPES, 150 mM NaCl, 0.05% Tween20) at pH 8.0 for 15 min and flood exposed as described in [7]. The foci of these biotin molograms are easily detectable [Fig. 8(b)]. To prove that indeed biotin molograms and not just random speckles are measured, a real-time video of the focal point is recorded while the grooves are backfilled with biotin as described in the description of Video 1. One can clearly see the intensity gradually decreasing until the focal spot becomes indistinguishable from the speckle background. The molographic mass density (sinusoidal modulation) calculated from the median intensity via Eq. (11) of 12 molograms amounts to 11 pg/mm^2 $\pm$ 1.6 pg/mm^2 (FOM$_{WG}$ = 0.58 $\times$ 10^{30}/m^4, calculated from the attenuation constant assuming the same a_{ani}). The uncertainty is caused by an intrinsic property of speckle statistics. Every speckle has a non-negative intensity. However, the sign of the electric field of a background speckle can be positive or negative and is unknown, since the phase information cannot be measured. If the field of the speckle is negative with respect to the field of the molographic signal, some additional coherent matter is required to cancel the contribution from the roughness. Vice versa, less biological mass is required in the case when the molographic focus happens to be on a positive speckle. This physical property poses an intrinsic constraint on the accuracy of the molographic measurement, and therefore the limit of quantification (LOQ).

It has to be stressed that the detection of this low-molecular-weight compound is possible without any equilibration, referencing, or stabilization of the sensor. The amount of bound mass is determined in a reproducible fashion with an accuracy below 2 pg/mm^2 even if the chip is removed and reinserted. This robustness is a fundamental difference to common refractometric sensors, where similar detection limits can only be achieved with samples that are mounted and stabilized within the device and cannot be removed and reinserted.

B. The resolution of real-time measurements is determined by the temporal noise of the speckle background

In real-time detection, the intensity in the focal plane is continuously monitored. Contrary to the above-described end-point measurement, the location of the focal spot in the speckle background is known exactly. This can be realized by reference focal spots or by localizing the spot before backfilling as shown below. Then image processing can be applied to monitor the intensity at the location of the focal spot resulting in a binding trace as a function of time. Such a binding curve is the output of all real-time
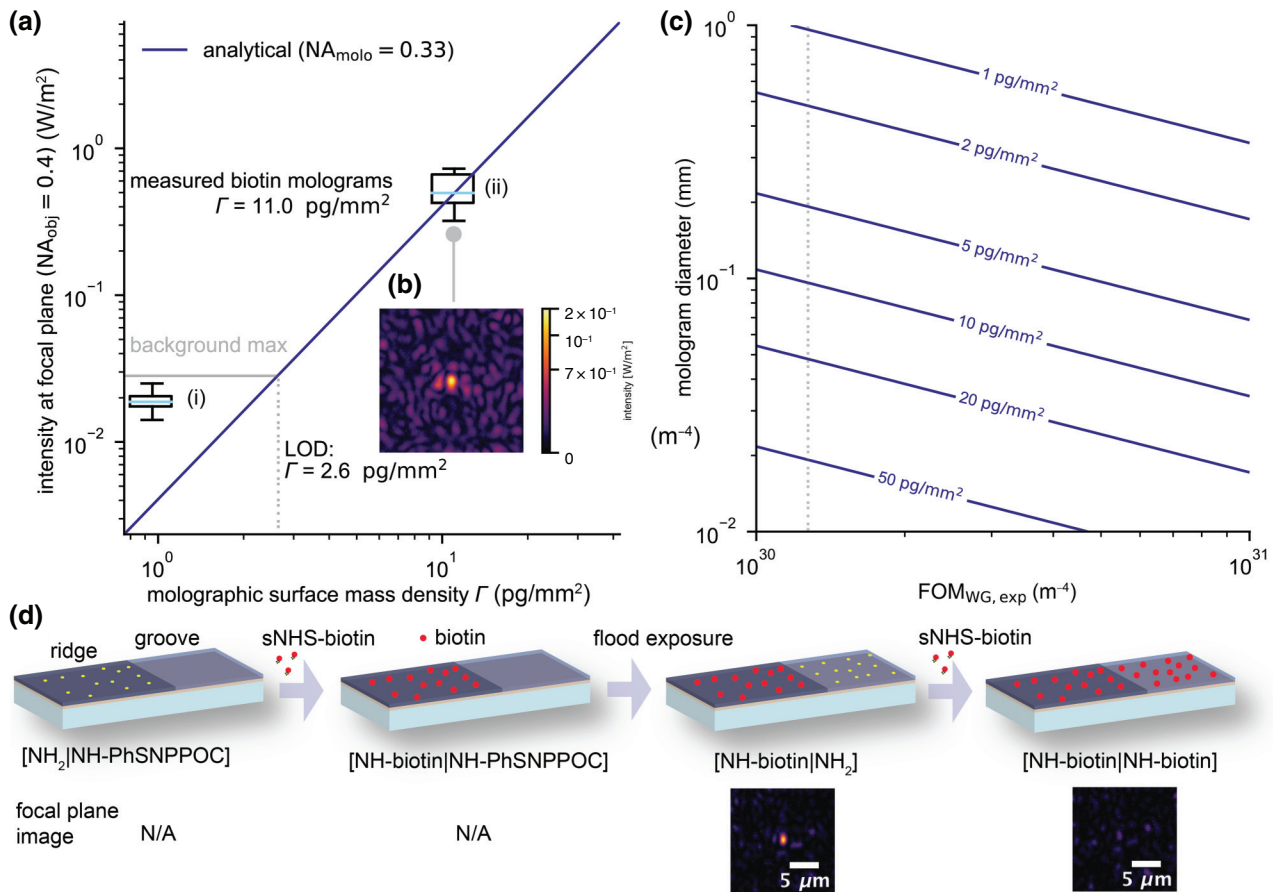
FIG. 8. Detection limit of focal molography without preknowledge of the position of the focal spot for the standard configuration in air (diameter mologram is 400 μm, focal length is 900 μm, NA of the mologram is 0.33, and NA of the objective is 0.4). (a) Detection limit for the standard configuration mologram. 180 background images (280×210 μm^2 with a 110-nm pixel size) focused 100 μm below the surface of the waveguide of three clean chips with a scattering leakage of 3.12/m (Figs. 18 and 19) are acquired, filtered with the shape of the Airy disk (NA = 0.33), and the maximum pixel of the convoluted image is extracted and summarized in the box plot (i). All measured intensity values are normalized to a standard power of 0.02 W/m in the waveguide. The solid gray line corresponds to the 99th percentile of the maximum pixels observed in the Airy disk convoluted background images defining the smallest intensity required for the coherent signal to be discriminated against surface roughness speckles. The solid blue line represents the scattered intensity as a function of the molographic mass density (sinusoidal mass distribution and calculated from Eq. (3) with the mass modulation replaced by the molographic mass density). The intersection point is indicated by the dashed gray line, which denotes the coherent mass that corresponds to the 99th percentile of the measured maximum background intensity for a field of view as specified above. The box plot (ii) corresponds to 12 measured biotin molograms in air. (b) Typical focal plane image of easily detectable biotin molographic foci with intensities roughly 10 times above the detection limit. (c) Detection limit in terms of molographic surface mass density for an ideal sinusoidal mass modulation as a function of the figure of merit of the waveguide (FOM$_{WG}$) and the diameter of the mologram. The detection limit decreases inversely with the diameter and decreases inversely with the square root of the FOM$_{WG}$. The dashed gray line indicates the FOM$_{WG}$ of the investigated waveguide. (d) Fabrication steps for biotin molograms and their subsequent backfilling (Video 1).

biosensors and the detection limit is commonly stated as the temporal rms noise (resolution) over a defined time span, typically a few minutes, of the signal before the analyte is injected. It should be noticed that this is a different definition than the 99.5% percentile described above. The goal of this section is to demonstrate the resolution of diffraction-limited focal molography theoretically and experimentally for our measurement system and compare it to SPR, the gold standard of refractometric sensing.

We start our discussion by appreciating the instrumental precision at which refractometric sensors are operated in order to achieve refractive index resolutions of 10^{-6}–10^{-7} and mass resolutions of 30 fg/mm^2–1 pg/mm^2 [59,60]. If one recalls that one monolayer of water molecules already gives rise to a signal of 300 pg/mm^2, one can appreciate this technological achievement. Only careful optimization of sensor design, referencing strategies, and signal processing over the past three decades has made

this possible [61]. The performance of real-time measurement devices is commonly characterized by two metrics, namely the baseline drift and the baseline noise. The former is expressed in pg/(mm^{2}min) or RU/min (where RU is response units), whereas the latter is expressed as an rms value in RU or pg/mm^2. Nowadays, commercial SPR instruments achieve a baseline noise of 15–30 fg/mm^2 (measured as a rms value after drift correction) and a baseline drift of around 300 fg/(mm^2/min) [28]. Over the course of a measurement, SPR sensors are usually limited by temperature drifts between reference and sensing channels [62]. As will be shown below, such drifts are virtually not present in diffractometric sensors. Therefore, we will compare molography to idealized SPR instruments, which are limited by the baseline noise.

The three most common readout modes of SPR are angle interrogation, wavelength interrogation, and intensity interrogation [10]. Independent of the interrogation mode, the readout of the SPR signal is a measurement of a relative intensity $R = I_{\text{reflected}}/I_{\text{in}}$. The noise in the measured intensity $\text{rms}_{R-\langle R \rangle}$ results in a noise of the detected surface mass density $\text{rms}_{\Gamma-\langle \Gamma \rangle}$, which determines the resolution of the technique [59]. The two noises are related by the sensitivity $S_\Gamma = \frac{\partial R}{\partial \Gamma}$:

$$\text{rms}_{\Gamma-\langle \Gamma \rangle} = \frac{\text{rms}_{R-\langle R \rangle}}{S_\Gamma}. \tag{13}$$

The sensitivity of SPR and a more detailed description of Eq. (13) is stated in Appendix K3 [63–65]. As described in Eq. (11), molography also measures a relative intensity ($I_{\text{sig}}/I_{\text{BG}}$) such that Eq. (13) is valid. For molography, the sensitivity can be described by (see Appendix K4 for the derivation)

$$S_{\Gamma_{\text{FM}}} = \frac{2}{\Gamma_0} \propto \sqrt{\text{FOM}_{\text{FM}}}. \tag{14}$$

Due to the quadratic nature of the sensor transfer function, Eq. (14) is only valid if the signal intensity is close to the reference intensity (in this case, the background intensity). It must be stated here that the values for the sensitivities of molography and SPR should not be compared since they depend on the chosen reference intensity (I_{BG} and I_{in}). Instead, the resolution ($\text{rms}_{\Gamma-\langle \Gamma \rangle}$) can be used for comparison. With the MoloReader, an intensity baseline noise of about 10% has been measured, resulting in a resolution of 90 fg/mm^2 (0.9 RU) for the waveguides used in this experiment (FOM = 0.63×10^{30}/m^4). This is close to the above-mentioned 30 fg/mm^2 (0.3 RU) resolution of the best SPR sensors [60]. If the waveguide background is the intensity reference, Eq. (14) can be used to improve the resolution of molography, which can be achieved by a higher FOM$_{\text{WG}}$ or by a larger diameter of the molo-gram. Further, any reduction of the intensity noise will also significantly improve it.

In order to verify these findings, real-time measurements are performed with the MoloReader. The setup is illustrated in Fig. 26 and the experiment in Fig. 9(d). First, it has to be stressed that all reported experimental results are without any kind of temperature stabilization. The sole effect of temperature in a molographic measurement is a slow drift of the location of the molographic spot within the focal plane, but its intensity is hardly affected by the temperature drift (Video 2). This movement can easily be compensated for by simple image registration algorithms, which are implemented in the readout algorithm (Fig. 27) [66]. As mentioned above, for a continuous measurement, the Airy disk must to be in the field of view and the optical system needs to be focused on the focal plane (see Appendix L2 for the protocol to accomplish this). Buffer baselines are acquired at a flow rate of 20 μl/min for 20 min with a syringe pump (NE-511L, PumpSystems Inc.) and 1 ml syringes (Henke Sass Wolf GmbH). The buffer solutions are degassed prior to use to avoid noise caused by micro-bubble formation. The exposure time of the camera is set to 500 ms and an image is acquired every 3 s. After 20 min, the syringe is exchanged by another one containing 500 pM SAv in PBS-T (spikes in the binding trace at around 22 min). Injection is continued at a flow rate of 20 μl/min. Finally, image processing is performed in order to obtain the baselines as described in Appendix L3 (Fig. 27).

Three binding curves are acquired and are displayed in Fig. 9(b). The signal change due to SAv binding is detectable almost instantaneously after injection. However, whereas in two measurements, the binding trace rose immediately after injection, it decreased at first in the third measurement. This is an example of the molographic focus lying on a background speckle with a negative electric field with respect to the focus itself, as explained before. The baseline noise over 20 min amounted to rms values of 0.074, 0.094, and 0.077 RU and in terms of normalized intensities, 7.1, 15.6, and 7.8%. These experimental noise values nicely agree with the theoretical prediction from Eqs. (13) and (14). Since the theory depends on the surface mass density, and not on the solution concentration, it can be applied to any other biological interaction pair resulting in the same limit of detection. As mentioned before, the mass density resolution is comparable to the best reported SPR results [60]. Yet, the molographic baselines are calculated without any baseline drift correction, unlike the common practice in refractometric sensing, demonstrating the robustness and sensitivity of focal molography.

Another fundamental requirement in label-free bimolecular interaction analysis is the ability to detect a distributed ensemble of molecules with a low surface density on a sufficiently large area. In other words, to detect low receptor (capture molecule) occupancies, a fact overlooked by most of today's nanosensing and single-molecular detection concepts [67]. In sensitive assays, the concentration
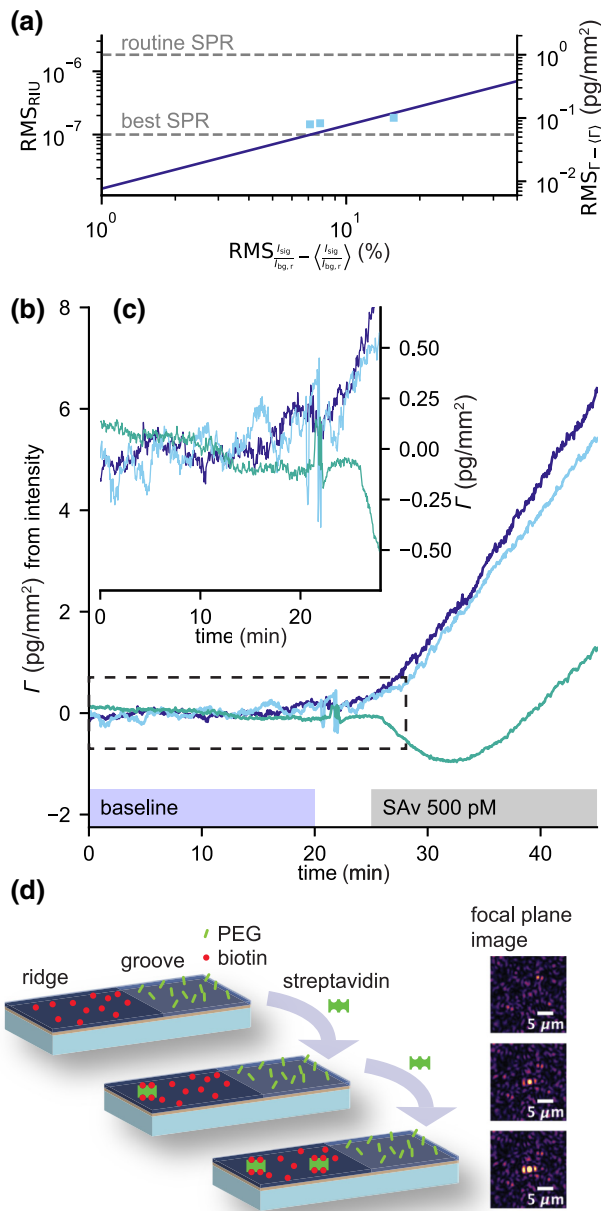
FIG. 9. Comparison of the resolution of focal molography to state-of-the-art resolution of SPR. (a) Best and routine resolution of SPR instruments is indicated. Blue squares are the resolution measured for focal molography. The dark blue line is the sensitivity of molography for a FOM_{FM} of $10^{23}/m^2$. (b) Label-free detection of 500 pM (26 ng/ml) of SAv and baseline noise levels. (c) Enlargement of the first part of the binding curve for better visualization of the baseline noise. (d) Illustration of the experiment: After acquiring the baseline noise from an invisible massless affinity modulation [NH-biotin|NH-PEG] injected SAv binds to the ridges and increases the intensity in the focal spot.

of the analyte is usually several orders of magnitude lower (10 fM–1 pM) [68] than the dissociation constant of the capture probe (10 pM–1 nM) [69], which leads to a receptor occupancy of typically 0.1–1% [67]. The molographic

focus of a 400-μm diameter mologram monitors the activity of roughly 1 billion recognition sites continuously and is, therefore, able to resolve low receptor occupancies as well as measuring a sufficient number of analyte molecules. For example, at the demonstrated resolution of 100 fg/mm^2 (roughly 1 million SAv molecules per mm^2), 100 000 proteins are bound to one biotin mologram [receptor density of 3×10^{10} molecules per mm^2 (11 pg/mm^2)]. Furthermore, taking into account that four biotin molecules bind one SAv molecule, the receptor occupancy in the experiments shown in Fig. 9 can be estimated to amount to 0.01%.

V. CONCLUSIONS

Focal molography visualizes an invisible affinity modulation upon molecular binding. The microscopic method is based on dark-field detection of biological molecules. The coherent arrangement of binding sites in the mologram and the resulting insensitivity to noncoherent noise sources enables robust and highly sensitive detection of biomolecular interactions. A quantitative analysis of these interactions is amenable through the analytical models presented in this paper. These models compute the amount of biological matter bound to the mologram from the intensity of the molographic focus. Their accuracy is proven by the excellent agreement with the presented numerical simulations and the discussed experiments. High sensitivity and a low background are achieved by a waveguide, providing field enhancement and a proper dark-field illumination. However, radiation due to scattering at waveguide imperfections remains the dominant source of background light for massless affinity modulations. Therefore, figures of merit are introduced to investigate the parameter dependencies of the signal to background ratio. They allow straightforward comparison of different molographic arrangements and waveguides. Two readout schemes, endpoint detection and real-time measurements, prove the intrinsic robustness and high sensitivity of focal molography. In an endpoint measurement, the low-molecular-weight compound vitamin B7 could be easily detected and the limit of detection in terms of surface mass is just a few pg/mm^2 by this simple readout scheme. The more elaborate real-time measurements exhibited a resolution below 100 fg/mm^2 over 20 min without any drift correction. This is comparable to the best commercially available refractometric sensors. With further optimization, it is, therefore, likely that the resolution of diffractometric sensors will surpass that of refractometric devices. Yet, by only detecting the coherent signal, the coherent detection scheme has unmet advantages over any established label-free biosensor. Its unique combination of robustness and high sensitivity will enable numerous new applications to analyze the interactions of biomolecules in their natural habitat—the crowded environments of body fluids, tissues, cells, and membranes.
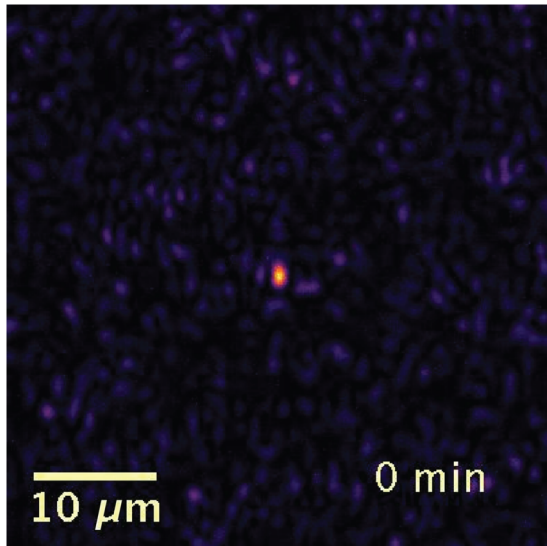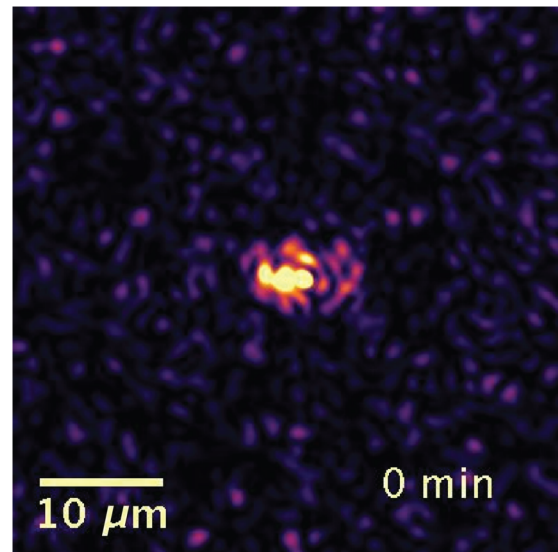
APPENDIX A: AUTHOR CONTRIBUTIONS

A.F., Y.B., J.V., and C.Fattinger planned the experiments, which were performed and evaluated by A.F. A.F and C.Fattinger designed the instrumentation. Y.B. and A.F derived the analytical expressions. S.B. and A.F implemented the simulation framework, which was adapted for GPU computing by C.Forró. M.L. conducted the AFM measurements and V.G. provided the foundations for many experiments. A.F. and Y.B. wrote the manuscript with C.Fattinger providing input for the content and the structure of the work.
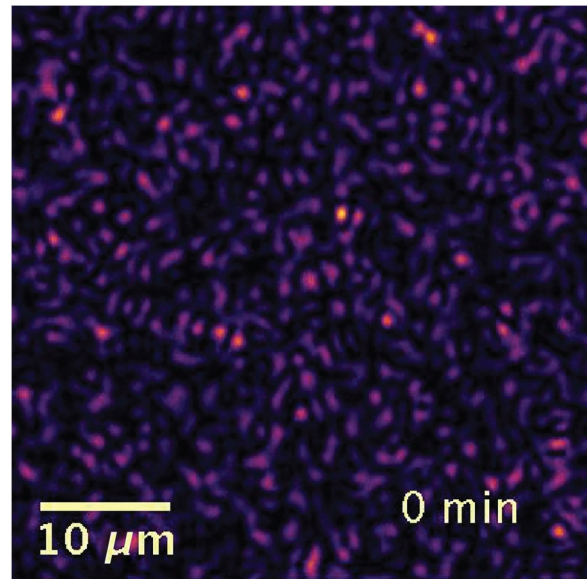
APPENDIX B: VIDEO DESCRIPTIONS

VIDEO 2. *Temperature effect on the molographic spot*. This video shows the influence of temperature on the speckles in the focal plane image. The speckles as well as the molographic spot shift as a function of temperature, but their intensities essentially remain constant. The chip is observed in PBS-T buffer without any flow and the entire chip assembly [Fig. 27(b)] is taken from the refrigerator before the measurement to induce a more visible temperature drift. This drift can easily be compensated by means of image registration.

VIDEO 1. *Real-time backfilling of biotin molograms*. This movie shows the real-time backfilling of a biotin mologram [NH-biotin|NH$_2$] with 1 mM sNHS-biotin at pH 8.0 in HBS-T buffer. The molographic spot fades away upon biotin binding because the grooves are also functionalized with biotin, essentially canceling the mass modulation [NH-biotin|NH-biotin]. This proves that our investigated molograms in Fig. 8 are indeed biotin molograms.

VIDEO 3. *Real-time binding of 500 pM SAv to [NH-biotin|NH-PEG] molograms*. The movie shows the real-time binding of 500 pM SAv in PBS-T pH 7.4 (0.05% Tween20) buffer to a [NH-biotin|NH-PEG] mologram. The movie corresponds to the light blue curve in Fig. 9(b).

APPENDIX C: DEFINITIONS

Here, we list a few definitions related to analyte mass that are used in the manuscript. Focal molography detects the diffracted signal of a coherent ensemble of biomolecules. Nevertheless, the exact arrangement of the analyte mass on the ridges and grooves influences the diffraction efficiency of the molecular hologram and needs to be taken into account to formulate a limit of detection. For instance, a sinusoidal affinity modulation will require more mass bound to it than a canonical affinity modulation to produce the same molographic signal. As the diffraction efficiency compares different grating structures [41], a similar figure of merit—the analyte efficiency—can be defined for diffractometric biosensors. To obtain this, we first outline the definitions that are required to formulate the concept of the analyte efficiency of a particular affinity modulation.

Affinity modulation: The affinity modulation defines the functional distribution of the binding affinity across one groove and ridge after common offset (minimum value on the grooves) subtraction. This determines the mass modulation and hence the refractive index modulation after molecular recognition. Different ideal affinity modulations (δ-distributed, sinusoidal, canonical) are shown in Fig. 10(a) and explained in more detail below.

Total mass on the mologram: The *total mass on the mologram* m_{tot} is the total mass that is bound to ridges and grooves.

Mass of the modulation: The *mass of the modulation* m_{mod} is the total mass on the mologram subtracted by the common mass up to the minimum value on the grooves.

Equivalent coherent mass: The *equivalent coherent mass* m_{coh} is the amount of mass that would need to be placed on the center of the ridges (δ-distribution) to obtain the same diffraction efficiency as a mass, which, for instance, is distributed sinusoidally or canonically.
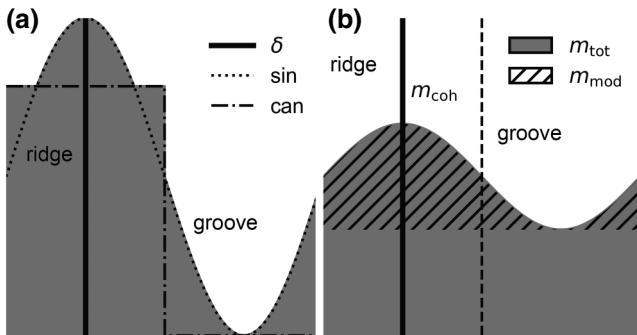

FIG. 10. Affinity modulations and different mass definitions. (a) δ-distributed, sinusoidal, and canonical affinity modulation on the mologram. The canonical and sinusoidal are scaled such that the resulting mass modulations would have the same diffraction efficiency. (b) Definition of the total mass m_{tot}, the mass of the modulation m_{mod}, and the equivalent coherent mass m_{coh}.

Equivalent coherent surface mass density: The *equivalent coherent surface mass density* Γ_{coh} is the equivalent coherent mass divided by the footprint of the mologram (Bragg area, grooves, and ridges included):

$$\Gamma_{\text{coh}} = \frac{m_{\text{coh}}}{A_{\text{mologram}}}. \tag{C1}$$

It is a measure of the sensitivity of molography, but practically not meaningful, since a δ-distributed affinity modulation would be required to obtain it (see discussion below). More appropriate for resolution comparison is the molographic surface mass density.

Molographic surface mass density: The *molographic surface mass density* Γ is defined as the mass of the modulation m_{mod} divided by the entire molographic footprint:

$$\Gamma = \frac{m_{\text{mod}}}{A_{\text{mologram}}}. \tag{C2}$$

In our eyes, it is the most appropriate measure to compare molography to surface plasmon resonance and other refractometric biosensors, which detect mass adlayer changes.

Analyte efficiency: In general, the *analyte efficiency* $\eta_{(A)}$ is the ratio between the equivalent coherent mass and the total mass m_{tot} that has bound to the affinity modulation. The analyte efficiency can be computed from

$$\eta_{[A]} = \frac{m_{\text{coh}}}{m_{\text{tot}}} = \frac{\int_0^\Lambda \sin(2\pi x/\Lambda) f(x)\, dx}{\int_0^\Lambda f(x)\, dx}. \tag{C3}$$

The sinusoid stems from coherent weighting and $f(x)$ is the average mass distribution function (averaged over all periods of the mologram), determined by the affinity distribution. The integration is performed over one representative period of the mologram Λ. In other words, it is the normalized Fourier coefficient of the spatial frequency of the mologram. (To be fully precise, since the mologram is composed of lines with different spatial frequencies, this would be a weighted sum over all periods in reality.) The concept of the analyte efficiency also incorporates imperfections of the lithography process such as finite mask contrast and nonlinear activation functions of the brushed polymer. However, for the discussion of the detection limit and resolution, it is more appropriate to use the analyte efficiency of the modulation, since the common offset is not measured by molography.

Analyte efficiency of the modulation: The *analyte efficiency of the modulation* $\eta_{\text{mod}[A]}$ is the ratio between the equivalent coherent mass and the mass of the modulation. It can be computed analogously to the analyte efficiency by simply subtracting the common offset C_G defined by the lowest value on the grooves from the affinity/mass

TABLE I. Analyte efficiency for different ideal affinity modulations.

Affinity modulation	$f(x) - C_G$	$\eta_{\text{mod}[A]}$
δ-distributed on ridge centers	$\delta\left(x - \dfrac{\Lambda}{4}\right) m_{\text{mod}}$	1
Sinusoidal (ideal)	$\dfrac{m_{\text{mod}}}{\Lambda} + \dfrac{m_{\text{mod}}}{\Lambda} \sin\left(\dfrac{2\pi x}{\Lambda}\right)$	0.5
Canonical (ideal)	$\dfrac{2m_{\text{mod}}}{\Lambda}, \quad 0 \leqslant x < \dfrac{\Lambda}{2}$	$\dfrac{2}{\pi}$

distribution function:

$$\eta_{\text{mod}[A]} = \frac{m_{\text{coh}}}{m_{\text{mod}}} = \frac{\int_0^\Lambda \sin\left(\frac{2\pi x}{\Lambda}\right)[f(x) - C_G]\,dx}{\int_0^\Lambda [f(x) - C_G]\,dx}. \quad (C4)$$

Table I summarizes the analyte efficiency of the modulation of important affinity modulations in molography and Fig. 10(a) depicts them.

δ-distributed affinity modulation: The δ-distributed affinity modulation binds all mass to the very center of the ridges. The analyte efficiency of this structure is 1. However, it is not practically feasible for two reasons: First and most importantly, as the width of the activated region on the ridges approaches zero, the amount of available receptors also does. If not enough receptors are there to capture the analyte, the biosensor is rendered useless [67]. Second, to fabricate such small feature sizes is not trivial with current lithographic technologies.

Sinusoidal affinity modulation: The sinusoidal affinity modulation is obtained by phase-mask lithography (interferometric exposure) under the assumption of a linear deprotection rate of the photoprotective group. To a first approximation, the affinity modulation of the molograms in this contribution is sinusoidal. The analyte efficiency is 50%, because a considerable amount of mass binds to the grooves, and therefore cancels some of the mass on the ridges.

Canonical affinity modulation: The canonical affinity modulation is the one used in the original CMT equations of Fattinger and Tamir [16,39]. The analyte efficiency of this modulation is slightly higher (63.6%), but not fabricatable by phase-mask lithography. Hence, the canonical mologram is only 1.3 ($4/\pi$) times better in terms of analyte efficiency than the sinusoidal mologram.

Surface mass density modulation: In the Supplementary Information of a previous manuscript [7], we introduced the definition of the surface mass density modulation as the difference in mass density between ridges and grooves $\Delta_\Gamma = \Gamma_+ - \Gamma_-$. This definition is not precise in that we did not specify Γ_+ and Γ_- unambiguously. Here, we refine it in the following way: The mass density modulation is referred to as the mass of the

modulation m_{mod} divided by the area of the ridges A_+:

$$\Delta_\Gamma = \frac{m_{\text{mod}}}{A_+}. \quad (C5)$$

For a sinusoidal mass density modulation, this corresponds to the peak-to-peak value $\Delta_{\Gamma_{\text{PTP}}} = m_{\text{mod}}/A_+$. For a canonical mologram, it corresponds to the difference in mass density between ridges and grooves $\Delta_{\Gamma_{\text{can}}}$. In addition, this definition is consistent for any mass distribution function.

APPENDIX D: DESCRIPTION OF THE SIMULATION FRAMEWORK

The following section describes the Python simulation framework that calculates the scattered intensity in the focal spot from coherent protein particles.

1. Waveguide and mologram calculation

The first step of the intensity calculation in the molographic spot is to solve the eigenvalue equations of the dielectric slab waveguide in order to determine the excitation field of the particles. This problem has been extensively investigated by Marcuse [29] in Chap. 1.3 and shall not be restated here. The framework allows the calculation of the fundamental TE and TM modes, whereas only the TE mode was investigated in this publication. Once the mode effective index is known, the molographic pattern with a defined focal length and numerical aperture can be generated.

2. Mologram creation and particle placement

The center of the molographic lines are given by formula 4 in Fattinger [16] and implemented in this way in the simulation framework (with $\xi = 0$). The molographic lines are hyperbolas. An intuitive understanding of the shape of the pattern can be obtained by recalling that the circles of a Fresnel zone plate become ellipses for tilted plane wave incidence and are finally rendered hyperbolic for large incidence angles.

Furthermore, the molographic structure is constrained by an outer circle and a central cutout (to avoid second-order Bragg reflection—approximated as two circles—in reality these are hyperbolas as well, but for molograms with an NA of 0.5 and below this approximation is sufficient). To place the particles efficiently, we make use of the symmetry of the molographic structure with respect to the x-z plane and seed particles only on the upper half of the mologram and invert the y coordinate with a probability of 0.5 for every particle after seeding them all. Every molographic line in the positive (upper) half plane can be seen as a curve $\{\vec{c}(x) = [x, f(x)]\}$ subsequent to the circular constraints mentioned above. However, in order to place the particles uniformly along the line, one cannot simply randomly seed x, since the curves move much

faster in $[x,y]$ space on certain intervals of x, depending on their gradient. Therefore, some regions would be more densely seeded than others. The solution is to sample the normalized arc length (can also be viewed as a cumulative distribution function) uniformly, given by

$$C(x) = \frac{\int_0^x \sqrt{1 + (\partial_x y)^2} \, dx}{\int_0^{x_0} \sqrt{1 + (\partial_x y)^2} \, dx}, \tag{D1}$$

where x_0 is the extent of the molographic line in x. Then, one simply needs to compute $x = C^{-1}(y)$, which is most easily done by interpolation. y is uniformly distributed on the interval [0,1]. Once placed on the center of the ridges, the particles need to be sinusoidally shifted in order to account for the sinusoidal affinity modulation as present from phase mask lithography.

In summary, the procedure is the following:

(1) Compute the mololines as a matrix of x and y coordinates. Calculate their lengths so as to have a vector of arc lengths $[l_0, l_1, \ldots, l_n]$. Normalize the cumulative arc length of this vector to 1. Multiply the normalized cumulative arc-length vector with the number of particles to be placed on the mologram to determine the number of particles that are placed on any line j.

(2) For a chosen particle to be placed on line j, which is discretized and a vector of normalized arc length generated, sample a uniform random number on [0,1]. Select the index of the normalized arc-length vector for this line that contains the entry that is closest to the sampled number. Use this index in order to get the position x, y where the particle should be placed, from the discretized line.

(3) Calculate the width δ of the line and seed the particle sinusoidally (with inverse CDF $C^{-1}(y) = \arccos(1 - 2x)/\pi, C^{-1} : [0, 1] \to [0, \pi]$) with respect to the middle of the line on an interval equivalent to the calculated ridge width.

Since no mass is seeded on the grooves in the framework, the mass density on the ridges needs to be adjusted to take into account the mass on the grooves that would cancel some of the mass density on the ridges. The average mass density on the ridges that needs to be seeded sinusoidally is

$$\frac{\int_0^\pi \Delta_{\Gamma_{PTP}} \sin(x) \, dx}{\int_0^\pi dx} = \frac{2}{\pi} \Delta_{\Gamma_{PTP}}. \tag{D2}$$

This times the area of the ridges gives the total number of particles that must be seeded on the mologram. It is further verified that canonical and sinusoidal placement give the same intensity when the particle number is adjusted accordingly (data not shown).

3. Model of scattering protein particles

Proteins can be modeled as dipole scatterers. This approximation is valid, since the particle radius of a typical protein is only 2.5 nm, which is well below the Rayleigh limit [36]. To use the dipolar approximation, the refractive index and the volume need to be calculated for the dry protein sphere as a function of the molecular mass of the protein. The refractive index of the dry protein sphere can either be calculated from tabulated refractive index data for amino acids [33] and using the procedure from Ref. [40] or via the refractive index increment of proteins in water [40,42]. The refractive index increment dn/dc for proteins in water is generally accepted to be 0.182 ml/g for red light and in the absence of solvation effects. This value for the refractive index increment is a good approximation for most proteins [16,34]. For a detailed discussion on the refractive index increment, see Ref. [40]. We chose the approach with the protein refractive index increment. The dry refractive index of a protein is amenable from

$$n_P = n_c \sqrt{\frac{3n_c + 4\rho_P \frac{dn}{dc}}{3n_c - 2\rho_P \frac{dn}{dc}}}, \tag{D3}$$

where the molecular weight-dependent mass density of the protein $\rho_P (M_P)$ is calculated as reported in Ref. [43]. For SAv ($M_P = 52.8$ kDa), we calculate and use a mass density of 1.413 g/ml and a refractive index value of 1.598 in the simulations (for both air and water). The volume of the protein and from this its radius are calculated according to

$$V_P = \frac{M_P}{N_A \rho_P} = \frac{M_P}{N_A} \frac{2}{3} \frac{dn}{dc} \frac{n_P^2 + 2n_c^2}{n_c (n_P^2 - n_c^2)}, \tag{D4}$$

with the Avogadro constant N_A. For SAv, the radius amounted to 2.46 nm.

4. Field calculation and dipole emission near planar interfaces

Dipole emission near planar interfaces was carefully investigated by Novotny and summarized in the book *Principles of Nanooptics*, Chap. 10 [30–32]. We briefly summarize the most important equations and steps.

The electric dipole moment of a single Rayleigh scatterer is given by

$$\vec{p} = \varepsilon_0 \varepsilon_c \alpha \vec{E}_0, \tag{D5}$$

with ε_0 being the vacuum permittivity and ε_c standing for the permittivity of the cover medium. In this equation, α is the static polarizability of the protein. This is not to be confused with the attenuation constant for which the symbol α is used throughout the rest of this publication. The

polarizibilty is defined as

$$\alpha = 4\pi a^3 \frac{\varepsilon_P - \varepsilon_c}{\varepsilon_P + 2\varepsilon_c}, \tag{D6}$$

with a being the radius and ε_P standing for the permittivity of the protein sphere.

The electric field of a single dipole in the allowed region in the substrate can be computed from the dipole potentials [30]:

$$\vec{E} = \begin{bmatrix} E_\theta \\ E_\phi \end{bmatrix}$$
$$= \frac{k_c^2}{4\pi\varepsilon_0\varepsilon_c} \frac{e^{ik_s r}}{r}$$
$$\times \begin{bmatrix} (p_x\cos\phi + p_y\sin\phi)\cos\theta\,\Phi_s^{(2)} - p_z\cos\theta\,\Phi_s^{(1)} \\ -(p_x\sin\phi - p_y\cos\phi)\,\Phi_s^{(3)} \end{bmatrix} \tag{D7}$$

and in Cartesian coordinates

$$\vec{E} = \begin{bmatrix} E_x \\ E_y \\ E_z \end{bmatrix} = \begin{bmatrix} \cos\theta\cos\phi & -\sin\phi \\ \cos\theta\sin\phi & \cos\phi \\ -\sin\theta & 0 \end{bmatrix} \begin{bmatrix} E_\theta \\ E_\phi \end{bmatrix}. \tag{D8}$$

p_x, p_y, and p_z are the components of the dipole moment vector. A dipole excited by the TE mode of a waveguide is oriented along p_x only, and therefore the above equations can be simplified.

The dipole potentials for the substrate half space were defined by Novotny and Hecht [30]:

$$\Phi_s^{(1)} = \frac{n_s}{n_c}\frac{\cos\theta}{\tilde{s}_z(\theta)} t^{(p)}(\theta) e^{ik_s[z_0\tilde{s}_z(\theta)+d\cos\theta]}, \tag{D9a}$$

$$\Phi_s^{(2)} = \frac{n_s}{n_c} t^{(p)}(\theta) e^{ik_s[z_0\tilde{s}_z(\theta)+d\cos\theta]}, \tag{D9b}$$

$$\Phi_s^{(3)} = \frac{\cos\theta}{\tilde{s}_z(\theta)} t^{(s)}(\theta) e^{ik_s[z_0\tilde{s}_z(\theta)+d\cos\theta]}. \tag{D9c}$$

Note: The second potential has an incorrect sign in [30]. The authors confirmed this and the equations stated here are correct. The Fresnel's transmission coefficients for a two interface structure were derived by Lukosz [44]:

$$t^{(p,s)} = \frac{t_{c,f}^{(p,s)} t_{f,s}^{(p,s)} \exp\left(ik_{z,f} t_f\right)}{1 + r_{c,f}^{(p,s)} r_{f,s}^{(p,s)} \exp\left(2ik_{z,f} t_f\right)}. \tag{D10}$$

With Fresnel's reflection and transmission coefficients for single interfaces defined as

$$t_{n,m}^{(s)}(\theta) = \frac{2k_{z,n}}{k_{z,n} + k_{z,m}}, \tag{D11a}$$

$$t_{n,m}^{(p)}(\theta) = \frac{2\varepsilon_m k_{z,n}}{\varepsilon_m k_{z,n} + \varepsilon_n k_{z,m}}\sqrt{\frac{\varepsilon_n}{\varepsilon_m}}, \tag{D11b}$$

$$r_{n,m}^{(s)}(\theta) = \frac{k_{z,n} - k_{z,m}}{k_{z,n} + k_{z,m}}, \tag{D11c}$$

$$r_{n,m}^{(p)}(\theta) = \frac{\varepsilon_m k_{z,n} - \varepsilon_n k_{z,m}}{\varepsilon_m k_{z,n} + \varepsilon_n k_{z,m}}, \tag{D11d}$$

the factor $\tilde{s}_z(\theta)$ is defined as

$$\tilde{s}_z(\theta) = \sqrt{\left(\frac{n_c}{n_s}\right)^2 - \sin^2\theta}. \tag{D12}$$

Finally, the total intensity in the molographic focal spot is then given by the coherent superposition of the scattered fields of individual dipoles:

$$I_{\text{avg}} = \frac{1}{2}\sqrt{\frac{\varepsilon_0\varepsilon_s}{\mu_0}} \sum_{i=1}^{N_P} \vec{E}_i(\vec{r}-\vec{r}_0) \sum_{i=1}^{N_P} \vec{E}_i^*(\vec{r}-\vec{r}_0), \tag{D13}$$

where N_P is the number of coherently assembled particles.

5. Implementation

The aggregated scattered intensities are calculated on a finite plane (screen) at regular, discrete space intervals (pixels). The computation consists of the coherent summation of the fields of independent Rayleigh scatterers according to Eq. (D13) for each individual pixel. Therefore, we make use of the parallel computation power of GPUs. We use a NVIDIA GTX 1080 Ti card and program it with its dedicated CUDA language (7.5 toolkit version). No additional library other than those provided per default by the toolkit are used. The complex computation library is not used, but the equations are broken down to their complex and imaginary parts and treated separately. Special care needs to be taken to use atomic additions when aggregating the fields on a given pixel. The low memory cost and the speed up provided by this approach allow the computation of 231 million protein scatterers on a screen of 150 x 150 pixels in only 1650 s. A further speed up could be achieved by optimizing the memory traffic during the computation, for example, by using shared memory and taking advantage of thread warping.

APPENDIX E: MOLOGRAPHIC SIGNAL INTENSITY FROM COUPLED MODE THEORY AND RAYLEIGH SCATTERING

The molographic signal (average intensity in the Airy disk) can be computed from CMT, but also via summation

of Rayleigh scatterers (neglecting the interface effects), which give basically the same result. Here, we outline the procedure to arrive at these expressions. To obtain an expression of the intensity in the focal spot via CMT, we start from Eq. (6) in Fattinger. By using $\Lambda \approx \lambda/n_c$ and substituting $\frac{\partial N}{\partial n_c}$ with Eq. (7) in Fattinger [16], we get

$$P_{\text{diff}} = 2\pi \frac{D^2}{\lambda^2} \left(\frac{dn}{dc}\right)^2 \Delta_{\Gamma_{\text{can}}}^2 \frac{n_c \left(n_f^2 - N^2\right)}{N t_{\text{eff}} \left(n_f^2 - n_c^2\right)} P_{\text{WG}}. \quad (E1)$$

There are a few differences compared to the equation in Fattinger [16]. First, P_{diff} is the total power coupled out in Watts, whereas Fattinger defined it as a power per line (W/m). We chose this definition because it is the total power that is then focused onto the Airy disk and not a power per line. We still keep the definition of P_{WG} as power per line (W/m). $\Delta_{\Gamma_{\text{can}}}$ denotes the canonical surface mass modulation. This definition is by a factor of two higher than the original one used in Fattinger [16], which corresponds to the molographic surface mass density here. In order to determine the intensity in the focal spot, we realize that the diffracted power is distributed over the Airy disk. From the properties of the Bessel function of the first kind that describe the shape of the Airy disk, it follows that 83.8% of the total diffracted power is contained within the first zeros of the Bessel function [37]:

$$I_{\text{avg}} = 0.838 \frac{P_{\text{diff}}}{A_{\text{Airy}}}. \quad (E2)$$

The area of the Airy disk is given by the readout wavelength and the NA of the system [70]:

$$A_{\text{Airy}} = 0.37\pi \frac{\lambda^2}{\text{NA}^2}, \quad (E3)$$

with the approximative numerical aperture NA $= Dn_s/2f$. Therefore, the average intensity on the Airy disk calculated via CMT reads

$$I_{\text{avg,CMT}} = 4.53 \text{NA}^2 \frac{n_c \left(n_f^2 - N^2\right)}{t_{\text{eff}} N \left(n_f^2 - n_c^2\right)} \left(\frac{dn}{dc}\right)^2$$

$$\times \frac{D^2}{\lambda^4} \Delta_{\Gamma_{\text{can}}}^2 P_{\text{WG}}. \quad (E4)$$

To see the analogy to Rayleigh scattering, we rewrite the power in the waveguide as the intensity on the waveguide

surface as

$$I_0 = \frac{n_c}{2Z_w} A^2 = n_c \frac{2 \left(n_f^2 - N^2\right)}{N t_{\text{eff}} \left(n_f^2 - n_c^2\right)} P_{\text{WG}}, \quad (E5)$$

and furthermore generalize the canonical surface mass density modulation with the analyte efficiency of the modulation $\eta_{\text{mod}[A]}$,

$$I_{\text{avg,CMT}} = 5.59 \text{NA}^2 \left(\frac{dn}{dc}\right)^2 \frac{D^2}{\lambda^4} \eta_{\text{mod}[A]}^2 \Delta_\Gamma^2 I_0. \quad (E6)$$

The expression for Rayleigh scattering in plane-polarized light [36] can be easily obtained from Eq. (1) in Ref. [16]:

$$I_{\text{Airy,max,RS}} = 9\pi^2 \frac{n_c^4}{f^2 \lambda^4} \frac{1}{\rho_P^2} \frac{M_P^2 N_P^2}{N_A^2} \frac{\left(n_P^2 - n_c^2\right)^2}{\left(n_P^2 + 2n_c^2\right)^2} I_0, \quad (E7)$$

where we have replaced the volume of a single protein according to $V_P = M_P/(\rho_P N_A)$ and generalized the expression to an ensemble of N_P coherent proteins. It has to be noted that Eq. (2) in Ref. [16], which is used to calculate the volume of the protein from the refractive-index increment, is an approximation. Equations (D4) should be used instead.

Equation (E7) describes the maximum intensity in the Airy disk and not the average intensity. The ratio between the maximum and the average can be readily calculated by integration over the Airy disk such that

$$I_{\text{Airy}_{\text{max}}} = \frac{\int_0^{3.83} r\,dr}{\int_0^{3.83} \left(\frac{2J_1(r)}{r}\right)^2 r\,dr} I_{\text{avg}} = 4.378 I_{\text{avg}}, \quad (E8)$$

where J_1 is the Bessel function of the first kind. Inserting Eq. (E7) for $I_{\text{Airy}_{\text{max}}}$ yields

$$I_{\text{avg,RS}} = 2.056\pi^2 \frac{n_c^4}{f^2 \lambda^4} \frac{1}{\rho_P^2} \frac{M_P^2 N_P^2}{N_A^2} \frac{\left(n_P^2 - n_c^2\right)^2}{\left(n_P^2 + 2n_c^2\right)^2} I_0. \quad (E9)$$

Using the definitions for the NA NA $= \frac{n_c D}{2f}$ and the conversion from protein number to surface mass density modulation $\frac{M_P^2 N_P^2}{N_A^2} = \frac{D^4 \pi^2}{64} \eta_{\text{mod}[A]}^2 \Delta_\Gamma^2$ results in

$$I_{\text{avg,RS}} = 1.268\pi^2 n_c^2 \frac{\left(n_P^2 - n_c^2\right)^2}{\left(n_P^2 + 2n_c^2\right)^2} \text{NA}^2$$

$$\times \frac{1}{\rho_P^2} \frac{D^2}{\lambda^4} \eta_{\text{mod}[A]}^2 \Delta_\Gamma^2 I_0. \quad (E10)$$
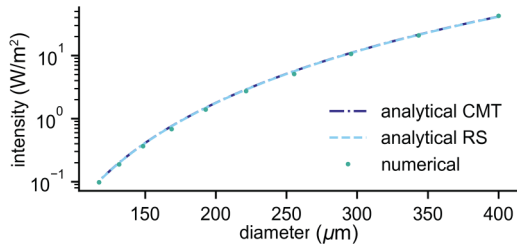
FIG. 11. Comparison of the different models in water to predict the average intensity in the Airy disk (simulation parameters are the same as the ones described in Fig. 5, which is in air). One can see that the two analytical models nearly perfectly match the numerical calculations.

And finally, with the definition of the refractive index increment for proteins (Eq. (4)), one obtains

$$I_{\text{avg,RS}} = 5.56 \text{NA}^2 \left(\frac{dn}{dc}\right)^2 \frac{D^2}{\lambda^4} \eta_{\text{mod}[A]}^2 \Delta_\Gamma^2 I_0, \quad \text{(E11)}$$

which is beside a small numerical error the same as the expression from CMT (Eq. (E6)).

Measurements in air are an unusual case in label-free investigations and are only chosen here for the sake of experimental simplicity. In the practical settings, namely in water, both of these models nearly perfectly coincide with the numerical computations that take the interfaces into account (Fig. (11)).

APPENDIX F: DEPHASING AS A LIMITATION ON MOLOGRAM SIZE

As described in the main manuscript, we expect the amount of sample volume to limit the mologram size to diameters <1 mm. However, it is already nontrivial to achieve diffraction-limited focusing for mologram up to 400-μm diameter on high-refractive-index waveguides, since gradients in effective refractive index can also result in a limitation on mologram size. The most prominent source for a gradient in effective index stems from the wedgedness of the waveguide [see Fig. 12(c)]. Due to the bell-shaped deposition curve of most sputtering systems used in research [71] or due to the thickness gradients introduced by the coating process in Sol-Gel waveguide fabrication methods (spin coating) [72], the resulting waveguides exhibit a certain degree of wedging. For sputtering fabrication, this is less of a problem if the deposition is performed by industrial-scale large-area sputtering systems. The wedging of the waveguide introduces a gradient in effective refractive index. This leads to an accumulated phase mismatch between the synthetic hologram (designed for constant effective refractive index) and the guided mode. After a certain propagation distance, which we call the dephasing length, light scattered at the first and
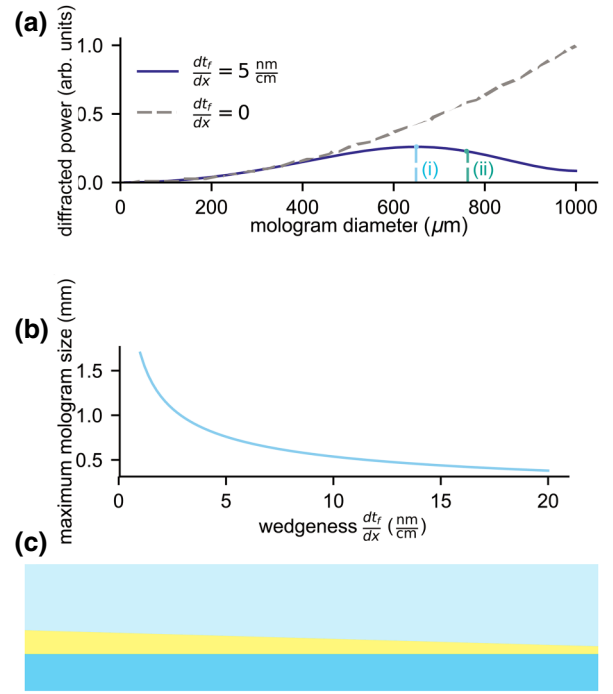

FIG. 12. (a) Diffracted power of a sinusoidal grating as a function of mologram size for a waveguide with no effective index gradient (gray) and with a wedgedness of 5 nm/cm: (i) maximum reasonable mologram size, (ii) dephasing length. (b) Maximum mologram size for different waveguide wedgedness. The calculations are performed for $n_f = 2.117$, $n_c = 1$, $n_s = 1.521$, and $\lambda = 632.8$ nm. (c) Illustration of a waveguide with a wedgedness.

the last lines of the mologram interferes destructively (see Fig. 12).

We can use CMT to address the effect of dephasing on the diffracted power. For simplicity, we model the mologram as a sinusoidal grating with a constant period. We start from Eqs. (3.4-8), (3.4-10), and (5.2-1) in Marcuse [29]. However, we adapt these expressions for the case of a guided mode with a nonconstant phase:

$$P_{\text{diff}} = \langle |c|^2 \rangle \propto \left\langle \left| \int_0^D f(z) e^{i[(\beta_{\text{in}} - \beta_{\text{out}})z + \phi(z)]} dz \right|^2 \right\rangle, \quad \text{(F1)}$$

with $f(z) = \sin(2\pi/\Lambda)z = e^{-i(2\pi/\Lambda)z}$ being the grating function of a sinusoidal grating. For a constant effective refractive index gradient, the accumulated phase mismatch reads

$$\phi(z) = -\int_0^z k_0 \frac{dN}{dz} z' dz' = -k_0 \frac{dN}{dz} \frac{z^2}{2}. \quad \text{(F2)}$$

We assume the detuning parameter $d = \beta_{\text{in}} - \beta_{\text{out}} - 2\pi/\Lambda$ over the entire waveguide to be constant. Then we

can write

$$P_{\text{diff}} \propto \left\langle \left| \int_0^D e^{idz-isz^2} \, dz \right|^2 \right\rangle, \tag{F3}$$

where $s = -(k_0/2)(dN/dz)$ is the dephasing parameter. After integration and multiplying with the complex conjugate we obtain

$$P_{\text{diff}} \propto \frac{\pi e^{id^2/2s}}{2s} \left\{ \left(\text{FC}\left[\frac{d}{\sqrt{2\pi}\sqrt{s}}\right] - \text{FC}\left[\frac{d-2Ls}{\sqrt{2\pi}\sqrt{s}}\right] \right)^2 \right.$$
$$\left. - \left(\text{FS}\left[\frac{d}{\sqrt{2\pi}\sqrt{s}}\right] - \text{FS}\left[\frac{d-2Ls}{\sqrt{2\pi}\sqrt{s}}\right] \right)^2 \right\}, \tag{F4}$$

where FS and FC represent the Fresnel integrals. Figure 12(a) shows the diffracted power as a function of the mologram diameter for a waveguide with a wedgedness of 5 nm/cm. There is no reason for fabricating a mologram larger than the size at which the diffracted power is at the maximum (blue). However, this term is laborious to calculate and we, therefore, define another parameter to determine the phase mismatch. We define the dephasing length as the mologram size at which the first

and the last molographic lines will interfere completely destructively. Because we assume the change in waveguide thickness to be the main source of the effective refractive index gradient, we substitute $dN/dz = (dN/dt_f)(dt_f/dz)$ into Eq. (F2). The first and the last lines will interfere destructively when the phase mismatch builds up to π. Solving

$$\frac{2\pi}{\lambda} \frac{dN}{dt_f} \frac{dt_f}{dz} \frac{D_{\text{deph}}^2}{2} = \pi \tag{F5}$$

yields an expression for the dephasing length:

$$D_{\text{deph}} = \sqrt{\frac{\lambda}{\frac{dN}{dt_f}\frac{dt_f}{dz}}}. \tag{F6}$$

The dephasing length is indicated in green. From Fig. 12(a), it is apparent that the waveguides used for our experiments exhibit a wedgedness of less than 5 nm/cm because we did not observe any effect of the dephasing, which would have led to a deviation for larger molograms between the analytical model and the experiments in (see Fig. 5). In Fig. 12(b), the maximum mologram size as a function of wedgedness is shown.
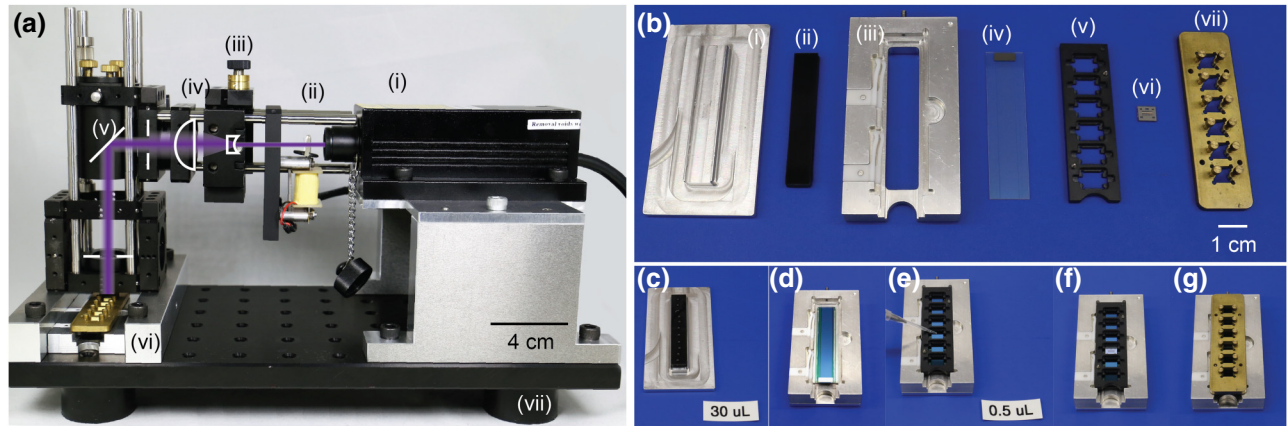

FIG. 13.   Reactive immersion lithography setup. (a) The second generation of the illumination setup for the creation of molograms. (i) A 405-nm solid-state laser beam with a coherence length of a few 10 μm (Model RLTMLL-405-200-5 200 mW, Roithner Lasertechnik) is expanded by a factor 8.3 (coherence lengths longer than this only create unwanted interferences with reflected beams). (ii) Home-built solenoid shutter that controls the illumination time. (iii) Plano-concave lens (PCV) $f = -6$ mm lens (x-y moveable, VIS 0° AR coated, #48265 Edmund optics). (iv) Plano-convex lens (PCX) lens ($f = 50$ mm, 405 nm V-Coat, #65465 Edmund optics), and deflected by an adjustable mirror (G063713000, Qioptiq) (v) to the chip and phase mask holder (vi). (vii) Sorbothane mounts (Edmund optics) for vibration isolation. (b) Parts of the chip phase mask assembly holder. (i) Aluminium support for black polydimethylsiloxane (PDMS) piece. (ii) Carbon black (Cabot Corporation) loaded (0.1 wt%) PDMS (Silgard 184, Dow Corning) piece in order to absorb transmitted radiation and avoid reflections. (iii) Aluminium holder with springs to secure molographic chip and mask holder. (iv) Waveguide molographic chip. (v) Machined polyetheretherketone (PEEK) mask holder for six phase masks. (vi) Phase mask chip featuring thirty molographic structures. (vii) Brass mask to fix the chip with two additional brass pillars to avoid lateral movement during illumination. (c)–(g) Assembly of the chip and phase mask holder for lithography. (c) 30 μl of DMSO are applied to the PDMS piece to provide index matching between the chip back surface and the PDMS. (d) Waveguide chip inserted into the aluminium holder. (e) Mask holder is inserted and 0.5 μl of the immersion medium (0.2% hydroxylamine in DMSO) is pipetted onto the molographic field. (f) One of the six phase mask inserted into the holder. (g) The final chip phase mask assembly as it is used in the illumination setup.
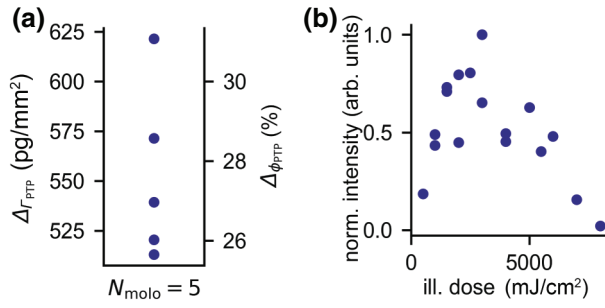
FIG. 14. Characterization of the illumination setup. (a) Activation and surface mass density modulation from STED images. The exposure dose is 2000 mJ/cm^2 and five different molograms are investigated in the same way as described in Supplementary Fig. 6 of our previous work [7]. The median of the activation modulation $\Delta_{\phi_{PTP}}$ amounts to 27% with a 95% confidence in the interval [25.6, 31]%. The median of the peak-to-peak surface mass density modulation $\Delta_{\Gamma_{PTP}}$ amounts to 540 pg/mm^2 when incubated with 1 μM of SAv (2000 pg/mm^2 surface coverage for 100% activation) with a confidence of 95% to be in the interval [513, 621] pg/mm^2. (b) The measured mologram intensity as a function of illumination dose at 405 nm for SAv molograms. The illumination dose of 2000 mJ/cm^2 used in this manuscript is close to the maximum of the curve and has the advantage that the ridges are not yet fully saturated.

APPENDIX G: ILLUMINATION SETUP AND FABRICATION OF MOLOGRAMS

We noted in our last publication that molograms with a better activation ratio could be fabricated by using a higher spatial and temporal coherence of the illumination source [7]. To achieve this, here, a laser is used as light source. A different wavelength of 405 nm is chosen since these laser sources are cheap thanks to the blue ray disk. We use the same phase masks as in the last publication (optimized for 390-nm illumination) and nevertheless obtain molograms with nearly twice the surface mass modulation compared to the previous results (27% vs 14% of the ideally achievable modulation). The optimal illumination dose lies between 2500–3000 mJ/cm^2 (compared to 284 mJ/cm^2 at 390 nm). However, the achieved mass modulation is only weakly dependent on the illumination dose in the range 1800–3500 mJ/cm^2, resulting in a robust and reproducible illumination process. The illumination setup (second generation) is displayed in Fig. 13. The mass density modulation for a dose of 2000 mJ/cm^2 determined by STED and the molographic signal as a function of illumination dose are shown in Fig. 14 (same procedure as in [7]).

APPENDIX H: MOLOGRAPHIC MEASUREMENTS IN AIR

Molographic measurements in air, waveguide characterizations, and speckle background analysis are carried out with the MoloReader in the configuration shown in Fig. 15. The entire setup is motorized such that all six measurement fields of a chip and all molograms on each of them can be separately measured. The arrangement of motors, microscope, and the fiber-coupled laser are displayed in Fig. 15(a). The arrangement can be separated in a moveable microscopy assembly [x,z] and a moveable chip assembly [y]. Each of the stepper motors has a small manual microstage (UMR3.5, Newport) on the other assembly for fine adjustments (microscope to beam alignment and coupling grating to center of rotation alignment). A MetalVelvet^{TM} Adhesive (ACM coatings)-coated aluminium cover with openings for the coupling beam is used to block the direct stray light [Fig. 15(b)]. This stray light mainly originates from the adjustable aperture whose function is to control the beam diameter. Figure 15(c) shows a propagating TE mode and the small MetalVelvet^{TM} Adhesive pieces that are used to protect the microscope from excessive stray light as well as the photodiode for measuring the out-coupled power.

1. Effective focal point distance and adjusted depth of field

The distance measured between the chip surface and the focal point differ from the focal distance that is expected by the design of the mologram (900 μm). This is a consequence of the refraction of light due to the optical interface. As evident from Fig. 16, two aspects contribute to the difference between the measured and the designed focal length: First, the distance seen by the objective appears to be shorter than the real distance due to the transmission into a denser medium [Fig. 16(a)]. By knowledge of the real distance (chip thickness: 700 μm), the apparent chip thickness can be calculated as

$$d_a = \frac{n_s}{n_o} d_r, \tag{H1}$$

which amounts to a value of 460 μm. Second, the scattered rays are refracted at the bottom interface of the chip. The remaining 200 μm from the bottom surface will be compressed by the refractive index of the glass substrate by $1/n_s$, yielding a focal distance of 831 μm with respect to the chip surface [see Fig. 16(b) and 16(c)], such that the apparent focal distance is approximately at 592 μm. Evidently, this exactly corresponds to the focal length of this mologram on the air side as requested in order to exhibit the same NA on both sides by Snells law.

2. Shape of the Airy disk with and without central Bragg recess area

Figure 17 shows that the distortion of the Airy disk compared to a diffraction-limited lens is mainly caused by the central Bragg area which breaks the radial symmetry.
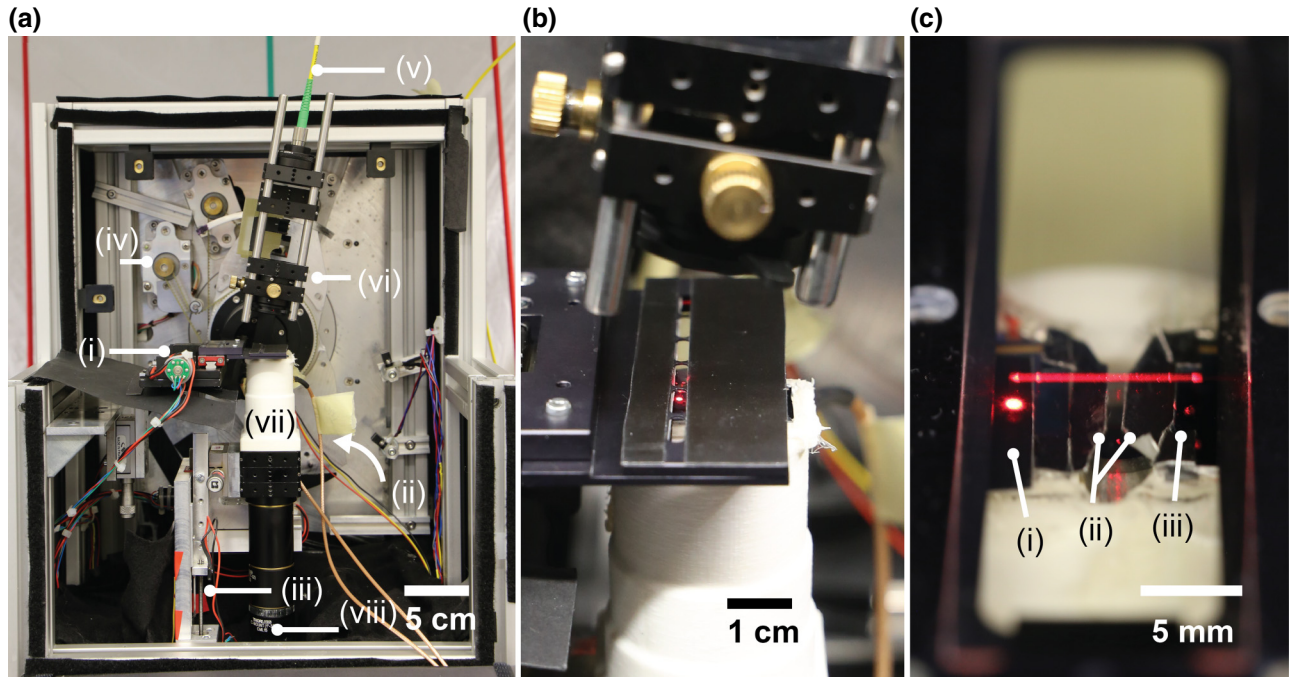
FIG. 15. MoloReader configuration for measurements in air. (a) Stepper motor for movement of chip y axis (i), movement of microscope x (ii), and z axes (iii), and for adjusting the in-coupling angle (iv). Motors (i)–(iii) were obtained from Faulhaber Inc, models (i)–(ii) AM1524 (M3x0.5 thread and M2x0.2 thread) and AM2224 (M3x0.5 thread) for (iii) controlled by MCST 3601 motor controllers, motor (iv) is an ancient RDM545 model (Berger Lahr) controlled by a Raspberry Pi 2 Model B and a D225 5 phase stepper motor control card. (v) Angled Physical Contact (APC) fiber-coupled He-Ne Laser (HNL020L-EC 2mW, Thorlabs). (vi) Moveable adjustable aperture (set to 1 mm for most of the experiments). Imaging system consisting of a standard DIN ×20 0.4 NA achromatic objective (Edmund Optics) (vii) and a C-Mount tube with CMOS camera (viii) (IDS Imaging Development Systems GmbH, Model UI-3480CP Rev.2, 2.2-μm pixel size, 2560 × 1920 pixels). (b) Chip holder with straylight cover. This setup configuration is used to acquire all the experimental data except the noise level measurements in Fig. 9 and movies. (c) Zeptochip with propagating TE mode. (i) Standard MetalVelvet$^{\text{TM}}$ Adhesive (ACM Coatings) to absorb most of the in-coupled light beneath the chip. (ii) Additional light-absorbing adhesive to protect the microscopy objective against stray light from the in-coupling and out-coupling grating. (iii) Anode-grounded photodiode to measure the out-coupled light (FDS1010, Thorlabs) via a bias module (PBM42, Thorlabs) and a data acquisition (DAQ) system (U6, Labjack). Load resistance is adapted to the light levels (1 kΩ).

3. Power calculations for the molograms in Fig. 5

The power in the waveguide is calculated from the in-coupled beam diameter (1 mm set by an adjustable aperture Fig. 15) and the out-coupled power measured by a photodiode. The distance of the mologram out-coupling grating is different for each mologram (spacing 400 μm, average distance to out-coupling grating is 6 mm) and the power in the waveguide at the location of the mologram is calculated by the determined damping constant of this waveguide (6.5 dB/cm) by the MoloReader.

APPENDIX I: MODELS FOR BACKGROUND SCATTERING

1. Approximative formula for the background intensity in the focal spot

The exact equation to calculate the intensity due to a distributed scattering mechanism giving rise to an isotropic scattering leakage α_{sca} is

$$I_{\text{BG}} = \alpha_{\text{sca}} P_{\text{WG}} \int_0^{2\pi} \int_0^{D/2} \frac{e^{-\alpha(\rho \cos \phi)}}{4\pi(f^2 + \rho^2)} \rho \, d\rho \, d\phi. \quad (\text{I1})$$

However, there is no analytical solution for this expression. To simplify the calculations, we consider the entire background power as a spherical wave originating from the middle of the mologram, which is then collected by an objective of a given NA. We essentially neglect free-space attenuation. However, this effect is small and we also neglect free-space attenuation of the signal. Signal and background are affected in the same way by free-space attenuation. To obtain an analytical expression, we approximate the total radiated background power (in watts) by

$$P_{\text{rad}} = D P_{\text{WG}} \left[1 - \exp\left(-\alpha_{\text{sca}} \frac{D\pi}{4} \right) \right] \approx \frac{D^2 \alpha_{\text{sca}} \pi}{4} P_{\text{WG}}. \quad (\text{I2})$$
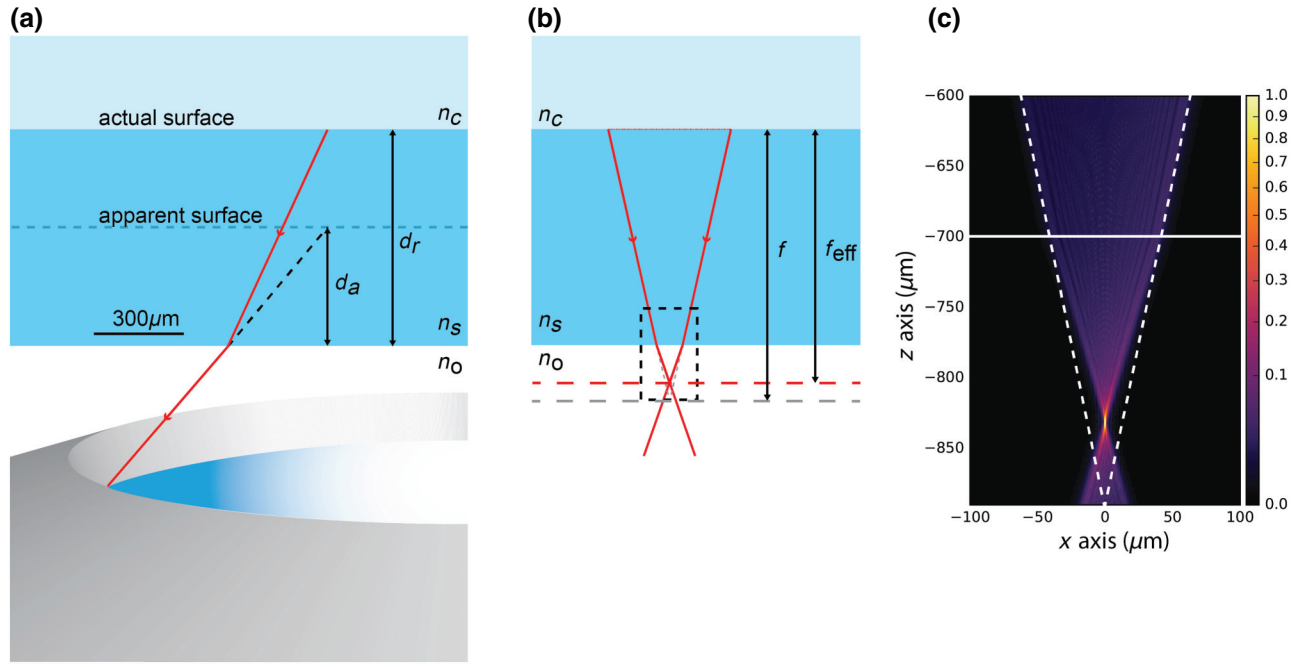
(a) **(b)** **(c)**

FIG. 16. Difference of the measured focal plane and the designed focal plane. (a) The optical interface induces a difference between the real and the apparent chip surface. The mologram is designed for a focal distance of 900 μm in a medium of $n_s = 1.521$. (b) Since the chip thickness is only 700 μm, the rays are refracted due to the refractive index change of the media ($n_s = 1.521$, $n_0 = 1$). (c) An upward shift of the focal point is clearly observable if the change of refractive index is introduced in the simulations. The distance measured between the chip surface and the focal point f_{eff} is approximately 592 μm, which coincides with the calculations.

Since we assume that this power is distributed over a solid angle of 4π, the intensity in the focal spot can be expressed in terms of the NA of the objective:

$$I_{\text{BG}} = \frac{P_{\text{rad}} n_s^2}{4\pi f^2} = \frac{D^2 n_s^2 \alpha_{\text{sca}}}{16 f^2} P_{\text{WG}} = \frac{1}{4} \text{NA}^2 \alpha_{\text{sca}} P_{\text{WG}}. \quad \text{(I3)}$$

This equation underestimates the total intensity in the focal spot. Surprisingly, the error does not depend on the leakage parameter. It only depends on the NA, with larger errors for larger NAs as expected. The error for a 0.1 NA objective is below 1% and rises to roughly 25% for 0.6 NA objectives. Yet, as described in the main manuscript, the scattering is not isotropic. Therefore, Eq. (I3) is corrected by the anisotropy parameter a_{ani} to yield Eq. (5).

2. Sidewall roughness scattering

In this section, we derive expressions for the attenuation constant and the scattering leakage. They can be computed from the statistical properties of the interfaces—namely the correlation length L_c and the rms roughness σ. The derivation is adopted from [52,56]. It is generalized to cover the case of an asymmetric slab waveguide. In these papers, the difference in notation compared to this publication is worth noticing. The thickness of the waveguide film t_f is defined as $2d$ and the coordinate system is defined in

a different manner (to transform from our coordinate system to theirs: $x \to z$, $y \to x$, $z \to y$). The refractive-index distribution of an asymmetric waveguide can be written in the following way:

$$n^2 = n_c^2 + \left(n_f^2 - n_c^2 \right) U[f(x) - z]$$
$$- \left(n_f^2 - n_s^2 \right) U\left[-\left(t_f + f(x) + z \right) \right], \quad \text{(I4)}$$

where U is the unit step function and $f(z)$ is the function of the sidewall roughness: $U[a] = 0$ if $a < 0$ and $U[a] = 1$ if $a > 0$. Following the procedure in Ref. [56], this leads to a slightly different solution for the perturbed fields of the asymmetric waveguide compared to the symmetric case (Eq. (7) in Ref. [56]):

$$E_y(x,z) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} k_0^2 \left(n_f^2 - n_c^2 \right) U[f(x') - z']$$
$$\times E_{y0}(x,z) G(x,x',z,z') dy dz'$$
$$- \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} k_0^2 \left(n_f^2 - n_s^2 \right) U\left[-\left(t_f + f(x') + z' \right) \right]$$
$$\times E_{y0}(x,z) G(x,x',z,z') dy dz', \quad \text{(I5)}$$

where G is the Green's function of the waveguide. As in Ref. [56], we will assume the two sidewalls to be noncorrelated. Thus, we can treat them separately and add their
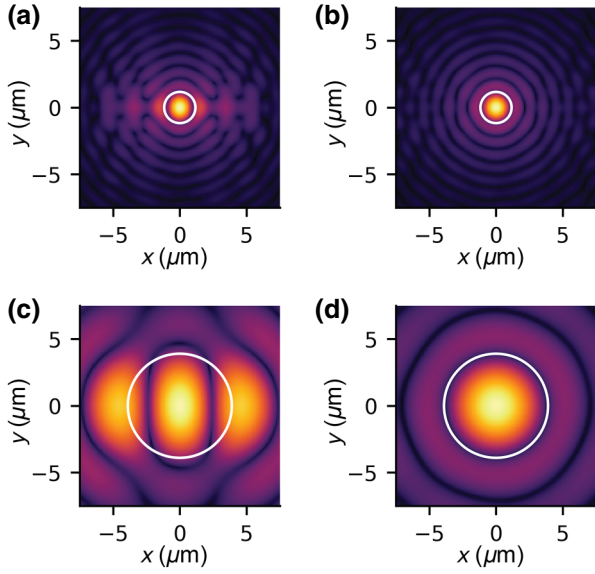
FIG. 17. Shape of the Airy disk for two molograms (0.33 and 0.1 NAs) with (a), (c) and without central (b), (d) Bragg recess area. The white circle is the size of the Airy disk of the diffraction-limited lens with the same NA.

contribution as done in Ref. [56]. Therefore, we can solve the integral over dz' [see Eq. (9) in Ref. [56]] using a first-order Taylor approximation resulting in

$$
E_y(x,z) = \frac{\sqrt{2}}{2}k_0^2 \left[\phi(0)\left(n_f^2 - n_c^2\right) + \phi\left(-t_f\right)\left(n_f^2 - n_s^2\right) \right]
$$
$$
\times \int_{-\infty}^{\infty} e^{-i\beta x'} G(x,x',z,z')dx', \tag{I6}
$$

where ϕ is the normalized electric field of the unperturbed waveguide:

$$
\phi^2(z) = \frac{\left|E_y(z)\right|^2}{\int_{-\infty}^{\infty}\left|E_y(z)\right|^2 dz}. \tag{I7}
$$

For the integration over dx', we make use of the Fourier transform of the Green's function. Thus, Eq. (I6) is Fourier transformed to the wavenumber domain [k_x, k_z]. Then the equation is transferred back into the spacial [x, z] domain by the inverse Fourier transform. For the inverse Fourier transformation we need to consider that, in contrast to the symmetric waveguide covered by Ref. [56], scattering into the upper half space is not identical to scattering into the lower half space. Thus, the integral over dk_z [dk_y in Ref. [56] Eq. (11)] needs to be split into two separate integrals. One covers scattering into the upper half space and the other one covers the scattering into the lower half space. However, because we are not interested in the phase information and due to the even character of the real part,

the result is not complicated very much:

$$
\mathrm{Re}\left\{ \int_0^\infty \frac{e^{ik_z z}}{n_c^2 k_0^2 - k_y^2 - k_z^2}dk_z + \int_{-\infty}^0 \frac{e^{ik_z z}}{n_s^2 k_0^2 - k_y^2 - k_z^2}dk_z \right\}
$$
$$
= \frac{1}{2}\mathrm{Re}\left\{ \int_{-\infty}^\infty \frac{e^{ik_z z}}{n_c^2 k_0^2 - k_y^2 - k_z^2}dk_z \right.
$$
$$
\left. + \int_{-\infty}^\infty \frac{e^{ik_z z}}{n_s^2 k_0^2 - k_y^2 - k_z^2}dk_z \right\}. \tag{I8}
$$

We end up with four different contributions to the attenuation constant, one for each combination of sidewall and upper, respectively, lower half-space scattering. We can apply all further steps in Ref. [56] to each of the four contributions, and therefore directly write our result in the form of Eq. (16) in Ref. [56]. The factorization of the four contributions reads as

$$
\alpha_{\mathrm{sca,r}} = \frac{1}{4}\frac{k_0^3}{4\pi n_f}
$$
$$
\times \left[\phi^2(0)\left(n_f^2 - n_c^2\right)^2 + \phi^2\left(-t_f\right)\left(n_f^2 - n_c^2\right)^2 \right]
$$
$$
\times \left(\int_0^\pi \int_{-\infty}^\infty R(u)e^{i(\beta - n_c k_0 \cos\theta)u}dud\theta \right.
$$
$$
\left. + \int_0^\pi \int_{-\infty}^\infty R(u)e^{i(\beta - n_s k_0 \cos\theta)u}dud\theta \right), \tag{I9}
$$

where the first bracket contains the two sidewall contributions and the second one contains the upper and lower half-space scattering. $R(u)$ is the correlation function of the surface roughness. The angle θ is defined such that it is zero in the forward scattering direction. Next, we use the expressions [Eqs. (1.3-16), (1.3-17), and (1.3-46, 47)] given in Ref. [29] for the electric field of the fundamental TE mode of the asymmetric waveguide to get an expression for the normalized fields at the two interfaces $\phi^2(0)$ and $\phi^2\left(-t_f\right)$:

$$
\phi^2(0) = \frac{A^2}{|E_{\mathrm{tot}}|^2} = \frac{A^2 N}{2\mu_0 c_0 P} = \frac{4\kappa^2 \omega N}{2|\beta| t_{\mathrm{eff}}\left(\kappa^2 + \delta^2\right)c_0}, \tag{I10a}
$$

$$
\phi^2\left(-t_f\right) = \frac{\left[A\cos(\kappa t_f) - B\sin(\kappa t_f)\right]^2}{|E_{\mathrm{tot}}|^2}
$$
$$
= \left[\cos(\kappa t_f) + \frac{\delta}{\kappa}\sin(\kappa t_f)\right]^2
$$
$$
\times \frac{4\kappa^2 \omega N}{2|\beta|\left(d + \frac{1}{\gamma} + \frac{1}{\delta}\right)\left(\kappa^2 + \delta^2\right)c_0}, \tag{I10b}
$$

where the relation of the power and the total field is just the integrated Poynting flux in z and γ, κ and δ are the z components of the wave vector k in the three media: substrate, film, and cover [29]. They can be expressed as a function of their corresponding refractive index and of the effective refractive index N:

$$\gamma^2 = \beta^2 \frac{N^2 - n_s^2}{N^2}, \tag{I11a}$$

$$\kappa^2 = \beta^2 \frac{n_f^2 - N^2}{N^2}, \tag{I11b}$$

$$\delta^2 = \beta^2 \frac{N^2 - n_c^2}{N^2}. \tag{I11c}$$

Furthermore, it is convenient to state the following relation:

$$\frac{\kappa^2}{\kappa^2 + \delta^2} = \frac{n_f^2 - N^2}{n_f^2 - n_c^2}, \tag{I12}$$

which allows the normalized fields at the two interfaces to be written in a more concise way.

$$\phi^2(0) = \frac{2\left(n_f^2 - N^2\right)}{t_{\text{eff}}\left(n_f^2 - n_c^2\right)}, \tag{I13a}$$

$$\phi^2(-t_f) = \frac{2\left(n_f^2 - N^2\right)}{t_{\text{eff}}\left(n_f^2 - n_c^2\right)} \left\{ \cos\left[t_f \beta \frac{\left(n_f^2 - N^2\right)^{1/2}}{N} \right] \right. $$
$$\left. + \frac{\left(N^2 - n_c^2\right)^{1/2}}{\left(n_f^2 - N^2\right)^{1/2}} \sin\left[t_f \beta \frac{\left(n_f^2 - N^2\right)^{1/2}}{N} \right] \right\}^2. \tag{I13b}$$

By substituting these values into Eq. (I9), we get for the damping constant due to surface roughness

$$\alpha_{\text{sca,r}} = \frac{1}{2} \frac{\left(n_f^2 - N^2\right)}{t_{\text{eff}}\left(n_f^2 - n_c^2\right)} \frac{k_0^3}{4\pi n_f}$$
$$\times \left[\left(n_f^2 - n_c^2\right)^2 + J^2\left(n_f^2 - n_s^2\right)^2 \right]$$
$$\times \left[\int_0^\pi \int_{-\infty}^\infty R(u)e^{i(\beta - n_c k_0 \cos\theta)u} du d\theta \right.$$
$$\left. + \int_0^\pi \int_{-\infty}^\infty R(u)e^{i(\beta - n_s k_0 \cos\theta)u} du d\theta \right], \tag{I14}$$

where we define J as the asymmetry parameter of the waveguide:

$$J = \cos\left[t_f \beta \frac{\left(n_f^2 - N^2\right)^{1/2}}{N} \right]$$
$$+ \frac{\left(N^2 - n_c^2\right)^{1/2}}{\left(n_f^2 - N^2\right)^{1/2}} \sin\left[t_f \beta \frac{\left(n_f^2 - N^2\right)^{1/2}}{N} \right]. \tag{I15}$$

It is important to note that attenuation due to scattering into waveguide modes is not accounted for by this equation. The integral is solved in Ref. [52] with the assumption of an exponential form of the roughness correlation function $R(u) = \sigma^2 \exp\left(-|u|/L_c\right)$. This can be used to determine the attenuation coefficient of the sidewall scattering $\alpha_{\text{sca,}r}$. For completeness, this is done at the end of this section. However, in order to determine the background intensity, we are interested in finding an expression for the scattering leakage $a_{\text{ani,}r}\alpha_{\text{sca,}r}$. The anisotropic scattering is illustrated in Fig. 6 in the manuscript.

Since we are interested only in the radiation into the substrate region, only the second integral needs to be considered in Eq. (I14). The boundaries of the integral over the angle θ need to be adapted, since only scattered rays captured by the NA matter. The scattering leakage contribution that radiates into the NA of the objective reads

$$\alpha_{\text{NA,real}} = \frac{1}{2} \frac{\left(n_f^2 - N^2\right)}{t_{\text{eff}}\left(n_f^2 - n_c^2\right)} \frac{k_0^3}{4\pi n_f}$$
$$\times \left[\left(n_f^2 - n_c^2\right)^2 + J^2\left(n_f^2 - n_s^2\right)^2 \right]$$
$$\times \int_{\pi/2-\text{NA}/n_s}^{\pi/2+\text{NA}/n_s} \int_{-\infty}^\infty R(u)e^{i(\beta - n_s k_0 \cos\theta)u} du d\theta. \tag{I16}$$

We have approximated the integral boundaries by $\arcsin(\pi/2 \pm \text{NA}/n_s) \approx \pm\text{NA}/n_s$. The anisotropy parameter is then simply this quantity divided by the damping constant weighted by the angular contribution of the NA. (This weighting factor is $(1/2\pi)(2\text{NA}/n_s)$ since the damping constant $\alpha_{\text{sca,}r}$ is obtained by integration over 2π in Eq. (I14).) In this normalization, we did not account for the different relative contribution of the angles within the NA (the center of the mologram with $\theta = \pi/2$ gives a larger contribution than the edges):

$$a_{\text{ani}} = \frac{\alpha_{\text{NA,real}}}{\alpha_{\text{NA,iso}}} = \frac{\alpha_{\text{NA,real}}}{\alpha_{\text{sca,}r} \frac{1}{2\pi} \frac{2\text{NA}}{n_s}}. \tag{I17}$$

As in Ref. [52], we assume an exponential form of the autocorrelation function. Thus, evaluating the integral as

described in Ref. [52],

$$
\int_{\pi/2-\mathrm{NA}/n_s}^{\pi/2+\mathrm{NA}/n_s} \int_{-\infty}^{\infty} R(u) e^{i(\beta - n_s k_0 \cos\theta)u} \, du \, d\theta
$$

$$
= 2\sigma^2 L_c \int_{\pi/2-\mathrm{NA}/n_s}^{\pi/2+\mathrm{NA}/n_s} \frac{1}{1 + [(\beta - k n_s \cos\theta) L_c]^2} \, d\theta.
$$

(I18)

The scattering leakage then reads

$$
a_{\mathrm{ani,r}} \alpha_{\mathrm{sca,r}} = \frac{\pi n_s}{\mathrm{NA}} \frac{1}{2} \frac{\left(n_f^2 - N^2\right)}{t_{\mathrm{eff}}\left(n_f^2 - n_c^2\right)} \frac{k_0^3}{4\pi n_f}
$$

$$
\times \left[\left(n_f^2 - n_c^2\right)^2 + J^2 \left(n_f^2 - n_s^2\right)^2 \right] 2\sigma^2 L_c
$$

$$
\times \int_{\pi/2-\mathrm{NA}/n_s}^{\pi/2+\mathrm{NA}/n_s} \frac{1}{1 + [(\beta - k n_s \cos\theta) L_c]^2} \, d\theta.
$$

(I19)

This expression can be rewritten such that

$$
a_{\mathrm{ani,r}} \alpha_{\mathrm{sca,r}} = \frac{\pi n_s}{\mathrm{NA}} \frac{\pi \left(n_f^2 - N^2\right)}{t_{\mathrm{eff}}\left(n_f^2 - n_c^2\right)} \frac{\sigma^2}{\lambda^2 N n_f}
$$

$$
\times \left[\left(n_f^2 - n_c^2\right)^2 + J^2 \left(n_f^2 - n_s^2\right)^2 \right] L_c \beta
$$

$$
\times \int_{\frac{\pi}{2}-\frac{\mathrm{NA}}{n_s}}^{\frac{\pi}{2}+\frac{\mathrm{NA}}{n_s}} \frac{1}{1 + [(\beta - k n_s \cos\theta) L_c]^2} \, d\theta.
$$

(I20)

The integral can be solved analytically. However, it yields an elaborate expression from which no practical information is gained. The integration is not performed here and the integral is evaluated numerically.

For the sake of completeness, we also derive the attenuation coefficient for the sidewall roughness scattering as mentioned above. As described in Ref. [52], the integrals in Eq. (I14) can be solved analytically:

$$
\int_0^{\pi} \int_{-\infty}^{\infty} R(u) e^{i(\beta - n_i k_0 \cos\theta)u} \, du \, d\theta = 2\sigma^2 L_c \int_0^{\pi} \frac{1}{1 + [(\beta - k n_i \cos\theta) L_c]^2} \, d\theta
$$

$$
= \sqrt{2} \pi \sigma^2 L_c \frac{\left(\left\{ 4\beta^2 L_c^2 + \left[1 - L_c^2 \left(\beta^2 - n_i^2 k_0^2\right)\right]^2 \right\}^{1/2} + 1 - L_c^2 \left(\beta^2 - n_i^2 k_0^2\right) \right)^{1/2}}{\left\{ 4\beta^2 L_c^2 + \left[1 - L_c^2 \left(\beta^2 - n_i^2 k_0^2\right)\right]^2 \right\}^{1/2}}
$$

(I21)

where i stands for the substrate s or the cover c, respectively. In order to get a dimensionless term, we expand with β/β and get

$$
\int_0^{\pi} \int_{-\infty}^{\infty} R(u) e^{i(\beta - n_i k_0 \cos\theta)u} \, du \, d\theta = \frac{\sqrt{2}\pi\sigma^2}{\beta} F_i(L_c, \beta),
$$

(I22)

where F_i is dimensionless and is expressed as

$$
F_i(L_c, \beta) = L_c \beta \frac{\left(\left\{ 4\beta^2 L_c^2 + \left[1 - L_c^2 \left(\beta^2 - n_i^2 k_0^2\right)\right]^2 \right\}^{1/2} + 1 - L_c^2 \left(\beta^2 - n_i^2 k_0^2\right) \right)^{1/2}}{\left\{ 4\beta^2 L_c^2 + \left[1 - L_c^2 \left(\beta^2 - n_i^2 k_0^2\right)\right]^2 \right\}^{1/2}}.
$$

(I23)

Thus, the attenuation constant yields

$$\alpha_{\text{sca},r} = \frac{\left(n_f^2 - N^2\right)}{t_{\text{eff}}\left(n_f^2 - n_c^2\right)} \frac{k_0^3}{4\pi n_f} \frac{1}{2} \frac{\sqrt{2}\pi\sigma^2}{\beta}$$

$$\times \left[\left(n_f^2 - n_c^2\right)^2 + J^2\left(n_f^2 - n_s^2\right)^2\right]$$

$$\times \left[F_c\left(L_c, \beta\right) + F_s\left(L_c, \beta\right)\right], \tag{I24}$$

which can be written as

$$\alpha_{\text{sca},r} = \frac{\sqrt{2}}{2} \pi^2 \frac{\sigma^2}{\lambda^2} \frac{\left(n_f^2 - N^2\right)}{N t_{\text{eff}}\left(n_f^2 - n_c^2\right) n_f}$$

$$\times \left[\left(n_f^2 - n_c^2\right)^2 + J^2\left(n_f^2 - n_s^2\right)^2\right]$$

$$\times \left[F_c\left(L_c, \beta\right) + F_s\left(L_c, \beta\right)\right]. \tag{I25}$$

APPENDIX J: WAVEGUIDE AND BACKGROUND CHARACTERIZATION

In this section, we first show the validation of the model for the scattering leakage due to waveguide surface scattering and the necessary characterization to assess the suitability of a waveguide for molography. The goal is to determine the total scattering leakage and, in addition, whether it is dominated by surface or volume scattering.

1. Validation of the model for scattering leakage due to waveguide surface scattering

To validate the model for the scattering leakage, the statistical properties of the surface L_c and σ must be acquired by a suitable technique—i.e., AFM. The AFM measurement results in an rms roughness of 0.6 nm. It is not trivial to determine the correlation length from the AFM measurement since there are several exponential decays superimposed on each other. This results in a large uncertainty for the model of the surface scattering leakage [Eq. (I20)]. The AFM data suggest a range between 100 and 900 nm for the correlation length, and the best fit yields a correlation length of 588 nm. Using Eq. (I20), this leads to a scattering leakage of 6.65/m, whereas 3.12/m has been measured experimentally (Figs. 18 and 19). A corresponding background intensity of 5.3 mW/m² for a P_{WG} of 0.02 W/m and a NA of 0.4 is calculated with the help of Eq. (5). Experimentally, 2.7 mW/m² is measured for an average waveguide (see next sections). When addressing $\alpha_{\text{sca},r}$ and $a_{\text{ani},r}$ separately, the model predicts for $\alpha_{\text{sca},r}$ 36.2/m (1.57 dB/cm), whereas 57.8/m (2.51 dB/cm) is measured (Fig. 18). For $a_{\text{ani},r}$, the model yields 0.22 compared to 0.054 being measured experimentally (Fig. 19). This divergence of absolute values between model and
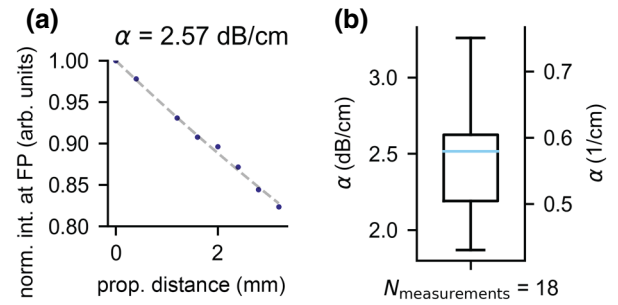
FIG. 18. Waveguide damping constant characterization. (a) Typical exponential decay of the scattered light of a waveguide mode. (b) Spread of the acquired damping constants. The median is 2.51 dB/cm or 0.58 1/cm.

measurements could be due to the following issues of the model: the uncertainty in the correlation length, neglecting guided to guided mode scattering, the model being two-dimensional, and the assumption of uncorrelated sidewalls. It is also possible that the model is sufficiently precise and
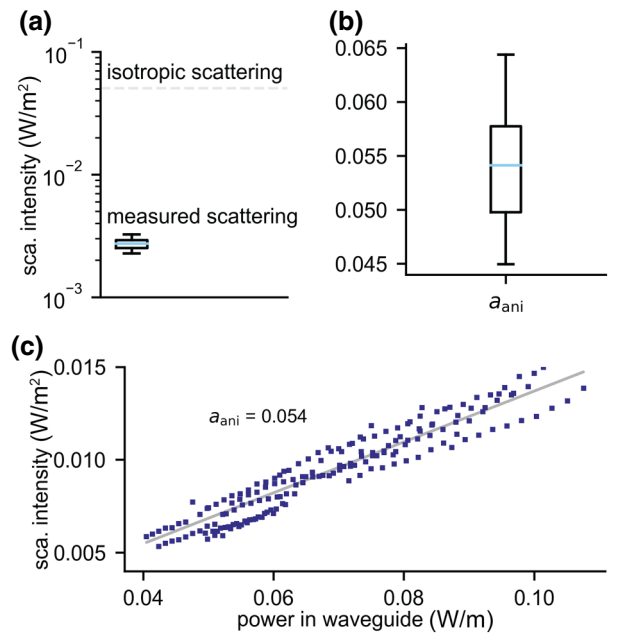
FIG. 19. Anisotropic scattering coefficient a_{ani}. (a) Measured scattered intensity with a 0.4 NA objective 100 μm away from the surface on the substrate side compared to isotropic scattering. The intensity is independent of the distance. Individual damping constants are normalized to a damping of 2.51 dB/cm (Fig. 18). Power in the waveguide is normalized to a value of 0.02 W/m. Box plots are based on 180 measurements of image size 280 × 210 μm² with 110-nm pixel size of three different chips. (b) a_{ani} for the investigated waveguide configuration. The median of the coefficient amounts to 0.054. (c) Intensity measured as a function of the power in the waveguide. The gray line is a least squares fit through the data and, as expected, yields the same value for the anisotropic scattering coefficient.

a fraction (30–40%) of the damping is caused by absorption. Also, some inaccuracies can arise from the limited precision of the measurement of the power carried by the waveguide mode. Nevertheless, the model can be used to provide a first estimate for the scattering leakage from an AFM measurement.

2. Relative importance of surface to volume scattering

The relative importance of volume-to-surface scattering can be investigated by measuring the ratio of the scattered intensity with two different cover media, since volume scattering will be hardly affected by the change in cover medium. Furthermore, the model for the surface roughness can be used to estimate the reduction of intensity from the surface scattering. We chose air and index matching oil (Immersol 518F, Zeiss) as the cover media. For the difference in scattered field intensity between oil and air, the model for the surface roughness predicts a reduction of a factor of 1.6 of the scattered intensity due to the lower index contrast at the cover interface. Experimentally, the intensity drops by a factor of 2.3. The higher experimental value can be explained by the change from an asymmetric to a symmetric waveguide upon oil immersion (roughly a reduction in factor 1.5 for a dipole). A stronger scattering into the optical denser half space in the asymmetric case is not considered in the model since the signal is affected in the same way [30]. Therefore, the reduction of a factor 2.3 is a strong indication that surface scattering is the dominant background source because volume scattering is hardly affected by the change in cover medium. If volume scattering were dominant, one would expect a much lower reduction of the background intensity only caused by the symmetry change of the waveguide (roughly 1.5).

3. Damping constant characterization

The damping constants of coating stripped (30 s O_2 Plasma, 10 min US in DI water) ZeptoMark chips are measured on the MoloReader. Prior to measurement, stage movement and mode propagation direction are aligned. Stray light images are acquired 100 μm below the chip surface and the mean of each image is used for exponential fitting. Per damping trace, six images spaced by 400 μm are acquired. The damping constants are measured on three chips at six different locations each. The median damping constant amounts to 2.51 dB/cm (Fig. 18). Other attenuation constants stated in the manuscript are measured by this method for the corresponding chips.

4. Experimental determination of scattering leakage and anisotropy of scattering

The scattering leakage can be determined by simultaneously measuring the power in the waveguide by the MoloReader and the intensity at a sufficient distance and using Eq. (5). We perform this characterization for the

Ta_2O_5 waveguide in air. The anisotropy coefficient can be determined by dividing the scattering leakage with the attenuation constant. Fig. 19(a) shows the measured intensity normalized to a power of 0.02 W/m in the waveguide compared to what is expected if the scattered power would be distributed uniformly over the solid angle. Fig. 19(b) shows the box plot of the anisotropic scattering coefficient from 180 measurements and Fig. 19(c) the non-normalized intensity values plotted against the corresponding power in the waveguide.

5. AFM characterization of surface roughness

AFM is used to determine the surface roughness [Fig. 20(a)]. The chip is mounted on the micrometer positioning stage of a Dimension Icon AFM (Bruker, Santa Barbara). The microscope is operated with Bruker Nanoscope V 9.1. Height images are recorded under tapping mode operation using etched silicon cantilevers with a nominal spring constant of $k = 20$–80 N/m (Bruker AFM Probes, RTESPA). An image area of 20×20 μm^2 is scanned with a resolution of 5210×256 pixels, and at limited z scanner range of <1 μm. The slow scanning axis [5210] is oriented perpendicular to the chip main axis. Height raw data are tilt-corrected followed by plane correction using a second-order polynomial fit (Nanoscope Analysis V1.8 R2). A rms roughness of 0.6 nm is measured. The one-dimensional autocorrelation function is calculated from the profile in the x direction on each line. The average autocorrelation function of all lines is taken and normalized. Due to the convolution with the tip roughness, the autocorrelation function exhibits a dominant exponential decay at extremely short length scales of 10 nm [Fig. 20(b)]. Since this roughness length scale is attributed to the tip and not the surface, the autocorrelation function is plotted from 50 to 3000 nm [Fig. 20(c)]. Several superimposing exponential decays at different length scales are visible. Therefore, using the power spectral density function (PSD) instead of fitting an exponential decay to the autocorrelation function would be more appropriate. However, this leads to a less intuitive model. The exponential decay with the shortest correlation length of Fig. 20(c) is extrapolated to 0 [Fig. 20(d)]. Then the extrapolated first 50 nm are concatenated with the autocorrelation from 50 to 20 000 nm. The normalized autocorrelation function and an exponentially decaying fit are shown in Fig. 20(e). The fit results in a correlation length of 588 nm.

6. Influence of waveguide parameters on signal and background

Waveguide parameters such as the refractive indices n_f, n_c, n_s, the thickness t_f, or the wavelength have a significant influence on signal and background intensities. Therefore, we explain some dependencies in this section. The intensity of the signal scales with the intensity of
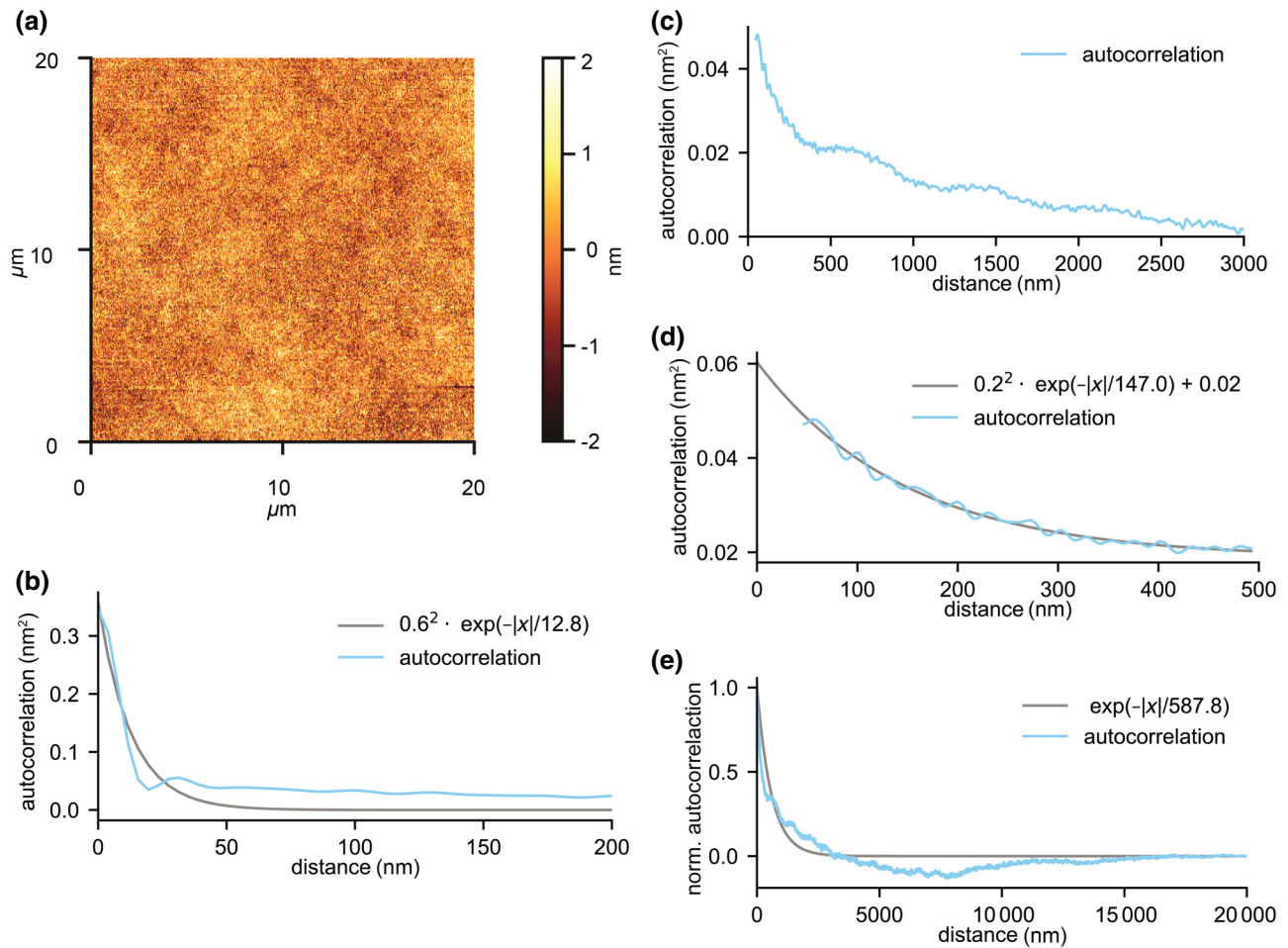
FIG. 20. Waveguide surface roughness characterization. (a) AFM image of the waveguide surface. (b) The average autocorrelation function of the surface roughness in the x direction normalized to its length is shown. An exponential fit suggests a correlation length of 12.8 nm. Yet, this is an artifact due to the convolution of the roughness with the AFM tip. (c) The autocorrelation function is, therefore, plotted from 50 to 3000 nm without the contribution from the tip. Several superimposed exponential decays are visible. (d) To obtain the missing data points that are lost due to the convolution with the tip, an extrapolation for the first 50 nm is performed with the exponential fit to the first decay (50–500 nm). (e) The extrapolation is concatenated with the normalized autocorrelation function from 50 to 20 000 nm and an exponential fit is performed to extract the correlation length of the surface roughness of 588 nm.

the electromagnetic field at the film-cover interface [see Fig. 21(a)]. This is closely related to the sensitivity of the waveguide to adlayer changes as described in Ref. [73]. However, while the sensitivity to adlayer changes accounts for the entire evanescent field, the molographic signal (in the configuration of this manuscript) stems from the surface only. It should be pointed out that scattering at non-specifically adsorbed molecules scales in the same fashion with respect to the waveguide parameters as the signal. The reason for this is that the scatterers have the same index contrast and that they are also located on the waveguide surface. Scattering at the surface roughness has a similar behavior with two exceptions [see Fig. 21(b)]. First of all, both interfaces need to be considered and it is affected by the optical contrast at the interfaces. Hence, a lower film refractive index increases the signal to surface

scattering ratio, although the signal itself decreases for a lower refractive index [see Fig. 21(d)]. In order to make a fair comparison between waveguides of different refractive indices, the thickness of the waveguide is adapted such that the ratio of the intensity at the respective thickness to the maximum intensity for the chosen refractive index is kept constant [see Fig. 21(a): The ratio of the maximum of the dark blue curve to the sensitivity at the intersection of the dashed green line (ii) and the dark blue curve is kept constant for any refractive index]. When calculating the FOM with a theoretical estimation of the background the absolute values must be seen as approximations. However, the formula can identify the dependency of the FOM with respect to different parameters. Yet, it must be noted that Fig. 21(d) neglects any volume scattering, which might become dominant for thicker waveguides.
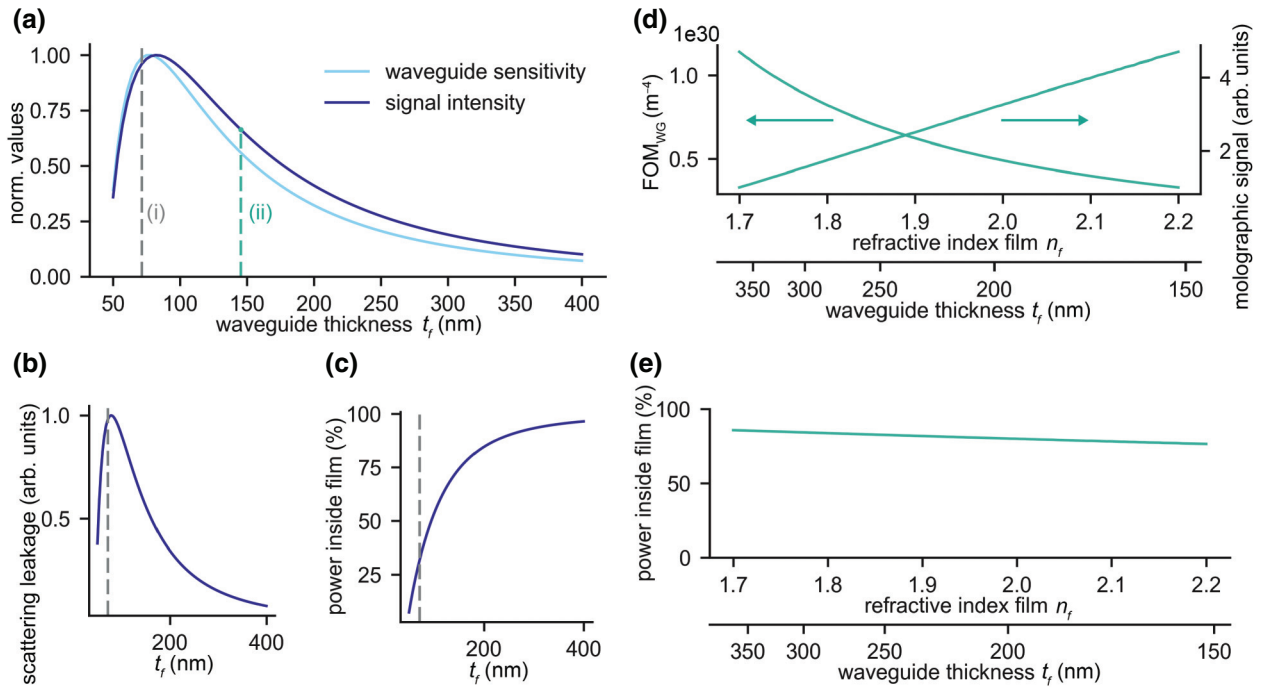
FIG. 21. (a) Normalized signal intensity (which corresponds to the intensity on the surface) and waveguide sensitivity to adlayer changes as a function of waveguide thickness for a Ta_2O_5 waveguide [73]. The dashed gray line (i) indicates the cut-off thickness. The dashed green line (ii) indicates the thickness for the waveguides used in this paper. (b) Surface scattering leakage against waveguide thickness. (c) Power inside the waveguide film against waveguide thickness. (d) Calculated figure of merit (neglecting any volume scattering) against the waveguide index (for $L_c = 588$ nm, $\sigma = 0.6$ nm and NA = 0.4). The waveguide thickness is adapted with respect to the refractive index to enable a fair comparison between different indices. The ratio of the signal intensity at the respective waveguide thickness to the maximal possible intensity for the chosen respective refractive index is kept constant. On the second y axis, the dependency of the molographic signal to the refractive index is illustrated. It should be noted that using the formula for the scattering results in an approximation of the FOM_{WG} and is, therefore, lower than the experimentally determined one. (e) Power inside the waveguide against the refractive index of the film. The thickness is adapted as described in (c). The power inside the waveguide slightly increases for the thicker lower refractive index films. For the calculations of the curves, the following waveguide parameters are used: $n_c = 1$, $n_s = 1.521$, $\lambda = 632.8$ nm for (a)–(c) $n_f = 2.117$.

Conversely, the scattering at inhomogeneities inside the waveguide depends on the power inside the waveguide. It scales in the opposite fashion with respect to the film thickness as the intensity on the surface [see Fig. 21(c)]. The power inside the waveguide film is basically independent of the refractive index of the waveguide, when the thickness is adapted in the same way as described above [see Fig. 21(e)]. However, this cannot be directly transferred to the volume scattering, since the refractive index inhomogeneities inside the volume depend on the refractive index of the film.

7. Noise estimation and speckle statistics

a. Estimation of I_{sig}/I_{BG} for endpoint measurements

In an endpoint measurement, the variance of the ratio of the maximum pixel to the mean of different images is the relevant quantity. Fig. 22 shows the variation of the mean and the maximum pixel of the image for 180 images, as well as the variation of their ratio. The maximum pixel is 8.1 (median is with 95% confidence between 7.9 and 8.3) times higher than the measured mean of the background intensity due to waveguide surface roughness scattering. The 99.5% percentile of the ratio of maximum/mean background that allows the estimation of the detection limit in an endpoint measurement with the specified field of view is 13.8. If the position of the molographic spot can be constrained to a smaller field of view, this ratio decreases. In the limit when the position of the molographic spot is precisely known, it reduces to the expected variation of a single speckle determined by speckle statistics [46]. Figure 23 shows the histogram of the speckle intensities before (a) and after a convolution operation with the Airy disk of the mologram (b). Before convolution, the distribution of the measured intensities follows a nearly perfect negative exponential distribution [46] and the ratio of the 99.7th percentile and the mean is close to the expected value from the negative exponential distribution $q_{99.7}/\mu = -\ln(0.003) = 5.8$. After convolution, the
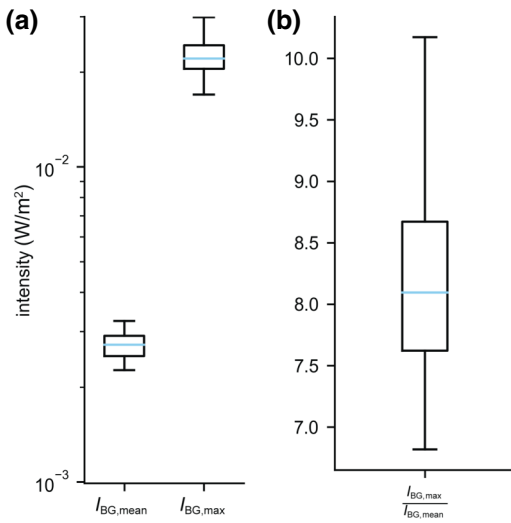
FIG. 22.   Estimation of $\frac{I_{\text{sig}}}{I_{\text{BG}}}$ for endpoint measurements. (a) Mean and maximum of the background intensity in the focal plane after convoluting the images with an Airy function of the numerical aperture of the objective (NA = 0.4), which is the equivalent of trying to detect a molographic spot from a mologram with NA = 0.4. (b) Ratio of maximum background intensity (maximum pixel) and mean background in the focal plane. Box plots are based on 180 measurements of image size $280 \times 210 \ \mu\text{m}^2$ with 110-nm pixel size of three different chips and are convoluted with the expected Airy disk of the mologram.

speckle statistics deviates from the exponential distribution, rendered more Gaussian [46], and the median of the ratio of the 99.7th percentile and mean is 3.33. Therefore,

the precise knowledge of the molographic spot would only give a slightly better detection limit of 1.5 pg/mm^2 in endpoint measurements.

b. Influence of large background scatterers

a. In terms of total power. To investigate the importance of large background scatterers, we acquire 20 images of the surface of one cleaned chip (30 s O$_2$ plasma followed by 10 min sonication in DI water). Images are thresholded (15 times the mean of the image) and everything above this is taken as power that originates from large scatterers. Figure 24 shows that below 1% of the total power leaving the waveguide is due to large scattering centers. Therefore, in terms of total power, this background contribution is negligible for properly cleaned chips.

b. As a function of distance. Despite the overall power scattered being small, a large particle on the waveguide surface will cause an inhomogeneous background that may interfere with the molographic measurement. Molographic foci should, therefore, be sufficiently distant to the waveguide surface. To obtain a rough estimate of this distance, we acquire images of the scattered field at various positions from the waveguide at a location where there is substantial particle contamination and calculate the value of the maximum pixel and the mean of the image (Fig. 25). One can see that the influence becomes negligible beyond 80–100 μm from the waveguide surface for a 0.4 NA objective and a scattering leakage of roughly 6 /m. The exact distance will be a function of the scatterer size and
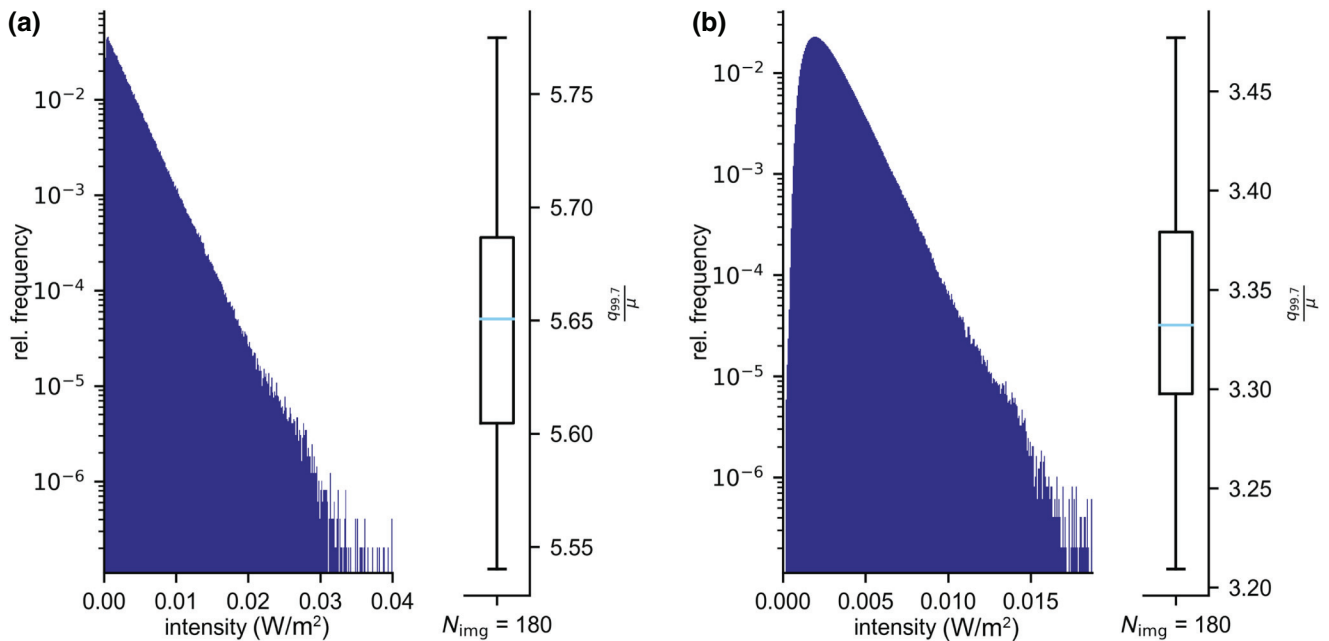

FIG. 23.   Distribution of speckle intensities and distribution of the 99.7th percentile divided by the mean. (a) Before convolution with the Airy disk and (b) after convolution with the Airy disk of the mologram (NA = 0.4). The speckle intensities have been normalized to a power per unit length in the waveguide of 0.02 W/m.
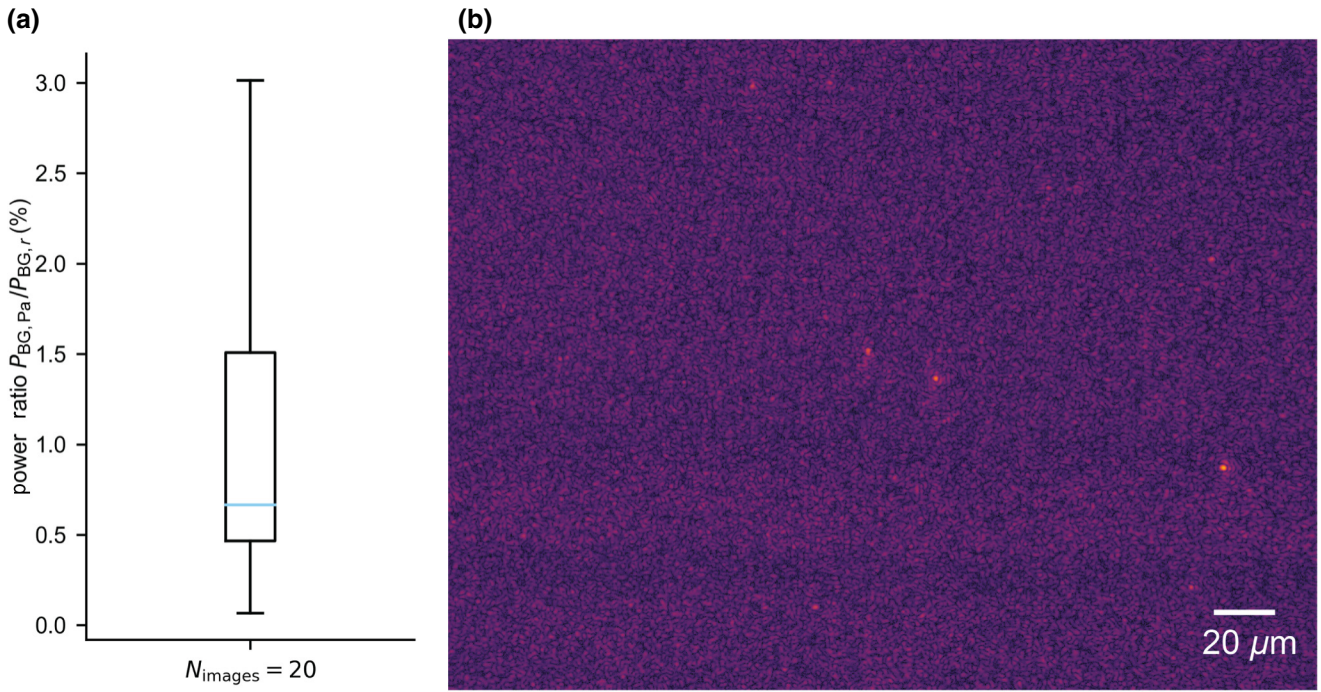
FIG. 24. Relative importance of the power scattered by large dust particles relative to power scattered from waveguide sidewall roughness (a) box plot of the power ratio $P_{BG,r}/P_{BG,Pa}$, median is 0.66% with 95% confidence in interval [0.47, 1.47]; (b) typical waveguide surface image with a few dust particles. The power norm chosen for plotting (inferno color map) is 0.25 in order to better visualize the waveguide background.

refractive index, the penetration depth of the evanescent field, the numerical aperture of the optics, and the intensity of the waveguide background. The precise derivation of this expression is beyond the scope of this publication, but as a rule of thumb, for a given scatterer size, the distance is inversely proportional to the NA of the mologram and the square root of the scattering leakage:

$$f \propto \frac{1}{\text{NA}} \frac{1}{\sqrt{a_{\text{ani}} \alpha_{\text{sca}}}}. \tag{J1}$$

To give an example, for a NA of 0.1 and a scattering leakage of 0.61/m, the focal plane should be roughly an order of magnitude further away than in our case (distance always in air).

APPENDIX K: RESOLUTION OF FOCAL MOLOGRAPHY AND COMPARISON TO SPR

1. Choice of the reference intensity for focal molography

It is not straightforward to choose the intensity reference in focal molography. One option would be to split the input beam and use this as the reference. Yet, this is not particularly useful since variations in the in-coupling efficiency would not be compensated. Also measuring the power in the waveguide via the out-coupling grating during a measurement will lead to drifts, since any change in scattering

over the entire propagation distance will affect the remaining power at the out-coupling grating. The most suitable intensity reference is, therefore, the waveguide background or a reference spot in the focal plane, since the majority of fluctuations will affect it in exactly the same way as the molographic focus. Here, we chose the waveguide background and provide the formulas for this reference.

2. Derivation of the limit of detection formula for focal molography

The mass density on the mologram can be computed from the average intensity in the Airy disk by starting from

$$I_{\text{avg,RS}} = 2.536\pi^2 \frac{A_{\text{mologram}}^2}{A_+^2} \text{NA}^2 n_c^2 \frac{(n_P^2 - n_c^2)^2}{(n_P^2 + 2n_c^2)^2}$$

$$\times \frac{D^2}{\lambda^4} \frac{\eta_{\text{mod}[A]}^2 \Gamma^2}{\rho_P^2} \frac{n_c (n_f^2 - N^2)}{N t_{\text{eff}} (n_f^2 - n_c^2)} P_{\text{WG}}. \tag{K1}$$

It is advantageous to use a different algorithm to compute the molographic signal and adapt this expression to the standard readout algorithm of the convolution with the normalized Airy disk kernel (see Appendix L3). The ratio of the convoluted intensity I_{sig} (evaluated at the center of the
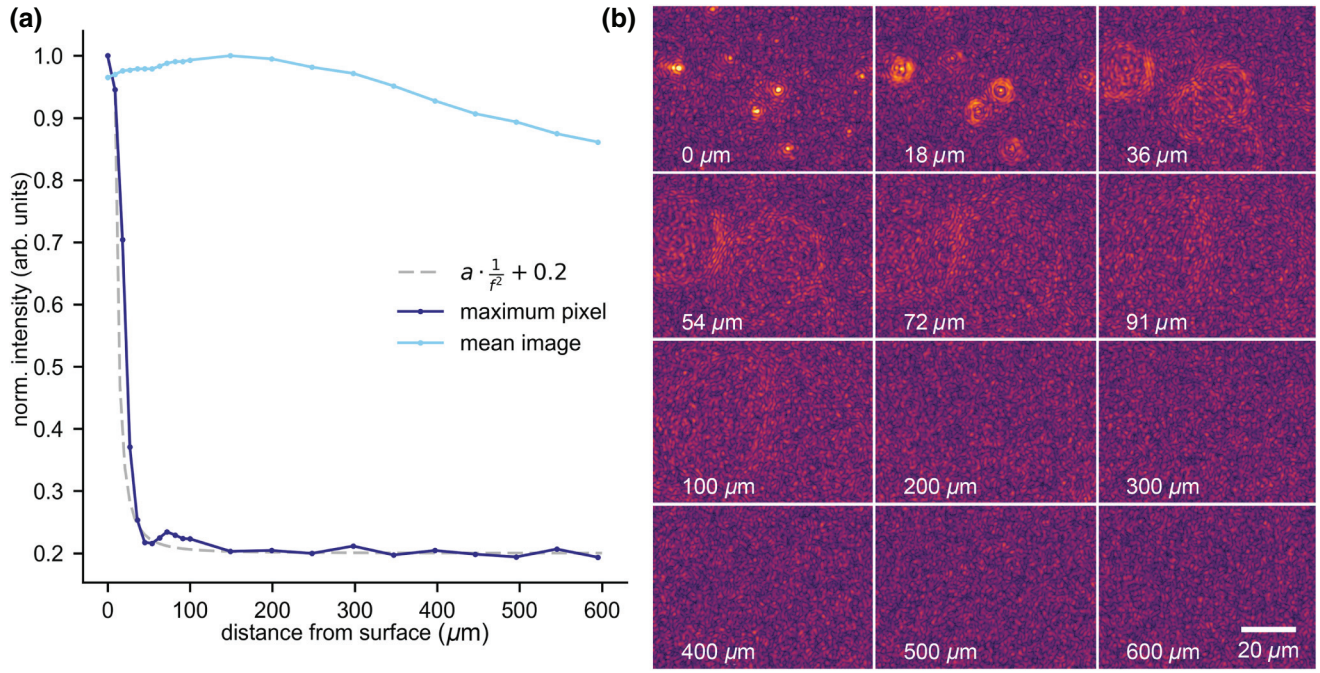
FIG. 25. Relative importance of large scatterer centers for the homogeneity of the speckle background. (a) Maximum pixel and mean of the image normalized to the value at the surface calculated from seven approach curves. Images are acquired by a 0.4 NA objective. The mean is nearly independent of distance as expected from the expression of the waveguide background intensity for a constant aperture. The decrease for larger distances is due to the finite beam width of 1 mm. The maximum pixel first decreases with a reciprocal quadratic relationship and then approaches the waveguide background (dashed line) as the function is normalized to the datapoint closest to the waveguide surface (9 μm). (b) Exemplary images highlighting the decrease of importance of large scatterers with increasing distance from the waveguide.

Airy disk $r = 0$) and the maximum intensity $I_{\text{Airy,max}}$ of the unconvoluted Airy disk reads:

$$\frac{I_{\text{sig}}}{I_{\text{Airy,max}}} = \left(\frac{2J_1(r)}{r}\right)^2 \left.\frac{\left(\frac{2J_1(r)}{r}\right)^2}{\int_0^{2\pi}\int_0^{\infty}\left(\frac{2J_1(r)}{r}\right)^2 r\,dr}\right|_{r=0}$$

$$= 0.4596. \tag{K2}$$

Hence, from Eqs. (E8) and (K2), it follows that $I_{\text{sig}} = 2.012 I_{\text{avg}}$. Inserting this into Eq. (K1) and rearranging gives

$$\Gamma = \frac{A_+}{A_{\text{mologram}}} \frac{1}{\eta_{\text{mod}[A]}} \frac{\lambda^2}{D} \frac{\rho_P \left(n_P^2 + 2n_c^2\right)}{n_c \left(n_P^2 - n_c^2\right)}$$

$$\times \sqrt{\frac{I_{\text{sig}}}{5.102\pi^2 \text{NA}^2 \frac{n_c\left(n_f^2 - N^2\right)}{Nt_{\text{eff}}\left(n_f^2 - n_c^2\right)} P_{\text{WG}}}}. \tag{K3}$$

The power in the waveguide can then be expressed from the average intensity of the background according to

Eq. (5), $P_{\text{wg}} = \frac{4I_{\text{bg}}}{\left(a_{\text{ani}}\alpha_{\text{sca}}\text{NA}^2\right)}$, which yields

$$\Gamma = 0.0704 \frac{A_+}{A_{\text{mologram}}} \frac{1}{\eta_{\text{mod}[A]}} \sqrt{\frac{a_{\text{ani}}\alpha_{\text{sca}}}{\frac{n_c\left(n_f^2 - N^2\right)}{Nt_{\text{eff}}\left(n_f^2 - n_c^2\right)}}}$$

$$\times \frac{\lambda^2}{D} \frac{\rho_P \left(n_P^2 + 2n_c^2\right)}{n_c \left(n_P^2 - n_c^2\right)} \sqrt{\frac{I_{\text{sig}}}{I_{\text{BG}}}}. \tag{K4}$$

This can also be written in terms of the refractive index increment for proteins:

$$\Gamma = 0.1056 \frac{A_+}{A_{\text{mologram}}} \frac{1}{\eta_{\text{mod}[A]}} \frac{1}{\frac{dn}{dc}} \sqrt{\frac{a_{\text{ani}}\alpha_{\text{sca}}}{\frac{n_c\left(n_f^2 - N^2\right)}{Nt_{\text{eff}}\left(n_f^2 - n_c^2\right)}}} \frac{\lambda^2}{D} \sqrt{\frac{I_{\text{sig}}}{I_{\text{BG}}}}, \tag{K5}$$

from which Γ_0 can be identified:

$$\Gamma_0 = 0.1056 \frac{A_+}{A_{\text{mologram}}} \frac{1}{\eta_{\text{mod}[A]}} \sqrt{\frac{a_{\text{ani}}\alpha_{\text{sca}}}{\frac{n_c\left(n_f^2 - N^2\right)}{Nt_{\text{eff}}\left(n_f^2 - n_c^2\right)}}} \frac{\lambda^2}{D\frac{dn}{dc}}. \tag{K6}$$

3. Derivation of the sensitivity of SPR

The aim of this section is to derive the sensitivity of SPR, which links the rms noise of the relative intensity (reflectance changes) to the rms noise of the adlayer mass density changes. It has to be noted that the reflectivity is normalized to the input beam power. However, since the reflectivity in intensity interrogation mode (at the location of maximum slope) is usually $R = 0.5$ [62], therefore, this equation is still a good estimate of the required precision of the intensity measurement at the detector:

$$\mathrm{rms}_{R-\langle R \rangle} = S_\Gamma \mathrm{rms}_{\Gamma-\langle \Gamma \rangle} \tag{K7}$$

or in terms of effective refractive index

$$\mathrm{rms}_{R-\langle R \rangle} = S_N \mathrm{rms}_{N-\langle N \rangle}. \tag{K8}$$

The sensitivities are defined by

$$S_N = \frac{\partial R}{\partial N}, \quad S_\Gamma = \frac{\partial R}{\partial \Gamma}. \tag{K9}$$

$\mathrm{rms}_{R-\langle R \rangle}$, $\mathrm{rms}_{N-\langle N \rangle}$ and $\mathrm{rms}_{\Gamma-\langle \Gamma \rangle}$ are rms noise of the reflectivity, the effective refractive index, and surface mass density, respectively.

We first aim to find an expression for the effective refractive index sensitivity S_N. From [63], we know that the maximum change in reflectivity of the SPR signal is given by

$$\left(\frac{dR}{dk_x} \right)_{\mathrm{max}} = 3\frac{\sqrt{3}}{2}\frac{\gamma_i \gamma_r}{(\gamma_i + \gamma_r)^3}, \tag{K10}$$

where k_x is the in-plane wavevector. γ_i is the intrinsic damping of the plasmon and γ_r is the radiation damping. The sensitivity is maximized when $\gamma_r = \frac{\gamma_i}{2}$. This is different from the perfectly matched case $\gamma_r = \gamma_i$ when the reflectivity is zero, which can be explained in that
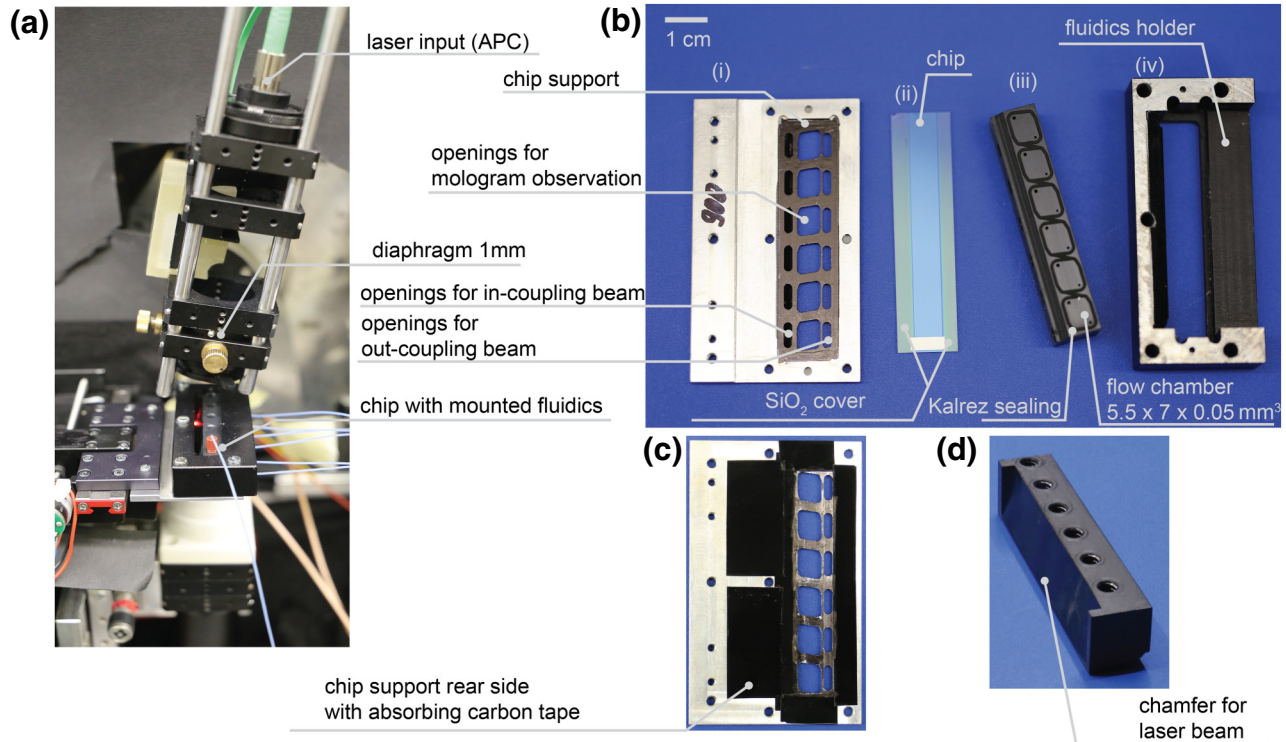

FIG. 26.   MoloReader configuration for real-time measurements. (a) Chip with mounted fluidics, connectors used are F-126-H (Ercat-ech AG) for 1/32 tubing (FEP009-031-B 225 μm internal diameter). (b) The chip fluidics assembly disassembled. From left to right: An aluminium chip holder (i) is machined with openings for in-coupling and out-coupling beams as well as for observation with the microscope. It is important that the sealing is supported everywhere with aluminium, otherwise the chip breaks upon tightening the screws. The rear side of the holder is shown in (c). Straylight is reduced at all important locations by means of a black waterproof pen or MetalVelvet[TM] Adhesive (ACM Coatings). The waveguide chips (ii) (IMT Masken und Teilungen AG, Greifensee, Switzerland) are described in the main text, for the real-time measurements the coupling gratings are covered with SiO_2 (approximately 1 μm) in order to avoid damping of the mode at the black Kalrez flat packing (laser cut from a sheet of Kalrez 0.5 mm thickness). Each flow chamber has a volume of roughly 2 μl. (iii) PEEK fluidics housing in bottom view [front view in (d)] six fluidic chambers are individually addressable. (iv) Aluminium holder for the PEEK fluidic body. (d) The PEEK body has a chamfer with a $-12°$ angle for the laser beam to reach the coupling grating.

although the reflectivity at resonance is not zero, the slope is maximized. Inserting this into Eq. (K10) yields

$$\frac{dR}{dk_{\mathrm{x}}} = \frac{2\sqrt{3}}{9\gamma_i}. \qquad (\mathrm{K}11)$$

The intrinsic damping of the film is

$$\gamma_i = \frac{n_c^3 k_0 \varepsilon_i}{2\varepsilon_r^2}, \qquad (\mathrm{K}12)$$
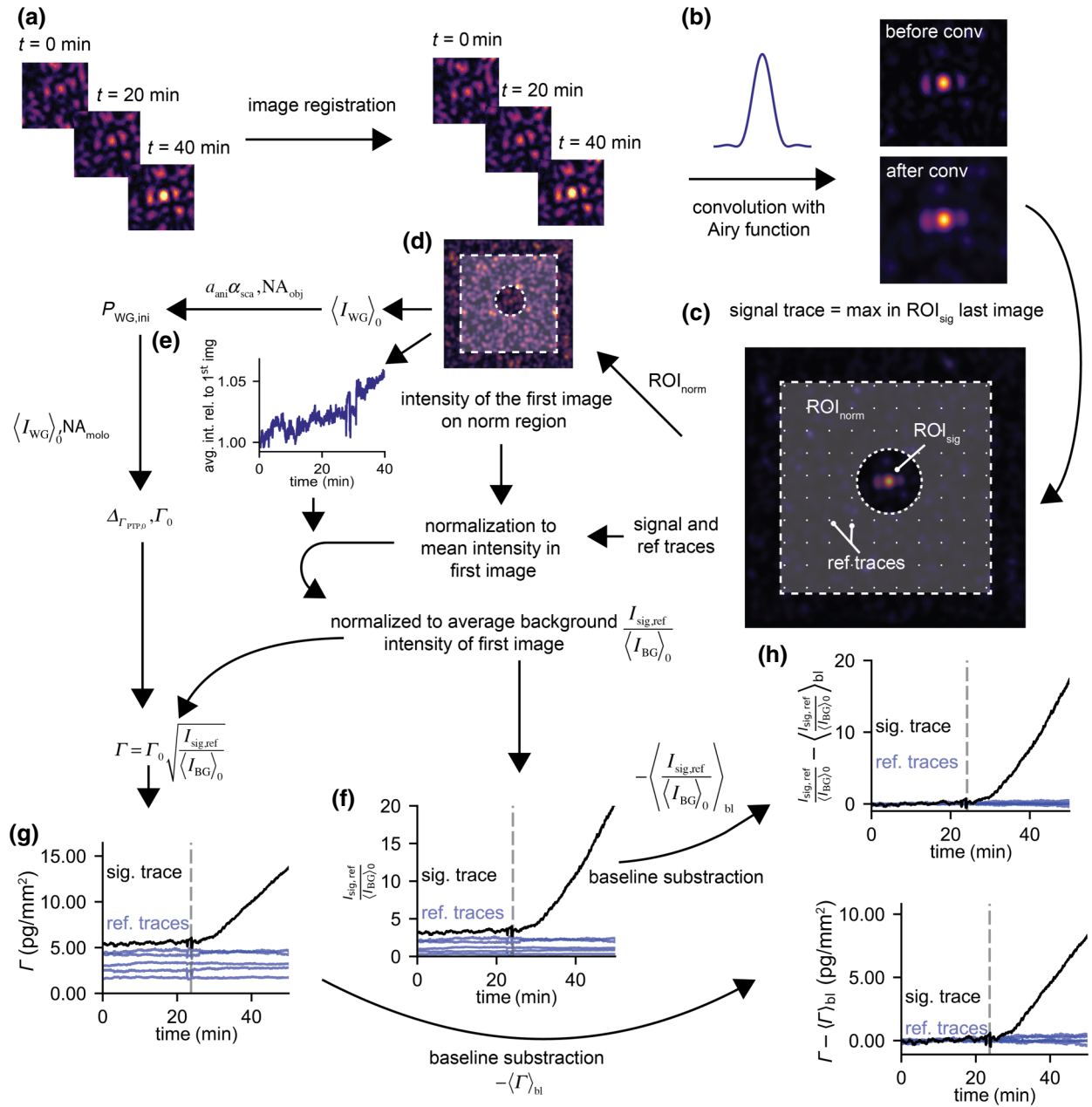

FIG. 27.　Processing of the real-time binding signals. (a) Raw images are registered with TurboReg [66] to compensate for the mostly temperature-induced lateral drift of the speckles in the focal plane. (b) Images are convoluted with the Airy disk of the mologram in order to decrease the influence of hot pixels and pixel noise. (c) The region of interest for normalization (ROI$_{\mathrm{norm}}$), signal detection (ROI$_{\mathrm{sig}}$) and the pixel chosen for reference binding traces are highlighted. The signal trace is determined from the location of the highest pixel in the last image of the binding experiment. A border of 50 pixels is discarded for the analysis due to shifts of the image registration algorithm. (d) Intensity on the ROI$_{\mathrm{norm}}$ of the first image is used to normalize the intensity of each image. (e) The intensity on the ROI$_{\mathrm{norm}}$ of every image is normalized to the first image in order to compensate for drifts of the laser source. (f) The normalized intensity of the signal and five reference traces as a function of time. (g) The molographic surface mass density that is equal to the measured intensity is calculated from the first image and by multiplying this value with the square root of the normalized intensity, the coherent mass density traces are obtained. (h) Baseline subtracted intensity and mass binding traces

whereas ε_r and ε_i are real and imaginary parts of the permittivity of the metal, respectively. By using $k_x = Nk_0$ and $\frac{dk_x}{dN} = k_0$, we can find for S_N,

$$S_N = \frac{dR}{dN} = \frac{dR}{dk_x}\frac{dk_x}{dN} \approx \frac{dR}{dk_x}k_0 = \frac{2\sqrt{3}}{9\gamma_i}k_0. \qquad \text{(K13)}$$

Inserting the expression for the damping constant (Chap. 4.1 in [10]) we obtain the sensitivity with respect to effective index changes:

$$S_N = \frac{4\sqrt{3}\varepsilon_r^2}{9n_c^3\varepsilon_i}. \qquad \text{(K14)}$$

This expression is maximized when the damping of the SPR film is minimized. The minimum damping of a gold film is reached at 700 nm with a permittivity of $\varepsilon = -16 + i1.1$ [60]. For gold at 700 nm, the sensitivity for refractive index changes is roughly 76. Hence, in order to resolve a refractive index change of $\Delta N = 10^{-7}$ as in the best SPR measurements reported [60], one needs to measure the intensity with a precision of roughly 10^{-5} or 0.001% reflectivity change.

Finally, we need to derive the sensitivity in terms of surface mass density. The sensitivity of the effective refractive index as a function of surface mass density was derived by Lukosz [44]:

$$\Delta N = \frac{4\pi N^4}{\lambda(-\varepsilon_r)^{\frac{1}{2}}}\frac{1}{n_c^2}\frac{dn}{dc}\Delta\Gamma. \qquad \text{(K15)}$$

With this expression and noting that in Eq. (K12) one should correctly use the effective index [64], we obtain the sensitivity equation of SPR with respect to surface mass changes for intensity interrogation:

$$S_\Gamma = \frac{\partial R}{\partial \Gamma} = \frac{16\sqrt{3}\varepsilon_r^2}{9\varepsilon_i(-\varepsilon_r)^{\frac{1}{2}}}\frac{\pi N}{\lambda n_c^2}\frac{dn}{dc}. \qquad \text{(K16)}$$

Inserting the values for gold at 700 nm and using the calculated effective index of 1.41, one obtains $S_\Gamma = 1.16 \times 10^5$ (mm^2/pg).

4. Derivation of the sensitivity of focal molography

The sensitivity of focal molography can be derived accordingly. The intensity reference in molography is not the input power that hits the grating coupler, but rather the intensity of the average background in the focal plane. We start from

$$\frac{I_{\text{sig}}}{I_{\text{BG}}} = 89.4\frac{A_{\text{mologram}}^2}{A_+^2}\left(\frac{dn}{dc}\right)^2\frac{D^2}{\lambda^4}$$

$$\times \eta_{\text{mod}[A]}^2\frac{n_c\left(n_f^2 - N^2\right)}{Nt_{\text{eff}}\left(n_f^2 - n_c^2\right)}\frac{1}{a_{\text{ani}}\alpha_{\text{sca}}}\Gamma^2, \qquad \text{(K17)}$$

where we have replaced the power in the waveguide P_{WG} by the average background intensity. This is a relative intensity in exactly the same way as the reflectivity in SPR. We take the derivative with respect to molographic surface mass density to obtain the sensitivity equation:

$$S_\Gamma = \frac{\partial\frac{I_{\text{sig}}}{I_{\text{BG}}}}{\partial\Gamma} = 178.8\frac{A_{\text{mologram}}^2}{A_+^2}\left(\frac{dn}{dc}\right)^2\frac{D^2}{\lambda^4}$$

$$\times \eta_{\text{mod}[A]}^2\frac{n_c\left(n_f^2 - N^2\right)}{Nt_{\text{eff}}\left(n_f^2 - n_c^2\right)}\frac{1}{a_{\text{ani}}\alpha_{\text{sca}}}\Gamma$$

$$= \frac{2}{\Gamma_0^2}\Gamma. \qquad \text{(K18)}$$

By assuming that the operation point is close to the background intensity ($\Gamma = \Gamma_0$), which is a valid approximation, the sensitivity of focal molography reads

$$S_{\Gamma_{\text{FM}}} = \frac{\partial\frac{I_{\text{sig}}}{I_{\text{BG}}}}{\partial\Gamma} = 18.9\frac{A_{\text{mologram}}\eta_{\text{mod}[A]}}{A_+}\frac{dn}{dc}$$

$$\times\frac{D}{\lambda^2}\sqrt{\frac{n_c\left(n_f^2 - N^2\right)}{Nt_{\text{eff}}\left(n_f^2 - n_c^2\right)a_{\text{ani}}\alpha_{\text{sca}}}}$$

$$= \frac{2}{\Gamma_0}\propto\sqrt{\text{FOM}_{\text{FM}}}. \qquad \text{(K19)}$$

Alternatively, the sensitivity can be derived from $\Gamma = \Gamma_0\sqrt{I_{\text{sig}}/I_{\text{BG}}}$. By writing $I_{\text{sig}} = I_{\text{BG}} + \Delta I$ and performing a Taylor expansion up to first order, one obtains

$$\Gamma - \Gamma_0 = \frac{\Gamma_0}{2}\frac{\Delta I}{I_{\text{BG}}}. \qquad \text{(K20)}$$

From the definition $I_{\text{sig}}/I_{\text{BG}} - \langle I_{\text{sig}}/I_{\text{BG}}\rangle = \Delta I/I_{\text{BG}}$ and again assuming that the operating point is close to the background intensity ($\Gamma_0 = \langle\Gamma\rangle$), it follows that the rms values of intensity and molographic surface mass density are connected accordingly:

$$\text{RMS}_{I_{\text{sig}}/I_{\text{BG}} - \langle I_{\text{sig}}/I_{\text{BG}}\rangle} = \frac{2}{\Gamma_0}\text{RMS}_{\Gamma - \langle\Gamma\rangle}. \qquad \text{(K21)}$$

The factor 2 stems from the quadratic sensor transfer function. Inserting the FOM of molography used in
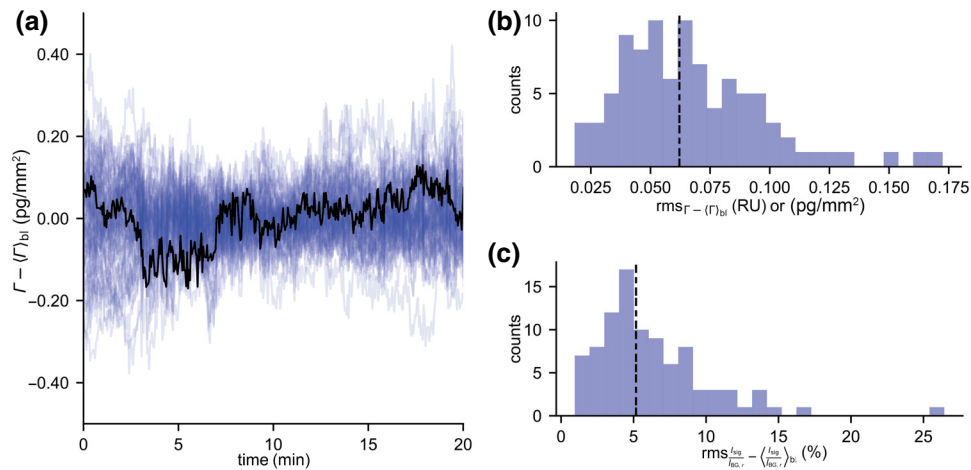
FIG. 28.   Noise levels of reference traces (a) 93 reference traces from one experiment are plotted as a function time (during the 20 min baseline measurement) in units of molographic surface mass density. In black, a typical curve is highlighted. (b) Histogram of the $\text{RMS}_{\Gamma - \langle \Gamma \rangle_{\text{bl}}}$ of the individual curves, the black dashed (–) vertical line is the median. (c) Histogram of the $\text{RMS}_{I_{\text{avg}}/I_{\text{BG},r} - \langle (I_{\text{avg}}/I_{\text{BG},r}) \rangle_{\text{bl}}}$ of the individual curves, the black dashed vertical line is the median. The rms values are calculated from the data and not via the sensitivity of molography. (Eq. (14)).

the real-time measurements, one obtains a sensitivity of $S_\Gamma = 1.1 \times 10^9$ (mm^2/pg). Thanks to this much higher sensitivity, molography can achieve the same resolution as SPR with much larger intensity noise. The sensitivity is, therefore, not suitable to compare the two techniques, since its value depends on the chosen intensity reference.

APPENDIX L: REAL-TIME MEASUREMENTS

1. Instrumentation for real-time measurements

For the real-time measurements, a flow cell and chips with SiO$_2$-covered coupling gratings are developed. Figure 26(a) shows the reader with the mounted flow cell. Figure 26(b)–26(d) display the unmounted flow cell assembly. The design of the flow chamber is similar as in our last publication [7]. However, there is one significant difference between the MoloReader and the ZeptoReader. The MoloReader couples in from the top and outside the flow chamber (compared to the bottom coupling of the ZeptoReader), and therefore at one point the guided mode needs to cross the sealing of the flow chamber. Here, we would like to emphasize that designing a fluidics for a waveguide where the mode has to cross the sealing is not straightforward at all. To our knowledge, the only solution is to encapsulate the evanescent part of the field in a silica coating where the mode crosses the sealing. Without any protective SiO$_2$ cover, any opaque sealing material (like Kalrez) completely extincts the mode. Transparent materials such as PDMS are also unsuitable since they act as lenses that alter the wavefront of the coupled mode. This leads to a sicklelike deformation of the focal spot, changing with the pressure in the flow chamber.

2. Localization of molographic foci for real-time measurements

The focal plane is determined by the NA and the diameter of the mologram and can be easily found by knowing the location of the surface. More challenging in our experimental setting is to have the Airy disk in the field of view. To accomplish this, biotin molograms [NH-biotin|NH$_2$] are fabricated as described in the main manuscript. Then HBS-T buffer is manually injected and the position of the now weaker focal spot is precisely determined in the speckle pattern. Subsequently, the grooves are backfilled by manually injecting 500 μl of 5 mM NHS-PEG$_{12}$-OMe in HBS-T buffer for 15 min and then rinsing with PBS-T. The backfilling results in the focal spot disappearing in the speckle pattern.

3. Processing of real-time binding signals

The processing of real-time signals is carried out offline. First, the images are registered in order to compensate for temperature-related movements of the focal spot in the focal plane [Fig. 27(a)]. After registration, the images are convoluted with the shape of the Airy disk to obtain a weighted averaging over the Airy disk to decrease the influence of hot pixels and pixel noise [Fig. 27(b)]. Each pixel in the convoluted image is related to the average intensity in the Airy disk surrounding it by $I_{\text{sig}} = 2.012 I_{\text{avg}}$. This form of averaging is a trade off between taking the maximum pixel only (not robust to hot pixels and pixel noise) or uniformly averaging over the Airy disk (too conservative). Subsequently, the images are segmented into the signal detection region of interest (ROI$_{\text{sig}}$), the region of interest for normalization (ROI$_{\text{norm}}$), and the

pixel chosen for reference binding traces Fig. 27(c)]. The signal trace is obtained from the location of the highest pixel in the signal region of the last image of the experiment (molographic signal is only visible after binding). In general, any pixel of the convoluted image is a potential candidate for a binding trace, however, they should be sufficiently distant from each other in order not to be correlated. Therefore, the reference binding traces are spaced such that their speckles do not overlap after the convolution operation, and therefore the traces are noncorrelated. This allows the extraction of more hypothetical binding traces and to characterize their noise levels (Fig. 28). Finally, the goal is to obtain signal and reference binding traces that are normalized to the mean intensity of the first image. The first step [Fig. 27(d)] is to compute the average intensity on the ROI_{norm} of the first image. This value is used for three different computations, which are indicated by different arrows. First, the signal and reference traces are normalized by the average intensity of the first image. These normalized traces still contain the drift of the power in the waveguide (due to coupling efficiency changes). By normalizing the average intensity of each image on the ROI_{norm} to the value of the first image [Fig. 27(e)] and multiplying this with the normalized traces, one obtains the drift-compensated relative intensity curves with respect to the average waveguide background of the first image [see Fig. 27(f)]. The reason for this rather complicated referencing procedure becomes apparent from the third use of the average intensity on the ROI_{norm} of the first image. Because the scattering leakage of the waveguide is measured, we can get the initial waveguide power $P_{WG,ini}$ and calculate the equivalent molographic mass density that would correspond to this intensity and then multiply it with the square root of the drift-compensated relative intensity curves. This finally yields the molographic mass binding traces [Fig. 27(g)]. We can only do this operation because they are all normalized to the average intensity in the first image and the equivalent coherent mass density Γ_0 is calculated from this value. From these curves, one can then subtract the baseline and plot the final intensity and binding curves that are used to calculate the rms values of intensity and molographic mass density [Fig. 27(h)].

---

[1] Nabil Mohammad, Monjurul Meem, Bing Shen, Peng Wang, and Rajesh Menon, Broadband imaging with one planar diffractive lens, Sci. Rep. **8**, 2799 (2018).

[2] James T. Early, Roderick Hyde, and Richard L. Baron, in *UV/Optical/IR Space Telescopes: Innovative Technologies and Concepts* (International Society for Optics and Photonics, San Diego, California, United States, 2004), Vol. 5166, p. 148.

[3] Yeonjoon Park, Laura Koch, Kyo D. Song, Sangjoon Park, Glen King, and Sang Choi, Miniaturization of a fresnel spectrometer, J. Opt. A: Pure Appl. Opt. **10**, 095301 (2008).

[4] Rita S. Rodrigues Ribeiro, Pabitra Dahal, Ariel Guerreiro, Pedro A. S. Jorge, and Jaime Viegas, Fabrication of fresnel plates on optical fibres by FIB milling for optical trapping, manipulation and detection of single cells, Sci. Rep. **7**, 4485 (2017).

[5] Istvan Mohacsi, Petri Karvinen, Ismo Vartiainen, Vitaliy A. Guzenko, Andrea Somogyi, Cameron M. Kewish, Pascal Mercere, and Christian David, High-efficiency zone-plate optics for multi-keV x-ray focusing, J. Synchrotron Radiat. **21**, 497 (2014).

[6] Benjamin A. Palmer, Gavin J. Taylor, Vlad Brumfeld, Dvir Gur, Michal Shemesh, Nadav Elad, Aya Osherov, Dan Oron, Steve Weiner, and Lia Addadi, The image-forming mirror in the eye of the scallop, Science **358**, 1172 (2017).

[7] Volker Gatterdam, Andreas Frutiger, Klaus-Peter Stengele, Dieter Heindl, Thomas Lübbers, Janos Vörös, and Christof Fattinger, Focal molography is a new method for the in situ analysis of molecular interactions in biological samples, Nat. Nanotechnol. **12**, 1089 (2017).

[8] Ângela Serrano, Stefan Zürcher, Samuele Tosatti, and Nicholas D. Spencer, Imparting nonfouling properties to chemically distinct surfaces with a single adsorbing polymer: A multimodal binding approach, Macromol. Rapid Commun. **37**, 622 (2016).

[9] Stephanie Fraser, Judy Y. Shih, Mark Ware, Edward O'Connor, Mark J. Cameron, Martin Schwickart, Xuemei Zhao, and Karin Regnstrom, Current trends in ligand binding Real-Time measurement technologies, AAPS J. **19**, 682 (2017).

[10] Jirí Homola, *Surface Plasmon Resonance Based Sensors*, edited by Jirí Homola, Series Springer Series on Chemical Sensors and Biosensors Vol. 4 (Springer, Berlin, Heidelberg, 2006).

[11] Peter Kozma, Florian Kehl, Eva Ehrentreich-Förster, Christoph Stamm, and Frank F. Bier, Integrated planar optical waveguide interferometer biosensors: A comparative review, Biosensors and Bioelectronics **58**, 287 (2014).

[12] Michelle J. Cannon, *et al.*, Comparative analyses of a small molecule/enzymeinteraction by multiple users of biacore technology, Anal. Biochem. **330**, 98 (2004).

[13] Peter Kozma, Andras Hamori, Kaspar Cottier, Sandor Kurunczi, and Robert Horvath, Grating coupled interferometry for optical sensing, Appl. Phys. B **97**, 5 (2009).

[14] Marek Piliarik, Markéta Bocková, and Jiří Homola, Surface plasmon resonance biosensor for parallelized detection of protein biomarkers in diluted blood plasma, Biosens. Bioelectron. **26**, 1656 (2010).

[15] Marek Piliarik, and Vahid Sandoghdar, Direct optical sensing of single unlabelled proteins and super-resolution imaging of their binding sites, Nat. Commun. **5**, 4495 (2014).

[16] Christof Fattinger, Focal molography: Coherent microscopic detection of biomolecular interaction, Phys. Rev. X **4**, 031024 (2014).

[17] Stephanie Pasche, Susan M. De Paul, Janos Voros, Nicholas D. Spencer, and Marcus Textor, Poly(-lysine)-graft-poly(ethylene glycol) assembled monolayers on niobium oxide surfaces: A quantitative study of the influence of polymer interfacial architecture on resistance to protein adsorption by ToF-SIMS and in situ OWLS, Langmuir **19**, 9216 (2003).

[18] J. Christopher Love, Lara A. Estroff, Jennah K. Kriebel, Ralph G. Nuzzo, and George M. Whitesides, Self-assembled monolayers of thiolates on metals as a form of nanotechnology, Chem. Rev. **105,** 1103 (2005).

[19] D. Falconnet, A. Koenig, F. Assi, and M. Textor, A combined photolithographic and Molecular-Assembly approach to produce functional micropatterns for applications in the biosciences, Adv. Funct. Mater. **14,** 749 (2004).

[20] N. Vigneswaran, Fahmi Samsuri, Balu Ranganathan, Padmapriya, Recent advances in nano patterning and nano imprint lithography for biological applications, Procedia Engineering **97,** 1387 (2014).

[21] Miquel Avella-Oliver, Javier Carrascosa, Rosa Puchades, and Ángel Maquieira, Diffractive protein gratings as optically active transducers for High-Throughput label-free immunosensing, Anal. Chem. **89,** 9002 (2017).

[22] Steve Cleverley, Irene Chen, and Jean-Francois Houle, Label-free and amplified quantitation of proteins in complex mixtures using diffractive optics technology, J. Chromatogr. B Analyt. Technol. Biomed. Life Sci. **878,** 264 (2010).

[23] Steven Lenhert, Falko Brinkmann, Thomas Laue, Stefan Walheim, Christoph Vannahme, Soenke Klinkhammer, Miao Xu, Sylwia Sekula, Timo Mappes, Thomas Schimmel, and Harald Fuchs, Lipid multilayer gratings, Nat. Nanotechnol. **5,** 275 (2010).

[24] Zhian Lai, Yuli Wang, Nancy Allbritton, G.-P. Li, and Mark Bachman, Label-free biosensor by protein grating coupler on planar optical waveguides, Opt. Lett. **33,** 1735 (2008).

[25] An error in Eq. (2) of Ref. [24] should be noted: Λ does not refer to the grating period but to the wavelength of the $n = -1$ wave inside the grating region in the direction normal to the waveguide surface as described in Ref. [39].

[26] Herwig Kogelnik, Coupled wave theory for thick hologram gratings, The Bell System Technical Journal **48,** 2909 (1969).

[27] Ian Poole, *Basic Radio: Principles and Technology* (Newnes, 1998).

[28] Selection guide biacoresystems, https://proteins.gelifesciences.com/, accessed:2017-12-1.

[29] Dietrich Marcuse, *Theory of Dielectric Optical Waveguides* (Academic Press, London, United Kingdom, 1974).

[30] Lukas Novotny, and Bert Hecht, *Principles of Nano-Optics* (Cambridge University Press, Cambridge, United Kingdom, 2011).

[31] Lukas Novotny, Allowed and forbidden light in near-field optics. II. Interacting dipolar particles, J. Opt. Soc. Am. A **14,** 105 (1997).

[32] Lukas Novotny, Allowed and forbidden light in near-field optics. I. A single dipolar light source, J. Opt. Soc. Am. A **14,** 91 (1997).

[33] Thomas L. McMeekin, Merton L. Groves, and Norbert J. Hipp, *Amino Acids and Serum Proteins*, Series Advances in Chemistry, Vol. 44 (American Chemical Society, Washington, United States, 1964), Chap. 4, p. 54.

[34] J. A. De Feijter, J. Benjamins, and F. A. Veer, Ellipsometry as a tool to study the adsorption behavior of synthetic and biopolymers at the air–water interface, Biopolymers **17,** 1759 (1978).

[35] Joseph W. Goodman, *Introduction to Fourier Optics* (Roberts and Company Publishers, Greenwood Village, United States, 2005).

[36] D. Sinclair, Light scattering by spherical particles, J. Opt. Soc. Am. **37,** 475 (1947).

[37] Eugene Hecht, *Optics* (Addison Wesley Publishing Company Incorporated, Boston, Massachusetts, United States, 2016).

[38] S. Monneret, P. Huguet-Chantôme, and F. Flory, M-lines technique: Prism coupling measurement and discussion of accuracy for homogeneous waveguides, J. Opt. A: Pure Appl. Opt. **2,** 188 (2000).

[39] T. Tamir, and S. T. Peng, Analysis and design of grating couplers, J. Phys. D Appl. Phys. **14,** 235 (1977).

[40] Huaying Zhao, Patrick H. Brown, and Peter Schuck, On the distribution of protein refractive index increments, Biophys. J. **100,** 2309 (2011).

[41] R. Magnusson, and T. K. Gaylord, Diffraction efficiencies of thin phase gratings with arbitrary grating shape, J. Opt. Soc. Am. **68,** 806 (1978).

[42] Wilfried Heller, Remarks on refractive index mixture rules, J. Phys. Chem. **69,** 1123 (1965).

[43] Hannesm Fischer, Igor Polikarpov, and Aldo F. Craievich, Average protein density is a molecular-weight-dependent function, Protein Sci. **13,** 2825 (2004).

[44] W. Lukosz, Light emission by multipole sources in thin layers. I. Radiation patterns of electric and magnetic dipoles, J. Opt. Soc. Am. **71,** 744 (1981).

[45] Structural Genomics Consortium, *et al.*, Protein production and purification, Nat. Methods **5,** 135 (2008).

[46] J. C. Dainty, in *Progress in Optics*, edited by E. Wolf (Elsevier, Amsterdam, Netherlands, 1977), Vol. 14, p. 1.

[47] S. Miyanaga, T. Asakura, and M. Imai, Scattering characteristics of a beam mode in a dielectric-slab optical waveguide. Part II, Opt. Quantum Electron. **12,** 23 (1980).

[48] Liangfang Zhang, Kevin Dammann, Sung Chul Bae, and Steve Granick, Ligand–receptor binding on nanoparticle-stabilized liposome surfaces, Soft. Matter. **3,** 551 (2007).

[49] Shu-Qun Liu, Xing-Lai Ji, Yan Tao, De-Yong Tan, Ke-Qin Zhang, and Yun-Xin Fu, *Protein Engineering* (InTech, London, United Kingdom, 2012).

[50] David W. Hahn, Light scattering theory, Department of Mechanical and Aerospace Engineering, University of Florida (2009).

[51] Francois Ladouceur, and John D. Love, *Silica-based buried channel wave guides and devices* (Chapman & Hall, Boca Raton, Florida, 1996).

[52] F. P. Payne, and J. P. R. Lacey, A theoretical analysis of scattering loss from planar optical waveguides, Opt. Quantum Electron. **26,** 977 (1994).

[53] G. Ames, and D. Hall, Attenuation in planar optical waveguides: Comparison of theory and experiment, IEEE J. Quantum Electron. **19,** 845 (1983).

[54] S. Miyanaga, T. Asakura, and M. Imai, Scattering characteristics of a beam mode in dielectric-slab optical waveguide, Opt. Quantum Electron. **11,** 205 (1979).

[55] Roncone Ronald Louis, PhD thesis, School The University of Arizona, 1992.

[56] J. P. R. Lacey, and F. P. Payne, Radiation loss from planar waveguides with random wall imperfections, IEE Proceedings J-Optoelectronics **137**, 282(1990).

[57] David A. Armbruster, and Terry Pry, Limit of blank, limit of detection and limit of quantitation, Clin. Biochem. Rev. **29**(S49 (2008).

[58] Carly A. Holstein, Maryclare Griffin, Jing Hong, and Paul D. Sampson, Statistical method for determining and comparing limits of detection of bioassays, Anal. Chem. **87**, 9795 (2015).

[59] Jirí Homola, Surface plasmon resonance sensors for detection of chemical and biological species, Chem. Rev. **108**, 462 (2008).

[60] Marek Piliarik, and Jirí Homola, Surface plasmon resonance (SPR) sensors: Approaching their limits? Opt. Express **17**, 16505 (2009).

[61] Richard B. M. Schasfoort, ed. *Handbook of Surface Plasmon Resonance* (The Royal Society of Chemistry, London, United Kingdom, 2017), p. P001.

[62] A. A. Kolomenskii, P. D. Gershon, and H. A. Schuessler, Sensitivity and detection limit of concentration and adsorption measurements by laser-induced surface-plasmon resonance, Appl. Opt. **36**, 6539 (1997).

[63] Eric M. Yeatman, Resolution and sensitivity in surface plasmon microscopy and sensing, Biosensors and Bioelectronics **11**, 635 (1996).

[64] W. Lukosz, Principles and sensitivities of integrated optical and surface plasmon sensors for direct affinity sensing and immunosensing, Biosensors and Bioelectronics **6**, 215 (1991).

[65] K. Tiefenthaler, and W. Lukosz, Sensitivity of grating couplers as integrated-optical chemical sensors, J. Opt. Soc. Am. B **6**, 209 (1989).

[66] P. Thevenaz, U. E. Ruttimann, and M. Unser, A pyramid approach to subpixel registration based on intensity, IEEE Trans. Image Process. **7**, 27 (1998).

[67] Todd M. Squires, Robert J. Messinger, and Scott R. Manalis, Making it stick: Convection, reaction and diffusion in surface-based biosensors, Nat. Biotechnol. **26**, 417 (2008).

[68] Silvia Surinova, Ralph Schiess, Ruth Hüttenhain, Ferdinando Cerciello, Bernd Wollscheid, and Ruedi Aebersold, On the development of plasma protein biomarkers, J. Proteome Res. **10**, 5 (2011).

[69] J. P. Landry, Yaohuang Ke, Guo-Liang Yu, and X. D. Zhu, Measuring affinity constants of 1450 monoclonal antibodies to peptide targets with a microarray-based label-free assay platform, J. Immunol. Methods **417**, 86 (2015).

[70] S. Martellucci, and A. N. Chester, eds. *Diffractive Optics and Optical Microsystems* (Springer, New York City, United States, 1997).

[71] S. Swann, Film thickness distribution in magnetron sputtering, Vacuum **38**, 791 (1988).

[72] Toshikazu Nishide, and Fujio Mizukami, Preparation and properties of TiO–SiO (1 : 1) films prepared by a complexing agent-assisted sol–gel process, Thin Solid Films **298**, 89 (1997).

[73] K. Tiefenthaler, and W. Lukosz, Sensitivity of grating couplers as integrated-optical chemical sensors, J. Opt. Soc. Am. B **6**, 209 (1989).