Bergmann, Ralf B. ; Fischer, Andreas ; Bockelmann, Carsten ; Dekorsy, Armin ; Garcia-Ortiz, Alberto ; Falldorf, Claas

The coherence function and its information content for optical metrology

Journal Article as: peer-reviewed accepted version (Postprint)

DOI of this document*(secondary publication): https://doi.org/10.26092/elib/3332

Publication date of this document: 20/09/2024

* for better findability or for reliable citation

Recommended Citation (primary publication/Version of Record) incl. DOI:

Bergmann, R., Fischer, A., Bockelmann, C., Dekorsy, A., Garcia-Ortiz, A. & Falldorf, C. (2022). The coherence function and its information content for optical metrology. tm - Technisches Messen, 89(6), 397-412. https://doi.org/10.1515/teme-2022-0048

Ralf B. Bergmann*, Andreas Fischer, Carsten Bockelmann, Armin Dekorsy, Alberto Garcia-Ortiz, and Claas Falldorf

The coherence function and its information content for optical metrology

Die Kohärenzfunktion und ihr Informationsgehalt für die optische Messtechnik

Abstract: The coherence function offers new possibilities for optical metrology that are not available with conventional wave field sensing. Its measurement involves a spatio-temporal sampling of the wave fields modulated by the object under investigation. Temporal sampling is well known e. g. by means of White Light Interferometry (WLI) and spatial sampling can e. g. performed by Computational Shear Interferometry (CoSI). The present paper describes an approach that combines both temporal and spatial sampling using a robust common-path setup. While the evaluation of the coherence function is more elaborate than approaches that either sample the temporal or the spatial domain, an information theoretical treatment shows that it also delivers more information about the object under investigation. Our approach is based on the mutual information that represents the reduction of uncertainty about the object as a consequence of the measurements performed. Using a simplified measurement case, we calculate the mutual information for different measurement situations and demonstrate that spatio-temporal sampling of the coherence function results in a higher mutual information as compared to classical approaches. Based on the proposed approach, we identify further open research tasks for an efficient information extraction from the coherence function to surpass current limitations of optical metrology.

Keywords: Optical metrology, interferometry, computational shear interferometry, coherence, spatio-temporal sampling, information, transinformation, compressed sensing.

Zusammenfassung: Die Kohärenzfunktion bietet neue Möglichkeiten für die optische Messtechnik, die mit konventioneller Wellenfeldsensorik nicht verfügbar sind. Die Messung der Kohärenzfunktion beinhaltet eine raumzeitliche Abtastung der vom Messobjekt veränderten Wellenfelder. Die zeitliche Abtastung z. B. mittels Weißlichtinterferometrie (WLI) ist wohlbekannt, eine räumliche Abtastung kann z. B. mittels Computational Shear Interferometry (CoSI) erfolgen. Die vorliegende Veröffentlichung beschreibt einen Ansatz, der sowohl eine zeitliche als auch eine räumliche Abtastung mit einem robusten Common-Path-Aufbau kombiniert. Während die Auswertung der Kohärenzfunktion aufwändiger ist als Ansätze, die entweder eine zeitliche oder eine räumliche Domäne abtasten, zeigt eine informationstheoretische Betrachtung, dass sie auch mehr Informationen über das Messobjekt liefert. Unser Ansatz basiert auf der Transinformation, die die Verringerung der Unsicherheit über das Messobjekt als Folge der durchgeführten Messungen darstellt. Anhand einer vereinfachten Messsituation berechnen wir die Transinformation für verschiedene Messsituationen und zeigen, dass die raum-zeitliche Abtastung der Kohärenzfunktion im Vergleich zu klassischen Ansätzen zu einer höheren Transinformation führt. Basierend auf dem vorgeschlagenen Ansatz identifizieren wir weitere offene Forschungsfragen für eine effiziente Informationsextraktion aus der Kohärenzfunktion, um derzeitige Beschränkungen der optischen Messtechnik zu überwinden.

Schlagwörter: Optische Messtechnik, Interferometrie, Computational Shear Interferometry, Kohärenz. Raumzeitliche Abtastung, Information, Transinformation, Compressed Sensing.

***Corresponding author: Ralf B. Bergmann,** Bremer Institut für angewandte Strahltechnik GmbH (BIAS), 28359 Bremen, Klagenfurter Str. 5, Germany, e-mail: bergmann@bias.de, ORCID: https://orcid.org/0000-0003-0214-2232
Andreas Fischer, Universität Bremen, Bremer Institut für Messtechnik, Automatisierung und Qualitätswissenschaft (BIMAQ), 28359 Bremen, Linzer Str. 13, Germany, ORCID: https://orcid.org/0000-0001-7349-7722
Carsten Bockelmann, Armin Dekorsy, Universität Bremen, Arbeitsbereich Nachrichtentechnik (ANT), 28359 Bremen, Otto-Hahn-Allee 1, Germany, ORCID: https://orcid.org/0000-0002-8501-7324 (C. Bockelmann)
Alberto Garcia-Ortiz, Universität Bremen, Integrated Digital Systems (IDS), 28359 Bremen, Germany, ORCID: https://orcid.org/0000-0002-6461-3864
Claas Falldorf, Bremer Institut für angewandte Strahltechnik GmbH (BIAS), 28359 Bremen, Klagenfurter Str. 5, Germany, ORCID: https://orcid.org/0000-0001-6481-5709

1 Introduction

Conventional interferometric measurement methods determine the phase or complex amplitude of a quasi-monochromatic optical wave field. Within scalar diffraction theory, light is described by means of the *time-dependent* complex amplitude, a scalar function $U(\vec{x}, t)$ which depends on space $\vec{x}$ and time t. Except for the polarization state, it contains all information of the wave field but cannot be measured directly because its oscillation in time is too fast for any detector based on electronics.

The *time-independent* complex amplitude $U(\vec{x})$ is usually briefly referred to as the complex amplitude. It can be measured by its correlation with a reference wave $R(\vec{x}, t)$ with the same temporal dependence by evaluating the intensity

$$I = (U^* + R)(U + R^*) = |U|^2 + |R|^2 + 2\mathcal{R}\{U^* R\} \qquad (1)$$

using a suitable detector, typically a CCD or CMOS camera. Here, $\mathcal{R}$ stands for the real part of $U^* R$. The complex amplitude is only well defined in the monochromatic case, because the reference wave R has to follow the time dependence of U which can only be realized at any point in space in a monochromatic regime.

In order to considerably expand the possibilities of interferometric measurement technology, we widen our scope to the coherence function of light $\Gamma(\vec{x}_1, \vec{x}_2, \tau)$ that describes the covariance of the wave field at different locations $\vec{x}_1$ and $\vec{x}_2$ and a time shift τ. It is, for a time-dependent wave field $U(\vec{x}, t)$ given by

$$\begin{aligned}\Gamma(\vec{x}_1, \vec{x}_2, \tau) &= \langle U^*(\vec{x}_1, t) U(\vec{x}_2, t + \tau) \rangle_T \\ &= \lim_{T \to \infty} \frac{1}{T} \int_{t=-T/2}^{t=T/2} U^*(\vec{x}_1, t) U(\vec{x}_2, t + \tau)\, dt, \end{aligned} \qquad (2)$$

with the time average $\langle \dots \rangle_T$ defined at the right side of the equation [1]. This description is valid, when the light is stationary, as it is the case in most metrology applications. Using the Γ-function allows us, however, not only to analyze a monochromatic wave field, but also partly coherent wave fields consisting of polychromatic light and even several independent light fields present at the same time. It can therefore be expected that using the Γ-function a higher degree of information can be obtained from optical measurements as compared to conventional wave field measurements. However, this advantage comes with a price: The Γ-function in the most general case maps a 7-dimensional (7D) space spanned by $\vec{x}_1$, $\vec{x}_2$ and τ to a complex-valued function. As light fields are usually detected using planar detectors, the space vectors $\vec{x}_1$ and $\vec{x}_2$ provide two dimensions each so that the detected Γ-function has a maximum of 5 dimensions.

We will start our discussion by determining a monochromatic wave function using the well known Michelson interferometry and then expand the discussion to the determination of the Γ-function by the use of Γ-Profilometry, a concept that will be introduced in greater detail later on. The discussion follows the leading question: Is the measurement of the Γ-function more informative as compared to measuring the wave function?

Figure 1 shows a simplified scheme of a Michelson interferometer that is frequently used for high precision optical metrology. In the highly idealized setting we disregard limiting effects such as diffraction, aberration or deviations from ideal plane waves. The observation plane corresponds to the x-axis. For simplicity, we assume a stepped surface with the step located at the z-axis along $y = 0$ and a step height of Δh as shown in Fig. 1(a). The measurement task is to determine the step height Δh that increases the optical path length with respect to light coming from a higher point of the sample surface corresponding

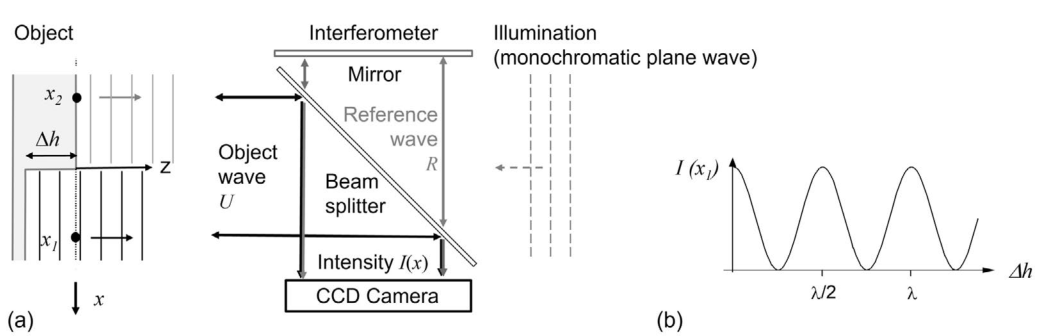

Figure 1: (a) Schematic drawing of a Michelson interferometer. The object wave interferes with the reference wave producing an intensity $I(x)$ at the detector. For simplicity, we assume that the paths that both reference and object wave coming from the plane that incorporates x_2 are identical. (b) Dependence of the intensity $I(x_1)$ on Δh scaled by the wavelength λ.

to x_2 at the observation plane compared to light coming from a lower part corresponding to x_1 at the observation plane by $2\Delta h$ resulting in a time delay of $\delta\tau = 2\Delta h/c$. In the setup (from right to left), monochromatic light is split in two parts, an object wave U and a reference wave R. In the general case, the path difference of object and reference arm is variable and is the sum of four components: A length Δl that allows to introduce a phase shift $\phi = 2\pi\,\Delta l/\lambda$, an adjustment error δl_i intrinsic to the measurement setup, an error δl_e introduced by external effects and a length ϵ corresponding to noise. In practical applications, δl_e is mainly caused by vibrations, a dominant concern for precise measurements. The waves coming from these arms interfere with each other at the detector plane.

For the dependence shown in Fig. 1(b), we neglect the above mentioned length differences and assume that the path for the light of the object beam traveling from x_2 to the detector has the same length as that of the reference wave and therefore results in a constant maximum intensity I at the detector. The length of the path of the light from x_1, however, depends on Δh and leads to the dependence of the intensity $I(x_1)$ at the detector as shown in Fig. 1(b). The relation between reference wave and object wave arriving from x_1 at the detector can thus be described by $U(x_1 - 2\Delta h) = R(x_1)$. The interference term $\mathcal{R}\{U^* R\}$ in Eq. 1 is thus given by the real part of a correlation function $U^*(x_1)\,U(x_1 - 2\Delta h)$, which is a correlation of $U(\vec{x}, t)$ with itself.

As a note important for the following line of arguments, please observe that the Michelson interferometer in the general case measures the *temporal correlation* expressed by

$$\Gamma(\vec{x}, \delta\tau) = \left\langle U^*(\vec{x}, t)\, U(\vec{x}, t + \delta\tau)\right\rangle_T . \tag{3}$$

with a time shift $\delta\tau$ resulting from the combined effect of Δh, Δl, δl_i, δl_e and ϵ as discussed above.

We will now extend our discussion to the concept of Γ-Profilometry. Figure 2 shows a simplified Γ-Profilometry setup [2] that comprises a temporal sampling unit and a spatial sampling unit. The common path configuration uses in its temporal sampling unit a Soleil-Babinet compensator as part of a $4f$-configuration that, dependent on the polarization set by the upper polarizer, creates two wavefronts with a temporal shift $\Delta\tau$ with respect to each other. The spatial sampling unit consists of a computational shear interferometer that uses a spatial light modulator within a $4f$-configuration to create a spatial shift, the so called shear $\vec{s}$, between two images that interfere at the CCD camera, once the polarization is properly set by the polarizer at the right side of the figure. The setup therefore samples the coherence function

$$\begin{aligned}\Gamma(\vec{x}, \vec{x} + \vec{s}, \Delta\tau) &= \left\langle U^*(\vec{x}, t) U(\vec{x} + \vec{s}, t + \Delta\tau)\right\rangle_T \\ &= U^*(\vec{x}) U(\vec{x} + \vec{s}, \Delta\tau).\end{aligned} \tag{4}$$

and allows to independently adjust $\Delta\tau$ and $\vec{s}$ for sampling the temporal and the spatial correlation, respectively. As there is no reference arm as in the case of the Michelson interferometer, errors only arise from errors in the setting of $\Delta\tau$ by the Soleil-Babinet compensator, errors in setting the shear $\vec{s}$ and noise represented by a suitable ϵ. Due to the common path principle, an otherwise often dominating error due to vibrations (comparable to δl_e) does not exist.

Figure 3 schematically illustrates the situation for a $\Delta\tau$-s plane where the multi-dimensional $\Delta\tau$-$\vec{s}$ space is reduced to a plane with a scalar s for simplicity. Although $\vec{s}$ is in principle a vector in three dimensions, we restrict $\vec{s}$ to a two-dimensional vector within the observation plane since a shift-component outside this plane would correspond to a temporal shift in non-dispersive media (compare the corresponding argument for the Michelson setup in Fig. 1).

Techniques such as White Light Interferometry (WLI), Computational Shear Interferometry (CoSI) and Γ-Profilometry appear well suited to probe subsamples of the Γ-function and have already been extensively explored by the authors [2–6]. Concerning the coherence properties of the light, shear interferometry merely demands the spatial coherence to be larger than the shear. The spatial coherence provided by light sources commonly used for WLI, e. g. light emitting or super-luminescence diodes (LEDs or SLDs), in general meets these demands [2].

A systematic investigation of the advantages of the Γ-function over conventional wave field measurements in terms of information content and sampling efficiency has not yet been conducted. Using WLI, light coming from an observation plane is detected using a time shift along the vertical $\Delta\tau$-axis. The horizontal axis of Fig. 3 represents spatial sampling using CoSI. Here, light comes from an observation plane that is shifted against itself by a vector $\vec{s}$. Finally, Γ-Profilometry enables sampling of the complete $\Delta\tau$-$\vec{s}$ space as shown in Fig. 3 by simultaneously using $\Delta\tau$ and $\vec{s}$ to shift the observation plane in time and space. The gray areas in the $\Delta\tau$-s plane schematically represent the fraction of the $\Delta\tau$-$\vec{s}$ space used for sampling by the approaches described above.

In the case of WLI as an example, any uncertainty that is caused by a reference arm, e. g. deviations of the refer-

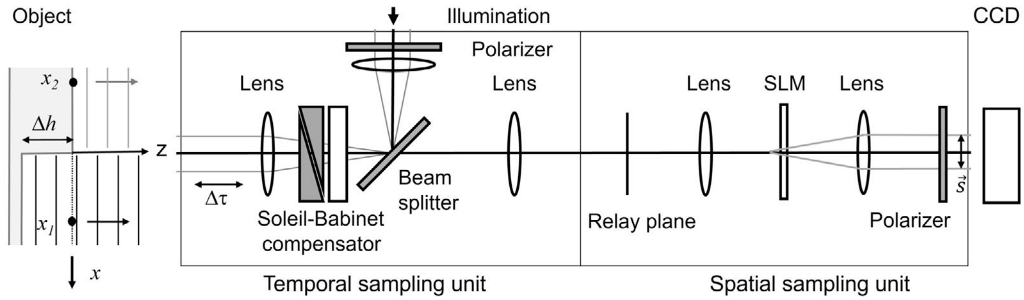

Figure 2: Simplified Γ-Profilometry setup. Light from the object is captured by the left unit, a $4f$-configuration formed by the two lenses in the left box. This temporal sampling unit enables temporal sampling by use of a Soleil-Babinet compensator that allows to introduce a time shift $\Delta\tau$ between the two differently polarized parts of the beam. The right box, the spatial sampling unit, consists of a shear interferometer (for simplicity in an unfolded configuration) based on a spatial light modulator (SLM) that allows to create an adjustable shear $\vec{s}$. The second $4f$-configuration is again formed by two lenses. The setup thus allows to sample the $\Delta\tau$-$\vec{s}$-space. Due to the common-path principle employed, the setup is fairly insensitive with respect to vibrations. For further description of the optical setup and its use see [2].

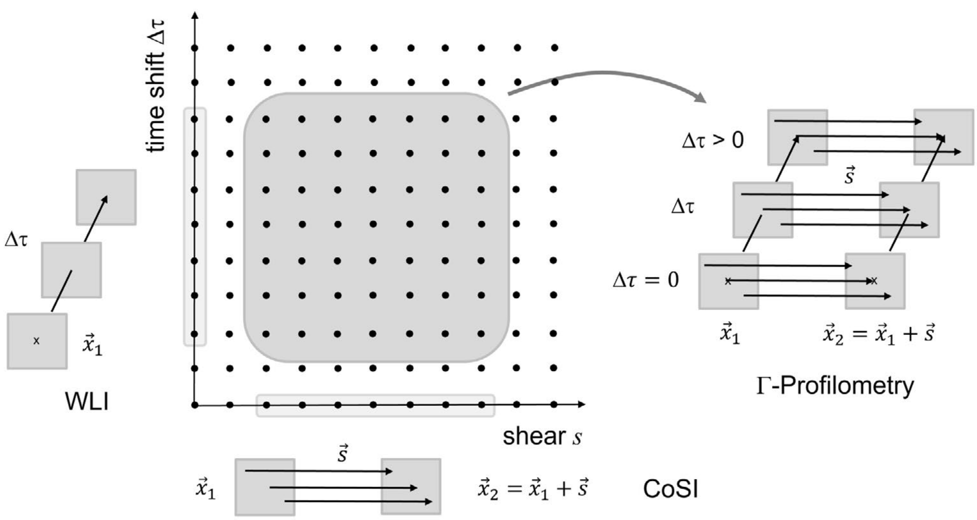

Figure 3: Schematic representation of the $\Delta\tau$-$\vec{s}$ parameter space reduced to a $\Delta\tau$-s plane. Gray areas schematically indicate the parameter space of measurements with i) White Light Interferometry (WLI) representing $\Delta\tau$-scanning, ii) Computational Shear Interferometry (CoSI) representing spatial scanning along the shear-axis and iii) Γ-Profilometry sampling the whole $\Delta\tau$-s plane using $\Delta\tau$ and $\vec{s}$ as variable measurement parameters. A complete measurement combines the analysis of the data from all measurements indicated by black dots within the respective gray area. Note that this conceptual illustration is strongly simplified. Rather than a scalar value, each sampling concerns the whole measurement plane providing information about the different parameters of the object. The information provided by a set of sampling points is not additive, but a complex function that depends on the correlation of the different points which cannot be evaluated individually.

ence mirror from a perfect flat (to just mention a trivial cause for simplicity), is always fully correlated with the entire object light and therefore affects the entire object. We can thus not distinguish between errors invoked by the reference arm and the actual topography of the object, no matter how many measurements we perform. In Γ-Profilometry however, we compare different regions of the object. Although the imaging system can also create unwanted deviations, we can easily separate them from the object information by performing measurements with different shears. Deviations which remain constant for different shears will not be regarded as object information through the reconstruction process. This is an intrinsic information advantage of the coherence function because the object information is encoded during multiple measurements and can thus be distinguished from artifacts created from the optical measurement system.

One of the major benefits of sampling the coherence function is therefore that only light returning from the object is investigated. All the collected information is based

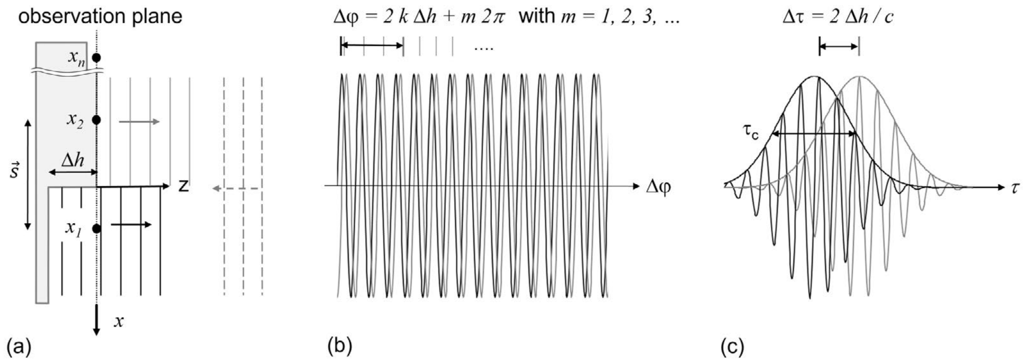

Figure 4: Idealized measurement situation for step height determination. (a) Surface with a step of height Δh at $z = 0$ and further steps indicated by further measurement points x_n. Dashed lines: incident wave front, black line: wave front reflected from lower part of the surface, gray line: wave front reflected from upper part of the surface. (b) Phase difference $\Delta\varphi$ of the two reflected wave fronts with wave number k. Phase differences differing by multiples of 2π can not be distinguished. (c) Using poly chromatic light with a suitable coherence time τ_c allows to uniquely evaluate the step height Δh from the shift $\Delta\tau$ of the two Γ-functions. For further details see text.

on correlations between light reflected by one area of the object with light from another area. The choice of the shear $\vec{s}$ thus introduces an important additional degree of freedom.

We will now describe several cases of optical metrology based on the above discussion using temporal as well as spatio-temporal sampling. *Using concepts from information theory, we demonstrate that spatio-temporal sampling delivers more information content than approaches restricted to sample either over space or time.*

2 Measurements with monochromatic, polychromatic and multiple light sources

Figure 4 shows an extension of the simplified measurement situation of the Figs. 1 and 2. Using the geometry shown in Fig. 4(a), we describe how to calculate the wave field and the coherence function for the cases of monochromatic and polychromatic light. As indicated in Fig. 4(b), the use of monochromatic light frequently creates ambiguities in the determination of the step height Δh from a phase shift $\Delta\phi$ which can be overcome by measuring the time shift $\Delta\tau$ of the maximum obtained from the use of polychromatic light, see Fig. 4(c). We first discuss the task of measuring a single step height and later on generalize it for many steps by introducing more points $\vec{x}_n$. A continuous surface profile can be dealt with by adding further points $\vec{x}_n$ and assume $n \to \infty$. The mathematical results obtained from these situations are used later on in Section 3 to determine the information content of the measurements.

2.1 Measurement of a step height using monochromatic light

Using a monochromatic plane wave

$$U(z,t) = u_0 \cdot \exp\left[\mathrm{i}(kz - \omega t)\right] \tag{5}$$

with the amplitude u_0, the magnitude k of the vector in z-direction and the angular frequency ω that is reflected at the surface, a phase shift is observed between light coming from $\vec{x}_1$ and $\vec{x}_2$. Using a shear interferometer with the shear $\vec{s}$ set according to the distance between the two points, the amplitude results in

$$\begin{aligned} U(z,t) &= \frac{u_0}{\sqrt{2}}\left(e^{\mathrm{i}[kz+\omega t+\phi]} + e^{\mathrm{i}[k(z+2\Delta h)+\omega t+\phi]}\right) \\ &= \frac{u_0}{\sqrt{2}}\, e^{\mathrm{i}(kz+\omega t+\phi)}\left(1 + e^{\mathrm{i}2k\Delta h}\right) \end{aligned} \tag{6}$$

with an additional unknown, but common phase shift ϕ. The factor $\sqrt{2}$ comes from the fact that the incoming field is split in two equal parts. We calculate the measured intensity

$$I(\Delta h) = |U \cdot U^*|^2 = \frac{u_0^2}{2}\left[2 + 2\cos(2k\Delta h)\right] = I_0\left[1 + \cos(2k\Delta h)\right] \tag{7}$$

which reads after normalization

$$\frac{I(\Delta h)}{I_0} = 1 + \cos(2k\Delta h). \tag{8}$$

The step height thus creates a phase shift $\Delta\varphi = 2k\Delta h$ between the two waves. The phase shift can be determined by using phase shifting techniques [7]. However, the periodicity of the cos-term introduces an ambiguity given by $m\,2\pi$ that results in ambiguous phase shifts

$$\Delta\varphi_n = 2k\,\Delta h + m\,2\pi \tag{9}$$

with m = 1, 2, 3, ... as can be seen in Fig. 4(b). As a consequence of this ambiguity, step heights differing by $m \cdot \lambda/2$ cannot be distinguished. This situation can be alleviated by using a two-wavelength approach with a synthetic wavelength of

$$\Lambda = \frac{\lambda_1 \cdot \lambda_2}{|\lambda_1 - \lambda_2|}. \tag{10}$$

which extends the unambiguity range to $\Lambda/2$. A generalization of this concept can be found in [8].

2.2 Measurement of a step height using polychromatic light

Now we consider the use of polychromatic light. The relation between the waves reflected from the surface is evaluated in the observation plane by measuring the coherence function $\Gamma(\vec{x}_1, \vec{x}_2, \tau)$ as defined in Eq. 2. In many situations the power spectral density $S(\omega)$ of the light source is a priori known. In this case, it is useful to determine the coherence function by exploiting its Fourier relationship with the cross spectral density (CSD) $S(\vec{x}_1, \vec{x}_2, \omega)$ given by

$$\Gamma(\vec{x}_1, \vec{x}_2, \tau) = \int_{-\infty}^{\infty} S(\vec{x}_1, \vec{x}_2, \omega) \cdot \exp(\mathrm{i}\omega\tau)\, d\omega, \tag{11}$$

see e. g. [9]. In [2] it is shown that spatial and spectral dependencies of the CSD can be separated, if the light field at the positions $\vec{x}_1$ and $\vec{x}_2$ shares the same spectral characteristics. In our simple example we can assume this requirement to hold true and find

$$\Gamma(\vec{x}_1, \vec{x}_2, \tau) = \int_{-\infty}^{\infty} S(\omega) \cdot \exp\left(\mathrm{i}\omega\left[\tau - \Delta\tau(\vec{x}_1, \vec{x}_2)\right]\right)\, d\omega \tag{12}$$

with

$$\Delta\tau(\vec{x}_1, \vec{x}_2) = \begin{cases} 2\Delta h/c & \text{for all } \vec{x}_1 = (x_1, y_1, 0) \text{ with } x_1 \geq 0 \text{ and} \\ & \vec{x}_2 = (x_2, y_2, 0) \text{ with } x_2 < 0 \\ 0 & \text{otherwise} \end{cases} \tag{13}$$

according to the stepped surface defined in Fig. 4(a). Equation 12 resembles the Wiener-Khinchin theorem [9], but with the time delay shifted by the additional time difference $\Delta\tau(\vec{x}_1, \vec{x}_2)$ which corresponds to the optical path difference between light reflected at the surface corresponding to $\vec{x}_1$ and $\vec{x}_2$. We now assume a Gaussian power spectral density of the wave field given by

$$S(\omega) = \frac{I_o}{\sqrt{2\pi}\Delta\omega} \exp\left(-\frac{(\omega - \omega_0)^2}{2\Delta\omega^2}\right). \tag{14}$$

with the total intensity of the distribution $I_0 = \int_{-\infty}^{\infty} S(\omega)\, d\omega$, the central frequency ω_0 and the standard deviation $\Delta\omega$. With the time τ that describes the time between sending and receiving the wave front reflected from $\vec{x}_2 = \vec{x}_1 \leq 0$ due to its propagation in the measurement setup and $\Delta\tau$ given in Eq. 13, the time shift between the two waves coming from $\vec{x}_1$ and $\vec{x}_2$ is $\tau - \Delta\tau$ and the coherence function results in

$$\begin{aligned}\Gamma(\tau - \Delta\tau) = I_0 \cdot {} & \exp\left(\mathrm{i}\omega_0\left[\tau - \Delta\tau(\vec{x}_1, \vec{x}_2)\right]\right) \\ & \cdot \exp\left(-\frac{1}{2}\left[\tau - \Delta\tau(\vec{x}_1, \vec{x}_2)\right]^2 \Delta\omega^2\right).\end{aligned} \tag{15}$$

With the incoming intensity I_0 being split into $I_0/2 = I_0(\vec{x}_1) = I_0(\vec{x}_2)$, the real part of Γ that enters into the measured intensity

$$I(\vec{x}_1, \vec{x}_2, \tau) = I_0(\vec{x}_1) + I_0(\vec{x}_2) + \mathcal{R}\{\Gamma(\vec{x}_1, \vec{x}_2, \tau)\} \tag{16}$$

is given by

$$\begin{aligned}\mathcal{R}\,(\Gamma) = I_0 \cdot {} & \cos\left(\omega_0\left[\tau - \Delta\tau(x_1, x_2)\right]\right) \\ & \cdot \exp\left(-\frac{1}{2}\left[\tau - \Delta\tau(\vec{x}_1, \vec{x}_2)\right]^2 \Delta\omega^2\right).\end{aligned} \tag{17}$$

As a result, the measured normalized intensity as function of Δh is finally given by

$$\frac{I(\Delta h)}{I_0} = 1 + \cos\left[\omega_0\left(\tau - \frac{2\Delta h}{c}\right)\right] \cdot \exp\left[-\frac{1}{2}\left(\tau - \frac{2\Delta h}{c}\right)^2 \Delta\omega^2\right], \tag{18}$$

where we have inserted $\Delta\tau(\vec{x}_1, \vec{x}_2) = 2\Delta h/c$ in accordance with the studied example and the speed of light c. Consequently, sampling of a monochromatic wave field leads to an ambiguity in the determination of a step height of $\Delta h \leq \lambda/2$, compare Eq. 7, or $\Lambda/2$ for the use of a synthetic wavelength, or correspondingly higher values for multi-λ approaches as discussed in the previous subsection. In contrast, using the coherence function given in Eq. 15 allows for an unambiguous determination of the step height provided that the coherence time

$$\tau_c = \frac{2\sqrt{2\ln 2}}{\Delta\omega}, \tag{19}$$

is properly chosen. As is obvious from the equation above, the coherence time τ_c depends on the choice of the spectral width $\Delta\omega$ of the light. Choosing a suitable value of τ_c as schematically depicted in Fig. 4(c) allows to uniquely determine the shift of the coherence function as a result of the step height Δh. Using the Full Width Half Maximum $\mathrm{FWHM} = 2\sqrt{2\ln 2}\,\Delta\omega$, Eq. 19 simplifies to $\tau_c = 1/\mathrm{FWHM}$.

For monochromatic light with $\Delta\omega \to 0$ and thus $\tau_c \to \infty$ the normalized intensity results in

$$\frac{I(\Delta h)}{I_0} = 1 + \cos\left[\omega_0\left(\tau - \frac{2\Delta h}{c}\right)\right]. \tag{20}$$

For the case of a monochromatic wave, no gain of information is obtained in using the Γ-function instead of the wave field. The advantage of using the Γ-function arises, as demonstrated above, for polychromatic light.

2.3 Measurements simultaneously using several light sources

The evaluation of the determination of the Γ-function using several light sources simultaneously has been shown in [6]. The light field thus reads

$$U_G(\vec{x},t) = \sum_n U_n(\vec{x},t) \tag{21}$$

for n independent light fields. The Γ-function then results in

$$\begin{aligned}\Gamma(\vec{x}_1,\vec{x}_2,t) &= \langle U_G^*(\vec{x}_1,t)U_G(\vec{x}_2,t)\rangle_T \\ &= \lim_{T\to\infty}\frac{1}{T}\int_{-T/2}^{T/2}\left(\sum_n U_n^*(\vec{x}_1,t)\right)\left(\sum_m U_m(\vec{x}_2,t)\right)dt.\end{aligned} \tag{22}$$

Due to the fact that only light from a common source can interfere we can use

$$\langle U_n^*(\vec{x}_1,t)U_m(\vec{x}_2,t)\rangle_T = \delta_{n,m}U_n^*(\vec{x}_1)U_m(\vec{x}_2) \tag{23}$$

to simplify the expression to

$$\Gamma(\vec{x}_1,\vec{x}_2) = \sum_n U_n^*(\vec{x}_1)U_n(\vec{x}_2), \tag{24}$$

which can again be measured by a procedure based on a shear interferometer as previously demonstrated [4]. We do not exemplify this case here any further, as the geometry of the employed light sources highly depends on the measurement situation, which is beyond the scope of this article. Measurement examples for the simultaneous use of many light sources are described in [4].

3 Information content of wave field and coherence function

A main goal of this paper is to compare from a fundamental point of view classical optical metrology techniques with the 7D-metrology concept based on the Γ-function. 7D-metrology provides rich information about the dimensions under measure, but at the same time requires a nontrivial post-processing. In order to focus the comparison at first w. r. t. the available information, an information-based analysis of both measurement concepts is presented. Therefore, an information-based framework is developed that allows us not only to reason about the fundamental properties of different measurement options, but also opens the possibility for using statistical learning and compressed signal processing post-processing techniques in future.

3.1 Concept of information content applied to optical metrology

We are now interested in the information transfer from the point of view of information theory. Figure 5 presents a block diagram of the setup shown in Fig. 2 that samples a τ-$\vec{s}$-space. To enable an analysis in terms of information theory, we need to cast the problem in a proper form. Information theory is concerned with the transfer of information from a source (object) to a sink (processing) through a channel (complete measurement system, including the camera) as shown in Fig. 5. This information transfer is best characterized by the so-called *mutual information*. The mutual information quantifies the amount of information that can be transmitted through the measurement system in a single number. In order to define it, we first require an intermediate concept, the so-called *differential entropy*. It uses the statistics of random variables (e. g. Δh) to express the information content independent of the specific statistical distribution. Hence, we start with modeling the physical characteristics of the object, the optical measurement system and the camera including any measurement errors, then define the differential-entropy and finally discuss the mutual information.

We now formalize the a priory knowledge about an object with the help of a probability distribution. First, we assume that the object is characterized by a set of parameters. Also, we assume the parameters to be random variables $\mathcal{G} = \{G_0, G_1, \ldots, G_{n-1}\} \sim p_{\mathcal{G}}(\vec{g})$ with realizations $\vec{g} = [g_0, g_1, \ldots, g_{n-1}]$, where $p_{\mathcal{G}}(\vec{g})$ denotes the probability distribution. Note that the degree of a-priory knowl-

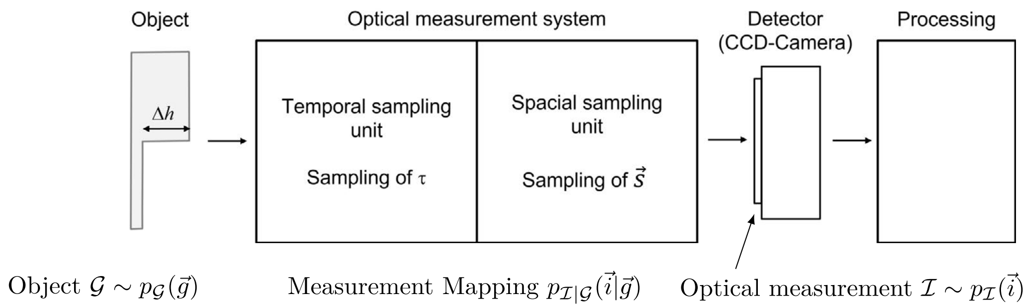

Figure 5: Block diagram of a measurement chain based on Γ-Profilometry as a conceptual model for the information-theoretical description of an optical metrology system. Light from the object with a step Δh is captured by a temporal and a spatial sampling unit that allow to sample the τ-$\vec{s}$-space. For further description of the optical setup and its use see [2]. The information theoretical modeling simplifies this optical model to a probability density function (pdf) for the object parameters $p_{\mathcal{G}}(\vec{g})$, a conditional pdf $p_{\mathcal{I}|\mathcal{G}}(\vec{i}|\vec{g})$ and a pdf for the measurements $p_{\mathcal{I}}(\vec{i})$.

edge captured by this distribution can vary considerably depending on the application. In some scenarios, almost no prior knowledge may exist besides a rough idea of the range of the parameters while in other scenarios, the information can be more precise. For example, it may be known that the objects under consideration are spherical lenses with an aberration that can be described with the coefficients of the first Zernike polynomials.

From the viewpoint of Shannon's information theory, uncertainty about a random variable is equivalent to information. The less is known about the outcome of a random experiment, say a coin flip, the more information is inferred. This is easily understood in the context of compression, where a stream of equally likely random characters cannot be compressed without loss of information. A natural language text, however, leads to different probabilities for different characters and is highly compressible without loss of information. Hence, from an information theoretical point of view $p_{\mathcal{G}}(\vec{g})$ determines the amount of information of the source or the uncertainty about the object (uncertainty in the general sense, not the measurement uncertainty defined in the GUM). This leads us to the differential entropy

$$\mathbb{H}(\mathcal{G}) = -\int p_{\mathcal{G}}(\vec{g})\, \log_2 p_{\mathcal{G}}(\vec{g})\, d\vec{g}\,, \tag{25}$$

as an important figure of merit. The logarithm to base 2 defines this quantity to be in unit bits (base e: nats, base 10: hartley, otherwise). While the entropy of discrete random variables is a measure of the quantitative average amount of information, such a direct interpretation is not available for the differential entropy of continuous random variables. A small differential entropy indicates the concentration in a small region in the space of the random variable, whereas a large differential entropy indicates that the variable is quite scattered. Hence, it is a useful figure of merit to qualitatively compare the information content of different objects. Further details on the fundamental concepts of information theory and it's measures can be found in [10] and details about the differential entropy are outlined in [11].

Now, we conceptualize the optical measurement system as a random mapping $\mathbb{M}$ of any random variable, the parameter, to any image of size $L_x \times L_y$. A set of L_c images is then to be interpreted as a set of random variables $\mathcal{I} = \{I_0, I_1, \ldots, I_{L_c-1}\} \sim p_{\mathcal{I}}(\vec{i})$ with image realizations $\vec{i} = [i_0, i_1, \ldots, i_{L_c-1}]$. This mapping will likely be highly non-linear and has to be considered random to take the different sources of uncertainty during the measurement process into account. That is, this mapping also includes measurement errors at different levels of the chain shown in Fig. 5, e. g. camera noise or vibrations. Statistically, it is possible to characterize this mapping by a conditional probability density function $p_{\mathcal{I}|\mathcal{G}}(\vec{i}|\vec{g})$ that describes the probability of a set of images given a set of parameters.

To finally answer the question how much information a set of measurements $\mathcal{I}$ contains about the parameters $\mathcal{G}$ of the object, we now introduce the mutual information $\mathbb{I}(\mathcal{G};\mathcal{I})$. The idea behind the mutual information is to characterize the information common in both random variables. Or in other words, the higher the mutual information, the more uncertainty about $\mathcal{G}$ is reduced by knowing the measurements $\mathcal{I}$ and vice versa, so

$$\mathbb{I}(\mathcal{G};\mathcal{I}) = \mathbb{H}(\mathcal{G}) - \mathbb{H}(\mathcal{G}|\mathcal{I}) = \mathbb{H}(\mathcal{I}) - \mathbb{H}(\mathcal{I}|\mathcal{G}), \tag{26}$$

where $\mathbb{I}$ and $\mathbb{H}$ are the mutual information and the conditional differential entropy, respectively. In contrast to the differential entropy discussed above, we now have a measure of the information about the object of interest that is contained in a set of measurements. A higher number of bits means more information about the object's statistical model is preserved in the measurements.

The conditional differential entropy $\mathbb{H}(\mathcal{G}|\mathcal{I})$ follows the same definition as in Eq. 25 and is defined by

$$\mathbb{H}(\mathcal{G}|\mathcal{I}) = -\int p_{\mathcal{I},\mathcal{G}}(\vec{i},\vec{g})\ \log_2 p_{\mathcal{I}|\mathcal{G}}(\vec{i}|\vec{g})\, d\vec{g}\, d\vec{i}\,, \qquad (27)$$

where $p_{\mathcal{I},\mathcal{G}}(\vec{i},\vec{g})$ is the joint distribution of the parameters $\mathcal{G}$ and images $\mathcal{I}$.[1]An intuitive interpretation of the second half of Eq. 26 is that the mutual information is determined by the differential entropy $\mathbb{H}(\mathcal{I})$ of the images reduced by the conditional differential entropy $\mathbb{H}(\mathcal{I}|\mathcal{G})$ in the images that is not included in the parameters (e. g. noise).

Note that the information theoretic description of the measurement process does not include any statement about the estimation of the parameters $\mathcal{G}$. A high mutual information shows that the measurements $\mathcal{I}$ are very informative about the parameters $\mathcal{G}$, but it does not tell how to exploit this information. The estimation task, which is solved by the block data processing as a part of the complete information processing chain shown in Fig. 5, needs to exploit the available measurement data. In the simplest but also typical case, we are interested in an estimate $\hat{\vec{g}}$ of the true value of the object parameters, which can be achieved by classical estimation methods such as minimum least squares estimation. Respective industry-relevant examples are object properties like the surface step height, a surface curvature or a surface roughness parameter. Additionally, we may be interested to derive values depending on $\hat{\vec{g}}$, e. g., a classification of objects into 'good' and 'faulty' parts that may require less information than a full estimate of $\vec{g}$.

With the perspective on the estimation tasks described above, which are in essence measurement tasks, from the viewpoint of information theory, it is possible to investigate and compare the potential of different measurement approaches regarding their fundamental information content at first. In a second future step, which is beyond the scope of this article, the extractable or accessible information content needs to be investigated. The aim is then to maximize the ratio of the extractable information content and the existing information content, i. e. to maximize the estimation efficiency or to determine a measurement uncertainty.

1 Overall, the mathematical expressions for the conditional differential entropy and the mutual information may be cumbersome and often cannot be solved analytically. Thus, approximations or numerical integration are often required for calculation.

3.2 Illustration of entropy and probabilistic object models

As an illustration, let us describe the application of the previous framework to the problem of measuring a step like in Fig. 4. The object is characterized by a single parameter, the step Δh. In the case that no a-priory knowledge about the object is known, but only the range of height variation, we can use a uniform distribution of $\Delta h \sim \mathcal{U}(h_{min}, h_{max})$. Note that while Δh is strictly denoted in meters, information theory is only concerned with the probability density function of Δh. Therefore we express every quantity in SI units without prefixes and ignore the unit in the following calculations which simplifies our notation. Then, the differential entropy is $\mathbb{H}(\Delta h) = \log_2(h_{max} - h_{min})$. For example, for $h_{max} - h_{min} = 50\,\mu\text{m}$ the differential entropy of the source is

$$\mathbb{H}(\Delta h) = \log_2(50 \cdot 10^{-6}) = -14.28\,\text{bits}\,, \qquad (28)$$

whereas for $h_{max} - h_{min} = \lambda/5$ with $\lambda = 500\,\text{nm}$ the differential entropy is

$$\mathbb{H}(\Delta h) = \log_2(100 \cdot 10^{-9}) = -23.25\,\text{bits}. \qquad (29)$$

As mentioned before the differential entropy cannot be interpreted as the information content of a random variable. However, a small differential entropy indicates the concentration in a small region in the space of the random variable, whereas a large differential entropy indicates that the variable is quite scattered [11]. This is reflected by the fact, that the result of Eq. 28 is larger than that of Eq. 29 due to the smaller dimensions of the second example. Again, this is in contrast to the mutual information $\mathbb{I}(\mathcal{G};\mathcal{I})$ that describes the common information in two random variables $\mathcal{G}$ and $\mathcal{I}$.

To continue the modeling of Δh as a random variable, we can also assume that some a-priory knowledge exist. For example, with a Gaussian distribution $N(\mu_h, \sigma_h)$ of known mean μ_h and standard deviation σ_h, i. e. $\Delta h \sim N(\mu_h, \sigma_h)$, the differential entropy is $\mathbb{H}(\Delta h) = \frac{1}{2}\log_2(2\pi e\,\sigma_h^2)$. To show that the volume in terms of the differential entropy of the Gaussian distribution can be matched to the uniform distribution used in Eq. 28, we calculate σ_h by the definition of the differential entropy. Then, for $\sigma_h = 12.16\,\mu\text{m}$ and independent of the mean, the differential entropy is also $\mathbb{H}(\Delta h) = -14.28$ bits.

Table 1: Comparison of the mutual information provided by different measurement schemes using a Michelson inter-ferometer (phase shift) and temporal sampling with monochromatic light. Simulation for $\lambda = 500$ nm, $\sigma_n = 0.1$ and a Normal distribution $\Delta h \sim N(0, \sigma_h)$ of the step height with $\sigma_h = 20$ or 80 nm.

Parameters		**Measurement scheme, monochromatic light** phase shift		**temporal sampling**		
phase- or τ-shift		$\phi = 0$	$\phi = \pi/2$	$\omega\tau = \pi/2$	$\omega\tau = \{-\pi/2, \pi/2\}$	$\omega\tau = \{0, \pi/2\}$
number of measurements		1	1	1	2	2
σ_h in nm	Results					
20	mutual Information in bits	0.721	2.167	2.167	2.774	2.486
80	mutual Information in bits	2.422	2.449	2.449	3.001	4.024

3.3 Temporal sampling

3.3.1 Information content for the case of monochromatic light

Here we consider the case of measurements using the Michelson-Interferometer setup described in Fig. 1 with monochromatic light and compare how informative this approach is with respect to the step height Δh of the measurement object. In this case, the intensity at the detector is given by Eq. 8. After trivial processing we can define an idealized measurement function by

$$\mathbb{M}(\Delta h) = \cos(2k\,\Delta h + \phi) + \epsilon\,, \tag{30}$$

where ϕ is a phase offset between the object and the reference wave and $\epsilon \sim N(0, \sigma_n)$ is the measurement error described as a mean-free normal distribution N. Taking a single measurement through $\mathbb{M}$ then gives a single one-dimensional measurement $\mathcal{I} = \{I_0\}$ determined by a probability density function given by the transformation of $p_{\Delta h}$ through $\mathbb{M}$ plus the Gaussian distributed error.

As follows from the discussion of Eq. 3, the phase shift ϕ given in Eq. 30 corresponds to a time shift $\tau = \phi/\omega$. Choosing a set of values $\tau_i = \phi_i/\omega$ with $i = 0 \ldots n-1$ therefore allows to scan along the τ-axis. In order to define a measurement process, we sample at the times $\tau_0, \ldots, \tau_{n-1}$ giving n measurements $\mathcal{I} = \{I_0, I_1, \ldots, I_{n-1}\}$ by

$$\mathbb{M}(\Delta h) = \{\cos(2k\,\Delta h - \omega\tau_0) + \epsilon_0, \cdots, \cos(2k\,\Delta h - \omega\tau_{n-1}) + \epsilon_{n-1}\}\,, \tag{31}$$

where τ_i are the different sampling points and $\epsilon_i \sim N(0, \sigma_n)$ is the independent measurement error of each sample. Note that this process is termed phase shifting in interferometry and allows to expand the unambiguity range of the measurement by a factor of 2 to $\lambda/2$ in comparison to a single measurement as described in Section 2.1. The transfer of information should therefore be increased by using multiple phase or τ-shifts as compared to a fixed value.

To highlight the limitations of monochromatic interferometry, firstly we analyze Eq. 30 for the case of a uniform distribution of Δh with $h_{max} - h_{min} = 50\,\mu$m and $\lambda = 500$ nm. The mutual information can be obtained by numerical calculation. More precisely we use 2^{20} random samples from Δh and the transformation $\mathbb{M}(\Delta h)$ to estimate the joint probability density function of the $(\Delta h, \mathbb{M}(\Delta h))$. The empirical joint probability (and its marginal) is used to compute numerically differential entropy and then the mutual information using Eq. 26. We find that it is almost negligible at $\mathbb{I}(\mathcal{G}; \mathcal{I}) = 0.145$ bits since there is a very large ambiguity in the estimation of Δh, as described by Eq. 9. We will see in Sec. 3.3.2 that this ambiguity can be solved by using polychromatic light. When Δh in confined to a much smaller range, the measurements provide more information. For example, when $h_{max} - h_{min} = 100$ nm the mutual information is $\mathbb{I}(\mathcal{G}; \mathcal{I}) = 4.181$ bits for $\sigma_n = 0.025$ and decreases to $\mathbb{I}(\mathcal{G}; \mathcal{I}) = 2.386$ bits when $\sigma_n = 0.1$ due to the increased noise.

Now we compare the mutual information provided by several measurement schemes with monochromatic light, firstly using a Michelson interferometer (phase shift) and secondly with temporal sampling. The results are compiled in Tab. 1.

First, we use a Normal distribution $\Delta h \sim N(0, \sigma_h)$ of the step height with $\sigma_h = 20$ nm. Further on, the width of the distribution of the measurement error as defined in Eq. 30 is $\sigma_n = 0.1$. Since $\mathbb{M}$ is normalized, σ_n is dimensionless. For these set of parameters, the main uncertainty therefore comes from the measurement errors, but not from the ambiguity associated with using monochromatic light. For the Michelson interferometer, the mutual information between the Δh and the measurements depends on ϕ and lies in the range of 0.721 bits for $\phi = 0$ and 2.167 bits for $\phi = \pi/2$. In case of the temporal sampling, one single sampling point should produce the same results than the Michelson-Interferometer. The simulations corresponding to $\omega\tau = \pi/2$ are 2.167 bits, which agrees with the expected results. Further on, the mutual information increases with two temporal samples. The actual gains depend of the sampling schema; in this particular experimental setting the best results are obtained for $\omega\tau = \{-\pi/2, \pi/2\}$ that produces 2.486 bits. This experiment illustrates that sampling of the τ-axis as in Eq. 31 allows to increase the mutual information.

Next, we repeat the previous experiment but increase the standard deviation of the step to $\sigma_h = 80\,\text{nm}$. The results confirm the previous observation regarding the improvement in the mutual information as the number of temporal samples increases. In this case, however, the improvement is larger (4.024 bits versus 2.449). It is interesting to note that the optimal sampling points are dependent on the problem: in the previous example the samplings $\omega\tau = \{-\frac{\pi}{2}, \frac{\pi}{2}\}$ were optimal, while in the current case $\omega\tau = \{0, \frac{\pi}{2}\}$ works better. The selection of a proper sampling schema thus plays an import role to increase the information gained from temporal sampling using monochromatic light.

3.3.2 Information content for the case of polychromatic light

Now we consider the case of measurements using the Γ-Profilometry setup described in Fig. 2 with polychromatic light. In this case, the measurement is a function of τ given by Eq. 18 and results in an idealized measurement function

$$\mathbb{M}(\Delta h) = \left\{\cos\left[\omega\left(\tau_i - \frac{2\Delta h}{c}\right)\right] \cdot \exp\left[-\frac{1}{2}\left(\tau_i - \frac{2\Delta h}{c}\right)^2 \Delta\omega^2\right] + \epsilon_i, \cdots\right\} \quad \text{for } i \in [0, \cdots, n-1], \tag{32}$$

where each of the single measurements corresponds to a different τ-sampling.

As shown in the previous section, the sampling scheme has an important impact in the achievable mutual information. To keep the discussion simple and easily comparable to WLI, we use a uniform τ-scan with 512 points, although more optimal sampling schemes are likely to exist. Further, we determine the value of τ_i that maximizes $\mathcal{R}(\Gamma(\Delta h - \frac{c\tau}{2}))$ and use this value as the result of the measurement. Such a procedure corresponds to the well known WLI. Note that by using this simple approach we may be limiting the full benefit of using the Γ-function.

As an example, we consider the case of a uniform distribution of Δh with $h_{max} - h_{min} = 50\,\mu\text{m}$ and $\lambda = 500\,\text{nm}$ and assume a coherence length of $l_c = 20\,\mu\text{m}$ for the polychromatic light. The results are compiled in Tab. 2. Recall from the previous Section 3.3.1 that the use of monochromatic light for this case provides a mutual information of just $\mathbb{I}(\mathcal{G}; \mathcal{I}_{\mathcal{K}}) = 0.145$ bits. For 512 repeated measurements one obtains an only slightly improved mutual information of 0.178 bits. However, when using polychromatic light, we observe that the ambiguity present in the monochromatic case disappears and the mutual information increases dramatically. For example, the mutual information $\mathbb{I}(\mathcal{G}; \mathcal{I}_{\mathcal{K}}) = 3.85$ bits for $\sigma_n = 0.1$ and increases to 4.78 bits for $\sigma_n = 0.025$. Temporal sampling using polychromatic light avoids measurement ambiguities encountered with monochromatic light and thus provides a significant increase in information content to recover the original step height.

Table 2: Comparison of the mutual information provided by different measurement schemes using temporal sampling with polychromatic light and monochromatic light. Example for $\Delta h \in [-25\,\mu\text{m}, 25\,\mu\text{m}]$, $\lambda = 500$ nm coherence length $l_c = 20\,\mu\text{m}$ and noise $\sigma_n = 0.1$.

Parameters and results	**Measurement scheme**			
width of distribution of measurement error σ_n	**monochromatic**		**polychromatic**	
	0.1	**0.1**	**0.1**	**0.025**
number of measurements	**1**	**512**	**512**	**512**
mutual Information in bits	0.145	0.178	3.85	4.78

3.4 Spatio-temporal sampling

In this section we show that from the perspective of information theory the correlation of two object points is superior to the correlation of the object with a reference wave, even in the simplistic case outlined there. For this purpose, we investigate the benefit of spatio-temporal sampling as is possible with the concept shown in Fig. 5 that allows us to sample any point of the spatio-temporal space of Fig. 3. This simple example points towards the large potential, which is hidden in the coherence function, waiting to be lifted.

In order to demonstrate the full potential of spatio-temporal sampling, we now extend the measurement to further steps on the sample as indicated by further points x_n in Fig. 4(a). For the following calculations we consider a surface profile consisting of R different steps. We assume that each individual step has a typical magnitude of Δh but presents some variations.

One example of this scenario would be the characterization after production of a diamond turned specular surface used for example in [2] as a reference. Each step is produced to have a given height, e. g. 50 μm. During production the height of the steps may, however, have some variations that need to be measured to ensure that they are within the limits of e. g. $\lambda/20$. We assume the use of polychromatic light as described in the previous Section 3.3.2 and compare the information provided by sampling τ compared to the Γ-sampling approach that we propose in this paper. Our goal is to determine if the use of spatio-temporal sampling (instead of just temporal sampling) can indeed improve the information transferred by the measurements.

Firstly we compare the number of measurements required by the two approaches from a qualitative point of view. In the case of WLI, the interference is obtained using a reference wave. Thus, a typical sampling approach would require to sample uniformly in the temporal domain, where the sampling range is proportional to the total height of the object, i.e, $R \cdot 50\,\mu\text{m}$. The sampling resolution determines the quality in the estimation of the envelope, and thus, the accuracy of the measurement. Note that the interference image obtained for each of the temporal sampling points provide a very reduced amount of information: the information is restricted to the single step of the object whose optical path is close to that of the reference wave. This limitation contrasts dramatically with the rich information provided by Γ-Profilometry. Here, by setting the shear parameter to one or more values suitable to obtain an interference image detailed information is provided for each of the individual steps. While the range of the temporal sampling is given by the variability of the step height instead of the total height of the object, the number of required shears provided by Γ-Profilometry depends on the number of steps R and their lateral distance. In the following, we will obviate this already substantial advantage and assume that the number of optical measurements required to obtain an estimation of a length is the same for WLI and Γ-Profilometry.

To model both methods we consider the case of four steps, $R = 4$, with a nominal step height of 50 μm, a standard deviation of the step height of 1 μm due to manufacturing tolerances. For the sake of comparability we assume a Gaussian distributed measurement error of $\sigma_m = 200$ nm in both cases, although WLI is currently further developed than Γ-Profilometry. The error σ_m results from an estimation of the difference of heights in two points using WLI or Γ-Profilometry with only sampling in τ-direction. Results provided for different measurements schemes correspond to one and two measurement batches. Each batch contains 512 measurements. Note that in most practical measurement setups for WLI, the fact that a reference wave travels a different path than the measurement wave causes larger errors than in the case of Γ-Profilometry that uses a common path principle with all waves having a similar path. To be conservative, we assume that the accuracy provided by WLI and Γ-Profilometry is equal.

The mutual information provided by several illustrative sampling schemes is reported in Tab. 3. The WLI approach that estimates the absolute location of the 4 steps provides an information of 5.783 bits while Γ-Profilometry using one shear provides a higher information content of 6.239 bits. Further on, if we allow two batches of measurements, the WLI measures 2 times each one of the absolute positions providing an information of 6.838 bits, while Γ-Profilometry achieves 7.325 bits for the same number of measurements. Here it becomes apparent that the possibility of combining temporal and spatial sampling has indeed the potential of providing better measurements. It is interesting to note that including the results for one and two shears provides more information than repeating the measurement twice using one shear. The flexibility provided by combining the temporal and spatial sampling is thus essential.

Table 3: Comparison of the mutual information provided by White Light Interferometry (WLI) and Γ-sampling when measuring a 4-step object with steps of 50 μm height, a standard deviation of the step height of 1 μm due to manufacturing tolerances and a measurement error of $\sigma_m = 200$ nm. Results provided for different measurement schemes correspond to one and two measurement batches. Each batch contains 512 measurements.

Parameters and results	**Measurement scheme, polychromatic light**				
	WLI		**Γ-sampling**		
number of τ-scans	**1**	**2**	**1**	**2**	**1**
number of $\vec{s}$-scans	**–**	**–**	**1**	**1**	**2**
number of measurement batches	**1**	**2**	**1**	**2**	**2**
mutual information in bits	5.783	6.838	6.239	7.267	7.325

4 Methods for efficient data collection and evaluation

This section illustrates how novel techniques such as Compressed Sensing (CS), Machine Learning (ML) and local pre-processing using Vision Systems on Chips (VSoCs) can be used to efficiently address the concept of Γ-Profilometry starting with the imaging hardware followed by the optical measurement scheme to the estimation of arbitrary object parameters.

4.1 Efficient sampling

The Γ-function offers a promising way to collect a maximum amount of information about the measured object. However, it is still unclear how the sampling of the τ-s space should be optimized for a given geometry or object structure. In the presented example a clear maximum of information has been pointed out, but in general this is unknown. Recent literature shows that sampling in the classical sense (i. e. Whittaker-Shannon-Kotelnikov sampling) is often wasteful and unnecessary, if knowledge about the structure of the problem is available. For example, the spatio-temporal Fourier transform of $\Gamma(\tau, \vec{s})$ leads to $\tilde{\Gamma}(\omega, \vec{k})$, which is sparse in many situations.[2] The spectrum can often be approximated by only a few frequencies ω_n while the structure of natural surfaces often lead to sparse Fourier representations in $\vec{k}$-space. Compressed Sensing (CS) [12] is a prominent sampling technique that explicitly exploits sparsity in a known basis to reduce the number of samples by randomized sampling. CS seems a promising method particular in this case, because the realization of random sampling strategies in the optical domain can be realized through a random selection of shears $\vec{s}$ and delays τ according to the measurement setup described in Fig. 2.

Additionally, adaptive measurement systems may be required to automatically minimize the number of samples through intelligent search in the τ-s space. In simple cases, this is akin to a gradient search maximizing the information content. While seemingly simple, this approach requires a valid estimation of said information content with a very limited amount of samples. For complex objects Machine Learning (ML) seems like a promising approach to adapt the sampling, but also depends on the availability of data. Machine learning with small data samples is a hot topic in current research [13].

4.2 Estimation methods

Casting the measurement problem from a probabilistic viewpoint is useful beyond the application of information theory. The field of signal processing offers a wide variety of estimation methods that are based on the Bayesian point of view with varying degrees of knowledge about the structure of the estimation problem in the form of statistical priors (e. g. a Gaussian distribution of the step size due to production processes). A very flexible way of casting problems is variational inference, where knowledge about an object can be formulated by families of probability functions and several hyperparameters that are learned along the estimation [14]. That way knowledge like it is assumed in our article (unitary or Gaussian distribution of Δh) can be further relaxed and adapted to the measurement problem at hand.

Furthermore, modern machine learning methods are often targeting the maximization of information theoretical measures to optimally tune deep neural networks for classification and estimation problems. Connections to information theory and various approaches from signal theory are still under heavy investigation, but the current state of the art shows that the inner workings of deep neural networks can be captured by information theoretic concepts [15].

4.3 Efficient hardware using vision systems on chips

In conventional systems, image acquisition and processing are not only logically but also physically separated. This separation, that limits the speed [16] and possibly also the accuracy of the measurement, represents a major bottleneck for a metrology system using the Γ-Function. A holistic approach that integrates optical detection and information processing is required. On the one hand, there is a notable increase in the volume of raw data since the Γ-Function has seven orthogonal dimensions that can be sampled; on the other hand, there is an increase in the processing demands because of the more complex decoding of the information intrinsic in the interpretation of the Γ-Function. Even worse, those requirements of data processing increase even further when on-line adaptive sampling methods are required, as the ones previously discussed. This significant increase in requirements demands the use of dedicated hardware architectures.

A promising alternative are Vision-Systems-on-Chip (VSoCs) [17]. They allow the integration of pre-processing close to the image detectors and enable a processing time several orders of magnitude faster than conventional techniques [16]. Furthermore, the integrated dedicated on-chip hardware accelerators can provide significant improvements in performance. In particular, the data processing required for the Γ-Function exhibits intrinsic large parallelism that can be efficiently exploited by dedicated massively parallel hardware architectures, a currently done for image processing [16]. Furthermore, since the accuracy

2 As $\tilde{\Gamma}$ depends on temporal and spatial frequency, it should not be confused with the spectral density of Eq. 11 that depends on temporal frequency and spatial coordinates.

requirements in the different processing stages are non-uniform, the dynamic reconfiguration of computational accuracy offers significant potential to adapt to changing requirements.

5 Conclusions and outlook

We have presented here a first comparative investigation of the information content of interferometrical measurements for a simplified case of shape measurements using optical metrology. The combined view of optical metrology and information theory given in this paper demonstrates that spatio-temporal sampling based on the Γ coherence function enables higher information gain as compared to wave field reconstruction solely using temporal or spatial reconstruction by techniques such as classical Michelson or White Light Interferometry (WLI) or Computational Shear Interferometry (CoSI).

In order to efficiently sample the high dimensional parameter space of the Γ-function, efficient sampling strategies are required. Compressed sensing (CS) using random sampling and gradient search methods may be used to obtain an efficient path to an optimum solution while machine learning (ML) based on small data sets and prior knowledge may be used to identify solutions in accord to prior knowledge about the structure of the object under investigation. The details of such procedures are a central subject of further work.

While the presented work is focused on shape measurements, the fundamental concept has a promising perspective to be applied also to surface roughness measurements [18]. Optical roughness measurements by means of auto- or cross-correlating the scattered light images from a coherent illumination are well known [19, 20], but only a single roughness parameter is typically measurable such as the root mean square S_q of the surface height distribution. The coherence function provides the appropriate starting point to design a new or to rethink existing measurement principles to maximize the accessible information on the surface roughness with coherent light scattering.

It is further noted that the structure of the coherence function which is basically a correlation, already appears in several state-of-the-art measurement techniques. For instance, a cross-correlation of the intensity values from image pairs is used in Particle Image Velocimetry (PIV) [21] to determine the motion of particles in flows, in Digital Image Correlation (DIC) [22] as well as in Digital Speckle Photography (DSP) [23] to measure surface movements and deformations. The reader interested in an information theoretic perspective on these principles is referred to [24–26].

As a consequence of the considerations given here, applying the Γ-function has the potential to significantly boost the accessible information with respect to a great variety of measurement tasks in the area of optical metrology.

Funding: Gamma-Profilometry, Grant No. 265388903, at BIAS and SFB/TRR 136 project C06 at BIMAQ both funded by the Deutsche Forschungsgemeinschaft (DFG).

References

1. J. W. Goodman, *Statistical Optics* (John Wiley & Sons, Inc., Hoboken, USA, 2000). p. 63.
2. C. Falldorf, M. Agour, A. F. Müller, and R. B. Bergmann, “Gamma-Profilometry: a new paradigm for precise optical metrology,” Opt. Exp. 29(22), 36100–36110 (2021).
3. R. Bergmann, M. Kalms, and C. Falldorf, “Optical In-Process Measurement: Concepts for Precise, Fast and Robust Optical Metrology for Complex Measurement Situations,” Appl. Sci. 11, 10533 (2021). https://doi.org/10.3390/app112210533.
4. A. F. Müller, C. Falldorf, M. Lotzgeselle, G. Ehret, and R. B. Bergmann, “Multiple Aperture Shear-Interferometry (MArS): a solution to the aperture problem for the form measurement of aspheric surfaces,” Opt. Exp. 28(23), 34677–34691 (2020).
5. C. Falldorf, M. Agour, and R. B. Bergmann, “Digital holography and quantitative phase contrast imaging using computational shear interferometry,” Opt. Engin. 54, 024110 (2015).
6. C. Falldorf, J.-H. Hagemann, G. Ehret, and R. B. Bergmann, “Sparse light fields in coherent optical metrology,” Appl. Opt. 56, F14–F19 (2017).
7. D. Malacara, *Interferogram Analysis For Optical Testing* (CRC Press, Hoboken, NJ, 2005).
8. C. Krause, R. B. Bergmann, and C. Falldorf, “Statistical analysis of phase values for the determination of step heights in multi-wavelength interferometry,” TM-Technisches Messen 89(6), 430–437 (2022). https://doi.org/10.1515/teme-2021-0139.
9. B. Saleh and M. C. Teich, *Fundamentals of Photonics* (Wiley, 2007).
10. T. M. Cover and J. A. Thomas, *Entropy, Relative Entropy, and Mutual Information*, chap. 2, pp. 13–55 (John Wiley & Sons, Ltd, 2005). https://onlinelibrary.wiley.com/doi/pdf/10.1002/047174882X.ch2, URL https://onlinelibrary.wiley.com/doi/abs/10.1002/047174882X.ch2.
11. T. M. Cover and J. A. Thomas, *Differential Entropy*, chap. 8, pp. 243–259 (John Wiley & Sons, Ltd, 2005). https://onlinelibrary.wiley.com/doi/pdf/10.1002/047174882X.ch8, URL https://onlinelibrary.wiley.com/doi/abs/10.1002/047174882X.ch8.
12. Y. C. Eldar and G. Kutyniok, eds., *Compressed Sensing - Theory and Applications* (Cambridge, 2012).
13. Y. Wang, Q. Yao, J. T. Kwok, and L. M. Ni, “Generalizing from a Few Examples: A Survey on Few-Shot Learning,” ACM Comput. Surv. 53(3) (2020). URL https://doi.org/10.1145/3386252.

14. C. Bishop, *Pattern Recognition and Machine Learning* (Springer, 2006). URL https://www.microsoft.com/en-us/research/publication/pattern-recognition-machine-learning/.
15. R. Shwartz-Ziv and N. Tishby, “Opening the Black Box of Deep Neural Networks via Information,” CoRR abs/1703.00810 (2017). 1703.00810, URL http://arxiv.org/abs/1703.00810.
16. M. Lefebvre, L. Moreau, R. Dekimpe, and D. Bol, “7.7 A 0.2-to-3.6TOPS/W Programmable Convolutional Imager SoC with In-Sensor Current-Domain Ternary-Weighted MAC Operations for Feature Extraction and Region-of-Interest Detection”, in *2021 IEEE Internat. Solid- State Circuits Conference (ISSCC), vol. 64*, pp. 118–120 (2021).
17. L. Alba, R. Domínguez Castro, F. Jiménez-Garrido, S. Espejo, S. Morillas, J. Listán, C. Utrera, A. García-Ortiz, M. D. Pardo, R. Romay, C. Mendoza, A. Jiménez, and Á. Rodríguez-Vázquez, “New Visual Sensors and Processors”, in *Spatial Temporal Patterns for Action-Oriented Perception in Roving Robots*, P. Arena and L. Patanè, eds., pp. 351–369 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2009). URL https://doi.org/10.1007/978-3-540-88464-4_8.
18. M. Kalms and R. B. Bergmann, “Structure function analysis of powder beds in additive manufacturing by laser beam melting,” Addit. Manuf. 36, 101396 (2020). https://Doi.org/10.1016/j.addma.2020.101396.
19. E. C. Teague, T. V. Vorburger, D. Maystre, and R. D. Young, “Light Scattering from Manufactured Surfaces,” CIRP Annals 30(2), 563–569 (1981).
20. A. Fischer and D. Stöbener, “In-process roughness quality inspection for metal sheet rolling,” CIRP Annals 68, 523–526 (2019).
21. J. Westerweel, “Fundamentals of digital particle image velocimetry,” Meas. Sci. Technology 8, 1379 (13 pp.) (1997).
22. Y. Barranger, P. Doumalin, J. C. Dupré, and A. Germaneau, “Strain measurement by digital image correlation: influence of two types of speckle patterns made from rigid or deformable marks,” strain 48, 357–365 (2012).
23. A. Tausendfreund, D. Stöbener, and A. Fischer, “In-process measurement of three-dimensional deformations based on speckle photography,” Applied Sciences 11, 4981 (11 pp.) (2021).
24. A. Fischer, “Limiting uncertainty relations in laser-based measurements of position and velocity due to quantum shot noise,” Entropy 21, 264 (19 pp.) (2019).
25. A. Fischer, “Fundamental flow measurement capabilities of optical Doppler and time-of-flight principles,” Experiments in Fluids 62(2), 37 (19 pp.) (2021).
26. A. Fischer, “Fundamental uncertainty limit for speckle displacement measurements,” Appl. Opt. 56, 7013–7019 (2017).

Bionotes

Ralf B. Bergmann
Bremer Institut für angewandte Strahltechnik GmbH (BIAS), 28359 Bremen, Klagenfurter Str. 5, Germany
bergmann@bias.de

Ralf B. Bergmann studied physics in Heidelberg and Freiburg, received his doctorate with his work at the Max Planck Institute for Solid State Research (MPI-FKF) from the University of Stuttgart, worked as a postdoc at of the University of New South Wales and habilitated at the University of Freiburg. After leading a research group at the University of Stuttgart he headed the department of Applied Physics at the Central Research and Advance Engineering facility of the Robert Bosch GmbH and later the Physical Analyses Laboratory in the Automotive Electronics division. Since 2008 he is a professor at the University of Bremen in the Faculty of Physics and Electrical Engineering and head of the Bremen Institute for Applied Beam Technology (BIAS) with the field of “Optical Metrology and Optoelectronic Systems”.
Foto: Marcus Windus / BIAS

Andreas Fischer
Universität Bremen, Bremer Institut für Messtechnik, Automatisierung und Qualitätswissenschaft (BIMAQ), 28359 Bremen, Linzer Str. 13, Germany
andreas.fischer@bimaq.de

Andreas Fischer studied electrical engineering, completed his PhD at the Technische Universität Dresden in 2009 and his habilitation in 2014. Since 2016, he is a professor at the University of Bremen in the Faculty of Production Engineering and head of the Bremen Institute for Measurement, Automation and Quality Science (BIMAQ). His research areas cover optical measurement principles for flow and production processes, in-process applications of model-based measurement systems, and the investigation of fundamental limits of measurability.

Carsten Bockelmann
Universität Bremen, Arbeitsbereich Nachrichtentechnik (ANT), 28359 Bremen, Otto-Hahn-Allee 1, Germany
bockelmann@ant.uni-bremen.de

Carsten Bockelmann received the Dipl.-Ing. and Ph. D. degrees in electrical engineering from the University of Bremen, Germany, in 2006 and 2012, respectively. Since 2012, he has been a Senior Research Group Leader at the University of Bremen within the Faculty of Physics and Electrical Engineering coordinating research activities regarding the application of compressive sensing and machine learning to communication problems. His research interests include massive machine-type communication, ultra-reliable low latency communications and industry 4.0, compressive sensing and channel coding.

Armin Dekorsy
Universität Bremen, Arbeitsbereich Nachrichtentechnik (ANT), 28359 Bremen, Otto-Hahn-Allee 1, Germany
dekorsy@ant.uni-bremen.de

Armin Dekorsy has more than ten years of industrial experience in leading research positions, such as an DMTS at Bell Labs Europe and the Head of Research Europe Qualcomm, Nuremberg. Since 2010 he is Professor for Communications Engineering and Head of the Department of Communications Engineering at the University of Bremen in the Faculty of Physics and Electrical Engineering. His current research focuses on distributed signal processing, compressed sampling, information bottleneck method, and machine learning leading to further development of communication technologies for 5G/6G, industrial wireless communications, and NewSpace satellite communications. He investigates wireless communication and signal processing for the baseband of transceivers of future communication systems, the results of which are transferred to the predevelopment of industry through political and strategic activities.

Alberto Garcia-Ortiz
Universität Bremen, Integrated Digital Systems (IDS), 28359 Bremen, Germany
agarcia@ids.uni-bremen.de

Alberto Garcia-Ortiz obtained the diploma degree in Telecommunication Systems from the Polytechnic University of Valencia (Spain) in 1998. After working for two years at Newlogic in Austria, he started the Ph. D. at the Institute of Microelectronic Systems, Darmstadt University of Technology, Germany and received his PhD in 2003. From 2003 to 2005, he worked as a Senior Hardware Design Engineer at IBM Deutschland Development and Research in Böblingen. He then joined the start-up AnaFocus (Spain), where he was responsible for the design and integration of AnaFocus' next generation Vision Systems-on-Chip. Since 2010 he is a full professor for the chair of integrated digital systems at the University of Bremen in the Faculty of Physics and Electrical Engineering. His interests include low-power design and estimation, communication-centric design, SoC integration, and hardware accelerators for machine learning.

Claas Falldorf
Bremer Institut für angewandte Strahltechnik GmbH (BIAS), 28359 Bremen, Klagenfurter Str. 5, Germany
falldorf@bias.de

Claas Falldorf studied physics at the University of Bremen, where he received his doctorate at the Faculty of Physics and Electrical Engineering in 2009. Since then he heads the group "Coherent Optics and Nano-Photonics" at BIAS – Bremen Institute of Applied Beam Technology. His research focusses on optical metrology, coherence theory, signal processing and optimization theory.
Foto: Marcus Windus / BIAS