Hindawi
International Journal of Spectroscopy
Volume 2017, Article ID 9265084, 29 pages
https://doi.org/10.1155/2017/9265084

Research Article

Fourier Spectroscopy: A Bayesian Way

Stefan Schmuck[1] **and Jakob Svensson**[2]

[1] *CCFE, Culham Science Centre, Abingdon OX14 3DB, UK*
[2] *Max-Planck-Institut für Plasmaphysik, Teilinstitut Greifswald, Wendelsteinstraße 1, 17491 Greifswald, Germany*

Correspondence should be addressed to Stefan Schmuck; schmuck@ifp.cnr.it

Received 6 February 2017; Revised 11 July 2017; Accepted 27 July 2017; Published 31 October 2017

Academic Editor: Wei Kong

The concepts of standard analysis techniques applied in the field of Fourier spectroscopy treat fundamental aspects insufficiently. For example, the spectra to be inferred are influenced by the noise contribution to the interferometric data, by nonprobed spatial domains which are linked to Fourier coefficients above a certain order, by the spectral limits which are in general not given by the Nyquist assumptions, and by additional parameters of the problem at hand like the zero-path difference. To consider these fundamentals, a probabilistic approach based on Bayes' theorem is introduced which exploits multivariate normal distributions. For the example application, we model the spectra by the Gaussian process of a Brownian bridge stated by a prior covariance. The spectra themselves are represented by a number of parameters which map linearly to the data domain. The posterior for these linear parameters is analytically obtained, and the marginalisation over these parameters is trivial. This allows the straightforward investigation of the posterior for the involved nonlinear parameters, like the zero-path difference location and the spectral limits, and hyperparameters, like the scaling of the Gaussian process. With respect to the linear problem, this can be interpreted as an implementation of Ockham's razor principle.

1. Introduction

Fourier spectroscopy is a diagnostic application which reveals information about spectral quantities like refractive index, absorption, and transmission of a medium under test. In addition, the characterisation in absolute terms is possible for broadband spectra, for example, emitted by electrons of a high-temperature plasma, being magnetically confined [1].

Commonly, an interferometer diagnostic, let us say of Michelson [2] or Martin-Puplett [3] design type, probes the Fourier transform of a spectral quantity. The corresponding interferometric data is a discrete set with finite length and includes noise contributions. Standard Fourier data analysis techniques [4–6] have been developed. These techniques lack describing and capturing properly several fundamental aspects, like the noisy nature of measured data and possible spectral limits and their impact on the spectral quantity to be inferred.

One misconception, arising from the standard formulation, is that certain spectral information must be lost inherently, because only a finite amount of data is acquired. This is proven in standard literature by evaluating the convolution function which has a finite full width at half maximum (FWHM), implying that only a finite amount of Fourier coefficients is accessible via measurements. While this conclusion remains valid when a continuous spectrum is probed, the reasoning does not hold in general for a discrete spectrum. This fact was exploited to develop a (self-) deconvolution procedure, so that some discrete lines which were separated by less than the FWHM width of the convolution function have been inferred [7].

Opposed to the standard data analysis techniques, a probabilistic ansatz was introduced to estimate model parameters, like amplitude spectra and frequencies, in the field of Fourier spectroscopy [8, 9]. For a spectroscopic problem, a model is formulated via Bayes' theorem which allows to state prior information about model parameters. Furthermore, a Gaussian likelihood connects functionally the parameters with the noisy interferometric data. Then, after having measured a noisy data set and framing the reality by a certain model, the knowledge about model parameters is expressed in probabilistic terms by the posterior. This approach [8, 9] demonstrated that the uncertainty on the posterior mean of a frequency to be estimated for a single-frequency problem can be orders of magnitude below the FWHM width of the equivalent convolution function. On top of that, a criterion has been derived how far frequencies need to be separated, so that the Bayesian approach is still able to make a distinction. This separation can be well below the FWHM width of the convolution function. These findings are a direct consequence of the use of probabilistic theory.

One of the advantages of the Bayesian approach is that different models and, hence, their assumptions can be compared with each other also known as Ockham's razor. This enables the identification of the best, that is, the most likely, model, complying with the data. Given this context, the (self-)deconvolution procedure mentioned above is interpreted here as an optimisation to find a minimising set of discrete frequencies to describe the data sufficiently. In general, the most fundamental issue is whether the spectrum to be inferred is discrete or continuous. If a discrete spectrum is more likely or follows by a physics model, how many discrete frequencies are involved and what are their estimates including uncertainties? Each frequency is associated with an amplitude and a phase which need to be estimated as well. If a continuous spectrum is at hand, then the spectral limits are of interest. In addition, in case, the underlying physics process is understood, what is the uncertainty on the spectrum and the phase following from experimentally inaccessible regions in the data domain.

The investigation of the fundamental issues listed above is quite challenging from numerical point of view. However, the computational effort is largely reduced when even and odd amplitudes are used instead of the phase and amplitude. This ensures a linear dependence between the even and odd amplitude parameters and the data. Formulating the prior information about all even and odd amplitudes as a multivariate normal with a specific prior mean and covariance gives straightforwardly the posterior mean and covariance. Furthermore, the marginalisation can be carried out analytically for these linear parameters. The remaining posterior quantity carries information about the nonlinear model parameters like frequencies or spectral limits and so-called hyperparameters, entering merely in the prior.

In Section 2, the basic equations for Fourier spectroscopy and their implications are generally investigated, and the main concepts of the standard analysis and their drawbacks are pointed out. Section 3 of this paper presents a Bayesian formalism, so that it lends itself to applications in the field of Fourier spectroscopy for Gaussian (white) noise. The fundamental information about the spectral quantity, that it must vanish at a lower and upper spectral limit, can be stated by the covariance function of a Brownian bridge process as described in Section 4. This covariance is used as prior in the example application of the Bayesian approach presented in Section 5 to infer continuous but band-limited spectral quantities given an actually measured interferometric data set. Thereby, some diagnostic imperfections like a drifting signal offset, the zero-path difference, and a nonuniform spatial sampling are also taken into account. Section 6 discusses the strategy for a plausibility study of models, using different priors for the spectral quantities, and attempts to compare results and computational efforts obtained with the Bayesian model and with a standard model. The last section presents the conclusions.

2. Fourier Spectroscopy

2.1. General Definitions. Commonly, the complementary coordinates used for Fourier transformations are the frequency f and time t, or the wavenumber σ and a spatial coordinate x. For the moment, the latter pair is used to state the basic operations of Fourier transformation. Afterwards, the wavenumber is replaced via $\sigma = f/c$ for convenience.

The real-valued continuous functions $V(x)$ and $S_\sigma(\sigma)$ form a Fourier transformation pair stated by

$$V(x) = \int_{-\infty}^{\infty} S_\sigma [\cos(2\pi\sigma x) + \sin(2\pi\sigma x)]\, d\sigma. \quad (1)$$

The inverse operation reads

$$S_{\sigma'} = \int_{-\infty}^{\infty} V \left[\cos\left(2\pi\sigma' x\right) + \sin\left(2\pi\sigma' x\right)\right] dx. \quad (2)$$

Note that so far $\sigma' \neq \sigma$ holds. To find a relation, one inserts (1) in the above expression. After applying trigonometric identities, the spatial integral becomes

$$\begin{aligned}&\int_{-\infty}^{\infty} \left[\cos\left(2\pi\left(\sigma - \sigma'\right)x\right) + \sin\left(2\pi\left(\sigma + \sigma'\right)x\right)\right] dx \\ &\quad = \delta\left(\sigma - \sigma'\right),\end{aligned} \quad (3)$$

because the sinusoidal contribution vanishes. Hence,

$$\begin{aligned}S_{\sigma'} &= \int_{-\infty}^{\infty} S_\sigma \delta\left(\sigma - \sigma'\right) d\sigma, \\ S_{\sigma'} &= S_\sigma\end{aligned} \quad (4)$$

follow, because the two wavenumber coordinates equal each other. In fact, the delta distribution occurs, because S_σ is a distribution.

Replacing σ with f/c gives in the spectral domain the quantity $S(f) = S_\sigma/c$. This gives

$$V = \int_{-\infty}^{\infty} S \left[\cos\left(2\pi\frac{f}{c}x\right) + \sin\left(2\pi\frac{f}{c}x\right)\right] df \quad (5)$$

and rescales the inverse operation like

$$S = \frac{1}{c}\int_{-\infty}^{\infty} V(x) \left[\cos\left(2\pi\frac{f}{c}x\right) + \sin\left(2\pi\frac{f}{c}x\right)\right] dx \quad (6)$$

to match the units $[S] = [V]/\mathrm{Hz}$.

2.2. Selection of Representation. The spectral domain over which the integration is performed in (5) includes the whole negative and the positive ranges. While the cosine transform acts on the even part $S_E = [S(f) + S(-f)]/2$ of S, the sine transform is linked only to the odd part $S_O = [S(f) - S(-f)]/2$. Thus, an alternative formulation reads

$$V = 2\int_0^\infty \left[S_E \cos\left(2\pi\frac{f}{c}x\right) + S_O \sin\left(2\pi\frac{f}{c}x\right)\right] df, \tag{7}$$

and it becomes clear that the even and the odd parts of V and S are connected.

Another representation uses the amplitude $|S|$ and the phase α by setting

$$\begin{aligned} |S| &= 2\sqrt{S_E^2 + S_O^2}, \\ \tan(\alpha) &= -\frac{S_O}{S_E} \end{aligned} \tag{8}$$

which gives

$$V = \int_0^\infty |S| \cos\left(2\pi\frac{f}{c}x + \alpha\right) df. \tag{9}$$

Because the model representations (5) and (7) are linear in S, S_E, and S_O, both are favoured over the formulation (9), when linear inversion techniques are to be applied. Of both favoured representations, the one, using even and odd parts, has preferred properties. Since the orders of magnitude of the amplitudes can be quite different for the even and odd functions, a separation is logical.

2.3. Finite Bandwidth. If the spectral domain is band-limited, such that S_E and S_O are finite for the range from f_L and f_U with the bandwidth $\Delta_F = f_U - f_L$ and the centre frequency $f_C = (f_L + f_U)/2$, (7) becomes

$$\begin{aligned} &V(x) \\ &= 2\int_{f_L}^{f_U} \left[S_E \cos\left(2\pi\frac{f}{c}x\right) + S_O \sin\left(2\pi\frac{f}{c}x\right)\right] df. \end{aligned} \tag{10}$$

Here, the assumption must be mentioned that both functions have the same spectral limits. This is assumed in the following but not mandatory. Furthermore, for any combination of f_C and Δ_F the relation $\Delta_F \leq 2f_C$ must be fulfilled to be meaningful.

2.3.1. Relation: Fourier Transform-Fourier Coefficients. When the bandwidth is finite, one can express the spectral functions by Fourier coefficients multiplied each with the associated sinusoidal basis function of order k ($\in N$). These coefficients are defined here by the integrals

$$\begin{aligned} &\begin{pmatrix} A_{E(O),0} \\ A_{E(O),k} \\ B_{E(O),k} \end{pmatrix} \\ &= \frac{1}{\Delta_F}\int_{f_L}^{f_U} S_{E(O)} \begin{pmatrix} 1 \\ 2\cos\left(2\pi k\dfrac{f-f_C}{\Delta_F}\right) \\ 2\sin\left(2\pi k\dfrac{f-f_C}{\Delta_F}\right) \end{pmatrix} df \end{aligned} \tag{11}$$

which carry the unit as S_E and S_O. The coefficients $A_{E(O),0}$ label the mean values of S_E and S_O in the spectral domain covered. Then, one can replace the even and odd functions with

$$\begin{aligned} S_{E(O)} = A_{E(O),0} &+ \sum_{k=1}^{\infty} A_{E(O),k} \cos\left(2\pi k\frac{f-f_C}{\Delta_F}\right) \\ &+ \sum_{k=1}^{\infty} B_{E(O),k} \sin\left(2\pi k\frac{f-f_C}{\Delta_F}\right) \end{aligned} \tag{12}$$

which allows performing the spectral integration in (10) analytically. Since the result

$$\begin{aligned} V(x) = 2\Delta_F A_{E,0} &\cos\left(2\pi\frac{f_C}{c}x\right) \operatorname{sinc}\left(\pi\frac{\Delta_F}{c}x\right) \\ &+ 2\Delta_F A_{O,0} \sin\left(2\pi\frac{f_C}{c}x\right) \operatorname{sinc}\left(\pi\frac{\Delta_F}{c}x\right) \\ &+ \Delta_F \sum_{k=1}^{\infty} A_{E,k} \cos\left(2\pi\frac{f_C}{c}x\right) \left[\operatorname{sinc}_{k-} + \operatorname{sinc}_{k+}\right] \\ &- \Delta_F \sum_{k=1}^{\infty} B_{E,k} \sin\left(2\pi\frac{f_C}{c}x\right) \left[\operatorname{sinc}_{k-} - \operatorname{sinc}_{k+}\right] \\ &+ \Delta_F \sum_{k=1}^{\infty} A_{O,k} \sin\left(2\pi\frac{f_C}{c}x\right) \left[\operatorname{sinc}_{k-} + \operatorname{sinc}_{k+}\right] \\ &+ \Delta_F \sum_{k=1}^{\infty} B_{O,k} \cos\left(2\pi\frac{f_C}{c}x\right) \left[\operatorname{sinc}_{k-} - \operatorname{sinc}_{k+}\right] \end{aligned} \tag{13}$$

with

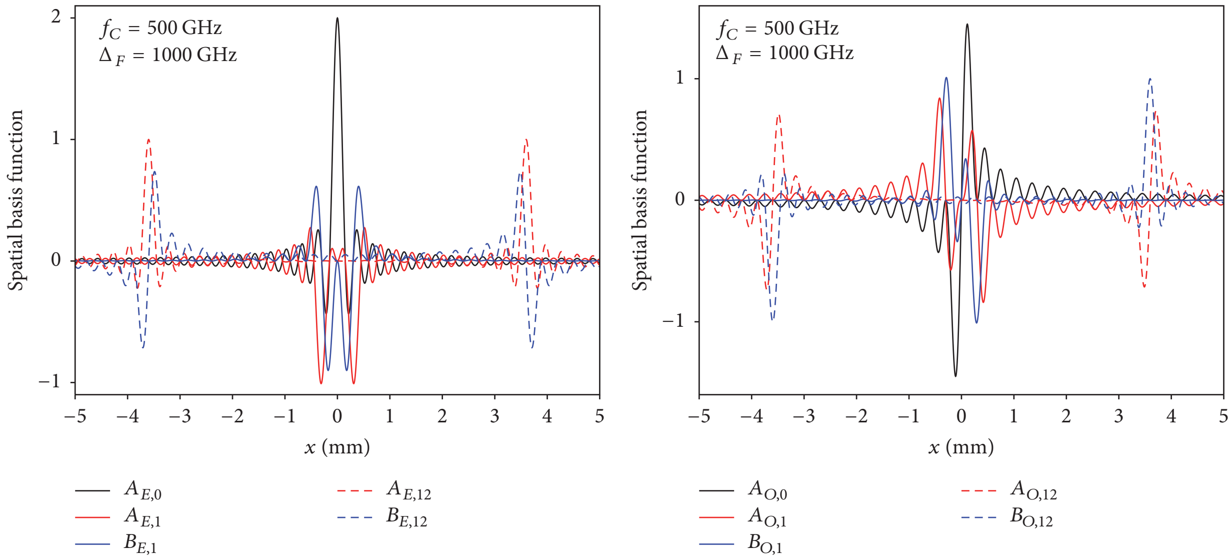
$$\operatorname{sinc}_{k\pm} = \operatorname{sinc}\left[\pi\left(k \pm \frac{x}{c}\Delta_F\right)\right]. \tag{14}$$

follows, it becomes clear that the Fourier transform of a band-limited function can be expressed by Fourier coefficients scaled with the bandwidth and multiplied each with the associated continuous basis function in the spatial domain. These basis functions have two contributions. The first is a sum/difference of two $\operatorname{sinc}_{k\pm}$ functions which depend on the order k, the spatial coordinate, and the bandwidth. The latter quantity determines the spatial width of $\operatorname{sinc}_{\pm}$. Furthermore, the localisation is permitted at $x_k = \pm kc/\Delta_F$, where a coefficient for a given k mainly acts. Hence, increasing the order implies the localisation at a larger distance from the spatial origin. This explains the occurrence of factor 2 for $A_{E(O),0}$ for which both sinc functions coincide.

The second contribution causes a modulation of $\operatorname{sinc}_{k\pm}$ and is given by a sine/cosine with the centre frequency and spatial coordinate in the argument. This dependency makes the basis function for $A_{O,0}$ vanish at the spatial origin.

With respect to the spatial origin, the transformed basis functions of the coefficients for S_E and S_O are symmetric and antisymmetric, respectively.

Some basis functions in the spatial domain are shown in Figures 1(a) and 1(b) for $f_C = 500$ GHz and $\Delta_F = 1000$ GHz.



(a) (b)

Figure 1: Basis functions in spatial domain of Fourier coefficients (of order $k = 0, 1$, and 12) $A_{E,k}$, $B_{E,k}$ for even spectral function (a) and $A_{O,k}$, $B_{O,k}$ for odd spectral function (b). The spectral functions are finite for the interval from 0 GHz to 1000 GHz for which the centre frequency $f_C = 500$ GHz and bandwidth $\Delta_F = 1000$ GHz follow. The basis functions in the spatial domain are given by a sum or difference of two sinc functions, depending each on the optical path difference x, Δ_F, and k. Furthermore, the sinc functions are modulated by sine or cosine with f_C and x in the argument. The sinc functions for fixed k become unity at $x_k = \pm kc/\Delta_F$, obviously increasing in k.

2.3.2. Embedding into Larger Spectral Domain. The functions S_E and S_O may be finite in the spectral domain with limits f_L and f_U or centre frequency f_C and bandwidth Δ_F. Embedding this domain in a larger one with limits $f'_L < f_L$ and $f'_U > f_U$ ($\Delta_F < \Delta'_F$), another set of Fourier coefficients $A'_{E(O),k'}$ and $B'_{E(O),k'}$ is obtained with associated basis functions for the spectral domain. Without going into more detail here, these coefficients can be evaluated from $A_{E(O),k}$ and $B_{E(O),k}$ and the scalar products of the basis functions labeled with k and k'. For instance, one finds for the ratio of the means $A'_{E(O),0}/A_{E(O),0} = \Delta_F/\Delta'_F < 1$. Basically, $A_{E(O),k}$ and $B_{E(O),k}$ maximise the information per coefficient when f_L and f_U are known. For example, a function which is constant inside a spectral domain and zero outside appears as a boxcar function from outside this domain. Thus, the only coefficient $A_{E(O),0}$ is mapped to an infinite number of coefficients $A'_{E(O),k'}$ and $B'_{E(O),k'}$ which are mandatory to capture both discontinuities.

In the spatial domain, the basis functions for $A'_{E(O),k'}$ and $B'_{E(O),k'}$ behave differently than the ones for $A_{E(O),k}$ and $B_{E(O),k}$. Important to mention is the effect which the larger bandwidth has; $\text{sinc}_{k'\pm}$ are spatially narrower than $\text{sinc}_{k\pm}$. In addition, the number of coefficients per spatial domain increases which is expressed by $k' = k\Delta'_F/\Delta_F$.

For $f'_C = 1873.7$ GHz and $\Delta'_F = 2f'_C$ Figure 2 shows some basis functions ($k' = 0, 1$, and 2) in the spatial domain. Indeed, the basis function for $A_{E,0}$ with $f_C = 500$ GHz and $\Delta_F = 1000$ GHz (see Figure 1(a)) is broader.

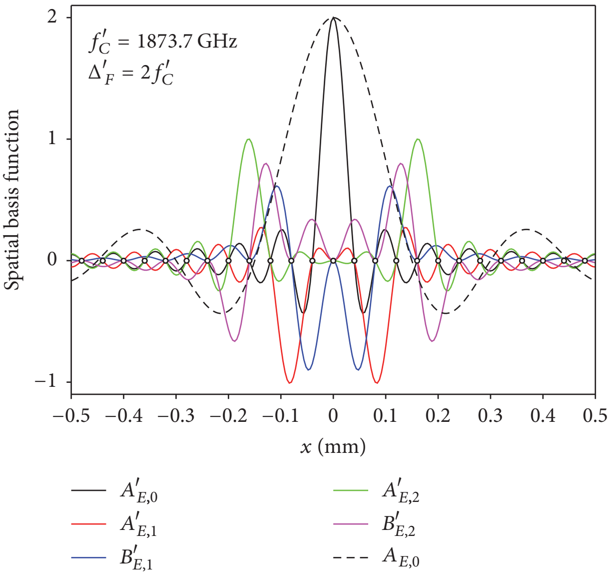


Figure 2: Basis functions in spatial domain of Fourier coefficients $A'_{E,k'}$, $B'_{E,k'}$ with order $k' = 0, 1$, and 2 for even spectral function, being finite in spectral domain with centre frequency $f'_C = 1873.7$ GHz and bandwidth $\Delta'_F = 2f'_C$. For this domain, the sinc functions, contributing to the basis functions, are narrower than for $f_C = 500$ GHz and $\Delta_F = 1000$ GHz (see Figure 1(a)). As an example, the basis function for $A_{E,0}$ (dashed-black) is given. Since $\Delta'_F \approx 4\Delta_F$ holds, approximately 4 times more Fourier coefficients locate inside a given spatial domain. Δ'_F was chosen to equal the Nyquist frequency $f_{\text{Ny}} = c/(2\Delta x)$, having set the spatial sampling increment $\Delta x = 40\,\mu$m (black dots).

2.4. Parseval's Theorem. Parseval's theorem states abstractly that the length of a function in the spectral domain equals the length of its Fourier transform counterpart in the spatial domain. The length l_S of the band-limited function S is stated by

$$l_S^2 = c\int_{f_L}^{f_U} S^2 df + c\int_{-f_U}^{-f_L} S^2 df = 2c\int_{f_L}^{f_U}\left(S_E^2 + S_O^2\right)df, \quad (15)$$

because only the term $S_E S_O$ remains odd and cancels by the integration. Furthermore, the scaling by the factor c appears which originates in $df/d\sigma = c$. Replacing $S_{E(O)}$ with the expression (12) and exploiting that the basis functions are perpendicular in the spectral domain leaves

$$l_S^2 = c\Delta_F \sum_{i=E,O}\left[2A_{i,0}^2 + \sum_{k=1}^{\infty}\left(A_{i,k}^2 + B_{i,k}^2\right)\right]. \quad (16)$$

Thus, the length is given by the sum of the Fourier coefficients squared. According to Parseval's theorem,

$$l_S^2 = \int_{-\infty}^{\infty} V^2(x)\, dx, \quad (17)$$

the length in the spectral and spatial domain remains unchanged. Inserting (13) in the above expression implies that the spatial basis functions must be orthogonal for $k_1 \neq k_2$, and the spatial integral yields $c/(2\Delta_F)$ for $k_1 = k_2 = 0$ and c/Δ_F for $k_1 = k_2 > 0$. Analytically, this is hard to prove; however, this was numerically investigated and is considered to be valid.

2.4.1. Square-Integrable Functions. The function S is said to be square-integrable, when the condition $l_S^2 < \infty$ holds. Furthermore, if S is square-integrable, then the Fourier series representations in (12) converges towards S_E and S_O almost everywhere in the spectral domain as the order k grows [10]. Hence, the requirement on S to be square-integrable seems reasonable.

2.5. Interferometric Data and Basic Model

2.5.1. Ideal and Real-World Interferometer. The Fourier transform can be performed by an interferometer, achieving an optical path difference between two partial beams, and the real-valued function $V(x)$ can be sampled. From theoretical point of view, with an ideal interferometer diagnostic, a purely symmetric and noiseless interferogram is acquired. However, a real-world interferometer suffers from diagnostic imperfections like, for example, dispersion of any kind and/or misalignment. As a consequence, any acquired interferogram is to some degree asymmetric, and, hence, an odd feature is inherent due to the measuring principle. Furthermore, a measurement involves noise, always.

2.5.2. Spatial Sampling and Implications. In the spatial domain, $V(x)$ is sampled at a finite set of optical path difference locations x_i with $i \in [1, N_D]$, and N_D marks the number of sample points. Usually, the sampling with constant increment $\Delta x = x_{i+1} - x_i$ between subsequent locations is preferred which puts constraints on the diagnostic design. Furthermore, the spatial origin is most likely missed by the sampling, and, thus, the absolute value of x_i might be unknown. If so, it is mandatory to introduce the zero-path difference x_0 which is in the following set that $x_1 = x_0$ holds.

The finite spatial sampling leaves $V(x)$ undetermined between the sampling nodes and outside the limits x_1 and x_{N_D}. Assuming $\Delta x =$ const. holds, the Nyquist theorem states that the maximum frequency accessible is given by the Nyquist frequency $f_{\text{Ny}} = c/(2\Delta x)$. Hence, for the spectral quantities S_E and S_O to be inferred, a maximum for the upper limit $f_{U,\text{Max}} = f_{\text{Ny}}$ follows from sampling theory. To prevent aliasing, Δx needs to be chosen small enough so that S_E and S_O vanish below f_{Ny}. In case, f_U can be acted on by reducing the diagnostic throughput via optical filters, the transmission line, the detector sensitivity, and postdetection amplifier settings. In fact, solely by these precautions, one can make sure that no other band well above f_{Ny} contributes to $V(x_i)$. If and only if no such band exists, then the interferogram is smooth with respect to the chosen sampling nodes, and missing to sample exactly at x_0 has no profound impact.

A diagnostic limitation is that the distance $x_{N_D} - x_1$ is finite, and, thus, no sampling is achieved below and above these limits. To gain information about S_E and S_O or the phase α (see (8)), $V(x)$ needs to be sampled on both sides of the spatial origin, so that the asymmetric feature in the interferogram is captured. Hence, in the following x_1 and, thus, x_0 are set to be negative, and the double-sided region is identified for the locations $|x_i| \leq |x_0|$. This diminishes the maximum optical path difference x_{N_D} achievable, being positive, and, thus, the length of the single-sided domain is identified by the relation $|x_0| < x_i \leq x_{N_D}$. Because x_{N_D} scales with the order k of the Fourier coefficients (see Section 2.3), only a finite number of coefficients can be probed. According to Parseval's theorem (see Section 2.4), information about the total length is missing. Furthermore, the Gibbs phenomenon, that is, a ringing, is present when S_E and S_O are inferred. Hence, x_{N_D} should be maximal, so that as many as possible coefficients can be probed to decrease the loss of information. However, a trade-off between lengths of the single-sided and double-side domains is inevitable, depending on the level of the asymmetric imperfection.

2.5.3. Noise Contribution. Since any measurement has a noise contribution, the noisy data value can be written as $D(x_i) = V(x_i) + \epsilon_i$, and the actual interferometric data is expressed by the vector $\vec{D} = \vec{V} + \vec{\epsilon}$. As spectral quantities are investigated, photons are involved in the measuring principle, and, hence, a part of ϵ_i has a Poissonian origin. However, the diagnostic under investigation later probes broadband spectra in the microwave and far-infrared range, and, thus, a large number of photons are present. Hence, the central limit theorem suggests that ϵ_i is a sample of a normal distribution with vanishing mean and a certain variance given by the squared noise level σ_D^2. In any case, dedicated diagnostics tests are mandatory to characterise ϵ_i for a given interferometer.

2.5.4. Basic Model. The combination of the relation (10) with the interferometric data, being noisy and sampled in a finite spatial domain, gives the most basic model. Formally, this model is stated here by

$$\begin{aligned} D(x_i) &= 2\int_{f_L}^{f_U}\left[S_E\cos\left(2\pi\frac{f}{c}x_i\right)+S_O\sin\left(2\pi\frac{f}{c}x_i\right)\right]df \\ &\quad+\epsilon_i \end{aligned} \tag{18}$$

with the additional information: the zero-path difference location $x_0 < 0$, the spatial limits are restricted by $x_1 = x_0$, $x_{N_D} > 0$ and $|x_i| < \infty$, the sampling increment $\Delta x = x_{i+1} - x_i$ is constant and known, S_E and S_O have the same spectral limits which obey $f_L \geq 0$ and $f_U \leq f_{\mathrm{Ny}} = c/(2\Delta x)$, and in case of Gaussian noise $p(\epsilon_i) = \mathcal{N}(0,\sigma_D^2)$. To be precise, the unknowns of the model are x_0, f_L, f_U, $S_E(f)$ and $S_O(f)$, and σ_D.

The basic model is a starting point and must be amended by diagnostic imperfections and specifics to the interferometer design type.

2.6. Inferring Spectra by Standard Analysis Techniques. To infer the spectral quantities S_E and S_O from an interferometric data set $\vec{D}$, the standard techniques rely on a noiseless model and follow a hierarchical ansatz. After making assumptions on the spectral limits f_L and f_U, the zero-path difference location x_0 is estimated. The next step evaluates a phase α which is a measure of the ratio S_O/S_E, relying on the data located in the double-sided region. Given the model, the spectral limits, the spatial origin, and the phase, S_E and S_O are estimated up to the Nyquist frequency from the whole data set. To reduce the Gibbs phenomenon on the inferred spectral quantities, window functions are multiplied to the interferometric data. In the following, the weak points of the standard analysis techniques are described.

2.6.1. Noiseless Model. The model used and stated by (18) lacks the noise contribution by definition, and, hence, treats the noisy data $\vec{D}$ as being not noisy. Strictly speaking, the model is not applicable for the problem at hand. Only by repeating the measurement sufficiently often so that the total noise contribution given by the vector $\vec{\epsilon}$ becomes small, implying $\vec{D} = \vec{V}$, the model applies. Because the integration time remains finite or only a single measurement is possible, the influence of the noise on the inferred quantities cannot be derived from the noiseless model.

2.6.2. Hierarchical Ansatz. Several steps are carried out to deduce the quantities of main interest S_E and S_O. Each step relies on model assumptions, which are not questioned or tested in any way, and results of previous steps, which carry an unstated uncertainty. This hierarchical ansatz lacks the uncertainty propagation onto S_E and S_O entirely.

2.6.3. Spectral Limits: Nyquist Assumptions. Two fundamental assumptions, called Nyquist assumptions in the following, are made by setting the spectral limits to 0 and the Nyquist frequency $f_{\mathrm{Ny}} = c/(2\Delta x)$ (see Section 2.5.2). Hence, the chosen spatial sampling would determine the bandwidth of the spectrum which is a misconception. Furthermore, the associated Fourier coefficients are located Δx apart via their basis functions in the spatial domain, and the maximum order probed is artificially blown up to $x_{N_D}/(2\Delta x)$ due to the Nyquist assumptions (see Section 2.3). However, if the functions S_E and S_O are finite in the spectral domain with limits $f_L > 0$ and $f_U < f_{\mathrm{Ny}}$, then the embedding of the smaller domain $[f_L, f_U]$ into the domain $[0, f_{\mathrm{Ny}}]$ leads to the reduction of the information content per Fourier coefficient as discussed in Section 2.3.2. Figure 2 compares the narrower spatial basis functions for the Nyquist assumptions with $\Delta x = 40\,\mu$m, that is, $f'_L = 0$ GHz and $f'_U = 3747.4$ GHz ($f'_C = 1873.7$ GHz, $\Delta'_F = 2f'_C$) with the wider basis function for the absolute term for $f_L = 0$ GHz and $f_U = 1000$ GHz ($f_C = 500$ GHz, $\Delta_F = 1000$ GHz).

The uncertainty of a Fourier coefficient, relying on the Nyquist assumptions, scales like σ_D/f_{Ny} (noise level/maximum bandwidth) which follows from the linear uncertainty propagation for (13). But for the band-limited case, the uncertainty would scale like $(\Delta_F/f_{\mathrm{Ny}})^{1/2}\sigma_D/\Delta_F$, where the square root term states that more than one data point is related to one coefficient. Hence, if a band-limitation exists but is not taken into account, then the uncertainty is maximised on the inferred spectral quantities.

2.6.4. Estimation of Spatial Origin. The spatial origin or zero-path difference x_0 is most likely missed by the spatial sampling. One of the standard approaches to estimate x_0 fits a parabola to the main interferogram peak without any information about the even and odd spectra itself. However, as one can see from (13), the basis functions for the even and odd absolute terms ($A_{E,0}$ and $A_{O,0}$) are of leading order close to the spatial origin. Hence, information about the zeroth-order coefficients and the spectral limits should be at hand for the estimation of x_0.

With x_0 available, the double-sided and single-sided regions are identified. Though, a systematically affected estimate of the origin causes an additional asymmetry in the interferometric data which would result in an increase of S_O and a decrease in S_E which is usually interpreted as a phase ramp feature. Hence, the origin should be determined with the criterion that it minimises the odd spectral function.

2.6.5. Windowing. Having only a finite amount of Fourier coefficients probed causes the Gibbs phenomenon to appear for the spectral quantities inferred. To reduce this ringing feature, window functions are applied in the spatial domain to bring the interferometric data smoothly to zero towards the sampling limits. More precise, probed Fourier coefficients of higher orders are damped out, and a window function corresponds to a certain convolution function in the spectral domain. Hence, a weighted averaging of the spectral quantities is carried out which reduces the ringing.

This approach can give a good global approximation of S_E and S_O for regions with no significant gradients. However, the damping of Fourier coefficients worsens the convergence of the inferred quantities in regions with considerable gradients.

Implicitly, the application of window functions excludes the investigation of the uncertainty on S_E and S_O introduced by nonprobed Fourier coefficients. Hence, the requirement of square-integrability of the spectral functions is not taken into account.

3. Bayesian Formalism

3.1. Bayes' Theorem. The joint probability density function (pdf) $p(d, v)$ captures the chance that the outcome d, let us say a data value or set, and the outcome v, a single model parameter or a set, are realised simultaneously.

The product rule

$$p(d, v) = p(d \mid v)\, p(v) = p(v \mid d)\, p(d) \tag{19}$$

introduces the conditional probabilities for finding the outcome d, if the outcome v were true and vice versa. By the theorem of Bayes

$$p(v \mid d) = \frac{p(d \mid v)}{p(d)} p(v), \tag{20}$$

one conditional probability can be expressed by the other, when the marginal distributions $p(d) = \int p(v \mid d) p(d) \mathrm{d}v$ and $p(v) = \int p(d \mid v) p(v) \mathrm{d}d$ are known. Hence, Bayes' theorem captures the information/knowledge gained for v when a certain outcome for d has manifested. For the pdfs occurring in Bayes' theorem, common names are used, that is, the posterior $p(v \mid d)$, the likelihood $p(d \mid v)$, the evidence $p(d)$, and the prior $p(v)$. The link or functional dependence $d = \mathrm{f}(v) + \epsilon$ enters in the likelihood which takes into account known uncertainties ϵ like, for example, measurement noise. Any knowledge about v before new data is available can be found in the prior $p(v)$.

Bayes' rule can be extended to

$$p(v, h \mid d) = \frac{p(d \mid v)}{p(d)} p(v \mid h)\, p(h), \tag{21}$$

introducing a set of hyperparameter h which enters per definition solely in the prior $p(v \mid h)$. The additional pdf $p(h)$ is called hyperprior which allocates trust in h. Apart from having the posterior for the parameters v, the marginalisation with respect to v reveals the posterior

$$p(h \mid d) = \int p(v, h \mid d)\, \mathrm{d}v = \frac{p(d \mid h)}{p(d)} p(h) \tag{22}$$

for h which measures the plausibility of an outcome of the hyperparameter given the data. Since $p(d)$ does not depend on h, the most likely hyperparameter set is identified by the maximum of $p(d \mid h)$, assuming $p(h)$ is uniform.

3.2. Formalism for Linear, Nonlinear, and Hyperparameter Problem for Gaussian Noise

3.2.1. Multivariate Normal. Let the joint pdf for the random vector $\vec{y}$ ($\in R^N$) be a multivariate normal with mean $\vec{\mu}$ ($\in R^N$) and covariance matrix $\mathbf{\Sigma}$ ($\in R^N \times R^N$); then the pdf becomes

$$\begin{aligned} p(\vec{y} \mid \vec{\mu}, \mathbf{\Sigma}) &= \mathcal{N}(\vec{y} \mid \vec{\mu}, \mathbf{\Sigma}) \\ &= \frac{\exp\left[-(1/2)(\vec{y} - \vec{\mu})^T \mathbf{\Sigma}^{-1} (\vec{y} - \vec{\mu})\right]}{(2\pi)^{N/2} |\mathbf{\Sigma}|^{1/2}} \end{aligned} \tag{23}$$

with the determinant $|\mathbf{\Sigma}|$.

3.2.2. Model for Linear Problem. If the dependency between the data and the parameters of interest is linear, and the likelihood and the prior can be expressed by multivariate normals, then the evaluation of the posterior is analytically straightforward. Such a model is the starting point for investigating a more complex model which includes parameters with a nonlinear mapping to the data domain and/or hyperparameters.

(a) Gaussian Likelihood. The data may be represented by the vector $\vec{D}$ ($\in R^{N_D}$). The parameters of interest $\vec{l}$ ($\in R^{N_l}$) map linearly to the data domain like

$$\vec{D} = \mathbf{M}\vec{l} + \vec{\epsilon}, \tag{24}$$

where the $N_D \times N_l$ dimensional matrix $\mathbf{M}$ encodes the linear mapping, and $\vec{\epsilon}$ captures the random noise contribution. When the data is acquired independently, and the noise is independent for each datum and follows a Gaussian $\epsilon_i = \mathcal{N}(0, \sigma_D^2)$ with vanishing mean and standard deviation σ_D (noise level), then the Gaussian likelihood

$$\begin{aligned} p(\vec{D} \mid \vec{l}) &= \mathcal{N}(\vec{D} \mid \vec{l}, \mathbf{\Sigma}_D) \\ &= \frac{\exp\left[-(1/2)(\vec{D} - \mathbf{M}\vec{l})^T \mathbf{\Sigma}_D^{-1} (\vec{D} - \mathbf{M}\vec{l})\right]}{(2\pi)^{N_D/2} |\mathbf{\Sigma}_D|^{1/2}} \end{aligned} \tag{25}$$

can be found with the covariance matrix $\mathbf{\Sigma}_D = \sigma_D^2 \delta_{ij}$.

(b) Gaussian Prior. The prior information about $\vec{l}$ may be expressed by the multivariate normal

$$\begin{aligned} p(\vec{l}) &= \mathcal{N}(\vec{l} \mid \vec{l}_{\mathrm{Pr}}, \mathbf{\Sigma}_{\mathrm{Pr}}) \\ &= \frac{\exp\left[-(1/2)(\vec{l} - \vec{l}_{\mathrm{Pr}})^T \mathbf{\Sigma}_{\mathrm{Pr}}^{-1} (\vec{l} - \vec{l}_{\mathrm{Pr}})\right]}{(2\pi)^{N_l/2} |\mathbf{\Sigma}_{\mathrm{Pr}}|^{1/2}} \end{aligned} \tag{26}$$

with the prior mean $\vec{l}_{\mathrm{Pr}}$ and the prior covariance $\mathbf{\Sigma}_{\mathrm{Pr}}$.

(c) Gaussian Posterior and Evidence. Formally, Bayes' theorem states the posterior by

$$p\left(\vec{l} \mid \vec{D}\right) = \frac{\mathcal{N}\left(\vec{D} \mid \vec{l}, \boldsymbol{\Sigma}_D\right)}{p\left(\vec{D}\right)} \mathcal{N}\left(\vec{l} \mid \vec{l}_{\mathrm{Pr}}, \boldsymbol{\Sigma}_{\mathrm{Pr}}\right). \tag{27}$$

After some algebra, one can show that the posterior

$$p\left(\vec{l} \mid \vec{D}\right) = \mathcal{N}\left(\vec{l} \mid \vec{l}_{\mathrm{Po}}, \boldsymbol{\Sigma}_{\mathrm{Po}}\right) \tag{28}$$

is a multivariate normal with posterior mean

$$\vec{l}_{\mathrm{Po}} = \vec{l}_{\mathrm{Pr}} + \boldsymbol{\Sigma}_{\mathrm{Po}} \mathbf{M}^T \boldsymbol{\Sigma}_D^{-1} \left(\vec{D} - \mathbf{M} \vec{l}_{\mathrm{Pr}}\right) \tag{29}$$

and covariance

$$\boldsymbol{\Sigma}_{\mathrm{Po}} = \left(\mathbf{M}^T \boldsymbol{\Sigma}_D^{-1} \mathbf{M} + \boldsymbol{\Sigma}_{\mathrm{Pr}}^{-1}\right)^{-1} \tag{30}$$

which are both analytically obtained. Furthermore, the evidence reads

$$\begin{aligned} p\left(\vec{D}\right) = {} & \frac{\exp\left[-(1/2)\,\vec{D}^T \boldsymbol{\Sigma}_D^{-1} \vec{D}\right]}{(2\pi)^{N_D/2} \left|\boldsymbol{\Sigma}_D\right|^{1/2}} \frac{\left|\boldsymbol{\Sigma}_{\mathrm{Po}}\right|^{1/2}}{\left|\boldsymbol{\Sigma}_{\mathrm{Pr}}\right|^{1/2}} \\ & \cdot \frac{\exp\left[(1/2)\,\vec{l}_{\mathrm{Po}}^T \boldsymbol{\Sigma}_{\mathrm{Po}}^{-1} \vec{l}_{\mathrm{Po}}\right]}{\exp\left[(1/2)\,\vec{l}_{\mathrm{Pr}}^T \boldsymbol{\Sigma}_{\mathrm{Pr}}^{-1} \vec{l}_{\mathrm{Pr}}\right]}, \end{aligned} \tag{31}$$

where the first part depends explicitly on the measured data, and the second part, being dimensionless, incorporates the ratio dependent on the means and covariances of the prior and posterior.

3.2.3. Model for Linear, Nonlinear, and Hyperparameter Problem. The linear model is amended by hyperparameters, entering in some way in the prior, and parameters with a nonlinear connection to the data domain. Such a model is then applicable in the field of Fourier spectroscopy.

(a) Gaussian Likelihood. The linear mapping $\mathbf{M}$ of the parameters $\vec{l}$ to the data domain, as stated by (24), should remain valid. However, the mapping itself may depend on the parameters $\vec{n}$ in a nonlinear way, so that $\mathbf{M} = \mathbf{M}(\vec{n})$. This leaves the Gaussian likelihood in (25) formally unchanged but is symbolically stated as $p(\vec{D} \mid \vec{l}, \vec{n})$.

(b) Priors. The Gaussian prior for $\vec{l}$ should be given by $p(\vec{l} \mid \vec{h}) = \mathcal{N}(\vec{l}_{\mathrm{Pr}}(\vec{h}), \boldsymbol{\Sigma}_{\mathrm{Pr}}(\vec{h}))$, where the prior mean and covariance depend on some of the hyperparameters $\vec{h}$. Similarly, a prior $p(\vec{n} \mid \vec{h})$ follows for the nonlinear parameters. Finally, the hyperparameters have an assigned prior $p(\vec{h})$.

(c) Posteriors and Evidence. According to Bayes' theorem, one can write the joint posterior like

$$\begin{aligned} & p\left(\vec{l} \mid \vec{n}, \vec{h}, \vec{D}\right) p\left(\vec{n}, \vec{h} \mid \vec{D}\right) \\ & \quad = \frac{p\left(\vec{D} \mid \vec{l}, \vec{n}\right)}{p\left(\vec{D}\right)} p\left(\vec{l} \mid \vec{h}\right) p\left(\vec{n} \mid \vec{h}\right) p\left(\vec{h}\right), \end{aligned} \tag{32}$$

and the conditional amplitude posterior for $\vec{l}$ becomes a multivariate normal

$$\begin{aligned} & p\left(\vec{l} \mid \vec{n}, \vec{h}, \vec{D}\right) \\ & \quad = \mathcal{N}\left(\vec{l} \mid \vec{l}_{\mathrm{Po}}\left(\vec{n}, \vec{h}\right), \boldsymbol{\Sigma}_{\mathrm{Po}}\left(\vec{n}, \vec{h}\right)\right). \end{aligned} \tag{33}$$

Thus, both, the conditional posterior mean $\vec{l}_{\mathrm{Po}}$ and covariance $\boldsymbol{\Sigma}_{\mathrm{Po}}$ evaluated by (29) and (30), depend on the nonlinear parameters and hyperparameters. After the trivial marginalisation with respect to $\vec{l}$, the joint posterior for $\vec{n}$ and $\vec{h}$

$$\begin{aligned} p\left(\vec{n}, \vec{h} \mid \vec{D}\right) = {} & \frac{\exp\left[-(1/2)\,\vec{D}^T \boldsymbol{\Sigma}_D^{-1} \vec{D}\right]}{(2\pi)^{N_D/2} \left|\boldsymbol{\Sigma}_D\right|^{1/2}} \frac{1}{p\left(\vec{D}\right)} \\ & \cdot \frac{\left|\boldsymbol{\Sigma}_{\mathrm{Po}}\right|^{1/2}}{\left|\boldsymbol{\Sigma}_{\mathrm{Pr}}\right|^{1/2}} \frac{\exp\left[(1/2)\,\vec{l}_{\mathrm{Po}}^T \boldsymbol{\Sigma}_{\mathrm{Po}}^{-1} \vec{l}_{\mathrm{Po}}\right]}{\exp\left[(1/2)\,\vec{l}_{\mathrm{Pr}}^T \boldsymbol{\Sigma}_{\mathrm{Pr}}^{-1} \vec{l}_{\mathrm{Pr}}\right]} \\ & \cdot p\left(\vec{n} \mid \vec{h}\right) p\left(\vec{h}\right) \end{aligned} \tag{34}$$

remains. By expressing the posterior, named settings posterior in the following, like

$$\begin{aligned} p\left(\vec{n}, \vec{h} \mid \vec{D}\right) = {} & \frac{1}{K'} \frac{\left|\boldsymbol{\Sigma}_{\mathrm{Po}}\right|^{1/2}}{\left|\boldsymbol{\Sigma}_{\mathrm{Pr}}\right|^{1/2}} \frac{\exp\left[(1/2)\,\vec{l}_{\mathrm{Po}}^T \boldsymbol{\Sigma}_{\mathrm{Po}}^{-1} \vec{l}_{\mathrm{Po}}\right]}{\exp\left[(1/2)\,\vec{l}_{\mathrm{Pr}}^T \boldsymbol{\Sigma}_{\mathrm{Pr}}^{-1} \vec{l}_{\mathrm{Pr}}\right]} \\ & \cdot p\left(\vec{n} \mid \vec{h}\right) p\left(\vec{h}\right), \end{aligned} \tag{35}$$

the evidence is identified with

$$p\left(\vec{D}\right) = K' \frac{\exp\left[-(1/2)\,\vec{D}^T \boldsymbol{\Sigma}_D^{-1} \vec{D}\right]}{(2\pi)^{N_D/2} \left|\boldsymbol{\Sigma}_D\right|^{1/2}}. \tag{36}$$

Note, that the dimensionless constant K', and, thus, the evidence depend on the chosen model, including likelihood and priors. Hence, K' is of importance, when the model is even further abstracted or compared with alternative models.

(d) Role of Settings Posterior. The optimisation, that is, the finding of the maximum of the settings posterior $p(\vec{n},\vec{h}\mid\vec{D})$, can be interpreted as an implementation of the Ockham's razor principle and/or as a regularisation procedure. This is essential when the number of parameters exceeds the number of data points.

Unfortunately, a general analytical expression is not available for this posterior, and, thus, it needs to be investigated numerically for the problem at hand. In order to do so, the quantity

$$p_{\times K'} = p\left(\vec{n},\vec{h}\mid\vec{D}\right)\times K' \tag{37}$$

is of interest, because it is numerically accessible. In case, $p_{\times K'}$ has a well distinguishable global maximum, $p_{\times K'}$ can be approximated by a multivariate normal which is estimated by evaluating the Hessian matrix. Thus, one finds the posterior means $\vec{n}_{\text{Po}}$ and $\vec{h}_{\text{Po}}$ with the associated posterior covariance. This allows an approximate marginalisation with respect to $\vec{n}$ and/or $\vec{h}$. This can be understood as a propagation of the posterior uncertainties in $\vec{n}$ and $\vec{h}$ to the marginalised posterior $p(\vec{l}\mid\vec{D})$ for the parameters of interest.

(e) Simplifications. For the remainder of this paper, some simplifications are made which modify (34), (35), and (36) accordingly. The prior mean $\vec{l}_{\text{Pr}}$ is set to 0, and the priors $p(\vec{n}\mid\vec{h})$ and $p(\vec{h})$ are chosen to be uniform. Then, one can set $K' = Kp(\vec{n}\mid\vec{h})p(\vec{h})$ which modifies (37) to

$$p_{\times K} = \frac{\left|\boldsymbol{\Sigma}_{\text{Po}}\right|^{1/2}}{\left|\boldsymbol{\Sigma}_{\text{Pr}}\right|^{1/2}}\exp\left[\frac{1}{2}\vec{l}_{\text{Po}}^{T}\boldsymbol{\Sigma}_{\text{Po}}^{-1}\vec{l}_{\text{Po}}\right]. \tag{38}$$

4. Brownian Bridge Covariance

The continuous even and odd spectral functions to be inferred can be modelled each by a Gaussian process [11]. Thereby, the Brownian bridge process is a good starting point, because it exploits a fundamental condition to prevent aliasing for Fourier spectroscopy applications. This condition states that the spectrum S and, thus, S_E and S_O must vanish at the spectral origin and at an upper limit which is smaller than the Nyquist frequency (see Section 2.6.3). However, this information is usually not taken into account any further in the analysis. On the contrary, a Brownian bridge and its associated covariance function fulfil the boundary conditions for any lower and upper limit. Hence, the covariance can be used in the Gaussian prior for S_E and S_O. In addition, this process has only one scaling hyperparameter which makes it attractive from data analysis point of view. This scaling can be estimated as well from the Fourier coefficients probed. In fact, this reveals information about the nonprobed coefficients and gives an additional uncertainty on S_E and S_O. After presenting some properties of the Brownian bridge covariance, it is used as prior covariance in the example application (see Section 5).

4.1. Standard Definition. The Brownian bridge is a continuous stochastic process for an interval, say from 0 to T. This bridge is constructed by tying-down a Brownian motion process to 0 at the end of the interval in question. Furthermore, the tie-down at the beginning of the interval is inherited from the Brownian motion process. The covariance function for the bridge is defined in standard literature by

$$\boldsymbol{\Sigma}'_{\text{BB}}(s,t) = -\frac{st}{T} + \begin{cases} s: & s \le t \\ t: & s > t \end{cases} \tag{39}$$

for s and $t \in [0,T]$.

4.2. Adapted Definition. The even and odd spectral functions, being finite for the interval $[f_L, f_U]$, demand some adaption of the standard covariance expression (39) when modelled by a Brownian bridge process. To keep the same properties on spectral scale, a shift by f_L and the interval length $T = f_U - f_L = \Delta_F$ need to be set. Thus, one gets

$$\begin{aligned}\boldsymbol{\Sigma}'_{\text{BB}}\left(f,f'\right) = {} & -\frac{\left(f-f_L\right)\left(f'-f_L\right)}{\Delta_F} \\ & + \begin{cases} f-f_L: & f\le f' \\ f'-f_L: & f>f' \end{cases}\end{aligned} \tag{40}$$

with the unit $[\boldsymbol{\Sigma}'_{\text{BB}}] = \text{Hz}$. With the normalisation

$$\int_{f_L}^{f_U}\int_{f_L}^{f_U}\boldsymbol{\Sigma}'_{\text{BB}}\left(f,f'\right)df\,df' = \frac{\Delta_F^3}{12}, \tag{41}$$

the modified covariance becomes

$$\boldsymbol{\Sigma}_{\text{BB}}\left(f,f'\right) = \frac{12}{\Delta_F^3}\boldsymbol{\Sigma}'_{\text{BB}}\left(f,f'\right), \tag{42}$$

where $[\boldsymbol{\Sigma}_{\text{BB}}] = \text{Hz}^{-2}$ has now the proper unit with respect to the spectral scale.

The parameters a_E and a_O of unit $[a_{E(O)}] = \text{V}^2$ are introduced which are defined each as a scaling factor for the associated process. With these scalings the covariances $\boldsymbol{\Sigma}_{\text{BB},E(O)} = a_{E(O)}\boldsymbol{\Sigma}_{\text{BB}}$ are obtained, so that the units $[\boldsymbol{\Sigma}_{\text{BB},E(O)}] = [S_{E(O)}]^2 = \text{V}^2/\text{Hz}^2$ match.

4.3. Covariance for Fourier Coefficients. The Brownian bridge covariance function for the spectral domain can be studied in the domain of the Fourier coefficients via the coordinate transform stated in (11) [11]. Compactly written, one finds the infinite-dimensional covariance matrix for the Fourier coefficients analytically by

$$\begin{aligned}\mathbf{\Sigma}_{\mathrm{BB},E}\left(k,k'\right) &= \frac{a_E}{\Delta_F^2}\int_{f_L}^{f_U}\int_{f_L}^{f_U}\begin{pmatrix}1\\ 2\cos\left(2\pi k\dfrac{f-f_C}{\Delta_F}\right)\\ 2\sin\left(2\pi k\dfrac{f-f_C}{\Delta_F}\right)\end{pmatrix}\begin{pmatrix}1\\ 2\cos\left(2\pi k'\dfrac{f'-f_C}{\Delta_F}\right)\\ 2\sin\left(2\pi k'\dfrac{f'-f_C}{\Delta_F}\right)\end{pmatrix}^T \Sigma_{\mathrm{BB}}\left(f,f'\right)df\,df' \\ &= \frac{a_E}{\Delta_F^2}\frac{6}{\pi^2}\begin{pmatrix}\dfrac{\pi^2}{6} & \dfrac{(-1)^{k'+1}}{k'^2} & \vec{0}^T\\ \dfrac{(-1)^{k+1}}{k^2} & \dfrac{\delta_{k,k'}}{kk'} & \mathbf{0}\\ \vec{0} & \mathbf{0} & \dfrac{\delta_{k,k'}}{kk'}\end{pmatrix} = \begin{pmatrix}\Sigma\left(A_{E,0},A_{E,0}\right) & \vec{\Sigma}^T\left(A_{E,0},A_{E,k'}\right) & \vec{0}^T\\ \vec{\Sigma}\left(A_{E,k},A_{E,0}\right) & \mathbf{\Sigma}\left(A_{E,k},A_{E,k'}\right) & \mathbf{0}\\ \vec{0} & \mathbf{0} & \mathbf{\Sigma}\left(B_{E,k},B_{E,k'}\right)\end{pmatrix}\end{aligned} \tag{43}$$

for all $k, k' \geq 1$, and similarly $\mathbf{\Sigma}_{\mathrm{BB,O}}(k,k')$ follows. The only finite off-diagonal elements occur for the absolute term in connection with the higher order terms for the even coefficients captured by the infinite-dimensional row and column vectors $\vec{\Sigma}(A_{E(\mathrm{O}),k}, A_{E(\mathrm{O}),0})$ and $\vec{\Sigma}^T(A_{E(\mathrm{O}),0}, A_{E(\mathrm{O}),k'})$, respectively. This is caused by the condition that $S_{E(\mathrm{O})}$ vanishes at the spectral boundaries where the sine vanishes intrinsically for any k but the cosine takes values either 1 or −1 for even and odd orders, respectively. Hence, the covariance imposes the boundary condition.

For orders greater than 0, $\mathbf{\Sigma}_{\mathrm{BB},E(\mathrm{O})}(k,k')$ has no off-diagonal elements due to the Kronecker delta $\delta_{k,k'}$, meaning that these coefficients are independent on each other. Furthermore, for the infinite-dimensional matrices $\mathbf{\Sigma}(A_{E(\mathrm{O}),k}, A_{E(\mathrm{O}),k'}) = \mathbf{\Sigma}(B_{E(\mathrm{O}),k}, B_{E(\mathrm{O}),k'}) = \delta_{k,k'}/(kk')$ holds, and the amplitudes for even and odd coefficients drop equally with the square of the order.

4.4. Square-Integrable Property. According to Parseval's theorem (see (16)), l_S^2 is evaluated by summing the squares of the Fourier coefficients. Because the entries of the main diagonal in the covariances $\mathbf{\Sigma}_{\mathrm{BB},E}(k,k')$ and $\mathbf{\Sigma}_{\mathrm{BB,O}}(k,k')$ drop with the order squared, the Brownian bridge process ensures square-integrability of S_E and S_O as long as the scalings a_E and a_O remain finite.

4.5. Signal Envelope. For the even process, the signal level can be estimated by the envelope $\sigma'_{D,E}(x_k)$ in the data domain. Starting point is the square root of the main diagonal of $\mathbf{\Sigma}_{\mathrm{BB},E}(k, k' = k)$. Since the argument of $\mathrm{sinc}_{k\pm}$ (see (14)) localises the even and odd Fourier coefficient at a fixed $k = \pm x_k \Delta_F / c$ in the same data domain, the even and odd contributions of $\mathbf{\Sigma}_{\mathrm{BB},E}(k, k' = k)$ must be added for $k \geq 1$. As can be seen by (13), the mapping of the absolute term to the data domain includes already the factor 2. In addition, the mapping comprises the bandwidth Δ_F. In total, one finds the envelope as

$$\sigma'_{D,E}\left(x_k\right) = \pm 2a_E^{1/2}\begin{cases}2: & x_k = 0\\ \dfrac{12^{1/2}}{\pi}\dfrac{c}{\left|x_k\right|\Delta_F}: & x_k \neq 0.\end{cases} \tag{44}$$

In the above equation, factor 2 in front of a_E was chosen, so that $\sigma'_{D,E}(x_k)$ captures most of the signal. An approximation might be convenient, because $12^{1/2}/\pi \approx 1$.

For the envelope $\sigma'_{D,\mathrm{O}}(x_k)$ of the odd process, the same reasoning can be applied with one modification. The mapping demands that the contribution of the absolute term at the spatial origin vanishes (see (13)). Hence, one finds

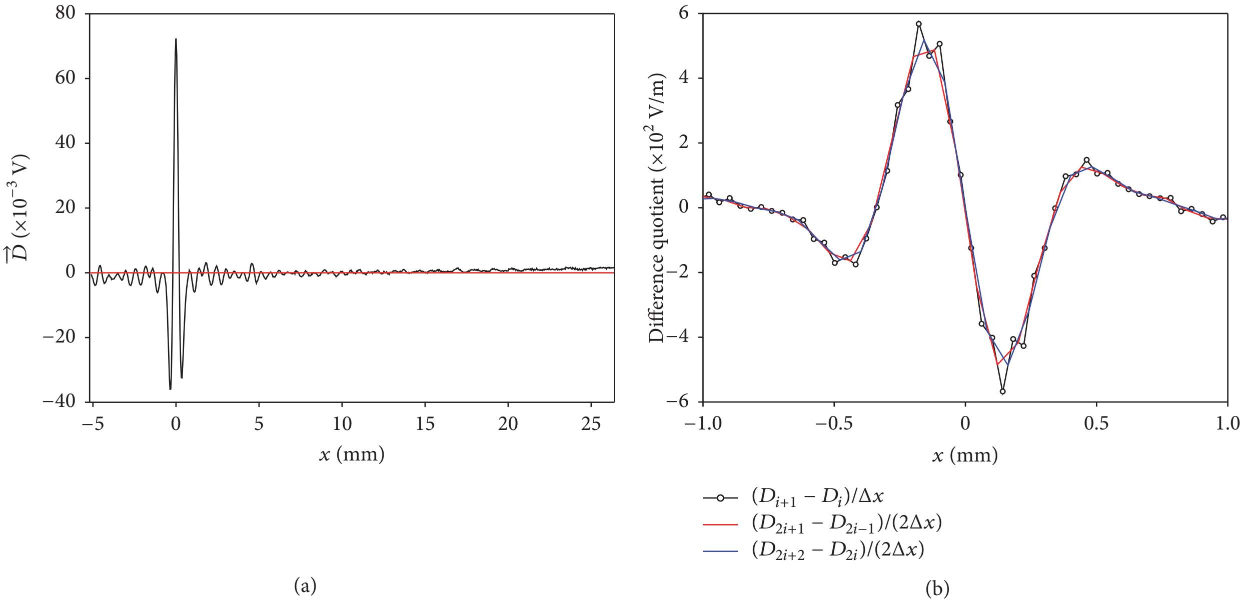
$$\sigma'_{D,\mathrm{O}}\left(x_k\right) = \pm 2a_\mathrm{O}^{1/2}\begin{cases}0: & x_k = 0\\ \dfrac{12^{1/2}}{\pi}\dfrac{c}{\left|x_k\right|\Delta_F}: & x_k \neq 0.\end{cases} \tag{45}$$

Both envelops drop with $1/|x_k|$, and, thus, most of the signal associated with each process would in the data domain.

5. Example Application

5.1. Formulation of Model

5.1.1. Martin-Puplett Interferometer at JET. The Martin-Puplett interferometer diagnostic [12] at the fusion device JET (Culham, UK) probes the spectrum emitted by a broadband source and performs the Fourier transform. The interferogram data $\vec{D}_1$ is acquired in terms of Volts dependent on the optical path difference x. However, two different sources are probed for 20 minutes subsequently to remove a class of diagnostic imperfections not treated here any further. By subtraction of the corresponding two interferograms, the



(a)

(b)

Figure 3: (a) Difference interferogram $\overrightarrow{D}$ (black) versus optical path difference x. The data set includes $N_D = 788$ values, and the noise level for each D_i reads $\sigma_D = 132.29\,\mu$V. The values of the spatial grid are determined by the standard model $x_i = (i-1)\Delta x + x_0$ with $i \in [1, N_D]$, the constant increment $\Delta x = 40\,\mu$m, and the origin x_0. This origin is estimated as −5.118 mm by a fit of a quadratic polynomial to the maximum max($\overrightarrow{D}$) and its neighbouring values to either side. Clearly, $\overrightarrow{D}$ deviates from the baseline (red). (b) Difference quotient for different subsets of $\overrightarrow{D}$. Since the zig-zag-pattern in the point-to-point variation appears only for the first case (black), the standard model $x_{i+1} - x_i =$ const. seems to be inadequate. Instead, a model seems appropriate for which $x_{2i+1} - x_{2i-1} = x_{2i+2} - x_{2i} =$ const. holds.

data becomes available in form of the difference interferogram $D(x_i) = D_2(x_i) - D_1(x_i)$ acquired at the spatial grid node x_i. Then, the abstract model for the Martin-Puplett interferometer is stated by

$$D(x_i) = \frac{A}{2}\left[2\int_{f_L}^{f_U}\left(S_E \cos\left(2\pi\frac{f}{c}x_i\right) + S_O \sin\left(2\pi\frac{f}{c}x_i\right)\right)df + V_{\text{Off}}(x_i)\right] + \mathcal{N}_i\left(0, \sigma_D^2\right), \tag{46}$$

using the total amplification A of the detection system. Furthermore, the offset V_{Off} marks a diagnostic imperfection which varies with x_i. The Gaussian noise contributes to each data sample described by $\mathcal{N}_i(0, \sigma_D^2)$. The unknown quantities in the diagnostic model are the spatial grid x_i, the lower and upper spectral boundaries f_L and f_U of the Fourier transform integral, the even and odd functions S_E and S_O dependent on frequency f, and the offset.

5.1.2. Interferometric Data. The data set $\overrightarrow{D}$, that is, the difference interferogram consists of $N_D = 788$ values (see Figure 3(a)). Merely for graphical presentation a certain x is chosen derived from the standard approach (see Section 5.1.3). Globally, the data shows an upward trend with respect to the zero baseline.

The components D_i are measured independently on each other, and the noise level for each D_i is captured by $\sigma_D = 132.29\,\mu$V, and, thus, the variance of the whole data vector is stated by the matrix $\Sigma_D = \sigma_D^2 \delta_{ij}$.

5.1.3. Optical Path Difference. The diagnostic is set up, so that the sampling of the interferogram is triggered ideally when the optical path difference has changed by the increment $\Delta x = 40\,\mu$m. Hence, the standard model $x_i = (i-1)\Delta x + x_0$ is obvious with $i = 1, \ldots, N_D$, and the zero-path difference x_0 is a free parameter. However, this model is not accurate. Applying the standard approach which fits a second-order polynomial to the maximum max($\overrightarrow{D}$) and its two nearby values, $x_0 = -127.95\Delta x = -5.118$ mm, is inferred for the data set shown in Figure 3(a). Furthermore, the difference quotient evaluated by $(D_{i+1} - D_i)/\Delta x$ is presented versus the optical path difference $x = (x_{i+1} + x_i)/2$ in Figure 3(b). The point-to-point variation of the quotient has a zig-zag pattern which implies that the assumption $\Delta x =$ const. is incorrect. Indeed, each of two the difference quotients $(D_{2i+1} - D_{2i-1})/(2\Delta x)$ and $(D_{2i+2} - D_{2i})/(2\Delta x)$ is smoother. Hence, the model

$$x_i = (i-1)\Delta x + x_0 + \begin{cases} 0: & \text{if } i \text{ is odd} \\ \delta x: & \text{if } i \text{ is even} \end{cases} \tag{47}$$

seems more appropriate, making use of two free parameters: the zero-path difference x_0 and a shift δx for every other grid value. The priors $p(x_0)$ and $p(\delta x)$ are set to be uniform.

5.1.4. Offset. The upward trend of $\overrightarrow{D}$ (see Figure 3(a)) is modelled by the offset $V_{\mathrm{Off}}(x_i)$. Here, a second-order polynomial is chosen to capture the offset by $V_{\mathrm{Off}}(x_i) = b_0 + b_1 x_i + b_2 x_i^2$ with the free parameters b_0, b_1, and b_2 being summarised by the vector $\overrightarrow{b} = (b_0, b_1, b_2)^T$. The corresponding mapping to the data domain can be expressed as

$$\mathbf{M}_{\mathrm{Off}} \overrightarrow{b} = \frac{A}{2} \begin{pmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_{N_D} & x_{N_D}^2 \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \\ b_2 \end{pmatrix} \qquad (48)$$

with the $N_D \times 3$ matrix $\mathbf{M}_{\mathrm{Off}}$. The joint prior is expressed by the factorisable multivariate normal distribution:

$$p\left(\overrightarrow{b}\right) = \prod_{i=0}^{2} \mathcal{N}\left(0, \sigma^2_{b_i,\mathrm{Pr}}\right) = \mathcal{N}\left(0, \mathbf{\Sigma}_{\overrightarrow{b},\mathrm{Pr}}\right), \qquad (49)$$

where $\sigma_{b_i,\mathrm{Pr}}$ are considered as hyperparameters to which a uniform prior is assigned.

5.1.5. Spectral Quantities. The kernel of the spectral integration in (46) is finite only for frequencies $f \in \,]f_L, f_U[$. Thereby, f_L and f_U are free parameters. The spectral domain is dicretised using the constant increment Δ_f which is considered as a free parameter as well. Then, the spectral domain is represented by the set $f_j = f_L + (j - 0.5)\Delta_f$ with $j = 1, \ldots, N_f$, covering the band $N_f \Delta_f = (f_U - f_L) = \Delta_F$ centred at $f_C = (f_U + f_L)/2$. To be clear, since f_L, f_U, and Δ_f are free parameters and will be inferred, the number N_f is variable. This number determines the dimensionality of the vectors $\overrightarrow{S}_E$ and $\overrightarrow{S}_O$ which represent the discretised functions $S_E(f)$ and $S_O(f)$. The mapping of $\overrightarrow{S}_E$ and $\overrightarrow{S}_O$ to the data domain in (46) is written as

$$\begin{aligned} &\frac{A}{2} 2 \int_{f_L}^{f_U} \left(S_E \cos\left(2\pi \frac{f}{c} x_i\right) + S_O \sin\left(2\pi \frac{f}{c} x_i\right)\right) df \\ &\quad \approx \left(\mathbf{M}_E, \mathbf{M}_O\right) \begin{pmatrix} \overrightarrow{S}_E \\ \overrightarrow{S}_O \end{pmatrix} \end{aligned} \qquad (50)$$

with the two $N_D \times N_f$ matrices

$$\begin{aligned} &\mathbf{M}_E \\ &= A \begin{pmatrix} \cos\left(2\pi \frac{f_1}{c} x_1\right) & \cdots & \cos\left(2\pi \frac{f_{N_f}}{c} x_1\right) \\ \vdots & \ddots & \vdots \\ \cos\left(2\pi \frac{f_1}{c} x_{N_D}\right) & \cdots & \cos\left(2\pi \frac{f_{N_f}}{c} x_{N_D}\right) \end{pmatrix} \Delta_f, \\ &\mathbf{M}_O \\ &= A \begin{pmatrix} \sin\left(2\pi \frac{f_1}{c} x_1\right) & \cdots & \sin\left(2\pi \frac{f_{N_f}}{c} x_1\right) \\ \vdots & \ddots & \vdots \\ \sin\left(2\pi \frac{f_1}{c} x_{N_D}\right) & \cdots & \sin\left(2\pi \frac{f_{N_f}}{c} x_{N_D}\right) \end{pmatrix} \Delta_f. \end{aligned} \qquad (51)$$

The joint prior for the two vectorial quantities $\overrightarrow{S}_E$ and $\overrightarrow{S}_O$ is factorised, and each prior is chosen as a multivariate normal distribution with vanishing mean. Since the Brownian bridge covariance $\mathbf{\Sigma}_{\mathrm{BB}}$ (see Section 4) describes functions which vanish at the boundaries f_L and f_U and are square-integrable, and its signal envelope decays with the optical path difference like the interferometric data at hand, the priors are chosen by

$$\begin{aligned} &p\left(\overrightarrow{S}_{E(\mathrm{O})} \mid \Delta_f, f_L, f_U, a_{E(\mathrm{O})}\right) \\ &\quad = \mathcal{N}\left(0, a_{E(\mathrm{O})} \mathbf{\Sigma}_{\mathrm{BB}}\left(\Delta_f, f_L, f_U\right)\right) \end{aligned} \qquad (52)$$

with the two hyperparameters a_E and a_{O} for scaling. For each of these hyperparameters a uniform prior $p(a_{E(\mathrm{O})}) = 1/\Delta_{a_{E(\mathrm{O})}}$ is applied.

Finally, the joint prior

$$p\left(\Delta_f, f_L, f_U\right) = \frac{1}{K_{\Delta_f, f_L, f_U}} \qquad (53)$$

should be constant, so that any combination of Δ_f, f_L, and f_U has the same probability. Furthermore, the conditions $0 \le \Delta_f \le f_U - f_L$, $0 \le f_L \le f_U$, and $0 \le f_U \le \Delta_{f_U}$ (global upper limit) must be fulfilled. For example, the upper limit of f_U is set here to the Nyquist frequency $\Delta_{f_U} = f_{\mathrm{Ny}} = c/(2\Delta x) =$ 3747.4 GHz.

5.2. Bayes' Theorem. In the following, the linear parameters are summerised by the set $\overrightarrow{l} = (\overrightarrow{S}_E, \overrightarrow{S}_O, \overrightarrow{b})^T$, the nonlinear parameters by $\overrightarrow{n} = (f_L, f_U, x_0, \delta x)^T$, and the hyperparameters by $\overrightarrow{h} = (a_E, a_{\mathrm{O}}, \sigma_{b_0,\mathrm{Pr}}, \sigma_{b_1,\mathrm{Pr}}, \sigma_{b_2,\mathrm{Pr}})^T$.

The matrix $\mathbf{M}(\overrightarrow{n}, \Delta_f) = (\mathbf{M}_E, \mathbf{M}_{\mathrm{O}}, \mathbf{M}_{\mathrm{Off}})$ maps the parameters $\overrightarrow{l}$ to the data domain. Hence, the Gaussian likelihood is written as

$$\begin{aligned} &p\left(\overrightarrow{D} \mid \overrightarrow{l}, \overrightarrow{n}, \Delta_f\right) \\ &\quad = \frac{\exp\left[-(1/2)\left(\overrightarrow{D} - \mathbf{M}\overrightarrow{l}\right)^T \mathbf{\Sigma}_D^{-1} \left(\overrightarrow{D} - \mathbf{M}\overrightarrow{l}\right)\right]}{(2\pi)^{N_D/2} \left|\mathbf{\Sigma}_D\right|^{1/2}}. \end{aligned} \qquad (54)$$

The prior for the full problem takes the form

$$\begin{aligned} p\left(\vec{l}, \Delta_f, \vec{n}, \vec{h}\right) &= p\left(\vec{l} \mid \Delta_f, f_L, f_U, a_E, a_O, \sigma_{b_0,\mathrm{Pr}}, \sigma_{b_1,\mathrm{Pr}}, \sigma_{b_2,\mathrm{Pr}}\right) \\ &\quad \cdot p\left(\Delta_f, f_L, f_U\right) p\left(a_E\right) p\left(a_O\right) p\left(x_0\right) p\left(\delta x\right) \\ &\quad \cdot p\left(\sigma_{b_0,\mathrm{Pr}}\right) p\left(\sigma_{b_1,\mathrm{Pr}}\right) p\left(\sigma_{b_2,\mathrm{Pr}}\right) \end{aligned} \tag{55}$$

with the multivariate normal prior

$$\begin{aligned} p\left(\vec{l} \mid \Delta_f, f_L, f_U, a_E, a_O, \sigma_{\vec{b},\mathrm{Pr}}\right) &= \frac{\exp\left[-(1/2)\,\vec{l}^{\,T} \boldsymbol{\Sigma}_{\mathrm{Pr}}^{-1} \vec{l}\,\right]}{(2\pi)^{(2N_f+3)/2} \left|\boldsymbol{\Sigma}_{\mathrm{Pr}}\right|^{1/2}} \end{aligned} \tag{56}$$

for $\vec{l}$, using the $(2N_f+3)\times(2N_f+3)$ dimensional covariance matrix

$$\boldsymbol{\Sigma}_{\mathrm{Pr}} = \begin{pmatrix} a_E \boldsymbol{\Sigma}_{\mathrm{BB}} & 0 & 0 \\ 0 & a_O \boldsymbol{\Sigma}_{\mathrm{BB}} & 0 \\ 0 & 0 & \boldsymbol{\Sigma}_{\vec{b},\mathrm{Pr}} \end{pmatrix}. \tag{57}$$

One gets the joint posterior

$$\begin{aligned} p\left(\vec{l}, \vec{n}, \vec{h}, \Delta_f \mid \vec{D}\right) &= p\left(\vec{l} \mid \vec{n}, \vec{h}, \Delta_f, \vec{D}\right) p\left(\vec{n}, \vec{h}, \Delta_f \mid \vec{D}\right) \end{aligned} \tag{58}$$

with the conditional amplitude posterior given by the multivariate normal

$$\begin{aligned} p\left(\vec{l} \mid \vec{n}, \vec{h}, \Delta_f, \vec{D}\right) &= \frac{\exp\left[-(1/2)\left(\vec{l} - \vec{l}_{\mathrm{Po}}\right)^T \boldsymbol{\Sigma}_{\mathrm{Po}}^{-1} \left(\vec{l} - \vec{l}_{\mathrm{Po}}\right)\right]}{(2\pi)^{(2N_f+3)/2} \left|\boldsymbol{\Sigma}_{\mathrm{Po}}\right|^{1/2}}, \end{aligned} \tag{59}$$

using the posterior covariance matrix $\boldsymbol{\Sigma}_{\mathrm{Po}} = (\mathbf{M}^T \boldsymbol{\Sigma}_D^{-1} \mathbf{M} + \boldsymbol{\Sigma}_{\mathrm{Pr}}^{-1})^{-1}$ and the mean $\vec{l}_{\mathrm{Po}} = \boldsymbol{\Sigma}_{\mathrm{Po}} \mathbf{M}^T \boldsymbol{\Sigma}_D^{-1} \vec{D}$. Furthermore, one obtains the settings posterior

$$\begin{aligned} p\left(\vec{n}, \vec{h}, \Delta_f \mid \vec{D}\right) &= \frac{p_{\times K}}{K} \\ &= \frac{\left|\boldsymbol{\Sigma}_{\mathrm{Po}}\right|^{1/2}}{\left|\boldsymbol{\Sigma}_{\mathrm{Pr}}\right|^{1/2}} \frac{\exp\left[(1/2)\,\vec{l}_{\mathrm{Po}}^{\,T} \boldsymbol{\Sigma}_{\mathrm{Po}}^{-1} \vec{l}_{\mathrm{Po}}\right]}{K}, \end{aligned} \tag{60}$$

where the constant K is unknown so far.

5.3. *Investigation of Posterior*

5.3.1. Conditional Amplitude Posterior for Chosen Settings. To give some insight, the conditional posterior for the amplitudes is evaluated given the specific set of values for x_0, δx, f_L, f_U, Δ_f, a_E and a_O, $\sigma_{b_0,\mathrm{Pr}}$, $\sigma_{b_1,\mathrm{Pr}}$, and $\sigma_{b_2,\mathrm{Pr}}$.

The values for $x_0 = -127.95\Delta x = -5.118$ mm and $\delta x = 0$ mm are chosen to form an optical path difference grid x_i for which all subsequent nodes are separated by Δx. Then, $|x_i| \le 5.118$ mm identifies the double-sided domain (256 data points), and the single-sided domain is marked by 5.118 mm $\le x_i \le 26.362$ mm (532 data points).

The spectral domain is covered from $f_L = 0$ GHz to the Nyquist frequency $f_U = c/(2\Delta x) = 3747.4$ GHz $(=\Delta_F)$. With the bandwidth and the optical path difference set, the conversion $k = |x|\Delta_F/c$ (see Section 2.3.1) to the order of Fourier coefficient follows. Thus, the double-sided and full spatial domain coefficients up to the orders 64 and 330 are located (see Figure 4(b)), respectively. Some examples of the spatial basis functions associated with this set of Fourier coefficients are shown in Figure 2.

To match the maximum order (330) of probed Fourier coefficient, the discretisation is set to $\Delta_f = c/(2\Delta x(256/2 + 532)) \approx 5.68$ GHz which reflects the classical increment. Then, the spectral grid has $N_f = 660$ elements.

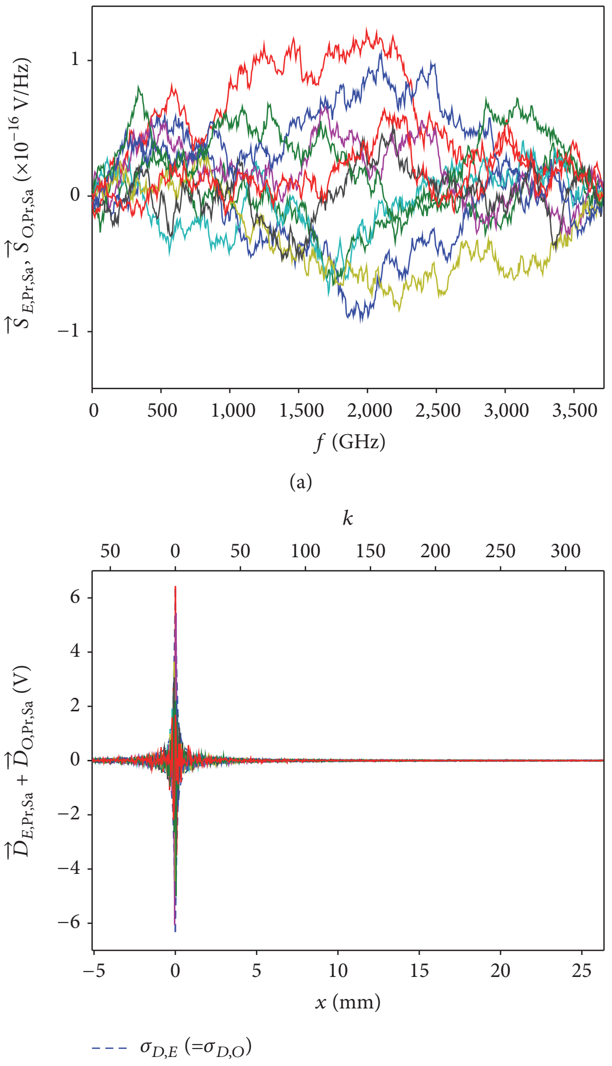
The hyperparameters are set to large values like $a_E = a_O = 10^{-8}\,\mathrm{V}^2$ to prevent from a prior determined posterior. This unwanted feature would be present, if one would choose $a_{E(O)}$ too small, so that the signal envelops do not include the measured data (see next paragraph).

With the chosen spectral priors, sample functions/vectors $\vec{S}_{E,\mathrm{Pr,Sa}}$ and $\vec{S}_{\mathrm{O,Pr,Sa}}$ are drawn and shown in Figure 4(a). The function values are of the order of 10^{-16} V/Hz decaying towards the spectral boundaries. The mapping of the spectral prior samples to the spatial domain via $\vec{D}_{E,\mathrm{Pr,Sa}} + \vec{D}_{\mathrm{O,Pr,Sa}} = \mathbf{M}_E \vec{S}_{E,\mathrm{Pr,Sa}} + \mathbf{M}_{\mathrm{O}} \vec{S}_{\mathrm{O,Pr,Sa}}$ shows that $\vec{D}_{E,\mathrm{Pr,Sa}} + \vec{D}_{\mathrm{O,Pr,Sa}}$ can be of the order of some Volts close to the spatial origin (see Figure 4(b)) which exceeds the measured data by about two orders of magnitude. Furthermore, the amplitude of $\vec{D}_{E,\mathrm{Pr,Sa}} + \vec{D}_{\mathrm{O,Pr,Sa}}$ drops with the distance from the origin. All these data samples are well bounded by the signal envelopes $\sigma_{D,E(O)}$ obtained by $\sigma_{D,E(O)} = (A/2)\sigma'_{D,E(O)}$ which is the scaled version of the expressions (44) and (45).

For $\vec{b}$ the prior covariance is determined by setting $\sigma_{b_0,\mathrm{Pr}} = 1$ V, $\sigma_{b_1,\mathrm{Pr}} = 1$ V/m and $\sigma_{b_2,\mathrm{Pr}} = 1\,\mathrm{V/m}^2$.

(a) Posterior Mean. For the above values, the mean $\vec{l}_{\mathrm{Po}}$ of the amplitude posterior is given by $\vec{b}_{\mathrm{Po}} = (-61.41\,\mathrm{nV}, 5.75\,\mu\mathrm{V/m}, 2.42\,\mu\mathrm{V/m}^2)^T$, $\vec{S}_{E,\mathrm{Po}}$ and $\vec{S}_{\mathrm{O,Po}}$ as presented in Figure 5(a). Clearly, $\vec{S}_{E,\mathrm{Po}}$ peaks below 1000 GHz with the maximum of approximately 6×10^{-18} V/Hz around 400 GHz. Furthermore, $\vec{S}_{\mathrm{O,Po}}$ is much smaller than $\vec{S}_{E,\mathrm{Po}}$ below 1000 GHz.

The offset $\vec{D}_{\mathrm{Off,Po}} = \mathbf{M}_{\mathrm{Off}} \vec{b}_{\mathrm{Po}}$ contribution to the data has an upwards trend but is small (see Figure 5(b)). The



(a)

(b)

FIGURE 4: Examination of Gaussian prior for even and odd spectra modelled by Brownian bridge process. The spectral discretisation $\Delta_f = 5.68$ GHz is used, and the lower and upper spectral limits are set to $f_L = 0$ GHz and $f_U = 3747.4$ GHz (Nyquist assumptions), respectively. The scalings of the processes are chosen by $a_E = a_O = 10^{-8}\,\mathrm{V}^2$. (a) Prior samples $\vec{S}_{E,\mathrm{Pr,Sa}}$ and $\vec{S}_{\mathrm{O,Pr,Sa}}$ for even and odd spectra. The spectra vanish towards the spectral limits. (b) Prior samples mapped to data domain to give even (odd) data-like quantity $\vec{D}_{E(\mathrm{O}),\mathrm{Pr,Sa}}$ for optical path difference obtained for $x_0 = -5.118$ mm and $\delta x = 0$. With increasing distance to the origin, the amplitude drops which is well bounded by the signal envelopes $\sigma_{D,E} = \sigma_{D,O}$ for the chosen bandwidth $\Delta_F = f_U$. Fourier coefficients up to order $k = 64$ and 330 settle in the double-sided domain ($|x| \leq |x_0|$) and single-sided domain ($x > |x_0|$), respectively.

even and odd contributions $\vec{D}_{E(\mathrm{O}),\mathrm{Po}} = \mathbf{M}_{E(\mathrm{O})}\vec{S}_{E(\mathrm{O}),\mathrm{Po}}$ behave differently in the double- and single-sided domains (see Figure 5(b)). While in the double-sided domain $\vec{D}_{E,\mathrm{Po}}$ is much larger than $\vec{D}_{\mathrm{O,Po}}$, $\vec{D}_{E,\mathrm{Po}}$ and $\vec{D}_{\mathrm{O,Po}}$ equal each other in the single-sided domain. This indicates an underestimation (overestimation) of the Fourier coefficients above the order 64 for $S_{E,\mathrm{Po}}$ ($S_{\mathrm{O,Po}}$). The cause of this finding relies in the choice $a_E = a_O$ ($\sigma_{D,E} = \sigma_{D,O}$) which becomes more clear in Section 5.3.3.

Since the quantity $\vec{D}_{\mathrm{Po}} = M\vec{l}_{\mathrm{Po}}$ equals almost the data set $\vec{D}$ (see residuals in Figure 5(c)), a nearly perfect match is achieved. Hence, all noise contributions are captured by the mean $\vec{l}_{\mathrm{Po}}$, and an overfitting of the data is obvious.

(b) Posterior Covariance. The posterior covariance $\mathbf{\Sigma}_{\mathrm{Po}}$ is not characterised in detail. From the elements of the main diagonal of $\mathbf{\Sigma}_{\mathrm{Po}}$, one finds the values $\sigma_{b_0,\mathrm{Po}} = 73.30$ nV, $\sigma_{b_1,\mathrm{Po}} = 12.58\,\mu$V/m, and $\sigma_{b_2,\mathrm{Po}} = 446.67\,\mu\mathrm{V/m}^2$. These posterior uncertainties are quite large when compared to the values of the corresponding mean $\vec{b}_{\mathrm{Po}}$, especially for the quadratic coefficient.

For the spectral quantities, the square root of the main diagonal elements of $\mathbf{\Sigma}_{\mathrm{Po}}$ is of the order of some 10^{-18} V/Hz (for one sample function drawn from the conditional amplitude posterior see Figure 5(a)). Hence, a considerable deviation from the posterior mean is possible.

(c) Posterior Samples. 100 samples $\vec{l}_{\mathrm{Po,Sa}} = (\vec{S}_{E,\mathrm{Po,Sa}}, \vec{S}_{\mathrm{O,Po,Sa}}, \vec{b}_{\mathrm{Po,Sa}})^T$ are drawn from the amplitude posterior and mapped to the data domain (see Figure 5(b)). Thereby, the contributions $\vec{D}_{E,\mathrm{Po,Sa}} = \mathbf{M}_E\vec{S}_{E,\mathrm{Po,Sa}}$, $\vec{D}_{\mathrm{O,Po,Sa}} = \mathbf{M}_{\mathrm{O}}\vec{S}_{\mathrm{O,Po,Sa}}$, and $\vec{D}_{\mathrm{Off,Po,Sa}} = \mathbf{M}_{\mathrm{Off}}\vec{b}_{\mathrm{Po,Sa}}$ are split to investigate their interplay. $\vec{D}_{\mathrm{Off,Po,Sa}}$ has a noticeable width but remains in the vicinity of $\vec{D}_{\mathrm{Off,Po}}$ for the whole spatial domain.

For the double-sided region $\vec{D}_{E,\mathrm{Po,Sa}}$ and $\vec{D}_{\mathrm{O,Po,Sa}}$ form a narrow band around the corresponding posterior mean quantities $\vec{D}_{E,\mathrm{Po}}$ and $\vec{D}_{\mathrm{O,Po}}$. On the contrary, for the single-sided region $\vec{D}_{E,\mathrm{Po,Sa}}$ and $\vec{D}_{\mathrm{O,Po,Sa}}$ increase drastically but are restricted by $\sigma_{D,E}$ and $\sigma_{D,\mathrm{O}}$. The deviation is large between the individual and the posterior mean quantities; though, the sum $\vec{D}_{\mathrm{Po,Sa}} = \vec{D}_{E,\mathrm{Po,Sa}} + \vec{D}_{\mathrm{O,Po,Sa}} + \vec{D}_{\mathrm{Off,Po,Sa}}$ remains well inside the 2-σ_D band surrounding $\vec{D}$ (see Figures 5(c) and 5(d)). Hence, the posterior covariance $\mathbf{\Sigma}_{\mathrm{Po}}$ captures properly the correlations.

With the specific values of the settings one gets the number $\ln p_{\times K} \approx 1.0488 \times 10^6$.

5.3.2. Settings Posterior. As the problem is formulated, the settings posterior is proportional to $p_{\times K}$. Its optimisation, that is, finding the global maximum of the settings posterior is a 10-dimensional problem. Since large numbers are involved, one has to investigate $\ln p_{\times K}$. In general, if only one parameter is changed, a well distinguishable peak in $\ln p_{\times K}$ is found. Hence, the optimisation is currently carried out by varying each parameter separately (coordinate descent algorithm). Thereby, $\ln p_{\times K}$ increases when Δ_f becomes smaller. This implies an increase of the dimensionality of the involved covariances $\mathbf{\Sigma}_{\mathrm{Pr}}$ and $\mathbf{\Sigma}_{\mathrm{Po}}$ and, thus, a prolonging of the

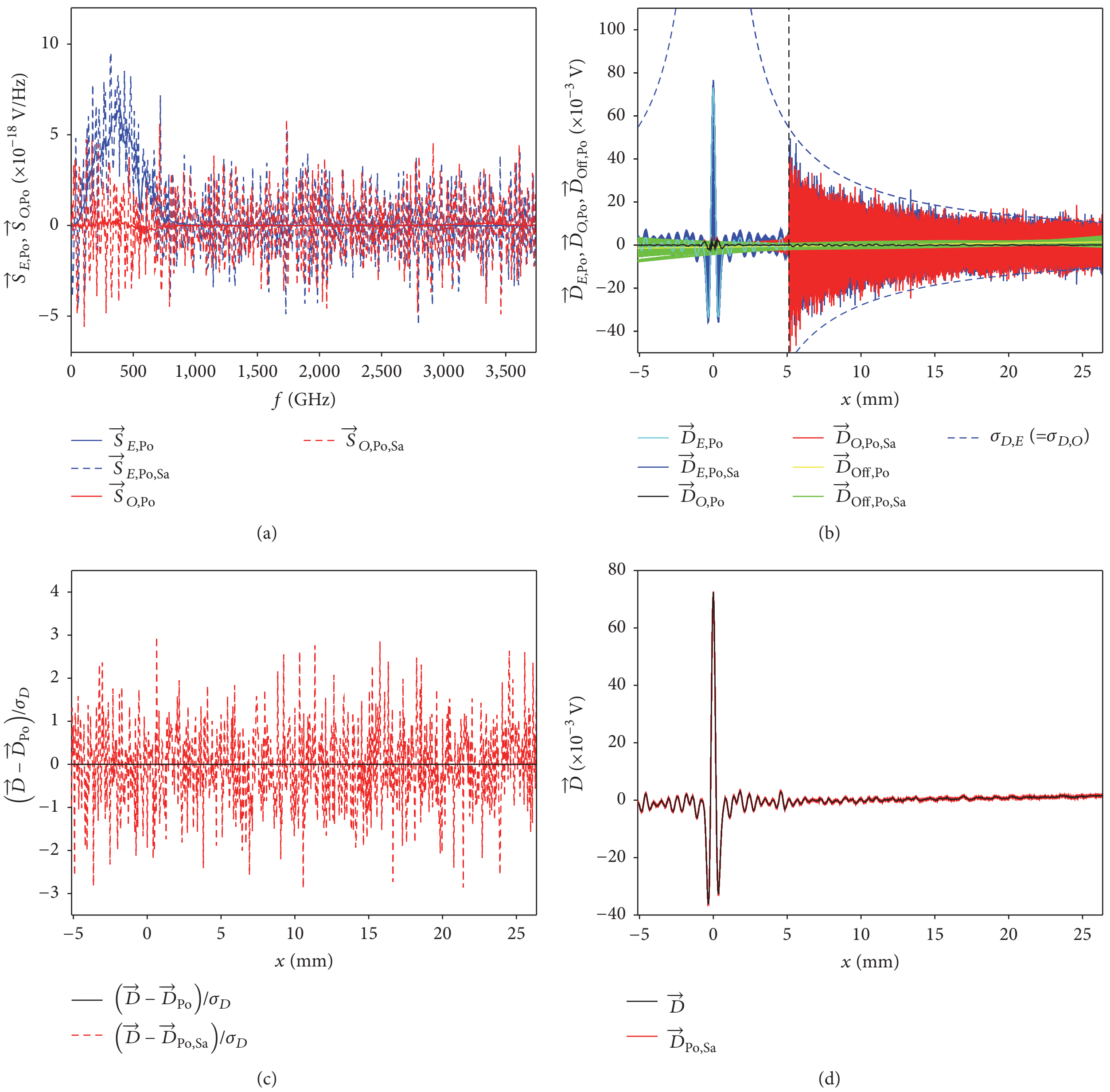
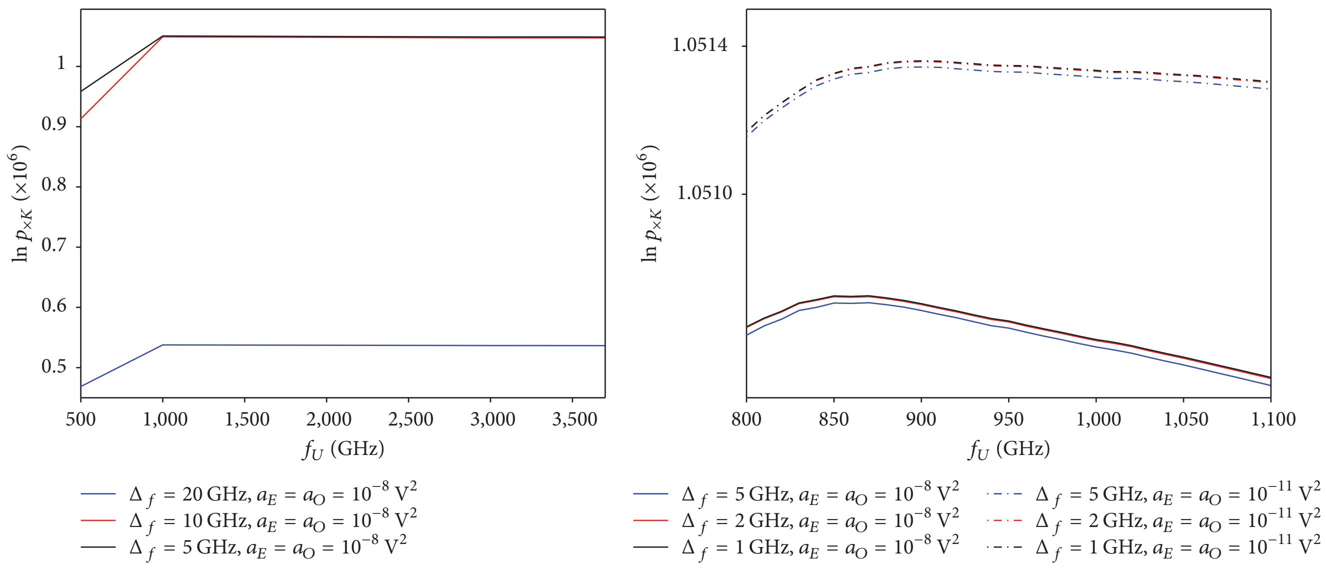


FIGURE 5: Investigation of conditional amplitude posterior $p(\vec{l} \mid \Delta_f, f_L, f_U, a_E, a_O, x_0, \delta x, \sigma_{b_0,\mathrm{Pr}}, \sigma_{b_1,\mathrm{Pr}}, \sigma_{b_2,\mathrm{Pr}}, \vec{D})$ for linear parameters $\vec{l} = (\vec{S}_E, \vec{S}_O, \vec{b})^T$. The posterior is obtained for the settings $x_0 = -5.118$ mm and $\delta x = 0$, $f_L = 0$ GHz, $f_U = 3747.4$ GHz, $\Delta_f = 5.68$ GHz, $a_E = a_O = 10^{-8}\ \mathrm{V}^2$, $\sigma_{b_0,\mathrm{Pr}} = 1$ V, $\sigma_{b_1,\mathrm{Pr}} = 1$ V/m, and $\sigma_{b_2,\mathrm{Pr}} = 1\ \mathrm{V/m}^2$. (a) Spectral posterior means $\vec{S}_{E(O),\mathrm{Po}}$ and one posterior sample. Below 1000 GHz, the means show a finite amplitude. Each sample deviates strongly from its corresponding mean. (b) Posterior mean $\vec{l}_{\mathrm{Po}}$ and 100 samples $\vec{l}_{\mathrm{Po,Sa}}$ mapped to give even, odd, and offset contributions in data domain. While in the double-sided region ($|x| \leq |x_0|$) most of the data described by $\vec{S}_{E,\mathrm{Po}}$, $\vec{S}_{E,\mathrm{Po}}$ and $\vec{S}_{O,\mathrm{Po}}$ give same contributions in the single-sided region ($x > |x_0|$). In the single-sided region, the spread of the even (odd) contributions for the posterior samples is large but bounded by the signal envelops $\sigma_{D,E(O)}$. (c) Residuals for $\vec{l}_{\mathrm{Po}}$ (black) and one sample $\vec{l}_{\mathrm{Po,Sa}}$ (red) both mapped to data domain. An almost perfect description of the noisy data is achieved by l_{Po}, indicating an overfitting of the data. (d) Measured data $\vec{D}$ and 100 samples of amplitude posterior mapped to data domain. The individual contributions have a much wider spread (see (b)), but the total sum $\vec{D}_{\mathrm{Po,Sa}}$ agrees well with $\vec{D}$ and its uncertainties. This is caused by the posterior covariance of $\vec{l}$ with suited off-diagonal entries.



(a) (b)

FIGURE 6: Ten-dimensional quantity $p_{\times K}$ scanned at different values for f_U, Δ_f, a_E, and a_O while keeping $f_L = 0$ GHz, $x_0 = -5.118$ mm, $\delta x = 0$ mm, $\sigma_{b_0,\mathrm{Pr}} = 1$ V, $\sigma_{b_1,\mathrm{Pr}} = 1$ V/m, and $\sigma_{b_2,\mathrm{Pr}} = 1$ V/m^2. (a) Rough scan in f_U and Δ_f at $a_E = a_O = 10^{-8}$ V^2. A peak for $\ln p_{\times k}$ with the amplitude of the order of 10^6 is found close to $f_U = 1000$ GHz. Small Δ_f are in favour. (b) Scan in f_U and Δ_f close to peaking for two sets of a_E and a_O. The peak locates in the vicinity of 910 GHz when a_E and a_O reduce by some orders of magnitude.

TABLE 1: Characterisation of maxima of $p_{\times K}$ dependent on Δ_f. The maxima locate at similar nine-dimensional parameter sets. The odds are evaluated by the ratio $p_{\times K}(\Delta_f)/p_{\times K}(\Delta_f = 1/2\,\mathrm{GHz})$ and relate the maxima globally. Since the odds are rising, the global maximum locates below $\Delta_f = 1/2$ GHz.

Δ_f (GHz)	≈5.68	4	3	2	1	1/2
$f_{L,\mathrm{Po}}$ (GHz)	33.499	33.243	33.277	32.817	32.445	32.425
$f_{U,\mathrm{Po}}$ (GHz)	913.572	913.243	912.277	912.817	913.445	913.425
$a_{E,\mathrm{Po}}$ $(\times 10^{-12}\,\mathrm{V}^2)$	0.9152	1.0127	1.0623	1.1002	1.1224	1.1299
$a_{O,\mathrm{Po}}$ $(\times 10^{-15}\,\mathrm{V}^2)$	5.7381	5.7731	5.7776	5.797	5.8126	5.8145
$x_{0,\mathrm{Po}}$ (mm)	−5.115293	−5.115297	−5.115288	−5.115291	−5.115295	−5.115295
δx_{Po} (μm)	−4.814	−4.814	−4.814	−4.814	−4.814	−4.814
$\sigma_{b_0,\mathrm{Pr},\mathrm{Po}}$ (nV)	62.150	62.148	62.146	62.143	62.142	62.142
$\sigma_{b_1,\mathrm{Pr},\mathrm{Po}}$ (μV/m)	6.031	6.003	6.003	6.000	5.998	5.998
$\sigma_{b_2,\mathrm{Pr},\mathrm{Po}}$ (μV/m^2)	11.541	9.884	9.913	9.725	9.575	9.573
$\ln p_{\times K}$	1052035.05	1052037.54	1052038.52	1052039.15	1052039.48	1052039.55
Odds	0.011	0.134	0.357	0.668	0.929	1

optimisation procedure. To demonstrate this procedure, the parameters $f_L = 0$ GHz, $a_E = a_O = 10^{-8}$ V^2, $x_0 = -5.118$ mm, $\delta x = 0$ mm, $\sigma_{b_0,\mathrm{Pr}} = 1$ V, $\sigma_{b_1,\mathrm{Pr}} = 1$ V/m, and $\sigma_{b_2,\mathrm{Pr}} = 1$ V/m^2 are set to the values used for the nonoptimised case (see Section 5.3.1). Scanned roughly in f_U, $\ln p_{\times K}$ shows a peak close to 1000 GHz (see Figure 6(a)). Furthermore, this peak increases for $\Delta_f \leq 5$ GHz to $\ln p_{\times K} \approx 10^6$. The peak is localised at 860 GHz which moves to about 910 GHz for the reduced values $a_E = a_O = 10^{-11}$ V^2 (see Figure 6(b)).

To ease the computational effort but still being able to characterise $\ln p_{\times K}$, its maximum is determined dependent on Δ_f which is scanned in the values 5.68, 4, 3, 2, 1, and 1/2 GHz. Each maximum is captured by the sets $f_{L,\mathrm{Po}}$, $f_{U,\mathrm{Po}}$, $a_{E,\mathrm{Po}}$, $a_{O,\mathrm{Po}}$, $x_{0,\mathrm{Po}}$, δx_{Po}, $\sigma_{b_0,\mathrm{Pr},\mathrm{Po}}$, $\sigma_{b_1,\mathrm{Pr},\mathrm{Po}}$, and $\sigma_{b_2,\mathrm{Pr},\mathrm{Po}}$ summarised by Table 1. All maxima locate at very similar sets which seem to converge as Δ_f becomes smaller. Although, in relative terms, the maximum for a smaller Δ_f has higher odds (see Figure 7). For example, the odds read 1 : 0.011 when the maxima at $\Delta_f = 1/2$ GHz and 5.68 GHz are related. Furthermore, taking the values for the maximum at $\Delta_f = 1/2$ GHz to evaluate $\ln p_{\times K}$ at $\Delta_f = 1/3$ GHz and 1/4 GHz gives the odds 1 : 1.013 : 1.018. Thus, the global maximum locates somewhere in the range below $\Delta_f \leq 1/4$ GHz. But this range cannot be investigated in more detail from numerical point of view. However, the increase in $\ln p_{\times K}$ when Δ_f is decreased below 1/2 GHz is interpreted as confirmation that a continuous spectrum is indeed probed.

TABLE 2: Posterior standard deviations dependent on Δ_f for spectral boundaries f_L and f_U, scalings a_E and a_O of prior covariances, zero-path difference x_0, and shift δx. The standard deviations are estimated from the inverse Hessian matrix at the corresponding maximum of $p_{\times K}$.

Δ_f (GHz)	≈5.68	4	3	2	1	1/2
$\sigma_{f_L,\mathrm{Po}}$ (GHz)	1.742	1.795	1.746	1.762	1.754	1.718
$\sigma_{f_U,\mathrm{Po}}$ (GHz)	3.101	3.155	2.927	2.869	2.858	2.810
$\sigma_{a_E,\mathrm{Po}}$ ($\times 10^{-12}$ V^2)	0.1064	0.1179	0.1242	0.1290	0.1316	0.1327
$\sigma_{a_O,\mathrm{Po}}$ ($\times 10^{-15}$ V^2)	1.5912	1.6064	1.6112	1.6164	1.6199	1.6211
$\sigma_{x_0,\mathrm{Po}}$ (μm)	1.207	1.207	1.207	1.206	1.205	1.205
$\sigma_{\delta x,\mathrm{Po}}$ (μm)	0.170	0.170	0.170	0.170	0.170	0.170

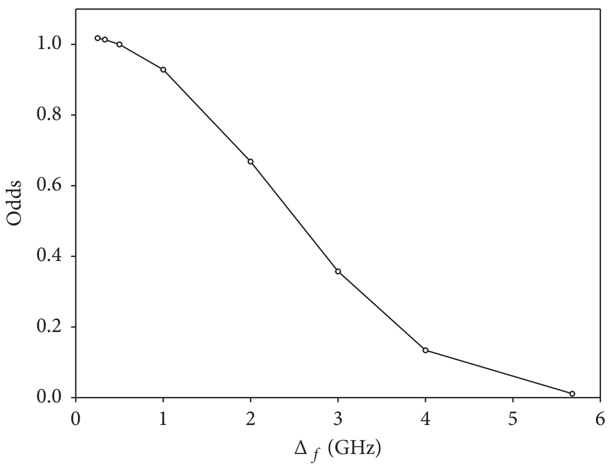


FIGURE 7: Odds for maxima of $p_{\times K}$ dependent on spectral increment Δ_f. The maximum at $\Delta_f = 1/2$ GHz is used as a reference (unity odd). For $\Delta_f = 1/3$ GHz and 1/4 GHz, the maximising parameter set (see Table 1) for $\Delta_f = 1/2$ GHz is used to evaluate the associated odds. The global maximum of $p_{\times K}$ is obtained for very small spectral increments.

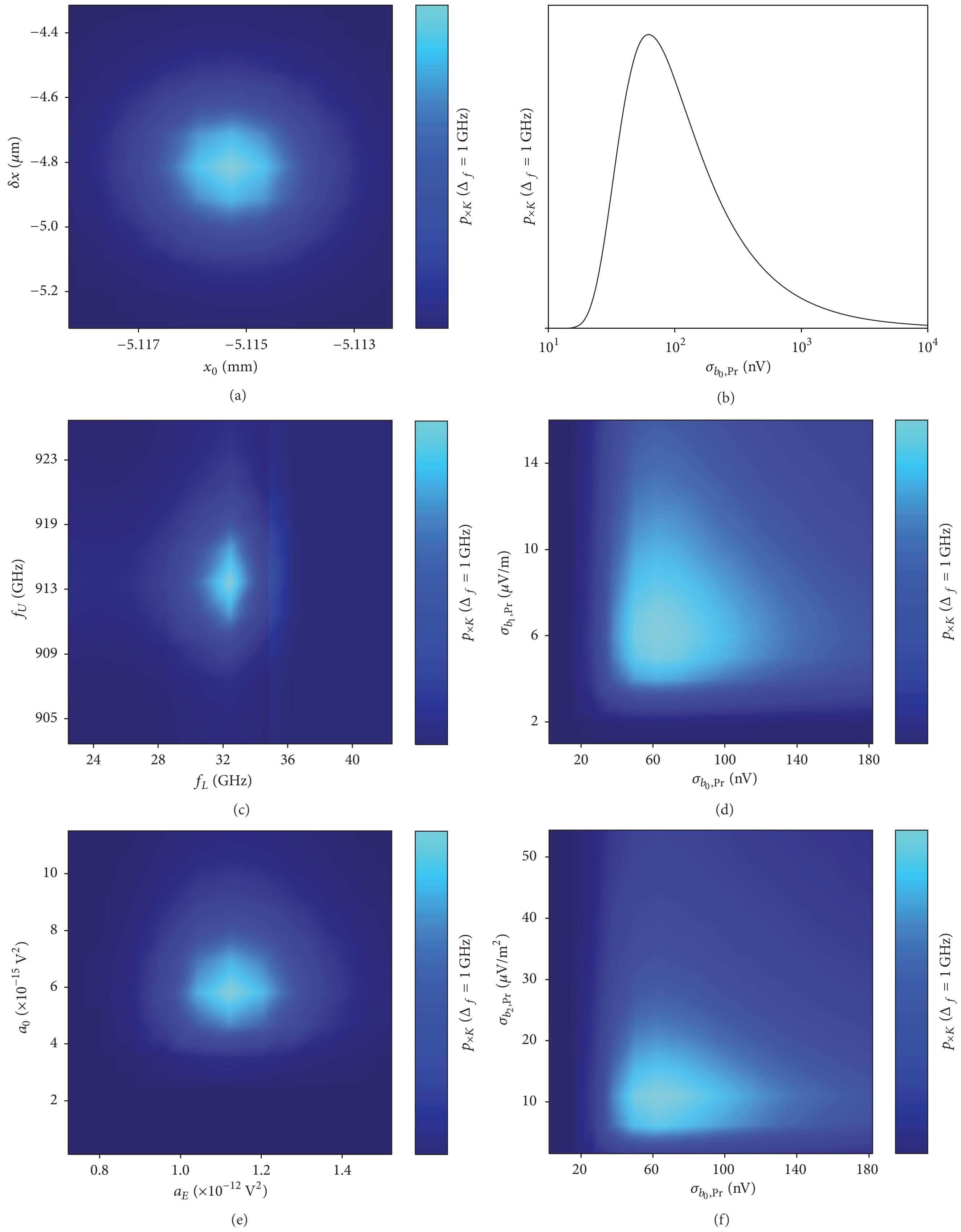
For $\Delta_f = 5.68$ GHz, the odds read $1:10^{-1390}$ when the associated maximum ($\ln p_{\times K} \approx 1.0520 \times 10^6$) is related to the nonoptimised case ($\ln p_{\times K} \approx 1.0488 \times 10^6$) investigated in Section 5.3.1. This is caused by having chosen some parameters like $a_{E(O)}$ several orders of magnitude too large with respect to the maximum settings values.

For $\Delta_f = 1$ GHz, Figures 8(a)–8(f) summarise scans in the parameter pairs (a_E, a_O), (f_L, f_U), $(x_0, \delta x)$, $(\sigma_{b_0,\mathrm{Pr}}, \sigma_{b_1,\mathrm{Pr}})$, $(\sigma_{b_0,\mathrm{Pr}}, \sigma_{b_2,\mathrm{Pr}})$ and in $\sigma_{b_0,\mathrm{Pr}}$ while the remaining parameters are held at the maximum values. In $\sigma_{b_0,\mathrm{Pr}}$, $\sigma_{b_1,\mathrm{Pr}}$, and $\sigma_{b_2,\mathrm{Pr}}$, a skewed distribution is found. The remaining six-dimensional posterior distribution has a high probability in a narrow region for a given Δ_f. This distribution is well approximated by a multivariate normal. Its mean is given by the maximum values listed in Table 1. The posterior covariance is estimated from the inverse of the Hessian matrix evaluated numerically via the second-order partial derivatives in the vicinity of the maximum. The off-diagonal elements of this covariance are negligible, so that one can factorise the posterior as a product of individual Gaussians. The posterior standard deviations $\sigma_{f_L,\mathrm{Po}}$, $\sigma_{f_U,\mathrm{Po}}$, $\sigma_{a_E,\mathrm{Po}}$, $\sigma_{a_O,\mathrm{Po}}$, $\sigma_{x_0,\mathrm{Po}}$, and $\sigma_{\delta x,\mathrm{Po}}$ vary little when Δ_f is changed (see Table 2). The spectral boundaries are well determined within an interval of some GHz. While the uncertainty in the scaling a_E is small compared to its posterior mean value, a_O is more uncertain. The uncertainty in the zero-path difference is of the order of some μm, and the shift δx is quite certain within some hundreds of nanometers.

5.3.3. Conditional Amplitude Posterior for Maximising Settings. In the following, the amplitude posterior is investigated for the maximising settings given $\Delta_f = 1/2$ GHz (see Table 1). For the listed settings, the double-sided domain is identified by $|x| \leq |x_{0,\mathrm{Po}}| \approx 5.115$ mm; the single-sided domain is bounded by the lower limit $x > |x_{0,\mathrm{Po}}|$ and the upper limit of about 26.36 mm, the centre frequency $f_{C,\mathrm{Po}}$, and the bandwidth $\Delta_{F,\mathrm{Po}}$ of the spectral domain read 472.925 GHz and 881 GHz, respectively. In addition, the conversion of the spatial coordinate to the order of the Fourier coefficients $k = |x|\Delta_{F,\mathrm{Po}}/c$ reveals that the double- and single-sided domain contain information up to the 15th and 77th order (see Figures 9(c) and 9(d)), respectively. From this follows that the interferometric part of the 256/788 data values in the double-sided/full domain are modelled best by 62/310 Fourier coefficients and their associated basis functions dependent on $f_{C,\mathrm{Po}}$, $\Delta_{F,\mathrm{Po}}$, and x. The signal envelopes $\sigma_{D,E}$ and $\sigma_{D,O}$ are evaluated with $a_{E,\mathrm{Po}}$ and $a_{O,\mathrm{Po}}$.

(a) Posterior Mean. Given the boundaries and increment, the spectral dimension reads $N_f = 1762$, and, thus, the amplitude posterior mean $\vec{l}_{\mathrm{Po}} = (\vec{S}_{E,\mathrm{Po}}, \vec{S}_{O,\mathrm{Po}}, \vec{b}_{\mathrm{Po}})^T$ is 3527-dimensional. The posterior mean values for the three coefficients $b_{0,\mathrm{Po}} = -62.142$ nV, $b_{1,\mathrm{Po}} = 5.997\,\mu$V/m, and $b_{2,\mathrm{Po}} = -8.641\,\mu$V/m^2 have magnitudes which are close to the corresponding prior standard deviations at the maximum of the settings posterior (see Table 1). Furthermore, the absolute and the linear mean values are similar to the ones obtained for the nonoptimised case (see Section 5.3.1). The offset $\vec{D}_{\mathrm{Off,Po}}$ which follows from mapping $\vec{b}_{\mathrm{Po}}$ is shown in Figures 9(b) and 9(d).

The posterior means $\vec{S}_{E,\mathrm{Po}}$ and $\vec{S}_{O,\mathrm{Po}}$ presented in Figure 9(a) approach zero towards the spectral boundaries as expected. Furthermore, $\vec{S}_{E,\mathrm{Po}}$ is much larger in amplitude than $\vec{S}_{O,\mathrm{Po}}$ anticipated by the quite different posterior mean scalings $a_{E,\mathrm{Po}}$ and $a_{O,\mathrm{Po}}$. Hence, the even process describes most of the interferometric data in the double-sided domain.



(a) (b) (c) (d) (e) (f)

FIGURE 8: $p_{\times K}$ scanned in individual nonlinear parameters and hyperparameters in vicinity of maximum of $p_{\times K}$ for $\Delta_f = 1$ GHz (see Table 1). (a) Zero-path difference position x_0 and shift δx. (b) Prior standard deviation $\sigma_{b_0,\mathrm{Pr}}$ for absolute coefficient. (c) Lower and upper limits f_L and f_U of spectral domain. (d) $\sigma_{b_0,\mathrm{Pr}}$ and prior standard deviation $\sigma_{b_1,\mathrm{Pr}}$ for linear coefficient. (e) Scaling a_E and scaling a_O of Brownian bridge priors for $\overrightarrow{S}_E$ and $\overrightarrow{S}_O$. (f) $\sigma_{b_0,\mathrm{Pr}}$ and prior standard deviation $\sigma_{b_2,\mathrm{Pr}}$ for quadratic coefficient.

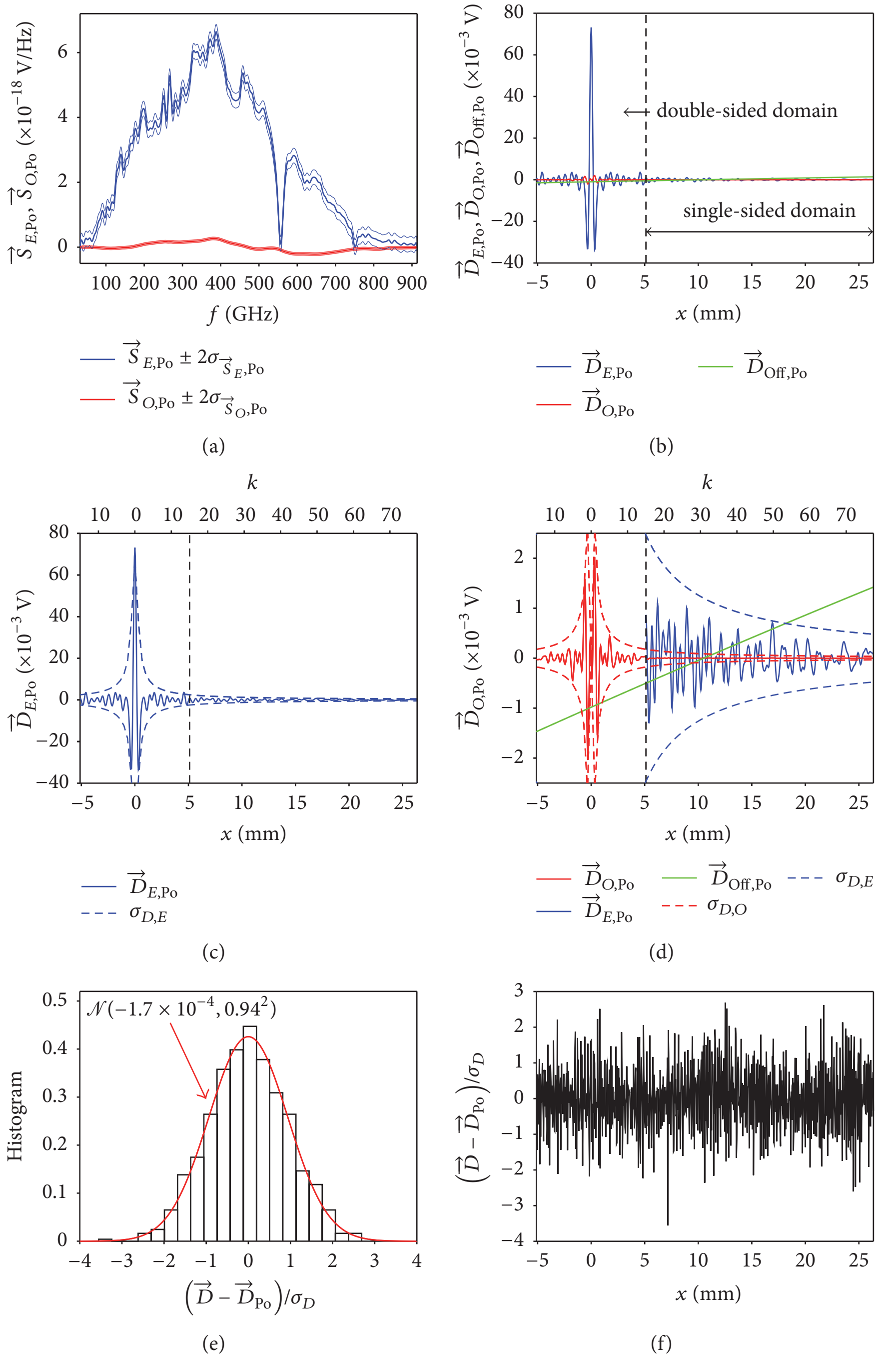


Figure 9: (a) Posterior means $\vec{S}_{E,\mathrm{Po}}$ and $\vec{S}_{\mathrm{O,Po}}$ with associated 2-σ uncertainty bands evaluated at maximum of settings posterior for $\Delta_f = 1/2$ GHz (see Table 1). Most likely, the spectral quantities $\vec{S}_{E,\mathrm{Po}}$ and $\vec{S}_{\mathrm{O,Po}}$ are finite only for the domain between $f_{L,\mathrm{Po}} = 32.425$ GHz and $f_{U,\mathrm{Po}} = 913.425$ GHz. This upper limit is approximately 4 smaller than the Nyquist frequency $f_{\mathrm{Ny}} = 3747.4$ GHz. The dips in $\vec{S}_{E,\mathrm{Po}}$ at 557 and 752 GHz are caused by absorption due to atmospheric water vapour. (b) $\vec{D}_{E,\mathrm{Po}}$, $\vec{D}_{\mathrm{O,Po}}$, and $\vec{D}_{\mathrm{Off,Po}}$ obtained by mapping posterior means $\vec{S}_{E,\mathrm{Po}}$, $\vec{S}_{\mathrm{O,Po}}$, $b_{0,\mathrm{Po}} = -62.142$ nV, $b_{1,\mathrm{Po}} = 5.997\ \mu$V/m, and $b_{2,\mathrm{Po}} = -8.641\ \mu\mathrm{V/m}^2$ to data domain. (c) Cosine transform $\vec{D}_{E,\mathrm{Po}}$ of $\vec{S}_{E,\mathrm{Po}}$ dependent on optical path difference x and order $k = |x|\Delta_{F,\mathrm{Po}}/c$ (bandwidth $\Delta_{F,\mathrm{Po}} = f_{U,\mathrm{Po}} - f_{L,\mathrm{Po}}$) of Fourier coefficients. The envelope $\sigma_{D,E} \propto a_{E,\mathrm{Po}}^{1/2}/k$, incorporating the scaling $a_{E,\mathrm{Po}} = 1.1299 \times 10^{-12}\ \mathrm{V}^2$, bounds $\vec{D}_{E,\mathrm{Po}}$ best. (d) Zoom of (b) dependent on x and k. Globally, the sine transform $\vec{D}_{\mathrm{O,Po}}$ of $\vec{S}_{\mathrm{O,Po}}$ is best summarised by the envelope $\sigma_{D,O} \propto a_{\mathrm{O,Po}}^{1/2}/k$ with the scaling $a_{\mathrm{O,Po}} = 5.8145 \times 10^{-15}\ \mathrm{V}^2$. Since $\sigma_{D,E} \gg \sigma_{D,O}$ holds, the single-sided domain is determined mainly by $\vec{S}_{E,\mathrm{Po}}$. (e) Histogram of residuals, that is, difference between measured data $\vec{D}$ and posterior mean prediction $\vec{D}_{\mathrm{Po}} = \vec{D}_{E,\mathrm{Po}} + \vec{D}_{\mathrm{O,Po}} + \vec{D}_{\mathrm{Off,Po}}$ scaled by noise level σ_D. The data is well described, because the mean -1.7×10^{-4} and the variance 0.94^2 of the residuals are almost vanishing and close to unity, respectively. (f) Residuals versus optical path difference.

This is seen by mapping both spectral means to the data domain which gives the even and odd quantities $\vec{D}_{E,\mathrm{Po}}$ and $\vec{D}_{\mathrm{O,Po}}$ (see Figures 9(b)–9(d)). In addition, the single-sided domain is described mainly by $\vec{S}_{E,\mathrm{Po}}$, because in this domain only the envelope $\sigma_{D,E}$ for the even process is of the order of the interferometric data (see Figure 9(d)). This differs from the findings for the nonoptimised case ($a_E = a_O$) for which $\vec{S}_{E,\mathrm{Po}}$ and $\vec{S}_{\mathrm{O,Po}}$ determine the single-sided region almost equally.

The histogram of the residuals (see Figure 9(e)) is approximated very well by the normal distribution $\mathcal{N}(-1.7 \times 10^{-4}, 0.94^2)$. Since the mean vanishes almost, and the standard deviation is very close to unity, the data set is well described by the model and the posterior means. Most likely, the data point located at $x \approx 7\,\mathrm{mm}$ (see Figure 9(f)) is an outlier, because its residual is outside the 3.5-σ_D band.

(b) Posterior Covariance. For the linear parameters the posterior covariance matrix is written like

$$\boldsymbol{\Sigma}_{\mathrm{Po}} = \begin{pmatrix} \boldsymbol{\Sigma}_{\vec{S}_{E,E},\mathrm{Po}} & \boldsymbol{\Sigma}_{\vec{S}_{E,O},\mathrm{Po}} & \boldsymbol{\Sigma}_{\vec{S}_E,\vec{b},\mathrm{Po}} \\ \boldsymbol{\Sigma}^T_{\vec{S}_{E,O},\mathrm{Po}} & \boldsymbol{\Sigma}_{\vec{S}_{O,O},\mathrm{Po}} & \boldsymbol{\Sigma}_{\vec{S}_O,\vec{b},\mathrm{Po}} \\ \boldsymbol{\Sigma}^T_{\vec{S}_E,\vec{b},\mathrm{Po}} & \boldsymbol{\Sigma}^T_{\vec{S}_O,\vec{b},\mathrm{Po}} & \boldsymbol{\Sigma}_{\vec{b},\vec{b},\mathrm{Po}} \end{pmatrix}, \tag{61}$$

and taking the square root of an element of the main diagonal gives the posterior standard deviation σ_{P_i} for the ith parameter. The correlation coefficient between two parameters P_1 and P_2 can be evaluated by $\rho_{P_1,P_2} = \boldsymbol{\Sigma}_{P_1,P_2,\mathrm{Po}}/(\sigma_{P_1}\sigma_{P_2})$.

The standard deviations $\sigma_{b_0,\mathrm{Po}} = 0.501\,\mathrm{nV}$, $\sigma_{b_1,\mathrm{Po}} = 0.093\,\mu\mathrm{V/m}$, and $\sigma_{b_2,\mathrm{Po}} = 4.119\,\mu\mathrm{V/m}^2$ of the three coefficients are reduced by about two orders of magnitude with respect to the uncertainties obtained for the nonoptimised case. Though, the quadratic coefficient is inferred with small confidence. Coefficients of neighbouring degree have the highest correlation in magnitude and are anticorrelated. For example, one finds $\rho_{b_0,b_1} = -0.55$ and $\rho_{b_1,b_2} = -0.93$, while $\rho_{b_0,b_2} = 0.29$ remains small.

The covariances $\boldsymbol{\Sigma}_{\vec{S}_{E,E},\mathrm{Po}}(f, f')$ and $\boldsymbol{\Sigma}_{\vec{S}_{O,O},\mathrm{Po}}(f, f')$ and their associated correlations $\rho_{E,E}(f, f')$ and $\rho_{O,O}(f, f')$ resemble sinc functions centred along the line $f = f'$ plus a tip at this condition (see Figures 10(a)–10(c) for cross section at $f' = 100.175\,\mathrm{GHz}$). Though, the FWHM widths of 1.5 GHz and 10 GHz are very different for the even and odd covariances/correlations. Regarding the covariances, the tip heights account for $1.3 \times 10^{-38}\,\mathrm{V}^2/\mathrm{Hz}^2$ and $4.4 \times 10^{-40}\,\mathrm{V}^2/\mathrm{Hz}^2$, and towards the spectral boundaries the heights are approaching zero which can be seen by the standard deviations $\sigma_{\vec{S}_E,\mathrm{Po}} = \boldsymbol{\Sigma}^{1/2}_{\vec{S}_{E,E},\mathrm{Po}}(f, f)$ and $\sigma_{\vec{S}_O,\mathrm{Po}} = \boldsymbol{\Sigma}^{1/2}_{\vec{S}_{O,O},\mathrm{Po}}(f, f)$ (see Figure 10(d)).

At the condition $f = f'$, the covariance $\boldsymbol{\Sigma}_{\vec{S}_{E,O},\mathrm{Po}}$ and the correlation $\rho_{E,O}$ vanish and are asymmetrically close to this line (see Figures 10(b) and 10(c) for cross section at $f' = 100.175\,\mathrm{GHz}$). This means that at a given frequency f the amplitudes $\vec{S}_E$ and $\vec{S}_O$ are independent but have a finite and opposite correlation with the neighbouring spectral domains to either side.

The correlations $\rho_{E,\vec{b}}$ and $\rho_{O,\vec{b}}$ are oscillatory and rise towards the spectral origin (see Figures 10(e) and 10(f)).

(c) Posterior Samples. From the conditional amplitude posterior samples $\vec{S}_{E,\mathrm{Po,Sa}}$, $\vec{S}_{\mathrm{O,Po,Sa}}$ and $\vec{b}_{\mathrm{Po,Sa}}$ are drawn. While Figure 11(a) shows one sample for each of the even and odd spectral functions, Figure 11(b) presents 100 samples. The samples form a band around the corresponding posterior mean $\vec{S}_{E(O),\mathrm{Po}}$ with the width of about twice the standard deviation $\sigma_{\vec{S}_{E(O)},\mathrm{Po}}$ as shown in Figure 10(d). Hence, the band for the odd function is much smaller.

The samples mapped to the data domain give $\vec{D}_{E,\mathrm{Po,Sa}}$, $\vec{D}_{\mathrm{O,Po,Sa}}$, and $\vec{D}_{\mathrm{Off,Po,Sa}}$ (see Figure 11(c)) which form much more narrow bands when compared to the nonoptimised case (see Figure 5(b)). In particular, the transition from the double- to the single-sided domain is smooth. This is a consequence of having used the most likely scalings $a_{E,\mathrm{Po}}$ and $a_{\mathrm{O,Po}}$ and the associated but quite different signal envelops inferred from the data. Furthermore, this explains why the band for $\vec{D}_{\mathrm{O,Po,Sa}}$ is much smaller than the one for $\vec{D}_{E,\mathrm{Po,Sa}}$.

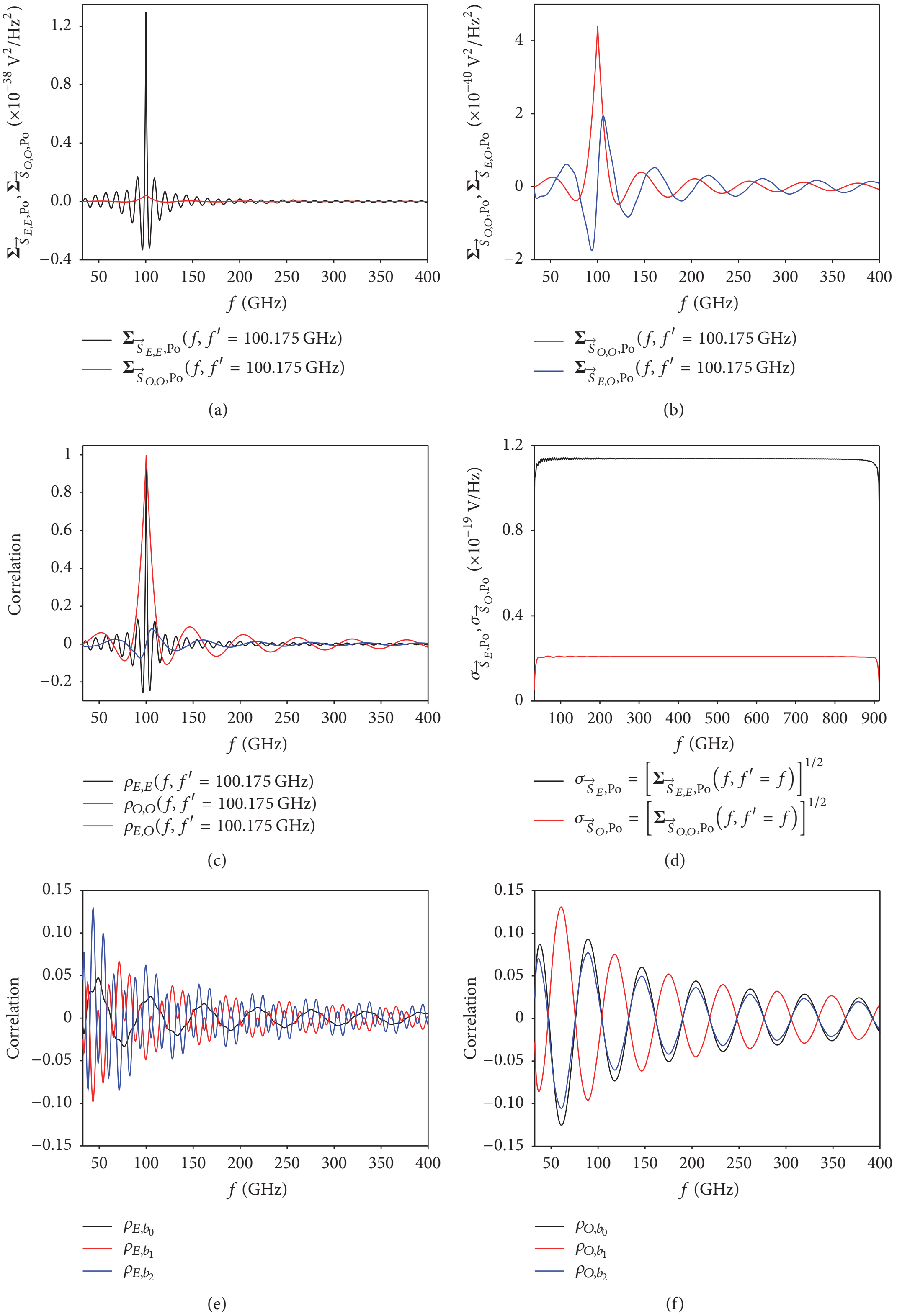
For each sample, the sum $\vec{D}_{\mathrm{Po,Sa}} = \vec{D}_{E,\mathrm{Po,Sa}} + \vec{D}_{\mathrm{O,Po,Sa}} + \vec{D}_{\mathrm{Off,Po,Sa}}$ complies with the measured data $\vec{D}$ and the noise level (see Figure 11(d)).

With the samples $\vec{S}_{E,\mathrm{Po,Sa}}$ and $\vec{S}_{\mathrm{O,Po,Sa}}$ at hand, $\vec{D}_{E,\mathrm{Po,Sa}}$ and $\vec{D}_{\mathrm{O,Po,Sa}}$ are predicted outside the probed spatial domain. Up to the optical path difference of 70 mm, Figures 11(e) and 11(f) show $\vec{D}_{E,\mathrm{Po,Sa}}$ and $\vec{D}_{\mathrm{O,Po,Sa}}$ for one sample and 100 samples, respectively. No difference is obvious in $\vec{D}_{E,\mathrm{Po,Sa}}$ and $\vec{D}_{\mathrm{O,Po,Sa}}$ for an individual prediction in the probed and nonprobed domain. However, many samples take values between $\sigma_{D,E(O)}$ outside the probed region. Since the spread in the probed domain is smaller than in the nonprobed domain, the main uncertainty in $\vec{S}_{E,\mathrm{Po}}$ and $\vec{S}_{\mathrm{O,Po}}$ originates in Fourier coefficients not probed rather than the noise on measured data.

5.3.4. Marginal Amplitude Posterior. Since only the marginal posterior $p(\vec{S}_E, \vec{S}_O \mid \Delta_f, \vec{D})$ of the spectral functions is of interest for this specific Fourier spectroscopy problem, formally the marginalisation

$$\begin{aligned} p\left(\vec{S}_E, \vec{S}_O \mid \Delta_f, \vec{D}\right) &= \int p\left(\vec{S}_E, \vec{S}_O, \vec{b} \mid \vec{n}, \vec{h}, \Delta_f, \vec{D}\right) \\ &\quad \cdot p\left(\vec{n}, \vec{h} \mid \Delta_f, \vec{D}\right) d\vec{b}\, d\vec{n}\, d\vec{h} \end{aligned} \tag{62}$$

needs to be carried out. In general, the pdf $p(\vec{S}_E, \vec{S}_O \mid \Delta_f, \vec{D})$ is not Gaussian but can be approximated by a multivariate normal. Performing the marginalisation, that is,

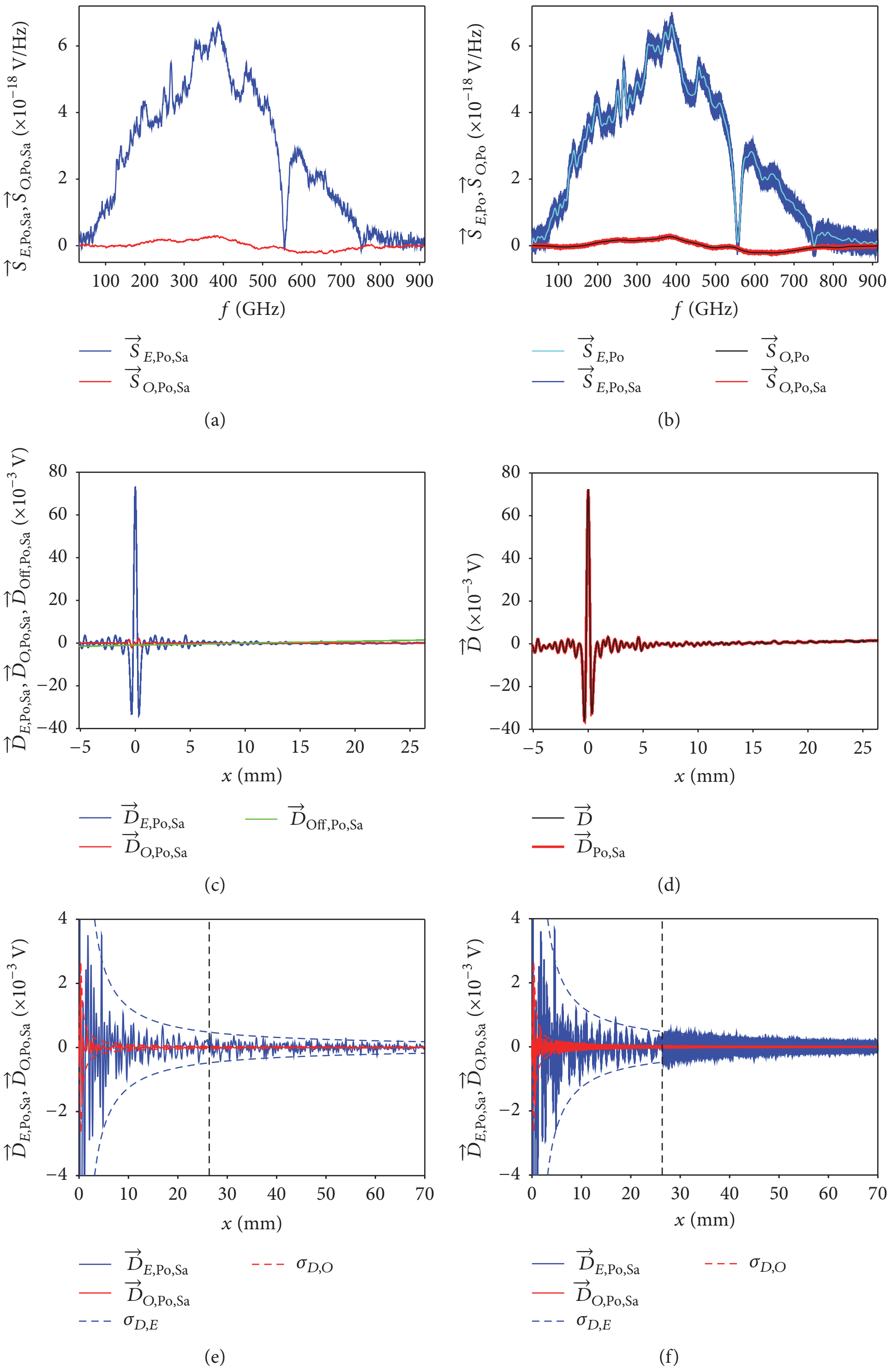


(a) (b)

(c) (d)

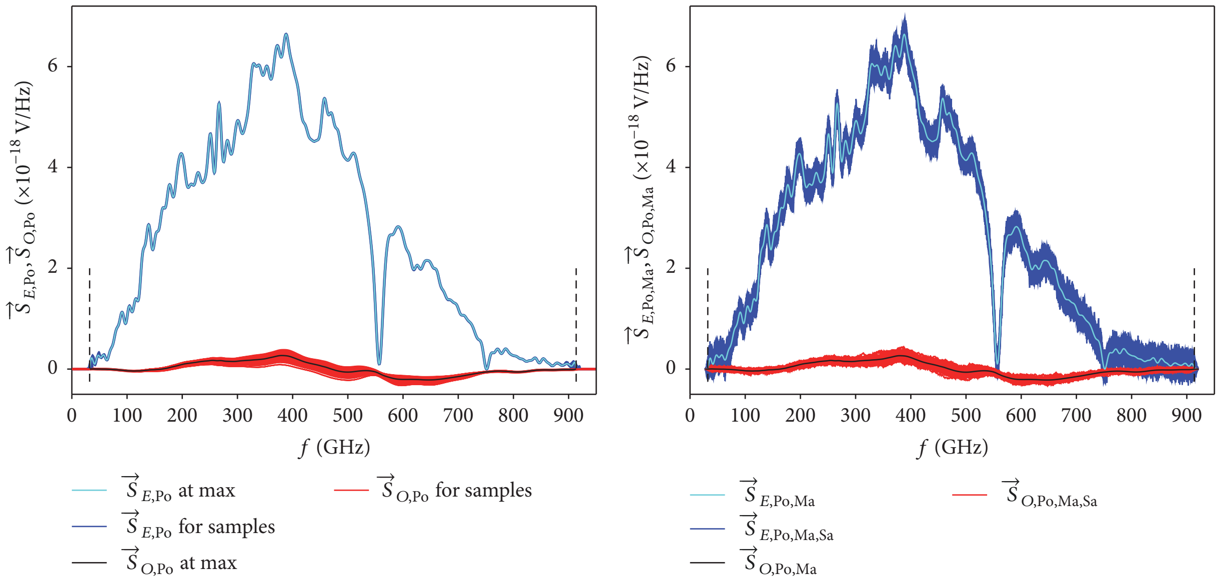
(e) (f)

FIGURE 10: Characterisation of posterior covariance for amplitude parameters obtained at maximum of settings posterior for $\Delta_f = 1/2$ GHz (see Table 1). (a) Cross section of covariances $\boldsymbol{\Sigma}_{\vec{S}_{E,E},\mathrm{Po}}$ and $\boldsymbol{\Sigma}_{\vec{S}_{O,O},\mathrm{Po}}$ for even and odd function. Each covariance is similar to a sinc function centred at $f = f'$, where a tip is added on top. (b) Cross section of covariances $\boldsymbol{\Sigma}_{\vec{S}_{O,O},\mathrm{Po}}$ and $\boldsymbol{\Sigma}_{\vec{S}_{E,O},\mathrm{Po}}$. At a fixed frequency, $\vec{S}_E$ and $\vec{S}_O$ are independent, because $\boldsymbol{\Sigma}_{\vec{S}_{E,O},\mathrm{Po}} = 0$. But finite correlation remains with the neighbouring spectral domain. (c) Correlation. The even (odd) parameters are correlated ($\rho_{E(O),E(O)} > 0.5$) over the spectral scale of about 1.5 GHz (10 GHz), while the correlation between even and odd parameters is small and asymmetric with respect to $f = f'$. (d) Standard deviations $\sigma_{\vec{S}_{E},\mathrm{Po}}$ and $\sigma_{\vec{S}_{O},\mathrm{Po}}$. At the spectral limits the uncertainties approach zero. (e) and (f) Correlations of $\vec{S}_{E(O)}$ and $\vec{b}$. The correlations are oscillating and decrease generally with the frequency.



(a) (b) (c) (d) (e) (f)

FIGURE 11: (a) Individual posterior sample functions $\vec{S}_{E,\text{Po,Sa}}$ and $\vec{S}_{O,\text{Po,Sa}}$. (b) 100 posterior sample functions $\vec{S}_{E,\text{Po,Sa}}$ and $\vec{S}_{O,\text{Po,Sa}}$. The uncertainty at a given frequency is larger for the even function, because for the associated scalings of the two processes $a_{E,\text{Po}} \gg a_{O,\text{Po}}$ holds. (c) 100 spectral samples and 100 offset coefficients mapped to probed data domain. Unlike for the nonoptimised case (see Figure 5(b)), the spread is much smaller, and the quantities connect smoothly at the transition between double- and single-sided regions. (d) 100 samples $\vec{D}_{\text{Po,Sa}}$ and measured data $\vec{D}$. The samples are in agreement with the uncertainty of the measurements. (e) One sample $\vec{D}_{E,\text{Po,Sa}}$ and $\vec{D}_{O,\text{Po,Sa}}$ evaluated inside probed domain ($x < 26.36$ mm) and outside. With respect to the envelopes $\sigma_{D,E(O)}$, the contributions behave similarly in both domains. (f) 100 samples $\vec{D}_{E,\text{Po,Sa}}$ and $\vec{D}_{O,\text{Po,Sa}}$ evaluated inside the probed domain and outside. Outside the probed domain, the area between the envelops $\sigma_{D,E(O)}$ is filled by $\vec{D}_{E,\text{Po,Sa}}$ and $\vec{D}_{O,\text{Po,Sa}}$. Because the spread in the nonprobed domain is wider than the one in the probed domain, the main uncertainty in the even and odd spectra originates in nonprobed Fourier coefficients.



(a)

(b)

FIGURE 12: (a) Means $\vec{S}_{E,\mathrm{Po}}$ (blue) and $\vec{S}_{O,\mathrm{Po}}$ (red) obtained for 1000 samples of settings posterior at $\Delta_f = 1/2$ GHz. Only at the spectral limits the means deviate from the most likely mean $\vec{S}_{E,\mathrm{Po}}$ (cyan) evaluated at the maximum of the settings posterior. The uncertainty in the zero-path difference x_0 causes a considerable deviation from the most likely mean $\vec{S}_{O,\mathrm{Po}}$ (black). (b) Means $\vec{S}_{E,\mathrm{Po,Ma}}$ (cyan) and $\vec{S}_{O,\mathrm{Po,Ma}}$ (black) after performing numerical marginalisation of settings parameters at $\Delta_f = 1/2$ GHz. The marginalised means are similar to the means $\vec{S}_{E,\mathrm{Po}}$ and $\vec{S}_{O,\mathrm{Po}}$ (see (a)) at the maximum of the settings posterior except for minor differences towards the spectral boundaries. Having the marginalised covariance at hand, 100 samples $\vec{S}_{E,\mathrm{Po,Ma,Sa}}$ (blue) and $\vec{S}_{O,\mathrm{Po,Ma,Sa}}$ (red) are drawn from the marginalised amplitude posterior. Only in the centre of the spectral domain the spread of $\vec{S}_{O,\mathrm{Po,Ma,Sa}}$ increases to capture the uncertainty in x_0 as well (see (a)).

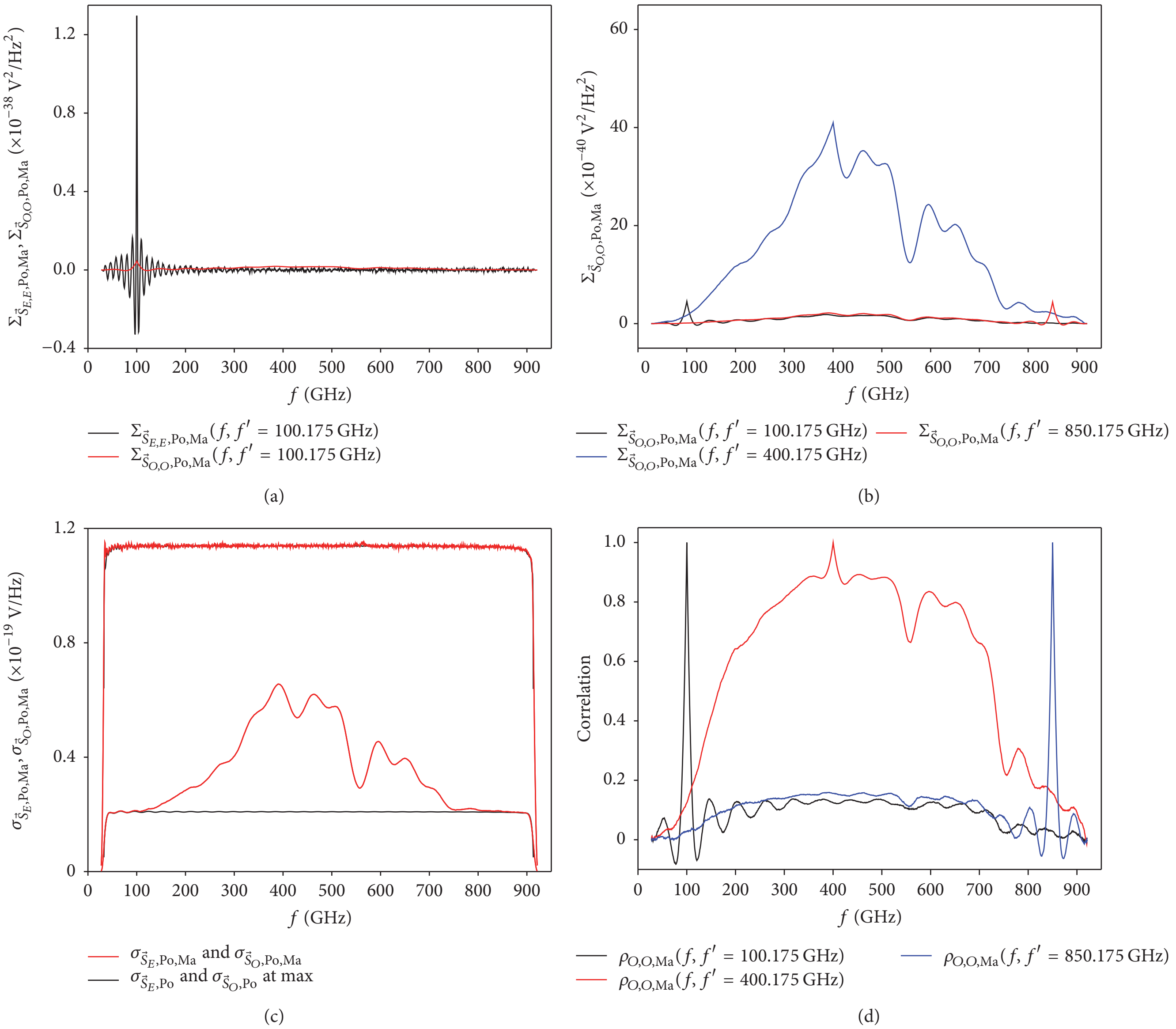
the uncertainty propagation, is analytically not possible but achievable numerically described below for $\Delta_f = 1/2$ GHz.

To reduce the numerical efforts, only the most important parameters f_L, f_U, x_0, δx, a_E, and a_O are taken into consideration for the marginalisation, and the remaining three parameters $\sigma_{b_0,\mathrm{Pr}}$, $\sigma_{b_1,\mathrm{Pr}}$, and $\sigma_{b_2,\mathrm{Pr}}$ are held at the maximum values listed in Table 1. The remaining pdf, labeled by $p'(\vec{n}, \vec{h} \mid \Delta_f, \vec{D})$ in the following, is well approximated by a factorisable multivariate normal, because the corresponding covariances are negligible (see Section 5.3.2). The posterior mean values and standard deviations are listed in Tables 1 and 2, respectively.

From the six-dimensional Gaussian posterior $p'(\vec{n}, \vec{h} \mid \Delta_f, \vec{D})$, 1000 samples are drawn. For each sample, mean and covariance for the conditional amplitude posterior are evaluated. Thereby, a difficulty arises due to the lower and upper spectral limits f_L and f_U. For given increment Δ_f, f_L and f_U determine the spectral grid. To keep this grid fixed, the spectral grid obtained for $f_{L,\mathrm{Po}} = 32.425$ GHz and $f_{U,\mathrm{Po}} = 913.425$ GHz is extended by some tens of GHz on either side. Furthermore, the posterior samples for f_L and f_U are compared with the extended but fixed spectral grid points and changed to the closest grid point. With the modified boundaries and the remaining samples unchanged, a conditional amplitude posterior follows. For example, Figure 12(a) shows the mean functions $\vec{S}_{E,\mathrm{Po}}$ and $\vec{S}_{O,\mathrm{Po}}$ for the 1000 posterior samples of the settings. While $\vec{S}_{E,\mathrm{Po}}$ changes little but most at the spectral boundaries, $\vec{S}_{O,\mathrm{Po}}$ varies considerably, mostly in the centre of the spectral domain covered. This is caused mainly by the posterior uncertainty of the zero-path difference x_0. However, the spread is symmetric around the most likely mean.

For each settings sample, 100 samples are drawn from $p(\vec{S}_E, \vec{S}_O, \vec{b} \mid \vec{n}, \vec{h}, \Delta_f, \vec{D})$. Overall, 100000 samples are available for $\vec{b}$, $\vec{S}_E$, and $\vec{S}_O$ at each spectral grid point. From this large sample set, the marginalised means $\vec{S}_{E,\mathrm{Po,Ma}}$, $\vec{S}_{O,\mathrm{Po,Ma}}$ and covariance $\mathbf{\Sigma}_{\vec{S},\mathrm{Po,Ma}}$ are evaluated which capture the first and second moments of the pdf $p(\vec{S}_E, \vec{S}_O \mid \Delta_f, \vec{D})$. In doing so, the marginalisation with respect to $\vec{b}$ is achieved implicitly.

(a) Posterior Mean. The means $\vec{S}_{E,\mathrm{Po,Ma}}$ and $\vec{S}_{O,\mathrm{Po,Ma}}$ (see Figure 12(b)) look similar to $\vec{S}_{E,\mathrm{Po}}$, $\vec{S}_{O,\mathrm{Po}}$ at the maximum of the settings posterior. However, $\vec{S}_{E,\mathrm{Po,Ma}}$ and $\vec{S}_{O,\mathrm{Po,Ma}}$ approach zero some GHz below and above $f_{L,\mathrm{Po}}$ and $f_{U,\mathrm{Po}}$, respectively.



(a) (b) (c) (d)

FIGURE 13: Characterisation of posterior covariance $\boldsymbol{\Sigma}_{\vec{S},\mathrm{Po,Ma}}$ numerically marginalised with respect to settings parameters. $\boldsymbol{\Sigma}_{\vec{S},\mathrm{Po,Ma}}$ is decomposed, so that $\boldsymbol{\Sigma}_{\vec{S}_{E(O),E(O)},\mathrm{Po,Ma}}$ captures the covariance of the even (odd) spectral quantities. (a) Cross section at frequency $f' = 100.175$ GHz. Apart from the extension in the spectral domain, $\boldsymbol{\Sigma}_{\vec{S}_{E,E},\mathrm{Po,Ma}}$ is similar to the covariance $\boldsymbol{\Sigma}_{\vec{S}_{E,E},\mathrm{Po}}$ evaluated at the maximum of the settings posterior (see Figure 10(a)). $\boldsymbol{\Sigma}_{\vec{S}_{O,O},\mathrm{Po,Ma}}$ shows a significant deviation from $\boldsymbol{\Sigma}_{\vec{S}_{O,O},\mathrm{Po}}$ (see also (b)). (b) Cross sections of $\boldsymbol{\Sigma}_{\vec{S}_{O,O},\mathrm{Po,Ma}}$ at frequencies $f' = 100.175$ GHz, 400.175 GHz, and 850.175 GHz. Towards the spectral limits $\boldsymbol{\Sigma}_{\vec{S}_{O,O},\mathrm{Po,Ma}}$ behaves like $\boldsymbol{\Sigma}_{\vec{S}_{O,O},\mathrm{Po}}$. But in the central region a broadband additive contribution is evident originating in the posterior uncertainty of the zero-path difference x_0. (c) Standard deviations evaluated from main diagonal of marginalised covariance $\boldsymbol{\Sigma}_{\vec{S},\mathrm{Po,Ma}}$ (red) and covariance $\boldsymbol{\Sigma}_{\vec{S},\mathrm{Po}}$ (black, see Figure 10(d)) evaluated at maximum of settings posterior. The odd quantity increases by a factor of three in the central region after the marginalisation. (d) Correlation $\rho_{O,O,\mathrm{Ma}}$ at three frequencies. Compared to $\rho_{O,O}$ (see Figure 10(c)), a broadband (several 100s GHz) elevation of $\rho_{O,O,\mathrm{Ma}}$ up to 0.8 follows from the uncertainty in x_0 for the central domain.

(b) Posterior Covariance. The posterior covariance after marginalisation with respect to the setting parameters is decomposed like

$$\boldsymbol{\Sigma}_{\vec{S},\mathrm{Po,Ma}} = \begin{pmatrix} \boldsymbol{\Sigma}_{\vec{S}_{E,E},\mathrm{Po,Ma}} & \boldsymbol{\Sigma}_{\vec{S}_{E,O},\mathrm{Po,Ma}} \\ \boldsymbol{\Sigma}^T_{\vec{S}_{E,O},\mathrm{Po,Ma}} & \boldsymbol{\Sigma}_{\vec{S}_{O,O},\mathrm{Po,Ma}} \end{pmatrix}. \tag{63}$$

$\boldsymbol{\Sigma}_{\vec{S}_{E,E},\mathrm{Po,Ma}}$ (see Figure 13(a) for cross section at $f' = 100.175$ GHz) and is very similar to $\boldsymbol{\Sigma}_{\vec{S}_{E,E},\mathrm{Po}}$ evaluated at the maximum of the settings posterior (see Section 5.3.3). The same holds for $\boldsymbol{\Sigma}_{\vec{S}_{E,O},\mathrm{Po,Ma}}$ and $\boldsymbol{\Sigma}_{\vec{S}_{E,O},\mathrm{Po}}$. However, a significant change is found for $\boldsymbol{\Sigma}_{\vec{S}_{O,O},\mathrm{Po,Ma}}$ (see Figures 13(a) and 13(b) for some cross sections) when compared to $\boldsymbol{\Sigma}_{\vec{S}_{O,O},\mathrm{Po}}$. Due to the posterior uncertainty in x_0, a broadband additive contribution increases $\boldsymbol{\Sigma}_{\vec{S}_{\mathrm{O,O}},\mathrm{Po,Ma}}$ up to about one order of magnitude in the centre of the spectral domain. Comparing the standard deviations $\sigma_{\vec{S}_E,\mathrm{Po,Ma}}$ with $\sigma_{\vec{S}_E,\mathrm{Po}}$, and $\sigma_{\vec{S}_\mathrm{O},\mathrm{Po,Ma}}$ with $\sigma_{\vec{S}_\mathrm{O},\mathrm{Po}}$ (see Figure 13(c)) shows that towards the spectral limits the propagation of the settings posterior uncertainties has little effect. Furthermore, the correlation $\rho_{\mathrm{O,O,Ma}}$ is high for a broad spectral range in the central region (see Figure 13(d)).

(c) Posterior Samples. From the marginalised Gaussian posterior $p(\vec{S}_E, \vec{S}_\mathrm{O} \mid \Delta_f, \vec{D})$ 100 samples ($\vec{S}_{E,\mathrm{Po,Ma,Sa}}$ and $\vec{S}_{\mathrm{O,Po,Ma,Sa}}$) are drawn and shown in Figure 12(b). As expected, the spread of $\vec{S}_{E,\mathrm{Po,Ma,Sa}}$ around the mean $\vec{S}_{E,\mathrm{Po,Ma}}$ is almost equal to the one for $\vec{S}_{E,\mathrm{Po,Sa}}$ around the mean $\vec{S}_{E,\mathrm{Po}}$ (see Figure 11(b)). The samples $\vec{S}_{\mathrm{O,Po,Ma,Sa}}$ have a wider spread in the centre of the spectral domain than $\vec{S}_{\mathrm{O,Po,Sa}}$ mainly caused by the uncertainty in the zero-path difference.

5.4. Figure of Merit for Real-World Interferometer. For the ideal Martin-Puplett interferometer, the odd spectral function, and, hence, the odd process must vanish from theoretical point of view. However, imperfections of a real-world interferometer leave the odd contribution finite in general. This imperfection is captured by the scaling a_O and in terms of signal by the envelope $\sigma_{D,\mathrm{O}} \propto a_\mathrm{O}^{1/2}$. By relating the square root of the scalings like

$$F_S = \frac{a_E^{1/2}}{a_E^{1/2} + a_\mathrm{O}^{1/2}}, \tag{64}$$

one can define the figure of merit F_S which expresses by a number the signal deviation of a real-world instrument from the ideal case. For an ideal interferometer, $F_S = 1$ holds ($a_\mathrm{O} = 0$). For the interferometer investigated here, the settings posterior (see Section 5.3.2) carries the information to state the mean as $F_S = 0.933$ with the uncertainty of about 0.01. Hence, about 7% of the signal is converted from the ideal even process to the odd process by the real-world diagnostic.

6. Discussion

6.1. Choice of Spectral Priors or Model Plausibility. The results obtained in the previous section rely on the model M_{BB} presented here with the assumption that the even and odd spectral functions can be described each by a Brownian bridge process and its associated prior covariance. An alternative model, let us say M_A, with certain assumptions on the spectral functions, leads to different prior covariances and to a different posterior. In addition, the model M_A might be more or less plausible when compared to the model M_{BB}.

In principle, the plausibility of a model relative to an alternative can be investigated within the Bayesian framework by rising the abstraction level. Starting from (36), a further factorisation needs to be carried out with respect to the used models M_{BB} or M_A. Basically, one can assign the model posterior by $p(M_{\mathrm{BB}} \mid \vec{D}) \propto K'_{\mathrm{BB}} p(M_{\mathrm{BB}})$ and $p(M_A \mid \vec{D}) \propto K'_A p(M_A)$ with the model priors $p(M_A)$ and $p(M_{\mathrm{BB}})$. The dimensionless constants K'_{BB} and K'_A may be obtained by marginalising over all linear and nonlinear parameters and hyperparameters. Then, the model plausibility is captured by the ratio $p(M_{\mathrm{BB}} \mid \vec{D})/p(M_A \mid \vec{D})$. This ratio becomes K'_{BB}/K'_A, if no model is preferred a priori for which one sets $p(M_A) = p(M_{\mathrm{BB}}) = 1/2$. Such a model plausibility study was not carried out, because it demands an investigation on its own alongside with a costly numerical treatment. However, some aspects of a plausibility study will be discussed below.

The signal envelope, corresponding to the used prior covariance for the even and odd spectral functions, is expected to be an indicator for a competitive model. This envelope should be able to resemble the global trend of the interferometric data (see Figure 3(a)). The investigated model M_{BB} seems to have these desired characteristics (see Figure 9).

For the even and odd spectral functions, an alternative prior choice could be $\Sigma_A(f, f') = g(f)\delta(f - f')$ which assigns no correlation. For example, if the function $g(f)$ is chosen to remain constant or as a triangular function, centred at f_C and approaching 0 towards the spectral limits, then the corresponding covariance $\Sigma_A(k, k')$ for the Fourier coefficients has a constant value along the main diagonal for all orders. This can be shown analytically by performing the operation stated in (43) on $\Sigma_A(f, f')$. As a consequence, no decay is imposed on the Fourier coefficients with rising order which is incompatible with square-integrability. Furthermore, the alternative signal envelop is constant in the optical path difference domain, opposing the fall-off in the interferometric data (see Figure 3(a)). Thus, the data should not be described better by either of the two suggested alternative models.

A competitive model could use a prior covariance $\Sigma_C(k, k')$ for the Fourier coefficients which has a dropping amplitude when the order of the coefficients rises. For instance, the main diagonal of $\Sigma_C(k, k')$ could be chosen like $1/(kk')^e$, and a transformation back to the spectral domain would allow further investigations of the properties of $\Sigma_C(f, f')$. For different but positive exponents e, the model plausibility could be examined.

6.2. Comparison with Standard Model. The standard analysis approach for the interferometer investigated here relies on a different model M_S which is set up hierarchically [12], and no covariances of the parameters involved are available. Basically, the quantity of interest, that is, the spectrum, is determined conditionally on the inferred voltage offset V_{Off}, the zero-path difference x_0, while setting the shift $\delta x = 0$ and the phase. Furthermore, spatial window functions are multiplied to the data to retrieve the phase and spectrum, using discrete fast Fourier transformation (DFFT) routines, for the Nyquist assumptions. A consideration of nonprobed Fourier coefficients, locating outside the experimentally accessible spatial domain, is missing completely as well as the influence of the measurement noise. Hence, a comparison of the standard model M_S with the Bayesian model M_{BB}

Table 3: Comparison of standard model M_S and Bayesian model M_{BB} for spectral increment $\Delta_f = 3.66$ GHz. M_S (second column) is well approximated by M_{BB} when the settings of the third column are chosen. However, an aliasing feature is present in the spectral domain 3000–3500 GHz (see Figure 14(a)). This feature disappears when the spatial settings are changed (fourth column) to the values at the maximum of the settings posterior of M_{BB} (fifth column). The maximum of the settings posterior reveals that the Nyquist assumptions ($f_L = 0$ GHz and $f_U = 3747.4$ GHz) are not plausible, because an overfitting of the data is made.

Model	M_S	M_{BB}	M_{BB}	M_{BB}
f_L (GHz)	0	0	0	33.614
f_U (GHz)	3747.4	3747.4	3747.4	911.912
a_E ($\times 10^{-12}$ V^2)	N/A	19.8182	19.8182	1.0304
a_O ($\times 10^{-15}$ V^2)	N/A	129.8670	129.8670	5.7641
x_0 (mm)	−5.118	−5.118	−5.115285	−5.115285
δx (μm)	0	0	−4.814	−4.814
$\sigma_{b_0,\mathrm{Pr}}$ (nV)	N/A	61.735	61.735	62.146
$\sigma_{b_1,\mathrm{Pr}}$ (μV/m)	N/A	5.822	5.822	6.007
$\sigma_{b_2,\mathrm{Pr}}$ (μV/m^2)	N/A	0.000087	0.000087	10.102
$\ln p_{\times K}$	N/A	1051396.55	1051400.41	1052037.88
Odds	N/A	$10^{-278.53}$	$10^{-276.84}$	1
Aliasing feature	Yes	Yes	No	No
Overfitting data	Yes	Yes	Yes	No

is not straightforward. In order to perform a reasonable comparison, the spectral grid with the increment $\Delta_f = 3.66$ GHz, which follows from the use of DFFT by M_S, and the limits $f_L = 0$ GHz, $f_U = 3747.4$ GHz are the same for both models. In addition, settings are determined for M_{BB} which come close to the settings for M_S (see second column of Table 3) to resemble the standard model. Given these settings, the even and odd spectra are compared. Then, the plausibility of these settings can be obtained with respect to the maximum of the settings posterior for M_{BB} by evaluating $\ln p_{\times K}$ and the corresponding odds.

The only window function applied here in the standard analysis weighs the single-sided data domain twice as large as the double-sided domain. Furthermore, the settings of the standard approach $x_0 = -5.118$ mm and vanishing δx are transferred to the model M_{BB}, and the remaining settings of M_{BB} are optimised (see third column of Table 3). The even and odd quantities $\vec{S}_{E,S}$ and $\vec{S}_{O,S}$ are inferred with M_S, and the means $\vec{S}_{E,\mathrm{Po}}$ and $\vec{S}_{O,\mathrm{Po}}$ of the conditional amplitude posterior of the model M_{BB} are evaluated. Figure 14(a) shows the results which are similar in amplitude. For both models, an aliasing feature can be found for the odd spectral quantity in the spectral domain from 3000 GHz to 3500 GHz. This feature originates in the assumed uniform optical path difference grid ($\delta x = 0$). The differences $\vec{S}_{E,S} - \vec{S}_{E,\mathrm{Po}}$ and $\vec{S}_{O,S} - \vec{S}_{O,\mathrm{Po}}$ settle mostly in the range $\pm 0.1 \times 10^{-18}$ V/Hz (see Figure 14(b)).

Keeping fixed all settings in M_{BB} but choosing x_0 and δx which are present at the maximum of the settings posterior for $\Delta_f = 3.66$ GHz (see forth column of Table 3), the aliasing feature vanishes completely, because the nonuniformity in the optical path difference grid is taken into account properly. This small change in the spatial settings would make the conditional amplitude posterior about 47 times more likely.

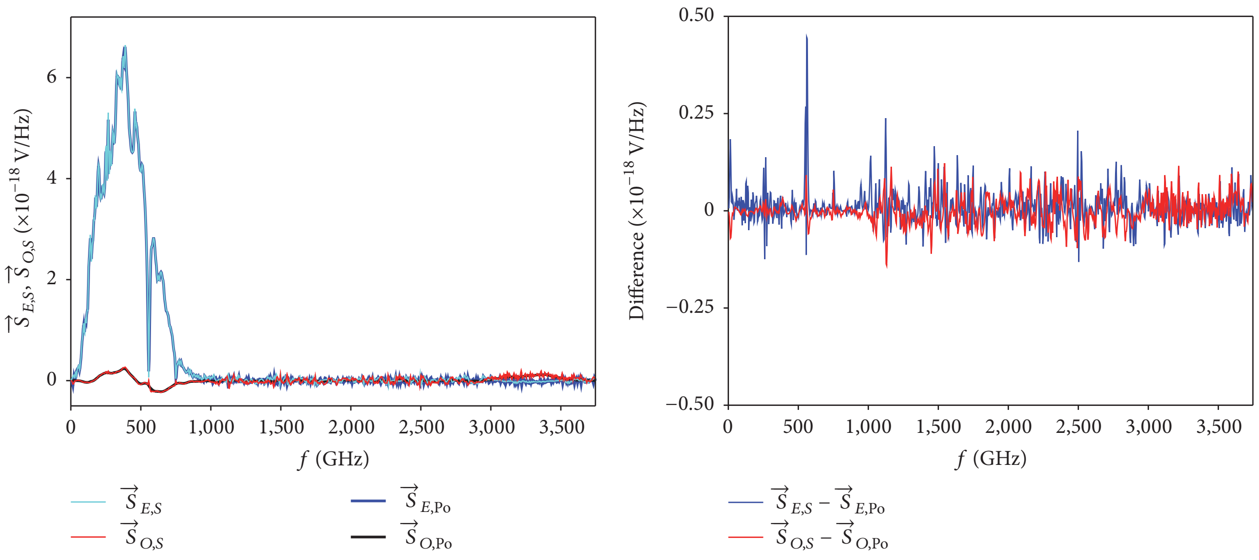
The odds $1 : 10^{-278.53}$ for the values at the maximum of the settings posterior (see fifth column of Table 3) with respect to the settings for the model M_{BB} which mimics M_S mark the Nyquist assumptions made by the standard approach very unlikely. This is explained when the corresponding residuals are investigated which show an overfitting of the data if the Nyquist assumptions are used (similar to Figure 5(c)).

6.3. Computational Time. The algorithms for the model M_{BB} and the standard model M_S are implemented in Scilab [13]. To obtain the even and odd spectra with M_S via DFFT routines, a computational time of about 4 ms is measured. This fast analysis time is exceeded by several orders of magnitude, when the problem is investigated with the Bayesian model.

For the model M_{BB}, the number $N_l = 2(f_U - f_L)/\Delta_f + 3$ of linear parameters, dependent on the spectral domain and increment, gives the dimension of the prior and posterior covariance matrices. Hence, N_l determines the computational times t_l, to evaluate the mean and covariance of the conditional amplitude posterior, and $t_{\ln p \times K}$, to investigate the settings posterior at one point in the parameter space. For the implemented algorithm and the maximising settings listed in Table 1 with decreasing Δ_f (N_l increases by about one order of magnitude), t_l and $t_{\ln p \times K}$ measured increase at least quadratically with N_l (see Table 4). This is caused mainly by the need to invert numerically prior and posterior covariance matrices.

The characterisation of the settings posterior (see Section 5.3.2) requires a duration of $10^3 t_{\ln p \times K}$ at least which adds up to hours for small Δ_f (large N_l) and to much less than one hour for $\Delta_f = 4$ GHz.

The numerical marginalisation described in Section 5.3.4 takes about half a day for $\Delta_f = 1/2$ GHz.



(a) (b)

Figure 14: Comparison of even and odd spectral quantities inferred from standard model M_S and Bayesian model M_{BB}. M_S exploits discrete fast Fourier transformation routines to evaluate the even and odd spectra $\vec{S}_{E,S}$ and $\vec{S}_{O,S}$ conditionally on the spectral limits $f_L = 0$ GHz and $f_U = 3747.4$ GHz (Nyquist assumptions), the spectral increment $\Delta_f = 3.66$ GHz, vanishing shift δx, and the zero-path difference $x_0 = -5.118$ mm. Conditional on these settings, the model M_{BB} obtains the means $\vec{S}_{E,\mathrm{Po}}$ and $\vec{S}_{O,\mathrm{Po}}$ of the conditional amplitude posterior when the remaining settings are taken at their maximum values (see third column of Table 3). (a) Even and odd spectra. Similar results follow from both models. The elevated amplitude in the odd spectra between 3000 GHz and 3500 GHz is caused by aliasing due to the assumption that the optical path difference grid is uniform. (b) Absolute differences of spectral quantities inferred by both models. Except in the vicinity of the water vapour absorption line at 557 GHz, the difference is small in absolute terms.

Table 4: Measured computational times t_l and $t_{\ln p \times K}$ of used algorithm for Bayesian model M_{BB} dependent on number N_l of linear parameters and discretisation increment Δ_f for maximising settings listed in Table 1. t_l and $t_{\ln p \times K}$ rise approximately with N_l^2. The investigation of the settings posterior demands a time of $t_{\ln p \times K}$ multiplied with a factor of at least 10^3 which becomes hours for $\Delta_f = 1/2$ GHz.

Δ_f (GHz)	≈5.68	4	3	2	1	1/2
N_l	313	443	589	883	1765	3527
t_l (s)	0.072	0.097	0.180	0.421	2.084	11.231
$t_{\ln p \times K}$ (s)	0.079	0.108	0.200	0.471	2.406	13.648

7. Conclusions

The Fourier transform is the heart of Fourier spectroscopy applications. Thereby, the interferometric data has a linear dependence on the even and odd continuous spectra to be inferred. Standard analysis techniques lack appropriate handling of fundamental aspects like noisy measurements, the influence of nonprobed spatial domains linked to Fourier coefficients above a certain order, the estimation of spectral limits, and the propagation of uncertainties of additional parameters like the zero-path difference onto the inferred spectra. For instance, the Nyquist assumption implies the fundamental misconception that the upper spectral limit of spectra to be inferred would depend on the spatial sampling. In addition, a broad spectral bandwidth would follow which increases artificially the number of Fourier coefficients necessary to describe the data. On the contrary, it can be shown analytically that a band-limitation causes spatially extended basis functions (modulated sinc functions) assigned to the Fourier coefficients in the data domain. Thus, several nearby data points are captured sufficiently by less coefficients. This example demonstrates that interferometric data contains more information than usually extracted.

As an alternative to the standard analysis techniques, a probabilistic ansatz, relying on Bayes' theorem, was proposed which is able to capture the fundamental aspects listed above. In general, Bayes' theorem relates the posterior probability density function of model parameters to the product of the likelihood and the prior probability density function for these parameters. The ansatz presented here uses multivariate normal distributions for the likelihood and the prior for parameters which map linearly to the data domain. This gives straightforwardly an analytical solution for the posterior of these linear parameters in form of a multivariate normal. Though, this amplitude posterior is conditional on the settings parameters, summarising all nonlinear model parameters and hyperparameters. After the trivial marginalisation over the linear parameters, the remaining quantity can be scanned in the settings parameters to investigate their joint posterior. This can be understood as a means of applying Ockham's razor for the linear problem. With the settings posterior at hand, the marginalisation projects the uncertainties in the settings onto the linear parameters.

The example application for the Bayesian approach infers even and odd spectra, which qualify as linear parameters, in the microwave and far-infrared spectral domain and several settings parameters, like the spectral discretisation increment, the spectral limits, the scalings of the even and odd processes, the zero-path difference, and a shift correction to the spatial sampling, given a measured interferometric data set. Each spectrum is modelled by a scaled Brownian bridge process which is able to capture a band-limitation, and the associated covariance is used in the Gaussian prior. This covariance assigns a broadband correlation, but its transform to the domain of Fourier coefficients reveals no correlation (vanishing off-diagonal elements except in connection with the zeroth-order term) between the coefficients. Furthermore, the diagonal elements drop with the square of the order of the coefficients. Hence, the prior information stated by the Brownian bridge covariance considers functions which are square-integrable and, thus, converge globally in the limit when the discretisation increment approaches zero and the order of the Fourier coefficients tends to infinity. In addition, these functions vanish smoothly at the lower and upper spectral boundaries. In the data domain, a signal envelope follows from the Brownian bridge process. This envelope decays with the optical path difference and the spectral bandwidth.

For the linear parameters like the even and odd spectra, a conditional amplitude posterior was briefly examined, relying on the Nyquist assumptions. Due to the large upper spectral limit, all noise contributions to the interferometric data are captured by the posterior mean of the linear parameters. This implies an overfitting. Because large and equal values are taken for the two Brownian bridge scalings (large signal envelops), the mapped posterior means of the spectra describe the even and odd parts of the interferometric data to equal parts in the single-sided domain, while the even part dominates in the double-sided domain. This is an indicator that the Fourier coefficients located in the single-sided domain are underestimated and overestimated for the even and odd spectra, respectively. The posterior samples for both spectra show large deviations from the means, and the even and odd contributions, obtained by mapping the samples, form much wider bands than the measurement uncertainty, especially in the single-sided domain. This indicates an unnecessary expanded solution space for the problem. Only by the posterior covariance of the linear parameters, the sum of the mapped samples complies with the data and its uncertainty band. The listed features mark a very unlikely conditional amplitude posterior which is revealed by the settings posterior.

The settings posterior for the most important settings is well approximated by the product of individual normal distributions, because no significant correlations could be found. The corresponding posterior means and standard deviations take reasonable values. These values are affected little by the discretisation increment which tends to be small, confirming the proposition that continuous spectral quantities are probed. The upper spectral limit is about a factor four smaller than the Nyquist frequency, and the lower limit is well separated from zero. This reduction of the bandwidth implies that the interferometric data can be described by a number of Fourier coefficients with associated spatial basis functions which is about one-quarter of the amount of data points. The scaling of the even process exceeds the one for the odd process by about two orders of magnitude.

For values corresponding to the maximum of the settings posterior at a small discretisation increment, the conditional amplitude posterior was investigated. By the discretisation, the number of the linear parameters exceeds the one of the Fourier coefficients, which is mandatory to describe the data points within the measurement uncertainty, by one order of magnitude. However, the Ockham's razor principle implemented by the settings posterior limits the solution space, so that the posterior means and samples for the linear spectral parameters, mapped to the data domain, have a smooth transition between the double- and single-sided regions. Due to the much larger scaling with respect to the odd process, the even process and, thus, the even spectrum describe most of the single-sided and double-sided interferogram region. While the probed interferometric data is well described within the uncertainty by the means and samples, the nonprobed data domain, corresponding to Fourier coefficients above a certain order, is filled broadly by these mapped samples. This filling decays with increasing optical path difference and is limited by the signal envelopes which follow from the estimated scalings of the Brownian bridge processes and the bandwidth. Because the spread of the mapped samples in the nonprobed domain exceeds the one in the probed region, the main uncertainties for the even and odd spectra originates in nonprobed Fourier coefficients.

The numerically costly marginalisation over some of the settings shows that the zero-path difference changes the covariance for the odd spectral parameters significantly. Basically, a broad increase of the posterior uncertainties and correlation was found.

A figure of merit was introduced which states the deviation of a real-world interferometer from an ideal diagnostic. By relating the scalings of the even and odd processes, the used interferometer is characterised as being close to the ideal case for which the odd process must vanish.

Disclosure

The views and opinions expressed herein do not necessarily reflect those of the European Commission.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

The first author acknowledges the institutions CCFE (Abingdon, United Kingdom), IPP (Greifswald and Garching, Germany), IFP (Milan, Italy), and KTH (Stockholm, Sweden) and thanks C. Marchetto, Drift, U. May, C. and J. Hastie, C. Giroud, S. Jachmich, K. Anke-Pense and E. Pense, A. Dinse, M. Domin, W. and T. Schlett, W. Mühlenbeck and J. Vieweg, D. and E. and H. Förster, S. and J. Ferdinand, and K. and U. Schmuck for the continuous support and help, especially during a period of severe sickness of his mother Sabine. This work has been carried out within the framework of the Contract for the Operation of the JET Facilities and has received funding from the European Unions Horizon 2020 research and innovation programme.

References

[1] A. E. Costley, "50 years of electron cyclotron emission research," *Fusion Science and Technology*, vol. 55, no. 1, pp. 1–15, 2009.

[2] A. Michelson and E. W. Morley, "On the relative motion of the Earth and the luminiferous ether," *American Journal of Science*, vol. 34, no. 3, pp. 333–345, 1887.

[3] D. H. Martin and E. Puplett, "Polarised interferometric spectrometry for the millimetre and submillimetre spectrum," *Infrared Physics*, vol. 10, no. 2, pp. 105–109, 1970.

[4] J. Kauppinen and J. Partanen, *Fourier Transforms in Spectroscopy*, Wiley-VCH Verlag GmbH, Berlin, Germany, 2001, Federal Republic of Germany.

[5] S. P. Davis, M. C. Abrams, and J. W. Brault, *Fourier Transform Spectrometry*, Academic Press, San Diego, Calif, USA, 2001.

[6] C. D. Porter and D. B. Tanner, "Correction of phase errors in fourier spectroscopy," *International Journal of Infrared and Millimeter Waves*, vol. 4, no. 2, pp. 273–298, 1983.

[7] J. Kauppinen and J. Partanen, *Fourier Transforms in Spectroscopy*, Wiley-VCH Verlag GmbH, Berlin, Germany, 2001, Federal Republic of Germany, p. 196 and 205.

[8] E. T. Jaynes, "Maximum entropy and bayesian spectral analysis and estimation problems," in *Fundamental Theories of Physics*, vol. 21, 1987, 1.

[9] G. L. Bretthorst, *Bayesian Spectrum Analysis And Parameter Estimation*, vol. 48 of *Lecture Notes in Statistics*, Springer-Verlag, New York, NY, USA, 1988.

[10] L. Carleson, "On convergence and growth of partial sums of Fourier series," *Acta Mathematica*, vol. 116, pp. 135–157, 1966.

[11] C. E. Rasmussen and C. K. I. Williams, *Gaussian Process for Machine Learning*, The MIT Press, Boston, Mass, USA, 2006.

[12] S. Schmuck, J. Fessey, J. E. Boom et al., "Electron cyclotron emission spectra in X- and O-mode polarisation at JET: Martin-Puplett interferometer, absolute calibration, revised uncertainties, inboard/outboard temperature profile, and wall properties," *Review of Scientific Instruments*, vol. 87, no. 9, Article ID 093506, 2016.

[13] Version 5.5.2, "Open source software," 2015, http://www.scilab.org.