Fundamentals of optical fiber sensing schemes based on coherent optical time domain reflectometry: Signal model under static fiber conditions

Leonid B. Liokumovich, Nikolai A. Ushakov, Student Member, IEEE, Oleg I. Kotov, Mikhail A. Bisyarin and Arthur H. Hartog, Senior Member

Abstract—The paper develops a statistical model for the signals received in phase-sensitive optical time domain reflectometry (OTDR) probed by highly coherent sources. The backscattering process is modelled by a set of discrete scatterers with properly chosen parameters. Explicit equations for calculating the amplitude and the phase of the backscattered signal are obtained. The developed model predicts spectral and autocorrelation characteristics of the amplitude signals that are validated by experimental results. Characteristics of the phase signals, practicable for studying the sensing applications of the OTDR system, are presented and studied as well, demonstrating good correspondence with experiment. A more detailed modelling of distributed vibration sensing systems and their response to disturbances along an optical fiber will be possible as an extension of the developed formalism.

Index Terms—Optical fiber sensors, reflectometry, time domain analysis, optical fiber devices, Rayleigh scattering.

I. INTRODUCTION

Optical time domain reflectometry (OTDR) is a widely used technique for probing the integrity of optical fiber links [1] as well as assessing the uniformity [2]–[4] of optical fibers.

The shift from multimode to single-mode fiber in the early 1980s led to the exploration of coherent detection in OTDR as a means of extending the range of the technique [5], [6]. Early experiments led to the appreciation of the phenomenon of fading [7], [8] when OTDR is carried out with coherent sources. Fading was initially regarded as a nuisance [9]–[11] because it introduces a large uncertainty in the intensity of the backscatter signal unless special measures, such as frequency diversity, are employed to mitigate its effects [12].

The use of Rayleigh backscatter for distributed sensing was proposed in the early 1990s [13], [14] and tested during the following 5 to 6 years [15]–[17], using illumination with a coherent source. The technique responds to a dynamic strain applied locally to any section of the sensing fiber. It is distributed in the sense that each and every part of the sensing fiber responds to the measurand. The vibration is converted from an external vibration to a dynamic strain of the fiber through a variety of mechanisms, depending on the siting of the fiber. In the simplest configuration [14], [17], variations of the intensity of the backscatter signal are used to provide an indication of the presence of a disturbance at any position along an optical fiber, and this approach has been used for intrusion detection [18]. Although the transfer function of the backscatter intensity as a function of fiber strain is highly nonlinear, research continues in its application [19], [20], and it has been used extensively for the detection of third-party interference in, for example, pipelines [21].

The phase of the backscatter signal, differentiated over a sufficient fiber section, provides a more linear transfer function of strain, and several configurations have been proposed and/or demonstrated [22]. These include launching pairs of pulses having a defined frequency difference [13], using an interferometric detection of the difference in the phase of backscatter returning from two distinct fiber sections [23], and a direct measurement of the phase of backscatter signal translated to RF, followed by differentiation in the digital domain [24]. The differential phase methods are particularly valuable when the application requires a linear response, such as in seismic acquisition, in which sweep-frequency sources are commonly used. In this case, the measured output is converted to an impulsive response by a correlation or a deconvolution operation that requires that the detected signal is reasonably linear.

Distributed vibration sensing (DVS), sometimes termed distributed acoustic sensing (DAS), is attracting considerable interest in oil and gas exploration and production for the acquisition of borehole seismic data and flow determination, in particular. In the case of the borehole seismic application, the DVS permits a complete seismic profile of the well to be acquired in a single shot (impulsive source) or sweep (sweep-
frequency source) and this improves the acquisition productivity considerably [25]–[27]. Nonetheless, the traditional electrical sensor arrays used for borehole seismic acquisition presently offer better signal-to-noise ratio than the distributed optical techniques. There is, therefore, a strong motivation for improving the sensitivity as well as the dynamic range and linearity of the optical methods.

To improve the performance of DVS systems, it is essential to understand the physical processes giving rise to the signals that are detected and their sensitivity to the design of the acquisition system.

Several papers have proposed backscattering models based on discrete [28], [29] and continuous [30] scattering, a number have discussed the statistical properties of the Rayleigh backscatter [31], [32]. However, most have had the purpose of understanding the fading statistics and, although one of these [33] discusses the effect of a disturbance, this is only in the context of an intensity measurement. In this paper we examine the theory and performance of distributed measurement systems based on coherent OTDR, and in particular phase-measuring coherent OTDR systems, in the case of no external perturbation. Given the random nature of the signals detected, a statistical model is essential, and this is what the present paper addresses.

II. BASIC PRINCIPLES FOR BUILDING A MODEL OF THE BACKSCATTER PROCESS

Firstly, it should be appreciated that an OTDR device really measures only the photocurrent of the detector, generated by the optical signal arriving at the detector at the sampling time. The received optical signal is defined by the electric field of the optical wave that results from the probe pulse undergoing the processes of backscattering and guidance back to the launching end.

It can be characterized by the following phasor:

$$\tilde{E}(\tau) = A(\tau)e^{i\phi(\tau)} = \tilde{E}(\tau) \exp \left[ i \arg\tilde{E}(\tau) \right],$$

(1)

where $\tau$ is the time delay between the moment when probe pulse enters the fiber line and the moment when the scattered wave with amplitude $A(\tau)$ and phase $\phi(\tau)$ arrives back at the launching end of the fiber.

The presented model is developed in a way that is most suitable for further relation of the backscattered wave parameters $A(\tau)$ and $\phi(\tau)$ (or $|\tilde{E}(\tau)|$ and $\arg[\tilde{E}(\tau)]$) to external distributed perturbations.

The main characteristics required to create the proposed model are the distributed parameters of the fiber that define its scattering properties and the properties of the propagation of the optical wave along the fiber. These characteristics are properties of the fiber itself, they are independent of light propagating along the fiber, and they form an axial profile of the fiber. In turn, the axial fiber profile may be altered by external perturbations applied to the fiber.

Thus the central consideration is the relation between the fiber axial profile and the backscattered wave $\tilde{E}(\tau)$ at the fiber launching end at some time $\tau$. A model that includes a consideration of mechanisms for changes to the axial profile by fiber perturbations will allow, in future work, the effect of such perturbations to be modeled and analyzed.

One additional consideration is the fact that modern systems generally involve analog-to-digital conversion (ADC), and it is therefore sensible to take account of the discrete character of the acquisition in the model from the outset.

We assume that the sensing fiber is probed with pulses of duration $T_p$ that are launched at pulse repetition frequency (PRF) $f_o$. The signal of the photodetector can be structured as a set of reflectograms or traces $u_i(\tau)$, where index $k$ labels each reflectogram and, with some assumptions, each $u_i(\tau)$ corresponds to the wave $E_i(\tau)$ arriving at the photodetector. Thus, $k$ corresponds to what we refer to as a “slow-time” scale $t_k = kf_o$.

In contrast, we describe as “fast-time” another discrete time scale formed by times $\tau_i$, where $i$ denotes the sample index in the trace ($\tau_i$ is the time between the center of probe pulse crossing the fiber input face and a particular part of the backscattered wave returning to the launching end of the fiber). The fast-time scale corresponding to numbers $i$ is determined by the ADC sampling frequency $f_o$ (discretization frequency), i.e., $\tau_i = if_o$. We avoid using the term “sample rate,” because for such systems, the term sample rate can apply to $f_o$ or to $f_c$ depending on which signal is under consideration.

Thus, the general discrete time scale is described by a two-dimensional scale defined as $t_a = t_k + \tau_i$, and, by fixing one of the index dependencies, we can explore either fast or slow time: $\tau_i = k = cT_p = const$ and $t_k = const$. In Fig. 1, the mechanism of transforming the initial continuous time scale into a two-dimensional time scale $t_a$ (matrix notation of signal samples $u_a$) is shown. In Fig. 1, $P(\tau)$ denotes the optical power launched into the fiber, and $\tau_i$ denotes the light travel time along the fiber and back.

![Fig. 1. Demonstration of time-scale transformations in OTDR systems.](image)

It is well known that in OTDR systems, the backscattered wave $E_i(t_a)$ that is received at any one time is the result of backscattering over some limited scattering zone with length $\Delta z = cT_p/2n_g$, where $c$ and $n_g$ are, respectively, the speed of light and the group refractive index of the fiber. The position
and length of this zone are related to \( \tau \), as considered below. However, taking into account the finite bandwidth from the photodetector, a signal \( E_{\phi} \) is determined by the average of the field \( E(\tau) \) over a time interval surrounding time \( t_d \) instead of just the instantaneous value \( E(\tau) \). This limitation does not apply where certain conditions are met and, in particular, if \( f_d > f_{\phi} \) and \( f_{\phi} \) are much greater than the upper limit of the frequency content of \( E(\tau) \). In the remainder of this article, \( u_{\phi} \) samples correspond to values of backscattered wave \( E_{\phi} = E(t_d) \) acquired at times \( t_d \).

As can be seen from the reasoning above, further steps are required to develop a complex model that allows a consistent and clear description of the forming of the signals in coherent OTDR systems intended to measure external fiber perturbations. These further steps are as follows:

1. It is necessary to determine the fiber characteristics describing the amplitude and phase profile of the scattering properties along the fiber (axial profile). If modeling is needed, rules for constructing the specific realizations of a suitable model should be formulated.

2. The relationship between observed \( \tau \)-dependencies of the received backscattered wave \( A(\tau), \phi(\tau) \), and current parameters of the axial profile of the fiber must be determined.

3. Rules for calculating the variations of the axial profile characteristics caused by perturbations of the fiber must be defined.

In the current paper, the first two points are developed, whereas the third will be considered separately.

III. INTRODUCTION TO THE FIBER AXIAL PROFILE CONCEPT

In common with other approaches, we define a distribution of discrete scatterers, and we characterize each scatterer with amplitude \( a \) and coordinate \( z \), so all scatterers are defined by set \( \{a_m, z_m\} \). It is convenient to imply a ranking \( z_m > z_{m+1} \). Distances \( z_m - z_{m+1} \) and sample separation \( \tau_i - \tau_{i+1} \) are assumed to be independent.

In our model, we assume \( z_m - z_{m+1} \gg \lambda \). We can connect this assumption to the “classical” understanding of Rayleigh scatterers by assigning each of the scatterers that we define on a “miniscale” (scatterer separation >> \( \lambda \)) to the aggregate of all the “microscale” (on a scale << \( \lambda \)) within the region separating adjacent miniscale scatterers. It should be noted that substitution of a number of microscatterers situate in the interval \( [z_m; z_{m+1}] \) by one equivalent scatterer with \( a_m \) and \( z_m \) corresponding to amplitude and phase of the backscattered wave is a formal operation that can be done to simplify the consideration. This does not affect the recorded \( \tau \)-dependencies in the case of \( z_m - z_{m+1} \ll (\tau_i - \tau_{i+1}) \lambda / \lambda \) and can be made without additional approximations, provided that the coherence length of the source is much larger than \( (z_m - z_{m+1}) \).

However, approximations will be introduced when the positions of scatterers, rather than their phase because this representation is independent of the optical frequency and it simplifies the further consideration of the variation of the phase of the scatterers when the fiber is perturbed. It is also more representative of the physics of the scattering process.

In an alternative approach, one could consider a set of random phases \( \phi_m \) instead of placing equivalent scatterers at local coordinates \( z_m \) to calculate the amplitude and phase of the full backscattered wave. However, we chose to define the positions of scatterers, rather than their phase because this representation is independent of the optical frequency and it simplifies the further consideration of the variation of the phase of the scatterers when the fiber is perturbed. It is also more representative of the physics of the scattering process.

The set of the amplitudes and coordinates of the scatterers is not sufficient for a complete definition of the fiber axial profile. To form the necessary relations between scattering and the properties of received wave \( E_{\phi} \), the longitudinal distribution of the refractive index \( n(z) \) should also be considered. Moreover, the group refractive index \( n_g(z) \) must be also used; however, further on we will use a simple relationship between \( n(z) \) and \( n_g(z) \).

As a result, the axial profile can be defined with follow two types of characteristics:

\[
\begin{align*}
\text{set} & \{ a_m; z_m \}; \\
\text{distribution} & \{ n(z) \}.
\end{align*}
\]

It is interesting that both parts of the axial profile have strong connections to the material density and its fluctuations. The first factor defining the axial profile corresponds to scattering properties of the fiber (described by the scatterers), and it is related in a physical sense to fiber density fluctuations over distances much smaller than the wavelength. So, the set of amplitudes and coordinates of these scatterers is related to the statistical distribution of these fluctuations on a microscopic scale. The second factor, namely the refractive index \( n \) and its axial dependence \( n(z) \), by definition corresponds to the mean value of density over intervals much greater than the wavelength. It therefore models the fiber on a macroscopic scale.

A. Rules for forming of \( \{a_m; z_m\} \)

It is not necessary to consider specific properties of the set \( \{a_m; z_m\} \) in addition to the conditions mentioned above to obtain a system of general expressions relating the fiber to the backscatter signal. However, to make practical calculations and to set up a model, some concrete rules for generating \( \{a_m; z_m\} \) must be used. Ideally, these rules are best defined in such a way that the general expressions are simplified and that more applicable equations for numeric calculations are obtained. We set out the following main rules.

1) The following rules will be used to form the set of coordinates of the scatterers. The spatial \( z \)-axis is divided into intervals of length \( d \), and the \( m \)-th scatterer is assumed to be positioned inside the \( m \)-th interval, which has a range: \( md < z_m < (m+1)d \). The set of initial scatterer coordinates (before any effects of external perturbations are applied) will be defined as

\[
z_m = md + \delta z_m
\]

where \( \delta z_m \) are independent random values with uniform distribution in an interval \([0; d]\).
For the proposed rules, it is clear that \( \zeta_m - \zeta_{m-1} = d \). Taking into account the previously mentioned limit for \( \zeta_m - \zeta_{m-1} \), limitations for \( d \) can be formulated as \( d < \frac{\epsilon T_P (2n_p) \Delta \zeta}{\alpha} \).

As follows from the physical meaning of the Rayleigh scattering, the scatterer size \( d \) must be much smaller than the wavelength of the irradiating light \( \lambda \). However, for practical cases (see Section 2.3 for the values of the typical OTDR systems parameters), this will result in a vast number of scatterers in the scattering zone, which will make the modeling of such systems very computationally expensive. However, when the condition \( d \ll \Delta \zeta \) is fulfilled, the statistics of the modeled backscattered wave remain nearly independent on the particular number of scatterers (\( 10^3 \) to \( 10^6 \)). Therefore, the condition \( \lambda \ll d \ll \Delta \zeta \) (where the second inequality is more crucial) will still ensure that the statistics of the backscattered wave amplitude follow Rayleigh statistics on the one hand and require a reasonable amount of required computations on the other hand.

2) The \( a_n \) are independent random quantities with equal distributions, each of them characterizes the magnitude of the backscattered signal from the \( m \)-th scatterer. A mean square value of scatterer’s amplitude could be defined directly proceeding from the relationship between power flow and amplitude of the travelling electromagnetic wave. Let \( P_f \) be the optical power of the probe pulse, then the power of radiation propagating in the opposite direction is determined based on the following considerations. The power of the Rayleigh scattering from a fiber segment of length \( d \) is \( \alpha_d P_f \), with a loss coefficient \( \alpha_d \) being on the order of 0.15 dB/km at the 1.55 \( \mu \)m wavelength [34]. Not all the scattered power propagates backwards, and calculation of the fraction \( B \) of the scattered power, captured by the optical fiber and travelling in the opposite direction is the subject of thorough investigations [35], [36]. Thus the power flow \( P_{bs} \) of the backscattered radiation produced by the Rayleigh scattering can be represented as \( P_{bs} = \alpha_d B P_f \), with the coefficient \( \alpha_d B \) being on the order of \( 10^{-7} \) m\(^{-1} \) [35].

If power flow density of the wave is given, then its electric field can be calculated using, e.g. formula (2.4-25) of [37]. Finally, the scatterer’s amplitude is obtained from the relationship

\[
\langle a^2 \rangle = \langle P \rangle \alpha_d B d \frac{4\pi}{c n_p S_{cm}},
\]

where \( S_{cm} \) is the effective cross-section of the fiber mode. This formula reflects the physical stages of the backscattered radiation formation and allows the absolute value of the recorded signal to be estimated.

Nevertheless, when modelling OTDR systems within the scope of the current paper, there is no necessity to handle absolute signal levels. For convenience of comparing various reflectograms it suffices that the \( a_n \) proportional to the distance between scatterers and inversely proportional to the probe pulse duration. Henceforth the mean square value of scatterers’ amplitudes is assumed to be

\[
\langle a^2 \rangle = \frac{2 \pi \sigma_d}{c T_P} = \frac{d}{\Delta \zeta},
\]

as it provides a stable, usable, level of calculated signal independently of the values of \( d \) and \( T_P \) used in computations. This level corresponds neither to absolute amplitude of the electric field nor optical power, but its advantage will be manifested in unification of the signal level when studying alterations of signal characteristics with respect to the pulse duration by numerical modeling.

B. Additional considerations

When considering probe pulse propagation in optical fibers, the formal description must take into account dispersion phenomena. Coherent OTDR systems imply that the laser source coherence length is much greater than \( \Delta \zeta \) (coherence time much greater than \( T_p \)), so the optical spectral width is determined by a relatively narrow pulse. However, in the applications of distributed sensing with a coherent OTDR setup, the effects of dispersion (for instance pulse broadening) for practical cases of \( T_p \) about 10 to 100 ns and fiber length less than 10 km are negligible. Therefore, to simplify the consideration, dispersion effects, either material or waveguide, are not considered in the model that is presented here.

Nonetheless, we ought to use the effective group and phase refractive indices to describe the time delay and the phase shift that are caused by light propagation in fiber. Analysis of these refractive index values presented in [38], [39] for a wide wavelength range leads to the conclusion that they are closely interrelated and can be assumed to be connected by some factor \( K_g \):

\[
n_g = K_g n
\]

where \( K_g \) is close to 1.01 and varies only slightly with wavelength within the vicinity of 1.55 \( \mu \)m ([\( K_g - 1.01 \)~10\(^{-5} \)], but for long fiber paths this difference can influence the actual relationship between time delay, spatial positions, and phase shift, for example in the selection of which scatterers are included in the electric field summation.

It is helpful to set some limits on the wide range of parameters that define the characteristics of coherent OTDR measurement systems, based on the typical applications that these systems are used for. We will assume that the duration of probe pulse \( T_p \) is in the approximate range of 10 to 100 ns, and, in turn, that the spatial resolution \( \Delta \zeta \) is in the range 1 to 10 m. We will also assume that the sensing fiber length is in the range 0.1 to 10 km, so the two-way transit time through the fiber is about 1 µs to 100 µs. The pulse repetition frequency is in the range of 1 kHz to tens of kilohertz. Finally, the sample rate of the acquisition is in the range from 100 megasamples per second (MSPS) to several gigasamples per second. So, the spatial interval corresponding to adjacent samples on the fast-time trace is less than \( \Delta \zeta \) and ranges from 1 m down to below 0.1 m.

These ranges will be used in the examples and for some approximations in the following sections. They are driven by applications in distributed vibration sensing [21] and are rather shorter than those in the context of coherent OTDR for very long distance monitoring of optical communications links [12].
IV. DESCRIPTION OF THE RECEIVED SCATTERED WAVE UNDER THE CONDITIONS OF FIXED FIBER CHARACTERISTICS

Because a nonperturbed fiber is considered in the current paper, i.e., no parameters of the fiber axial profile are changed from trace to trace, the slow-time argument \( k \) can be omitted and only the dependency on the reflectogram point (sampling time \( \tau \)) will be considered.

In this paper, we also omit from consideration the noise from the active devices, such as the laser and receivers. These points are important, but can be addressed separately. The model developed here is isotropic (the evolution of the state of polarization of the probe pulse and the backscatter are omitted). Formally, it may be incorrect in most practical cases, but the consideration of the polarization effects significantly complicates the theoretical description. Omitting the polarization effects is appropriate when any fiber part of length \( cT_p/2n \) is close to isotropic and no polarization filtering is present in the acquisition. Polarization effects merit a separate theoretical extension to this work.

A. General model of the backscattered wave definition

As was mentioned in Section II, the backscattered wave arriving at the photodetector at some time is a superposition of the light waves, scattered from some particular fiber section (a scattering zone) that corresponds to this time. Thus, in the models with discrete scatterers, the phasor \( E_i \) may be written in the form

\[
E_i = \sum_{\tau_i \in \tau_{0} \in \tau_{2}} a_m \cdot \exp \left( \frac{4\pi i}{\lambda} \int_{0}^{T_p/2} n(x) dx \right)
\]

(7)

where an assumption of rectangular probe pulse shape was used.

In most of the literature, the position and boundaries of the scattering zone are considered using simple approximations, namely the position is \( c\tau/2n \) and the length of the scattering zone is \( cT_p/2n \). However, stronger relations can be formulated and analyzed. Taking into account the light propagation over the fiber, the scattering zone borders \( z_{i1} \) and \( z_{i2} \) are related to \( \tau_i \) using the integral equation

\[
\tau_i = \frac{2}{c} \int_{0}^{z_{i1}} n_g(x) dx \pm \frac{T_p}{2}
\]

(8)

where \( n_g(x) \) is the spatial distribution of the group refractive index along the fiber line, and \( \tau_{i1} \) and \( \tau_{i2} \) are, respectively, the time delays between the front and the back of the probe pulse slopes passing the input fiber end and the sample time of the backscattered wave. However, the scattering zone center \( z_{i} \) together with the zone width \( \Delta z = z_{i2} - z_{i1} \) provides a more convenient definition of the zone. Formal relations for \( z_{i} \), \( \tau_i \) and \( \tau_{i1}, \tau_{i2} \) are

\[
\tau_i = \frac{2}{c} \int_{0}^{z_{i1}} n_g(x) dx - \frac{T_p}{2} = \frac{2}{c} \int_{z_{i2}}^{z_i} n_g(x) dx - \frac{T_p}{2} = \frac{2}{c} \int_{z_{i1}}^{z_{i2}} n_g(x) dx.
\]

(9)

Due to spatial fluctuations of \( n \), the zone width can depend on the zone position \( \Delta z(z_i) \), and also \( z_i \) is not necessarily the geometrical center of the zone, but it corresponds to its temporal center.

We note that in practice, sensing systems contain, in addition to the main sensing fiber, auxiliary fibers and other optical elements that must be taken into account. However, generally this involves only effecting an additive shift in the index \( i \). Further, we will continue to imply that the \( i \)-scale corresponds to the main sensing fiber.

These expressions for \( E_i \) and \( z_i \) (\( E_A \) and \( z_A \) when considering external perturbations and taking slow-time into account) define the backscattered wave, i.e., constitute the general background for consideration of the wave in terms of the proposed axial-profile model.

It is desirable to simplify the expressions that have been introduced to improve readability and to make clearer the physical background to the model. One way of achieving this is to rely on objective factors, such as certain relations between parameters of the model (e.g., \( T_p, f_o \), etc.) or assumptions about the properties of objective relationships (e.g., the spatial rate of \( n(z) \) deviations). On the other hand, subjective factors, such as the rules of coordinates \( \{z_m\} \) formation, etc., may be used to simplify the expressions.

An expression for a zone’s phasor (7) can be rewritten as follows:

\[
E_i = \sum_{z_{i1} < z_i < z_{i2}} a_m \cdot \exp \left( \frac{4\pi i}{\lambda} \int_{0}^{T_p/2} n(x) dx \right) \cdot \exp \left( j \omega \tau_i / K_g \right)
\]

(10)

where the following relation is taken into account:

\[
\frac{4\pi i}{\lambda} \int_{0}^{T_p/2} n(x) dx = \left( \frac{2\pi c}{\lambda} \right) \left( \frac{2}{cK_g} \right) \int_{0}^{T_p/2} n_g(x) dx = \omega \tau_i / K_g.
\]

(11)

Assuming that the wave \( E_i \) is formed by some imaginary mirror, located at the point \( z_i \), then the sum in square brackets [\( \ldots \) ] in (10) specifies the relative reflectivity of the mirror (which is the absolute value of sum, i.e., \( |E_i| \)) and an intrinsic phase shift of the mirror (argument of the sum in square brackets). The mirror's intrinsic phase is a specific characteristic feature of each zone. The phase factor \( \omega \tau_i / K_g \) has the meaning of the phase delay associated with the wave backward propagation from the mirror’s position to the beginning of the line. It is worth mentioning that the regular phase delay \( \omega \tau_i / K_g \) does not depend on the \( n(z) \) spatial fluctuations, even when considering perturbation-induced \( n \) changes. However, that \( n(z) \) dependency is still represented, since it affects the \( z_i \) coordinate, related to \( i \) by expression (9).

One important point, relevant to the practice of coherent OTDR systems and their correct modeling, is considering the shape of the probe pulse. Firstly, in real systems, which use relatively short pulses, the pulse shape differs substantially from the rectangular. Also, in modelling, the smoothly falling edges can smooth the fluctuations of amplitude and phase of \( \dot{E} \) along \( z \) that are caused by the approximate model with discrete local scatterers. Let us consider the probe pulse shape defined by the \( w_p(\tau) \) function, with support on the \([-T_p/2; T_p/2]\) interval. \( w_p \) can be also specified as a function of the spatial argument; however, owing to the deviations \( n(z) \), this relation can be quite complicated. For a zone with center \( z_i \), the
exciting weighting function caused by pulse shape \( w_p(\tau) \) can be rewritten as a function of argument \( (z - z_i) \): \( w_p(z - z_i) = w_p(\tau(z - z_i)) \), where the correspondence of \( \tau(z - z_i) \) is set by expression

\[
\tau = \frac{2}{c} \int_{\Delta z} n_g(x)dx
\]

(12)

Considering the envelope of the probe pulse, the expression (10) should be corrected and rewritten as follows:

\[
E_i = \left[ \sum_{z_i < z < z_{i+1}} a_m \cdot w_p(z_m - z_i) \cdot \exp \left( \frac{4\pi}{\lambda} \int n(x)dx \right) \right] \cdot \exp(\jmath \omega \tau_i / K_g)
\]

(13)

The phase distribution of the probe pulse can also be introduced in the model, however, in contrast to the amplitude distribution, changing the impacts of the central and the boundary scatterers on the statistics of the backscattered wave, the phase distribution will result only in effective shifts of the scattered waves, not changing their impacts on the statistics of the backscattered wave. Therefore, this feature isn’t considered in this article in order to avoid excessive unwieldiness.

B. Model in the case of integer \( d \)-sectors in zone

When the series \( \{z_m\} \) is formed as was mentioned in Section III.A, the whole fiber length is divided into small elements of length \( d \). In this case, the selection of value of \( d \) may be made according to the following condition

\[
\Delta z(z_i)/d = M + 1
\]

(14)

where \( M \) is even.

The following equation is preferred when the described fiber is long (and includes several zones) to avoid double indexing:

\[
E_i = \left[ \sum_{m=M_1}^{M+M_2/2} a_m \cdot w_p(z_m - z_i) \cdot \exp \left( \frac{4\pi}{\lambda} \int n(x)dx \right) \right] \cdot \exp(\jmath \omega \tau_i / K_g)
\]

(15)

where \( m \) is the index of the fiber segment of length \( d \), containing the scatterer and \( M_1 = \text{floor}(\Delta z/d) \) is the number of the segment, containing the zone center. The relative position of scatterers and the scattering zone is shown in Fig. 2.

So, with (15) we provide a description for a long piece of fiber (several/many zones) but avoid the use of double indexing for scatterers (although we may possibly need to use very large values of \( m \)). For simplicity of calculation and programming, it is sometimes convenient to transform (15) using internal variable \( m' \) and rewrite the expression for the backscattered wave as follows:

\[
E_i = \left[ \sum_{m'=M_1}^{M+M_2/2} a_{M_1 + m'} \cdot w_p(z_{M_1 + m'} - z_i) \cdot \exp \left( \frac{4\pi}{\lambda} \int n(x)dx \right) \right] \cdot \exp(\jmath \omega \tau_i / K_g)
\]

(16)

Index \( m' \) is the internal dummy index running from \(-M/2\) to \( M/2 \), such that \( m' = M + m \) (see Fig. 2).

If only one zone is the subject of interest, then only \( M \) scatterers can be used, and modeling is performed according to (16) with \( M=0 \).

C. The case of refractive index with slow spatial variations

It is possible to simplify the equations (13), (15) and (16) in the case of a relatively short pulse duration (zone length), when the value of \( n(z) \) can be assumed constant within the zone. In this case, an explicit expression for the scattering zone length can be considered:

\[
\Delta z(z_i) = \frac{T_p}{2} \cdot \frac{c}{n_g}, z_{d,i} = z_i \pm \Delta z(z_i)/2.
\]

(17)

In (17) and below, the refractive index values in the zone centered on \( z_i \) are written as \( n_i \) and \( n_{gi} \); the integrals in the arguments of the exponents in (15) and (16) are transformed to the form \( n_i(z_{M_1 + m'} - z_i) \); the function \( w_p(z_{M_1 + m'} - z_i) \) can be written in the form \( w_{m'} \); and the condition (14) can be rewritten as follows:

\[
d = \frac{1}{2}(M + 1) \cdot \frac{T_p}{2} \cdot \frac{c}{n_g}.
\]

(18)

Under these conditions, the expression for the received wave phasor in case of integer \( d \)-sectors in a zone can be written as

\[
E_i = \left[ \sum_{m'=M_1}^{M+M_2/2} a_{M_1 + m'} \cdot w_{m'} \cdot \exp \left( \frac{4\pi}{\lambda} n_i \left( z_{M_1 + m'} - z_i \right) \right) \right] \cdot \exp(\jmath \omega \tau_i / K_g)
\]

(19)

where

\[
w_{m'} = w_p(2\pi n_g \cdot d \cdot m'/c).
\]

(20)
The use of expressions involving summation over $M$ scatterers, which the scattering zone consists of, is associated with the selection of the $d$ value according to (18). Therefore, the value of $d$ depends on several model parameters, but also (18) includes the value of $n_i(z)$, which can vary along the fiber length. Thus, the $d$ value, chosen for some specific $z$, may turn out to be inexact when summing partial waves produced by scatterers situated in some other fiber sections. However, in practical cases, the deviation of $n_i(z)$ is expected to be less than $10^{-3}$. Considering typical values of parameters ($\Delta z \sim 1 - 10^{-3} m$ and $M \sim 1000$), resulting in $d > 1$ mm, the $n_i$ deviation of the order of $10^{-3}$ causes a change of the value of $M$ of approximately 1. Therefore, the error due to not accounting strictly for scatterers while summing over a fixed number of scatterers $M$ will be negligibly small.

Then, substituting (18) into (5) and noting that $\Delta z = 2\pi c T_p$, one can obtain the following simple normalization rule:

$$\langle a^2 \rangle = d/\Delta z = M^{-1}. \tag{21}$$

When considering signals in practical devices, the following details listed below must be taken into account. 

1. When specific optical arrangements and acquisition types are modeled, such as a direct-detection receiver or phase detection (e.g., using heterodyne or homodyne methods [22]), these expressions will provide the basis for more particular expressions for the recorded signal at the photodetector output $a_{th}$ or the signal produced by the further processing.

2. Expression (7) and the consequent ones define the backscattered wave, neglecting different noise sources in the active elements. However, their independence of the noise in the photodetector or elsewhere in the electronics should be mentioned as well as the existence of noise caused by conversion of the laser irradiation amplitude and phase deviations in heterodyne and some other optical arrangements. These types of noise could be correctly taken into account by including additional components in expressions for the recorded optical signal.

The laser frequency is included in the expressions above because the optical wavelength and amplitude values are reflected in the expressions via a common normalization factor. Therefore, the effect of variations of laser frequency and amplitude from pulse to pulse can be studied using this model.

3. Intrinsic fluctuations of phase and amplitude within the probe pulse are not involved in this model. The expressions would need to be significantly modified to take into account random changes in the amplitude or the phase of the light in the probe pulse. However, since in the systems that are considered here the coherence time of the source is much greater than the pulse duration, the intrinsic amplitude and phase deviations of the light within the pulse are negligible compared to other fluctuation sources.

V. STUDYING PROPERTIES OF Z-DEPENDENCIES OF AMPLITUDE AND PHASE OF RECEIVED BACKSCATTERED WAVE

A. Experimental setup and mathematical modeling

Even without considering changes of axial profile induced by external perturbations, the proposed model enables one to perform numerical modeling of dependencies $\tilde{E}(z)$ and investigate those of their local properties that are needed for the analysis of the performance of phase-measuring coherent OTDR systems. In the current section, we use the $z$ scale, normalized by $\Delta z$ owing to its instructiveness, despite the fact that in practical systems dependencies $\tilde{E}(z)$ are measured in the time domain. For convenience of the analysis, the temporal axis can be normalized by the probe pulse duration $T_p$.

The system of equations used to describe $\tilde{E}(z)$ allows some common behavior of backscattered waves to be studied by analytical means. However, there are wider possibilities for studying the properties of OTDR signals using adequate numerical modelling. So, in this section we will consider the results of a numerical modelling of the backscattered wave according to the mathematical apparatus developed in the previous paragraphs. Here the goal was to demonstrate the application of the modeling rules that were developed and to prove the adequacy of the calculation results by comparing them with experimental data.

As was already mentioned, in a number of papers, such key characteristics of backscattered wave as amplitude spectrum $S_E$ and autocorrelation function $C_E$ have been provided [40].

Particularly, for monochromatic light and rectangular pulses, it is shown from the analysis of $E_z(z)$ expressions that $S_E$ has the form of $S_E(f) = \text{sinc}(f T_p)$ and $C_E$ has a triangular shape with full width $2T_p$ at half-height width. However, in practical situations, the measurement system does not deal with such a broad statistical ensemble of trace realizations. In most cases, OTDR signals acquired during several seconds remain correlated and, hence, in practice, the properties of $E(z)$ fluctuations over limited spatial intervals are of a great practical interest. Thus, most of the assumptions made during earlier analytical derivations are not applicable. Moreover, the probe pulse shape can significantly differ from rectangular, changing spatial characteristics of OTDR signals compared to the rectangular one. Hence, for a more detailed analysis of backscattered wave properties, modeling these signals is essential.

In our calculations, according to Section 2, the fiber profile was formed by two arrays, describing amplitudes and spatial positions of the scatterers: $\{a_{th}\}$, $\{z_{th}\}$ and the refractive index spatial distribution, considered uniform. Refractive index fluctuations are not considered in the present paper because of their small effect on the statistical properties of the spatial oscillations of the backscattered wave. Scatterer coordinates $z_{th}$ were formed according to (2) and scatterer amplitudes $a_{th}$ as Gaussian random values according to (21).

Each reflectogram point $\tilde{E}_p$ was calculated according to (19) with the probe pulse envelope defined as function $w_{m^*}(p)$ (p the shape parameter) with edges smoothed by quadrants of the sine function period given by

$$w_{m^*}(p) = \begin{cases} \sin(\pi/2 \cdot m^*/pM), & \text{if } m^* < pM \\ 1, & \text{if } pM \leq m^* < (1-p)M \\ \sin(\pi/2 \cdot (M - m^*)/pM), & \text{if } m^* \geq (1-p)M \end{cases} \tag{22}$$

Typical simulation parameters were the following: scattering zone length $\Delta z$=10 m, spatial zone step 0.1 m, and $\Delta z$=20 mm (number of scatterers in the scattering zone in most
calculations was \( M = \Delta z/d = 500 \). Calculations for different values of \( \Delta z, d, \) and \( M \) were also performed. However, it was found that increasing of the number of scatterers \( M \) beyond several hundred did not affect the statistical (autocorrelation) properties of the modeled backscattered wave. Moreover, the spatial scale of the \( E(z) \) changes remained the same and were strictly related to the effective length of the scattering zone taking into account the probe pulse apodization function \( w_{m^*}(t) \) (22).

To prove the adequacy of the developed model, we analyzed real-world experimental data produced by a heterodyne OTDR setup [22] (Fig. 3). The main characteristics of the experimental data were the following: probe pulse width 50 and 90 ns (corresponding zone lengths \( \Delta z \sim 5 \) and 9 m), shifted by 110 MHz relative to the laser output frequency, discretization frequency 300 MSPS (with corresponding spatial step \( \sim 0.34 \) m), and pulse repetition frequency 5 kHz.

Since the edges of the probe pulse were nearly invariant with the pulse length, we concluded that the pulse shape for \( T_p=90 \) ns and \( T_p=50 \) ns is adequately modeled by the function \( w_{m^*}(p) \) with parameter values \( p=0.15 \) and \( p=0.5 \), respectively.

It is worth mentioning that the experimental data were acquired during a finite time interval (a few seconds), and the external fiber conditions were changing slightly. Nonetheless, even traces acquired at significantly different times within this interval demonstrated good correlation of the amplitude traces; this correlation limits the possibilities of ensemble averaging over this data set.

For simulated signals, amplitude and phase of the backscattered wave were found directly as absolute value and argument of the calculated phasor \( E(z) \). For the experiment results, it is necessary to extract the amplitude and phase of the backscattered wave; in the heterodyne setup, this means detecting the amplitude and phase of the intermediate frequency. In fact, a coherent detection arrangement, in which the backscatter signal is mixed with a local oscillator taken from the source laser, is believed to be the only way of acquiring complex backscatter wave prior to any further processing; other techniques, such as the dual-pulse method [13] or the interferometric phase-comparison method [23] provide a measurement that is the result of mixing the backscatter from two separate zones, so the detail of the original backscatter signal is lost. To minimize the influence of amplitude and phase detection operations on the signal statistics, we have calculated the envelope and phase of the analytical signal, since the influence of this method on signal statistics is minimal. For a more detailed study of particular OTDR systems, heterodyning and further signal processing operations can be included in the modeling; however, the main aim of the current paper is the study of backscattered wave itself.

B. Comparison of experimental and modeled results

Fig. 4 (upper plots) demonstrates spatial dependencies of the backscattered light amplitude for simulated and experimental data (presented example is for \( T_p=90 \) ns, dependencies for 50 ns pulse show the same behavior in normalized spatial scale \( z/\Delta z \).

Fig. 4. Comparison of spatial dependencies of amplitude and phase of two realizations of experimental (left plots) and simulated (right plots) backscattered waves (r.u., relative units). Solid line: first trace in data set; dotted line: trace acquired 10 s later. Some of the amplitude fading points and corresponding phase jumps are highlighted by vertical lines.
Note that the plots are proportional to the electric field, because the varying component of the output intensity oscillation in the considered systems is proportional to the electric field of the backscattered wave, which is what is analyzed and modeled in this paper.

The experimental traces that are illustrated were acquired with a time interval about 10 s, resulting in small variations of the external conditions of the fiber (or of the laser frequency) and hence, there is a small difference in the appearance of amplitude (and phase) signals; in contrast, the simulated traces were statistically independent. The experimental plots demonstrate somewhat less sharp oscillations owing to additional smoothing during the acquisition of the optical signal.

To verify the basic properties of the developed model, the probability distribution functions (PDF) of experimental and simulated backscattered wave amplitudes were compared (see Fig. 5). For simulated wave an excellent correspondence with Rayleigh fit can be observed, whereas for experimental data, the distribution differs slightly from the Rayleigh. The deviation of the experimental PDF from the simulated one could be caused by polarization effects in optical fibers (which would increase the number of degrees of freedom) and additional elements in optical setup.

Fig. 6 illustrates spectra of the backscattered wave amplitude. Since all spatial characteristics are related to \( \Delta z \), the abscissa in Figs. 6 and 7 are normalized by \( \Delta z \). It should be noted that experimental data were obtained with a heterodyne setup; hence, zero spatial frequency in experimental spectra corresponds to the intermediate frequency (110 MHz) in the initially recorded signal. As we have mentioned, it can be shown analytically that the spectrum of the backscattered wave amplitude is described by a sinc function (for rectangular probe pulse). The thick dotted curve in Fig. 6a is a statistical average over an ensemble of independent simulated amplitude spectra calculated with parameter \( \rho \) in (22) equal to 0. On the other hand, taking into account apodization and setting \( \rho = 0.15 \) changes the averaged amplitude spectrum to the one illustrated by the thick solid curve in Fig. 6a. Without statistical averaging, the spectrum for a single trace demonstrated chaotic (although decreasing) behavior (see thin solid curve in Fig. 6a). For comparison, experimental results for \( T_p = 90 \) ns are also shown. The thin solid curve in Fig. 6b demonstrates spectrum for single trace, while the thick dotted one is an example of averaging over ensemble of experimental traces. As we have mentioned, the data set does not contain enough statistically independent traces to obtain a properly averaged spectrum, as would be determined by the pulse shape; however, the coincidence of modeled and experimental results can be observed.

The correlation function is strongly related to the spectrum and demonstrates nearly the same properties of the backscattered wave: strong relation of slightly offset amplitude reflectograms and loss of correlation for offsets greater than the zone length. Sidelobes of the autocorrelation function calculated for a single trace (without any averaging over the ensemble) have the same character for experimental and simulated signals (Fig. 7, thin solid and dotted curves, respectively). Since all simulated traces were independent, we were able to obtain the average over an ensemble of traces resulting in a nearly triangular correlation function with well-suppressed sidelobes; in contrast, the experimental traces were correlated and did not form a sufficiently representative ensemble for proper averaging.

For both the amplitude spectrum and the spatial autocorrelation function, the width of the main lobe is related to the zone length and probe pulse apodization. Processing of experimental data with \( T_p = 50 \) ns shows a similar appearance with a slight spectrum widening (and narrowing of \( C_2 \)) caused by the smaller effective width of the probe pulse.

In a wide variety of practical situations, the characteristics considered above are not the only objects of interest. For instance, in some cases, we need to understand how the amplitude (or phase) of the backscattered wave differs as a result of small spatial offsets \( \delta z \) of the scattering zone along the fiber. On the one hand, these dependencies also

Fig. 5. Comparison of probability density functions of backscattered wave amplitude. (a) Experimental. (b) Simulated. Solid line: Rayleigh distribution fitted to the histogram in each case.

Fig. 6. Comparison of amplitude reflectogram spectra. (a) Simulated data. (b) Experimental data.

Fig. 7. Autocorrelation functions of amplitude traces: single realizations (experimental, thin solid line; simulated, thin dotted line; simulated signal averaged over 2000 realizations, thick dashed line), \( T_p = 90 \) ns.

For both the amplitude spectrum and the spatial autocorrelation function, the width of the main lobe is related to the zone length and probe pulse apodization. Processing of experimental data with \( T_p = 50 \) ns shows a similar appearance with a slight spectrum widening (and narrowing of \( C_2 \)) caused by the smaller effective width of the probe pulse.

In a wide variety of practical situations, the characteristics considered above are not the only objects of interest. For instance, in some cases, we need to understand how the amplitude (or phase) of the backscattered wave differs as a result of small spatial offsets \( \delta z \) of the scattering zone along the fiber. On the one hand, these dependencies also
characterize spatial properties of the backscattered wave, and, on the other hand, they are particularly of interest when the coherent OTDR is used as a sensor [22] or for analysis in the influence of the different noise sources.

These increments, by their nature, randomly change in different zones, but for a statistical evaluation the rms values of the amplitude increments \( |\bar{E}_z| - |E_z| \), normalized to its mean \( |\bar{E}_z| \), \( \sigma_a \), and the phase \( \arg(\bar{E}_z) \), \( \sigma_p \) can be used. As for the spatial dependencies above, spatial offset was normalized to the length of the scattering zone \( \Delta z \). Every considered value of zone offset increments for all possible combinations of zone pairs were calculated, and then the rms value of these increments was calculated for each trace. Values obtained for different traces were averaged, resulting in the quantities of interest \( \sigma_a \) and \( \sigma_p \). Dependencies of \( \sigma_a \) and \( \sigma_p \) on \( \Delta z/\Delta z \) are difficult to address analytically, especially for an arbitrary apodization; however, the model developed here enables one to calculate these increments and study their properties.

Simulations were carried out for several types of the probe pulse apodizations: for rectangular pulse (\( p = 0 \)), for slightly smoothed pulse (\( p = 0.05 \)), and for two cases when a correspondence with experimental pulse shape of different durations was obtained: \( p = 0.15 \) and a limiting case \( p = 0.5 \). Fig. 8 shows dependencies of \( \sigma_a \) (Fig. 8a) and \( \sigma_p \) (Fig. 8b) on the normalized shift \( \delta z/\Delta z \) for simulated and experimental traces.

The point step in the experimental data is 0.34 m, which corresponds to a step of a few hundreds of the normalized shift interval and is limited by the sample rate. In simulated data, the minimal resolution is related to length \( \delta z \), which can be reduced arbitrarily. We have used \( M \approx 1000 \) which corresponds to resolution \( -0.001 \) of the normalized shift interval (a few millimeters in absolute spatial scale). However, when considering the coherent OTDR as a sensor, analysis of smaller zone shifts (of the order of the wavelength) is usually necessary. For that purpose, the resulting spatial dependencies \( \sigma_a \) and \( \sigma_p \) can be approximated by power fits

\[
\sigma_a = a_A \left( \frac{\delta z}{\Delta z} \right)^{k_A}, \quad \sigma_p = a_p \left( \frac{\delta z}{\Delta z} \right)^{k_p}
\]

with parameters \( k_A, k_p, a_A, \) and \( a_p \) related to the apodization parameter \( p \).

One can see that for a rectangular pulse, these dependences are the square root functions, and this fact is expected from an analysis of the number of common and different summands in equations (10) for two zones having a gap between their centers. However, even a small apodization changes the behavior of dependence, and for maximal apodization (\( p = 0.5 \)), the \( \sigma_a \) and \( \sigma_p \) are linear functions of \( \delta z/\Delta z \). Empirically, we find that the fit parameters for phase, \( k_p \) and amplitude, \( k_A \), have the following dependencies on the pulse shape parameter \( p \):

\[
k_A = 0.5 + 0.6 p^{0.4}, \quad k_p = 0.5 + 0.43 p^{0.4}.
\]

The performed investigation shows that in further works, the following rough approximations can be used: both dependencies \( \sigma_a(\delta z) \) and \( \sigma_p(\delta z) \) are square root for a rectangular pulse, and linear for a cosine-like pulse shape (\( p = 0.5 \) in (22)).

VI. CONCLUSION

The results shown here illustrate how the model that we have presented allows signals of coherent OTDRs to be simulated and their properties to be analyzed. The results are in extremely good agreement with experimentally-obtained signals.

The model sets the ground rules for simulating the phase detection and extraction steps that are critical for modeling further stages in distributed vibration sensing. It can be extended to examining the fading, transfer function, and noise properties of differentiated phase measurements as are commonly used in distributed sensors and to understanding how perturbations of the fiber are transferred to the signals acquired by phase-measuring coherent OTDRs. Finally, the model can be used for understanding how noise processes in the system, e.g., laser phase noise, transfer onto the measured phase signals and degrade them.

REFERENCES


