Analysis and Reduction of Large Errors in Rayleigh-Based Distributed Sensor

Li Zhang, Luis Duarte Costa, Zhisheng Yang, *Member, OSA*, Marcelo A. Soto, *Member, OSA*, Miguel Gonzalez-Herráez, *Senior Member, OSA*, and Luc Thévenaz, *Fellow, IEEE, Fellow, OSA*

***Abstract*—Commonly, the frequency shift of back-reflection spectra is the key parameter to measure quantitatively local temperature or strain changes in frequency-scanned Rayleigh-based distributed fiber sensors. Cross correlation is the most common method to estimate the frequency shift; however, large errors may take place, particularly when the frequency shift introduced by the temperature or strain change applied to the fiber is beyond the spectral width of the main correlation peak. This fact substantially limits the reliability of the system, and therefore requires careful analysis and possible solutions. In this paper, an analytical model is proposed to thoroughly describe the probability of large errors. This model shows that the cross correlation intrinsically and inevitably leads to large errors when the sampled signal distribution is finite, even under perfect signal-to-noise ratio. As an alternative solution to overcome such a problem, least mean squares is employed to estimate the frequency shift. In addition to reducing the probability of large errors, the proposed method only requires to measure a narrow spectrum, significantly reducing the measurement time compared to state-of-the-art implementations. Both the model and the solution are experimentally verified using a frequency-scanned phase-sensitive optical time-domain reflectometry system, achieving a spatial resolution of 5 cm, with a sensing range of 860 m and an acquisition time below 15 s, over a measurable temperature range of more than 100 K with a repeatability of 20 mK, corresponding to a temperature dynamic range of 5000 resolved points.**

***Index Terms*—Fiber optics, fiber optics sensors, Rayleigh scattering.**

I. Introduction

DISTRIBUTED optical fiber sensors (DOFS) are capable of spatially resolving the distributed profile of a physical variable, such as temperature or strain, over a long sensing range using only a single interrogating unit and a single optical fiber. These systems provide a simple and cost-effective alternative to the use of complex arrays of several discrete point sensors. As such, DOFS have been extensively studied and widely used in the past decades in domains such as structural health monitoring, non-destructive evaluation of infrastructures and industrial monitoring, among many others [1]–[6].

The first investigated distributed optical fiber measurement system was optical time-domain reflectometry (OTDR) based on Rayleigh scattering [7]. Owing to the relatively high scattering coefficient of Rayleigh compared to Raman and Brillouin scatterings in optical fibers and its simple configuration, OTDR has been commercialized and widely used for fiber link characterization. In 1993, a system similar to OTDR, namely phase-sensitive OTDR (φ-OTDR) [8], was demonstrated to be able to detect intrusions near the fiber. Compared to normal OTDR, which uses a low coherence optical source, φ-OTDR is based on fully coherent light signals. The simplest implementation of φ-OTDR [8] measures the intensity resulting from the interference of light backscattered at different scattering points along the fiber. The resulting time-domain trace typically shows a jagged shape as a consequence of the random distribution of scattering centers along the length, but remains stationary and reproducible if no stimulus is applied to the fiber. Localized temperature or strain changes in the fiber cause an effective change of the optical path differences between the different scattering centers within the pulse length, thus leading to a modification of the recorded trace. Several recent works reported in the literature have focused on developing φ-OTDR since it shows an extremely high sensitivity compared to other distributed sensing approaches [9]–[12]. In particular, a growing number of investigations have focused on recovering the phase of the Rayleigh back-scattered light to obtain quantitative measurements, relying on different principles such as interferometric or coherent detection, the use of chirped pulses, etc. [13]–[19]. All these schemes allow fast phase recovery; hence, dynamic measurements can be achieved, at the expense of either a complex phase demodulation in the receiver [13]–[17] or the requirement of high bandwidth detection [18], [19].

Alternatively, for static measurements, phase changes of the coherent Rayleigh scattering can be retrieved by tuning the optical frequency of the interrogating pulse used in φ-OTDR and quantifying the frequency shift (FS) of the position-resolved spectra between different measurements [20]. Based on a layout

Manuscript received January 15, 2019; revised April 30, 2019 and May 15, 2019; accepted May 15, 2019. Date of publication May 20, 2019; date of current version September 18, 2019. The work was performed in the framework of ITN-FINESSE, funded by European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement no. 722509, as well as Swiss Commission for Technology and Innovation (18337.2 PFNM-NM). *(Corresponding author: Li Zhang.)*

L. Zhang, Z. Yang, and L. Thévenaz are with the Institute of Electrical Engineering, Swiss Federal Institute of Technology of Lausanne, Lausanne CH-1015, Switzerland (e-mail: li.zhang@epfl.ch; zhisheng.yang@epfl.ch; luc.thevenaz@epfl.ch).

L. D. Costa and M. Gonzalez-Herráez are with the Departamento de Electronica, Universidad de Alcalá, Madrid 28805, Spain (e-mail: luis.duarte@uah.es; miguel.gonzalezh@uah.es).

M. A. Soto is with the Institute of Electrical Engineering, Swiss Federal Institute of Technology of Lausanne, Lausanne CH-1015, Switzerland, and also with the Department of Electronic Engineering, Universidad Técnica Federico Santa María, Valparaíso 2390123, Chile (e-mail: marcelo.sotoh@usm.cl).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JLT.2019.2917746

similar to conventional φ-OTDR but with correlation between two spectra with orthogonal polarization states, a technique for distributed birefringence measurements has been proposed and demonstrated [21]. Combining such a scheme and conventional φ-OTDR, the discrimination of temperature and strain has also been recently demonstrated [22].

On the other hand, the spectra of the Rayleigh backscattered light can also be interrogated by optical frequency-domain reflectometry (OFDR) [23]. Compared to the OTDR approach, higher spatial resolution can be typically achieved with OFDR; however, it requires a coherent detection scheme, limiting the sensing range to the coherent length of the interrogating light.

In both frequency-tuned φ-OTDR and OFDR, cross-correlation (CC) of back-reflected spectra is generally used to estimate the FS. However, as pointed out and explained in this study, outlier peaks or multi-peaks may appear in the correlated spectrum, leading to large errors on the frequency shift estimation. This reduces the reliability of the system and limits the measurable temperature/strain range. While this is a problem experienced by most systems employing CC-based FS estimation, to the best of our knowledge, neither study on the origin nor impact of this issue have been reported for Rayleigh-based distributed sensors. Alternative frequency shift estimation schemes have been proposed recently to avoid the large errors [24]–[27], however, no throughout explanation and analysis have been addressed regarding this issue.

This paper reports on a thorough study of the origin and impact of large errors in the frequency shift estimation when employing the cross-correlation approach. A mathematical model as a function of the pulse width and frequency scanning range of the interrogating light is proposed to analyze the probability of large errors (PLE) in frequency–scanned φ-OTDR, which shows a good agreement with the experimental situation. It is demonstrated that the PLE is purely due to the stochastic nature of the Rayleigh scattering coherent response and is inherent to the cross-correlation method even in absence of noise. To overcome this problem, an alternative solution based on similarity using least mean squares (LMS) is proposed. Benefiting from the significantly looser condition for the appearance of large errors, the proposed approach enables the use of a narrower measurement spectrum while preserving a large measurable FS range, remarkably reducing the measurement time. A frequency-tuned φ-OTDR with 5 cm spatial resolution is experimentally demonstrated along an 860 m-long fiber, totally immune to large errors. A temperature repeatability of 0.026 K is demonstrated over a total temperature measuring range of 100 K.

II. Principle of CC Based φ-OTDR

The working principle of frequency-scanned φ-OTDR is shown in Fig. 1. In this technique, highly-coherent optical pulses are launched into an optical fiber. As the pulse propagates, light is backscattered by the presence of small refractive index changes along the fiber (Rayleigh scattering centers), caused by thermodynamic fluctuations frozen after solidification during the fiber drawing process. This generates a backscattered φ-OTDR signal, whose optical intensity is measured as a function of time,

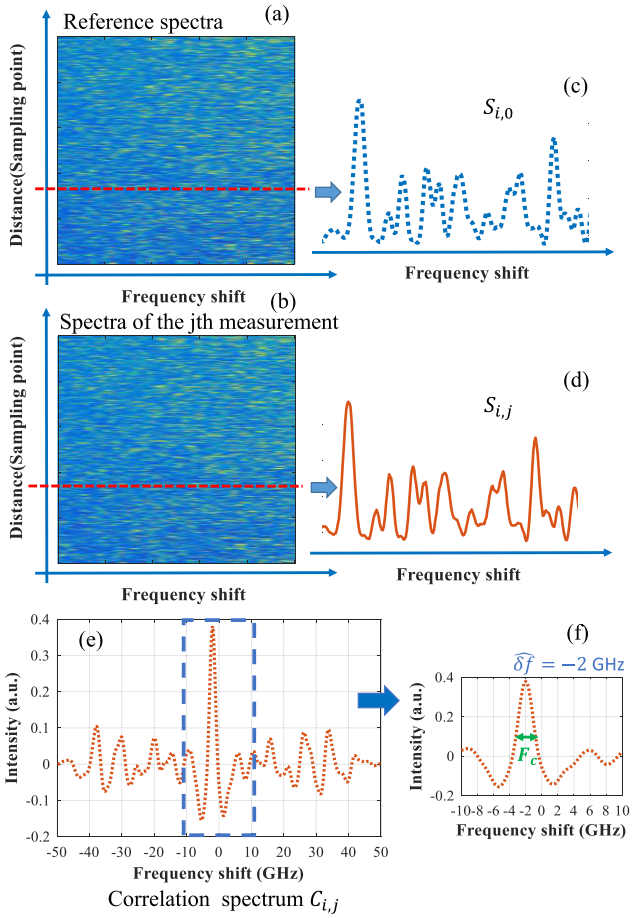

Fig. 1. Principle of cross-correlation based φ-OTDR: (a) the reference spectra and (b) the j^{th} measurement spectra of the Rayleigh scattering along the fiber; (c) the spectrum at the i^{th} position point in the reference measurement; (d) the spectrum at the i^{th} position point of the j^{th} measurement; (e) the local correlation spectrum at the i^{th} position point; (f) close-up view of the correlation peak.

and subsequently converted to position along the fiber. At a certain location, the optical intensity pattern results from the interferences between light scattered by distinct centers and is therefore dependent on the optical path differences (OPD) between the multiple scattering centers seen by the pulse at that position. Any temperature/strain change applied on such section will lead to the change of both the refractive index and section length, resulting in an OPD change that deforms the back scattering intensity pattern. Knowing that such OPD change can be perfectly compensated by shifting the light frequency, the original pattern over such fiber section can be perfectly recovered in this way [20].

The conventional measurement technique in frequency-scanned φ-OTDR requires acquiring Rayleigh intensity traces as a function of the input laser frequency within a given frequency range. The refractive index distribution is fixed along the fiber once for ever after solidification, so that at a given position the intensity observed as a function of frequency will not change and can be regarded as a reference spectrum before the fiber is subjected to any strain or temperature change. In this case, a spectral shift of this local intensity versus frequency distribution is caused by the induced refractive index variation.

Cross correlating the new measured local spectral response with the reference spectrum will result in a correlation peak placed at a frequency shift proportional to the local temperature and strain variations. The proportionality is a direct function of the thermo-optic and elasto-optic responses of silica and is strictly equal to the coefficients scaling the spectral response of fiber Bragg gratings.

Let assume that the spectral response measured at the ith sampling point along the position axis of the fiber is $S_{i,0}$ for the reference, and $S_{i,j}$ for a subsequent measurement at the same sampling point, as shown in Fig. 1(c) and (d). The procedure to obtain the CC spectrum is then:

$$C_{i,j}(\delta f) = \frac{1}{F_t}\int_0^{F_t} S_{i,j}(f)S_{i,0}(f-\delta f)df, \quad -F_t \le \delta f \le F_t \tag{1}$$

where F_t is the total frequency scanning range over the spectrum. The estimated FS $\widehat{\delta f}$ is located at

$$\widehat{\delta f_{i,j}} = \underset{-F_t \le \delta f \le F_t}{\max \arg}\ C_{i,j}(\delta f) \tag{2}$$

representing the spectral location of the highest peak of the correlation spectrum (Fig. 1(e) and (f)). In presence of noise, the accuracy of the method is directly related to the width of the correlation peak (F_c), which is in turn function of the shape and width of the interrogating pulse. For example, for a transform-limited square pulse with a width τ, Fc should be $1/\tau$ [28]. Incidentally, it means that a sharper spatial resolution (smaller τ) decreases proportionally the accuracy on the measurement. The width F_c of the correlation peak will be used later in the definition of the large errors.

III. Model to Analyze the Probability of Large Errors

The estimation error can be reasonably defined as the difference between the estimated FS $\widehat{\delta f}$ and the true FS δf_{true}:

$$\varepsilon_{i,j} = \left|\widehat{\delta f_{i,j}} - \delta f_{i,j\,true}\right| \tag{3}$$

When the error is larger than the half width at half-maximum of the correlation peak, i.e., $\varepsilon_{i,j} \ge \frac{1}{2}F_c$, it is here defined as *large error*. For example, considering that the temperature and strain sensitivities are ~1.5 GHz/°C and ~0.15 GHz/$\mu\varepsilon$ respectively for a standard single mode fiber, the use of a 1 ns-long rectangular pulse, i.e., $F_c = 1/\tau = 1$ GHz, will lead to large temperature or strain errors that will be over ~0.33 °C and ~3.3 $\mu\varepsilon$, respectively.

Obviously, poor values of signal-to-noise ratio (*SNR*) or insufficient sampling rate may lead to errors of this type. However, in this work we show that for CC, these large errors occur even for perfect *SNR* and sufficient oversampling of the frequency spectrum, due to pure statistical reasons. We will restrict our study to the error originated from the shift estimation method itself (cross-correlation), while the impact of other parameters (such as *SNR* and sampling rate) are out of the scope. Hence, all over our theoretical derivation, we will always assume that the *SNR* is sufficiently high, and we will ensure this high *SNR* in the experiments by performing a large number of trace averaging.

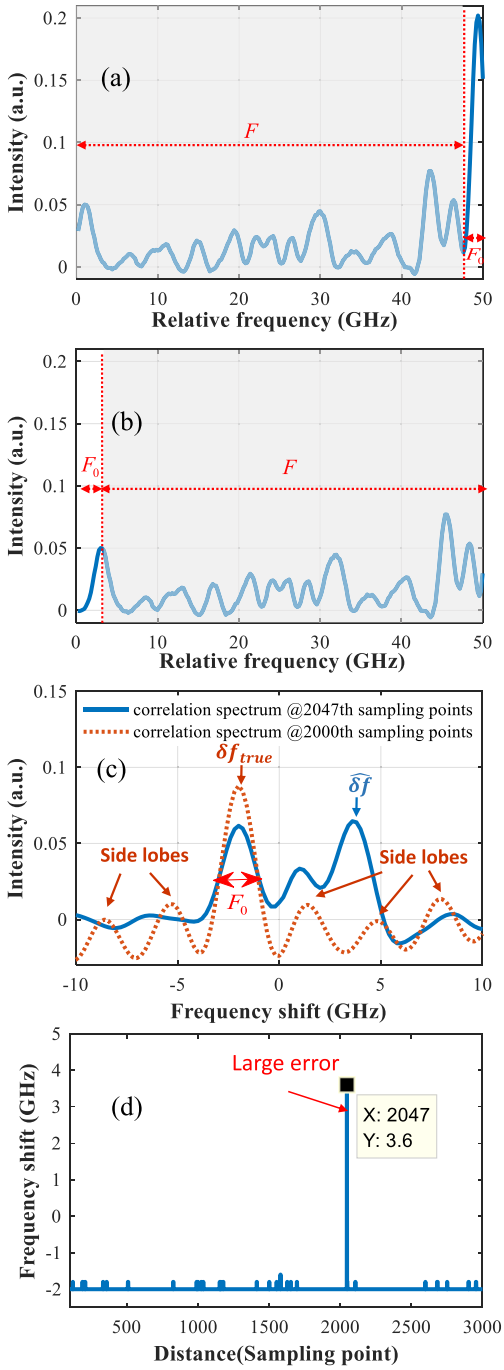

Fig. 2. (a) Reference spectrum and (b) measurement spectrum at the position of the 2047th sampling point. (c) Correlation spectra at the 2047th sampling point (with a large error) and at the 2000th sampling point (without large error). (d) Estimated FS profile obtained by CC showing the occurrence of a large error.

An example of the occurrence of large errors in the estimation of frequency shift by cross-correlation is illustrated in Fig. 2. We scan a total frequency range F_t of 50 GHz for the reference and then deliberately change the temperature of the fiber, so that δf_{true} should be around -2 GHz in the measurement spectrum.

Fig. 2(a) and (b) show, respectively, the reference and measurement spectra at the 2047th sampling point along the fiber. We can see that most part of the reference spectrum is preserved in the measurement spectrum (see shaded spectral section with a frequency range of *F* in and (b) respectively). Fig. 2(c) shows the cross-correlated spectra between the reference and measurement at two positions along the fiber (points 2000th and 2047th). Although only a small shift of around -2 GHz is introduced in the measurement spectrum, the peak position in the correlation spectrum at position 2047th is not located at the right position, unlike the 2000th sampling point (see Fig. 2(c)). Thus, upon the FS profile shown in Fig. 2(d), a huge spike at the 2047th sampling point can be observed, representing a large estimation error (much larger than ½ F_c).

From Fig. 2(c) it can be also observed that, at the 2047th point where the large error takes place, the spectral location of this misleading peak matches the frequency difference between the highest peaks in both the reference and measurement spectra, respectively. This can be explained by the fact that the CC spectra corresponds to the windowed convolution of the reference and measurement spectra (i.e., the integral of multiplied vectors, as a function of the relative shift). Hence, the spectral data with higher intensity contributes much more to the correlation magnitude. In other words, the CC process gives more weight to the high-intensity points and almost no weight to the low-intensity ones. From the point of view of shift estimation, however, all points carry the same amount of information. In the case of Rayleigh spectra, the probability density function (PDF) of the scattered intensity obeys a negative exponential distribution [29], which means that most of the points in the spectra are low-intensity points. Whenever a high intensity point appears in the finite analyzed spectral window, the peak cross-correlation value may be biased towards this new position, giving rise to large errors.

However, this feature is not restricted to exponential distributions. Any data statistical distribution would lead to a similar failing. The key issue is the finite analyzed spectral window. As observed in Fig. 2(a) and Fig. 2(b), the measurement spectrum at a given position can be divided into a correlated signal within a spectral range *F* and an uncorrelated spectral section appearing in the measurement due to an induced temperature or strain change (within a spectral range F_0). Thus the total measured spectral range can be described as $F_t = F_0 + F$. Only the correlated spectral section will contribute to the FS estimation through cross-correlation while the uncorrelated spectra will essentially give irrelevant contributions. If the spectral response in this region is high enough, this uncorrelated area may induce a spurious peak higher than the true correlation peak. This is the essence of the large errors when using CC as shift estimation method.

In fact, CC is a common method for shift/time delay estimation found in several applications, such as sonar and radar [30], [31]. The appearance of large errors imposed by CC for stochastic signals has already been observed and modeled in time delay estimation [31]–[33]. The method for FS estimation in φ-OTDR and OFDR is mathematically equivalent to the time delay estimation problem, which means that we can use a similar approach to model it. By adapting the model from time delay estimation literature [33], the PLE in φ-OTDR can be written as:

$$PLE \approx (M-1)\sqrt{\frac{1}{2\pi C}}\exp\left(-\frac{C^2}{2}\right) \tag{4}$$

where

$$C = \sqrt{B_s F}\frac{SNR}{\left[(SNR)^2 + 2(SNR+1)^2\right]^{1/2}} \tag{5}$$

and $M = 2F_0/F_c$, a unitless coefficient scaling the relative importance of the uncorrelated spectral range F_0 with respect to the width of the correlation peak F_c. B_s is the statistical bandwidth which is defined as:

$$B_s = \left[\int_{-\infty}^{\infty}\rho^2(f)df\right]^{-1} \tag{6}$$

where $\rho(f)$ is the normalized correlation function of the signals [26]. For example, for a rectangular pulse of width τ and considering an exponential delay distribution, the calculated $B_s = 1.38\tau$. Note that the SNR in the above expressions is the ratio between the variance of the spectrum and the variance of the noise across the full measured spectral window, hence the impact of fading signals is averaged out. Under the assumption of $SNR \gg 1$ we have:

$$\frac{SNR}{\left[(SNR)^2 + 2(SNR+1)^2\right]^{1/2}} \approx 1/\sqrt{3} \tag{7}$$

Therefore, the expression of PLE can be simplified as

$$PLE \approx \frac{\sqrt{\frac{3}{2\pi}}\left(2F_0/F_c - 1\right)\exp\left(-\frac{B_s F}{6}\right)}{\sqrt{B_s F}} \tag{8}$$

For a rectangular pulse, $F_c = 1/\tau$ and $B_s = 1.38\tau$, so that Eq. (8) can be rewritten as

$$PLE \approx \frac{\sqrt{\frac{3}{2\pi}}\left(2F_0\tau - 1\right)\exp\left(-\frac{1.38\tau F}{6}\right)}{\sqrt{1.38\tau F}} \tag{9}$$

From Eq. (9), it is possible to observe that: with a certain pulse width τ, the PLE is linearly related to F_0 when $F_0\tau \gg 1$, while the frequency range *F* decreases the PLE exponentially. Intuitively speaking, the spectra in the range of *F* for reference and measurement are correlated and therefore contribute to the same peak in the correlation spectrum. Consequently, this peak, placed at δf_{true}, grows exponentially with the increase of *F*, leading to an exponential reduction of the PLE. On the other hand, any perturbation applied on the fiber (such as temperature or strain change) leads to a loss of correlation in the data points, which are replaced by uncorrelated data over a range F_0. This feature contributes to the occurrence of random side lobes in the cross-correlation spectrum, showing an increasing probability to turn higher than the true correlation peak. As a result, the increase of F_0 increases the PLE linearly (when $F_0\tau \gg 1$).

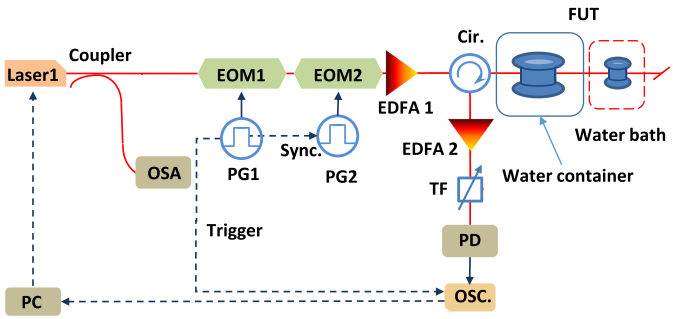

Fig. 3. Experimental setup of frequency-scanned φ-OTDR (EOM: electro-optical modulator, EDFA: erbium-doped fiber amplifier, Cir: circulator, FUT: fiber under test, TF: tunable filter, PD: photodetector, OSC: oscilloscope, PC: personal computer, PG: pulse generator, Sync, synchronization).

IV. Comparisons of the Theoretical and Experimental PLE

To investigate the impact of the measurement parameters, such as the pulse width, the scanned frequency range and the scanning frequency step on the PLE, a direct-detection based φ-OTDR is implemented to obtain the experimental spectra along a sensing fiber.

The experimental setup is depicted in Fig. 3. A DFB laser with 1 MHz linewidth is used as a coherent light source. Pulses are generated with an extinction ratio high enough for the number of resolved points along the fiber with two cascaded electro-optical modulators (EOMs). The pulse width is set to 0.5–1 ns, thereby allowing a spatial resolution of 5–10 cm. An erbium-doped fiber amplifier (EDFA) is then inserted to boost the pulse peak power, though at a level below the onset of nonlinear optical effects in the sensing fiber. The coherent pulse is launched into the sensing fiber and the back-captured Rayleigh scattering is then pre-amplified using another EDFA before photo-detection. The amplified spontaneous emission (ASE) noise introduced by this EDFA is filtered out by an optical filter with a bandwidth of 1 nm. The bandwidth of the PD is 3 GHz. An oscilloscope operating at a sampling rate of 5/10 Gs/s is used to digitize the electrical signal for further analysis. The sensing fiber here is a single-mode fiber with a length of about 860 m, from which 5 m are immersed into a water bath to apply temperature changes (see later Section V).

Here, the sweep of the optical frequency is performed by directly modulating the temperature of the laser to cover a relatively large scan range of 100 GHz, using a remotely controlled laser power supply. In the later experiment using least mean squares (LMS) for the frequency shift estimation, a laser current scan is implemented for the live measurements to achieve a fast frequency sweep. The frequency step Δf here is 100 MHz and 200 MHz when using 1 ns pulse and 500 ps pulse, respectively. To secure a high enough SNR, each retrieved trace is averaged 4000 times.

Using this scheme, the coherent Rayleigh spectrum over the entire fiber is measured with spatial resolutions of 5 cm and 10 cm. To obtain the experimental PLE for different scanning ranges F_t, the following procedure was applied:

a) Obtain the Rayleigh spectrum of the fiber over a range of 100 GHz: $X_{i,0}(f)$, $f \in$[0, 100 GHz]. Here, i denotes the position along the fiber and j denotes the measurement sequence;
b) Choose 'reference' and 'measurement' data from $X_{i,j}(f)$. For a certain scanning range F_t and frequency shift F_0, let the reference spectrum be $S_{i,0}(f) = X_{i,0}(f)$, $f \in [0,\ F_t]$ and the measurement spectrum be $S_{i,j}(f) = X_{i,0}(f)$, $f \in [F_0,\ F_0 + F_t]$. Therefore, the spectrum of the reference $S_{i,0}(f)$ and the measurement $S_{i,j}(f)$ are subsets from the same data set, will span over a total range F_t and the true FS is $\delta f_{true} = F_0$. Since both $S_{i,0}(f)$ and $S_{i,j}(f)$ are taken from $X_{i,0}(f)$, the two spectra can be considered to have perfect SNR (effectively there is no variance between the two correlated spectra);
c) Cross correlate $S_{i,0}(f)$ and $S_{i,j}(f)$ to get the correlation spectrum $C_{i,j}(\delta f)$ and obtain the estimation using Eq. (2);
d) Determine if $\varepsilon_{i,j}$ is a large error by determining if $\varepsilon_{i,j} = |\delta f_{i,j} - \delta f_{i,jtrue}| \geq \frac{1}{2}F_c$;
e) Repeat steps a-d for each sampled position to get the total number of large errors. PLE is estimated as the ratio between the number of large errors and the total number of sampled positions.

The experimental results are obtained through the aforementioned process, and the theoretical results are calculated by Eq. (9) by replacing F with $F_t - F_0$.

An exponentially decaying tendency of the PLE as a function of the frequency scanning range F_t is shown in Fig. 4(a) and (b), as predicted. Fig. 4(c) shows a comparison of the PLE (in log scale) obtained with two different pulse widths. It can be seen clearly that, with the same frequency scanning range F_t, the PLEs of the longer pulse decrease faster than that of the shorter pulse, demonstrating that the implementation of a robust Rayleigh sensor with high spatial resolution becomes extremely challenging due to the increased probability of large errors. Fig. 4(d) shows that the PLE grows with increasing F_0, but grows much faster when using shorter pulses. This means that the measurable temperature and strain change must be kept smaller using high spatial resolution schemes.

It must be noted that similar PLEs will be obtained at high spatial resolution by simply rescaling proportionally the frequency scanning range, since eventually this is the width of the correlation peak normalized to the frequency scanning range that scales the PLE.

Note that in all these cases, the experimental results show a good agreement with the theoretical results, validating the model proposed in Section III. This means that the PLE can be reliably predicted using Eq. (9) when all the measurement parameters are known. For instance, Fig. 4(c) shows that PLE $= 10^{-3}$ when the spectral scanning range is 50 GHz and the pulse width is 500 ps. This means that, in principle, in a 50 m-long sensor with 5 cm spatial resolution (i.e., having only 1,000 sensing points) and high enough SNR, there will be statistically one large error for each measurement. If the scanning range increases up to 70 GHz, the PLE reduces to 10^{-4}, which is statistically still not low enough to allow reliable measurements if the number of sensing points is large.

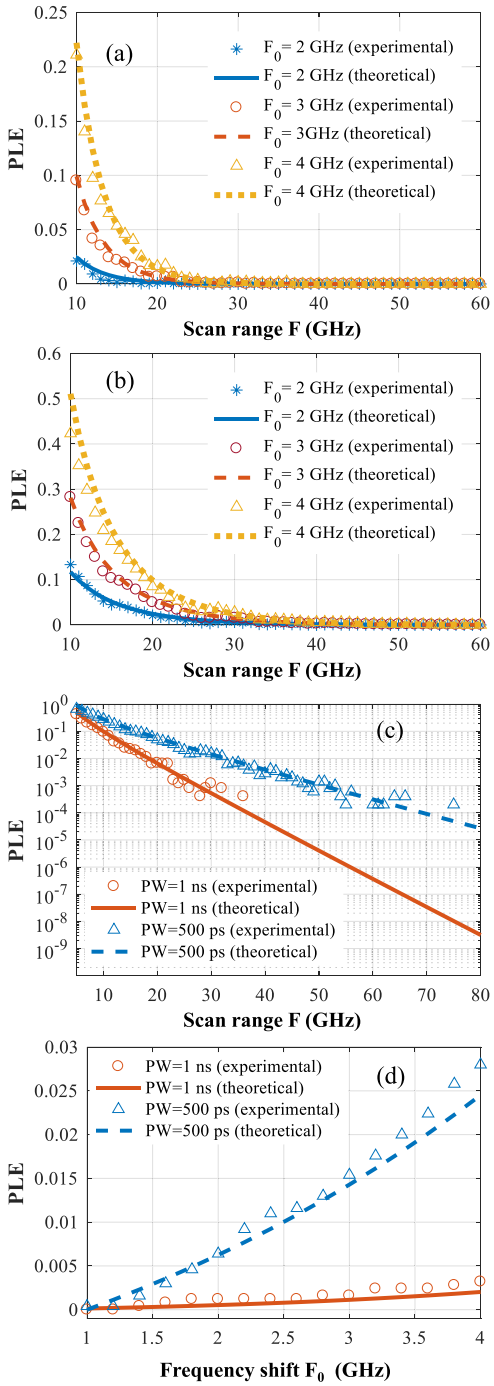

Fig. 4. Comparisons between experimental and theoretical results. The probability of large error v.s. scan range Ft with a pulse width of (a) 1 ns and (b) 500 ps, F_0 is the frequency offset between measurement and reference spectra; (c) comparative PLE using different pulse widths ($F_0 = 3$ GHz); (d) PLE dependence on F_0 ($F_t = 30$ GHz).

It may be argued that the above mentioned large error problem can be alleviated by extending the frequency scanning range of the reference only, while keeping the measurement spectrum bounded to the same value. Let the scanning ranges of reference and measurement be F_t and F_s, respectively, F_t being larger than F_s, as shown in Fig. 5. In this case, the range relevant for correlating the signals can be kept as F (i.e., $F = F_s$) even

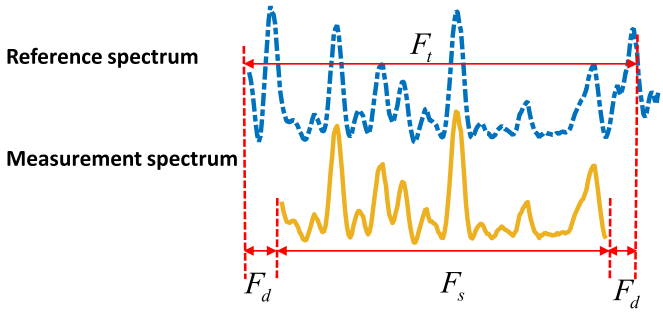

Fig. 5. Diagram to show the different scan ranges of reference and measurement.

through there is a shift δf_{true} in the measurement spectrum, as long as $|\delta f_{true}| \le F_d$, where $F_d = \frac{1}{2}(F_t - F_s)$. However, since the ranges of the reference and measurement are different in this case, there will always exist some spectral samples in the reference that are not contained in the measurement set (thus being uncorrelated) even when there is no temperature or stain change to the fiber. These samples are expected to contribute to the generation of random spurious peaks, increasing the PLE. The FS estimation in this case is made using the equations as below:

$$C_{i,j}(\delta f) = \frac{1}{F_s}\int_0^{F_s} S_{i,j}(f)S_{i,0}(f-\delta f)df, \quad -F_d \le \delta f \le F_d \tag{10}$$

$$\delta f_{i,j} = \max_{-F_0 \le \delta f \le F_0} \arg\, C_{i,j}(\delta f) \tag{11}$$

The experimental results are calculated following the same procedure as before, but in step b, the data used for cross-correlation have been changed into:

b) For a certain measurement scan range F_s and range difference F_0, Let $S_{i,0}(f) = X_{i,0}(f)$, $f \in [0,\ F_0 + F_s]$ and $S_{i,j}(f) = X_{i,0}(f)$, $f \in [0,\ F_s]$;

In this case, the correlated spectral range F equals to the measurement scanning range F_s, while the uncorrelated spectral range F_0 equals to $2F_d$. The results of the calculated PLE through experimental data and the presented analysis method are shown in Fig. 6. An exponential decay tendency of PLEs versus the increase of F_s is again visible, indicating that the PLE grows as the range difference F_0 between measurement and reference increases. It suggests that the larger the difference between the reference scan range F_t and the measurement scan range F_s, the larger the PLE, due to the presence of a larger number of uncorrelated points.

In short, as long as there are different samples entering the spectral scanning windows of measurement and reference, the PLE remains larger than 0, no matter if the new sampled points are due to temperature or strain-induced frequency shift or simply due to differences in the reference and measurement scanning ranges. In principle, only when both reference and measurement scanning ranges are the same and temperature or strain do not change, the PLE reaches zero (under the assumption of no

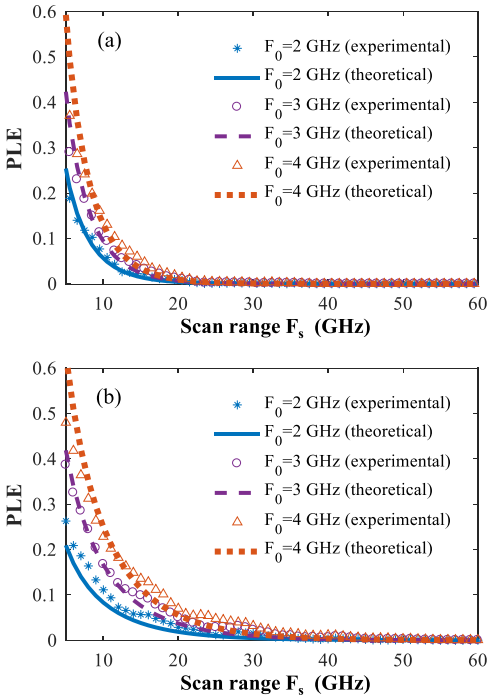

Fig. 6. PLE vs. spectral scanning range of measurement F_s with a pulse width of (a) 1 ns and (b) 500 ps when the scanning ranges of the reference and measurement are different. $F_t = F_s + F_0$, $F_0 = 2F_{\rm d}$ is the frequency range difference between the measurement and reference spectra.

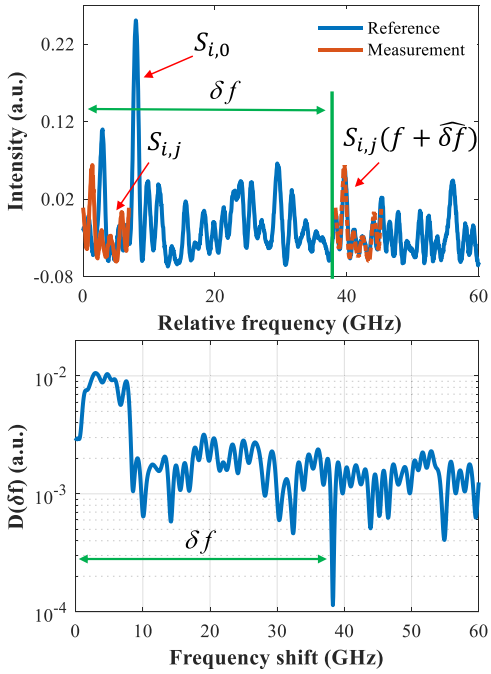

Fig. 7. Principle of the frequency shift estimation using the proposed least mean squares (LMS) method. (a) The measurement spectrum at a given point (red) is swept over the broad reference spectrum (blue) and LMS are calculated for each relative spectral position δf. (b) Mean square errors as a function of the frequency shift δf showing a clear minimum value which is the estimation of the best similarity.

statistical noise). This condition is of course in total contradiction with the purpose of a sensing system.

It must be noted that, all calculations are based on the same measurement, which has been obtained under a very high SNR condition. In normal conditions, under the presence of significantly higher noise levels, the level of correlation between common spectral sections will be reduced, thus increasing the PLE.

V. FS Using Least Mean Squares (LMS) in φ-OTDR

As discussed in the previous section, the PLE using cross-correlation is often not negligible to secure confident sensing conditions. The question arises if another method is more effective to reliably estimate the frequency shift. Here, an approach based on least mean squares (LMS) is proposed for retrieving the frequency shift in Rayleigh sensing, by evaluating the degree of similarity between particular sections of the spectra and estimating the offset between them. Assuming that the scanning ranges of reference and measurement are F_t and F_s respectively, the mean square error between reference and measurement is defined as

$$D_{i,j}(\delta f) = \frac{1}{F}\int_0^F (S_{i,j}(f) - S_{i,0}(f-\delta f))^2 df, \quad -F_d \le \delta f \le F_d \tag{12}$$

and the FS estimation is

$$\delta f_{i,j} = \min_{\delta f \in (-F_d, F_d)} \arg\, D_{i,j}(\delta f). \tag{13}$$

LMS measure the amplitude difference between two spectral regions, so that the estimated frequency shift is the one associated to the minimum LMS. This way the weight of each spectral points is identical–even the most frequently occurring spectral points of low amplitude equally contribute to the estimation–thereby the seldom-occurring high-amplitude points would not lead to large error. In reality, the minimum value will not be zero due to the presence of noise; however, the minimum LMS will be located only at the frequency shift corresponding to the true spectral shift, as illustrated in Fig. 7.

Fig. 8 compares the PLE obtained when using cross-correlation and the proposed LMS method using the same datasets. In Fig. 8(a), PLE is calculated with infinitely high SNR (when $S_{i,j}$ and $S_{i,0}$ are obtained from the same dataset $X_{i,0}(f)$, using different scanning ranges for the reference and measurement). It can be seen that, with no statistical variations (noise), the PLE is always zero using LMS even if the scanning range of the measurement F_s is only 2 GHz. In Fig. 8(b), the $S_{i,j}$ and $S_{i,0}$ used for calculation are taken from different datasets ($S_{i,0}$ is taken from $X_{i,0}(f)$, and $S_{i,j}$ from $X_{i,j}(f)$ $(j > 0)$), hence noise is present. The figure shows that, in presence of noise, the PLE is not always zero when using LMS, as expected, but still, it is significantly reduced by the LMS estimation compared to cross-correlation. The results also clearly indicate that when the scanning range F_s reaches 8 GHz, the PLE becomes negligible in our experiment using LMS. In other words, only a scanning range of 8 GHz is required using the proposed method, significantly speeding the acquisition.

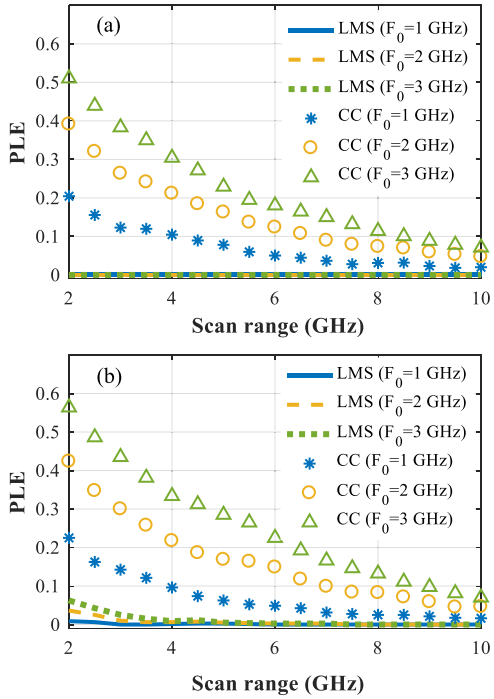

Fig. 8. Comparison of PLE between LMS and correlation with increasing scan range of measurement F_s ($\tau = 1$ ns), (a) with noise; (b) without noise.

To explain the mathematical difference between LMS and CC, we can open the brackets in Equation (12), and rewrite it as:

$$D_{i,j}(\delta f) = \frac{1}{F}\left(\int_0^F S_{i,j}^2(f)df + \int_0^F S_{i,0}^2(f-\delta f)df - 2\int_0^F S_{i,j}(f)S_{i,0}(f-\delta f)df \right), \quad -F_d \le \delta f \le F_d \tag{14}$$

Comparing Eq. (14) with Eq. (10) shows that the minimum of $D_{i,j}(\delta f)$ will be co-located with the maximum of $C_{i,j}(\delta f)$, provided that the first two terms on the right hand of Eq. (14) are constant. This condition is fulfilled when the mean energy of the process does not depend on the size of the integration window, proving the theoretical convergence. Practically, this co-location can only be strictly true in absence of CC large errors, i.e., when the signal and reference windows are infinitely large as stated by Eqs. (8)–(9).

In order to validate the performance of the proposed LMS-based method, a temperature sensing experiment is set up by placing the last few meters of fiber into a temperature controlled water bath to create a hotspot. A 500 ps pulse is used, corresponding to a spatial resolution of 5 cm. The temperature sensing experiments are carried out with a 200 MHz frequency scanning step. A reference measurement over a ∼140 GHz range is first acquired, which is realized by tuning the laser temperature. This lengthy acquisition is made once forever under controlled conditions (stable temperature). In contrast, the spectral scanning for the live measurements is realized by laser current tuning, so that

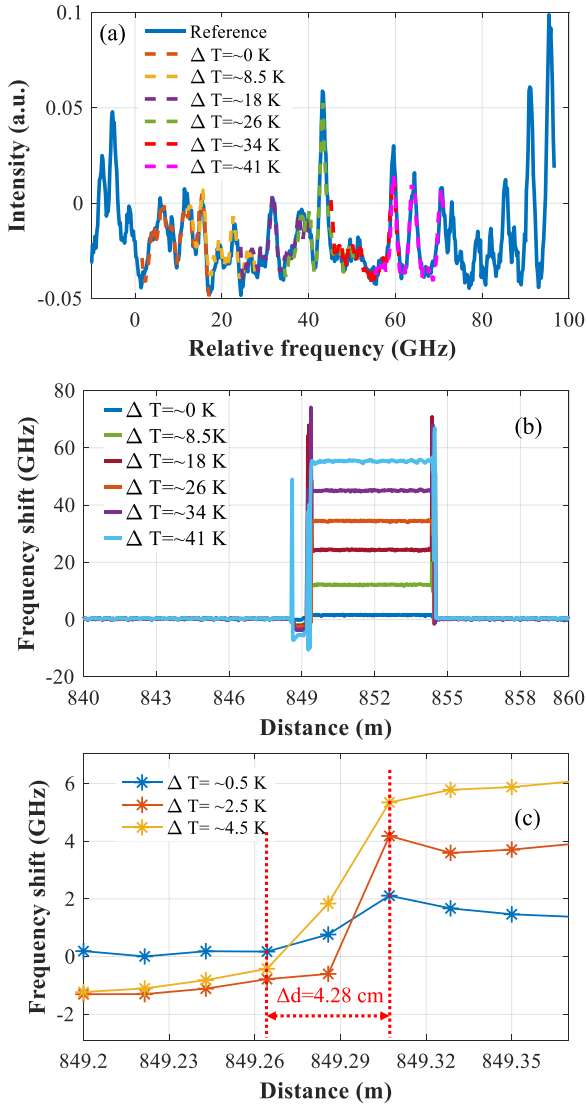

Fig. 9. Experimental results of temperature sensing using the proposed LMS-based φ-OTDR: (a) Spectra of the reference (blue line, obtained by tuning the temperature of the laser) and 6 measurements at different temperatures (at a fiber position of ∼852 m, obtained by tuning the current of the laser); (b) Retrieved frequency shift profiles; (c) frequency shift distributions around the edge of the hotpot which confirms a spatial resolution of less than 5 cm.

the acquisition can be way faster (∼15 s). The scanning range of the measurement spectrum is set to cover 16 GHz, which is much shorter than the reference spectrum. A 100X averaging is performed at acquisition before any data processing. The temperature at the hotspot is varied from 0 K to 41 K with respect to room temperature.

Fig. 9(a) shows the measured spectra for different temperatures at a particular sensing point (distance ∼852 m), along with the reference spectrum (solid blue curve). It clearly shows that the measurement spectra match well with the corresponding sections of the reference spectrum. These results also suggest that:

1) The maximum measureable temperature range is actually limited by the spectral range of the reference but not that of the measurement. Therefore, a large FS range can be attained by only extending the reference spectrum while

maintaining the high sensitivity of the system and a fast acquisition for sequential measurements. This is in clear contrast with the CC estimation.

2) A measurement scanning range containing only few peaks (e.g., 2∼3) in the spectral window is sufficient to estimate the frequency shift without large errors. This provides a crucial advantage in terms of measurement speed, since the number of scanning points for the measurement spectrum can be reduced. Besides, the measurement time is further reduced since the minimum scan range of the measurement is only a few tens of GHz (16 GHz in this experiment), the scan of the optical frequency is performed by tuning the current of the laser, which is practically faster settled than tuning the temperature of the laser.

The profiles of the frequency shift along the end section of the fiber are shown in Fig. 9(b), estimated using LMS at different temperatures. In the hotspot (starting from ∼849.5 m), clear frequency shifts can be seen without large errors, being also in good agreement with the preset water temperature. A few glitches are present at fiber positions corresponding to transitions to and from the heated water bath, where the temperature is unstable and non-uniform over the interrogating spatial resolution. A spatial resolution of less than 5 cm can be verified in Fig. 9(c), showing the frequency shift distributions around the front edge of a hot spot and confirming a transition in the measured profiles of 4.28 cm.

Lastly, the frequency shift repeatability is calculated to be 0.026 K, estimated by obtaining the standard deviations of each local estimated FS from 10 consecutive independent measurements, which shows a temperature dynamic range as large as 36.5 dB (an uncertainty of about 1/5000 of the measureable temperature change).

VI. Conclusions

The probability of large errors (PLE) occurring in φ-OTDR employing frequency shift estimation through cross-correlation is addressed for the first time in this paper. It turns out to be a critical limitation and PLE occurrences are essentially unpredictable. According to the proposed model, the PLE decreases exponentially with the frequency scanning range and grows linearly with the applied frequency shift (when $F_0\tau \gg 1$). The results of our analysis agree well with the experimental results. The analytical model clearly shows that large errors on the FS estimation of φ-OTDR unavoidably occur when using cross-correlation techniques due to purely stochastic reasons related to the limited spectral window of analysis, even in conditions of perfect signal to noise ratio. Moreover, we have proposed a method to circumvent these large errors, employing a least mean squares estimation of the similarity between reference and measurement spectra. Using this method, the stochastic cause of large errors can be totally suppressed and becomes only limited by noise. The maximum temperature measurement range can be safely enlarged while keeping a limited spectral scan for running acquisitions. As a result, a φ-OTDR system with a spatial resolution of 5 cm and sensing range of about 860 m is demonstrated, which has a dynamic range of temperature change of about 36.5 dB and very robustly immune to large statistical errors.

Finally, it should be noted that large errors do not only occur in correlation-based φ-OTDR. A similar model can be developed for any Rayleigh-based distributed sensor employing frequency shift estimation of the backscattered spectra based on cross-correlation, such as coherent OFDR.

References

[1] L. Thévenaz, "Brillouin distributed time-domain sensing in optical fibers: State of the art and perspectives," *Front. Optoelectron. China*, vol. 3, no. 1, pp. 13–21, 2010.

[2] K. Yuksel, M. Wuilpart, V. Moeyaert, and P. Mégret, "Optical frequency domain reflectometry: A review," in *Proc. 11th Int. Conf. Transp. Opt. Netw.*, 2009, pp. 1–5.

[3] X. Bao and L. Chen, "Recent progress in distributed fiber optic sensors," *Sensors*, vol. 12, no. 7, pp. 8601–8639, 2012.

[4] A. Masoudi and T. P. Newson, "Contributed review: Distributed optical fibre dynamic strain sensing," *Rev. Sci. Instrum.*, vol. 87, no. 1, 2016, Paper no. 011501.

[5] Y. Muanenda, "Recent advances in distributed acoustic sensing based on phase-sensitive optical time domain reflectometry," *J. Sens.*, vol. 2018, pp. 1–16, 2018.

[6] X. Bao, D. P. Zhou, C. Baker, and L. Chen, "Recent development in the distributed fiber optic acoustic and ultrasonic detection," *J. Lightw. Technol.*, vol. 35, no. 16, pp. 3256–3267, Aug. 2017.

[7] M. K. Barnoski and S. M. Jensen, "Fiber waveguides: A novel technique for investigating attenuation characteristics," *Appl. Opt.*, vol. 15, no. 9, pp. 2112–2115, 1976.

[8] H. F. Taylor and C. E. Lee, "Apparatus and method for fiber optic intrusion sensing," U.S. Patent 5 194 847, 1993.

[9] Z. N. Wang *et al.*, "Ultra-long phase-sensitive OTDR with hybrid distributed amplification," *Opt. Lett.*, vol. 39, no. 20, pp. 5866–5869, 2014.

[10] H. F. Martins, S. Martin-Lopez, P. Corredera, J. D. Ania-Castanon, O. Frazao, and M. Gonzalez-Herraez, "Distributed vibration sensing over 125 km with enhanced SNR using Phi-OTDR over a URFL cavity," *J. Lightw. Technol.*, vol. 33, no. 12, pp. 2628–2632, Jun. 2015.

[11] F. Peng, N. Duan, Y. J. Rao, and J. Li, "Real-time position and speed monitoring of trains using phase-sensitive OTDR," *IEEE Photon. Technol. Lett.*, vol. 26, no. 20, pp. 2055–2057, Oct. 2014.

[12] G. Yang, X. Fan, Q. Liu, and Z. He, "Frequency response enhancement of direct-detection phase-sensitive OTDR by using frequency division multiplexing," *J. Lightw. Technol.*, vol. 36, no. 4, pp. 1197–1203, Feb. 2018.

[13] C. Wang, C. Wang, Y. Shang, X. Liu, and G. Peng, "Distributed acoustic mapping based on interferometry of phase optical time-domain reflectometry," *Opt. Commun.*, vol. 346, pp. 172–177, 2015.

[14] A. Masoudi, M. Belal, and T. P. Newson, "A distributed optical fibre dynamic strain sensor based on phase-OTDR," *Meas. Sci. Technol.*, vol. 24, no. 8, 2013, Paper no. 085204.

[15] Z. Wang *et al.*, "Coherent Φ-OTDR based on I/Q demodulation and homodyne detection," *Opt. Express*, vol. 24, no. 2, pp. 853–858, 2016.

[16] Z. Pan, K. Liang, Q. Ye, H. Cai, R. Qu, and Z. Fang, "Phase-sensitive OTDR system based on digital coherent detection," *Proc. SPIE*, vol. 590, no. 8, 2011, Paper no. 83110S.

[17] G. S. Fang, T. W. Xu, S. W. Feng, and F. Li, "Phase-sensitive optical time domain reflectometer based on phase-generated carrier algorithm," *J. Lightw. Technol.*, vol. 33, no. 13, pp. 2811–2816, Jul. 2015.

[18] S. Liehr, Y. S. Muanenda, S. Münzenberger, and K. Krebber, "Relative change measurement of physical quantities using dual-wavelength coherent OTDR," *Opt. Express*, vol. 25, no. 2, pp. 720–729, 2017.

[19] J. Pastor-Graells, H. F. Martins, A. Garcia-Ruiz, S. Martin-Lopez, and M. Gonzalez-Herraez, "Single-shot distributed temperature and strain tracking using direct detection phase-sensitive OTDR with chirped pulses," *Opt. Express*, vol. 24, no. 12, pp. 13121–13133, 2016.

[20] Y. Koyamada, M. Imahama, K. Kubota, and K. Hogari, "Fiber-optic distributed strain and temperature sensing with very high measurand resolution over long range using coherent OTDR," *J. Lightw. Technol.*, vol. 27, no. 9, pp. 1142–1146, May 2009.

[21] M. A. Soto, X. Lu, H. F. Martins, M. Gonzalez-Herraez, and L. Thévenaz, "Distributed phase birefringence measurements based on polarization correlation in phase-sensitive optical time-domain reflectometers," *Opt. Express*, vol. 23, no. 19, pp. 24923–24936, 2015.

[22] X. Lu, M. A. Soto, and L. Thévenaz, "Temperature-strain discrimination in distributed optical fiber sensing using phase-sensitive optical time-domain reflectometry," *Opt. Express*, vol. 25, no. 14, pp. 16059–16071, 2017.
[23] A. K. Sang, M. E. Froggatt, D. K. Gifford, S. T. Kreger, and B. D. Dickerson, "One centimeter spatial resolution temperature measurements in a nuclear reactor using Rayleigh scatter in optical fiber," *IEEE Sens. J.*, vol. 8, no. 7, pp. 1375–1380, Jul. 2008.
[24] L. Zhang, Z. Yang, F. Gyger, M. A. Soto, and L. Thévenaz, "Rayleigh-based distributed optical fiber sensing using least mean square similarity," in *Proc. 26th Int. Conf. Opt. Fiber Sens.*, Lausanne, Switzerland, 2018, Paper no. ThE29.
[25] K. P. Feng *et al.*, "Improvement of the strain measurable range of an OFDR based on local similar characteristics of a Rayleigh scattering spectrum," *Opt. Lett.* vol. 43, no. 14, pp. 3293–3296, 2018.
[26] S. Liehr, S. Münzenberger, and K. Krebber, "Wavelength-scanning coherent OTDR for dynamic high strain resolution sensing," *Opt. Express*, vol. 26, no. 8, pp. 10573–10588, 2018.
[27] S. Liehr, L. A. Jäger, C. Karapanagiotis, S. Münzenberger, and S. Kowarik, "Real-time dynamic strain sensing in optical fibers using artificial neural networks," *Opt. Express*, vol. 27, no. 5, pp. 7405–7425, 2019.
[28] M. D. Mermelstein, R. Posey, G. A. Johnson, and S. T. Vohra, "Rayleigh scattering optical frequency correlation in a single-mode optical fiber," *Opt. Lett.*, vol. 26, no. 2, pp. 58–60, 2007.
[29] X. Lu, "Coherent Rayleigh time domain reflectometry: Novel applications for optical fiber sensing," Ph.D. dissertation, School of Eng., EPFL, Lausanne, Switzerland, 2016, doi: 10.5075/epfl-thesis-7104.
[30] J. Chen, J. Benesty, and Y. Huang, "Time delay estimation in room acoustic environments: An overview," *EURASIP J. Appl. Signal Process.*, vol. 2006, no. 1, pp. 1–19, 2006.
[31] G. Jacovitti and G. Scarano, "Discrete time techniques for time delay estimation," *IEEE Trans. Signal Process.*, vol. 41, no. 2, pp. 525–533, Feb. 1993.
[32] J. P. Ianniello, "Large and small error performance limits for multipath time delay estimation," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-34, no. 2, pp. 245–251, Apr. 1986.
[33] J. Ianniello, "Time delay estimation via cross-correlation in the presence of large estimation errors," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-30, no. 6, pp. 998–1003, Dec. 1982.

Li Zhang received the B.Sc. degree in optical information science and technology from the Ocean University of China, Qingdao, China, in 2013, and the M.E. degree in optical engineering from the University of Electronic Science and Technology of China, Chengdu, China, in 2016. In 2016, she joined the Group for Fibre Optics, Swiss Federal Institute of Technology of Lausanne, Switzerland, as a Doctoral Research Assistant. Her main research interest is focused on distributed optical fiber sensing based on Rayleigh scattering.

Luis Duarte Costa received the M.Sc. degree in engineering physics from the University of Porto, Porto, Portugal, in 2015. He is currently a Doctoral Research Fellow with the Photonics Engineering Group, University of Alcalá, Alcalá de Henares, Spain. His research activities focus on the research and development of post processing methods for the improvement of distributed acoustic sensing systems.

Zhisheng Yang received the B.E. degree in optoelectronics from the Beijing Institute of Technology, Beijing, China, in 2010, and the Ph.D. degree in communication and information system from the Beijing University of Posts and Telecommunications, Beijing, in 2016. His Ph.D. research work, performed at the Institute of Information Photonics and Optical Communications, was focused on distributed optical fiber sensors.

In 2016, he joined the Group for Fibre Optics, Swiss Federal Institute of Technology of Lausanne, Switzerland, as a Postdoctoral Researcher. His main research interests include optical fiber sensing and nonlinearities in optical fiber.

Dr. Yang is a Member of the Optical Society of America (OSA).

Marcelo A. Soto received the M.Sc. degree in electronic engineering from the Universidad Técnica Federico Santa María, Valparaíso, Chile, in 2005, and the Ph.D. degree in telecommunications from Scuola Superiore Sant Anna, Pisa, Italy, in 2011.

During 2010–2011, he was a Research Fellow with Scuola Sant'Anna, where he worked on distributed optical fiber sensors based on Raman and Brillouin scattering. Later, he was a Postdoctoral Researcher with the EPFL Swiss Federal Institute of Technology of Lausanne, Switzerland, where he worked on high-performance Brillouin and Rayleigh distributed fiber sensing, nonlinear fiber optics, optical signal processing, and optical Nyquist pulse generation. Since March 2018, he has been a Tenure-Track Assistant Professor with Universidad Técnica Federico Santa María, Valparaíso. He also has an invited position as one of the "100 distinguished invited professors" at Guangzhou University, China. He is author or coauthor of more than 160 scientific publications in international refereed journals and conferences, 3 book chapters and 8 patents in the fields of optical communications and optical fiber sensing.

Dr. Soto is a Member of the Optical Society of America (OSA), and he is in the Board of Reviewers of major international journals in photonics.

Miguel Gonzalez-Herráez received the M.Eng. and D.Eng. degrees from the Polytechnic University of Madrid, Madrid, Spain, in 2000 and 2004, respectively. While working toward the D.Eng. degree, he was first as a Research Assistant and then as a Postdoctoral Fellow with the Applied Physics Institute, Spanish Council for Research, Madrid, Spain, and had several long stays in the Nanophotonics and Metrology Laboratory, the Swiss Federal Institute of Technology of Lausanne, Switzerland. In October 2004, he was appointed an Assistant Professor with the Department of Electronics, University of Alcalá, Alcalá de Henares, Spain, where he was promoted to an Associate Professor in June 2006, and to a Full Professor in January 2018. His research interest focuses on the wide field of nonlinear interactions in optical fibers.

Luc Thévenaz (F'17) received the M.Sc. and Ph.D. degrees in physics from the University of Geneva, Geneva, Switzerland.

In 1988, he joined the Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland, where he currently leads a research group (Group for Fibre Optics) involved in photonics, namely fiber optics and optical sensing. Research topics include Brillouin-scattering fiber sensors, slow and fast light, nonlinear fiber optics and laser applications in gases. He achieved with his collaborators the first experimental demonstration of optically controlled slow and fast light in optical fibers, realized at ambient temperature and operating at any wavelength since based on stimulated Brillouin scattering. He also contributed to the development of Brillouin distributed fiber sensing by proposing innovative concepts pushing beyond barriers. During his career, he stayed at Stanford University, at the Korea Advanced Institute of Science and Technology, at Tel Aviv University, at the University of Sydney, and at the Polytechnic University of Valencia. In 2000, he cofounded the company Omnisens that is developing and commercializing advanced photonic instrumentation based on distributed fiber sensing.

Prof. Thévenaz chaired the International Conference on Optical Fiber Sensors, is a co-Executive Editor-in-Chief of *Nature Light: Science & Applications*, and is a Fellow of OSA.