Headspace Oxygen Concentration Measurement for Pharmaceutical Glass Bottles in Open-path Optical Environment Using TDLAS/WMS

Qiwu Luo, *Member, IEEE*, Cao Song, Chunhua Yang, *Member, IEEE*, Weihua Gui, *Member, IEEE*, Yichuang Sun, *Senior Member, IEEE* and Zoe Jeffrey, *Member, IEEE*

Abstract—**Accurate measurement of oxygen concentration for pharmaceutical glass bottles is of great significance to ensure the asepsis of medicine and stability of ingredients. With merits of high sensitivity, low cost, non-contact, and real time response, the wavelength-modulation-based tunable diode laser absorption spectroscopy (TDLAS/WMS) technology shows great potential to achieve in-site oxygen concentration detection by the single-line spectrum measurement. This paper focuses on headspace oxygen concentration measurement in open-path optical environment, which is extremely challenging owing to the short light path length and random ambient noises. First, a signal reconstruction method is established based on discrete-time wavelet packet transform (DWPT), where random noise suppression is implicitly achieved. Then, oxygen concentration is inversed among the data between two adjacent valley values of the demodulated 2nd harmonic signal by multiple linear regression (MLR), and the linear discriminant analysis (LDA) is imported to enhance the information sparsity of 2nd harmonic signal matrix and to address the multi-collinearity problem. Simulation results prove our proposed method achieved considerable detection accuracy with average absolute error of 0.05%. This paper also designed a TDLAS/WMS prototype, the experimental results aiming at glass bottles with different oxygen concentration of 0%, 5%, 10% and 21% in open path optical environment indicate our method has achieved an encouraging average absolute error of 0.54%, and can survive well when the normalized SNR is within 0.85 to 1. These results promise that the proposed methodology can be widely applied in in-site AOI instrumentation of headspace oxygen concentration measurement for glass vials.**

Index Terms—**AOI, TDLAS, WMS, oxygen concentration detection, optical measurement, optical instrumentation, signal reconstruction.**

I. INTRODUCTION

In the pharmaceutical industry, glass medicine bottles are in great demand. In order to ensure the stability and asepsis of the medicine in glass bottles, vacuum or nitrogen is usually used in the process to isolate the medicine from the outside air. However, the sealing integrity of glass medicine bottles would be inevitably degraded due to unstable plugs and covers, the damage of packaging materials or man-made misoperation, leading to gas exchange occurring inside and outside the glass medicine bottles. The resulting entry of microorganisms and oxygen contaminate the sterile drugs accompanied by oxygen reaction, which causes great risks to patients. In order to monitor the quality during the drug encapsulation, it is urgently needed to measure the oxygen concentration in glass medicine bottles. At present, many pharmaceutical companies are sampling to determine whether the oxygen concentration is qualified during the batch of products, or adopting the relevant oxygen analyzer. There are some traditional testing methods, such as magnetic oxygen gas analyzer [1], this is a contact measurement that will damage test samples. Electrochemical oxygen analyzer [2] is used in extensive range, but the lifetime of this sensor is limited. When the oxidation reduction of metals occurs on the poles, this sensor will have an adverse impact on the measurement results. Moreover, it is difficult to realize online detection as such sensing method is also not non-contact. Further, the gas chromatographic oxygen analyzer [3] shows a wide range of applications due to the high detection sensitivity, but its implementation process is relatively complex. With the emerging development of laser chips and electronic techniques, near-infrared absorption spectrum measurement method has been used in the material composition detection [4], [5], [6], [7]. But most typical infrared sources (such as lamps and thermal sources) have very wide (usually up to tens of nanometers) spectral ranges. Nevertheless, this method is still enjoying much popularity in gas sample composition measurement for its smart size and low cost. However, in real-world application, the actual gas composition is often complex and diverse interferes are regularly presented. It is worth noting that the headspace oxygen concentration measurement should be carried out under short distance and open-path optical condition to adapt to the rapid production of drug bottles and the specific mechanical structure of production line. Consequently, how to achieve a long-term, stable and accurate headspace oxygen concentration detection for glass vials in open-path optical environment is still an open issue to both industry and academia.

Tunable diode laser absorption spectroscopy (TDLAS) is a kind of light absorption spectrum detection technology [8], which uses different absorption bands of different molecules to achieve qualitative measurement of the detected gas. And the degree of light absorption by gas molecules is proportional to the concentration of gas according to Lambert-Beer's law. Therefore, the quantitative analysis of gas concentration is converted to a problem of extracting and calculating some

relevant features from the absorption spectrum of gas, which is essential to establish an inversion model between the detected signal and gas concentration. Comparing with other spectral detection techniques, TDLAS possesses prominent natures of extremely narrow spectral line width, adjustable wavelength, preeminent coherence and high spectral power density, making it very popular in online detection of trace gases [9][10][11]. It is not easy to use direct absorption spectroscopy (DAS) for accurate gas concentration measurement especially in actual industrial applications, because noises always stochastically distribute among all frequency bands [12]. Further, a theoretical scheme named wavelength modulation spectroscopy (WMS) has been investigated by Arndt [13] and experimentally proved by Reid and Labrie [14]. The key idea of WMS is to realize a narrowband system for achieving noise suppression by shifting the absorption signal in time domain to a certain carrier frequency point (i.e., kHz grade) in frequency domain, where the characteristic signal (e.g., 2nd harmonic signal) can be precisely and reliably fetched from a low-pass filter closely after coherent demodulation with the corresponding frequency or multiplication frequency. With such a merit, the TDLAS/WMS has increasingly attracted attentions in recent reports for composition measurement [15][16]. Jatana *et al.* [17] designed a TDLAS-based sensor with a Herriott cell based on multi-pass arrangement to solve the challenge of transient oxygen measurement under high gas temperature (up to 800K) environment. Svensson *et al.* [18] constructed a minimalistic single-beam instrumentation based on TDLAS/WMS and applied it on structural analysis of pharmaceutical solids. Zhou *et al.* [19] designed a sensitive oxygen detection system based on TDLAS/WMS, which achieved an uncertainty of 0.05% and a detection sensitivity of 350 ppmv. These considerable results are contributed mainly from its diffused integrating cavity (a gas absorption cell in essence), which permits the absorption optical path can be even as long as several kilometers.

As mentioned above, nearly most of the recent reports on oxygen concentration measurement all focus on the theoretical validation researches with long-path optical length in confined gas absorption cells. Fortunately, open-path-related studies are beginning to emerge. To suppress the optical interference of bottle wall under open-path optical condition, Zhu *et al.* [20] claimed that changing the incident angle of laser could improve the signal integrity of the 2nd harmonic signal. However, the whole study roadmap is based on multiple large and expensive instruments, which limits the wide industrial application [21][22]. For clearer understanding, we summarized the challenges on headspace oxygen concentration detection for glass vials in open-path optical environment as follows:

1) Seriously weak signal analysis. The optical path length within glass bottles is even shorter (12 mm ~ 32 mm for penicillin bottles) than the sum (at least 30mm limited by mechanical structure of production line) of the distances from the bottle to the laser and from the laser to the photodetector. It means that we need to identify the oxygen concentration in a short optical path between two relatively long optical paths filled with air. More than usual difficulty, the frequent environmental parameter (e.g., temperature, humidity, pressure, etc.) variations [23] of outside air bring severe degradation of signal-to-noise ratio (SNR) on detected signal.

2) Multiple stochastic noise impacts. Optical noises caused by etalon effect occur frequently and accumulate on several demarcation surfaces among the light-emitting area of laser, collimating lens, the inside and outside walls of glass vial, and the sensitive surface of photodetector [24][25][26]. These optical interference fringes will directly act as complex noises suffering on absorption spectrum signals then on demodulated harmonic signals [27][28]. Despite all-round signal integrity consideration, random electronic noises still inevitably exist in the electronic devices such as signal acquisition board, signal conditioning circuit, photodetector, demodulator, connecting cables, etc. These diverse and stochastic noises challenge the accurate oxygen concentration identification jointly.

This paper investigates that the discrete-time wavelet packet transform (DWPT) method [29] is noise-robust on harmonic signal decomposition and the multiple linear regression (MLR) method [30] is efficient on data fusion inversion. An oxygen concentration measurement scheme is proposed to aim at the headspace of pharmaceutical glass bottles in open-path optical environment. The contributions are as follows.

1) A DWPT-based 2nd harmonic signal reconstruction method is established to suppress stochastic noises effectively. The recommended parameter settings such as wavelet function and decomposition level are carefully tested and analyzed.

2) An MLR-based oxygen concentration inversion scheme is proposed through data fusion between two adjacent valley values of the demodulated 2nd harmonic signal. A dimension reduction method based on linear discriminant analysis (LDA)[31] is imported to reduce computational complexity, and the multi-collinearity problem is also suppressed to a large extent.

The rest of this paper is organized as follows. Section II includes the related theory preliminaries and the application challenges. Section III elaborates our methodology details and the relative parameter configurations. Section IV presents extensive simulation results, the advantages brought by WMS, DWPT and LDA are progressively analyzed as well. Further, Section V describes the realization details of our TDLAS/WMS prototype, and several actual experiments are carried out to verify our proposed oxygen concentration detection scheme. Finally, Section VI concludes this paper.

II. TDLAS/WMS-BASED MEASUREMENT SCHEME AND APPLICATION CHALLENGES

The basic principle of WMS has been introduced in several articles [13],[14],[15]. According to (A.1) to (A.12) in Appendix A, the concentration (ρ) of absorbing gas obeys a linear relationship to the maximum amplitude (i_{2fmax}) of the 2nd-harmonic absorption signal at center of the absorption line. Therefore, we design an oxygen concentration measurement scheme based on TDLAS/WMS, which consists of *six* modules. As shown in Fig. 1, the laser controller (I) insures that the laser emitter (II) operates at a constant temperature, and its current controller generates a modulating current to drive the laser diode stimulating laser spectrums of around 760 nm under the controlling of the DDS core in the demodulator (IV). After the light beams travel through the open space including the headspace of glass vial, the photodetector (III) gathers the transmission intensity, then the recessed spectrum absorbed by
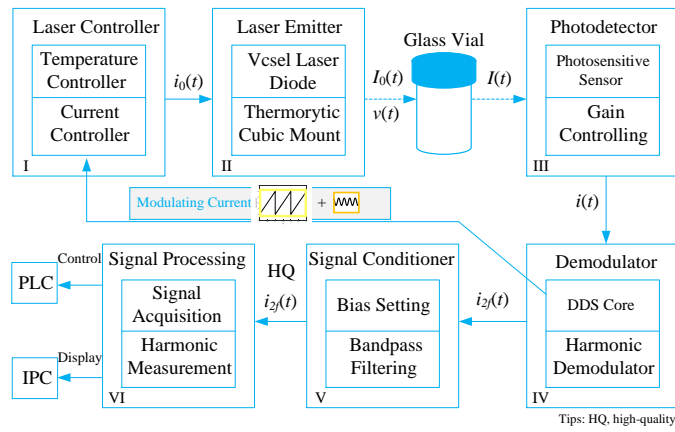
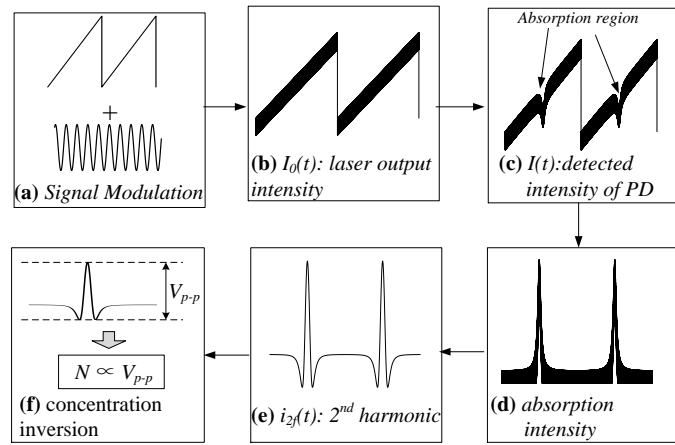Fig. 1. TDLAS/WMS-based oxygen concentration measurement scheme.

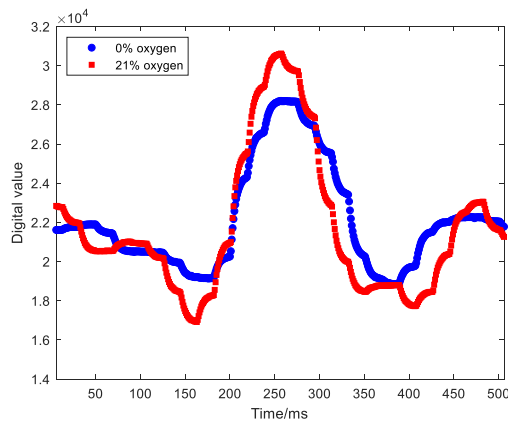Fig.2. Signal flow chart of the TDLAS/WMS realization scheme.

Fig.3. The 2nd harmonic signal of glass bottles with 0% and 21% oxygen concentration obtained from our realized TDLAS/WMS prototype.

O_2 can be fetched after harmonic demodulation in the digital demodulator (IV). The weak demodulated 2nd-harmonic signal will be amplified by the band limiting signal conditioner (V), the bandwidth is set at 0.01Hz~150Hz for robust noise suppression. The signal processing unit (VI) plays a key role of feature extraction and calculation, undertaking the tasks of retrieval of oxygen concentration and the decision of medical glass vial qualification. The decision results drive PLC to control the corresponding transmission devices to locate the glass vials with different quality to different transfer channels.

Finally, a measuring interface is provided on IPC, where data analysis and backup are completed at the same time.

The signal flow chaw of Fig.1 is further illustrated in Fig. 2. As shown in Fig. 2(a) and Fig. 2(b), the laser diode is excited by a superposition waveform of a wavelength scanning ramp wave (Hz grade) and a cosine wave with relatively high frequency (kHz grade). After the laser light beam with the intensity of $I_0(t)$ travels through the headspace of glass vial, the absorbed intensity (i.e., Fig. 2(d)) can be fetched from the detected signal (i.e., Fig. 2(c)), then the absorbed 2nd harmonic signal (i.e., Fig. 2(e)) can be further demodulated from the absorbed intensity. Finally in Fig. 2(f), the oxygen concentration ρ can be inversed by using the absolute peak value of the 2nd harmonic signal according to (24). In real-world application, the absolute peak value is usually replaced by the peak-to-peak value (V_{p-p}) of the 2nd harmonic absorption signal as using relative amplitude (i.e., V_{p-p}) is more robust to bias voltage drift than only using the absolute peak value.

As shown in Fig. 3, the 2nd harmonic signal diagrams of glass bottles with different oxygen concentration of 0% and 21% are taken ahead from Sec. V for better understanding. The 2nd harmonic signal captured from real-world industrial field presents relatively worse signal quality than that of the ideal one which is shown in Fig. 2(e). As claimed in Sec. I, multiple stochastic optical and electronic noises make the 2nd harmonic signal unsmooth and the left and right valleys asymmetrical, which will lead to incorrect detection results if the V_{p-p} is directly used as the inversing criteria of oxygen concentration. This problem is what the whole article concentrates on, and it will be progressively analyzed and addressed by focused signal processing methods, which will be presented in Sec. III.

III. METHODOLOGY AND ANALYSIS

Our detection methodology is constructed by two key parts of signal reconstructor and concentration discriminator. As illustrated in Fig. 4, the bias disturbances and random noises entangled on the raw 2nd harmonic signal are significantly suppressed when it passes through the DWPT-based signal reconstructor. Then the oxygen concentration can be reliably identified after the filtered signal passes through the concentration discriminator enhanced by MLR and LDA. Following contents will present more details about our method.
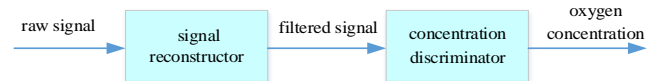

Fig.4. Data processing flow chat

A. Noise-robust signal reconstruction based on DWPT

In the open-path optical environment, the excited noises superimposed on the 2nd harmonic signal always tend to be random, the discrete wavelet transform (DWT) possesses some merits to refine key features in wavelet domain compared with method in frequency domain (e.g., DFT: discrete Fourier transform). But the DWT is also reported that it is suffering with poor frequency resolution in the high frequency band as it only decomposes the low frequency band. Discrete wavelet packet transform (DWPT) is a signal analysis method which divides the frequency band at multiple levels and selects the best basis function adaptively according to the signal

characteristics. By comparing with other signal analysis methods, DWPT possesses prominent natures of high accuracy and small amount of calculation, making it enjoy increasing popularity in analyzing nonstationary signals [32] [33]. Thus the DWPT is adopted and plays a key role in the signal reconstructor.

According to [29], the DWPT coefficients at any scale j is obtained from the convolution of the sampled raw signal with the infinite impulse response filters g and h, as follows:

$$s_j^{2z}(k) = \sum_{n=-\infty}^{+\infty} g(n) s_{j-1}^z (2k-n) \qquad (1)$$

$$s_j^{2z+1}(k) = \sum_{n=-\infty}^{+\infty} h(n) s_{j-1}^z (2k-n) \qquad (2)$$

where s_0^0 is the raw signal, z is the node number, the node zero component $s_j^0(k)$ represents the decomposition packet coefficient of lowest frequency band at scale j, whereas at any other node, i.e., for ($z \neq 0$), $s_j^z(k)$ represents the decomposition packet coefficient of the higher frequency bands at scale j.

The scaling and wavelet filters $g(\ )$ and $h(\ )$ present the following properties [34]

$$\sum_{n=-\infty}^{+\infty} g(n) = \sqrt{2}, \ \sum_{n=-\infty}^{+\infty} g^2(n) = 1, \ \sum_{n=-\infty}^{+\infty} g(n)h(n) = 0 \qquad (3)$$

$$\sum_{n=-\infty}^{+\infty} h(n) = 0, \ \sum_{n=-\infty}^{+\infty} h^2(n) = 1, \ \sum_{n=-\infty}^{+\infty} h(n)g(n) = 0 \qquad (4)$$

attending the finite energy signal and zero mean conditions and the admissibility.

The DWPT processing of the 2nd harmonic signal with oxygen concentration of 0% is illustrated in Fig. 5. The working process of signal reconstructor can be divided into two main parts: First, the raw 2nd harmonic signal sequences $x(n)$ are decomposed into low- and high-frequency components after several DWPT operations. Then the inverse DWPT (IDWPT) performs on some principal decomposed components to realize signal reconstruction. The detailed steps are presented as follows:

Step 1. We choose an adaptable wavelet function as mother wavelet function. Then, the raw discrete 2nd harmonic signal sequences $x(n)$ go through a set of discrete time low- and high-pass filters h and g.

Step 2. The down-sampling operation is applied to the signal passing through the filters. Thus, the raw signal sequences $x(n)$ is decomposed into its high- and low-frequency components at scale j, i.e., decomposition coefficient s_j^{2z} and s_j^{2z+1}, respectively.

Step 3. Decomposition is successively applied to s_j^{2z} and s_j^{2z+1} in order to obtain subsequent high- and low-frequency bands down to the desired scale L. The frequency of each band is set to $f_s/2^j$ and f_s represents the sampling frequency. And we suggest that an ideal L should support the central frequency component and some of its multiplicate frequency components of the raw 2nd harmonic signal can be completely accessed. Particularly in this paper, the sampling frequency f_s and the 2nd harmonic signal frequency (i.e., the frequency of the scanning ramp-wave) are respectively 12800Hz and 25Hz (refer to Sec. V.A), hence the L is set to 9, that is, the 2nd harmonic signal sequences $x(n)$ are decomposed into 512 equal frequency bands
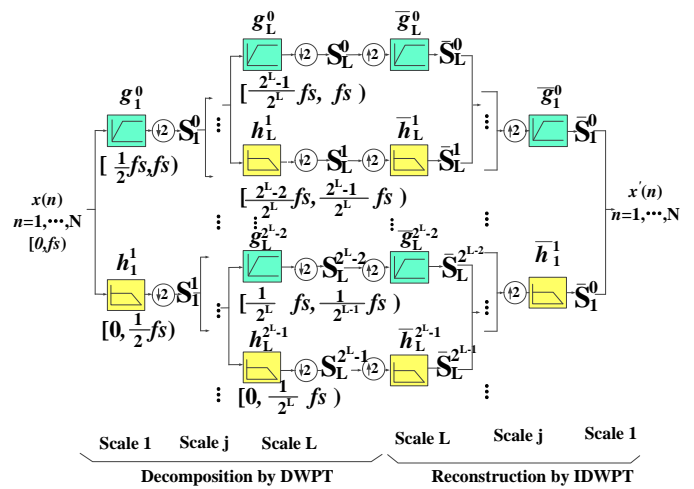

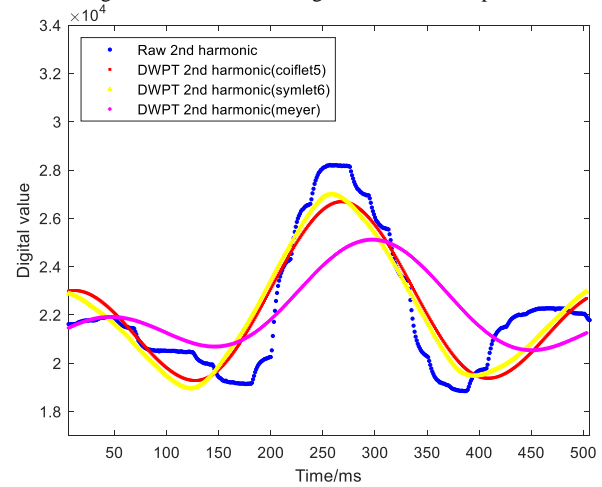Fig.5. The DWPT-based signal reconstruction process.

Fig.6. Comparison reconstruction effects with different wavelet functions.

with each band width of 25Hz, which ensures the multiplicate frequency components based on 25Hz are accessible.

Step 4. Step into the reconstruction: The series of s_j^{2z} and s_j^{2z+1} will go through a set of high- and low-pass filters $\bar{g}$ and $\bar{h}$ with distinct bandwidths. For obtaining principal decomposed components of the 2nd harmonic signal, the bandwidth of the filter array will be set properly to jointly preserve the signal components around the central frequency of the 2nd harmonic signal while suppressing those of others. In particular in this paper, the bandwidth is calculated by the cross-correlation operation between the raw 2nd harmonic signal and each wavelet packet coefficient. The larger the cross-correlation coefficient is, the broader its corresponding filter bandwidth will be set, then this band of packet coefficient will pass through the filter freely. In contrast, the smaller the coefficient is, most of the components will be suppressed in this wavelet band or even be abandoned. It is interesting to note that the three components (i.e., 0Hz, 25Hz, 50Hz) and some adjacent frequency multiplicated components have been approved while others have not in this paper. For more precise analysis, higher scale L can be set, and we set L equals 9 to support embedded implementation as this paper focuses more on in-site detection.

Step 5. The operations of up-sampling and summation are carried out successively from scale L, then to scale L-1, …, finally to scale 1 until the filtered 2nd harmonic signal $x'(n)$ is

reconstructed and most noises can be eliminated effectively.

It is worth mentioning that the wavelet function should be carefully selected for accurate signal reconstruction. Here we take the 2nd harmonic signal of 0 % oxygen concentration for illustration and three typical wavelet functions of Coiflets, Symlets and Meyer are chosen as mother wavelet function by turn for comparision. As shown in Fig. 6, it can be clearly observed that, when using Symlet6 function, the signal envelope is slightly asymmetric, that is the left and right valleys of the reconstructed signal are not equal. When we use Meyer, the signal amplitude drops sharply and the signal shape is still not so symmetric. In contrast, the final reconstructed signal using Coiflet5 function describes the raw principal trend well with a symmetrical envelope and small amplitude attenuation. The reconstructed (or say filtered) 2nd harmonic signal will act as the reliable data source for the following oxygen concentration measurement.

B. Lightweight concentration discrimination based on LDA

Linear discriminant analysis (LDA) is a feature extraction technique that projects high-dimensional pattern samples into the optimal discriminant vector space. In the traditional gas concentration inversion methods [15], the concentration is simply determined by the second harmonic peak-to-peak value (V_{p-p}) according to the linear relationship between the V_{p-p} and the oxygen concentration. However, the V_{p-p} will drift due to the random noise in the open-path optical environment, which makes the results unreliable. Multiple linear regression (MLR) can effectively reduce the identification error of oxygen concentration caused by V_{p-p} drift by the means of analyzing multiple data information between two valleys of 2nd harmonic signals. However, the multicollinearity of the data still brings calculation error. Therefore, in order to achieve high-accuracy measurement of oxygen with low concentration, this paper chooses the MLR coupled by the LDA (here after, abbreviated as D_{v-v}(LDA)) contributing our concentration discriminator, and the detection speed is also promoted to a large extent due to the dimension reduction brought by LDA. The detailed process is briefly given as follows.

The 2nd harmonic signal matrix X consists of 2nd harmonic signal with different concentrations.

$$X = \begin{bmatrix} I_{11} & I_{12} & \cdots & I_{1\beta} \\ I_{21} & I_{22} & \cdots & I_{2\beta} \\ \vdots & \vdots & \cdots & \vdots \\ I_{\alpha 1} & I_{\alpha 2} & \cdots & I_{\alpha\beta} \end{bmatrix}_{\alpha \times \beta} \quad (5)$$

where β is the number of data between the two adjacent valleys of 2nd harmonic signal; α is the number of sample concentrations; $I_{11}, I_{12},…, I_{1\beta}$ are the amplitudes of each point recorded at 1% oxygen concentration; $I_{21}, I_{22},…, I_{2\beta}$ are the amplitudes of each point recorded at 2% oxygen concentration, similarly, $I_{\alpha 1}, I_{\alpha 2},…, I_{\alpha\beta}$ are the amplitudes of each point recorded at α% oxygen concentration.

The matrix $S_w^{-1} \cdot S_b$ can be determined by calculating the within-class scatter matrix S_w and the between-class scatter matrix S_b of X. The loading matrix K can be determined by calculating the eigenvalues and eigenvectors of matrix $S_w^{-1} \cdot S_b$. In a certain experiment, there are in total k nonzero eigenvalues

and the corresponding eigenvectors are chosen to constitute the loading matrix K (more detailed calculations can be referred to [31]). The score matrix, T, can be obtained from:

$$T = X \square K \quad (6)$$

where T is given by

$$T = \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_\alpha \end{bmatrix} = \begin{bmatrix} I'_{11} & I'_{12} & \cdots & I'_{1k} \\ I'_{21} & I'_{22} & \cdots & I'_{2k} \\ \vdots & \vdots & \cdots & \vdots \\ I'_{\alpha 1} & I'_{\alpha 2} & \cdots & I'_{\alpha k} \end{bmatrix}_{\alpha \times k} \quad (7)$$

where, t_i ($i=1,…,\alpha$) is 2nd harmonic signal score vector reduced dimension by LDA at i% oxygen concentration, $I'_{i1}, I'_{i2},…, I'_{ik}$ are the element of t_i at i% oxygen concentration. Since there is little multi-collinearity in the score matrix T, the 2nd harmonic information is utilized as much as possible to calculate the oxygen concentration. Therefore, the MLR method can be used for quantitative analysis of the oxygen concentration by taking the score matrix T as the source of new 2nd harmonic signal data. In addition, the dimension of the matrix X is reduced to k dimension, that is, the dimension of data is reduced by k/β, and the corresponding calculation amount is reduced. The relationship between the concentration vector C and the score matrix T is given as

$$C = T \square B \quad (8)$$

where C is given by

$$C = [\rho_1 \ \rho_2 \cdots \rho_\alpha]^T \quad (9)$$

where, ρ_1 is 1%, ρ_2 is 2%, … , ρ_α is α%; B is the regression coefficients vector, which is given by

$$B = [b_1 \ b_2 \cdots b_n]^T \quad (10)$$

So, B can be calculated by

$$B = (T^T T)^{-1} T^T C \quad (11)$$

According to (7) - (11), the concentration, ρ_{detect}, of the gas sample to be detected can be given as:

$$\rho_{det ect} = t_{det ect} B = t_{det ect} (T^T T)^{-1} T^T C \quad (12)$$

where t_{detect} is the 2nd harmonic signal score vector calculated by the signal reconstructor of a gas sample under test.

IV. SIMULATION AND DISCUSSION

The section carries out a series of comparative simulation tests to evaluate the performance of our proposed framework of WMS + DWPT + D_{v-v}(LDA).

A. Simulation setup

Typical noising 2nd harmonic signals under average optical environments in the actual industrial field have been emulated here for performance evaluation. For fair comparison, identical optical and electronic noises are manually added to (A.6), where the *original* (A.6) and (A.6) *without cosine terms* are respectively fed into the WMS and DAS method as the absorption spectroscopy signals to be analyzed. Particularly, we add a 50Hz sinusoidal wave with slight amplitude fluctuation to emulate optical noises, then we mix white Gaussian noises with a mean value of 0 and the standard deviation of 1 to emulate electronic noises. And both the amplitudes of the optical and electronic noises are restricted within one tenth of the peak value of direct absorption signal.
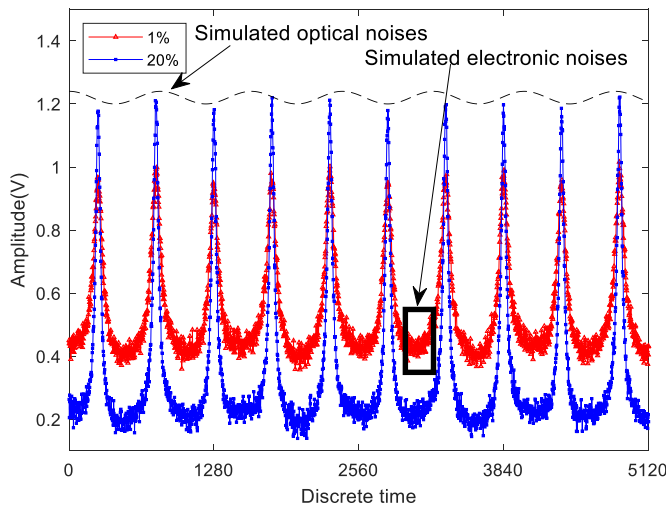
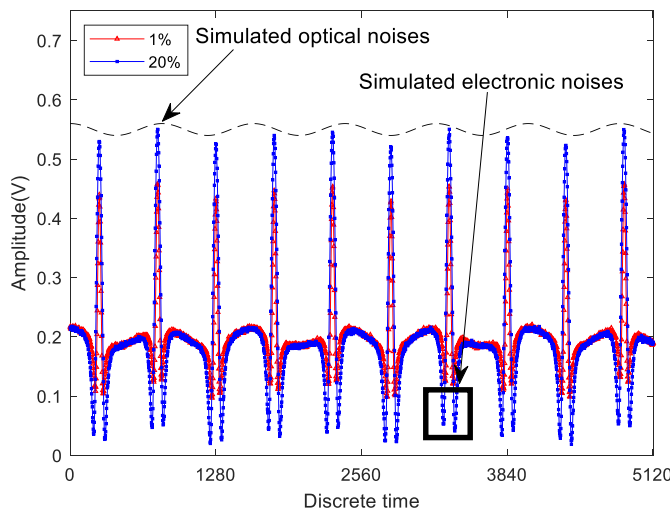Fig.7. Simulated direct absorption peak waveforms, for the DAS method.

Fig.8. Simulated 2nd harmonic waveforms, for the WMS method.

For convenience, the constant terms of $S(T)$, L and P in (A.6) are set to 1, the constant terms of Δv_c and v_0 in (A.7) are set to 2 and 0, respectively. And to consistent with the experiment setup in Sec. V, the frequency of the scanning ramp-wave is set to 25Hz, the frequency of the modulation cosine wave is set to 14.4 kHz, and the sampling frequency f_s is set to 12800Hz (that is, each period of simulated wave from is digitalized into discrete 512 points). Under these configurations, the oxygen concentration using DAS method can be inversed directly from (A.6) where its cosine terms are dropped. While the oxygen concentration using WMS method can be calculated (A.12). In this paper, we generate 20 groups (i.e., α=20 in (5)) of direct absorption peak waveforms and 20 groups of 2nd harmonic waveforms for evaluating DAS and WMS method, respectively. The each 20 groups of absorption waves simulate the detected signals acting as concentration inversion criterion for 20 kinds of glass vial samples with 20 discrete oxygen concentration (i.e., 1%, 2% ..., 20%). For more vivid illustration, Fig. 7 and Fig. 8 illustrate two groups (for 1% and 20% oxygen concentration) of the simulated direct absorption peak waveforms and the 2nd harmonic waveforms, respectively. In order to achieve accurate measurement, each round of simulation is repeated 23 times, thus, there are 460 inversed concentration values for each method. With these emulated wave signals, oxygen

concentration can be inversed by analyzing several signal features (V_p, V_{p-p}, etc.) flexibly. In the signal reconstruction process, as recommended in Sec. III, we select the Coiflet5 as the wavelet function and we set the decomposing scale L to 9. In the concentration discrimination process, the signal between two valleys covers 165 points (i.e., β=165 in (5)) of its resident full period with 512 points. The LDA method reduces the dimension of X in (5) from 20×165 to 20×15 of T in (7).

B. Results and discussion

1) DAS+V_{p-p} vs. WMS+V_{p-p}

We choose the scheme of DAS+V_{p-p} as the baseline in these simulation tests. As described above, there are 20 groups of scores according to 20 increase steps of oxygen concentration distributed in Fig. 9 for the detected concentration (ρ1, ρ2,…, ρ20 stands for 1%, 2%, …, 20% oxygen concentration of glass bottles under test, respectively) and Fig. 10 for the corresponding absolute errors (the absolute value of the difference between the real value and the measured value). In this round of simulation, we evaluate the performance difference between the methods of WMS and DAS. For straightforward comparison, we use the V_{p-p} of direct absorption spectroscopy signal and the 2nd harmonic signal as the decision criteria for inversing oxygen concentration. As shown in Fig. 9, the measuring results of the scheme of DAS+V_{p-p} exist extensive fluctuations to all the 20 groups of simulation samples. While the fluctuating range is narrowed by the scheme of WMS+V_{p-p} due to the noise suppression capability of its narrowband feature. More comparative results can be observed from the absolute error chart in Fig. 10. It can be clearly seen that the average relative error is nearly 1.14% when using the scheme of WMS+V_{p-p}, which is better than that of DAS+V_{p-p}, 1.38%. Nevertheless, the achieved scores of WMS+V_{p-p} are still far from the ground truth. The main reason is that the decision criteria of V_{p-p} is noise sensitive in time domain. Any noise occurring at the peak or valleys will lead to erroneous V_{p-p}, then a wrong judgement will be triggered with large probability.

2) DAS+DWPT+Vp-p vs. WMS+DWPT+Vp-p

In the second round of comparison, we further evaluate the performance difference between WMS and DAS improved by DWPT. The V_{p-p} of the simulated waveform continues to be used as the concentration discriminating criteria. As shown in Fig. 11, the previous fluctuations have been suppressed to a large extent for both schemes of DAS+DWPT+V_{p-p} and WMS +DWPT+V_{p-p}. As expected, our designed sub-scheme defeats its competitor again. Notably, for each simulation group, the measuring results of the scheme of WMS+DWPT+V_{p-p} have been optimized to be almost a horizontal line. And it can be clearly observed from Fig. 12 that the average absolute error is improved to within 0.45% when using the scheme of WMS+DWPT+V_{p-p}, which is competitive to 0.48% of DAS+DWPT+V_{p-p}. The promising results once again confirms our proposal that the DWPT can selectively extract the principal components from the noise-intensive signals with a considerable accuracy. Nevertheless, the achieved scores of WMS+DWPT+V_{p-p} still have space to improve. Especially,
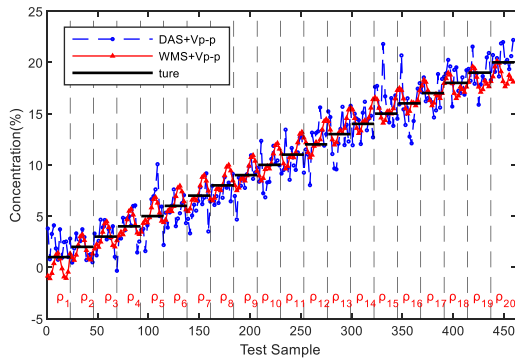
Fig. 9. Comparison results of oxygen concentration measurement by using the scheme of "DAS+V_{p-p}" and "WMS+V_{p-p}".
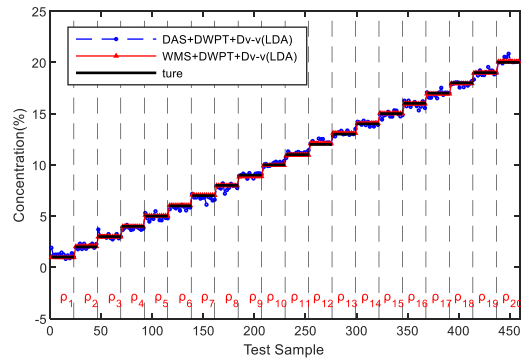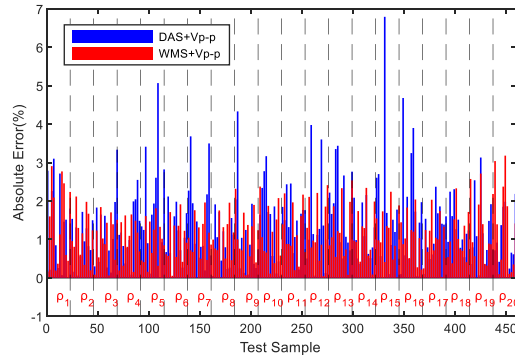

Fig. 10. Absolute errors of oxygen concentration measurement by using the scheme of "DAS+V_{p-p}" and "WMS+V_{p-p}".

Fig. 11. Comparison results of oxygen concentration measurement by using the scheme of "DAS+DWPT+V_{p-p}" and "WMS+DWPT+V_{p-p}".

Fig. 12. Absolute errors of oxygen concentration measurement by using the scheme of "DAS+DWPT+V_{p-p}" and "WMS+DWPT+V_{p-p}".

Fig. 13. Comparison results of oxygen concentration measurement by using the scheme of "DAS+DWPT+D_{v-v}(LDA)" and "WMS+DWPT+D_{v-v}(LDA)".
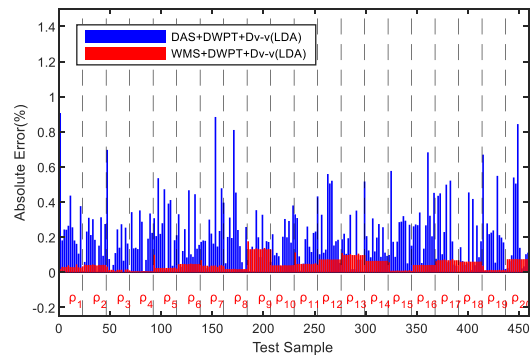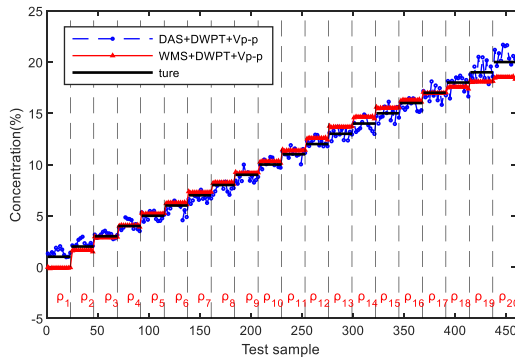

Fig. 14. Absolute errors of oxygen concentration measurement by using the scheme of "DAS+DWPT+D_{v-v}(LDA)" and "WMS+DWPT+D_{v-v}(LDA)".
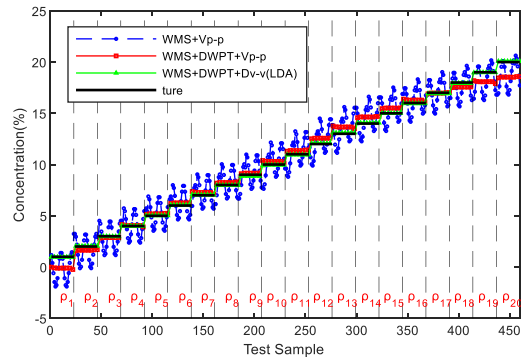

Fig. 15. Comparison results of oxygen concentration measurement by using the schemes of "WMS+V_{p-p}", "WMS+DWPT+V_{p-p}" and "WMS+DWPT+D_{v-v}(LDA)".
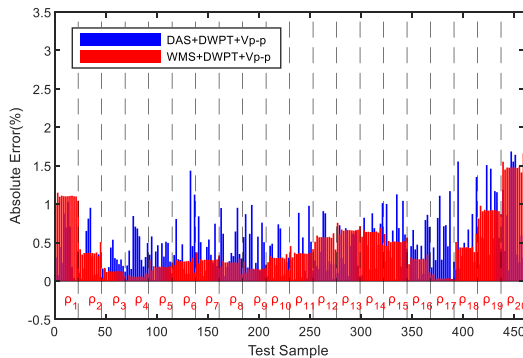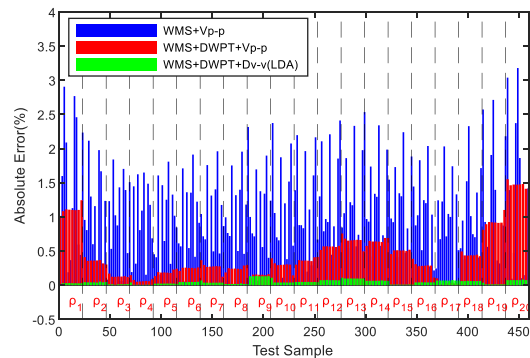

Fig. 16. Absolute errors of oxygen concentration measurement by using the scheme of schemes of "WMS+V_{p-p}", "WMS+DWPT+V_{p-p}" and "WMS+DWPT+D_{v-v}(LDA)".
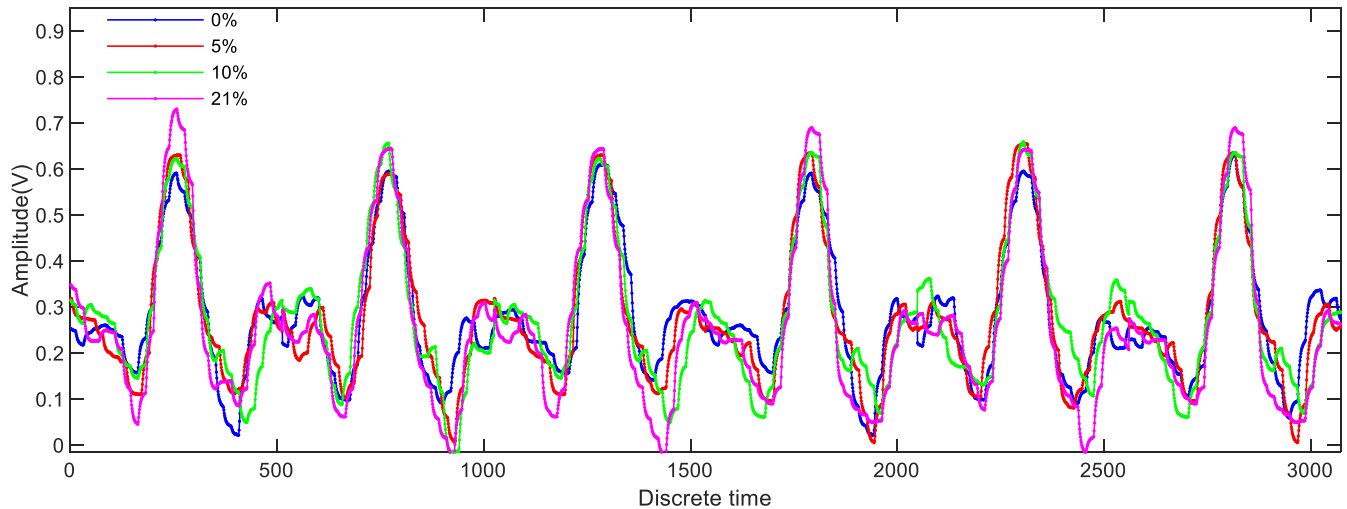
Fig.17. The raw 2nd harmonic signals of *four* glass bottle samples with oxygen concentration of 0%, 5%, 10% and 21% captured by our realized TDLAS/WMS prototype. Use color printing for better presentation.

some groups of test samples (e.g., ρ_1, ρ_{14}, ρ_{19}, ρ_{20}) cluster far from the ground truth, which can be more obviously observed from the relative error chart in Fig. 12. The fundamental reason is that the concentration inversion module only using criteria of V_{p-p} is essentially a simple one dimensional linear regression model. Any disturbance on the crest or trough will degrade the regression performance of the model. Therefore, despite the use of fine-grained wavelet filtering, residual noises still affect the accuracy of concentration identification.

3) DAS+DWPT+D_{v-v}(LDA) vs.WMS+DWPT+D_{v-v}(LDA)

In the third round of evaluation, we focus on the performance improvements brought by the multiple linear regression (MLR) based on data information between the left valley and right *v*alley (namely, D_{v-v}) in the 2nd harmonic signals, i.e., replacing V_{p-p} by D_{v-v}(LDA). As shown in Fig. 13, it is gratifying to note that the previous identification gaps in Fig. 11 (pay attention to check ρ_1, ρ_{14}, ρ_{19}, ρ_{20}) no longer exist to the scheme of WMS+DWPT+D_{v-v}(LDA). Even for the scheme of WMS+DWPT+D_{v-v}(LDA), the new round of results begin to fluctuate around the ground truth with a narrow average absolute error , 0.05%. Most notably in Fig. 14, the average absolute error of WMS+DWPT+D_{v-v}(LDA) has been restricted within 0.22%. These figures indicate that the MLR scheme has made intelligent use of the data information between two valleys of the absorbed 2nd harmonic signals. And the imported LDA has succeeded improving the information sparsity of the raw signal matrix of X in (17) to address the multi-collinearity problem.

4) Complementary contributions of DWPT and D_{v-v}(LDA)

For better comparison, Fig. 15 presents the measuring results of the TDLAS/WMS-series scheme distributed in Fig. 9 ~ Fig. 14. It can be clearly observed how our imported methods (i.e., DWPT, D_{v-v}(LDA)) improve the measurement accuracy step by step. The fluctuation range of measurement results is narrowed by the DWPT to a large extent. However, the one dimensional linear regression model only using the information in V_{p-p} still
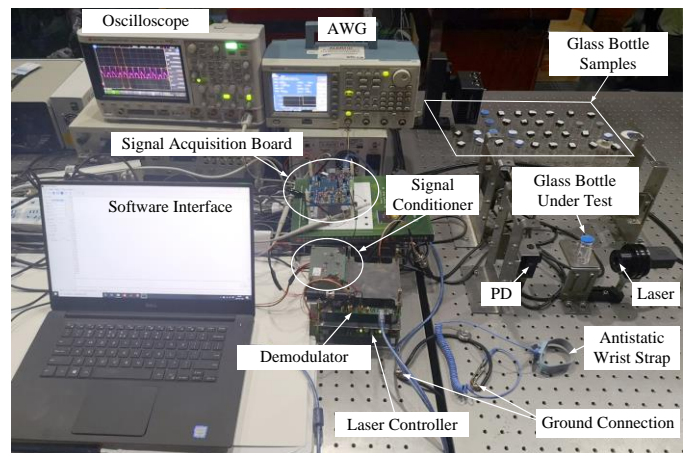

Fig.18. Test scenario of our realized TDLAS/WMS prototype.

suffers some detection errors for certain test groups because residual noises are inevitably pass through the DWPT-based signal reconstructor. Consequently, we launch another round of performance promotion by mining the implicit information around the V_{p-p}, that is, we adopt MLR to utilize the multiple information distributed between the two valleys of the 2nd harmonic waveform. Another side effect is that the adjacent data points have much linear correlation, the LDA is imported to address the multi-collinearity problem and to reduce the calculation complexity as well. This complementary concept is proved by the average absolute error of 0.05% achieved by our proposed scheme of WMS+DWPT+D_{v-v}(LDA) in Fig. 16.

V. EXPERIMENTAL VERIFICATION

In order to verify the accuracy and reliability of our proposed oxygen detection scheme of WMS+DWPT+D_{v-v}(LDA), this section implements a TDLAS/WMS prototype for real-world experimental tests.
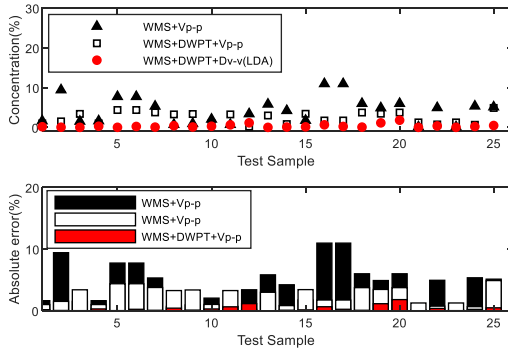
Fig. 19. Measured results and the absolute errors by using the schemes of "WMS+V_{p-p}", "WMS+DWPT+V_{p-p}" and "WMS+DWPT+D_{v-v}(LDA)", for standard glass bottles with 0% oxygen concentration.

Fig. 20. Measured results and the absolute errors by using the schemes of "WMS+V_{p-p}", "WMS+DWPT+V_{p-p}" and "WMS+DWPT+D_{v-v}(LDA)", for standard glass bottles with 5% oxygen concentration.

Fig. 21. Measured results and the absolute errors by using the schemes of "WMS+V_{p-p}", "WMS+DWPT+V_{p-p}" and "WMS+DWPT+D_{v-v}(LDA)", for standard glass bottles with 10% oxygen concentration.
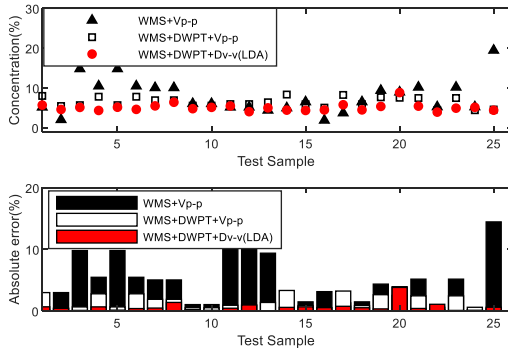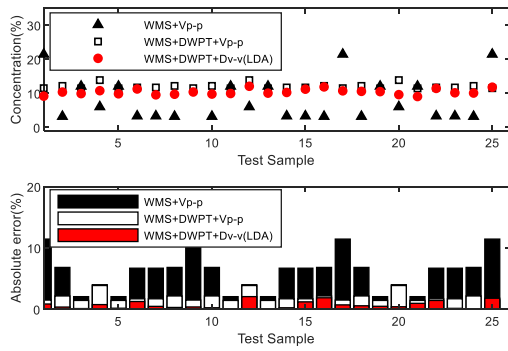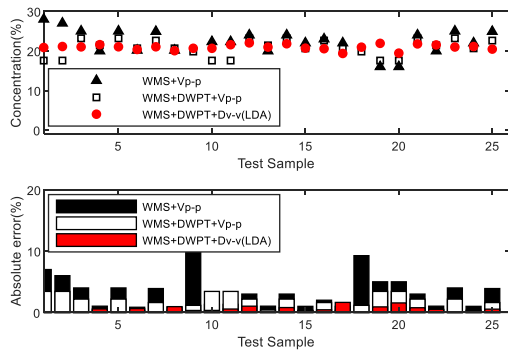

Fig. 22. Measured results and the absolute errors by using the schemes of "WMS+V_{p-p}", "WMS+DWPT+V_{p-p}" and "WMS+DWPT+D_{v-v}(LDA)", for standard glass bottles with 21% oxygen concentration.
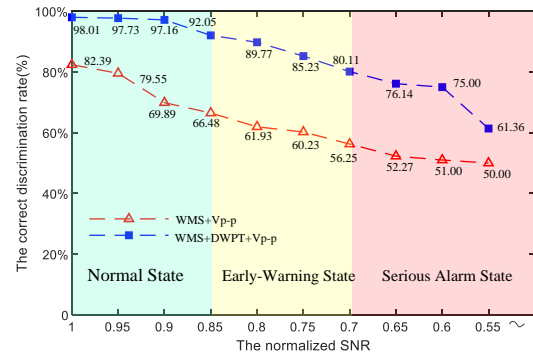
Fig. 23. The correct discrimination rates by using the schemes of "WMS+Vp-p" and "WMS+DWPT+Vp-p", for standard glass bottles with 5% oxygen concentration when the SNR decreases successively.
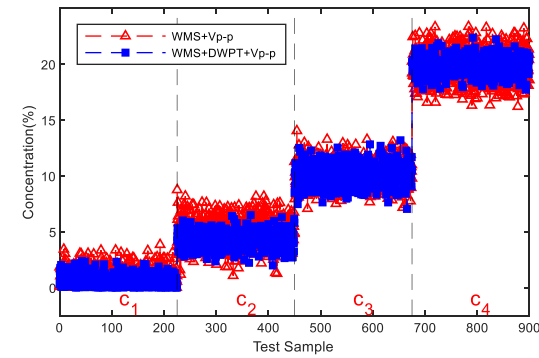
Fig. 24. Comparison results of oxygen concentration measurement by using the schemes of "WMS+Vp-p" and "WMS+DWPT+Vp-p", for standard glass bottles with different oxygen concentration when the normalized SNR is 0.7.
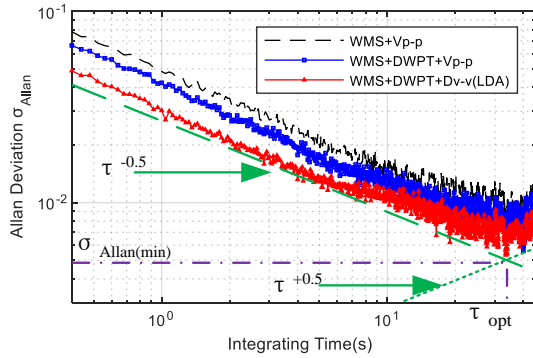
Fig. 25. Allan plots of different schemes for standard glass bot with 5% oxygen concentration.

TABLE I
ALLLAN ANALYSIS FOR STANDARD GLASS VIAL WITH
5% OXYGEN CONCENTRATION

Methods	σ Allan @1s	σ Allan @10s	σ Allan(min) @τopt(s)
WMS+Vp-p	0.0484	0.0152	0.0077@(36s)
WMS+DWPT+Vp-p	0.0423	0.0116	0.0056@(40s)
WMS+DWPT+ Dv-v(LDA)	0.0301	0.0105	0.0053@(33s)

A. Experiment setup

This section implements a TDLAS/WMS prototype for oxygen concentration measurement according to the scheme designed in Fig. 1. As show in Fig. 18, any module in Fig. 1 can be found in the test scenarios. This measurement test was performed on a Thorlabs optical platform in open-path optical environment, where the room temperature in the laboratory was around 25°C. In particular, a vertical-cavity surface-emitting laser diode (Laser component, single mode VCSEL, 760nm, TO5 footprint) with a free-running output power of 0.3mW was selected as the laser source. The laser temperature and exciting current were strictly controlled by a laser controller (Thorlabs, VITC002). The operation temperature of the laser diode is adaptively controlled at 21.20°C by PID fine-grained turning, and the laser injection current is a superimposing waveform, a 14.4 kHz balanced cosine wave with a peak-to-peak value of 5mV modulated on a 25Hz ramp-wave. The peak-to-peak value and the DC effective value of this ramp-wave were set to 10mV and 180mV, respectively, which controlled the current varying from 1.725mA to 1.875mA. Hence, the center frequency scanning around the spectral lines around 760.885nm for covering the scope of oxygen absorption spectral lines. The absorbed laser beams after travelling through glass medicine bottle can be received by the photodetector (Thorlabs, PDA36A, 350-1100 nm), this detected signals are transmitted to the FPGA-based demodulator for obtaining the 2nd harmonic signal, and the demodulated signal will then be sent to the signal acquisition board (MCU, C8051F064), and finally the collected data are uploaded to a host computer for oxygen concentration measurement by the WMS+DWPT+D_{v-v}(LDA) method.

As for the glass bottle samples, we chose the most commonly used penicillin vials, the detailed size is with 7mL volume, 22 mm bottom diameter and 22 mm bottle height. We prepared four groups of penicillin vials with four distinct oxygen concentrations of 0%, 5%, 10% and 21% by using the simple oxygen-nitrogen mixing method, and each group includes 25 independent bottle samples. These 100 penicillin vial samples were used as standard oxygen bottles for the experimental verification in this paper. More specifically, a standard bottle with an oxygen concentration of 0% means that it is filled fully with pure nitrogen, and a standard bottle with an oxygen concentration of 5% means that it is filled with 5% volume of pure oxygen and 95% volume of pure nitrogen. The oxygen and nitrogen are respectively taken from a steel oxygen tank and a steel nitrogen tank, both the gas purity values of the source tanks reach over 99.999%. Oxygen and nitrogen are lowered to the same pressure through a pressure relief valve before being fed into standard bottles. In order to ensure that the oxygen concentration in penicillin bottle meets its nominal value, we used a high-precision headspace oxygen analyzer (Lighthouse, FMS-760, maximum error (O_2%): 0.05%) to measure the actual oxygen concentration after that the standard oxygen bottles had been encapsulated. Under this strict configuration, the oxygen absolute concentration error of all the 100 standard bottles were controlled within 0.05%. Again, during the following experiments, all the parameter setups of the algorithms of DWPT, MLR and LDA are inherited from the simulation part in Sec. IV. One exception is the total number of the data points

between the two valleys of the actually acquired 2nd harmonic signal is about 230, not the 165 points in the simulation setup.

For quick glance to the quality of the acquired 2nd harmonic waveforms, Fig. 17 presents some signal segments of raw 2nd harmonic waveforms of four standard glass bottles with oxygen concentration of 0%, 5%, 10% and 21% captured by our realized TDLAS/WMS prototype. Comparing with those in Fig. 8, signals presented in Fig. 17 almost tend to be entanglement and overlapping at some point. And nearly all the challenges described in Sec. I can be found in Fig. 17, which indicates that the actually acquired 2nd harmonic signals in the open-path optical environment exist large intra-class variation and ambiguous inter-class distance. As mentioned in Sec. I, there are many factors contributing to this phenomenon, including the interference effect caused by different glass bottles, which will aggravate the contamination of the 2nd harmonic signal. Furthermore, we found in the experiment that the degree of interference is different when the laser enters exactly the same glass bottle from different incident point. The slight difference in diameter, the thickness and the smoothness of the glass's surface will cause the interference of varying degrees as the laser travels through the glass bottle with the same oxygen concentration. From these evidences, the instrument realization is much tougher than the measurement problem in the simulation stage in Sec. IV. It is worth emphasizing that the signal quality in Fig. 17 would be even worse without the hardware-implemented band-limiting and linear amplification offered by the signal conditioning board (refer to module V in Fig. 1).

B. Results and Evaluation

1) Measurement Accuracy

The measurement concentration results and corresponding absolute concentration errors of the four groups of standard bottles with oxygen concentration of 0%, 5%, 10% and 21% are presented in Fig. 19, Fig. 20, Fig. 21, and Fig. 22, respectively. For focused evaluation, the three WMS-based schemes discussed in Fig. 15 are verified here in this experiment. It can be clearly observed from these figures that our proposed scheme of WMS+DWPT+D_{v-v}(LDA) performs the best in all sets of tests on measurement stability and absolute error. To be specific, as shown in the upper subfigure in Fig. 19, the measured concentration results using the WMS+V_{p-p} scheme produce the most violent fluctuations, and some test bottle samples (e.g., the 3rd, 16th, and 17th sample) were even erroneously identified with concentration of more than 10%, that is far away from the 0% nominal concentration. And the DWPT-based signal reconstruction narrowed the negative fluctuations to a large extent. In addition, the last round of D_{v-v}(LDA)-driven enhancement pulled the absolute errors closer to zero. These promising results can be consistently found in the following experiments for the standard bottle samples with oxygen concentration of 5%, 10% and 21%. Notably, the average absolute errors for detected glass bottles with different oxygen concentration by using the scheme of WMS+DWPT+D_{v-v}(LDA) have been restricted as low as 0.54%. These encouraging experimental results prove that the proposed methodology is promising to be applied on the on-site aseptic guarantee and quality monitoring of glass bottle

encapsulated pharmaceuticals through accurate oxygen concentration measurement.

2) Noise-robustness

Here, we carried out a set of tests aiming at the sub-scheme of WMS+DWPT+V_{p-p} to prove the noise-robustness of our proposed methodology. The standard glass bottle with 0% and 5% oxygen concentration are measured to acquire 225 2nd harmonic signal, respectively. Then, the Gaussian white noises with different signal to noise ratio (SNR) are superimposed on the raw 2nd harmonic signal. As shown in Fig.23, according to different SNR, we define *three* states for headspace oxygen concentration measurement system: normal state ($0.85<SNR\leq1$), early-warning state ($0.7\leq SNR\leq0.85$), and serious alarm state ($SNR<0.7$). The V_{p-p} of all the different SNR 2nd harmonic signal of glass bottles with 0% and 5% oxygen concentration were extracted. Then, the average of the maximum of 225 V_{p-p} of 2nd harmonic signal of glass bottle with 0% oxygen concentration and the minimum of 225 V_{p-p} of 2nd harmonic signal of glass bottle with 5% oxygen concentration is used as the threshold, which is the criterion of correct or fault discrimination. The concentration discrimination is correct when the V_{p-p} of 2nd harmonic signal of glass bottle with 5% oxygen concentration is higher than the threshold. Inversely, the concentration discrimination is fault when the V_{p-p} of 2nd harmonic signal of glass bottle with 5% oxygen concentration is below the threshold. In this way, the correct discrimination rate of 5% oxygen concentration under different SNR can be obtained. The blue dashed line in the Fig.23 is the correct discrimination rate curve with DWPT, while the red dashed line is the correct discrimination rare curve without DWPT. Under the first state, the correct discrimination rate of the scheme of WMS+DWPT+V_{p-p} is higher than 90%, while that of the scheme of WMS+V_{p-p} is between 82.49% and 66.48%, which shows that the noises seriously contaminate these 2nd signal and the DWPT performs the noise-robust property. Under the second state, the correct discrimination rate of both schemes degenerate gradually with the decrease of signal quality, however, the scheme of WMS+DWPT+V_{p-p} can still suppress in this range to some extent(SNR=0.7: 80.11% *vs*. 56.25%). As you can see in Fig. 24 (c1, c2, c3, c4 stands for 0%, 5%, 10%, 21% oxygen concentration of glass bottles under test, respectively), the error range of concentration discriminated by V_{p-p} fluctuates greatly when the SNR is 0.7, this situation has closed to limitation of serious alarm. The last state is not allowed to the headspace oxygen concentration measurement system, which will trigger emergency alarm. In fact, at the second state, the system will continuously send out early-warning signals, reminding operators to investigate the potential failures of related equipment. The above results prove that the DWPT-based method can survive well when the normalized SNR is within 0.85 to 1.

3) System Stability

In order to evaluate the stability when adapting our proposed measurement scheme of WMS+DWPT+D_{v-v}(LDA), one of the glass bottles with 5% oxygen concentration is chosen as the vial sample and is measured for 15 min to acquire 22500 2nd harmonic signal. Three different schemes (as mentioned in Sec.V.B.1) are continued to be used to process these signals. The Allan variance analysis results corresponding to the concentration data are shown in Fig. 25. Allan variance is a method of measuring the frequency stability of the measurement system in the time domain. This method can be used to determine the intrinsic noise in a system as a function of the integrating time. It is one of most popular methods for identifying and quantifying the noise terms that exist in the inertial sensor data. As shown in Fig. 25, the white noise is dominant when the slope is -0.5 (dashed line) and the Brownian noise is dominant when the slope is 0.5 (dotted line). When σ_{Allan} is minimum ($\sigma_{Allan(min)}$), the optimum integrating time (τ_{opt}) and the detection limit is determined. It can be seen that in the white and Brownian noise range, the precision by using the method of WMS+DWPT+D_{v-v}(LDA) is higher than that of the other methods. For example, at 1s, σ_{Allan} by using the method of WMS+V_{p-p},WMS+DWPT+V_{p-p} and WMS+DWPT+D_{v-v}(LDA), it can obtain at 0.0484, 0.0423 and 0.0301, respectively, suggesting that DWPT and D_{v-v}(LDA) can effectively suppress the random noise and improve the detection precision again. For the concrete σ_{Allan} for 5% oxygen concentration glass bottle with three methods as listed in Table I, which conclusively shows that the method of WMS+DWPT+D_{v-v}(LDA) can obtain the lowest $\sigma_{Allan(min)}$ among three schemes. These results verified that the stochastic noises superimposed on the 2nd harmonic signal, to some extent, can be suppressed by using DWPT+D_{v-v}(LDA). Therefore, these Allan variance results verify the stability and the precision of our designed prototype, which provides the feasibility for the implementation in the industrial field in the future.

VI. CONCLUSION

This work concentrates on establishing TDLAS/WMS-based headspace oxygen concentration measurement method for pharmaceutical glass bottles, with the hope to offer solutions for achieving on-site inspection of sealing integrity under open-path optical environment. A noise robust and lightweight signal processing framework based on DWPT, MLR and LDA, namely WMS+DWPT+D_{v-v}(LDA), is constructed for accurate and reliable oxygen concentration measurement. Specifically, we equip two rounds of anti-noise measures: First, the DWPT signal reconstructor selectively extracts several sets of wavelet components that have strong cross-correlation to the raw 2nd harmonic signal, hence the stochastic and intensive optical and electronic noise are suppressed effectively. Second, the traditional one dimensional linear regression model only using the information in V_{p-p} is replaced by our D_{v-v}(LDA) scheme (e.t., LDA-enhanced MLR) to eliminate the influence of residual noises at peak or valley point on the measurement accuracy by mining the potential information distributed around the peak value. Extensive simulation tests are carried out to evaluate the performance of our proposed methodology and a modularized TDLAS/WMS hardware architecture considering flexibility and scalability is designed and then realized as an oxygen concentration measurement prototype, where the former experiment results are verified carefully to evaluate the accuracy and stability achieved by our proposed methodology.

Work done to date has been laboratory-based but designed with on-site application in mind. It can be expected that more challenges involving theoretical problems and engineering developments will run through the TDLAS/WMS application process of on-line oxygen concentration in real-world industrial environment. Work will need to be done to suppress the detection susceptibility resulted from the disturbances of pressure and temperature fluctuations in industrial environment. Additionally, the modular circuits can be integrated into an industrial instrument to be transplanted to a real bottle packaging production line. This represents an important start in transferring the TDLAS/WMS technology from the laboratory to practical industrial applications to solve some problems of accurate measurement of weak signals, which will provide a good foundation for the online quality supervision of pharmaceutical production.

APPENDIX A

Wavelength modulation spectroscopy (WMS) is a method for sensitive absorbance measurements which uses a significantly faster sinusoidal signal riding on a slowly varying diode laser injection current. Supposing the ramp-wave driving current to laser diode by using direct absorption spectroscopy (DAS) method is $i_0^d(t)$, the wavelength modulation spectroscopy (WMS) method converts the direct-absorption-signal based detection problem to signal demodulation in much narrower band by modulating the ramp-wave with a cosine wave $\cos(wt)$ under an intensity modulation index α. Thus the synthetic driving current of laser is

$$i_0(t) = i_0^d(t)\left[1 + \alpha\cos(2\pi f t)\right] \tag{A.1}$$

where w stands for the angular frequency corresponding to its modulation frequency f, and $w = 2\pi f$, then the exciting laser intensity $I_0(t)$ driven by $i_0(t)$ and the scanning optical frequency $v(t)$ can be respectively denoted as

$$I_0(t) = I_0^d(t)\left[1 + \alpha\cos(\omega t)\right] \tag{A.2}$$

$$v(t) = v_c(t) + v_m\cos(\omega t) \tag{A.3}$$

where, $I_0^d(t)$ represents the laser intensity corresponding to $i_0^d(t)$, $v_c(t)$ is the central frequency of laser diode, v_m is the amplitude of frequency modulation.

When a light beam with a laser intensity of $I_0(t)$ travels through an absorbing gas (oxygen in this paper) with a spectral feature line strength $s(T)$, where T stands for the gas ambient temperature, according to Beer–Lambert law, the output light intensity $I(t)$ and the input $I_0(t)$ can be linked as

$$I(t) = I_0(t)e^{-s(T)\rho LPg(v)} = I_0^d(t)\left[1 + \alpha\cos(wt)\right]e^{-s(T)\rho LPg[v_c(t)+v_m\cos(\omega t)]} \tag{A.4}$$

where, ρ is the concentration of absorbing gas under test, P is the total pressure of the absorbing gas in gas cell (glass vial in this paper), L is the overall absorption path length, and $g(v)$ represents the shape function of absorbing line.

Considering that gas absorption and intensity modulation are always very small ($-S(T)\rho LPg(v)\Box\ 1$, and $\alpha\Box\ 1$), and after expanding (A.4) using Taylor series, $I(t)$ may be approximated as

$$I(t) \approx I_0^d(t)\left\{1 + \alpha\cos(wt) - s(T)\rho LPg\left[v_c(t)+v_m\cos(wt)\right]\right\} \tag{A.5}$$

After the linear transformation of the photodetector on the detection side, the receiving current $i(t)$ corresponding to $I(t)$ is

$$i(t) \approx i_0^p(t)\left\{1 + \alpha\cos(wt) - s(T)\rho LPg\left[v_c(t)+v_m\cos(wt)\right]\right\} \tag{A.6}$$

Similarly, $i_0^d(t)$ stands for the detection current corresponding to $I_0^d(t)$. Under atmospheric pressure and temperature, if we choose the Lorentzian distribution (where Δv_c denotes the full width at half maximum of Lorentzian line shape, and v_0 represents the central frequency of gas spectral absorbing.)

$$g(v) = \frac{1}{2\pi}\frac{\Delta v_c}{(v-v_0)^2 + (\Delta v_c/2)^2} \tag{A.7}$$

to fit the absorbing shape line $g(v)$, $i(t)$ in (A.6) is then expanded by using Fourier series as

$$i(t) \approx i_0^p(t)\left\{1 + \alpha\cos(wt) - s(T)\rho LP\frac{2}{\pi\Delta v_c}S_n(x,m)\cos(nwt)\right\} \tag{A.8}$$

where, $S_n(x,m)$ represents the amplitude of the nth Fourier components of absorption signal, and its general formula and binomial expansion can be referred in [13] and [35], respectively. And $m = v_m/(\Delta v_c/2)$ stands for the frequency modulation degree, x holds the expression of $(v_c-v_0)/(\Delta v_c/2)$.

Any nth harmonic signal of (A.8) can be calculated by using digital demodulation with its modulation frequency and then obtained through a low-pass filter, and all these steps are very suitable embedded in SOC system (i.e., FPGA). Absorption coefficient of odd order harmonic component is zero at the central position, while even order is the maximum. Since the amplitude of harmonic signal decreases with increasing order, the 2nd harmonic component is selected as the detection signal in practice. Take the 2nd harmonic signal as an example, then demodulation can be expressed as

$$i_{2f}(t) = X_{LP}\left[i(t)\cos(2wt)\right] \tag{A.9}$$

where X_{LP} represents the operator of a digital low-path filter. After a series of trigonometric function's operations, (A.9) can be further deduced as

$$i_{2f}(t) = i_0^p(t)s(T)\rho LP\frac{2}{\pi\Delta v_c}S_2(x,m) \tag{A.10}$$

According to [35], $S_2(x,m)$ is only related to the central frequency of laser diode $v_c(t)$, and $i_{2f}(t)$ can be further expanded by using the binomial formula. In particular, when $v_c = v_0$, in other words, $v_c(t)$ is the central frequency stimulated by laser diode is overlapping with the central frequency of gas spectral absorbing v_0, we denote the $i_0^p(t)$ as $i_0^p|_{vc=v0}$, and $i_{2f}(t)$ gets the maximum as

$$i_{2f\max} = i_0^p\big|_{v_c=v_0}s(T)\rho LP\frac{2}{\pi\Delta v_c}\frac{2}{m^2}\left[\frac{2-(2+m^2)}{\sqrt{(1+m^2)}}\right] \tag{A.11}$$

then the concentration of absorbing gas N can be deduced from (A.11) to

$$\rho = \frac{i_{2f\max}\pi\Delta v_c m^2}{4i_0^p\big|_{v_c=v_0}s(T)LP}\left[\frac{2-(2+m^2)}{\sqrt{(1+m^2)}}\right]^{-1} \tag{A.12}$$

It is clear that all the parameters Δv_c, m, $i_0^p|_{vc=v0}$, $s(T)$, L and P are constants under certain experimental condition, thus the concentration (ρ) of absorbing gas obeys a linear relationship to the maximum amplitude (i_{2fmax}) of the 2nd-harmonic absorption signal at center of the absorption line.

REFERENCES

[1] L. Norbert, B. Thomas, and L. Matthias, "Cross -interferences of volatile anesthetic agents on the measurement of O2 utilizing an optical fiber sensor based on luminescence quenching," *Anesthesiologists*, vol. 93, no. 3A, pp. A-565, Sep. 2000.

[2] Y. Amao, "Probes and polymers for optical sensing of oxygen," *Microchim. Acta.*, vol. 143, no. 1, pp. 1–12, Aug. 2003.

[3] G. A. Eiceman, B. Tadjikov, E. Krylov, E. G. Nazarov, R. A. Miller, J. Westbrook, and P. Funk, "Miniature radio-frequency mobility analyzer as a gas chromatographic detector for oxygen-containing volatile organic compounds, pheromones and other insect attractants," *J. Chromatogr. A*, vol. 917, no. 1-2, pp. 205–217, May 2001.

[4] Z. Du, R. Qi, H. Zhang, X. Yin, and K. Xu, "Quantitative detection of propane and isobutene based on NIR spectroscopy," *J. Tianjin Univ.*, vol. 41, no. 5, pp. 589–592, May 2008.

[5] X. Liu, S. Cheng, H. Liu, S. Hu, D. Zhang, and H. Ning, "A survey on gas sensing technology," *Sensors*, vol. 12, no. 7, pp. 9635–9665, Jul. 2012.

[6] J. J. R. Rohwedder, C. Pasquini, P. R. Fortes, I. M. Raimundo, Jr., A. Wilk, and B. Mizaikoff, "iHWG-μNIR: A miniaturised nearinfrared gas sensor based on substrate-integrated hollow waveguides coupled to a micro-NIR-spectrophotometer," *Analyst*, vol. 139, no. 14, pp. 3572–3576, 2014.

[7] J. J. Harrison and P. F. Bernath, "Infrared absorption cross sections for propane (C3H8) in the 3 μm region," *J. Quant. Spectrosc. Radiat. Transf.*, vol. 111, no. 9, pp. 1282–1288, Jun. 2010.

[8] E. D. Hinkley, "High-resolution infrared spectroscopy with a tunable diode laser," *Appl. Phys. Lett.*, vol. 16, no. 9, pp. 351–354, Oct. 2003.

[9] Q. Gao *et al*., "Tunable multi-mode diode laser absorption spectroscopy for methane detection," *Sensor. Actuat. A-phys.*, vol. 199, pp. 106–110, Sep. 2013.

[10] A. Seidel *et al*., "Robust spatially scanning, open-path TDLAS hygrometer using retro-reflective foils for fast tomographic 2-D water vapor concentration field measurements," *Atmos. Meas. Tech.*, vol. 8, no. 5, pp. 2061–2068, May. 2015.

[11] T. Le Barbu *et al*., "TDLAS a laser diode sensor for the in situ monitoring of H2O, CO2 and their isotopes in the Martian atmosphere," *Adv. Space Res.*, vol. 38, no. 4, pp. 718–725, 2006.

[12] Q. Tang, Y. Zhang *et al*. "Laser absorption spectroscopy data processing method based on co-frequency and dual-wavelength and its application," Opt. Express., vol. 26, no.4,pp. 4448-4458,Feb.2018.

[13] R. Arndt, "Analytical line shapes for Lorentzian signals broadened by modulation," *J. Appl. Phys.*, vol. 36, no.8, pp. 2522-2524, Mar. 1965.

[14] J. Reid and D. Labrie, "Second-harmonic detection with tunable diode lasers-Comparison of experiment and theory," *Appl. Phys. B*, vol. 26, pp. 203-210, Nov. 1981.

[15] J. Jiang, M. Zhao, G. Ma, H. Song, C. Li, X. Han, and C. Zhang, "TDLAS-Based Detection of Dissolved Methane in Power Transformer Oil and Field Application," *IEEE Sens. J.*, vol. 18, no. 16, pp. 2318-2325, Mar. 2018.

[16] J. Jiang, G. Ma, H. Song, H. Zhou, C. Li, H. Wang, Y. Luo, and H. Wu, "Tracing methane dissolved in transformer oil by tunable diode laser absorption spectrum," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 23, no. 6, pp. 3435-3442, Dec. 2016.

[17] G. S. Jatana, A. K. Perfetto, S. C. Geckler, and W. P. Partridge, "Absorption spectroscopy based high-speed oxygen concentration measurements at elevated gas temperatures," *Sensor. Actuat. B-chem*., vol. 29, no. 3, pp. 173-182, May. 2019.

[18] T. Svensson, M. Andersson, L. Rippe, S. Svanberg, S. Andersson-Engels, J. Johansson, and S. Folestad, "VCSEL-based oxygen spectroscopy for structural analysis of pharmaceutical solids," *Appl. Phys. B*, vol. 90, pp. 345-354, Feb. 2008.

[19] X. Zhou, J. Yu, L. Wang, Q. Gao, and Z. Zhang, "Sensitive detection of oxygen using a diffused integrating cavity as a gas absorption cell," *Sensor. Actuat. B-chem*., vol. 241, pp. 1076-1081, Mar. 2017.

[20] G. Zhu *et al*., "The multi-beam interference suppression for measuring Penicillin Vial's Oxygen concentration based on tunable diode laser absorption," *Spectrosc. Spect. Anal*., vol. 38, no. 2, pp. 372-376, Feb. 2018.

[21] P. W. Werle, B. Scheumann, and J. Schandl, "Real-time signal-processing concepts for trace-gas analysis by diode-laser spectroscopy," *Opt. Eng.*, vol. 33, pp. 3093-3105, Sep.1994.

[22] D. P. Leleux, R. Claps, W. Chen, F. K. Tittel, and T. L. Harman, "Applications of Kalman filtering to real-time trace gas concentration

measurements," *Appl. Phys. B-lasers O*., vol. 74, no. 1, pp. 85-93, Jan. 2002.

[23] W. Liang, Q. Zhou, X. Dong, and T. Lv, "Influence of temperature-induced cavity length variation in wavelength modulation spectroscopy," *Optik*, vol. 172, pp. 220-224, 2018.

[24] E. Hecht and A. Zajac, *Optics (5th Edition)*, New Jersey: Addison-Wesley, 2017.

[25] A. Hartmann, R. Strzoda, R. Schrobenhauser, and R. Weigel, "Ultra-compact TDLAS humidity measurement cell with advanced signal processing," *Appl. Phys. B*, vol. 115, pp. 263-268, 2014.

[26] H. Ahlberg, S. Lundqvist, M. S. Shumate, and U. Persson, "Analysis of errors caused by optical interference effects in wavelength-diverse CO2 laser long-path systems," *Appl. Opt.*, vol. 24, no. 22, pp. 3917-3923, 1985.

[27] D.T.Cassidy and J. Reid, "Harmonic detection with tunable diode lasers two-tone modulation," *Appl. Phys. B*, vol. 29, pp. 279-285, 1982.

[28] L. Persson, F. Andersson, M. Andersson, and S. Svanberg, "Approach to optical interference fringes reduction in diode laser absorption spectroscopy," *Appl. Phys. B*, vol. 87, no. 3, pp. 523-530, May. 2007.

[29] S. Mallat, *A Wavelet Tour of Signal Processing*, New York: Academic Press 2008.

[30] Y. Wang, Y. Wei, J. Chang, T. Zhang, T. Liu, T. Sun, and K. T. Grattan, "Tunable diode laser absorption spectroscopy- based detection of propane for explosion early warning by using a vertical cavity surface enhanced laser source and principle component analysis approach," *IEEE Sensor. J.*, vol. 17, no. 15, pp. 4975-4982, Aug.2017.

[31] S. Ji and J. Ye, "Generalized linear discriminant analysis: A unified framework and efficient model selection," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 19, no. 10, pp. 1768-1782, Oct.2008.

[32] I. Daubechies, "The wavelet transform, time-frequency localization, and signal analysis," *IEEE Trans. Inf. Theory*, vol. 36, no. 5, pp. 961–1005, Sep. 1990.

[33] R. J. Romero-Troncoso, M. Pena-Anaya, E. Cabal-Yepez, A. Garcia-Perez, and R. A. Osornio-Rios, "Reconfigurable SoC-based smart sensor for wavelet and wavelet packet analysis," *IEEE Trans. Instrum. Meas.*, vol. 61, no. 9, pp. 2458-2468, Sep. 2012.

[34] D. B. Percival and A. T.Walden, *Wavelet Methods for Time Series Analysis(1st Edition)*. Cambridge: Cambridge Univ. Press, 2000.

[35] O. Axner, P. Kluczynski and A.M. Lindberg, "A general non-complex analytical expression for the n*th* Fourier component of a wavelength-modulated Lorentzian lineshape function," *J. Quant. Spectrosc. Ra.*, vol. 68, no. 3, pp. 299-317, Feb. 2001.