

Simultaneous Multifrequency Demodulation for Single-Shot Multiple-Path ToF Imaging

Peyman Fayyaz Shahandashti^{ID}, *Graduate Student Member, IEEE*, Paula López^{ID}, *Senior Member, IEEE*, Víctor Manuel Brea^{ID}, Daniel García-Lesta^{ID}, and Miguel Heredia Conde^{ID}, *Member, IEEE*

Abstract—Indirect Time-of-Flight (iTof) sensors measure the received signal’s phase shift or time delay to calculate depth. In realistic conditions, however, recovering depth is challenging as reflections from secondary scattering areas or translucent objects may interfere with the direct reflection, resulting in inaccurate 3D estimates. We propose a new measurement concept including a single-shot on-chip multifrequency demodulation method with periodically-repeated ultrashort-pulsed illumination using a novel pixel array architecture to address a main limitation of conventional iTof, the Multi-Path Interference (MPI). Due to the careful hardware/software codesign, the proposed single-shot architecture provides close-to-optimal Fourier measurements to a spectral estimation algorithm that retrieves the unknown parameters of the interfering return paths in a closed form. Electrical simulations of the on-chip multifrequency demodulation circuit demonstrate the feasibility of distance retrieval in double and triple bounce conditions in a single shot with high accuracy. Furthermore, we propose a set of methods for processing the resulting sensor measurements that exploit valuable a priori information and structural constraints of the data and observe that they yield a substantial increase in accuracy.

Index Terms—CMOS image sensor, depth image sensor, macro-pixel, multifrequency demodulation, multipath interference, single-shot, spectral estimation, time-of-flight.

I. INTRODUCTION

IN CONTRAST to typical digital cameras, which produce 2D images, ToF sensors capture 3D images, where the third dimension is the depth or distance [1], [2]. ToF sensors rely on a simple idea, albeit complex to implement: time and distance are

Manuscript received 26 January 2023; revised 5 August 2023 and 31 October 2023; accepted 14 December 2023. Date of publication 1 January 2024; date of current version 12 January 2024. This work was supported by the European Union’s H2020 Research and Innovation Programme under the Marie Skłodowska-Curie under Grant 860370, in part by the Spanish Ministry of Science, Innovation and Universities under Grant PID2021-128009OB-C32, in part by Xunta de Galicia-ConSELLERÍA de Cultura, Educación e Ordenación Universitaria Accreditation under Grant 2019-2022 ED431G-2019/04, and in part by Reference Competitive Group Accreditation under Grant 2021-2024 ED431C2021/048, co-funded through (ERDF/FEDER Programme). The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Jinli Suo. (*Corresponding author: Peyman Fayyaz Shahandashti*.)

Peyman Fayyaz Shahandashti, Paula López, Víctor Manuel Brea, and Daniel García-Lesta are with the Centro Singular de Investigación en Tecnologías Inteligentes (CITIUS), Universidad de Santiago de Compostela, 15705 Santiago de Compostela, Spain (e-mail: peyman.fayyaz@usc.es; p.lopez@usc.es; victor.brea@usc.es; daniel.garcia.lesta@usc.es).

Miguel Heredia Conde is with the Center for Sensor Systems (ZESS), University of Siegen, 57076 Siegen, Germany (e-mail: heredia@zess.uni-siegen.de).

Digital Object Identifier 10.1109/TCI.2023.3348758

linearly related [3]. While optical ToF sensors are a relatively recent phenomenon, other ToF systems, such as radar, sonar, and seismic technology, have been used for many years.

The ToF approach is based on the time difference between the emitted and reflected signals. Since the time delay is related to the distance, measuring the ToF is equivalent to measuring the object’s range. Even though ToF systems can be based on pulse waves, commercial products are mostly based on continuous wave demodulation [4], in which the light source intensity is modulated at radio frequencies (tens to hundreds of MHz). The sensor measurements allow for estimating the phase shift between the emitted and reflected signals. The modulation frequency converts phase into a time delay and the speed of light to convert time delay into distance [5], [6]. This method, known as amplitude-modulated continuous wave (AMCW) ToF, attains a sufficient signal-to-noise ratio (SNR) by integrating over many modulation cycles.

In recent years, substantial efforts have been made toward the coherent codesign of the image sensor and signal/image processing in the growing field of computational imaging, [7], [8], [9], [10]. However, to a great extent, it has been primarily limited to 2D scenes, despite the fact that capturing 3D information of a scene provides novel capabilities that can be used in many applications, such as autonomous driving, mobile robotics, industry, and augmented and virtual reality. Therefore, combining the circuit design and signal processing perspectives is essential to overcome challenges in ToF imaging that neither sophisticated circuits nor algorithms alone can tackle, such as the MPI problem.

MPI occurs due to inter-reflections in the scene, causing objects to be illuminated by both direct and indirect illumination, as shown in Fig. 1. Because of this superposition of two (or more) phase-shifted signals, the camera produces an incorrect measurement, which can result in considerable depth and amplitude errors. Although many topologies have been proposed in pixel design for overcoming today’s challenges in iTof, like background light [11], [12], motion artifacts [13], and limited speed [14], no architecture has been designed for acquiring optimal measurements to solve the MPI problem.

On the other hand, from the signal processing point of view, an active field of research deals with MPI [15], [16], [17]. Reliable depth estimation under MPI conditions has been addressed comprehensively in the iTof literature; nonetheless, it is still one of the main challenges in this technique [18], [19], [20], [21]. Resolving MPI generally requires formulating an

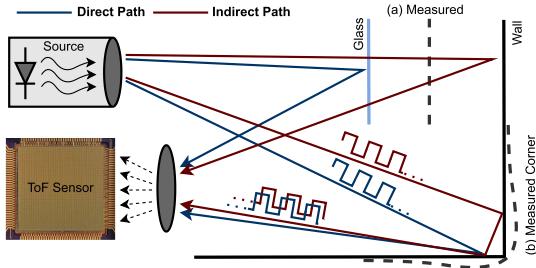


Fig. 1. Examples of Multi-Path Interference (MPI) scenarios where the sensor receives light from various reflections. The estimated depth from the conventional approach in these cases is erroneous.

appropriate mathematical model of the phenomenon and a separation method to disentangle the interfering bounces. Many solutions have been proposed to mitigate MPI that do not need hardware modifications but mostly at the cost of many frequencies or dense temporal sampling [22], [23], [24], [25].

This work presents a joint hardware/signal processing code-sign approach to tackle the MPI problem in iToF imaging for the first time. The proposed system can reduce computational costs and memory requirements, increasing speed and enabling real-time MPI solutions, which is crucial for various applications, including autonomous driving and robotics. From the signal processing perspective, the key novelty is the introduction of a unified pipeline that takes the single-shot raw data from the sensor as input, where a complete image or frame is captured in a single exposure or acquisition and delivers an independent depth image per interfering path. We use the measurements to estimate the lowest-frequency Fourier coefficients of the scene response function, which can be ideally modeled as a train of Dirac delta functions in the reflective MPI case. Crucially, for the reflective MPI model, it follows that the optimal measurement space is the Fourier space [26], [27]. Besides, we propose data refinement methods to improve depth estimation accuracy.

On the other hand, from the hardware perspective, the key novelty is the design of an iToF sensor architecture specifically designed to deliver the best possible measurements for inverting the reflective MPI problem. We propose a single-shot multifrequency approach to deal with MPI under different conditions. Conventional multifrequency methods rely on acquiring consecutive scene samples at different frequencies with the consequent frame rate reduction. The single-shot multifrequency ToF sensor proposed here can deliver Fourier measurements with minimal harmonic distortion. This is achieved by clustering groups of neighboring ToF pixels, each modulated with a different frequency, forming macro-pixels [28]. Increasing the macro-pixel size allows more raw data at different frequencies to be acquired in a single frame. On the other hand, owing to recent advancements in process technologies, a reasonable increase in the resolution of ToF imagers has been achieved (e.g., 1 Mpix sensor with $3.5\ \mu\text{m}$ pixel pitch in [29]). Therefore, even though implementing macro-pixels results in lower spatial resolution, state-of-the-art technologies can obtain a reasonable performance. The conceptualized idea resembles the utilization of RGB color filters in 2D imaging. Performance evaluation

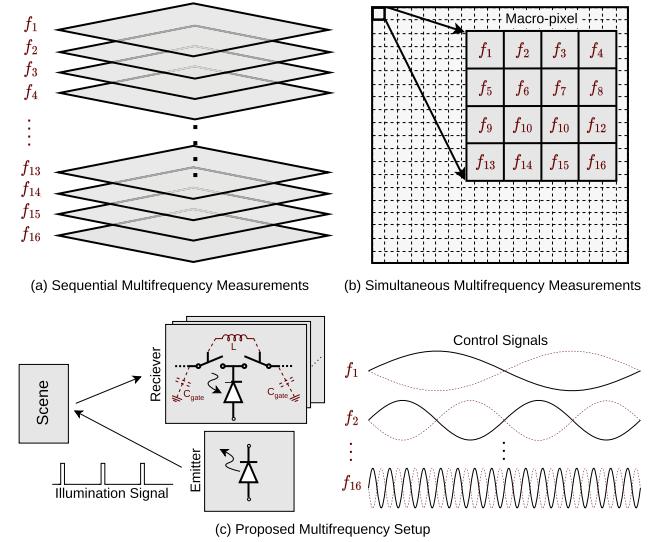


Fig. 2. Conventional multifrequency measurement (a) and simultaneous multifrequency demodulation concept (b). An array of macro-pixels, including 16 subpixels (magnified area), is shown. The proposed iToF imaging setup with multifrequency demodulation and impulse illumination (c).

is conducted using standard electrical simulation tools. The proposed codesign is compared to the state-of-the-art methods, including related methods leveraging sparsity (such as [18]) and deep learning approaches.

The work is organized as follows: in Section II, the sensor architecture is presented, while the depth retrieval algorithm and data refinement methods are explained in Sections III and IV, respectively. In Section V, experimental results and comparison with prior works are provided. The main conclusions of the work are summarized in Section VI.

II. IN-PIXEL SIMULTANEOUS MULTIFREQUENCY DEMODULATION

The first part of this section is dedicated to the array architecture (macro-pixel grouping), which enables the desired multifrequency demodulation, and the second part deals with the pixel architecture and the readout stages.

A. Simultaneous Multifrequency Demodulation

Two significant challenges hinder the possibility of real-time multiple-path ToF imaging. First, regarding the hardware implementations, the raw images taken at different frequencies need to be acquired sequentially, like in Fig. 2(a), whereas the number of required frequencies will increase linearly with the number of interfering paths. Second, most sparse retrieval algorithms are too slow in signal processing to work at an acceptable frame rate without resorting to neural network inference. To overcome these issues, we propose an iToF sensor architecture specifically designed to deliver undistorted Fourier measurements in a single shot that can be used to solve the inverse problem due to MPI in a closed form.

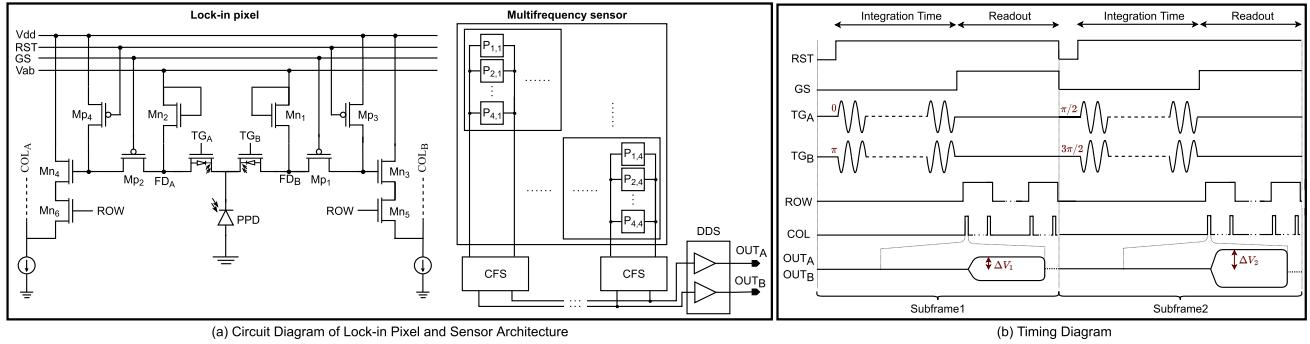


Fig. 3. Circuit diagram of the 2-tap pixel, the system architecture of the proposed iToF sensor (a), and the timing diagram of the system for image acquisition (b).

Fig. 2(b) shows the single-shot multifrequency acquisition architecture proposed in this paper, which can be achieved by grouping several ToF pixels into a macro-pixel [30]. Each pixel shown in the magnified macro-pixel has a different demodulation frequency, f_ℓ , which is an integer multiple of f_0 , the base modulation frequency ($f_\ell = \{\ell f_0\}_{\ell=1}^{16}$). These pixels, connected to different control demodulating signals, are called subpixels. This concept can be implemented with architectures that allow a programmable number of sub-pixels, e.g., up to 16 sub-pixels in this work. As a result, the arrangement of sub-pixels can be altered or modified to suit specific requirements or design considerations.

In addition, the measurement setup in Fig. 2(c) is considered. As seen there, it consists of a modulated illumination signal (LED or laser), a scene that reflects the light, and a ToF sensor that demodulates light echo. Single-shot multifrequency operation requires sufficient frequency spread in the illumination waveforms. We obtain this using a sharp light pulse that closely emulates a Dirac delta function. The pulse is periodically repeated during the exposure time to gain SNR. The received light's amplitude decays and is phase-shifted with the distance to the target. As mentioned, most conventional ToF systems operate with sine wave modulation. Generally, these systems are homodyne, meaning the pixels demodulate at the same frequency as the modulated light source. In our approach, each subpixel can demodulate at a different f_ℓ frequency, achieving close-to-zero harmonic distortion by implementing a resonant circuit at the desired frequency for each control signal, taking advantage of the intrinsic capacitance of the control gates and adding an inductor of adequate value [31], [32]. Thus, simultaneously acquired undistorted Fourier measurements at various frequencies are provided, as intended.

B. ToF Pixel Architecture

Typically, an iToF sensor works with two accumulation regions, called floating diffusion (FD) nodes, in each pixel to demodulate the photogenerated signal. The lock-in pixel architecture is commonly used in iToF imagers [33]. Different designs for lock-in pixels have been proposed, such as the Photonic Mixer Device (PMD) [34], lateral electric field charge

modulator [35], current-assisted photonic demodulator [36], and gate assisted photonic demodulator [37].

Fig. 3(a) depicts a circuit diagram of the two-tap lock-in pixel with a symmetrical active pixel sensor structure and the system architecture of the iToF sensor [31]. Each pixel consists of a $4\text{ }\mu\text{m} \times 4\text{ }\mu\text{m}$ Pinned Photo Diode (PPD), transfer gate (TG) transistors, diode-switched transistors performing anti-blooming, global shutters (GS), reset switches, source followers (SF), and row-select transistors [14]. Current sources are connected to two-tap pixel outputs, COL_A and COL_B, for the biasing of the source followers. The system architecture highlights the configuration of the main blocks of our proposed system. The Column Filtering Stage (CFS) is placed in each column to filter and measure the voltage difference of the outputs (COL_A and COL_B), ΔV . The outputs of the CFS are directly connected with a capacitive feedback amplifier circuit in the Double-Delta-Sampling (DDS) stage. The final outputs of the system are $V_0 \pm A\Delta V$, where V_0 is a DC offset and A is the gain of the amplifier. The system consists of an array of 128×128 pixels, the CFS, and the DDS stages. It has an array-shared modulating signal generator, row and column control circuits, and driver circuits [38].

Fig. 3(b) illustrates the timing diagram of the iToF pixel proposed in this work. The operation procedure starts with the pixel reset by setting the FD nodes to a specific value. The next step is pixel demodulation, in which the gates of the TG transistors start to demodulate. At the same time, they are synchronized with the light illumination (i.e., demodulating with 0° and 180° phase shift with respect to the emitter) so that the photo-generated electrons in the PPD are transferred to the corresponding FD during the exposure time. During this time, the voltage of both FDs starts to drop proportionally to the number of electrons they accumulate. The GS signal is ON during the exposure time to ensure that all pixels' measurements are synchronized. The final output is buffered by the SF and stored on its source, ready to be read to the output lines through row-select transistors. The final step is the readout, accomplished by sequentially reading the outputs of the pixels per row, repeating this as many times as the number of rows. To assess the performance and functionality of the proposed hardware architecture, we have conducted not only nominal but also post-layout electrical simulations using commercial CMOS

technology design kits. These simulations faithfully emulate the real operation of the 3D imager by taking into account offsets, nonlinearities, charge injections, and cross-coupling effects as well as parasitic capacitances, resistances, and other non-ideal effects caused by the physical implementation of the system [31], [38].

III. MEASUREMENT MODEL AND DEPTH RETRIEVAL ALGORITHM USING MULTIFREQUENCY MEASUREMENTS

In this section, the measurement model of iToF systems, as well as depth retrieval and accuracy improvement algorithms based on multifrequency measurements, will be presented.

A. Baseline Model

In iToF, the distance map of the scene can be calculated with

$$d = \frac{c\phi}{4\pi f} = c \frac{t_d}{2}, \quad t_d = \frac{\phi}{2\pi f}, \quad (1)$$

where d is the distance, c is the speed of light, ϕ is the phase shift, f is the modulation frequency, and t_d is the time delay. Phase ambiguity limits the total detectable distance range, inversely proportional to the base modulation frequency. For example, at 4 MHz, a range of 37.5 m can be unambiguously achieved. In the light echo signal, besides the phase, ϕ , two more parameters are unknown in the incoming signal: amplitude and offset. Therefore, a single measurement is insufficient to recover the phase information, and at least three measurements are needed to calculate a distance. Usually, four samples are measured, known as the Four Bucket Method [4], [39], corresponding to applied phase shifts of 0, $\pi/2$, π , and $3\pi/2$. These four measurements can be acquired in two cycles in a two-tap lock-in pixel structure. In the baseline mono-frequency approach, the phase shift can be calculated by,

$$\phi = \arctan \left(\frac{m(3\pi/2) - m(\pi/2)}{m(0) - m(\pi)} \right), \quad (2)$$

where $m(0)$ to $m(3\pi/2)$, as in (6), are proportional to the number of accumulated photo-generated carriers in each tap of the pixel (OUT_A and OUT_B in Fig. 3(a)) during the exposure time (t_{exp}). This method works well for single bounces but cannot resolve MPI because using a single modulating frequency means that the “bandwidth” of our measurements in the frequency domain is zero, and thus, more than one target can not be distinguished from a single resultant target in the range domain. Consequently, in AMCW ToF, the depth estimation from phase measurement suffers from MPI and ambiguity or phase-wrapping issues. Hence, a desirable phase estimation algorithm should jointly correct for any distortions and phase-wrapping.

In the following, we provide a general sensing model for the ToF imaging system, shown in Fig. 4. As mentioned, we consider an illumination pulse with a very low duty cycle, ideally a Dirac delta function. This allows for attaining high SNR while complying with eye safety regulations. Therefore, the illumination function $i(t)$ that interacts with a scene can be

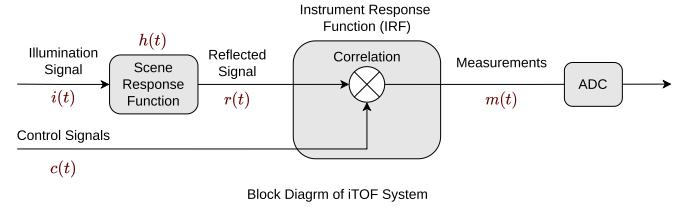


Fig. 4. Block diagram of the mathematical model of iToF system.

modeled as,

$$i(t) = \sum_{q=0}^Q \delta(t - qT), \quad T = 1/f, \quad (3)$$

where the number of impulses, Q , is t_{exp}/T . The emitted light will be reflected from the scene according to a scene response function that, for a single path, can be modeled as

$$h(t) = a \delta(t - t_d), \quad d = \frac{c}{2} t_d. \quad (4)$$

The aim is to estimate the amplitude, a , and t_d . The reflection function, $r(t)$, which represents the back-reflected illumination from the emitter, can be modeled as the convolution operation between the illumination and scene response function,

$$r(t) = (i * h)(t) = a \sum_{q=0}^Q \delta(t - t_d - qT). \quad (5)$$

Finally, the sinusoidal demodulation control signal, $c(t) = \cos(\omega t)$, synchronized with the illumination signal, is applied to the TG transistors in Fig. 3(a). As a result, the obtained measurements $m(\tau)$ are discrete samples of the cross-correlation between the reflected signal and the demodulation control signal applied to the pixel,

$$\begin{aligned} m(\tau) &= c(t) \otimes r(t) = c(t) \otimes (i * h)(t) = (i \otimes c) * h(t) \\ &= a \cos(\omega\tau - \phi) + b, \quad \phi \in [0, 2\pi], \end{aligned} \quad (6)$$

where $\phi = \omega t_d = 2d\omega/c$ is the frequency-dependant phase.

In reality, the sinusoidal hypothesis does not hold. The measurements are samples of a scaled and shifted Instrument Response Function (IRF), which models the aggregated electro-optical behavior of the iToF system, including the low-pass effect of the electronics, the rising and falling times of the light emitters, and inner multi-path effects in the lens system. In this work, we consider a realistic IRF model arising from electrical simulations, which introduces frequency-dependant scaling and phase-shifting in (6). The effect of the IRF can be counteracted by an element-wise calibration in the frequency domain equivalent to deconvolution in the time domain.

As mentioned, to estimate parameters ϕ and a , four samples are typically taken at $\omega\tau = [0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}]$. For a given modulation frequency, these four measurements are enough to form the complex number,

$$\begin{aligned} X &= \cos(\phi) - j \sin(\phi) \\ &= (m(0) - m(\pi)) + j(m(3\pi/2) - m(\pi/2)). \end{aligned} \quad (7)$$

where measurements $m(0)$ to $m(3\pi/2)$ are exactly the measurements used in (2). These measurements form Fourier coefficients of the scene response function.

B. Multipath Measurement Model

The previous formulation can be extended to account for MPI. In this case, the scene response function takes the form of a P-sparse filter,

$$h(t) = \sum_{k=1}^P a_k \delta(t - t_k), \quad t_k = 2 \frac{d_k}{c}, \quad (8)$$

where t_k and d_k are time delay and distance in the case of multiple paths, and P is the number of optical paths. Therefore, with the same illumination function $i(t)$, the reflection function can be formed:

$$r(t) = \sum_{k=1}^P \sum_{q=0}^Q a_k \delta(t - t_k - qT). \quad (9)$$

In (9), $r(t)$ is a weighted sum of impulses with varying time shifts (time delays). The interaction between the reflected signal and the ToF sensor results in the measured signal m_ℓ . Hence, the measurements take the form

$$m_\ell(\tau) = \sum_{k=1}^P a_k \cos(\omega_\ell \tau - \phi_k) + b_k, \quad \phi_k = \omega_\ell t_k, \quad (10)$$

where ℓ is the frequency index for the multiple measurements, taken at different modulation frequencies ($\omega_\ell = \ell \omega_0$). For a certain modulation frequency, four measurements are enough to form the complex number,

$$\begin{aligned} X_\ell &= \frac{1}{2} e^{j\omega_\ell \tau} \sum_{k=1}^P a_k e^{-j\omega_\ell t_k} \\ &= \frac{1}{2} e^{j\omega_\ell \tau} \sum_{k=1}^P a_k (\cos(\omega_\ell t_k) - j \sin(\omega_\ell t_k)). \end{aligned} \quad (11)$$

The Four Bucket method provides an estimate of the phasor in (11) from the measurements,

$$X_\ell = (m_\ell(0) - m_\ell(\pi)) + j(m_\ell(3\pi/2) - m_\ell(\pi/2)), \quad (12)$$

where the measurements $m_\ell(0)$ to $m_\ell(3\pi/2)$ are exactly the same as measurements in (2), but for different frequencies.

C. Depth Retrieval Algorithms Using Multifrequency Measurements

Fig. 5 shows the real part of the scene response function in the Fourier domain for representation simplicity for the case of three bounces. As seen, three different response functions, one for each time delay, are obtained at t_1 , t_2 , and t_3 , each represented with a complex phasor. In the Fourier domain, the sensor measures the sum of the three sinusoids from the three phasors. Ideally, sampling more points (Fourier coefficients) from the sum figure (sum of sinusoids) results in a more accurate retrieval of the multipath parameters, but with the cost of speed and power consumption. Hence, the idea is to take as few measurements as

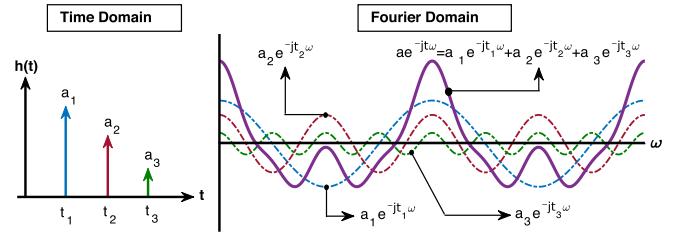


Fig. 5. Scene response function of the ToF system in the time (left) and the real part of Fourier (right) domains: the individual harmonic constituents arise from the three individual bounces and the composed result (purple figure).

possible on the sum curve, allowing us to calculate time delays with reasonable accuracy.

MPI in frequency-domain ToF mode can be reinterpreted as a spectral estimation problem consisting of recovering the parameters of a sum of complex exponentials from noisy samples. In the following, we will outline two fast procedures to recover MPI components from multiple frequency measurements to recover depth, namely the Matrix Pencil, [4], [40], and the Inverse Discrete Fourier Transform methods (IDFT) [41].

The matrix pencil method is a non-iterative algorithm of a parametric nature, yielding a high-speed execution. Instead of the standard sparsity-promoting ℓ_1 regularization strategy, this parametric model estimation approach uses a sparsity assumption translated into a matrix rank constraint in a Toeplitz matrix. This yields a closed-form, non-iterative solution to the multipath estimation problem in contrast to iterative optimization strategies, which might have problems with convergence and require much more computation time. The algorithm requires $2P$ nonzero frequency measurements ($2P + 1$ including the zero frequency) to recover P bounces.

The factor 2 in the $2P + 1$ accounts for positive and negative frequencies, while the negative frequencies can be retrieved from the positive ones in (12) by simple conjugation, provided that the scene response function is a real function. Thus, without model mismatch or noise, P paths can be retrieved from P measurements at P different nonzero frequencies. The algorithm scales to any modulation frequencies and interfering paths, where the number of detectable paths is only limited by SNR and modulation frequency bandwidth.

Our second approach uses the most basic spectral estimation tool, which is the Fourier transform. This method allows us to have rapid and straightforward 3D depth formation by means of the IDFT. To improve the accuracy, instead of typical IDFT calculation, the normalization factor can be increased, resulting in a higher number of discrete steps in the time domain, resulting,

$$x_n = \sum_{\ell=1}^N \frac{1}{N} X_\ell e^{j2\pi(\ell-1)n/N'}, \quad (13)$$

where n is a running index on the time domain, N is the number of steps in both domains, X_ℓ the vector of measurements as in (12), and $N' = \lambda \times N$ is the number of samples in the time domain, where λ is the “super-resolution factor,” i. e., the ratio between the number of measurements in the time domain and the number of measurements in the frequency domain. Therefore,

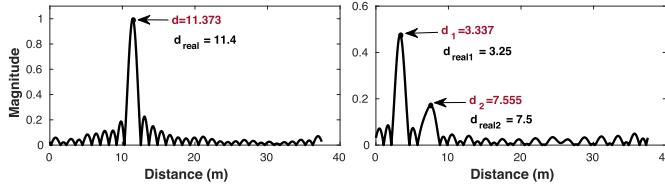


Fig. 6. Magnitude versus distance of IDFT of the 16 frequency measurements in the case of a single (left) and double bounces (right).

more points in the target domain (time domain in this case) can be generated. The optimum value of λ is the value beyond which the accuracy no longer improves. At this point, the result is no longer limited by the grid resolution, and refining the temporal grid is futile. The effect of λ on the retrieved distance error has been studied in our previous work [41]. Here, $\lambda = 1000$ is selected in most of the computations, a relatively good trade-off between speed and accuracy.

The magnitude of the x_n in (13) will show a single peak in the time domain in the case of a single bounce and secondary peaks when having more than a single bounce. These peaks are exactly where the Dirac delta function was in the scene response function. Since time and frequency are interrelated by the Fourier transform, the time delay value corresponding to the location of the detected peak in the time domain can be calculated. By estimating the time delay, the distance can be directly recovered from (1). Fig. 6 shows the magnitude with respect to the distance of the 16 frequency measurements after IDFT in the case of single (left) and double bounces (right). The targets are assumed to be at 11.4 m in the single bounce and 3.25 m and 7.5 m in the double bounce. As can be seen, the main peaks of the magnitude correspond to the position of the targets.

IV. DATA REFINEMENT

Here, we propose data refinement strategies to improve accuracy by minimizing the signals' residual harmonic content and noise. The methods integrate valuable information (e.g., low rank, Toeplitz structure, depth statistics, etc.) within the signal processing pipeline. The more a priori information is used, the more the solution space is restricted, thus counteracting the effect of disturbances and noise.

A. Self-Compensation

The main idea of this method is to use the measurements obtained at all frequencies to cancel out the possible distortion present in the measurements at each of the frequencies. This is a flexible procedure that can compensate for systematic effects such as harmonic distortion and can be carried out through a calibration process. The ideal model of measurements can be presented as follows,

$$\vec{X}_I = \Phi_I \vec{R}, \quad \vec{X} \in \mathbb{C}^\ell, \quad \vec{R} \in \mathbb{C}^\eta, \quad \Phi \in \mathbb{C}^{\ell \times \eta}, \quad \ell \ll \eta \quad (14)$$

where Φ_I is the partial Discrete Fourier Transform (DFT) matrix at chosen frequencies and \vec{R} is a sparse vector of reflectivities.

The real model is:

$$\vec{X}_\ell = \Phi_\ell \vec{R}, \quad (15)$$

where Φ_ℓ is the set of distorted complex sinusoids at chosen frequencies, affected by harmonic distortion, obtained from measurements or realistic simulations. The objective of Self-Compensation (SC) is to cancel the harmonics of measurements in a simple linear fashion, which means to retrieve \vec{X}_I from \vec{X}_ℓ utilizing a linear operator or compensation matrix:

$$\hat{\vec{X}}_I = C \vec{X}_\ell = C \Phi_\ell \vec{R}, \quad (16)$$

where $C \in \mathbb{C}^{\ell \times \ell}$ is the compensation matrix. On the other hand,

$$\hat{\vec{X}}_I = \vec{X}_I \Leftrightarrow C \Phi_\ell = \Phi_I, \quad (17)$$

Therefore,

$$C \Phi_\ell \Phi_\ell^* = \Phi_I \Phi_I^*, \quad (18)$$

Hence,

$$C = \Phi_I \Phi_I^* (\Phi_\ell \Phi_\ell^*)^{-1} = \Phi_I \Phi_\ell^\dagger, \quad (19)$$

where Φ_ℓ^\dagger is the Moore-Penrose Pseudoinverse of Φ_ℓ .

B. Self-Compensation With Probability Density Function

In order to give more importance to the depth locations where depth appears more often in reality (based on the system), the self-compensation method can be modified with a weighting matrix. This focuses on reducing distortion over those domain areas that are statistically most relevant. It is important to note that this approach relies on a priori scene knowledge and does not apply to scenarios where this does not hold. The diagonal weighting matrix, W , can be calculated from the Probability Density Function (PDF) of the system with respect to the distance:

$$W_{i,i} = f(d_{\text{PDF}}), \quad (20)$$

where f is a sublinear function, such as the square root (adopted in the experimental validation) and d_{PDF} is the empirical PDF. More probable depths are weighted more heavily, and close-to-zero probability values (e.g., lower than 10^{-4}) are masked. The new compensation matrix can be formed as,

$$C = \Phi_I W \Phi_I^* (\Phi_\ell W \Phi_\ell^*)^{-1}. \quad (21)$$

C. Cadzow Denoising

The Cadzow Denoising algorithm forces the Toeplitz matrix formed with the vector of measurements to be both low-rank and preserve the Toeplitz structure. The algorithm runs a predefined number of iterations with two sequential steps, forcing first the desired rank and later the Toeplitz structure. Algorithm 1 presents the Cadzow iterative denoising method implemented in this work, which follows the procedure outlined in [42]. The algorithm's input is the Fourier coefficients obtained from the measurements of our iTToF sensor as in (12). First, it builds a rectangular Toeplitz matrix (T) from the vector of Fourier coefficients in the first row and column of T and repeats them along diagonals. Then, it performs a Singular Value Decomposition

Algorithm 1: Cadzow Denoising.

Data: The Fourier coefficients $\{\mathbf{X}_\ell\}_{\ell=-L}^{\ell=L}$

Result: $\{\mathbf{X}'_\ell\}_{\ell=-L}^{\ell=L}$

- 1 **for** 15 iterations **do**
- 2 $T_{row1} \leftarrow \text{flip}(X_\ell(1:L+1))$;
- 3 $T_{col1} \leftarrow X_\ell(L+1:end)$;
- 4 $T \leftarrow \text{Toeplitz}(T_{col1}, T_{row1})$;
- 5 $[U, S, V] \leftarrow \text{SVD}(T)$;
- 6 $T' \leftarrow US'V^T$; // most significant diagonal elements of S
- 7 **for** $i = 0$ to L **do**
- 8 $T_{row1}(i+1) \leftarrow \text{mean}_k(T'(k, k+i))$;
- 9 $T_{col1}(1) \leftarrow T_{row1}(1)$;
- 10 **for** $j = 1$ to L **do**
- 11 $T_{col1}(j+1) \leftarrow \text{mean}_k(T'(k+j, k))$;
- 12 $X'_\ell(1:L+1) \leftarrow \text{flip}(T_{row1})$;
- 13 $X'_\ell(L+1:end) \leftarrow (T_{col1})$;

(SVD) of T and approximates this matrix by retaining the largest eigenvalues. Finally, it converts T again into a Toeplitz matrix via diagonal averaging and constructs a denoised approximation X'_ℓ of X_ℓ .

V. RESULTS AND DISCUSSION

In this section, we present the results of a series of experiments focused on evaluating the performance of the proposed joint design of the sensor architecture and a depth retrieval algorithm. We performed a series of electrical simulations under different conditions to demonstrate the concept of acquiring single-shot multifrequency demodulation for single, double, and triple bounces. Despite our system being optimized for solving the reflective MPI problem, its ability to mitigate the effect of diffusive MPI is also demonstrated. The base frequency modulation is $f_0 = \omega_0/2\pi = 4$ MHz, while the other subpixels reach up to 64 MHz for the sixteenth frequency (4 MHz increment). The integration time in simulations is 50 μ s, the pulse width of the illumination signal is 5 ns in all simulations, and its repetition frequency is equal to the base frequency (2% duty cycle). The pulse widths of the order of nanoseconds are feasible to implement considering the state-of-the-art commercially available illumination drivers and Vertical Cavity Surface Emitting Lasers (VCSEL).

Fig. 7 illustrates the electrical simulation results of the voltage difference between the pixel's taps, ΔV , for the first subframe, showcasing the ability of the system to provide quasi-sinusoidal correlation measurements with low harmonic distortion. The system's output corresponds to samples of the cross-correlation function between the incoming illumination signal and the demodulation signals controlling the gates. The demodulating signals were fixed to perform this simulation while the illumination signal was time-shifted. As expected from (10), since the illumination is an ultrashort-pulsed signal, the cross-correlation function must be a sinusoidal function, as demonstrated in Fig. 7, as opposed to existing ToF systems (cf. Fig. 3 of [43]).

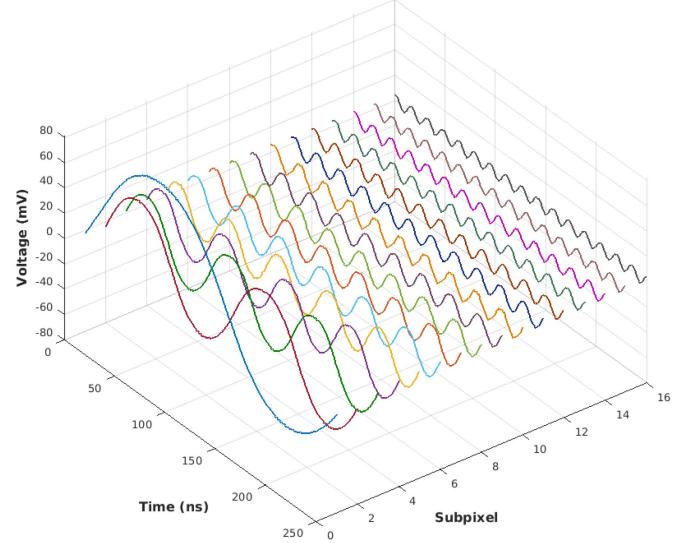


Fig. 7. Electrical simulation results showing the nominal output voltage difference as a function of the time delay for each subpixel.

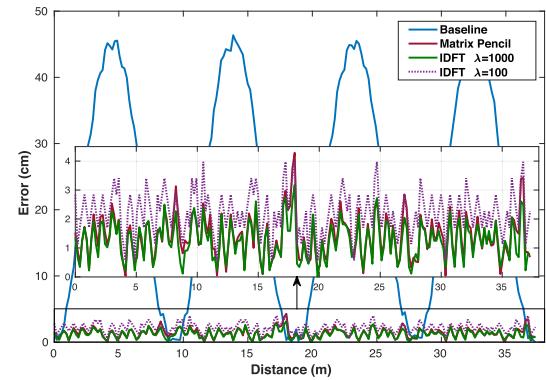


Fig. 8. Measured depth error with the baseline, matrix pencil, and IDFT methods for a single bounce.

A. Single-Bounce Depth Retrieval

The circuit implementation in the simulation setup of our iToF system is based on Fig. 3. The illumination is a low-duty cycle pulse, and the demodulation control signals at the transfer gates TG_A and TG_B are clean sinusoidal signals thanks to the resonant demodulation technique from [31]. To investigate the system's functionality and verify the validity of the mathematical models, a macro-pixel consisting of 16 subpixels is designed and integrated into a full iToF array.

Fig. 8 depicts the error estimated over the whole possible range of 0–37.5 m with the three methods introduced in Section III using the electrical simulations of the proposed ToF architecture. The extracted error values of matrix pencil and IDFTs, taken from 16 frequencies, are indicated. As can be seen, the matrix pencil and IDFT methods outperform the classical four-phase baseline algorithm in the single-path case thanks to the combination of measurements at different frequencies. Leveraging measurements at different frequencies allows

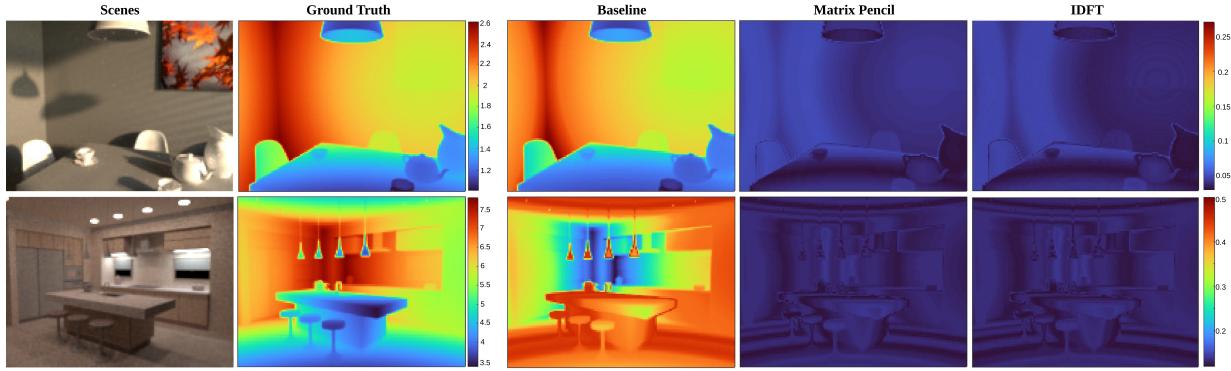


Fig. 9. Scenes and their ground truth from the *iToF2dTof* dataset and their corresponding depth errors with the baseline, matrix pencil, and IDFT depth estimation methods. All scales are in meters.

TABLE I
COMPARISON OF DISTANCE RMSE VALUES WITH DIFFERENT NUMBERS OF FREQUENCY MEASUREMENTS

# Frequencies	1	2	4	8	16
RMSE (cm)	Matrix Pencil	27.74	13.79	3.31	2.65
	IDFT	26.72	11.35	3.81	2.62

annihilating the wiggling effect since harmonic frequencies contribute to the estimation instead of disturbing it, as would be the case under the classical mono-frequency hypothesis. The baseline approach uses only the base frequency ($f = 4$ MHz), while the matrix pencil and IDFT use multifrequency in a single shot. The overlaid plot shows a magnification of the results for the matrix pencil and IDFT methods. While the performance of the matrix pencil and IDFT are very similar, the matrix pencil is faster in the case of $\lambda = 1000$, and it is slower in the case of $\lambda = 100$. This illustrates that, as explained in Section III, the IDFT method can trade accuracy for speed.

Additionally, the effect of the number of frequency measurements on the accuracy has been investigated. Table I compares the matrix pencil and IDFT methods' accuracy with different frequency measurements. The RMSE is measured for 200 equally separated target locations over the range of 0–37.5 m with $\{2^\alpha\}_{\alpha=0}^{4}$ measurements obtained using the depth retrieval algorithms. As expected, the error is reduced by increasing the number of frequency measurements. It is worth noting that while RMSE values for matrix pencil and IDFT are reported in the case of 16 frequency measurements in Fig. 8, the RMSE of all three methods is very close in mono-frequency measurement, where the baseline approach is the fastest and IDFT (with $\lambda = 1000$) is the slowest. Adding more measurements at higher frequencies helps but to some limits. A performance decay is anticipated at some point, especially when adding measurements at frequencies approaching the overall system bandwidth.

To compare these methods in more detail, the system's measurements are tested against the *iToF2dTof* 3D image dataset provided in [44]. The ground truth of two different scenes from the dataset is selected. Fig. 9 plots the original RGB images, the ground truth, and the retrieved depth errors using the three

TABLE II
COMPARISON OF DISTANCE RMSE VALUES OF THE BASELINE, MATRIX PENCIL, AND IDFT METHODS

Paths	Retrieved Distance RMSE (cm)		
	Baseline	Matrix Pencil	IDFT
One	17.04	1.39	1.27
Two	1st	-	2.24
	2nd	-	15.31
Three	1st	-	2.12
	2nd	-	17.92
	3rd	-	26.57

methods considered in this work. The measurement steps are adjusted using linear interpolation to solve the grid mismatch between our electrical simulation results and the ground truth from the *iToF2dTof* dataset. Then, the depth is recovered with the proposed methods for each pixel. As can be seen, the matrix pencil and IDFT show great performance compared to the baseline method, which shows poor performance in some parts (e.g., in the corner of the top image). The actual values of the RMSE of the top image can be found in Table II.

B. Reflective MPI: Two-Bounces Depth Retrieval

In reality, several reflections with different phase shifts can reach the pixel. Consequently, the measured voltage drop in the pixel, proportional to the accumulated charges during the exposure time, includes the sum of all these reflections. The retrieved phase shift under mono frequency assumptions, therefore, is erroneous. To demonstrate the capability of our proposed single-shot multifrequency architecture against the reflective MPI problem, we have performed electrical simulations for the case in which two bounces are received simultaneously.

Fig. 10 provides plots of the retrieved distance errors obtained using the matrix pencil and IDFT algorithms described in Section III for the two-bounces case. Simulations were carried out by fixing the first bounce at relatively close (1 m) and long (4 m) distances and shifting the position of the second bounce. In addition, different ratios of illumination signal amplitudes for the first, a_1 , and second, a_2 bounces were considered. Since

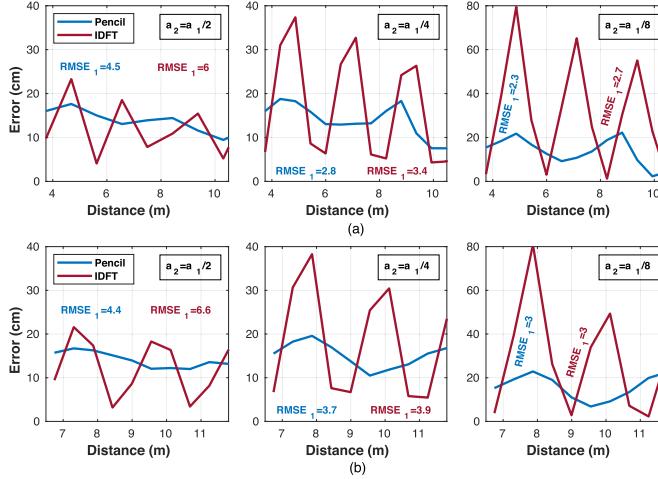


Fig. 10. Retrieved distance errors of the electrical simulations with the matrix pencil and IDFT methods in the presence of two bounces. The measured depth errors of the second bounce are shown for different amplitude ratios with respect to the first bounce. The first bounce is fixed at a constant distance of 1 m and 4 m for the first (a) and second (b) sets of experiments, respectively. The retrieved depth RMSE values (scales in centimeters) for the first bounce are also included in the figures, denoted as RMSE₁.

the distance error reduces with increasing the frequency measurements, as shown in Table I, only depth, retrievals from 16 frequencies (the best case) are calculated in Fig. 10. As can be seen, the proposed system can retrieve the distance in the presence of the second bounce in a single shot. Even though the super-resolution factor is selected to be relatively high in the IDFT method ($\lambda = 1000$), unlike the single bounce case, the matrix pencil is more accurate when dealing with MPI.

Here, we take advantage of our realistic electrical simulations to answer two critical questions: the minimum detectable illumination power and the minimum and maximum recoverable range for the second bounce. Different ratios of a_1 to a_2 are considered, and the retrieved depth from the outputs of electrical simulations is tested. The distance can be recovered perfectly from the system until a_2 is 5% of a_1 . Below this ratio, errors in some ranges are observed. To answer the second question, we repeated the same simulations where the first target is fixed at 1 m and the power of the second bounce is 1/8 of the first one. The minimum and maximum distances that could be retrieved for the second bounce were 2 m and 37.5 m for the matrix pencil method. Also, the distance of the first bounce (1 m) is correctly recovered when the second bounce is placed between 2–37.5 m. Therefore, the distances in the matrix pencil method can be accurately recovered when the targets are at least separated 1 m from each other. In the IDFT method, while the first target's distance (1 m) can be retrieved in the whole range (similar to the main peak in the right diagram of Fig. 6), the second distance can be only recovered from 3.5 m to 17 m. Even in this range, some distances can not be recovered. The problem with the IDFT in MPI recovery is that for some locations of the second target, a virtual secondary target arising from constructive interference of side lobes becomes dominant and appears at the wrong location. Apart from its lower accuracy, this is a problem that may indeed

happen in practice and one of the reasons why IDFT may not be very robust. More frequency measurements are needed for the IDFT method to perform more accurately.

For clarity, two scenes from the *iToF2dToF* dataset have been used by artificially adding another bounce at a distance of 1 m in some areas. These areas are marked with a red rectangle in Fig. 11 and all the pixels in this area are deep green, which represents 1 m distance. Fig. 11 shows images of the ground truth for the first and second bounce and the recovered depth with both methods. Please note that the backgrounds, which are the unmarked areas with single depth, are recovered from the single bounce simulations and are depicted in both images. As can be seen, the matrix pencil method outperforms the IDFT in the two-paths case. The actual values of the RMSE of the top image can be found in Table II.

To compare the computational time and the accuracy of the two methods, Fig. 12 provides plots of the Mean Absolute Error (MAE) of the distance to computational time in single (left) and double bounce (right) cases. While the MAE and computational time of the matrix pencil method are constant, these two parameters are in a trade-off that depends on the λ in the IDFT method. While the IDFT method can be a perfect candidate in a single-bounce depth recovery, the matrix pencil method performs better than the IDFT in MPI. Even though detecting the first target in MPI is the main goal in several applications, and the IDFT method might be interesting where the speed can be traded for accuracy, the matrix pencil shows superior results for the same computational budget results. Therefore, the matrix pencil is a stronger candidate for our codesign.

To measure the matrix pencil's robustness to noise, simulations were performed on the proposed algorithm using ideal outputs from the system (clean sinusoidal shapes) as inputs of the algorithm. In contrast to the other evaluations in this section that study accuracy, the performance metric evaluated here is precision, allowing for an assessment of the algorithm's response to noise under controlled conditions. The algorithm's performance is evaluated for different SNR of the measurements in Fig. 13 for single (left) and double bounce (right) cases. The noise source in the simulations is random Gaussian white noise. Fig. 13 provides plots of means and standard deviations of depth RMSE of the output signals extracted from the algorithm in over 100 samples versus SNR for different numbers of frequency measurements. In the two-bounce cases, only the RMSE of the first bounce is shown for clarity, while the results of the second bounce had the same behavior. The SNR of our hardware is roughly estimated and shown with a solid vertical line. The value of the RMSE is significantly reduced as the number of frequencies increases. These results show the importance of multifrequency measurement for reliable depth retrieval, especially in noisy environments. The ratio of the retrieved signal's amplitude for the first and second paths was $a_2 = a_1/4$. In general, the algorithm can recover the distances with less error when a_2 is closer to zero, meaning that the second path is weak compared to the dominant main path. However, this is not valid

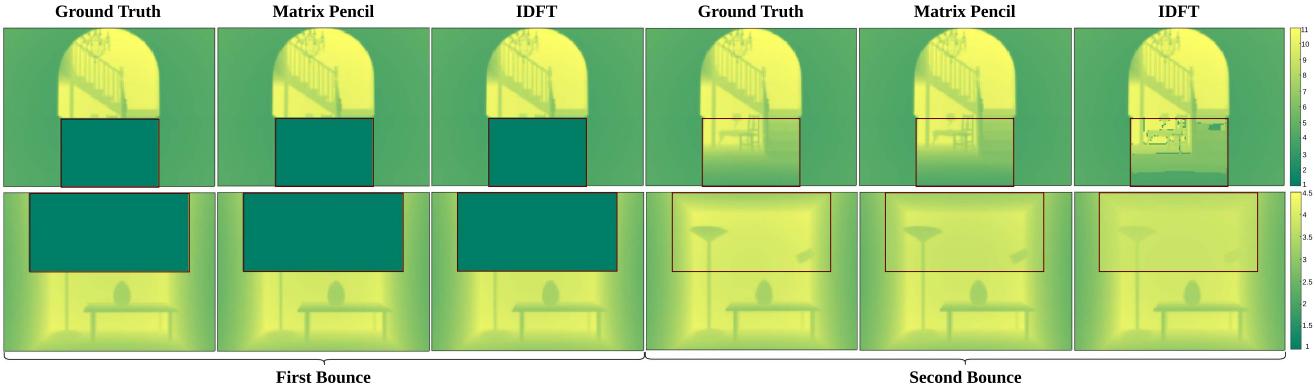


Fig. 11. Images of the retrieved depth of the first (left images) and second targets (red rectangles, right images) in case of two bounces. The matrix pencil method shows more accurate estimation, especially in retrieving the second target. Considering the super-resolution factor ($\lambda = 1000$) used in these experiments, it is also faster. The backgrounds, where the pixels receive a single path (unmarked areas), are depicted in both images. All scales in meters.

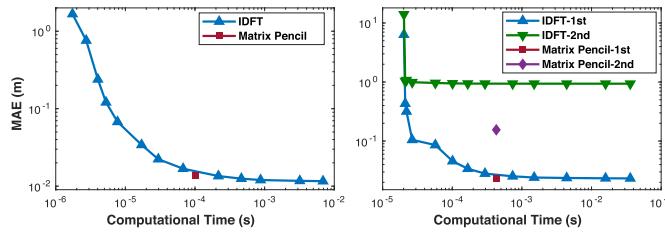


Fig. 12. The distance MAE (mean absolute error) versus computational time in single bounce (left) and two bounces (right) in the matrix pencil and the IDFT methods. Although the IDFT can be faster due to having a programmable super-resolution factor, the matrix pencil shows more promising results when dealing with MPI.

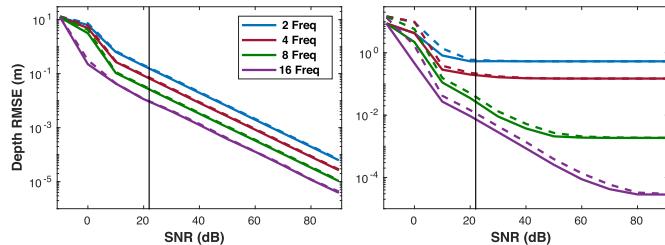


Fig. 13. RMSE of the reconstructed depth of single (left) and double (right) bounces with respect to the SNR for the different numbers of frequencies using the matrix pencil algorithm. The solid and dashed lines are mean and standard deviation values of RMSE computed from over 100 samples with random Gaussian white noise.

in practice because the system will not be able to detect a tiny portion of a second bounce, and the noise becomes dominant in those cases.

C. Reflective MPI: Three-Bounces Depth Retrieval

The simulations, in this case, consist of three reflected signals. The amplitude of the second and third bounces was set to $1/4$ and $1/16$ of the primary reflection, respectively. For simplicity in the simulation setup, the first and second bounces are considered at constant distances of 1 m and 4 m, respectively, while the third bounce shifts from a close range to a relatively long distance (5 m to 15 m). As in previous cases, 16 modulation frequencies

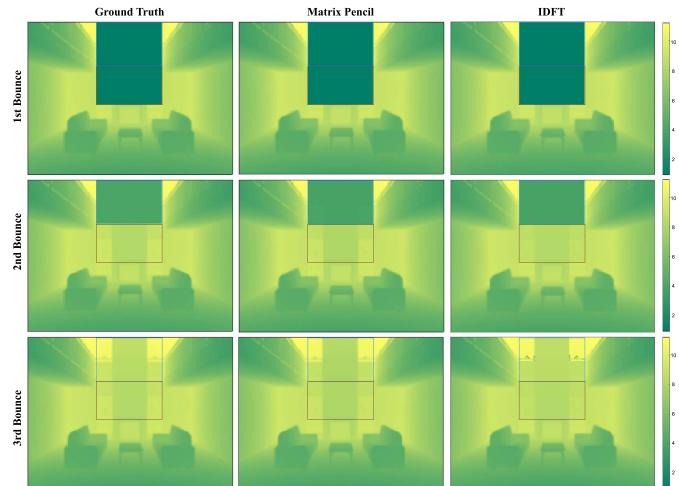


Fig. 14. Images of estimated depth with the matrix pencil and IDFT methods from the electrical simulations of the single-shot iToF system in the presence of three bounces and the ground truth as ideal values. The backgrounds, where the pixels receive a single path (unmarked areas), are depicted in both images. All scales in meters.

are acquired simultaneously, spaced 4 MHz apart. The minimum distance for the third bounce recovered in this experiment was 6 m, which means 2 m separation with respect to the second target.

Fig. 14 shows the estimated depth from the electrical simulations in the case of three bounces recovered from the ground truth. Two artificial layers are added at a distance of 1 m (blue and red rectangles, pixels [1:65, 55:105]) and 4 m (blue rectangle, pixels [1:35, 55:105]), respectively. In this experiment, the pixels' distances for the blue part are recovered from the three-bounce simulations, the red part from the two-bounce simulations, and the rest from the single-bounce simulation. As expected, the recovered depths for the first, second, and third targets match the ground truth with centimeter accuracy in all pixels for the matrix pencil method and most pixels for the IDFT method.

Table II shows the depth RMSE of the shown images in Fig. 9 (top image), Fig. 11 (top image), and Fig. 14 to evaluate the depth

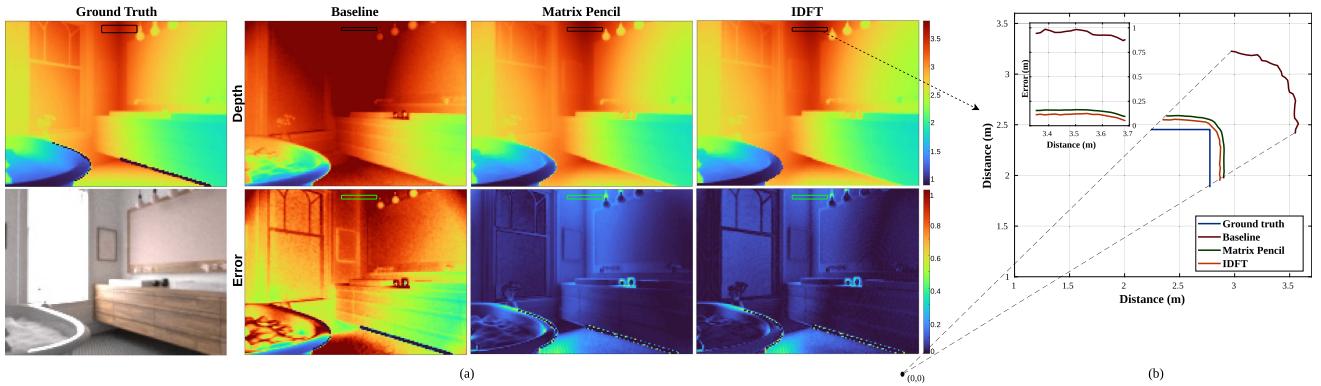


Fig. 15. Depth and error images reconstructed from our iToF imaging system (a). The ground truth is the “bathroom scene” of the *iToF2dToF* dataset. The results are extracted from transient images of the dataset. All scales are in meters and restricted to the ground truth range. The corner area affected by MPI is highlighted with a box and analyzed (b). The matrix pencil and IDFT can provide a faithful depth reconstruction while coping with MPI.

retrieval capabilities of mentioned three methods quantitatively. The results demonstrate the superiority of the matrix pencil method for depth recovery.

A critical aspect regarding our implementation is the maximum possible number of bounces that can be retrieved from the measurements. Ideally, the algorithm can retrieve as many bounces as frequency measurements. To analyze the practical limits of the hardware implementation, we performed several experiments with a different number of bounces and relative amplitudes and evaluated the retrieved distances. While the accuracy depends drastically on the noise, the amplitudes of the bounces, and their separation from each other, the algorithm performed well, having four, eight, and sixteen bounces in many cases. For example, in an experiment for 16 bounces, the first, second, and third targets were fixed in 1 m, 4 m, 8 m, respectively, and the other 13 targets were separated 2.25 m from each other. In this case, without the presence of noise, when the amplitudes of the bounces were $\{\frac{1}{4^{2\alpha}}\}_{\alpha=0}^{15}$ fifteen distances were correctly recovered with errors lower than 15 cm and one target (third bounce) was retrieved with an error of 1.5 m.

D. Diffusive MPI

Even if the main advantage of the proposed approach, the single-shot multi-frequency acquisition with low harmonic distortion, is best exploited in the case of reflective MPI, the same system can also be used to mitigate the effect of diffusive MPI. Fig. 15 shows depth and error images of the “bathroom-scene” of the *iToF2dToF* dataset reconstructed from our iToF imaging system and a cut of the corner area affected by diffusive MPI with respect to transient images of the ground truth. For each scene, the data set includes 2000 transient images covering 66.66 ns with step size 33.33 ps. The measurements are interpolated and resampled to achieve equal step size in the measurements and transient samples. We leverage the realistic correlation measurements obtained through electrical simulations to emulate measurements obtained from our iToF sensor for each transient profile. This is achieved by convolving the cross-correlation function with the transient profile of each pixel. The depth map is then recovered with four-phase, matrix pencil, and IDFT

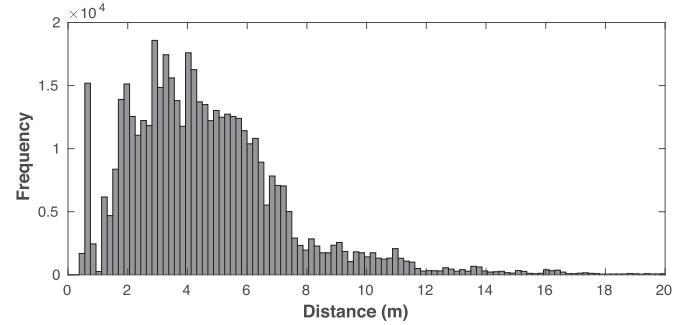


Fig. 16. Depth histogram calculated from the transients of all 25 scenes in the *iToF2dToF* dataset.

algorithms and compared. Pixels with no return due to distances exceeding the range considered in the dataset and pixels for which ray tracing experienced errors were excluded from the evaluation, as in [44]. As can be seen, both the IDFT and matrix pencil methods have better performance than the baseline method, particularly in areas of the image heavily affected by diffusive MPI, such as the top corner, shown in the inset of Fig. 15.

E. Depth Retrieval After Data Refinement

In this section, we quantitatively analyze the effect of using the data refining techniques described in Section IV on the depth retrieval’s accuracy. The proposed techniques are applied to the frequency measurements (Fourier coefficients) X_ℓ , and the modified data is used with the matrix pencil method for depth recovery.

Fig. 16 shows the histogram of distances computed from *iToF2dToF* 3D image dataset provided in [44]. Based on the step size in the time grid in our simulations (1.5 ns, with 201 steps) and the speed of light, a histogram of depth values from the entire dataset of 25 depth images is calculated. Then, the histogram is normalized to the unit ℓ_1 norm to obtain a PDF. The square root of the PDF is used as weights (diagonal elements of W in (20)). Then the compensation matrix “C” is calculated as in (21).

TABLE III
COMPARISON OF DISTANCE RMSE VALUES FOR DIFFERENT COMBINATIONS OF THE SELF-COMPENSATION (SC), CADZOW, AND PDF IMPROVEMENT METHODS

Paths	Matrix Pencil Retrieved Distance RMSE (cm)				
	Raw	SC	PDF	SC+Cadzow	PDF+Cadzow
One	1.62	0.83	0.64	0.85	0.69
Two	1st	2.3	0.95	0.89	0.88
	2nd	15.5	11.5	12.5	11.31
Three	1st	2.1	1.7	1.8	1.2
	2nd	17.5	15.1	14.9	15.1
	3rd	24.8	13.5	14.1	12.5
					13.1

The RMSE values are measured in the case of one, two, and three bounces from the same explained measurement setups from over 100 samples in their recoverable range. Table III shows the RMSE of the recovered distances from the raw data (the ones used to retrieve images of Figs. 9, 11, and 14) and compares them with the RMSE after applying data refinement. We assess the impact of refinement techniques on the overall reconstruction performance through a series of ablation experiments. Here, only the matrix pencil’s results are shown as our selected approach. However, similar improvements are found for the IDFT method.

Among the proposed improvement methods, SC dramatically reduces errors. The SC modified by PDF also improves the results in all cases. It is worth mentioning that the PDF is only measured and applied in the first half of the range (18.75 m). The reason is that since most of the data set depths were in short distances, the PDF could not be reliably estimated for long ranges. The weighting heavily penalizes errors in the short range, where most of the probability mass of the PDF is allocated, while being less sensitive to errors in the long range. The combination of SC and Cadzow denoising showed the greatest improvements in most cases among the experiments. Also, applying PDF+Cadzow methods improved the errors, closely approaching those obtained for SC+Cadzow. However, despite the error reduction, it yields no steady advantage over SC+Cadzow.

Overall, these improvement methods had an outstanding performance in reducing error. Although the recovered images from the matrix pencil were reasonable, and there was no significant error in the depth map, applying these methods allows further improving these results and reach RMSEs below or close to 1 cm in single bounce and MPI cases over the whole depth range.

F. Comparison Against Other Methods

In this section, we compare the performance of our codesigned ToF system to other state-of-the-art techniques dealing with MPI. Table IV shows a comprehensive qualitative comparison encompassing a spectrum of state-of-the-art MPI correction techniques. This evaluation provides insight into the strengths and limitations inherent to each approach. It is noteworthy that while single-frequency techniques suffer from high computational times, multifrequency measurements, despite their merits, encounter challenges linked to frame rate limitations. Even though these methods exhibit potential for mitigating MPI, their practical application may be hampered by constraints

associated with speed. On the other hand, Convolutional Neural Network (CNN)-based techniques introduce a distinct set of requirements. These methods necessitate extensive training procedures to learn the intricacies of MPI correction effectively and often involve subsequent post-processing steps to refine the outcomes. As the table demonstrates, each technique presents a distinctive trade-off between efficacy, computational demands, and practical feasibility.

As a quantitative comparison, we conducted a study to assess the performance of the proposed method in relation to existing data-driven iToF learning models. The compared methods in this study include Phasor [47], SRA [18], Depth2Depth [48], iToF2Depth [50], iToF2iTof [44], and iToF2dTof [44]. For detailed information about the algorithms and the testing process, further elaboration can be found in [44].

We performed simulations using the transient profiles of two real indoor scenes, namely “hot-living” and “kitchen-2,” from the *iToF2dTof* dataset. For each image, the depth errors are arranged in ascending order, and subsequently, they are divided into four percentile groups, specifically 0–75%, 75–85%, 85–95%, and 95–99%. The chosen performance metric is the percentile mean absolute errors (MAE), used in [19], [44], [50]. The MAE is then calculated within each percentile group. This strategy enables a comprehensive assessment of the performance in regions characterized by high SNR and low MPI (0–75% percentile) and regions with low SNR and high MPI (85–99% percentile). For the computation of the percentile MAE on synthetic data, edge pixels containing depth discontinuities, referred to as flying pixels, are masked using Canny edge detection to ensure the reliability of the ground truth depth information as provided in *iToF2dTof* dataset. Also, areas with depths greater than 7.5 m are masked, as in [44]. Furthermore, pixels with the largest 1% error are excluded from consideration to avoid any erroneous effects that might arise from unmasked flying pixels or infinite ray pixels.

Table V presents a quantitative analysis of our design against the mentioned baselines. The data-driven approaches show a capacity for mitigating MPI to a certain extent. On a comprehensive scale, the best of *iToF2dTof* distinguish themselves by surpassing all the models scrutinized in terms of performance. In our study, we conducted four distinct sets of experiments for the selected testing scenes. The results showcased in this table are derived from the unprocessed system outputs employing both the Matrix Pencil and IDFT methods, along with the self-compensation and Cadzow denoising refinement techniques. The PDF method has not been included in the tests to ensure a fair comparison with other methods, as it relies on prior scene knowledge. The depth retrieval process utilizing the IDFT method was executed with $\lambda = 100$, chosen for its computational efficiency [41]. Our proposed models demonstrate remarkable performance, particularly in scenarios with low SNR and high MPI, without requiring additional post-processing steps, a distinctive advantage over certain neural network-based techniques. The method’s effectiveness in these conditions can be attributed to the availability of single-shot multifrequency measurements instead of deep frequency extrapolation. The

TABLE IV
QUALITATIVE COMPARISON OF MPI CORRECTION METHODS

Work	MPI Type	Input	Solution Type	Computation	Constraints
Heide [22]	Diffuse	Multiple frequencies	Iterative	Low	Multiframe multifrequency
Kadambi [45]	K-sparse	Single frequency	Iterative	High	Custom code
Freedman [18]	K-sparse	Multiple frequencies	Iterative	Low	Multiframe multifrequency
Bhandari [16]	K-sparse	Multiple frequencies	Closed-form	Low	Multiframe multifrequency
Peters [46]	K-sparse	Multiple frequencies	Closed-form	Low	Multiframe multifrequency
Gupta [47]	Diffuse	Multiple frequencies	Closed-form	Low	Multiframe multifrequency
Marco [48]	General	Single depth	CNN	High	GHz-freq sensor and illumination
Wang [49]	General	Single depth	CNN	Medium	Requires GPU, training
Simonetto [21]	General	Multiple depths	CNN	High	Requires GPU, training
Our Work	K-sparse	Single-shot multifrequency	Closed-form	Low	Pixel-domain multiplexing

TABLE V
COMPARISON OF PERCENTILE MAE (MM) WITH THE STATE-OF-THE-ART METHODS

Synthetic Test Set Percentile MAE (mm)				
Model	0-75%	75-85%	85-95%	95-99%
Phasor (No Noise)	9.53	29.58	46.37	94.79
SRA (No Noise)	13.97	41.75	62.70	126.51
Depth2Depth Baseline	13.40	42.21	69.07	136.96
iToF2iTof Baseline	10.56	31.93	49.71	101.82
iToF2Depth Baseline	7.49	21.86	34.99	88.03
Best of iToF2dTof	7.19	20.40	32.17	71.56
Matrix Pencil	9.92	40.94	54.03	75.98
Matrix Pencil (SC+Cadzow)	7.87	37.90	50.99	70.14
IDFT	11.76	33.35	47.63	79.17
IDFT (SC+Cadzow)	9.72	31.22	45.97	76.90

proposed approach also outperforms the non-learning alternatives regarding acquisition time and procedure. Additionally, it is worth highlighting that the MAE results from the other methods are obtained based on two input frequency measurements at 20 MHz and 100 MHz, whereas our results correspond to the 4–64 MHz range. Our proposed approach exhibits the potential to achieve superior results compared to other methodologies when operating with increased bandwidth.

G. Discussion and Limitations

The proposed system offers a novel approach to imaging by employing pulsed illumination and spatially multiplexing multiple frequencies across different pixels. Since the proposed approach requires a single image and uses the algorithms with the lowest possible computational cost, we anticipate their potential integration into forthcoming ToF imaging systems. Even though this strategy introduces significant advantages, it also presents noteworthy challenges at the circuit design level. Including extra connections required to route the 16 channels within individual pixels and throughout the system demands meticulous consideration.

Regarding MPI resolution, the algorithm's performance is intricately linked to the quality of the generated cross-correlation function of the system. This quality is inherently contingent upon the SNR of the designed hardware. Ensuring a high SNR becomes paramount to achieving superior MPI resolution. The proposed harmonic canceling technique is pivotal in mitigating the wiggling effect that can arise from measurements, further enhancing the system's performance.

The operational range of the sensor depends on the base frequency and illumination power of the system. Increasing the base frequency significantly enhances the precision of measurements, albeit potentially constraining the operational range to shorter distances. On the other hand, higher illumination power enables the sensor to capture signals from objects at greater distances.

The most notable limitation of our system lies in the trade-off between spatial and temporal resolution. This poses certain constraints, particularly in applications where high spatial resolution is paramount. It's worth highlighting that leveraging advanced fabrication technologies, such as backside illumination and smaller process technology, can compensate for this limitation.

VI. CONCLUSION

In this work, we have presented a novel joint hardware/software codesign for effectively and efficiently solving the MPI problem in iToF depth imaging using a single-shot on-chip multifrequency demodulation concept. The proposed system can capture multiple Fourier coefficients of the scene response function in a single shot with minimal harmonic distortion. This is combined with robust depth retrieval algorithms to solve the inverse problem parametrically in a closed form. The Fourier sensing functions are optimal for sensing Dirac delta functions, so the proposed sensor is optimal for sensing scenes with opaque objects and reflective multi-path.

The data processing method is general and holds for any number of reflections. The proposed approach has been validated using ground truth transient images from the *iToF2dTof* database for multiple bounces, showing great accuracy in the depth estimation. The effect of multifrequency measurements is apparent at low SNR and, of course, when considering more than a single path. Moreover, refinement techniques have been proposed that exploit otherwise ignored a priori information to boost the depth accuracy. The proposed data techniques increase the depth accuracy in single and multiple path cases.

Future work includes coupling our computational ToF hardware to emerging deep frequency extrapolation methods [44] for estimating transient profiles in a single shot.

REFERENCES

- [1] F. Piron, D. Morrison, M. R. Yuce, and J.-M. Redouté, "A review of single-photon avalanche diode time-of-flight imaging sensor arrays," *IEEE Sensors J.*, vol. 21, no. 11, pp. 12654–12666, Jun. 2021.

- [2] M. Heredia Conde, T. Kerstein, B. Buxbaum, and O. Loffeld, "Near-infrared, depth, material: Towards a trimodal time-of-flight camera," *IEEE Sensors J.*, vol. 22, no. 12, pp. 11271–11279, Jun. 2022.
- [3] M. Heredia Conde, *Compressive Sensing for the Photonic Mixer Device*. Berlin, Germany: Springer, 2017.
- [4] A. Bhandari and R. Raskar, "Signal processing for time-of-flight imaging sensors: An introduction to inverse problems in computational 3-D imaging," *IEEE Signal Process. Mag.*, vol. 33, no. 5, pp. 45–58, Sep. 2016.
- [5] R. Lange and P. Seitz, "Solid-state time-of-flight range camera," *IEEE J. Quantum Electron.*, vol. 37, no. 3, pp. 390–397, Mar. 2001.
- [6] C. Bamji et al., "A review of indirect time-of-flight technologies," *IEEE Trans. Electron Devices*, vol. 69, no. 6, pp. 2779–2793, Jun. 2022.
- [7] P. E. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *Proc. ACM SIGGRAPH Classes*, 2008, pp. 1–10.
- [8] R. Raskar, A. Agrawal, and J. Tumblin, "Coded exposure photography: Motion deblurring using fluttered shutter," in *Proc. ACM SIGGRAPH*, 2006, pp. 795–804.
- [9] M. F. Duarte et al., "Single-pixel imaging via compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 83–91, Mar. 2008.
- [10] A. Bhandari, M. Heredia Conde, and O. Loffeld, "One-bit time-resolved imaging," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 7, pp. 1630–1641, Jul. 2020.
- [11] T.-H. Hsu, T. Liao, N.-A. Lee, and C. -C. Hsieh, "A CMOS time-of-flight depth image sensor with in-pixel background light cancellation and phase shifting readout technique," *IEEE J. Solid-State Circuits*, vol. 53, no. 10, pp. 2898–2905, Oct. 2018.
- [12] C. Niclass et al., "A 100-m Range 10-Frame/s 340 96-Pixel Time-of-Flight Depth Sensor in 0.18- μ m CMOS," *IEEE J. Solid-State Circuits*, vol. 48, no. 2, pp. 559–572, Feb. 2012.
- [13] D. Kim et al., "Indirect time-of-flight CMOS image sensor with on-chip background light cancelling and pseudo-four-tap/two-tap hybrid imaging for motion artifact suppression," *IEEE J. Solid-State Circuits*, vol. 55, no. 11, pp. 2849–2865, Nov. 2020.
- [14] D. Stoppa, N. Massari, L. Pancheri, M. Malfatti, M. Perenzoni, and L. Gonzo, "A range image sensor based on 10- μ m lock-in pixels in 0.18- μ m CMOS imaging technology," *IEEE J. Solid-State Circuits*, vol. 46, no. 1, pp. 248–258, Jan. 2011.
- [15] R. Whyte, L. Streeter, M. J. Cree, and A. A. Dorrington, "Review of methods for resolving multi-path interference in time-of-flight range cameras," in *Proc. IEEE SENSORS*, 2014, pp. 629–632.
- [16] A. Bhandari et al., "Resolving multipath interference in time-of-flight imaging via modulation frequency diversity and sparse regularization," *Opt. Lett.*, vol. 39, no. 6, pp. 1705–1708, 2014.
- [17] M. Heredia Conde, T. Kerstein, B. Buxbaum, and O. Loffeld, "Fast multipath estimation for PMD sensors," in *Proc. 5th Int. Workshop Compressed Sens. Theory Appl. Radar, Sonar, Remote Sens.*, 2018.
- [18] D. Freedman, Y. Smolin, E. Krupka, I. Leichter, and M. Schmidt, "SRA: Fast removal of general multipath for ToF sensors," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 234–249.
- [19] A. Adam, C. Dann, O. Yair, S. Mazor, and S. Nowozin, "Bayesian time-of-flight for realtime shape, illumination and albedo," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 5, pp. 851–864, May 2017.
- [20] G. Agresti and P. Zanuttigh, "Deep learning for multi-path error removal in ToF sensors," in *Proc. Eur. Conf. Comput. Vis. Workshops*, 2018, pp. 410–426.
- [21] A. Simonetto, G. Agresti, P. Zanuttigh, and H. Schäfer, "Lightweight deep learning architecture for MPI correction and transient reconstruction," *IEEE Trans. Comput. Imag.*, vol. 8, pp. 721–732, 2022.
- [22] F. Heide, M. B. Hullin, J. Gregson, and W. Heidrich, "Low-budget transient imaging using photonic mixer devices," *ACM Trans. Graph.*, vol. 32, no. 4, pp. 1–10, 2013.
- [23] J. Lin, Y. Liu, J. Suo, and Q. Dai, "Frequency-domain transient imaging," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 5, pp. 937–950, May 2017.
- [24] A. Jarabo, B. Masia, J. Marco, and D. Gutierrez, "Recent advances in transient imaging: A computer graphics and vision perspective," *Vis. Inform.*, vol. 1, no. 1, pp. 65–79, 2017.
- [25] H. Wang, H. Qiao, J. Lin, R. Wu, Y. Liu, and Q. Dai, "Model study of transient imaging with multi-frequency time-of-flight sensors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 10, pp. 3523–3539, Oct. 2021.
- [26] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
- [27] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proc. Nat. Acad. Sci.*, vol. 106, no. 45, pp. 18914–18919, 2009.
- [28] K. Kagawa et al., "Multi-tap macro-pixel based compressive ultra-high-speed CMOS image sensor," in *Proc. Int. Image Sensor Workshop*, 2019, pp. 270–273.
- [29] C. S. Bamji et al., "Impixel 65 nm BSI 320 MHz demodulated ToF image sensor with 3 μ m global shutter pixels and analog binning," in *Proc. IEEE Int. Solid-State Circuits Conf.*, 2018, pp. 94–96.
- [30] M. Horio et al., "Resolving multi-path interference in compressive time-of-flight depth imaging with a multi-tap macro-pixel computational CMOS image sensor," *Sensors*, vol. 22, no. 7, 2022, Art. no. 2442.
- [31] P. F. Shahandashti, P. López, V. M. Brea, D. García-Lesta, and M. Heredia Conde, "Proposal of a single-shot multi-frame multi-frequency CMOS ToF sensor," in *Proc. IEEE 28th Int. Conf. Electron., Circuits, Syst.*, 2021, pp. 1–4.
- [32] R. M. Conroy, A. A. Dorrington, and A. D. Payne, "A power saving modulation technique for time-of-flight range imaging sensors," in *Proc. SPIE. Videometrics, Range Imag., Appl.*, vol. 8085, 2011, Art. no. 808506.
- [33] J. Ohta, *Smart CMOS Image Sensors and Applications*. Boca Raton, FL, USA: CRC Press, 2017.
- [34] B. Buxbaum, R. Schwarte, and T. Ringbeck, "PMD-PLL: Receiver structure for incoherent communication and ranging systems," in *Proc. Int. Conf. SPIE Opt. Wireless Commun. II*, 1999, pp. 116–127.
- [35] S. Kawahito et al., "CMOS lock-in pixel image sensors with lateral electric field control for time-resolved imaging," in *Proc. Int. Image Sensor Workshop*, Snowbird, UT, USA, 2013.
- [36] Y. Kato et al., "320 \times 240 back-illuminated 10- μ m CAPD pixels for high-speed modulation time-of-flight CMOS image sensor," *IEEE J. Solid-State Circuits*, vol. 53, no. 4, pp. 1071–1078, Apr. 2018.
- [37] Y. Ebiko et al., "Low power consumption and high resolution 1280 \times 960 gate assisted photonic demodulator pixel for indirect time of flight," in *Proc. IEEE Int. Electron Devices Meeting*, 2020, pp. 33.1.1–33.1.4.
- [38] P. F. Shahandashti, P. López, V. Brea, D. García-Lesta, and M. Heredia Conde, "A 2-tap macro-pixel-based indirect ToF CMOS image sensor for multi-frequency demodulation," in *Proc. IEEE Int. Symp. Circuits Syst.*, 2022, pp. 1430–1434.
- [39] S. Foix, G. Alenya, and C. Torras, "Lock-in time-of-flight (ToF) cameras: A survey," *IEEE Sensors J.*, vol. 11, no. 9, pp. 1917–1926, Sep. 2011.
- [40] Y. Hua and T. K. Sarkar, "Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 5, pp. 814–824, May 1990.
- [41] P. F. Shahandashti, P. López, V. Brea, D. García-Lesta, and M. Heredia Conde, "Fast time-domain super-resolution for single-shot multi-path ToF imaging," in *Proc. 29th IEEE Int. Conf. Electron., Circuits, Syst.*, 2022, pp. 1–4.
- [42] T. Blu, P.-L. Dragotti, M. Vetterli, P. Marziliano, and L. Coulot, "Sparse sampling of signal innovations," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 31–40, Mar. 2008.
- [43] M. Heredia Conde, K. Hartmann, and O. Loffeld, "Subpixel spatial response of PMD pixels," in *Proc. IEEE Int. Conf. Imag. Syst. Techn.*, 2014, pp. 297–302.
- [44] F. Gutierrez-Barragan, H. Chen, M. Gupta, A. Velten, and J. Gu, "iTof2dToF: A robust and flexible representation for data-driven time-of-flight imaging," *IEEE Trans. Comput. Imag.*, vol. 7, pp. 1205–1214, 2021.
- [45] A. Kadambi et al., "Coded time of flight cameras: Sparse deconvolution to address multipath interference and recover time profiles," *ACM Trans. Graph.*, vol. 32, no. 6, pp. 1–10, 2013.
- [46] C. Peters, J. Klein, M. B. Hullin, and R. Klein, "Solving trigonometric moment problems for fast transient imaging," *ACM Trans. Graph.*, vol. 34, no. 6, pp. 1–11, 2015.
- [47] M. Gupta, S. K. Nayar, M. B. Hullin, and J. Martin, "Phasor imaging: A generalization of correlation-based time-of-flight imaging," *ACM Trans. Graph.*, vol. 34, no. 5, pp. 1–18, 2015.
- [48] J. Marco et al., "Deepoft: Off-the-shelf real-time correction of multipath interference in time-of-flight imaging," *ACM Trans. Graph.*, vol. 36, no. 6, pp. 1–12, 2017.
- [49] X. Wang, W. Zhou, and Y. Jia, "Attention GAN for multipath error removal from ToF sensors," *IEEE Sensors J.*, vol. 22, no. 20, pp. 19713–19721, Oct. 2022.
- [50] S. Su, F. Heide, G. Wetzstein, and W. Heidrich, "Deep end-to-end time-of-flight imaging," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 6383–6392.



Peyman Fayyaz Shahandashti (Graduate Student Member, IEEE) received the M.Sc. degree in electrical engineering, and electronics from the Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran, in 2019. He is currently a Ph.D. Student with the Centro Singular de Investigación en Tecnologías Intelixentes (CITIUS), the University of Santiago de Compostela, Santiago, Spain, under the Marie Curie research fellowship of Horizon 2020 (the MENELAOS^{NT} Project). His research interests include the Time-of-Flight imaging systems, analog/mixed-signal integrated circuits design, and sensors.



Daniel García-Lesta received the Physics degree from the the University of Santiago de Compostela (USC), Santiago, Spain and the master's degree in electronics Systems for Information and Communication from the National Distance Education University, Madrid, Spain, in 2014 and 2016, respectively. He received the Ph.D. degree from USC, in 2021. He is currently a Postdoc Researcher with Centro Singular de Investigación en Tecnologías Intelixentes, Santiago, Spain. His research focuses on the design of mixed-signal circuits for efficient computation, especially in video processing applications.



Paula López (Senior Member, IEEE) received the Ph.D. degree in physics from the University of Santiago de Compostela, Santiago, Spain, in 2003. She is currently an Associate Professor with the Centro Singular de Investigación en Tecnologías Intelixentes, the University of Santiago de Compostela. Later, she held a 2-year postdoctoral position with the Fraunhofer Institute for Integrated Circuits, Erlangen, Germany. Her research interests include the design of mixed-signal integrated circuits, particularly CMOS imagers, and the physical modeling of electronic devices. Her current research interests include the design of smart image sensors and low-power circuits for energy harvesting and IoT applications.



Miguel Heredia Conde (Member, IEEE) received the Dr. Eng. degree in the field of sensor signal processing and Habilitation degree, both from the University of Siegen, Siegen, Germany, in 2016 and 2022, respectively. Since 2016, he has been the Leader with the Research Group Compressive Sensing for the Photonic Mixer Device and since 2020, also the General Manager with the H2020-MSCA-ITNMENELAOS^{NT}. His research interests include Time-of-Flight imaging systems, such as those based on the Photonic Mixer Device, compressive sensing, computational imaging, and unconventional sensing. Miguel Heredia was one of the recipients of the 2006 Academic Excellence Prices, awarded by the Government of Galicia, Spain. In 2017, he was the recipient of the University of Siegen Prize for International Young Academics, for his excellent performance in his doctoral studies. He is a member of the ITG/VDE and the IEEE/SPS.



Víctor Manuel Brea received the Ph.D. degree in physics in 2003. He is currently an Associate Professor with Centro Singular de Investigación en Tecnologías da Información (CITIUS), University of Santiago de Compostela, Santiago, Spain. His research interests include the design of efficient architectures and CMOS solutions for computer vision, especially in early vision, as well as micro-energy harvesting.