Depth Dynamics via One-Bit Frequency Probing in Embedded Direct Time-of-Flight Sensing

Seth Lindgren *Student Member, IEEE*, Benjamin R. Johnson, and Lucas J. Koerner, *Member, IEEE*

Abstract—Time-of-flight (ToF) sensors with single-photon avalanche diodes (SPADs) estimate depth by accumulating a histogram of photon return times, which discards the timing information required to measure depth dynamics, such as vibrations or transient motions. We introduce a method that transforms a direct ToF sensor into a depth frequency analyzer capable of measuring high-frequency motion and transient events using only lightweight, on-sensor computations. By replacing conventional discrete Fourier transforms (DFTs) with one-bit probing sinusoids generated via oversampled sigma-delta modulation, we enable in-pixel frequency analysis without multipliers or floating-point operations. We extend the lightweight analysis of depth dynamics to Haar wavelets for time-localized detection of brief, non-repetitive depth changes. We validate our approach through simulation and hardware experiments, showing that it achieves noise performance approaching that of full-resolution DFTs, detects sub-millimeter motions above 6 kHz, and localizes millisecond-scale transients. Using a laboratory ToF setup, we demonstrate applications in oscillatory motion analysis and depth edge detection. This work has the potential to enable a new class of compact, motion-aware ToF sensors for embedded deployment in industrial predictive maintenance, structural health monitoring, robotic perception, and dynamic scene understanding.

Index Terms—Computational Photography, Depth sensing, SPAD LiDAR, Spectral analysis, Signal processing in embedded systems

◆

1 INTRODUCTION

TIME-of-flight (ToF) depth imaging systems that combine a pulsed laser with a sensor that counts and time-stamps individual photon arrivals [1], [2] are becoming increasingly precise and widely deployed, including in consumer devices such as smartphones [3]. These sensors measure the travel time of photons to estimate the distance of objects in a scene. Traditionally, ToF measurements are applied to static depth imaging. Yet, objects often move or vibrate, and capturing these dynamics can reveal properties relevant to applications such as industrial predictive maintenance, remote respiration monitoring, and acoustics. Beyond these domains, dynamic depth measurements present opportunities for further exploration in ToF computer vision — for example, using camera motion to infer surface topology or leveraging velocity cues for motion segmentation. Incorporating dynamics is a powerful extension, especially given the small size, low cost, and increasing integration of ToF sensors into everyday devices, which may enable broad adoption across diverse applications.

Although ToF sensors offer picosecond-scale timing precision, practical constraints limit their ability to capture temporal dynamics. The first bottleneck is the output data interface – reading out every individual photon timestamp from a large-format array would require data rates beyond 100 Gbits/second Ethernet [4], which is impractical for power-constrained devices. The second bottleneck is pixel size, which limits the area for storing and processing timestamps on-chip. Current ToF sensors address these challenges by accumulating arrival times into per-pixel histograms [1]. After many laser pulses, the histogram data is transferred and processed to determine the average depth. This histogram compression reduces output bandwidth and fits within the area of a pixel, but discards the association of the photon travel time with the laser cycle which prevents the analysis of how depth changes with time. A method that preserves the *dynamics* of photon arrival times while remaining practical in terms of output bandwidth and in-pixel compute is needed.

To address these challenges, we develop new on-sensor storage and processing methods that preserve the temporal dynamics of photon arrival times to create a *depth frequency analyzer*. Our frequency-domain analysis extracts the amplitude of motion at each frequency to reveal vibrations relevant to structural health monitoring, acoustics, or the rotational speed of moving objects such as drone rotors. To capture non-repetitive or transient motions, we complement Fourier analysis with Haar wavelet analysis, which localizes motion in time and estimates velocity.

Our approach respects the limitations of in-pixel computation and storage, yet approaches the resolution and noise of a conventional discrete Fourier transform that would impractically require readout of all timestamps. For in-pixel frequency analysis, we leverage the significant difference between the ToF laser repetition rate and scene motion frequencies to perform in-pixel frequency probing with one-bit oversampled sinusoids. These binary basis functions eliminate the need for multiplications or sinusoidal evaluations, making this approach well-suited for future large-scale ToF arrays and energy-constrained edge-computing imaging systems.

While high-end systems such as laser Doppler vibrometers [5] or high-speed cameras [6] can capture fine motion dynamics they are typically bulky, power-hungry, and unsuitable for integration into embedded or portable systems. Conventional camera-based methods are constrained by frame rate, require significant computation, and depend on ambient texture or lighting. In contrast, our approach

---

- *S.L, B.R.J., and L.J.K are with the Department of Electrical and Computer Engineering, University of St. Thomas, St. Paul, MN, 55105.*
  e-mail: koerner.lucas@stthomas.edu.

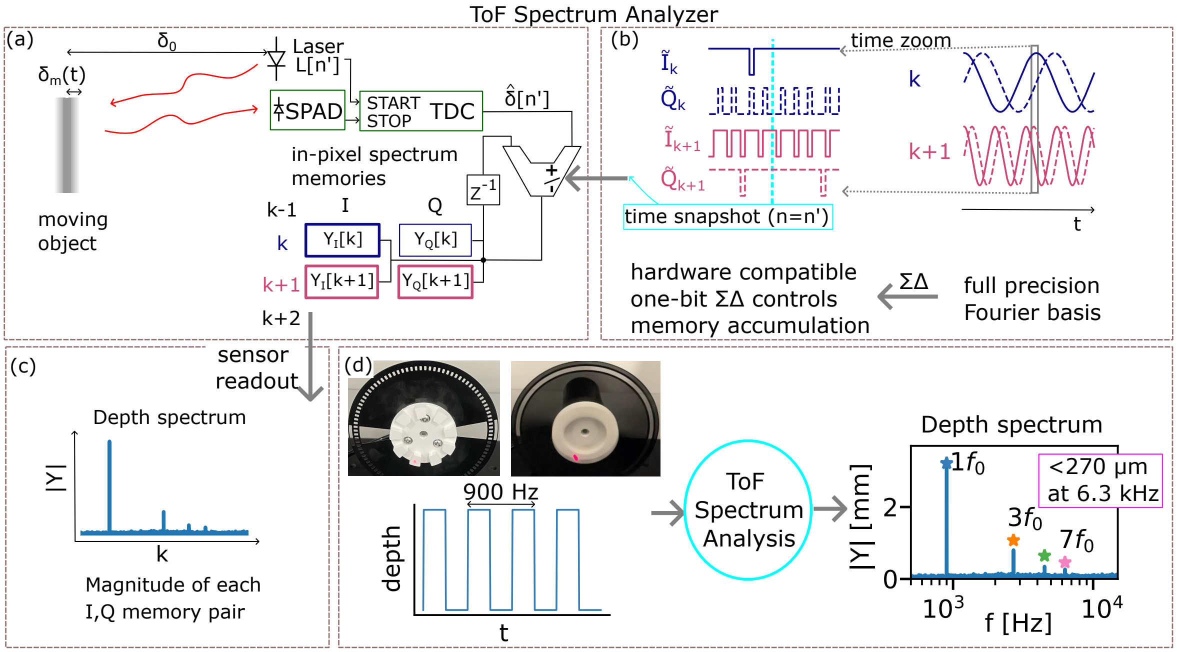

Fig. 1. Overview of a time-of-flight sensor designed to operate as a *frequency analyzer of object depth* while respecting hardware limitations. (a) Object depth $\delta(t)$ varies over time as $\delta_0 + \delta_m(t)$. The ToF system pulses a laser at discrete time intervals n and the returning photons are detected by a single-photon avalanche diode (SPAD), with the photon return time digitized by a time-to-digital converter (TDC) to estimate object depth, $\hat{\delta}[n']$. Each depth sample is accumulated into spectrum memory bins $Y_{I,Q}[k]$ based on one-bit (± 1) probing waveforms $\tilde{I}_k[n]$ and $\tilde{Q}_k[n]$ generated by a sigma-delta ($\Sigma\Delta$) modulator. (b) Full-precision Fourier basis functions for frequencies k and $k+1$ are shown alongside a short segment of their one-bit $\Sigma\Delta$ representations (indicated by a vertical gray box). At a particular snapshot in time ($n = n'$; ╍ vertical line) the sign of the one-bit waveform determines whether the current depth sample is added to or subtracted from the corresponding $Y_{I,Q}[k]$ memory elements (bold and thin outlines, respectively). In this example, $\tilde{I}_k = 1$ so that the depth measurement adds to $Y_I[k]$ and $\tilde{Q}_k = -1$, which causes the depth measurement to be subtracted from $Y_Q[k]$. Simultaneously, in the $k+1$ frequency bin, the depth measurement is added into both $Y_I[k+1]$ and $Y_Q[k+1]$. (c) After sensor readout, a discrete Fourier transform of the object depth is calculated with each I/Q pair determining the amplitude at a discrete frequency. (d) Experimental results using a pulsed-laser ToF system measure a rapid depth square wave. Frequency-domain analysis recovers the 7th harmonic ($7f_0 = 6.3\,\mathrm{kHz}$) with an amplitude of $270\,\mu\mathrm{m}$, which is not visible in a super-slo-mo video recorded at $240\,\mathrm{fps}$.

enables high-temporal-resolution motion sensing, albeit with less amplitude resolution, using direct depth measurements and minimal on-sensor resources making it complementary to camera-based systems. Table 1 summarizes key differences between conventional cameras and our ToF sensor for measuring dynamics. Sec. 8 explores future applications enabled by the miniaturization and spatial coverage of our embedded ToF frequency analyzer, inaccessible to conventional LDVs.

This manuscript introduces methods for ToF frequency analysis and describes the proposed hardware compatible one-bit probing method in Sec. 3. Challenges unique to ToF frequency analysis including timing jitter, background photons, and missed pulses are evaluated with methods for mitigation in Sec. 5. Simulations (Sec. 6) evaluate the performance dependence on measurement parameters and validate an analysis of the noise limits. Laboratory experiments of Sec. 7 detect dynamic depth features at frequencies beyond 5 kHz, align with analytically established noise limits, and localize depth transient signals using wavelet capture. In total, this work proposes and evaluates new designs that transform a direct ToF sensor into a *depth frequency analyzer* while maintaining compatibility with future large-scale arrays and embedded platforms.

Scope and Limitations: This work demonstrates frequency analysis of depth dynamics using synthetic and laboratory time-stamp data from a single-pixel direct ToF sensor. While the approach is designed to be compatible with in-pixel implementation, it does not address full-array scalability or real-time operation. Although the impact of background photons is analyzed, the current design is not expected to operate robustly under high ambient light, such as outdoors; future work may address this limitation through timing window optimization or adaptive filtering.

2 RELATED WORK

Single-photon frequency extraction: Wei et al. [7] demonstrate frequency extraction of photon flux from DC to 31 GHz using single-photon avalanche diodes (SPADs) in a passive setup, unsynchronized with illuminators. They achieve this by transforming photon arrival timestamps through a Fourier probing basis, identifying spectral content, and suppressing bands without significant signal, showcasing the potential of photon sensing at picosecond-scale temporal resolution for GHz videography. Frequency domain analysis of single-photon ToF data has also been proposed to extract the Doppler shift for velocity estimates [8], [9]. Our work emphasizes practical sensor designs for synchronized ToF frequency analysis, focusing on spectral analysis of depth variations from object motion.

Compressive ToF sensing and resource-constrained signal processing: Compressive acquisition strategies for static depth measurements by direct ToF sensors aim to reduce in-pixel memory usage, limit output data bandwidth, and improve robustness to background light. Compressive histograms that accumulate on the fly use linear matrix

TABLE 1
Comparison of conventional camera-based motion sensing and the proposed dynamic-aware ToF sensing.

Specification	Camera-Based Motion Sensing	Dynamic-Aware ToF (Proposed Method)
Temporal resolution	Frame-limited (30-240 fps)	100+ kHz probing rate
Motion estimation method	Indirect or learned	Direct
Power consumption	Moderate to high	Low (on-sensor; integer-only logic)
Lighting / texture	Required	Not required

encodings that are hand-crafted [4] or learned [10] to exploit the limited bandwidth of the laser pulse and compress the depth data without considerable information loss. A sketching framework extracts summary statistics from the characteristic function of the photon returns to minimize in-pixel memory [11]. Fourier domain histograms have been developed for low-latency non-line-of-sight imaging [12]. Our method uses one-bit signal processing to enable low-power and low-complexity computation. Digital filters operating on oversampled single-bit signals have demonstrated power and area savings as compared to operating on a full resolution signal [13], [14].

Non-contact optical sensing of motion: High-speed videography combined with laser speckles or physical markers enables 3D point tracking and digital image correlation (DIC) [15]. Marker-less methods have been used for modal analysis [16], optical microphones [6], or to estimate material properties [17]. These video-based methods often leverage motion magnification [18] or phase-based detection [19] to resolve subtle sub-pixel motions. However, these camera-based methods are limited by frame-rate, are only sensitive to in-plane motion, and require texture [16]. A dual-camera system expanded the frequency range to 63 kHz by exploiting the fast line times of rolling shutter exposures [20]. Outside of cameras, a triangulation-based LiDAR system demonstrated audio extraction to $1.25\,\mathrm{kHz}$ [21]. Recently, the iPhone's direct ToF sensor was used for nodal vibration analysis and sensed mm amplitude oscillations up to $7.5\,\mathrm{Hz}$, constrained by the frame rate [22]. In contrast, our proposed methods perform in-pixel processing to decouple measurement frequency from the sensor frame rate.

Interferometric and phase-based vibrometry: Laser Doppler vibrometers are sensitive detectors of object velocity with applications in structural health and modal analysis [5] and are typically custom and expensive. If spatial resolution is required the laser spot must be scanned which slows acquisition times. Here, we investigate the feasibility of out-of-plane vibration measurements without scanning using inexpensive and lightweight depth sensors that are appropriate for wide deployment in smartphones and robots.

3 METHODS FOR FREQUENCY ANALYSIS

3.1 Problem Definition

This work investigates how dynamics of object depth can be extracted by pulsed-laser time-of-flight sensors. The object depth may be decomposed into fixed and time varying components as: $\delta(t) = \delta_0 + \delta_m(t)$ (Fig. 1(a)). The variation may be from oscillatory vibrations, motion along the optical axis, or camera motion that varies the topology of the object surface within the field of view of the sensor. The pulsed laser transmitter probes the object depth at discrete time intervals with an inter-pulse interval of T_s which is set by capabilities of the laser driver, eye-safety, and by the desired maximum depth of the measurement. The laser probe can be pulsed (on) or gated (off) at each discrete time interval and is defined as a bit stream as: $L[n] \in \{0, 1\}$, with time discretized as $t = nT_s$. We assume that continuous variations in the laser intensity are not practical.

The object displacement during photon travel is reasonably approximated to be insignificant. Because of this assumption, the system measures object position $\delta(t = nT_s)$ when the laser is pulsed at cycle n. When the SPAD receiver detects a returned photon the arrival time is digitized by a time-to-digital-converter (TDC). This estimate of the object depth, $\hat{\delta}[n]/(2c)$, where c is the speed of light, is spread by the instrument timing jitter with variance σ_I^2 and quantized. Readout of each time-stamp is impractical due to output bandwidth limitations. So strategies must be developed to store each TDC result into in-pixel memory slot(s) that preserve the spectral content (frequency and amplitude) of the depth measurement. To do so, we consider a (two-dimensional) array of memory elements $Y_m[k]$ that store the depth measurements. The memory has M slots indexed by m and each slot has a set of frequency bins indexed by k. The general update strategy is

$$Y_m[g(\hat{\delta}; k, m, n)] \leftarrow Y_m[g(\hat{\delta}; k, m, n)] + h(\hat{\delta}; k, m, n).$$

where $h(\hat{\delta}; k, m, n)$ modifies the measured depth value and $g(\hat{\delta}; k, m, n)$ determines the set of memory slots and frequency bins that are updated. In practice, the memory $Y_m[k]$ could be represented as a single extended histogram. However, the concept of a 2D memory organization is retained here for clarity.

3.2 Discrete Fourier Transform

The baseline measurement of the frequency content in the object depth from N discrete measurements (or laser pulses) is the discrete Fourier transform (DFT), defined as:

$$Y[k] = \frac{2}{N} \sum_{n=0}^{N-1} \hat{\delta}[n] e^{-j\frac{2\pi}{N}kn}, \quad k = 0, 1, \ldots, N-1, \qquad (1)$$

where the transform frequency $f_k = \frac{k}{NT_s} = \frac{kf_s}{N}$ with $f_s = 1/T_s$. For this calculation, multiplying by the frequency basis $e^{-j\frac{2\pi}{N}kn}$ is $h(\delta; k, m, n)$. The DFT result is complex with an amplitude of $|Y[k]|$ and phase of $\angle Y[k]$ for the frequency component k. To store both real and imaginary parts each measured frequency requires two distinct on-chip storage elements for a total of $2N$. The calculation of each frequency bin is independent of other bins so a design may skip particular bins if, for example, the hardware is constrained by memory or if a limited range of frequencies are of interest. For on-chip streaming DFT calculations the record length, N, is predetermined, and the update policy following each laser pulse, n, is

$$Y_I[k] \leftarrow Y_I[k] + \hat{\delta}[n]\cos\left(\frac{2\pi}{N}kn\right) \tag{2}$$

$$Y_Q[k] \leftarrow Y_Q[k] + \hat{\delta}[n]\sin\left(\frac{2\pi}{N}kn\right) \tag{3}$$

where $m = 0$, $m = 1$ are renamed to I for in-phase and Q for quadrature, respectively, to follow the conventions of spectrum analysis (see Fig. 1(a)). Next, we consider capture schemes that respect hardware limitations yet allow for the measurement of the amplitude at specified frequencies k. The performance of the proposed capture schemes will be evaluated in relation to this DFT baseline.

3.3 Oversampled Modulation

The DFT algorithm of (2) and (3) requires continuous calculation or lookup of sines/cosines and full-precision multiplications which may be too complex for in-pixel computation. Instead, a lighter weight alternative is pursued that replaces the sine/cosine calculation and multiplication with single bit modulation that controls accumulation into memory location(s) and/or laser pulsing. The key insight is that the frequencies of object motion ($\sim$1 kHz) are typically considerably less than the frequency of laser pulsing ($\sim$10 MHz) so that probing sinusoids can be generated by an oversampling one-bit modulator such as a sigma-delta ($\Sigma\Delta$) modulator[1]. One-bit modulated signals replace the full-precision complex multiplications of the DFT with simple switching operations to increase the feasibility of on-sensor calculation.

Sigma-delta modulators are defined by the oversampling ratio (OSR), modulator order and topology, and the bit depth of the quantizer [23]. The oversampling ratio is set by the ratio of the sampling frequency (f_s) and the frequency of the generated baseband waveform (f_0) as $\mathrm{OSR} = 2^d = \frac{f_s}{2f_0}$. A greater OSR reduces the quantization noise in the baseband by shifting it to higher frequencies. The SNR of a two-level modulator in decibels can be approximated based on the power-of-2 OSR (d) and the modulator order (Θ) as:

$$\mathrm{SNR_{dB}} \approx 7.78 + 10\log_{10}(2\Theta+1) - 9.94\Theta + 3.01(2\Theta+1)d \tag{4}$$

with the effective number of bits (ENOB, see Fig. 2(d)) calculated as $\log_2(\mathrm{SNR})$ [23].

Fig. 2 demonstrates a $\Sigma\Delta$ modulator with a 50 kHz input signal and a sampling rate of $f_s = 10\,\mathrm{MHz}$ to align with typical laser pulsing rates of a ToF sensor. The Fourier transform of Fig. 2(c) shows that the modulator shapes quantization noise out of the signal band and to high frequencies. The high frequency quantization noise requires a signal pre-filter to exploit the noise shaping of the modulator to be discussed in Sec. 5.3.

The $\Delta\Sigma$ modulator is a causal system that generates an oversampled, lower resolution representation of the input signal in real-time. In this application, real-time operation is not required, and a non-causal noise-shaping approach could be theoretically used. However, we chose to leverage the highly developed technology of causal $\Sigma\Delta$ modulators.

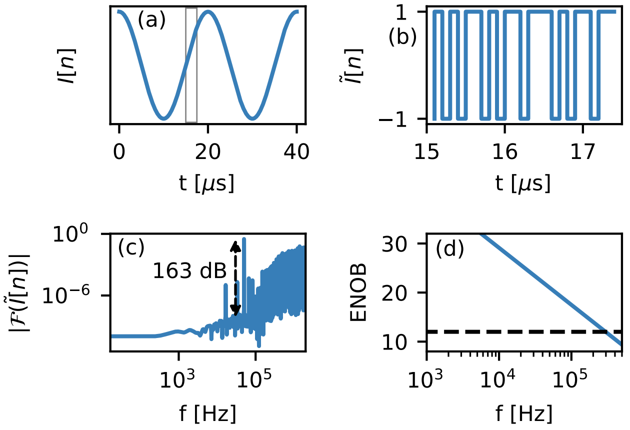

Fig. 2. (a) A full-precision in-phase (I; cos) waveform for probing at 50 kHz. The vertical rectangle shows the time-span of the one-bit modulated representation. (b) $\tilde{I}[n]$ is the one-bit $\Sigma\Delta$ representation of the full-precision waveform generated by a 3rd order modulator with a sampling rate of $f_s = 10\,\mathrm{MHz}$. (c) The discrete Fourier transform (DFT) of the one-bit modulated waveform of (b) with a duration of 33 ms. The DFT shows a signal-to-noise ratio of 164 dB. (d) The effective number of bits (ENOB) calculated using (4) remains above 12 bits (dashed black line) for signal frequencies less than 297 kHz.

3.3.1 *Two-level I/Q Modulation*

For each frequency, $f_k = \frac{kf_s}{N}$, to be measured, a one-bit representation, ± 1, of the continuous cosine and sine waveforms are generated via sigma-delta modulation as $\tilde{I}_k[n]$ and $\tilde{Q}_k[n]$, respectively. As shown in Fig. 1, the update algorithm of the storage elements is

$$Y_I[k] \leftarrow Y_I[k] + \hat{\delta}\tilde{I}_k[n] \tag{5}$$

$$Y_Q[k] \leftarrow Y_Q[k] + \hat{\delta}\tilde{Q}_k[n]. \tag{6}$$

At the end of a frame the amplitude at frequency index k is calculated as

$$|Y[k]| = \frac{2}{N}\sqrt{Y_I[k]^2 + Y_Q[k]^2}. \tag{7}$$

The average depth value is available from the frequency bin of $k = 0$ with an amplitude of $|Y[k{=}0]| = \frac{1}{N}Y_I[k{=}0]$. Each frequency may be probed simultaneously with the limit set by the number of storage elements and the number of measurements N.

3.4 Walsh-Hadamard Transform

Walsh functions form a complete orthogonal set and take values exclusively of -1 and +1 [24]. The Walsh-Hadamard transform does not require multiplications, which offers implementation benefits similar to the proposed $\Sigma\Delta$ modulation of the Fourier basis functions. While the Hadamard transform, computed using Walsh functions, represents the signal in terms of piecewise constant basis functions rather than sinusoidal frequency components, it does not directly decompose the signal into frequency bins. However, it can be converted into the discrete Fourier transform (DFT) [25]. This conversion typically requires computing the Hadamard transform up to at least the bandwidth limit of the signal [26] which may place significant demands on on-chip storage resources. In contrast, direct frequency analysis using $\Sigma\Delta$ modulation allows for a more flexible selection of frequency

---

[1] In this work we do not extensively explore the parameter space of one-bit modulators.

bins, making it advantageous when storage is limited. If measuring the DFT, the system can be configured to allocate storage resources specifically to frequencies associated with important dynamics.

4 METHODS FOR NON-REPETITIVE DYNAMICS

Fourier analysis is well-suited for analyzing repetitive and periodic signals. However, certain applications require measuring transient, or single-shot, dynamics of scene depth. Such applications include characterizing scene topology, detecting depth edges in the presence of ToF sensor motion, and estimating object velocity. In these scenarios, it is beneficial to both quantify the amplitude of transient depth variations and also localize them in time. Wavelets provide a joint time-frequency representation that enables the analysis of such transient components.

For hardware compatibility and background rejection, we select finite-support Haar wavelets, generated as dilated and shifted versions of the Haar mother wavelet:

$$\Psi_{0,0}(t') = \begin{cases} 1, & 0 \leq t' < 0.5 \\ -1, & 0.5 \leq t' < 1 \\ 0, & \text{otherwise} \end{cases}$$

where t′ is normalized time, $t' = {}^{t}/_{(NT_s)} = {}^{n}/_{N}$. Wavelets at different dilation levels and shifts are defined as:

$$\Psi_{u,v}(t') = \Psi_{0,0}(2^u t' - v) \tag{8}$$

where u is the integer dilation factor and v is the shift index, with $0 \leq v \leq 2^u - 1$ (Fig. 3(a)). We adopt the convention where $u = 0$ corresponds to the coarsest scale, i.e., a wavelet spanning the full signal duration. This makes it straightforward to truncate overly fine temporal resolutions.

This two-level wavelet-based approach is compatible with our proposed on-sensor electronics, which accumulate the depth signal using ±1 modulated control signals (Fig. 1(a,b)). The memory is updated as:

$$Y_u[v] \leftarrow Y_u[v] + \hat{\delta}[n]\Psi_{u,v}({}^{n}/_{N}). \tag{9}$$

To analyze up to a scale of U, a total of $2^{U+1} - 1$ memories are read out. Wavelet coefficients $w_{u,v}$ are calculated by normalizing $Y_u[v]$ by the length of the wavelet. These coefficients can be processed into a scalogram that localizes and quantifies depth transients (Fig. 3(c)), with data at coarse scales interpolated to match the resolution of finer scales. The synthetic depth signal of Fig. 3(b) contains 100,000 timestamps (e.g., a 10 ms capture at 10 MHz repetition rate) while the scalogram is generated from only 127 discrete wavelet transform coefficients for an output data compression of 787×. A hardware implementation could toggle between Fourier and wavelet analysis with a simple reconfiguration of the same sensor.

5 NOISE LIMITS AND OTHER CHALLENGES

In this section, we present the noise limits of spectrum analysis using photon time-stamps, examine degradation from background photons and possible mitigation techniques, and describe methods for handling pulses without a detection, windowing for the Fourier transform, and prefiltering needed to exploit the noise shaping properties of sigma-delta modulation.

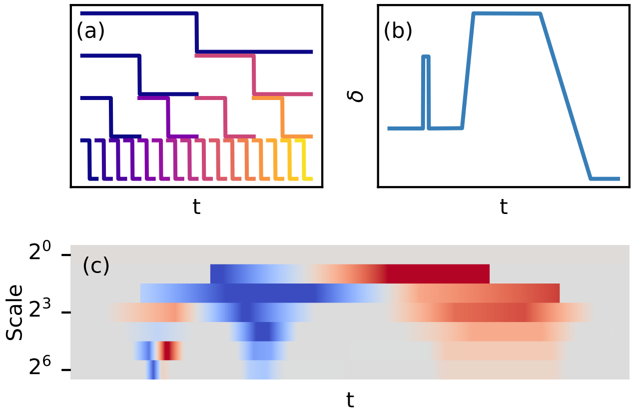

Fig. 3. (a) The non-zero values of the first four scales (u) of the Haar wavelet (offset vertically for clarity). At each scale the shifts takes a different color (e.g., at scale $u = 3$, 16 shifts span the total range). (b) A synthetic depth signal versus time with a brief transient impulse, a rising edge, and a slower falling edge. (c) A scalogram from Haar wavelet analysis of the dynamic depth signal localizes the impulse, shows the moderate rate rising edge at middle scales (blue), and the gradual falling edge at lower scales (red).

5.1 Noise

5.1.1 Timing Jitter

The noise floor of the Fourier transform is determined by the variance in the time domain, which is related to the noise power spectrum through the discrete form of Parseval's theorem:

$$\sigma^2 = \frac{1}{N} \sum_{k=0}^{N-1} S[k] \tag{10}$$

where $S[k]$ is a noise power spectrum and N is the number of measurements. Since the return time samples are independent and identically distributed $S[k]$ is frequency-independent (white noise). The time domain variance is influenced by the temporal spread of the timing measurement, which follows a Gaussian distribution, as well as quantization effects from the time to digital conversion, and background photons [27]. The temporal spread termed the instrument response function (IRF), contributes a variance of σ_I^2 due to laser, SPAD, and TDC timing jitter. The measurement precision, σ^2, improves as more signal photons, p, are collected following: $\sigma^2 = \sigma_I^2/p$. With a TDC bin size of, a, the quantization noise, given by $a^2/12$, is similarly reduced by the number of photons collected. The combined precision, accounting for both timing jitter and quantization is

$$\sigma^2 = \frac{1}{p}\left(\sigma_I^2 + \frac{a^2}{12}\right) = \frac{\sigma_s^2}{p} \tag{11}$$

where the subscript s denotes the noise in a single measurement. To simplify the analysis, we set the number of detected photons, p, equal to the number of laser pulses, N in equation 10 leading to the relationship $\sigma^2 = S[k]$. Later, in Sec. 5.2, we expand the analysis to account for pulses that do not detect a photon. The variance of the real and imaginary components of the FFT are independent, each have a variance of σ_s^2/N.

The magnitude of the DFT is then Rayleigh distributed with an expected value of

$$\mathbb{E}(|Y[k]|) = \sqrt{\frac{\pi\sigma_s^2}{N}} \tag{12}$$

where this result assumes the normalization factor of $^2/_N$ used in (1). Fig. 4 verifies this relationship between the noise floor of the DFT and the variance of instrument timing described by (12). The agreement between simulation and the theoretical expression informs design and sensor configuration. Using realistic but aggressive instrument parameters guided by state-of-the-art sensors (see the table in Suppl. Sec. S2), we predict that oscillation amplitudes as small as $5\,\mu\text{m}$ can be detected with a signal-to-noise ratio (SNR) of at least 1 using ToF spectrum analysis.

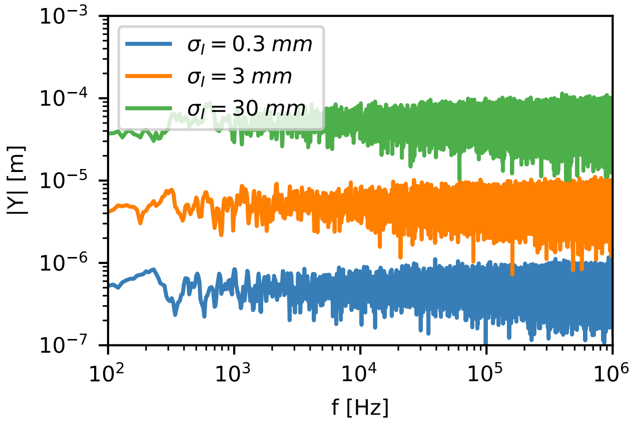

Fig. 4. Simulated noise floor set by the instrument response function. The total acquisition time is $100\,\text{ms}$ with $T_s = 100\,\text{ns}$ for 10^6 laser pulses. The average noise calculated from the simulated DFTs from 1 kHz and beyond is $0.53\,\mu\text{m}$, $5.3\,\mu\text{m}$, and $53\,\mu\text{m}$ for instrument response function standard deviations of $\sigma_I = 2\,\text{ps}$, $20\,\text{ps}$, and $200\,\text{ps}$, respectively. The IRF noise is converted to equivalent depths in the legend.

5.1.2 Background Photons

Background photons degrade ToF spectrum analysis in two primary ways: they increase the measurement variance and attenuate spectral amplitudes through pile-up. To model the impact of background photons, we assume background photon arrival times are uniformly distributed over a window of $[t_L, t_H]$, with variance $\sigma_b^2 = (t_H - t_L)^2/12$. The observed composite timestamp distribution is modeled as a mixture of signal and background photons, with average arrival number per pulse as ρ_s and $\rho_b = \lambda_b(t_H - t_L)$, respectively, where λ_b is the background arrival rate. Signal photon arrival times are assumed to follow a Gaussian distribution with variance of σ_s^2. The combined variance is then given by the law of total variance as [28]

$$\sigma^2 = \hat{\rho}_b\sigma_b^2 + \hat{\rho}_s\sigma_s^2 + \hat{\rho}_b\hat{\rho}_s(\mu_s - \mu_b)^2, \tag{13}$$

where $\hat{\rho}_s$ and $\hat{\rho}_b$ are the normalized background and signal average arrival number so that $\hat{\rho}_s + \hat{\rho}_b = 1$; μ_s is the mean of the signal timestamps; and $\mu_b = (t_H - t_L)/2$ is the mean arrival time of the background.

Background photons also attenuate the signal when they arrive earlier in the laser cycle. This pile-up effect is quantified using a survival model of

$$\rho_s^* = \rho_s e^{-\lambda_b(t_H - t_L)/2} = \rho_s e^{-\rho_b/2} \tag{14}$$

where ρ_s^* is the reduced average signal per laser pulse, assuming the signal is centered in the timing window. Fig. 5 verifies (14) using Monte Carlo simulations that generate signal and background photons (described in Sec. 6.1) and a DFT transform of the composite result; the impact of pileup is also observed in simulations using ΣΔ probing in Table 2.

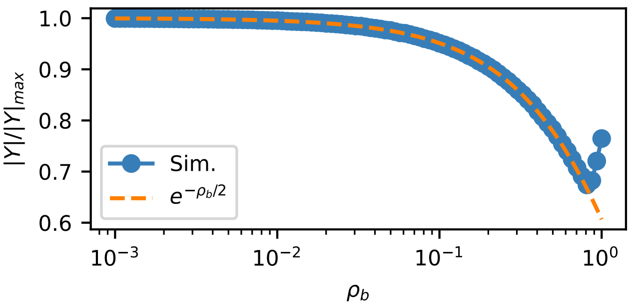

Fig. 5. Simulated normalized spectral magnitude ('Sim.') versus the average background arrival rate with the analytical survival model ($e^{-\rho_b/2}$) overlaid. At the highest background arrival rates the considerable noise in the DFT makes the amplitude extraction unreliable.

The challenges to ToF spectrum analysis from background photons emphasize the importance of reducing background rate, minimizing the span of the timing window, and centering the timing window on the average signal return time. Direct ToF sensors have incorporated some of the required capabilities including methods to shift and/or zoom a time gate which reduce the impact of background photons and efficiently utilize limited histogram bins [1], [29], [30].

5.2 Missed Pulses

Some laser pulses will detect zero returning photons so that a depth estimate is not generated. If all timestamps could be read out, methods for frequency analysis of non-uniformly spaced data, such as the Lomb-Scargle method, could be used [31], [32], [33]. However, since readout bandwidth constraints require in-pixel accumulation, alternative approaches are needed. When a laser pulse does not produce a timestamp, not accumulating into a memory element is equivalent to measuring a depth of zero ($\hat{\delta}[n] = 0$) which injects significant high frequency content into the measured spectrum due to artificial step changes. This section presents a lightweight method to fill in this missing data—a step important for both full-precision DFT and one-bit probing— and evaluates the impact on the noise floor and signal bandwidth.

To limit artifacts from zeros we replace missed measurements with the most recent valid depth measurement:

$$\text{If } \hat{\delta}[n] = \varnothing, \text{ then } \hat{\delta}[n] \leftarrow \hat{\delta}[n-1].$$

This approach effectively implements a zero-order-hold (ZOH) circuit with a hold time that varies with the stochastic photon arrivals. The magnitude of the transfer function of a ZOH with a hold time of T_h is $|H(f)| = \frac{\sin(\pi f T_h)}{\pi f T_h}$

which, when analyzed numerically, is found to have a 3dB bandwidth of $f_{3dB} = \frac{0.44}{T_h}$. This method is termed nearest neighbor (NN) imputation and can be extended so that the substituted value is an average of k prior values (kNN) [34]. Last valid replacement approach requires the sequence to be initialized with a valid depth measurement which may be captured by a brief priming sequence before beginning spectrum capture.

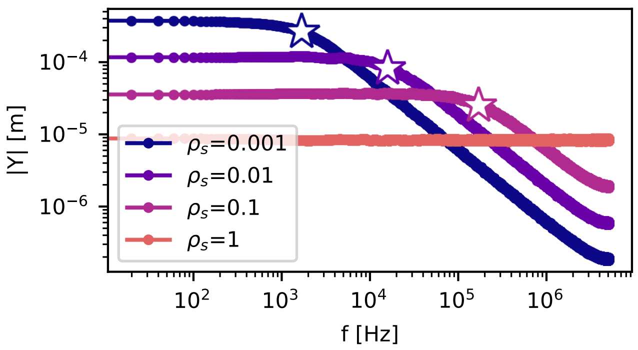

Fig. 6. Impact of the pulse detection rate (ρ_s) on the noise floor and the bandwidth using the last valid depth hold algorithm. The signal was generated with a repetition rate of $T_s = 100\,\mathrm{ns}$, and a total acquisition time of $1\,\mathrm{s}$ with a timing jitter of $\sigma_I = 100\,\mathrm{ps} = 15\,\mathrm{mm}$. The 3dB bandwidth is indicated for each curve as a ☆. The spectrum is smoothed using a Savitzky-Golay filter to ease viewing.

Fig. 6 evaluates the frequency response and noise floor when the 1NN algorithm is applied to return times with jitter and randomly generated missed detections, for varying average signal rates. From simulations, the 3dB bandwidth is 168.4, 15.8, and 1.69 kHz for rates $\rho_s = 0.1, 0.01$, and 0.001, respectively which shows that with only 1% of laser pulses returning a photon the measurement bandwidth remains above $15\,\mathrm{kHz}$. The expected run-length of missed pulses is a Bernoulli process with a success probability of ρ_s and follows a geometric distribution with mean $1/\rho_s$ and standard deviation $\sqrt{1-\rho_s}/\rho_s$; which approaches ${}^1/_{\rho_s}$ for $\rho_s \ll 1$. The run-length distribution allows us to empirically approximate the effective hold time of the 1NN algorithm as $(2/\rho_s + 1)T_s$ which corresponds to one standard deviation above the mean of the expected run-length of missed pulses. The DFT noise floor is also degraded by a lower signal rate due to fewer detections, increasing proportionally to $\sqrt{1/\rho_s}$. Additional modest impacts to the DFT noise floor from the 1NN frequency response and a windowing function are discussed alongside simulations in Sec. 6.2. Importantly, if the measured signal frequency is within the bandwidth of the last valid replacement algorithm, the spectral amplitude can still be correctly quantified despite missed pulses.

5.3 Signal Pre-Filter

Sigma-delta modulation shapes quantization noise toward high frequencies, leaving the signal band relatively noise free. However, timing jitter in the depth estimates introduces broadband noise that overlaps with the rising noise spectrum of the $\Sigma\Delta$ modulator (see Fig. 2(c)). To limit excess noise, the depth signal must be low-pass filtered before accumulation into spectrum memory. We implement a lightweight first-order recursive filter:

$$x[n] = (1-\alpha)x[n-1] + \alpha\hat{\delta}[n-1]$$

where the smoothing parameter α sets the filter bandwidth and the filter output x is the input to the spectrum analysis algorithm. In our design, the modulator operates at a sampling frequency of $f_s = 10\,\mathrm{MHz}$ with an oversampling ratio (OSR) of 32, which targets a signal bandwidth of ${}^{f_s}/_{2OSR} = 156\,\mathrm{kHz}$. We chose a smoothing parameter for efficient implementation using bit-shifts as $\alpha = 1/2^5$. This creates a filter bandwidth of

$$f_{\mathrm{LP}} = \frac{-f_s \ln\alpha}{2\pi} = 50.5\,\mathrm{kHz}.$$

which is selected conservatively to highly attenuate the quantization noise. Without this pre-filter, the simulated integrated noise using $\Sigma\Delta$ probing was around $3.4\times$ that of the full-precision DFT (see Suppl. Fig. S1.1). A signal pre-filter is required when probing waveforms are created by $\Sigma\Delta$ modulation but would not be needed if a full precision DFT is calculated.

5.4 Windowing and Frequency Resolution

The discrete Fourier transform (DFT) of a finite-duration sinusoid at frequency f_0 produces energy at other frequencies due to spectral leakage from the implicit rectangular window. To reduce leakage, standard window functions can be applied [35], and importantly, these can be integrated into the full-resolution Fourier basis prior to one-bit modulation, making the one-bit $\Sigma\Delta$ probing method compatible with standard windowing. We employ a Blackman window for both $\Sigma\Delta$ probing and full-resolution DFTs in subsequent simulations and experiments. The window impacts the noise floor of (12) [35]; this degradation is quantified in Sec. 6.2.

The DFT frequency resolution is set by the total number of samples N as $\Delta f = f_s/N$, and is influenced by the window choice. The appropriate resolution depends on the motion bandwidth of the scene (e.g., narrowband versus broadband) and the in-pixel memory capacity. When memory constraints limit the number of frequency bins, it may be desired to intentionally increase Δf by reducing N. Bartlett's method enables shorter acquisitions with coarser frequency resolution while averaging the resulting periodograms to recover the noise floor [36]. In our implementation, this corresponds to reading out multiple short frames and averaging the squared magnitudes of the spectra to produce a low-noise estimate at reduced frequency resolution.

6 SIMULATION EXPERIMENTS

6.1 Simulation Methods

Oscillating depth signals are generated and then degraded by 1) adding Gaussian distributed timing jitter, 2) dropping timestamps randomly to match the signal rate (ρ_s), 3) generating background timestamps uniformly within the zoom window, and, 4) for each laser pulse, keeping the first timestamp and discarding any subsequent measurements. All Fourier analysis is truncated to $20\,\mathrm{kHz}$ due to computation time limits associated with creating the $\Sigma\Delta$ probing waveforms. Nearest neighbor imputation is used (1NN) to fill pulses without a detection before spectral and wavelet analysis. To reduce spectral leakage, both the DFT and $\Sigma\Delta$ analysis use a Blackman window (coherent gain of 0.42) and the resulting magnitudes are normalized by the window gain. The $\Sigma\Delta$ modulated probing waveforms are generated using the Python *deltasigma* package [37] with a modulator of order 3 and oversampling ratio of 32. Suppl. Sec. S2 includes a table and discussion of performance specifications of recent ToF sensors that we used to guide simulation parameters.

6.2 Simulation Results

Fig. 7 compares the noise floor of a DFT spectrum and a two-level $\Sigma\Delta$ modulation with modest ToF sensor jitter ($\sigma_s = 3.0\,\mathrm{mm}$). The DFT result (average noise of $2.5\,\mu\mathrm{m}$) closely follows the analytical expression of (12) after scaling the number of signal counts N by the Blackman window ($w[n]$) coherent gain ($G = \frac{1}{N}\sum w[n]$) and a noise power factor ($\alpha = \frac{1}{N}\sum w[n]^2$) as $N_{eff} = N\frac{G^2}{\alpha}$ [35]. This correction factor for the Blackman window is $G^2/\alpha = 0.578$. Simulations show an average noise value equivalent to $2.25\,\mu\mathrm{m}$ and calculations predict $2.21\,\mu\mathrm{m}$.

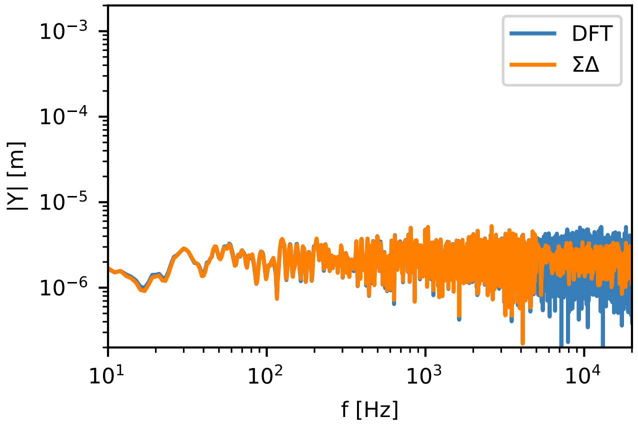

Fig. 7. The noise floor of a spectrum derived from a DFT and two-level $\Sigma\Delta$ probing. In this simulation ToF sensor jitter was $\sigma_I = 3.0\,\mathrm{mm}$ with a count rate of $\rho_s = 1$ no background photons, an integration time of $1\,\mathrm{s}$, and a pre-filter smoothing parameter $\alpha = 1/2^5$. For ease of viewing the spectrum is smoothed with a Savitzky-Golay filter (length of 13, 3rd order polynomial).

Table 2 summarizes results of Monte Carlo simulations and is segmented into experiments that test dependence on specific parameters. The first four rows assess the noise floor versus the instrument time jitter with a stationary object ($A = 0$). Probing using 1-bit $\Sigma\Delta$ waveforms demonstrates a minimal increase in the noise floor using the recursive low-pass pre-filter described in Sec. 5.3. Eq. (12) (with window correction) is verified and demonstrates that an oscillation amplitude of $\sim 5.5\,\mu\mathrm{m}$ is detectable with an instrument timing jitter of $50\,\mathrm{ps}$ and 10^6 laser pulses. The second section of Table 2 demonstrates accurate measurement of single tone depth oscillations at $4.5\,\mathrm{kHz}$ — amplitudes at or above $30\,\mu\mathrm{m}$ are measured with less than 3% error and do not increase the noise floor. The third section shows how the signal rate impacts the noise floor and amplitude measurement. The noise floor empirically increases by $C\sqrt{\frac{1}{\rho_s}}$ since fewer signal photons are measured. The additional factor C, due to the frequency response of nearest neighbor imputation, plateaus to a maximum of $\sqrt{2}$ for low signal counts. As expected the data imputation preserves the signal amplitude, even at a count rate of 3%. The final section of Table 2 assesses the impact of background photons. The SNR modestly degrades until $\rho_s = \rho_b = 0.3$ at which point pile-up diminishes the signal photons and reduces the extracted amplitude.

Fig. 8 demonstrates the frequency spectrum of a depth signal oscillating at $4.5\,\mathrm{kHz}$ with an amplitude of $0.1\,\mathrm{mm}$ and shows accurate extraction of the signal magnitude. We recognize the absurdity of this acceleration, $A\omega^2 \approx 8150g!$, but chose these parameters to exercise the high frequency capabilities and verify a minimal impact on the noise floor from a large signal. Fig. 9 shows $\Sigma\Delta$ probing results for the same frequency and amplitude as the signal rate is changed. The 1NN missing pulse fill-in algorithm maintains the signal amplitude (see inset) while the noise floor of the spectrum is increased.

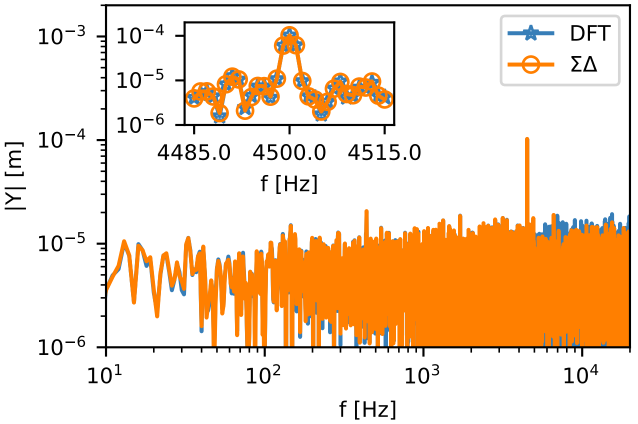

Fig. 8. Simulation of the frequency spectrum of a $0.1\,\mathrm{mm}$ amplitude depth signal oscillating at $4.5\,\mathrm{kHz}$ measured by probing with both the reference DFT and $\Sigma\Delta$ waveforms. The laser pulse repetition rate is $f_s = 10\,\mathrm{MHz}$, with a signal rate of $\rho_s = 1$, a 1-second acquisition time, and a noise of $\sigma_I = 7.5\,\mathrm{mm}$. The inset spans ± 15 Hz around the signal frequency and shows that both methods accurately measure the magnitude.

7 LABORATORY EXPERIMENTS

7.1 Methods

We evaluated the spectral analysis methods experimentally using a picosecond pulsed laser at $660\,\mathrm{nm}$ (LDH-P-C-660), operated at a repetition rate of $10\,\mathrm{MHz}$; a single-pixel SPAD detector with a diameter of $50\,\mu\mathrm{m}$ (Micro Photon Devices), paired with a 660 ± 10 nm bandpass filter (Thorlabs FBH660-10); and a time-correlated single-photon counter (TCSPC, PicoQuant PicoHarp 300), operated in time-tagged mode to individually record each photon event. The TCSPC bin width was set to $4\,\mathrm{ps}$, and the total system RMS timing jitter was measured to be $130\,\mathrm{ps}$. The laser intensity was adjusted to vary the count rate reported by the TCPSC which combines both signal and background photons. Experiments were run at low background rates, around $1.5 \times 10^3\,\mathrm{cps}$ or less. Dynamic depth scenes were generated using 3D-printed depth-feature disks attached to an optical chopper (Thorlabs, MC2000B) (see Fig. 10) and by scanning the laser beam across a discrete step in object depth. The measured

Configuration					Noise [μm]		A meas. [μm]		SNR [dB]	
σ_s [mm]	ρ_s	ρ_b	f_0 [Hz]	A [μm]	DFT	$\Sigma\Delta$	DFT	$\Sigma\Delta$	DFT	$\Sigma\Delta$
3.0	1	0	-	0	2.25	2.26	7.13	7.24	-	-
7.5	1	0	-	0	5.47	5.49	21.84	21.81	-	-
15.0	1	0	-	0	10.85	10.88	31.83	31.91	-	-
30.0	1	0	-	0	22.24	22.32	68.52	70.40	-	-
3.0	1	0	4500	10	2.26	2.27	8.27	8.25	10.24	10.20
3.0	1	0	4500	30	2.23	2.23	30.49	30.45	21.69	21.66
3.0	1	0	4500	100	2.19	2.19	97.65	97.67	31.89	31.87
3.0	1	0	4500	300	2.23	2.23	298.34	298.30	41.47	41.46
7.5	1	0	4500	100	5.52	5.54	102.30	102.63	24.29	24.29
7.5	0.30	0	4500	100	13.11	13.12	102.22	102.60	16.80	16.83
7.5	0.10	0	4500	100	23.57	23.57	114.76	114.82	12.71	12.71
7.5	0.03	0	4500	100	44.42	44.42	131.34	131.34	8.37	8.37
7.5	0.30	0.010	4500	100	13.95	13.95	99.63	99.60	15.98	15.98
7.5	0.30	0.025	4500	100	14.59	14.59	95.71	95.31	15.26	15.22
7.5	0.30	0.050	4500	100	15.29	15.29	92.61	92.66	14.57	14.57
7.5	0.30	0.075	4500	100	16.12	16.12	88.75	88.47	13.77	13.74
7.5	0.30	0.100	4500	100	16.33	16.34	98.31	98.18	14.56	14.54
7.5	0.30	0.300	4500	100	16.93	16.93	55.14	54.66	9.17	9.09

TABLE 2

Monte Carlo simulations that configure the sensor noise (σ_s), the signal rate (ρ_s), the background rate (ρ_b), the injected signal frequency (f_0), and amplitude (A). The total acquisition time is $1\,\mathrm{s}$ with a laser pulse frequency of $f_s = 10\,\mathrm{MHz}$. Results are shown from both DFT processing and two-level $\Sigma\Delta$ frequency probing. Noise is calculated as the mean of the amplitude magnitude spectrum away from the signal frequency, the measured amplitude (A meas.) is found from a 3-point quadratic interpolation of the largest value in the spectrum, and the SNR is calculated as $20\log_{10}(^{A\ meas.}/_{Noise})$. Horizontal lines divide experiments that assess the impact as different configurations are varied, from top: instrument noise, signal amplitude, signal rate, and background rate.

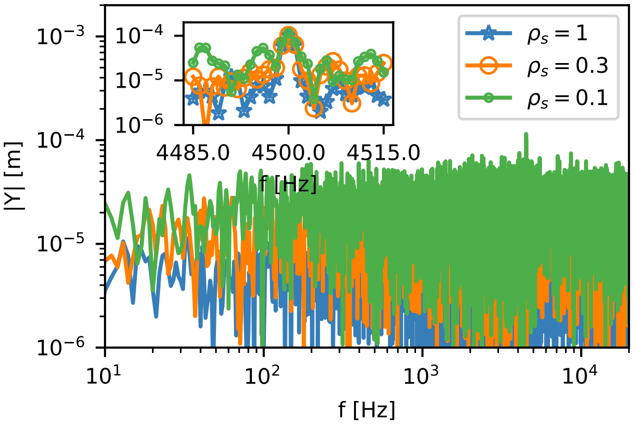

Fig. 9. Simulation of the frequency spectrum of a $0.1\,\mathrm{mm}$ amplitude depth signal oscillating at $4.5\,\mathrm{kHz}$ a with three values of signal rates (ρ_s) measured by $\Sigma\Delta$ probing over a 1-second acquisition with instrument timing jitter of $\sigma_I = 7.5\,\mathrm{mm}$. The full frequency span of the spectrum shows the noise floor increase with decreasing signal rate. The inset, centered around $4.5\,\mathrm{kHz}$, demonstrates accurate peak magnitude for all three signal rates.

time-stamp sequence was pre-processed using time-window zooming, 1NN missing pulse fill-in, the recursive pre-filter with a smoothing parameter of $\alpha = 1/2^5$, and is finally transformed to a frequency spectrum using the proposed $\Sigma\Delta$ probing waveforms.

7.2 Experiment Results

7.2.1 Experimental Spectrum Analysis

To assess the accurate extraction of high-frequency depth oscillations we measured the ToF from a depth square wave of $2.5\,\mathrm{mm}$ amplitude with a frequency of $900\,\mathrm{Hz}$. This was created using a depth feature disk with 9 periods rotated at $100\,\mathrm{Hz}$. Fig. 10 shows the amplitude spectrum obtained after processing the experimental timestamps with depth zooming, 1NN imputation, and a $50\,\mathrm{kHz}$ pre-filter, followed by spectral extraction using $\Sigma\Delta$ probing waveforms. The experimental spectrum accurately locates the frequencies of the primary components. However, the finite laser spot size occasionally causes measurements to originate from either the peak or valley of the depth profile, softening the apparent step edges and reducing the energy of high-frequency harmonics. This attenuation is visible in Fig. 10, where the predicted harmonic amplitudes exceed the experimental measurement. This experiment shows that the proposed ToF spectrum resolves a frequency component at $6.3\,\mathrm{kHz}$ with an amplitude of $270\,\mu\mathrm{m}$. The noise floor from 1 kHz - 2 kHz averages $46\,\mu\mathrm{m}$ which is consistent with (12) when corrected for the window and the signal rate.

To evaluate a more complex frequency response, a pseudo-random depth pattern disk was printed and attached to the chopper. Fig. 11 shows an image of this disk (upper left) and a single cycle of the repeating depth pattern as a function of time (upper left) when spinning at $100\,\mathrm{Hz}$. The experimentally measured spectrum ('Expt.') is compared to the DFT of the designed disk pattern ('Disk') and shows encouraging concordance. These measurements were captured with a count rate of 39.4%. Results at a lower count rate are provided in Suppl. Fig. S1.2, which shows the expected increase in the noise floor while still resolving the high-magnitude frequency components.

7.2.2 Experimental Wavelet Analysis

Non-oscillatory motions may be detected and localized using the Haar wavelet analysis described in Sec. 4. To investigate this, we scanned the laser spot across a $25\,\mathrm{mm}$ depth step (see Suppl. Fig. S1.3) using a galvanometer with a small-angle step response of $400\ \mu$s (Thorlabs, GVS012). The resulting normalized wavelet coefficients are shown as a scalogram in

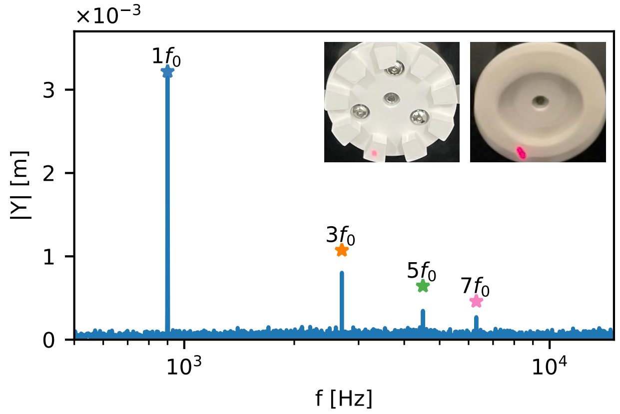

Fig. 10. The depth spectrum of a depth square wave (900 Hz with $A = 2.5$ mm) captured for 1 second. The odd harmonics of the fundamental frequency of $f_0 = 900\,\mathrm{Hz}$ are annotated. The system clearly resolves the 7th harmonic at 6300 Hz. In the upper right, the depth feature disk is shown stationary and while spinning as captured by a slo-mo video (240 fps). The TCSPC count rate was 28.3%.

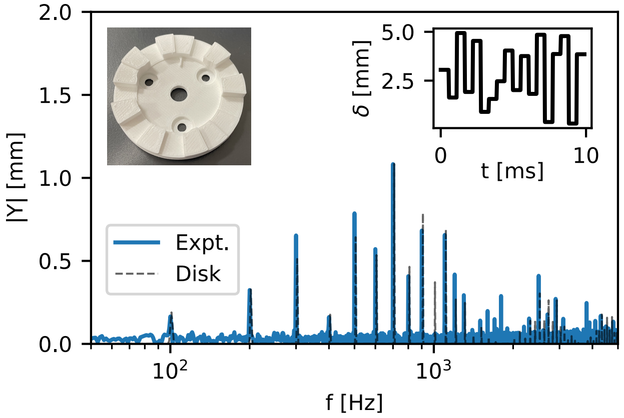

Fig. 11. The depth spectrum of a pseudo-random depth pattern shown in the inset captured for 1 second with a count rate of 39.4%. 'Expt.' is the experimentally measured spectrum and 'Disk' is the DFT of the disk design repeated for 100 cycles.

Fig. 12 (bottom). The scalogram localizes the depth transient in time and provides an estimate of its duration based on the scale of the maximum coefficient. The ramp rate of the transient can also be estimated as $4w_{u',v'}/(T/2^{u'})$ where $T/2^{u'}$ is the extent of the wavelet and u', v' denote the scale and time shift of the maximum coefficient (see Suppl. Sec. S5 for a derivation). This yields an estimate of $27.7\,\mathrm{m/s}$, which roughly matches the velocity anticipated from the scene and the galvanometer step response. On-chip wavelet compression may be advantageous over full-frame readout when the latter cannot keep pace with the scene motion. The conventional histogram and an extended sequence (5 second duration) for this configuration are shown in Suppl. Figs. S1.5 and S1.4.

8 DISCUSSION

On-sensor feasibility, memory requirements, and compression ratios. To assess the practicality of on-sensor

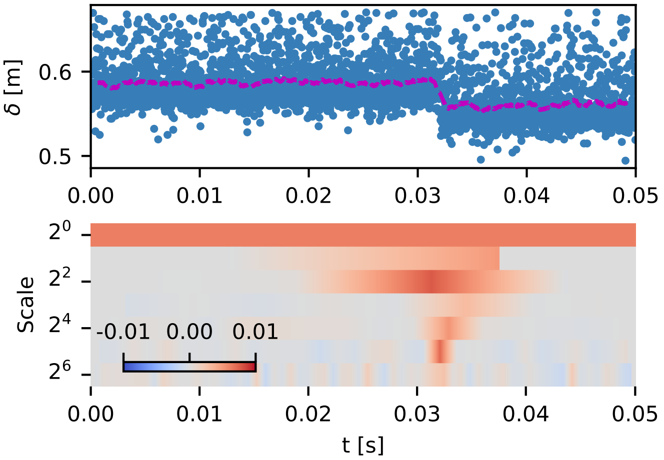

Fig. 12. The ToF laser spot is quickly moved across a depth discontinuity of 25 mm to create a transient depth-edge due to object or camera motion. (top) Measured time stamps (for viewing only 1 of every 128 are shown) (•) and a length 100 moving average (---). (bottom) The scalogram generated from 127 discrete wavelet coefficients computed with the proposed on-chip compatible Haar wavelet transform. The transient decrease in depth is revealed most prominently at a scale of 2^5 (feature duration of $50\,\mathrm{ms}/2^5$ = $1.5\,\mathrm{ms}$) and is localized in time (≈$0.032\,\mathrm{s}$). Note that the Haar wavelets are defined such that a transient decrease in depth produces a positive wavelet coefficient. This on-chip discrete wavelet transform requires the readout of 127 coefficients as compared to the 500,000 total time stamps (reduction of $\times 3937$).

implementation, we hand-coded the pixel processing algorithms in Verilog and synthesized the design using the SkyWater $130\,\mathrm{nm}$ CMOS library to evaluate area utilization. By scaling the area usage from the $130\,\mathrm{nm}$ CMOS to a modern 40 nm CMOS process we estimate that 6 and 13 frequency measurements could be supported in a $30\,\mu\mathrm{m}\times 30\,\mu m$ and $40\,\mu\mathrm{m}\times 40\,\mu m$ footprint, respectively, see Suppl. Sec. S3 for details. Analysis of the maximum clock frequency suggests that the adder and other logic can be shared among at least 16 frequency bins with a laser pulse rate of $10\,\mathrm{MHz}$, see Suppl. Sec. S3.3. In-pixel construction of a frequency spectrum reduces output bandwidth in comparison to reading out individual timestamps, provided that frequency bins are selected judiciously. The required I/Q memory depth can be estimated from (7), based on the number of laser pulses N, the maximum motion amplitude $|Y|_{max}$, and the resolution of the TDC. The maximum anticipated value stored per memory element is $Y_{I/Q} = |Y|_{max}\left(\frac{N}{2}\right)$. For example, to capture motion up to $10\,\mathrm{mm}$ with a TDC resolution of $6.67\,\mathrm{ps}$ (equivalent to $1\,\mathrm{mm}$ depth resolution) using a $10\,\mathrm{MHz}$ laser pulse rate over a 1-second interval, a 26-bit memory is required for each of I and Q at a given frequency bin. Notably, the data readout scales as $\log_2(N)$ for in-pixel frequency spectra, compared to N for full timestamp readout. For a spectrum of 128 frequency bins, the total readout is 832 bytes per pixel, versus 17.5 MB/pixel for reading all 14-bit timestamps — yielding a compression factor of approximately $22400\times$.

Noise improvements: The one-photon per laser pulse may be relaxed by using multiple SPADs to view the same object point, similar to "superpixel" depth sensor architectures where multiple SPADs share processing logic. In this superpixel configuration, individual depth measurements can be averaged per pulse before updating the spectral analysis memories to reduce the effective timing jitter. Similarly, post-readout averaging of spectra from superpixels observing the same surface region could further reduce noise.

Comparison with Laser Doppler Vibrometers. Laser Doppler Vibrometers (LDVs) are exceptionally sensitive instruments that measure in-plane vibration (or velocity) with sub-nanometer displacement sensitivity at a single point [5]. Because LDVs rely on interferometric detection of Doppler shifts, they require coherent lasers and, conventionally, free-space optics [38], making LDVs bulky, power-hungry, and sensitive to alignment. Multi-point LDV sensing typically uses mechanical scanning and/or duplicated photodetectors and processing electronics which limits speed and scaling. In contrast, our method may enable simultaneous motion analysis at multi-kHz rates across an array with <10 μm sensitivity, using a miniature 20 milligram-scale SPAD-ToF sensor. This capability is well suited for embedded applications such as monitoring the motion of a microdrone or soft robot motion sensing [39], biomechanics and body tracking [40] or human-device interaction [41], [42], and robot touch and grasping [39], [43], where displacement amplitudes are relatively large and spatial mapping is required. In these scenarios, the combination of high probing rate, per-pixel parallel capture, and low size, weight, and power (SWaP) offers advantages in form factor and scalability over traditional LDV systems. We acknowledge active research efforts to miniaturize LDVs using photonic integrated circuits and extend LDV systems to multi-point or array-based architectures [38], [44]. These developments, alongside our work on ToF-based frequency sensing, represent complementary approaches toward embedded, high-resolution motion sensing systems suitable for resource-constrained platforms.

More advanced imputation for low-SNR operation. This work uses a zero-order-hold (1NN) to approximate missing depth measurements, which enables simple hardware implementation and performs well at moderate signal rates. However, as signal rates decrease or scene dynamics become more complex, the assumption of constant depth between valid measurements becomes increasingly limiting. Future work could explore more advanced imputation methods, such as a motion model and a Kalman filter to estimate missing data. In our experiments, 1NN imputation did not significantly impact spectral accuracy, suggesting that any improvement from more sophisticated methods must be carefully balanced against their hardware and implementation costs.

ACKNOWLEDGMENTS

S.L. was supported by a Young Scholars grant from the University of St. Thomas Undergraduate Research Opportunities Program (UROP) and the Scalable Asymmetric Lifecycle Engagement (SCALE) program (U.S. DoD contract W52P1J-22-9-3009). B.R.J. was supported by the University of St. Thomas Center for Applied Mathematics (CAM). L.J.K. acknowledges the School of Engineering for equipment support and the Minnesota Supercomputing Institute (MSI) at the University of Minnesota for providing resources. Thanks to Ryan Doerr for designing and printing the depth feature disks.

REFERENCES

[1] I. Gyongy, N. A. W. Dutton, and R. K. Henderson, “Direct Time-of-Flight Single-Photon Imaging,” *IEEE Transactions on Electron Devices*, pp. 1–12, 2021.

[2] C. Niclass, M. Soga, H. Matsubara, S. Kato, and M. Kagami, “A 100-m Range 10-Frame/s 340x96-Pixel Time-of-Flight Depth Sensor in 0.18-um CMOS,” *IEEE Journal of Solid-State Circuits*, vol. 48, no. 2, pp. 559–572, Feb. 2013.

[3] G. Luetzenburg, A. Kroon, and A. A. Bjørk, “Evaluation of the apple iPhone 12 pro LiDAR for an application in geosciences,” *Scientific reports*, vol. 11, no. 1, pp. 1–9, 2021.

[4] F. Gutierrez-Barragan, A. Ingle, T. Seets, M. Gupta, and A. Velten, “Compressive Single-Photon 3D Cameras,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 17 854–17 864.

[5] S. J. Rothberg, M. S. Allen, P. Castellini, D. Di Maio, J. J. J. Dirckx, D. J. Ewins, B. J. Halkon, P. Muyshondt, N. Paone, T. Ryan, H. Steger, E. P. Tomasini, S. Vanlanduit, and J. F. Vignola, “An international review of laser Doppler vibrometry: Making light work of vibration measurement,” *Optics and Lasers in Engineering*, vol. 99, pp. 11–22, Dec. 2017.

[6] D. Abe, R. Michael, W. Neal, G. J. Mysore, D. Frédo, and W. T. Freeman, “The Visual Microphone: Passive Recovery of Sound from Video,” *ACM Transactions on Graphics*, vol. 33, no. 4, pp. 1–10, Jul. 2014.

[7] M. Wei, S. Nousias, R. Gulve, D. B. Lindell, and K. N. Kutulakos, “Passive Ultra-Wideband Single-Photon Imaging,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 8135–8146.

[8] R. Kitichotkul, J. Rapp, Y. Ma, and H. Mansour, “Doppler Single-Photon Lidar,” in *ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Apr. 2025, pp. 1–5.

[9] ——, “Simultaneous range and velocity measurement with Doppler single-photon lidar,” *Optica*, vol. 12, no. 5, pp. 604–613, May 2025.

[10] F. Gutierrez-Barragan, F. Mu, A. Ardelean, A. Ingle, C. Bruschini, E. Charbon, Y. Li, M. Gupta, and A. Velten, “Learned Compressive Representations for Single-Photon 3D Imaging,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 10 756–10 766.

[11] M. P. Sheehan, J. Tachella, and M. E. Davies, “A Sketching Framework for Reduced Data Transfer in Photon Counting Lidar,” *IEEE Transactions on Computational Imaging*, vol. 7, pp. 989–1004, 2021.

[12] J. H. Nam, E. Brandt, S. Bauer, X. Liu, M. Renna, A. Tosi, E. Sifakis, and A. Velten, “Low-latency time-of-flight non-line-of-sight imaging at 5 frames per second,” *Nature Communications*, vol. 12, no. 1, p. 6526, Nov. 2021.

[13] T. D. Memon, P. Beckett, and A. Z. Sadik, “POWER-Area-Performance Characteristics of FPGA-based Sigma-Delta FIR Filters,” *Journal of Signal Processing Systems*, vol. 70, no. 3, pp. 275–288, Mar. 2013.

[14] A. Pathan, T. D. Memon, S. Raza, and R. Aziz, “An autocorrelation-less single-bit Weiner filter on FPGA,” *Biomedical Signal Processing and Control*, vol. 86, p. 105166, Sep. 2023.

[15] A. Lavatelli, R. Balcaen, E. Zappa, and D. Debruyne, “An Experimental Investigation on Uncertainty in Measuring Vibration Deflection Shapes With Digital Image Correlation,” *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 6, pp. 2826–2836, Jun. 2020.

[16] J. G. Chen, N. Wadhwa, Y.-J. Cha, F. Durand, W. T. Freeman, and O. Buyukozturk, “Modal identification of simple structures with high-speed video using motion magnification,” *Journal of Sound and Vibration*, vol. 345, pp. 58–71, Jun. 2015.

[17] A. Davis, K. L. Bouman, J. G. Chen, M. Rubinstein, F. Durand, and W. T. Freeman, “Visual Vibrometry: Estimating Material Properties From Small Motion in Video,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 5335–5343.

[18] C. Liu, A. Torralba, W. T. Freeman, F. Durand, and E. H. Adelson, “Motion magnification,” *ACM Trans. Graph.*, vol. 24, no. 3, pp. 519–526, Jul. 2005.

[19] N. Wadhwa, M. Rubinstein, F. Durand, and W. T. Freeman, “Phase-based video motion processing,” *ACM Trans. Graph.*, vol. 32, no. 4, pp. 80:1–80:10, Jul. 2013.

[20] M. Sheinin, D. Chan, M. O’Toole, and S. G. Narasimhan, “Dual-Shutter Optical Vibration Sensing,” in *Proceedings of the IEEE/CVF*

Conference on Computer Vision and Pattern Recognition, 2022, pp. 16 324–16 333.

[21] K. Sarkar, R. M. Corey, and A. C. Singer, “Utilizing Time-of-Flight LIDARs For Spatial Audio Processing,” in *2021 55th Asilomar Conference on Signals, Systems, and Computers*, Oct. 2021, pp. 1388–1392.

[22] G. R. Tondo, C. Riley, and G. Morgenthal, “Characterization of the iPhone LiDAR-Based Sensing System for Vibration Measurement and Modal Analysis,” *Sensors*, vol. 23, no. 18, p. 7832, Jan. 2023.

[23] J. D. Reiss, “Understanding sigma-delta modulation: The solved and unsolved issues,” *Journal of the Audio Engineering Society*, vol. 56, no. 1/2, pp. 49–64, 2008.

[24] H. F. Harmuth, “Applications of Walsh functions in communications,” *IEEE Spectrum*, vol. 6, no. 11, pp. 82–91, Nov. 1969.

[25] Kunz, “On the Equivalence Between One-Dimensional Discrete Walsh-Hadamard and Multidimensional Discrete Fourier Transforms,” *IEEE Transactions on Computers*, vol. C-28, no. 3, pp. 267–268, Mar. 1979.

[26] R. Kitai, “Walsh-to-Fourier Spectral Conversion for Periodic Waves,” *IEEE Transactions on Electromagnetic Compatibility*, vol. EMC-17, no. 4, pp. 266–269, Nov. 1975.

[27] L. J. Koerner, “Models of Direct Time-of-Flight Sensor Precision That Enable Optimal Design and Dynamic Configuration,” *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–9, 2021.

[28] J. Behboodian, “On a mixture of normal distributions,” *Biometrika*, vol. 57, no. 1, pp. 215–217, Apr. 1970.

[29] I. Gyongy, A. Erdogan, N. Dutton, H. Mai, F. M. Della Rocca, and R. K. Henderson, “A 200kFPS, 256× 128 SPAD dToF sensor with peak tracking and smart readout,” in *International Image Sensor Workshop 2021*, 2021.

[30] F. Taneski, T. A. Abbas, and R. K. Henderson, “Laser Power Efficiency of Partial Histogram Direct Time-of-Flight LiDAR Sensors,” *Journal of Lightwave Technology*, vol. 40, no. 17, pp. 5884–5893, Sep. 2022.

[31] N. R. Lomb, “Least-squares frequency analysis of unequally spaced data,” *Astrophysics and Space Science*, vol. 39, no. 2, pp. 447–462, Feb. 1976.

[32] J. D. Scargle, “Studies in astronomical time series analysis. III-Fourier transforms, autocorrelation functions, and cross-correlation functions of unevenly spaced data,” *Astrophysical Journal, Part 1 (ISSN 0004-637X), vol. 343, Aug. 15, 1989, p. 874-887.*, vol. 343, pp. 874–887, 1989.

[33] M. Zechmeister and M. Kürster, “The generalised Lomb-Scargle periodogram - A new formalism for the floating-mean and Keplerian periodograms,” *Astronomy & Astrophysics*, vol. 496, no. 2, pp. 577–584, Mar. 2009.

[34] L. Stankovic, S. Stankovic, and M. Amin, “Missing samples analysis in signals for applications to L-estimation and compressive sensing,” *Signal Processing*, vol. 94, pp. 401–408, Jan. 2014.

[35] F. Harris, “On the use of windows for harmonic analysis with the discrete Fourier transform,” *Proceedings of the IEEE*, vol. 66, no. 1, pp. 51–83, Jan. 1978.

[36] M. S. Bartlett, “Smoothing Periodograms from Time-Series with Continuous Spectra,” *Nature*, vol. 161, no. 4096, pp. 686–687, May 1948.

[37] G. Venturini and Y. Fukuda, “Python-deltasigma,” https://www.python-deltasigma.io/, Aceessed: 2025-04-01, 2013.

[38] Y. Li, E. Dieussaert, and R. Baets, “Miniaturization of Laser Doppler Vibrometers—A Review,” *Sensors*, vol. 22, no. 13, p. 4735, Jan. 2022.

[39] M. Hofer, C. Sferrazza, and R. D’Andrea, “A Vision-Based Sensing Approach for a Spherical Soft Robotic Arm,” *Frontiers in Robotics and AI*, vol. 8, Feb. 2021.

[40] D. Balta, “Markerless movement analysis based on RGB and Depth sensor Technology for clinical applications,” Ph.D. dissertation, Politecnico di Torino, 2024.

[41] S. Sridhar, A. Markussen, A. Oulasvirta, C. Theobalt, and S. Boring, “WatchSense: On- and Above-Skin Input Sensing through a Wearable Depth Sensor,” in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, ser. CHI ’17. New York, NY, USA: Association for Computing Machinery, May 2017, pp. 3891–3902.

[42] C. Keskin, F. Kıraç, Y. E. Kara, and L. Akarun, “Real time hand pose estimation using depth sensors,” in *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, Nov. 2011, pp. 1228–1234.

[43] D. F. Gomes, Z. Lin, and S. Luo, “GelTip: A Finger-shaped Optical Tactile Sensor for Robotic Manipulation,” in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Oct. 2020, pp. 9903–9909.

[44] J. Li, V. Aranchuk, and I. Aranchuk, “Multi-beam full-field laser Doppler vibrometer based on a high-speed area scan CMOS camera,” *Applied Optics*, vol. 64, no. 17, pp. 4934–4940, Jun. 2025.

Seth Lindgren (Student Member, IEEE) is pursuing the B.S. degree in Electrical Engineering and the B.A. degree in Physics at the University of St. Thomas, St. Paul, MN, USA. His research interests include imaging systems and biomedical devices.

Benjamin R. Johnson is pursuing the B.S. degree in Electrical Engineering with minors in Physics and Mathematics at the University of St. Thomas, St. Paul, MN, USA. His research interests include applied mathematics and computational methods for algorithm development.

Lucas J. Koerner (Member, IEEE) received the B.A. (Hons.) degree in integrated science, physics, and mathematics from Northwestern University, Evanston, IL, USA, and the Ph.D. degree in physics from Cornell University, Ithaca, NY, USA. Since 2023, he has been an Associate Professor of Electrical and Computer Engineering with the University of St. Thomas, St. Paul, MN, USA. His research interests include electrical instrumentation, time-of-flight sensing, image sensors, and resource-efficient imaging systems.