Centimeter-Wave Free-Space Neural Time-of-Flight Imaging

SEUNG-HWAN BAEK*, Princeton University
NOAH WALSH*, Princeton University
ILYA CHUGUNOV, Princeton University
ZHENG SHI, Princeton University
FELIX HEIDE, Princeton University

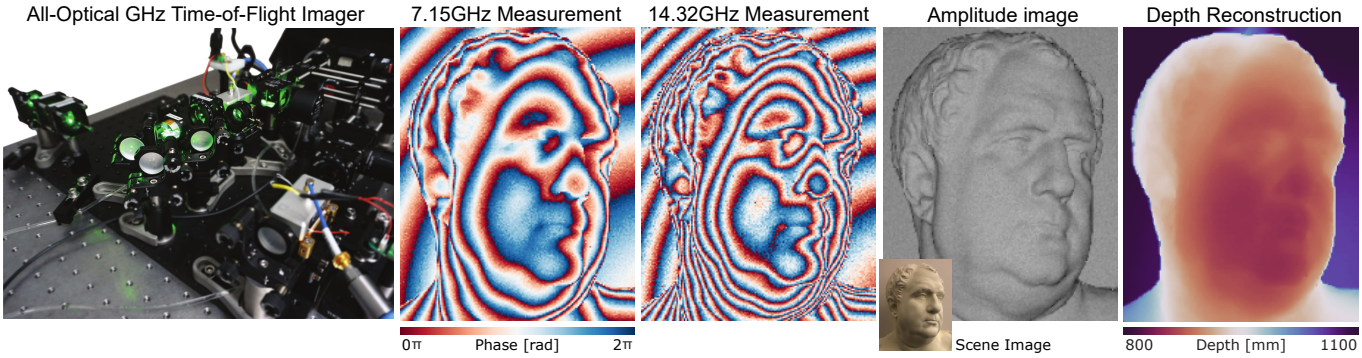

Fig. 1. We propose an all-optical neural time-of-flight (ToF) imaging system with centimeter-wave intensity-modulated illumination. To this end, we repurpose electro-optical modulators used in optical communication to compute GHz frequency ToF correlation signals in free space, avoiding photo-conversion and fiber coupling. The proposed system provides correlation measurements at 7.15 GHz and 14.32 GHz modulation frequencies (second to left and center), which results in dozens of phase wraps over meter-scale scenes that cannot be estimated accurately by existing phase unwrapping methods. To this end, we propose a segmentation-inspired neural phase unwrapping network that recovers accurate scene depth (right) from the correlation measurements and scene amplitude (second to right). See scene photograph (inset second to right) as reference. We demonstrate a robust method for all-optical GHz-frequency correlation ToF depth imaging of macroscopic scenes.

Depth sensors have emerged as a cornerstone sensor modality with diverse applications in personal hand-held devices, robotics, scientific imaging, autonomous vehicles, and more. In particular, correlation Time-of-Flight (ToF) sensors have found widespread adoption for meter-scale indoor applications such as object tracking and pose estimation. While they offer high depth resolution at competitive costs, the precision of these indirect ToF sensors is fundamentally limited by their modulation contrast, which is in turn limited by the effects of photo-conversion noise. In contrast, optical interferometric methods can leverage short illumination modulation wavelengths to achieve depth precision three orders of magnitude greater than ToF, but typically find their range restricted to the sub-centimeter.

In this work, we merge concepts from both correlation ToF design and interferometric imaging; a step towards bridging the gap between these methods. We propose a computational ToF imaging method which optically computes the GHz ToF correlation signal in free space before photo-conversion. To acquire a depth map, we scan a scene point-wise and computationally unwrap the collected correlation measurements. Specifically, we repurpose electro-optical modulators used in optical communication for ToF imaging with centimeter-wave signals, and achieve all-optical correlation at 7.15 and 14.32 GHz modulation frequencies. While GHz modulation frequencies increase depth precision, these high modulation rates also pose a technical challenge. They result in dozens of wraps per meter which cannot be estimated robustly by existing phase unwrapping methods. We tackle this problem with a proposed segmentation-inspired *phase unwrapping network*, which exploits the correlation of adjacent GHz phase measurements to classify regions into their respective wrap counts. We validate this method in simulation and experimentally, and demonstrate precise depth sensing using centimeter wave modulation that is robust to surface texture and ambient light. Compared to existing analog demodulation methods, the proposed system outperforms all of them across all tested scenarios.

CCS Concepts: • **Computing methodologies**;

Additional Key Words and Phrases: Time-of-flight imaging, 3D imaging

ACM Reference Format:
Seung-Hwan Baek, Noah Walsh, Ilya Chugunov, Zheng Shi, and Felix Heide. 2021. Centimeter-Wave Free-Space Neural Time-of-Flight Imaging . *ACM Trans. Graph.* 39, 4, Article 1 (July 2021), 18 pages. https://doi.org/http://dx.doi.org/10.1145/8888888.7777777

1 INTRODUCTION

From interactive gaming to precision industrial manufacturing, depth sensors have enabled advances in a broad set of consumer and research applications. Their ability to recover 3D data at scale [Chang et al. 2015; Dai et al. 2017; Silberman et al. 2012] and produce high-fidelity scene reconstructions [Izadi et al. 2011; Tulsiani et al. 2018] drives developments in 3D scene understanding [Dai et al. 2018;

*Authors contributed equally to this work.

Authors' addresses: Seung-Hwan Baek, Princeton University; Noah Walsh, Princeton University; Ilya Chugunov, Princeton University; Zheng Shi, Princeton University; Felix Heide, Princeton University.

Hickson et al. 2014; Song et al. 2015], which in turn influence the fields of augmented reality, virtual reality, robotic scanning, autonomous vehicle guidance, and path planning for delivery drones.

Some of the most successful depth acquisition approaches for wide operating ranges are based on active time-of-flight sensing, as they offer high depth precision at a small sensor-illumination baseline [Hansard et al. 2012]. Passive approaches, that infer distance from parallax [Mahjourian et al. 2018; Subbarao and Surya 1994] or visual cues in monocular images [Bhat et al. 2021; Saxena et al. 2005], do not offer the same range and depth precision as they struggle with textureless regions and complex geometries [Lazaros et al. 2008; Smolyanskiy et al. 2018]. Active sensing approaches tackle this challenge by projecting light into the scene and reconstructing depth from the returned signal. Structured light methods such as active stereo systems use spatially patterned light to aid stereo matching [Ahuja and Abbott 1993; Baek and Heide 2021]. While being robust to textureless scenes, their accuracy is limited by illumination pattern density and sensor baseline, resulting in a large form factor. ToF depth sensing approaches avoid these limitations by estimating depth from the travel time of photons leaving from and returning to the device, allowing for co-axial sensor setups with virtually no illumination-camera baseline.

Direct ToF systems, such as light detection and ranging (LiDAR) sensors [Schwarz 2010], *directly* measure the round-trip time of emitted light pulses to estimate point depths, and can theoretically provide accuracy over a long range. However, this direct acquisition approach demands fast pulsed lasers, accurate synchronization, narrow-band filters, and picosecond-resolution time-tagged detectors such as single-photon avalanche diodes (SPADs) [Aull et al. 2002; Bronzi et al. 2015; Niclass et al. 2005; Rochas et al. 2003]. Though affordable SPADs have recently entered the market, these have only 20cm depth resolution [Callenberg et al. 2021], more than 50× lower than their costly picosecond-resolution counterparts.

Amplitude-modulated continuous-wave (AMCW) ToF methods – which we hereon refer to as *correlation* ToF methods – [Gupta et al. 2015; Lange and Seitz 2001; Shrestha et al. 2016; Su et al. 2018] flood a scene with periodic amplitude-modulated light and *indirectly* infer depth from the phase shift of returned light. In contrast to direct ToF sensing approaches, this modulation and correlation does not require ultra-short pulse generation and time-tagging, this lowers sensor and laser complexity requirements. Correlation ToF sensors that demodulate the amplitude-modulated flash-illumination on-sensor have been widely adopted, for example, the Microsoft Kinect One camera. These sensors implement multiple charge buckets per pixel and shift a photo-electron to an individual bucket by applying an electrical potential between the individual quantum wells [Lange and Seitz 2001]. Though amplitude modulation allows for depth precision *comparable* to picosecond-pulsed direct ToF at meter-scale distances, while remaining low-cost thanks to scalable CMOS technology, it is also this sensing mode that fundamentally limits the sensor. Specifically, *modulation after photo-electric conversion limits the maximum achievable modulation frequency to a few hundred MHz* in practice, restricted by the photon absorption depth in silicon [Lange and Seitz 2001]. This has limited the depth precision of existing correlation ToF sensors to the sub-centimeter regime. Fiber-coupled modulation approaches from optical communication which bypass this limit suffer from low modulation contrast due to coupling loss [Bandyopadhyay et al. 2020; Kadambi and Raskar 2017; Marchetti et al. 2017; Rogers et al. 2021].

In this work, we co-opt free-space electro-optic modulators (EOMs) from optical communication and combine them with a phase unwrapping neural network to build a GHz correlation ToF system. EOM-based ranging systems are known to offer fast intensity modulation and can be integrated with conventional intensity sensors and a continuous-wave laser, bypassing the more complex hardware requirements of time-tagged ToF devices [Froome and Bradsell 1961]. Inspired by existing EOM-based ranging methods, we devise a two-pass EOM-based GHz ToF sensing system that achieves a 7 GHz modulation frequency with > 50% contrast. Our system inherits the benefits of EOM-based systems – large-area freespace modulation, single-digit driving voltage – using conventional intensity sensors and continuous-wave lasers.

Although a higher modulation frequency can increase phase contrast and allow for more precise depth measurement, it also greatly complicates the task of phase unwrapping, a major obstacle in applying EOMs to depth sensing. At 7 GHz, even a 2 cm depth change results in a phase wrap, in contrast to 3 m of unambiguous depth for a 100 MHz ToF camera. In addition to a few dozens of wraps, imaging noise and the small modulation bandwidth of EOMs – only a few MHz – imposes a further challenge for conventional look-up table approaches. We tackle this challenge with a segmentation-inspired neural phase unwrapping network, where the problem is decomposed into ordinal classification, mapping regions of measured data to their wrap count. Trained in an end-to-end fashion on simulated ToF data and fine-tuned on a small set of experimental measurements, the proposed network exploits the correlation of adjacent measurements to robustly unwrap them.

We validate the proposed ToF system in simulation and experimentally, and demonstrate robust depth imaging for macroscopic diffuse scenes with freespace centimeter-wave modulation at mW laser powers, corresponding to < 100 femtosecond temporal resolution. Jointly with the learned unwrapping, the all-optical modulation without coupling losses allows for robustness to low-reflectance texture regions and highly specular objects with low diffuse reflectance components. We assess the neural phase unwrapping network extensively on real and simulated data, and validate that it outperforms existing conventional and learned unwrapping approaches across all tested scenarios. We further validate precision and compare extensively against post-photoconversion modulation, which fails in low flux scenarios, and interferometric approaches, that are limited to small ranges. As our free-space modulation is all-optical, we demonstrate that it can be readily combined with interferometric modulation, allowing us to narrow the gap between interferometry and correlation ToF imaging, with the future potential for photon-efficient imaging of macro-scale ultrafast phenomena.

Specifically, we make the following contributions in this work:

- We introduce computational ToF imaging with fully optical free-space correlation and an EOM-based two-pass intensity modulation that allows for ≥ 10 GHz frequencies.

- To tackle phase-unwrapping at centimeter wavelengths, we introduce a segmentation-based phase unwrapping network that poses phase recovery as a classification problem.
- We validate the proposed method experimentally with a prototype, achieving robust depth imaging with freespace centimeter-wave modulation for macroscopic scenes.

To ensure reproducibility, we will share the schematics, code, and optical design of the proposed method.

2 RELATED WORK

In this section, we seek to give the reader a broad overview of the current state of depth imaging to better illustrate the gap our work fills in the 3D vision ecosystem.

Depth Imaging. The wide family of modern depth imaging methods can be broadly categorized into passive and active systems. Passive approaches, which leverage solely image cues such as parallax [Baek et al. 2016; Hirschmuller 2005; Meuleman et al. 2020] or defocus [Subbarao and Surya 1994], can offer low-cost depth estimation solutions using commodity camera hardware [Garg et al. 2019]. Their reliance on visual features, however, means they struggle to achieve sub-cm accuracy in favorable conditions, and can fail catastrophically for complex scene geometries and textureless regions [Smolyanskiy et al. 2018]. Active methods, which first project a known signal into the scene before attempting to recover depth, can reduce this reliance on visual features. For example, structured light approaches, such as those used in the Kinect V1 and Intel D415 depth cameras, improve local image contrast with active illumination patterns [Ahuja and Abbott 1993; Baek and Heide 2021; Scharstein and Szeliski 2003], at a detriment to form-factor and power consumption. Even active stereo methods, however, still cannot disambiguate mm-scale features, as they are smaller than the illumination feature size itself and make finding accurate stereo correspondences infeasible. Time-of-flight (ToF) imaging is an active method that does not rely on visual cues, and so avoids the pitfalls of stereo matching completely. ToF cameras instead directly or indirectly measure the travel time of light to infer distances [Hansard et al. 2012; Lange and Seitz 2001], with modern continuous-wave correlation ToF systems achieving sub-cm accuracy for megahertz-scale modulation frequencies. Interferometry extends this principle to the terahertz range, measuring the interference of electromagnetic waves to estimate their travel time. These systems can achieve micron-scale accuracy at the cost of mm-scale operating ranges [Hariharan 2003]. In this work we seek to bridge the gap between commodity MHz-frequency correlation ToF systems and THz frequency interferometry with a GHz-frequency correlation ToF system for meter-scale imaging.

Pulsed ToF. Pulsed ToF systems, such as LiDAR, are direct ToF acquisition methods which *directly* measure the travel time of photon packets to infer depth. They send discrete laser pulses into the scene and detect their reflections with avalanche photodiodes [Cova et al. 1996; Pandey et al. 2011] or single-photon detectors [Gupta et al. 2019a,b; Heide et al. 2018; McCarthy et al. 2009]. These sensors can extract depth from measured pulse returns without phase wrap ambiguities. Their depth precision is limited by their temporal resolution, however, and the complex detectors and narrow-band filters, used to filter out ambient light, contend with high cost as a result of fabrication complexity when compared to conventional intensity sensors. Recently, low-cost pulsed sensors have appeared, however at the cost of coarse 20 cm depth precision [Callenberg et al. 2021]. In this work, we revisit indirect ToF with amplitude modulation paired with learned phase unwrapping as an approach to precise depth imaging that does not mandate time-resolved sensors and time-tagging electronics.

Correlation ToF. Amplitude-modulated continuous-wave ToF, which we refer to as simply correlation ToF, floods the scene with periodically modulated illumination and infers distance from phase differences in the returned light [Hagebeuker and Marketing 2007; Lange and Seitz 2001; Remondino and Stoppa 2013]. These systems, such as cameras in the prolific Microsoft Kinect series [Tölgyessy et al. 2021], can rely on affordable CMOS sensors and conventional CW laser diodes to produce dense depth measurements. This flood illumination can lead to multipath interference, though there exists a large body of work to mitigate this [Achar et al. 2017; Bhandari et al. 2014; Freedman et al. 2014; Fuchs 2010; Jiménez et al. 2014; Kadambi et al. 2013; Kirmani et al. 2013; Naik et al. 2015]. Correlation ToF measurements can also be used to resolve the travel-time of light in flight [Heide et al. 2013; Kadambi et al. 2013]. These time-resolved transient images have found a number of emerging applications, such as non-line-of-sight imaging [Heide et al. 2014; Kadambi et al. 2016], imaging through scattering media [Heide et al. 2014], and material classification [Su et al. 2016], which have also been solved with pulsed ToF systems [Heide et al. 2019; O'Toole et al. 2018] and interferometric methods [Gkioulekas et al. 2015]. All these methods, however, are restricted to working with modulation frequencies of only a few hundred MHz due to photon absorption depth in silicon [Lange and Seitz 2001], which governs how these devices perform photo-electric conversion. This limit places the depth resolution of modern correlation ToF sensors at mm- to cm-scale for operating ranges of up to several meters. Previous attempts at pushing this modulation frequency to the GHz regime struggle with low modulation contrast due to the energy loss from fiber coupling within eye-safe laser power levels [Kadambi and Raskar 2017; Li et al. 2018]. Li et al. [2018] overcome some of these limitations but solely rely on interferometric modulation, making the method susceptible to speckle, vibration, laser frequency drift, and other common interferometry errors. Notably, Bamji et al. [2018] achieve 200 MHz modulation frequency at high contrast, but are limited to single-frequency modulation. Gupta et al. [2018] achieve 500MHz modulation frequency with a fast photodiode and analog radio-frequency (RF) modulation, but contend with low modulation contrast at the GHz regime due to modulation after photo-conversion.

Interferometry and Frequency-Modulated Continuous-Wave ToF. Optical interferometry leverage the interference of electromagnetic waves to infer their path lengths, which is encoded in the measured amplitude and/or phase patterns. A detailed review of interferometry can be found in [Hariharan 2003]. Methods such as optical coherence tomography (OCT) [Huang et al. 1991] have found prolific use in biomedical applications [Fujimoto and Swanson 2016] for their ability to resolve micron-scale features in optical scattering media. This, however, comes with the caveat of a mm-scale operating

range as diffuse scattering leads to a sharp decline in SNR. In graphics, OCT approaches have been successfully employed to achieve micron-scale light transport decompositions [Gkioulekas et al. 2015] and light transport probing [Kotwal et al. 2020]. Fourier-domain OCT systems mitigate some of the sensitivity to vibration by using a spectrometer and a broadband light source [Leitgeb et al. 2003]. While these methods provide high temporal resolution, they are also limited to cm-scale scenes. Frequency-modulated continuous-wave (FMCW) ToF systems employ an alternative interferometric approach to measuring distance. These methods continuously apply frequency modulation to their output illumination, which when combined in a wave-guide with the delayed returned light from the scene produces constructive and destructive interference patterns from which travel-time (and thereby depth) can be inferred. Experimental FMCW LiDAR setups can achieve millimeter precision for scenes at decimeter range [Behroozpour et al. 2016], but require complex tunable laser systems [Amann 1992; Gao and Hui 2012; Sandborn et al. 2016]. We revisit continuous-wave intensity modulation, which allows us to use conventional continuous-wave lasers modulated and demodulated in free-space.

Phase Unwrapping. In correlation ToF systems, the analog correlation signal can experience phase shifts of more than one wavelength. To recover the true phase, and thereby accurately reconstruct depth, phase unwrapping algorithms are required [An et al. 2016; Crabb and Manduchi 2015; Dorrington et al. 2011; Lawin et al. 2016]. Single phase unwrapping approaches are only able to recover the relative depth, and require a-priori assumptions to estimate scale [Bioucas-Dias et al. 2008; Bioucas-Dias and Valadao 2007; Crabb and Manduchi 2015; Ghiglia and Pritt 1998]. Multi-frequency phase unwrapping methods overcome this limitation by unwrapping high-frequency phases with their lower-frequency counterpart. Wrap count is recovered by either weighing Euclidean division candidates [Bioucas-Dias et al. 2009; Droeschel et al. 2010; Freedman et al. 2014; Kirmani et al. 2013; Lawin et al. 2016], or using frequency-space lookup table [Gupta et al. 2015]. All of these methods, while powerful for MHz ToF imaging, fail in the presence of noise for the dozens of wrap counts observed in the GHz correlation imaging. To tackle this challenge, in this work we introduce a neural network capable of unwrapping GHz frequency ToF correlation measurements.

Electro-optic Modulators. EOMs control the refractive index of a crystal with an electric field to modulate the phase, frequency, amplitude, and polarization of incident light [Yariv and Yeh 2007]. As such, they have been employed in diverse applications, including fiber communications [Phare et al. 2015], frequency modulation spectroscopy [Tai et al. 2016], laser mode locking [Hudson et al. 2005], and optical interferometry [Minoni et al. 1991]. In particular, EOMs have been used in LiDAR systems to change the optical-carrier frequency for FMCW sensing [Behroozpour et al. 2017] or facilitate pulsed sensing [Chen et al. 2018]. Instead, we repurpose these EOMs for continuous-wave correlation ToF imaging. We employ a two-pass modulation scheme for our ranging system that, instead of optical frequency, modulates *intensity* with high contrast. We combine this acquisition scheme with a neural phase unwrapping method to then unwrap the dozens of phase wraps we encounter in the GHz regime.

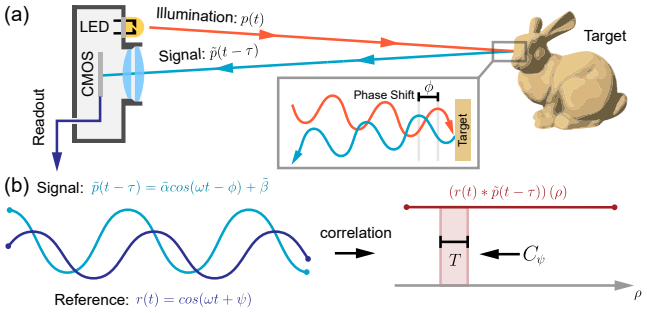

Fig. 2. Principle of correlation ToF. (a) Typical correlation ToF imagers emit coded illumination into a scene with time-varying sinusoidal intensity modulation. The reflected light then encodes travel time via its phase shift. (b) Homodyne detectors measure the correlation between the reflected sinusoidal signal and a reference signal with the same frequency, which produces a DC value as a function of the reference-signal phase and the phase shift from the scene.

3 CORRELATION TOF IMAGING

In this section we review the principles of correlation ToF imaging; for a detailed introduction, see [Lange 2000].

Image Formation. Correlation ToF cameras start by sending an amplitude-modulated light into the scene

$$p(t) = \alpha \cos(\omega_p t) + \beta, \tag{1}$$

where ω_p is modulation frequency, α is amplitude, and β is a DC offset. After traveling through the scene and reflecting off a target, the measured return signal

$$\tilde{p}(t-\tau) = \tilde{\alpha} \cos(\omega_p t - \phi) + \tilde{\beta}, \quad \phi = 2\pi\omega_p \tau \tag{2}$$

is a time-delayed $p(t)$ by time τ with an observed attenuation in amplitude $\tilde{\alpha}$, a shift in bias $\tilde{\beta}$, and a time-dependent phase shift ϕ. This measured signal is then correlated with a reference

$$r(t) = \cos(\omega_r t + \psi) + 1/2, \tag{3}$$

where ω_r and ψ are the demodulation frequency and phase, respectively. In existing multi-bucket imagers, this correlation occurs during exposure via photonic mixer device pixels [Foix et al. 2011; Lange and Seitz 2001], which are modulated according to the reference function $r(t)$. When we modulate and demodulate at the same frequency, that is $\omega_p = \omega_r = \omega$, this is called *homodyne* imaging. Integrating this signal over exposure time T, we get a correlation measurement

$$C_\psi = \int_0^T \tilde{p}(t-\tau) r(t)\, \mathrm{d}t = \frac{\tilde{\alpha}}{2} \cos(\psi - \phi) + TK, \tag{4}$$

where K is a general constant offset, meant to model a non-zero modulation bias on the sensor. Given this measurement, we aim to estimate the phase delay ϕ from which the scene depth can be computed. As illustrated in Fig. 2 (b), the correlation measurement C_ψ is a constant dependent on the demodulation phase offset ψ (achieving its max at $\psi = n\phi, n \in \mathbb{N}$). In practice, this means we never have to explicitly sample $\tilde{p}(t-\tau)$, which would require expensive ultrafast detectors and modulation electronics. Although the

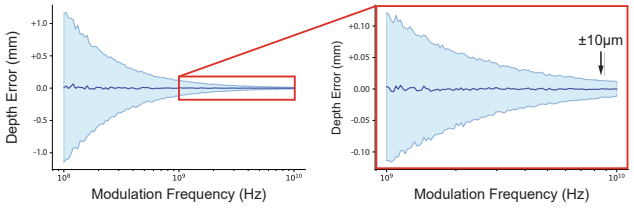

Fig. 3. Illustration of depth estimation error versus modulation frequency. For each frequency we simulate 1000 samples with added Poisson-Gaussian noise of constant magnitude for an indoor scenario. We quantize the simulated measurement to 14 bits (mimicking a 14-bit digital-to-analog conversion), reconstruct the estimated depth, and plot the resultant mean measurement and standard deviation envelope. We see that as we increase modulation frequency from 100MHz to 10GHz, our expected precision similarly increases 100×.

correlation measurement C_ψ does not directly give us access to the true phase ϕ, by sampling this function for multiple demodulation phase offsets ψ we can make use of Fourier analysis to discern the true phase ϕ. Existing correlation imagers typically acquire four equally-spaced correlation measurements at $\psi \in [0,\ \pi/2,\ \pi,\ 3\pi/2]$. Using these, we can estimate the phase offset $\hat{\phi}$ wrapped to the 2π range as $\hat{\phi} = \arctan\left(\frac{C_\pi - C_{\pi/2}}{C_0 - C_{3\pi/2}}\right)$. Phase unwrapping amounts to estimating the integer factor n to recover the unwrapped phase $\phi = \hat{\phi} + 2\pi n$. If successful, we can convert this phase estimate ϕ to depth as $z = \phi c / 4\pi\omega_p$, where c is the speed of light.

Modulation Frequency. As we noted earlier in Eq. (2), the round-trip path of the amplitude-modulated illumination imparts on it a ϕ phase shift. Setting $t = 0, \tilde{\beta} = 0$ and $\omega = 100MHz$ (a common modulation frequency in conventional ToF cameras) in Eq. (2), we observe a 0.0009% signal difference for a 1mm change in depth z. See Fig. 3. This means, with realistic imaging noise and quantization in existing sensors, we would practically not be able to discern millimeter scale features on object surfaces for a setup with this modulation frequency. To achieve higher precision we go to higher frequency, the same experiment repeated for $\omega = 8GHz$ leads to a more detectable 5.6% difference in signal amplitude. In practice, there are many factors that affect signal contrast, which we explore in the remainder of this work.

4 OVERVIEW

Realizing correlation imaging at two orders of magnitude higher frequencies than existing systems is prohibited by two technical challenges: modulating at GHz rates, and unwrapping the measured phase estimates, see Fig. 4. Stable GHz demodulation is challenging as analog modulation after photo-conversion or with fiber-coupling suffers from the high noise of ultra-fast photodiodes or large coupling losses, respectively. Phase unwrapping becomes a challenge as the increase in modulation frequency results in multiple dozens of wraps instead of a handful of wraps. The proposed computational imaging system tackles both limitations as follows.

First, we present a convolutional network for high-frequency phase unwrapping, motivated by recent learning-based segmentation methods. Our approach represents wrap counts as class labels

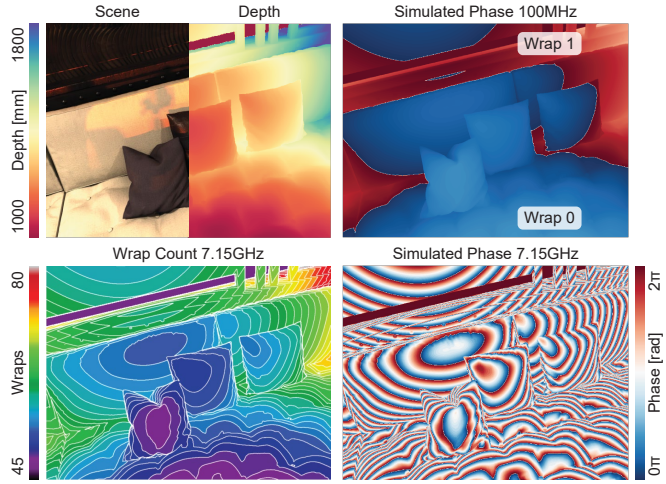

Fig. 4. Simulated measurements for a 100MHz ToF system, which exhibits only a single phase wrap, and a 7.15GHz system which experiences 35.

and segments measurements into their corresponding wrap regions, wherein we exploit the fact that proximal measurements that are highly correlated are likely to also be similarly phase wrapped.

Second, we introduce a two-pass EOM-based system with frequency doubling to tackle the problem of GHz frequency intensity modulation. The proposed method performs correlation computation *optically in free-space* rather than in the conventional analog domain. In this way, we avoid photo-conversion artifacts and energy loss from fiber-coupling, enabling high modulation contrast ToF imaging at 7.15 GHz and 14.32 GHz.

5 NEURAL PHASE UNWRAPPING

Phase unwrapping methods estimate the wrap count n of the wrapped phase $\hat{\phi}$ to recover the unwrapped phase ϕ for depth estimation. Our GHz ToF system presents two main challenges for unwrapping. First, the high modulation frequencies (7.15GHz and 14.32GHz) result in dozens of wraps over meter-scale scenes, as opposed to one or two for conventional MHz systems, see Fig. 4. Second, the modulation bandwidth of our GHz correlation ToF system is limited to ±10MHz, limiting the available sets of frequencies for multi-frequency approaches [Gutierrez-Barragan et al. 2019]. These challenges lead to lackluster performance from prior phase-unwrapping approaches including analytical solutions [Xia and Wang 2007], kernel methods [Lawin et al. 2016], and newer neural-network designs [Su et al. 2018; Zhang et al. 2019]. Here, we present a novel segmentation-inspired neural network tailored for high-frequency phase unwrapping. Rather than synthesizing the unwrapped phase directly, we *pose this as an ordinal classification problem of wrap counts.* Our network outputs N class weights for each input pixel, each corresponding to a candidate wrap count. Here, N is determined by the minimum and maximum expected wrap counts for the lowest modulation frequency, 7.15 GHz, to reduce class count.

5.1 Segmentation-based Fourier Phase Unwrapping

For our architecture, we modify the Fast SCNN [Poudel et al. 2019] image segmentation network. First, to encourage the network to learn local frequency unwrapping, rather than overfitting to global

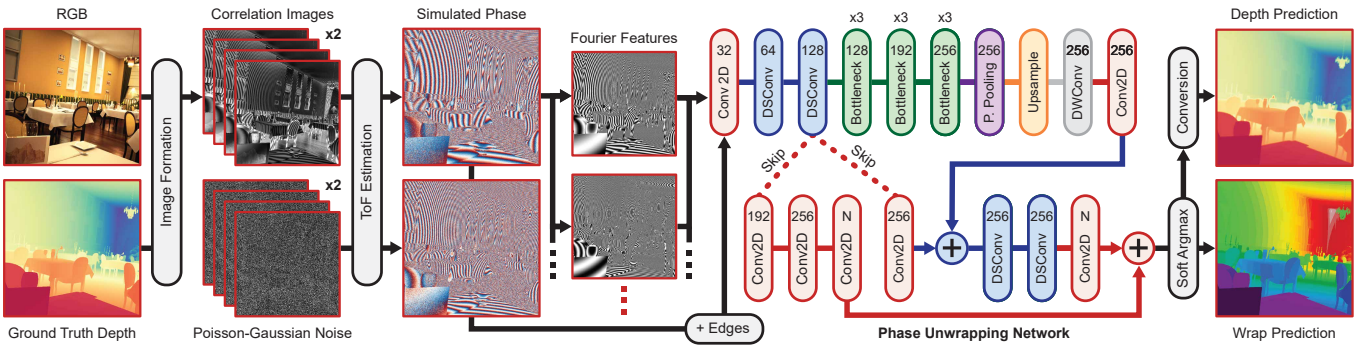

Fig. 5. We pose the phase unwrapping problem as an ordinal classification problem, and train a neural network to map phase and phase edge measurements to wrap counts. We use Fourier feature encodings of the phase measurements to allow the network to perform a rudimentary frequency analysis of the underlying ToF signal, and add realistic levels of Poisson-Gaussian noise to the simulated data to promote the learning of noise-robust unwrapping.

scene structure, we reduce its receptive field and add a full resolution skip layer directly to the output. We refer to the Supplemental Material for details on the network architecture. Second, as input to our network, in addition to measured amplitude, we use a Fourier feature encoding [Tancik et al. 2020] of wrapped phase $\hat{\phi}$

$$\gamma(\hat{\phi}) = [cos(2^0\hat{\phi}), sin(2^0\hat{\phi}), cos(2^1\hat{\phi}), ..., sin(2^{EC}\hat{\phi})]^\intercal. \quad (5)$$

This was used to great success in [Mildenhall et al. 2020] as a positional encoding method, mapping x,y,z coordinates to a higher dimensional space and improved training for their multilayer-perceptron representation. For our phase unwrapping network the purpose is two-fold. This encoding increases the dimensionality of the input multi-frequency measurements to facilitate learning of high-frequency features, and effectively modulates the correlation values with a new set of sinusoids, as seen in Fig. 5, which allows the network to perform a rudimentary frequency analysis of the underlying ToF signal.

5.2 Ordinal Classification Loss

We calculate our final estimate *unwrapped* phase ϕ' with

$$\phi' = \sum_{n=0}^{N-1} n \left(\frac{\gamma e^{\hat{\phi}_n}}{\sum_{m=0}^{N-1} e^{\gamma\hat{\phi}_m}} \right), \quad (6)$$

a differentiable argmax. Here ϕ_n is the predicted weight for phase class n, corresponding to n wraps, and γ adjusts the *hardness* of the argmax function. Predicted depth is as before, $z' = \phi' c/4\pi\omega$. This differentiable argmax allows for back-propagation through our phase-unwrapping network, meaning we are able to use both entropy-based classification losses on the output class weights and standard image losses on estimated phase or depth. Taking into consideration the ordinal nature of wrap counts – that is, predicting one wrap for a twice wrapped measurement is better than predicting twenty – we opt for a mixed cross-entropy $\mathcal{L}_{CE}$ and ℓ_1 loss $\mathcal{L}_{L1}$

$$\begin{aligned} \mathcal{L} &= \mathcal{L}_{CE} + w_{L1}\mathcal{L}_{L1} \\ \mathcal{L}_{L1} &= |z - z'| \\ \mathcal{L}_{CE} &= -\sum_{n=0}^{N-1} \phi_n log(\phi'_n), \end{aligned} \quad (7)$$

where z and ϕ are ground truth measurements. The cross-entropy loss allows us to train the network as a classifier, while the smooth ℓ_1-term provides a distance metric for the classes, penalizing the network for guessing wrap counts n' far from the true n.

5.3 ToF Simulation from RGBD

Given that there do not exist GHz ToF datasets, especially not ones with associated ground truth, we look to simulation to fill our need for training data. We simulate our measurements from the Hypersim RGB-D dataset [Roberts and Paczan 2020], containing 77,400 ground truth depth maps z (in mm) and images I from 461 computer-generated indoor scenes. We first calculate ground truth phase as $\phi = (z4\pi\omega)/c$, where $\omega \in \{7.15\text{GHz}, 14.32\text{GHz}\}$, and c is the speed of light. We simulate ToF correlation images C_ψ for $\psi \in \{0, \pi/2, 3\pi/2, \pi\}$ as

$$C_\psi = GTI_g(0.5 + cos(\phi + \psi)/\pi) + \eta_P + \eta_G, \quad (8)$$

where G is sensor gain, T is integration time, and I_g is the green channel of the image, meant to emulate the green laser in the experimental prototype. To simulate measurement fluctuations, we add Poisson noise η_P and Gaussian noise η_G with mean μ and standard deviation σ. We note that a typical correlation ToF camera follows a Skellam-Gaussian noise model [Hansard et al. 2012], however our all-optical correlation ToF design has no photon bucketing and subsequently encounters only Poisson-Gaussian noise.

5.4 Correlation Images to Wrapped Phase and Amplitude

From the correlation images C_ψ obtained either from the training dataset or our GHz ToF imaging system, we recover the wrapped phase $\hat{\phi}$ and amplitude $\hat{a}$ using a per-pixel Fourier transform

$$\hat{\phi} = \text{angle}(\mathcal{F}_2(C_\psi)), \hat{a} = 2|\mathcal{F}_2(C_\psi)|, \quad (9)$$

where angle$(\cdot)$ is the phase angle of a complex number and $\mathcal{F}_i(\cdot)$ is the i-th complex value of the Fourier transformed signal (e.g.$\mathcal{F}_0$ is the DC component). This process is repeated for the two modulation frequencies, 7.15 GHz and 14.32 GHz, and the arrays are stacked to form the raw multi-frequency measurements. As a result of the above Fourier recovery, $\hat{\phi} \in [0, 2\pi]$ is phase wrapped, and is passed into our phase-unwrapping network.

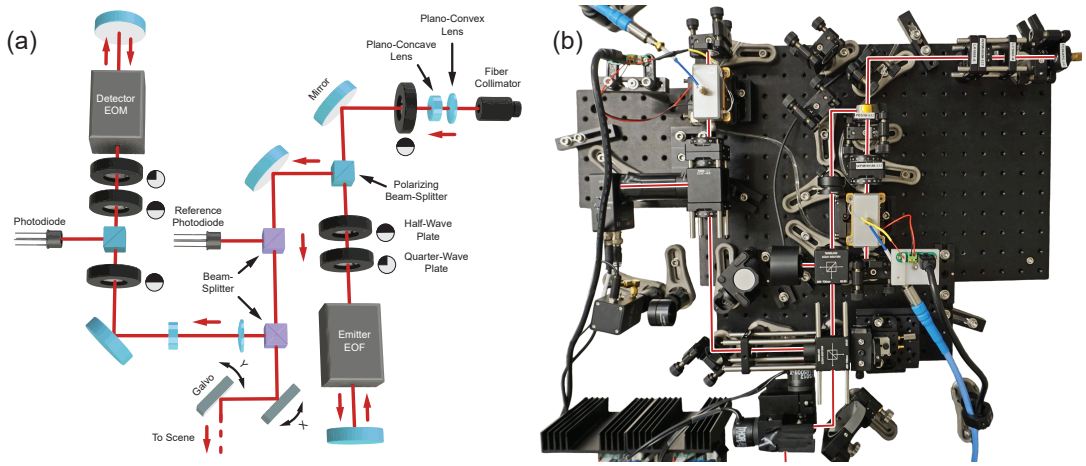

Fig. 6. We propose an all-optical free-space correlation ToF imaging system using polarizing optics and resonant EOMs. (a) The schematic diagram of our imager as realized in (b) shows the light path from a laser source to a scene, and back to the photodiode. See text for details.

6 ALL-OPTICAL GHZ CORRELATION TOF IMAGING

GHz modulation frequencies for correlation ToF can allow for high-precision depth imaging as illustrated in Fig. 3. However, practically realizing this idea has been challenging due to the limited photon absorption depth in silicon [Lange and Seitz 2001] and inefficacy of fiber coupling [Kadambi and Raskar 2017]. In this work, we take a different approach by co-opting free-space EOMs, mainly used for optical communication, and introducing a two-pass intensity modulation system with polarizing optics. Our method *optically* performs correlation computation, and, as such, permits the use of intensity sensors and continuous-wave lasers as compared to the more complex hardware requirements of pulsed LiDAR.

6.1 Backgrounds on EOM and Jones Calculus

We briefly review EOM and Jones calculus before describing our novel two-pass intensity modulation scheme. Modern EOMs modulate the phase, amplitude, and polarization of light by applying an electric field to control the refractive indices of a bulk crystal, perpendicular to the direction of light propagation, according to the electro-optic Pockel's effect [Yariv 1967]. To mathematically model the effect of an EOM, we rely on a Jones vector and Jones matrix. The Jones vector is a 2×1 vector that describes the amplitude and phase of horizontal and vertical polarization components. As such, the corresponding Jones matrix describes the change of the polarization state of light with a 2×2 matrix that can be multiplied by a Jones vector. We refer the reader to Collet [2005] for a review on Jones calculus. The Jones matrix describes how an EOM shifts the horizontal and vertical polarization waves of light by an amount dependent on the applied voltage V as

$$B(V) = \begin{bmatrix} e^{-iV/2} & 0 \\ 0 & e^{iV/2} \end{bmatrix}, \quad V = \eta \cos(\omega t - \phi), \tag{10}$$

where V is a time-varying voltage function, η is the modulation power, ω is the voltage modulation frequency, and ϕ is the modulation phase. Our custom-designed resonant EOM is capable of generating phase differences in two orthogonally-polarized components of light at a 7.15 GHz frequency with 20 MHz of bandwidth. See Supplemental Material for additional information on our EOM.

6.2 Two-pass GHz Intensity Modulation

For GHz ToF imaging, we propose a two-pass GHz intensity modulation method which uses polarizing optics and custom phase-modulating EOMs. This allows us to achieve a 7.15 GHz modulation frequency at the EOM's native resolution as well as a higher frequency 14.32 GHz enabled by intensity modulation with a combination of two-pass phase modulation and polarization changes. The two modulation frequencies provide high depth resolution, with an effective wavelength of 2.1 cm for the 14.32 GHz frequency, and overcome the small native bandwidth of 20 MHz of the EOMs, resulting in a frequency difference of 7.17 GHz for phase unwrapping. We note that our method enables high-frequency EOM-based *intensity* modulation, an entirely different concept than conventional optical-frequency doubling using EOMs. We describe the working principle of our method below.

6.2.1 Intensity Modulation. Our custom resonant EOM delays the phase of horizontal and vertical components of light at a frequency $\omega = 7.15\,GHz$. We exploit these polarization-dependent phase shifts to perform intensity modulation of incident light. Specifically, we use the following polarization optics: a polarizing beamsplitter (PBS), a half-wave plate (HWP), a quarter-wave plate (QWP), and a mirror as shown in Fig. 7. Incident light enters a PBS, turning light into vertical linear polarization as

$$E_0 = A \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \tag{11}$$

where A is the amplitude of the incident light. The polarization state of the light is then modulated by a HWP and a QWP followed by a EOM at a given voltage V as

$$E_1 = B(V)Q(\theta_q)H(\theta_h)E_0, \tag{12}$$

where $H(\theta_h)$ and $Q(\theta_q)$ are the Jones matrices of the HWP and the QWP oriented at angles θ_h and θ_q. The light then propagates in free-space for half a modulation wavelength c/ω to where a mirror is placed, resulting in the change of Jones vector as

$$E_2 = ME_1, \tag{13}$$

where M is the Jones matrix of a mirror. Light travels again back to the EOM, the QWP, and the HWP. The PBS picks up the vertical linear polarization component of this light. Setting the HWP and the QWP angles as $\theta_q = 11.25°$ $\theta_q = 45°$, we obtain the output light

$$\begin{aligned} E_3 &= L_h H(-\theta_h)Q(-\theta_q)B(V)MB(V)Q(\theta_q)H(\theta_h)E_0 \\ &= A\begin{bmatrix} \frac{i(\cos V - \sin V)}{\sqrt{2}} \\ 0 \end{bmatrix}. \end{aligned} \tag{14}$$

We square the magnitude of E_3, and observe a modulated intensity

$$I(V) = |E_3|^2 = \frac{A^2}{2}(1 - \sin 2V). \tag{15}$$

Eq. (15) indicates that the output intensity of light is a function of the voltage V applied to the EOM. As we supply a time-varying sinusoidal voltage to the EOM as in Eq. (10), we arrive at the time-varying intensity-modulated light as

$$\begin{aligned} I(t) &= \frac{A^2}{2}(1 - \sin(2\eta \sin(\omega t - \phi))) \\ &\approx \frac{A^2}{2}(1 - 2\eta \sin(\omega t - \phi)). \end{aligned} \tag{16}$$

The last approximation is based on the Taylor expansion given that the modulation power η of our EOM is small. The applied voltage to the EOM has GHz modulation frequency $\omega = 7.15\,GHz$, enabling effective all-optical GHz modulation of light intensity. We refer to the Supplemental Material for detailed derivation.

Eq. (16) describes the high-frequency intensity modulation realized by the proposed free-space two-pass phase modulation with polarizing optics shown in Fig. 7. This optical configuration serves as a *building block for both illumination and detection modules* in our imaging system. In the illumination module, we input continuous laser light into the EOM, resulting in sinusoidal intensity-modulated light emitted into the scene. For the detection module, the returned amplitude-modulated light from the scene is demodulated by an additional intensity modulation with the reference signal r, recall Eq. (4), and we *optically* multiply r and $\tilde{p}$ before integration on the detector.

6.2.2 Doubled Intensity-Modulation Rate. Even though the voltage modulation frequency ω is limited to a narrow modulation band 20 MHz in our resonant EOM, we can modulate at the double frequency of 2ω by adjusting the angle of the HWP, θ_h, in front of the EOM, achieving 14.32 GHz modulation rate. While doubling the frequency of the optical carrier is well known in optics, we note that the proposed frequency doubling of the intensity modulation is novel. In the original operating mode, we set the HWP angle θ_h as 11.25° resulting in the intensity modulation at the original frequency ω. For frequency doubling, we rotate the HWP to $\theta_h = 22.5°$. To derive the modulation behavior, we rely on the same Jones calculus

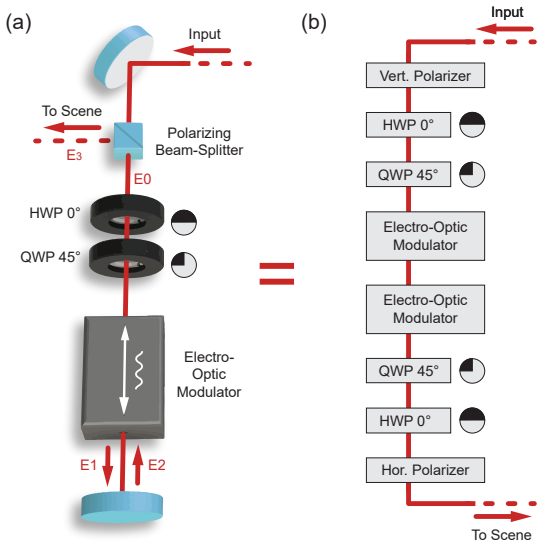

Fig. 7. Two-pass GHz intensity modulation with polarizing optics and an EOM. (a) We implement the GHz intensity modulation of incident light by using polarizing optics and a EOM. Incident light becomes linearly polarized after a PBS and further polarization modulated by a HWP and a QWP. An EOM with a sinusoidal voltage applied shifts phases of the horizontal and vertical polarization components. The light then is returned by a mirror distanced at half the modulation wavelength, and returns back to the EOM, the QWP, the HWP, and the PBS. The combination of forward and reverse paths results in (b) the optical intensity modulation of incident light at GHz frequency with unrolled polarization modulation, see the text for details.

from above. Specifically, changing the HWP angle θ_h results in the output light E_3 as

$$E_3 = A\begin{bmatrix} -i\sin V \\ 0 \end{bmatrix}. \tag{17}$$

The intensity $I(V)$ is the magnitude square of E_3 as

$$I(V) = |E_3|^2 = \frac{A^2}{2}(1 - \cos(2V)). \tag{18}$$

Note that the difference between Eq. (18) and Eq. (15) is that we have cos() instead of sin(). This single difference enables intensity modulation at a doubled frequency. After applying the time-varying voltage modulation of Eq. (10), the time-varying intensity of the output light is

$$\begin{aligned} I(t) &= \frac{A^2}{2}(1 - \cos(2\eta \sin(\omega t - \phi))) \\ &\approx \frac{A^2}{2}\eta^2 \sin^2(\omega t - \phi) \\ &= \frac{A^2}{4}\eta^2(1 - \cos(2\omega t - 2\phi)). \end{aligned} \tag{19}$$

We use the same Taylor expansion. Eq. (19) shows that we can obtain the doubled frequency 2ω at 1/4th amplitude compared to the single-frequency mode at ω – only by changing the polarization optics instead of the electro-optical modulation itself.

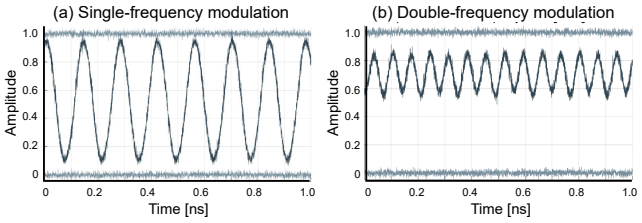

Fig. 8. We validate the GHz intensity modulation module by capturing the signal reflected from a mirror at a fixed position. (b) When the HWP is oriented at 11.25°, we achieve the single intensity modulation frequency of $\omega = 7.15$GHz. (c) Changing the HWP angle to 22.5° at the EOM base frequency $\omega = 7.16$ GHz enables frequency doubling: $2\omega = 14.32$GHz.

6.2.3 Validation of GHz Intensity Modulation. We validate our GHz intensity modulation module consisting of a PBS, a HWP, a QWP, a EOM, and a mirror. We emit laser light to a mirror at a fixed position and *directly* capture the intensity of the modulated light steered onto a GHz photodetector, see Supplemental Material for the measurement configuration. Fig. 8 demonstrates the effective GHz intensity modulation with high modulation contrast at two different HWP angles of 11.25° and 22.5°, corresponding to the modulation frequencies of $\omega = 7.15$ GHz and $2\omega = 14.32$ GHz.

6.3 Coaxial Spatial Intensity Imaging

Equipped with the intensity modulation block, we design a coaxial imaging system with an illumination and a detection module, see Fig. 6. For the illumination module, we opt for continuous-wave laser illumination at 532 nm (for eye-safe lab operation of the prototype) followed by a GHz intensity modulation block. A second GHz modulation block is used for the detection module, combined with an avalanche photodiode (APD) for intensity sensing. We employ an APD to allow for high gain at fast readout rates in low-flux scenarios, which is especially important for scene surfaces low reflectance and objects at long distances; see Supplemental Material. Using a non-polarizing beamsplitter, we share the same path for the output light to a scene and the detected light from a scene, improving the signal-to-ratio of the system. For 2D spatial scanning, we use a 2-axis galvonometer in front of the beamsplitter, shown in the bottom of Fig. 6. Although the proposed free-space modulation method is not limited to co-axial scanning, the beam-steered acquisition effectively eliminates most multi-path interference, which we neglect in the remainder of this work.

Analog Signal Integration. We use a conventional avalanche photodiode with a gain G to detect the correlation signals from the detection module without any quantization involved. This generates analog photocurrent which is then low-pass filtered with an electrical filter and a resistor-capacitor (RC) circuit that further integrates the constant correlation input signal over an exposure time T. We read out the analog signal with an ADC with 14bit quantization. This results in the digital read of

$$C_\psi = Q\left(G\int_0^T \tilde{p}(t-\tau)r(t)\,\mathrm{d}t\right) = Q\left(\frac{\tilde{\alpha}}{2}\cos(\psi-\phi)+TK\right), \quad (20)$$

where Q is the 14bit ADC quantization, ϕ is the illumination phase, and ψ is the phase of the reference r shown in Eq. (4).

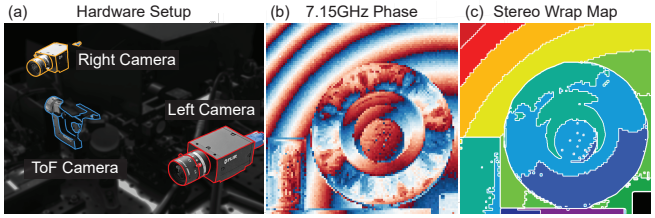

Fig. 9. Active stereo setup to generate ground truth depth for fine-tuning. (a) Stereo cameras mounted to the ToF system. (b) Sample 7.15 GHz ToF phase map. (c) Corresponding wrap map recovered from stereo depth.

6.4 Fine-tuning with Active Stereo Supervision

We fine-tune our phase-unwrapping network to allow for specialization to the specific noise characteristics of our experimental system and minor deviations of the modulation functions from ideal sinusoids. We acquire pseudo-ground-truth phase wrap maps by augmenting our system with stereo cameras (other ground-truth acquisition approaches are also possible, we chose stereo for ease of implementation). We build the acquisition system by mounting two CMOS cameras (FLIR Grasshopper3 GS3-U3-32S4C) to the ToF rig, with 8 mm lenses to match our system's FOV as shown in Fig. 9. This effectively creates an auxiliary active stereo system from which to recover coarse scene depth without additional captures. After geometric calibration, we triangulate the position of the ToF laser spot in 3D space with the stereo cameras as we scan the scene. The estimated depth for each laser spot allows us to generate a pseudo ground truth wrap map. We perform fine-tuning on a diverse dataset of captured stereo measurements, which are all *withheld from the experimental validation* section. For details on the fine-tuning, we refer to the Supplemental Document.

7 EXPERIMENTAL SETUP

For experimental validation, we implement the prototype system shown in Fig. 2. While we assemble the experimental prototype on an optical breadboard, we note that the EOMs and optics can be integrated in a small form factor similar to LiDAR sensors.

7.1 Illumination Module

We use a single transverse mode continuous-wave laser at 532 nm wavelength (Laser Quantum Gem 532). The laser beam is coupled with a custom-design single mode high power optical fiber (OZ Optics QPMJ-A3AHPCA3AHPC-488-3.5/125-3AS-1-1) which removes the higher order modal light and produces a uniform Gaussian beam at the output of the fiber, maintaining 20 – 30% laser output power. The light then enters a 2.5× inverse beam expander (Thorlabs LC1060-A and LA1608-A) that reduces the beam diameter from 1.25 mm to 0.5 mm, matching the desired beam size for our EOM. The reduced light becomes horizontally linearly polarized by passing through a first PBS (Thorlabs PBS101). Then, a pair of HWP (Thorlabs WPMH05M-532) and QWP (Thorlabs WPQ05M-532) modulates the polarization state of the beam. The polarization-modulated light passes through our custom GHz EOM that operates at the modulation frequency ω. The light is reflected by a mirror (Thorlabs PF10-03-P01), returning back to the EOM, the QWP, the

HWP, and the PBS. This procedure results in the GHz-frequency intensity modulation of light.

The light then passes through a mirror (Thorlabs PF10-03-P01) and a NBS (Thorlabs CCM1-BS013) dividing the incident beam into two beams of equal intensity. One beam is directed to an integrating sphere (Thorlabs S140C) which measures the intensity of emitted light for calibration purposes and the other beam passes through another NBS (Thorlabs CCM1-BS013). The purpose of this module is to calibrate intensity fluctuations from the laser by normalizing the signal incident on the detection module. The optical intensity modulation has higher frequency than the integration time of a few milliseconds, which allows compensation after the modulation without error. It splits the beam again into two paths with equal intensity where one half of the beam is used as the reference beam in interferometric measurement mode (used for precision comparison see Fig. 13) with a mirror; otherwise this beam is discarded in intensity-measurement mode. The other half of the beam is sent to a scene through a mirror (Thorlabs PF30-03-P01) and a 2-axis galvo mirror system (Thorlabs GVS012) for spatial scanning. The emitted CW laser power is 3 mW. For photon efficiency estimates, see Supplemental Document. To avoid speckle artifacts of the coherent laser illumination, we slightly defocus the projected beam.

7.2 Detection Module

The intensity-modulated light returns from a scene and passes through the galvo mirror system and the mirror followed by a NBS which redirects the beam to the detection module. We use an $1.6\times$ inverse beam expander (Thorlabs LA1213-A and Thorlabs LC1060-A) and a mirror (Thorlabs PF10-03-P01) resulting in a beam diameter of 0.5 mm and collimated beam accurately entering the detection EOM. Symmetric to the emission module, we mount a PBS, a HWP, a QWP, an EOM, and a mirror which in effect optically demodulate returned light from the scene. The intensity demodulated light is then captured by an avalanche photodiode (Thorlabs APD440A) with a focusing lens (Thorlabs LA1951-A). We use a 10 kHz lowpass filter (Thorlabs EF120) resistor capacitor (RC) low pass integrator circuit with RC time constant $t_{RC} = 100ms$ to integrate the detected photocurrent signal, then passed into an analog-digital-converter (LabJack T7) to sample the signal at up to 24K samples per second. We integrate over 20 samples for a single phase measurement and sample 16 phases corresponding to 13ms integration time a single galvo measurement point.

7.3 RF Driver

To operate the EOMs with a sinusoidal voltage input, we use two custom RF drivers with a high-frequency DDS which are synchronized to the external clock source of a function generator (Siglent SDG2042X). Our DSS signal generators support sinusoidal modulation only, leaving non-sinusoidal modulation as interesting future work [Gupta et al. 2018]. The external clock enables accurate control of the phase of the modulation signal ϕ. Our driver contains two RF modulators to output an RF signal provided to the EOMs. The RF driver performs frequency locking to significantly increase the output power and reduce frequency drifting in the EOM. For further details, refer to the Supplemental Document.

7.4 Comparison to Analog RF Demodulation

For comparison of the proposed system with demodulation of a signal after photo-conversion, we add a highspeed GaAs 12GHz photodetector (EOT GaAs PIN Detector ET-4000) connected to an RF demodulation circuit. This measurement setup can be enabled by flipping a flip-mirror in the optical path, redirecting the scene illumination to the fast photodiode instead of the proposed detection module. This photodiode offered the highest photon-detection efficiency and high-frequency response available to us. The captured photocurrent from the detector is sent as input to an I/Q demodulator consisting of analog microwave electronics as follows. The photodector signal is first amplified and band-pass filtered. Then it enters an RF mixer to be demodulated with the local oscillator (LO) signal from the RF driver. This produces a signal with the difference of the two frequencies and a signal with the sum of the two frequencies. These signals are passed through a low pass filter which removes the higher frequency signal. Then the remaining homodyne DC signal, is output as two signals, an in-phase component I, and a quadrature component Q shifted by 90 degrees. For a detailed circuit design, see the Supplementary Document.

8 ASSESSMENT

In this section, we validate the proposed computational ToF method in simulation and experimentally. Specifically, we first perform quantitative evaluation of our neural unwrapping approach on a synthetic dataset and compare with other baseline phase unwrapping methods. We then experimentally validate the proposed system quantitatively and qualitatively on unseen real-world measurements captured by our experimental prototype.

8.1 Simulated Analysis

Ablation Study. We conduct an ablation study to validate our choice of Fourier feature encoding and combined loss function. The different ablation configurations and corresponding quantitative results are shown in Tab. 1, and we refer to Supplemental Document for qualitative results. We observe that Fourier encoding leads to a 10 percentage point boost in correct wrap predictions, supporting the theory that the doubly modulated phases provide valuable features during training, possibly in the form of a learned frequency analysis of the underlying measurements. We concatenate phase edges to the input in order to aid the network's understanding of each wrap region. For the loss functions, we find the model trained on cross-entropy loss alone demonstrates competitive results, validating the choice to represent phase unwrapping as a classification problem. However, when we make use of the differentiable argmax function to directly introduce ℓ_1 loss on predicted depth, we see a reduction in outliers and an overall smoother final prediction. This reinforces the problem as ordinal classification, where the ordering of classes — in this case wrap counts — is significant.

Comparison to Phase Unwrapping Methods. We validate our proposed neural unwrapping approach on a synthetic test set and discuss the qualitative and quantitative results. As a baseline, we compare our work against traditional unwrapping methods including the approach used in phasor imaging [Gupta et al. 2015], the algebraic chinese-remainder theorem (CRT) solution [Pei et al. 1996;

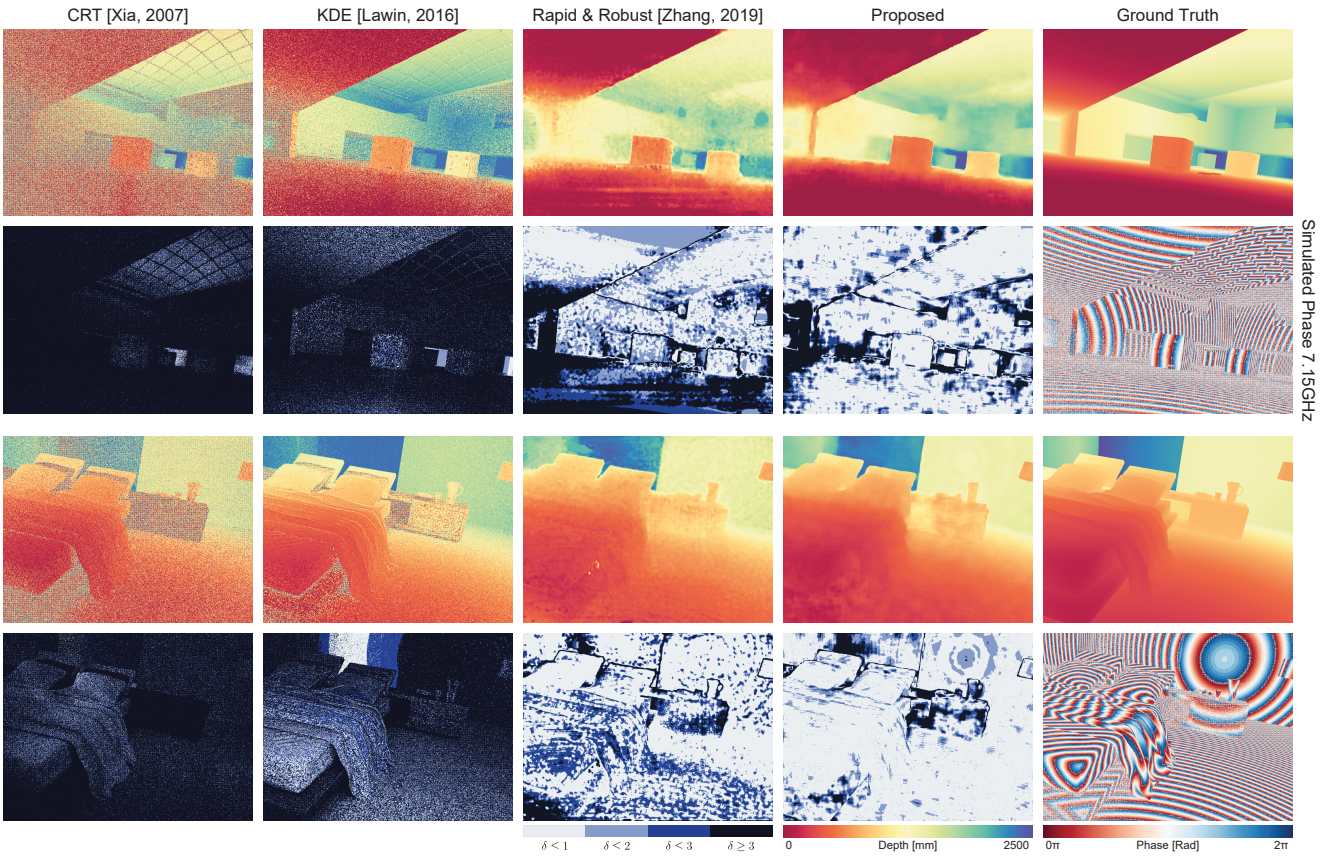

Fig. 10. Phase unwrapping results for comparison to existing conventional and learned methods and our proposed approach. Analytic solutions of CRT [Xia and Wang 2007] and KDE [Zhang et al. 2019] suffer from rapid phase wraps and phase noise. The state-of-the-art neural network method partly overcomes such problems at the cost of smoothed geometry and low-frequency depth artifacts. Our method outperforms the previous methods by recovering both accurate scale and geometric details. Error map below results corresponds to a visual representation of the δ metric, see text, in Tab. 2 and 1.

	Input			Loss		Performance (%)			
	$\hat{\phi}$	$\gamma(\hat{\phi})$	$\lvert\nabla\hat{\phi}\rvert$	$\mathcal{L}_{CE}$	$\mathcal{L}_{L1}$	$\uparrow \delta < 1$	$\uparrow \delta < 2$	$\uparrow \delta < 3$	$\downarrow \delta \geq 3$
Proposed	✓	✓	✓	✓	✓	**51.6%**	**69.1%**	**77.0%**	**23.0%**
$\mathcal{L}_{CE}$ Only	✓	✓	✓	✓	-	49.1%	65.0%	73.3%	26.7%
F. Features	✓	✓	-	✓	✓	40.2%	59.6%	68.8%	31.2%
Phase Only	✓	-	-	✓	✓	30.3%	52.3%	65.7%	34.3%

Table 1. Ablation study configurations and corresponding quantitative results. Here the δ metric represents the percent of pixels whose prediction is δ wraps from ground truth wrap count. Up arrow denotes "higher is better", down arrow means "lower is better".

Method	$\uparrow \delta < 1$	$\uparrow \delta < 2$	$\uparrow \delta < 3$	$\downarrow \delta \geq 3$	$\downarrow \delta \geq 10$
Phasor [2015]	0.74%	1.66%	3.50%	96.5%	84.4%
CRT [2007]	9.29%	14.7%	19.7%	80.3%	56.0%
KDE [2016]	9.46%	18.56%	27.0%	73.0%	8.93%
One-Step [2019]	19.9%	37.6%	52.2%	47.8%	14.6%
U-Net [2015]	21.8%	45.6%	64.4%	35.6%	10.0%
Deep-ToF [2018]	20.1%	47.5%	67.6%	32.4%	8.40%
Rapid. [2019]	23.1%	45.4%	61.1%	38.9%	9.74%
Proposed	**51.6%**	**69.1%**	**77.0%**	**23.0%**	**7.59%**

Table 2. Quantitative comparison table for proposed neural phase unwrapping method and baselines, as evaluated on the synthetic test scenes. $\delta \geq 10$ metric added to better quantify outlier performance.

Xia and Wang 2007], and the kernel density method (KDE) [Lawin et al. 2016], which is also used in the Kinect V2 software. We also compare to an unmodified U-Net [Ronneberger et al. 2015] baseline and three recent regression-based deep learning approaches [Su et al. 2018; Wang et al. 2019; Zhang et al. 2019].

We show qualitative performance of our proposed neural unwrapping method and baseline methods including CRT [Xia and Wang 2007], KDE [Lawin et al. 2016] and the next best network-based method [Zhang et al. 2019] in Fig. 10, and refer to the Supplemental Document for additional qualitative comparisons. Tab. 2 presents quantitative classification results for the full range of methods in increasingly widening error bands, as well as outlier percentages. Visually CRT and KDE achieve similar results, as they have similar underlying mechanics for wrap calculation, however KDE's spatial aggregation allows it better tackle noise and make significantly more correct estimates. This is quantitatively confirmed by the fact that more than half of CRT's predictions are outliers ($\delta \geq 10$) while for

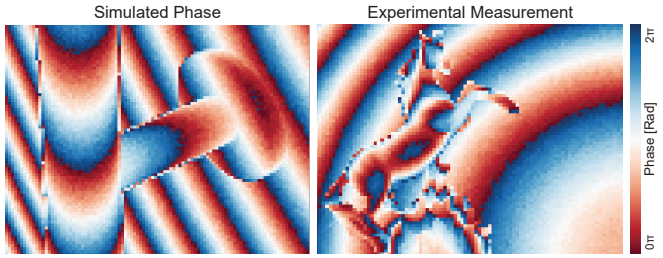

Fig. 11. Noise matching example. Left: simulated synthetic data. Right: experimental measurement.

KDE this number is less than 9%. The last conventional method, phasor imaging, struggles heavily under added noise and sub-optimal modulation frequencies, leading to nearly all the measurements being incorrectly unwrapped. The U-Net and three comparison deep learning methods produce similar spatially smoother predictions than the classical methods, however often bin entire patches of the image into the wrong wrap count, leading to a marginally higher rate of outliers than KDE while making more than double the number of correct predictions. Deep-ToF [Su et al. 2018] solves the phase unwrapping problem by directly regressing the raw correlation measurements to depth using the modified U-Net with skip connections. This regression-based method often results in globally inaccurate phase unwrapping as it hallucinates depth directly without the exact meaning of wrap counts. In contrast, our unwrapping network avoids this problem by estimating wrap counts with the segmentation-inspired network architecture and loss functions. The proposed neural unwrapping method more than doubles the rate of correct predictions when compared to Zhang et al.'s [2019] baseline. This is confirmed qualitatively in Fig. 10 with spatially-consistent outputs and object discontinuities that accurately align with the amplitude measurement. The proposed network outperforms all existing methods in GHz frequency unwrapping.

Impact of Measurement Noise. In addition to simulating the Poisson-Gaussian noise as described in Section 3, we further test our method with two different measurement distortions. First, we simulate global phase offsets between the two high modulation frequencies. Note that we used a phase-accurate clock (SDG2024X) to mitigate phase offsets, however, high GHz modulation frequencies can make the system sensitive to the phase shifts. We add ±0.01 radians global offset to one of the phase measurements. Tab. 3 shows that our neural phase unwrapping is robust to such phase shift, resulting in minor performance drop of less than 1% for all metrics. Second, we simulate a high noise level $\sigma = 2000$ instead of 1200 to mimic low-signal scenario with strong ambient light present. Again, we obtain a minor performance drop for 1.6× higher noise level than the training as shown in Tab. 3.

Training Details. For all methods except phasor imaging we simulate measurements for two modulation frequencies, a fundamental 7.15GHz signal and a shifted plus frequency doubled (7.15GHz + 10MHz) × 2 = 14.32GHz signal. We note that these frequencies correspond to the frequencies we can implement in the experimental setup. For the phasor imaging method, we input 7.15GHz and

Environment	$\uparrow \delta < 1$	$\uparrow \delta < 2$	$\uparrow \delta < 3$	$\downarrow \delta \geq 3$
Conventional	**51.6%**	**69.1%**	**77.0%**	**23.0%**
With Phase Offset	51.0%	68.9%	77.0%	23.0%
With Ambient Light	42.7%	60.4%	70.0%	30.0%

Table 3. Quantitative results for our phase unwrapping method against phase offset error and higher noise level due to the ambient light.

7.16GHz simulated measurements, as these are the locally optimal feasible shifts achieved by the optical amplitude modulation system. To be robust against real-world noise, we simulate measurements with sensor gain $G = 20$ and integration time $T = 1000$ms, with noise parameters $\mu = 0, \sigma = 1200$. The models are trained for 1000 epochs each, with 500 samples drawn per epoch, each consisting of a 512×512 image and ground truth depth patch (sampled randomly from the full RGB-D datum). We use a OneCycle learning rate schedule with a ratio of 0.995 per epoch and an initial rate of $1e-3$; training on 3 Nvidia V100 GPUs with a batch size of 12 takes approximately 24 hours. The synthetic test set consists of the 42nd frame of each simulated scene, withheld from both the training and validation sets. We balance the losses by setting $w_{L1} = 0.1$, which leads to noticable improvements in smoothness without the classifier's early training behavior. During inference, running on one Nvidia V100 GPU, we achieve an average runtime of 16.5ms≈60FPS per image of size 256×256, and 50ms≈20FPS with the full synthetic image size of 768x1024.

8.2 Experimental Assessment

In this section, we validate the proposed computational ToF imaging approach on experimental scenes.

Qualitative Reconstruction. We demonstrate depth captures on diverse real-world scenes as shown in Fig. 16. All scenes were captured with the galvo on the floor plane with respect to the scene, and swept through 16 phase shifts from $0 - \pi$, corresponding to 13ms integration time for a single galvo measurement point. Note that we perform this capture procedure under strong room ambient light for all captured scenes, demonstrating the robustness to ambient illumination. Operating outside the visible range (our system uses 532 nm for lab eye safety reasons), and employing narrowband spectral filters can further enhance this robustness. We use a single-frequency 7.15 GHz and double-frequency 14.32 GHz pair for depth measurement. Fig. 16 shows that combining the proposed free-space correlation acquisition and neural unwrapping method enables high-fidelity depth reconstruction of all the tested objects with *wide dynamic range*.

Compared to RF demodulation after photon conversion, using the highspeed GaAs 12GHz photodetector as described in Section 7, the proposed method drastically outperforms post-detection modulation across all experimental tests for an identical photon budget. We tested even for a 10× higher laser power of 30 mW with the same result, again validating the photon-efficiency of the proposed free-space modulation approach. Our measured phase maps clearly show depth-dependent contours for diverse surface reflectance types (see also Supplemental Material), demonstrating the robustness of the proposed system. Moreover, our imager handles large variations

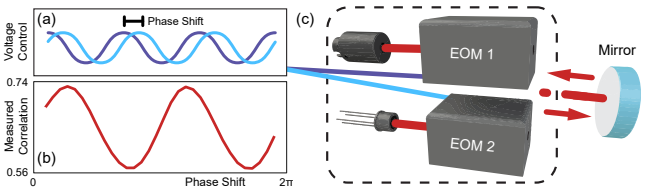

Fig. 12. (c) We validate our optical GHz modulation by capturing the instrument response of the complete system for a fixed mirror with varying phase shifts of the detection EOM, controlled by (a) the voltage-control RF drivers. (b) As predicted in the model, the amplitude measurements accurately follow a sinusoidal function, validating the effective GHz correlation mode.

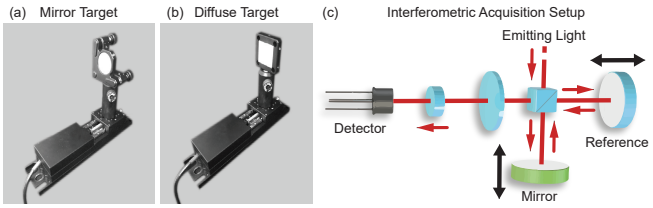

Fig. 13. We measure depth precision of our system for (a) a specular mirror and (b) a diffuse reflector mounted on a linear motion stage, placed at 60 cm distance from the system. Compared to the known positions of the target objects in a sweep of the mirror and diffuse reflector, we achieve a depth accuracy of 33.5 um with a standard deviation of 7.5 um, outperforming the post-photoconversion RF-based method and approaching (c) optical interferometry. Note that while interferometry is extremely sensitive to short travel distance and scene reflectance, the proposed method effectively estimates depth independently of these environmental influences.

in object reflectance. From a diffuse bust, a highly specular helmet with a very small diffuse component, to a textured owl object with low albedo components. We evaluate the impact of our neural phase unwrapping on these challenging scenes compared with the existing KDE [Lawin et al. 2016], recent learned network [Zhang et al. 2019] method, and micro ToF phasor unwrapping [Gupta et al. 2015] methods. KDE unwrapping [Lawin et al. 2016] struggles with the high frequencies of the proposed system and residual measurements noise, failing to provide high-quality residual measurements. The other two methods [Gupta et al. 2015; Lawin et al. 2016] also fail to recover correct meaningful geometric structures which can be found in the Supplemental Document. The lookup-table based unwrapping method from [Gupta et al. 2015] fails here due to the small modulation bandwidth available in our experimental system. We note that we use the optimal frequency settings for the phasor unwrapping [Gupta et al. 2015] in our operating bandwidth. Our neural phase unwrapping successfully handles high wrap counts in the GHz regime, enabling us to obtain accurate depth maps across all scenes. Thus, these experiments validate that the proposed method robustly performs high-frequency correlation depth imaging, outperforming existing approaches and phase unwrapping methods across all tested scenarios.

Validation of Correlation Profiles. We validate the functionality of the proposed imaging system by acquiring correlation measurements as figure of merit. Specifically, we capture measurements of a

Method	RMSE Mirror	MAE Mirror	RMSE Diffuse	MAE Diffuse
Interferometry	20 μm	20 μm	14 μm	14 μm
RF	49.5 μm	48.8 μm	11800 μm	11800 μm
Proposed	33.5 μm	32.5 μm	34.6 μm	32.9 μm

Table 4. Quantitative comparison of the proposed method for diffuse and specular objects corresponding to the measurements in Fig. 13.

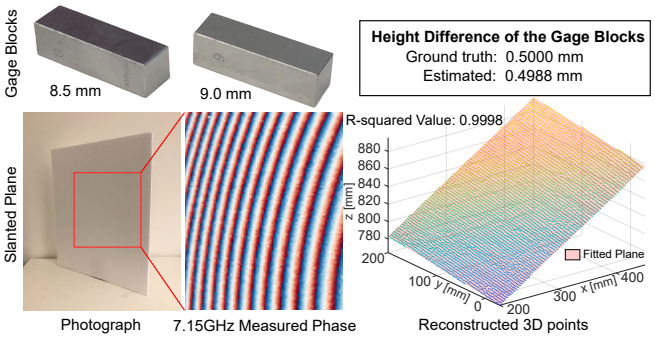

Fig. 14. As additional illustrative validation of the depth resolution provided by the proposed method, we capture gage blocks and a slanted flat plane. Our method accurately recovers the height difference of the gage blocks. For the plane object, we fit a plane equation to the acquired 3D points and achieves an R-squared value of 0.9998. This validates the effectiveness and precision of our method independent of phase unwrapping.

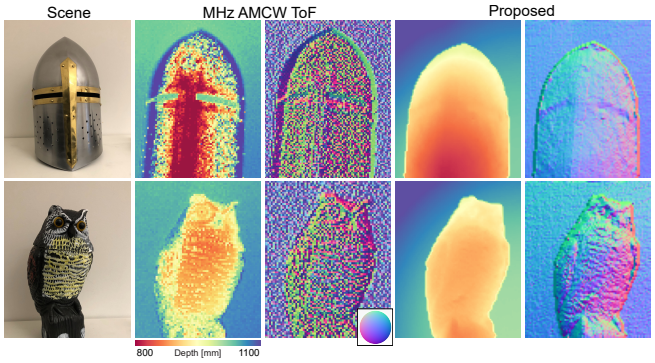

Fig. 15. Our all-optical GHz ToF imaging captures fine geometric details which cannot be revealed in MHz correlation ToF imaging. In the MHz regime, fundamentally limiting depth resolution and phase contrast by modulation in the analog domain instead of the optical domain, conventional correlation ToF fails to recover correct depth and fine-grained normals. Noticeably, we observe artifacts on the specular surface of the helmet, which only returns a very faint diffuse component, and the texture-dependent artifacts on the bright and dark spots of the owl object.

static target without galvo movement while sweeping the phase of the reference signal driven by the RF driver. To this end, we place a mirror (Thorlabs PF10-03-P01) at a fixed position and uniformly sample ψ over a range of 0 to 2π. Fig. 12 confirms that the measured correlation values accurately follow the sinusoidal image formation model from Eq. (20).

Quantitative Evaluation of Depth Precision. We quantitatively evaluate depth precision of our experimental prototype by capturing

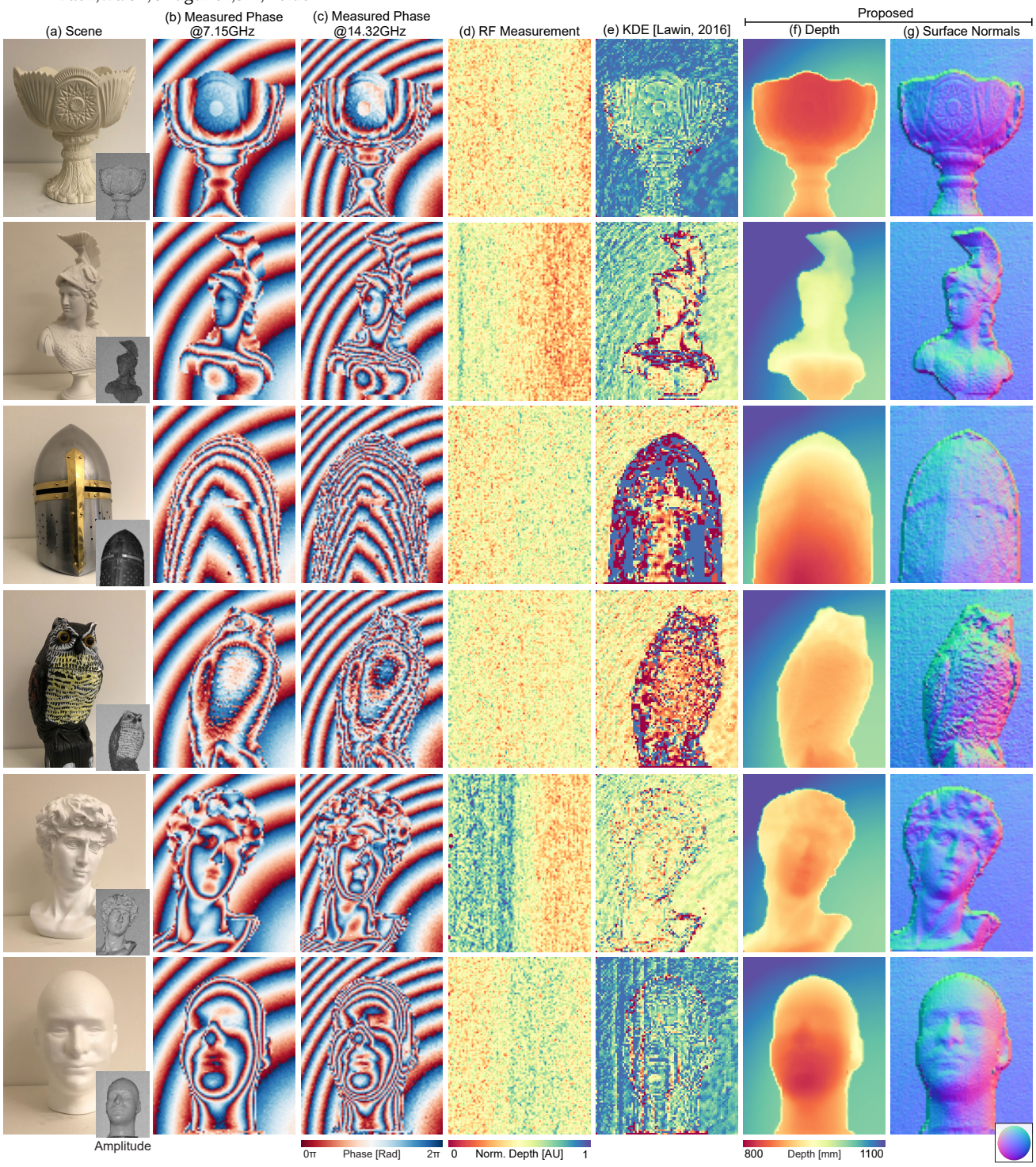

Fig. 16. (a) We experimentally validate the proposed method on challenging real-world scenes, for which we show photographs and recovered amplitude measurements (b) & (c). While our all-optical neural ToF system captures accurate phase at 7.15 and 14.32 GHz (frequency-doubled 7.16 GHz), (d) the existing GHz radio-frequency (RF) method electronically computes the correlation after photo-electron conversion (see text for details). As a result, this method struggles with low photon flux of the returned signal, producing noisy depth reconstructions. (e) We also evaluate existing phase-unwrapping methods for unwrapping measured GHz phase measurements which fail for most scenes due to the high wrapping counts of GHz frequencies (f) & (g). The proposed neural unwrapping method successfully resolves this issue and enables accurate geometry reconstruction visualized as depth and normals.

objects at known distances using a motion stage (Thorlabs MTS50/M-Z8) as shown in Fig. 13 and Tab. 2. We control the position of the target object that is placed at 60 cm distance from the setup. At this depth offset, we sweep over 1 mm travel distance with 50 μm step size (stage error 0.05 μm) and estimate corresponding depth values using the proposed method. Our imaging system achieves a mean depth error of 32.5 um and 32.9 um for a specular mirror and a diffuse reflector respectively.

Furthermore, we measure the height difference between the two metallic precision-fabricated gage blocks at 100 cm distance as shown in Fig. 14. The two gage blocks (ACCUSIZE DIN861 Metric, Grade2) are placed on a static mount. The difference of the measured depths, the height difference, is 0.4988 mm which is only 12 μm off from the ground truth 0.5 mm, demonstrating the precision of our depth acquisition. We also captured the shape of a large diffuse flat plane at a slanted angle. Once we obtain the depth map from our measurements, we fit a plane equation and the fitting R-squared value is 0.9998, demonstrating the accuracy of our method over longer travel distance than the translation-stage experiment.

Comparison with MHz Correlation ToF Imaging. The proposed method performs all-optical GHz modulation for high-resolution depth imaging. Fig. 15 compares our method with the conventional MHz correlation ToF imaging used in LUCID Helios Flex camera equipped with four VCSEL diodes of cumulatively 8 mW illumination module at 850 nm wavelength and 8 ms exposure, which is comparable to the effective photon budget of our system, although less susceptible to ambient light due to the wavelength filter. Our estimated depth contains fine geometric details for challenging scenes at correct depth scales, whereas the MHz correlation ToF suffers from low-precision depth with mm deviation on diffuse highly reflecting surfaces and larger cm to 10 cm deviation for surface areas of low reflectance. We ignore errors due to multipath effects (naturally suppressed by scanning in the proposed method) in this evaluation by focusing on convex object shapes. We note that we provide here qualitative comparisons with RGB frames as reference rather than ground truth depth,which is challenging to acquire for highly specular objects such as the helmet. Qualitatively, MHz correlation ToF fails to recover correct geometry for bright and dark spots on the owl, resulting in a 200 mm depth error (i.e. texture-dependent depth errors). The method also fails for the faint diffuse component returned from the specular helmet scene. While this trend is also confirmed in the estimated normals, we note that the holes in the helmet are "closed" by an incorrect wrapping estimate.

Comparison to RF Demodulation and Optical Interferometry. We compare the proposed method to RF demodulation after photoconversion and to interferometric depth estimation. To compare to RF demodulation, we use the same highspeed GaAs 12GHz photodetector as before. We note that this was the fastest photodiode available to us, see again Sec. 7 and the Supplement for additional details. To compare interferometric depth estimation with the proposed method, we add a moving reference mirror and an intensity detector so that interference can be detected with the superposed reference and scene beams as shown in Fig. 13(c). To implement this approach with the same proposed setup, we place a beam block in front of the reference mirror when we use the system in the proposed correlation mode. For fair comparison, we unwrap the interferometric data with sequential unwrapping which adds the smallest multiple of 2pi whenever the phase exceeds 2pi.

Tab. 4 shows that for a 1 mm sweep at 0.6 m distance, our proposed method with an emitter-decoder setup outperforms the RF demodulation in depth accuracy. The proposed method has a lower depth error than the RF method for RMSE and MAE for both a specular reflector shown in (a), and a diffuse reflector shown in (b). The depth estimates for all methods are shown for a 1 mm range compared to a ground truth for both specular and diffuse reflectors. We validate that, while post-photoconversion performs well for high flux levels, typical diffuse scenes results in low photon counts that are challenging to sense at high frequencies. As such, the RF demodulation approach *fails* for the diffuse scene object. We note that in contrast to direct fast sampling at rates higher than 10 GHz in the RF setup, our *all-optical sensing enables us to get away low-frequency kHz sampling (six orders of magnitude slower) with high SNR.*

The RSME and MAE for interferometry, RF and proposed methods are shown in Tab. 4 for specular and diffuse reflectors. The interferometric depth estimation performs best in terms of RMSE and MAE for specular and diffuse reflectors as expected. In this experiment, the proposed method achieves a depth precision of around 30 microns. We note that optical interferometry setup is extremely sensitive to scene scale, system vibrations, to the point where measurements had to be completed remotely from outside the lab and repeated multiple times due to tiny measurement fluctuations.

9 DISCUSSION

We have introduced a computational imaging method, that presents a complementary direction to existing ToF methods. Specifically, we have jointly designed the optics, sensing and neural network reconstruction such that computation that is typically done on the sensor, or digitally after sensing, is executed optically on incident photon stream. Doing so, we introduce concepts from optics on electro-optical modulation to the graphics and imaging community, while devising a new method for two-pass modulation and a new method for unwrapping high-frequency phase measurements. Although we experimentally and synthetically validate that our system performs effective centimeter-wave ToF depth imaging, as a nascent technology, our work also leaves the reader with some open questions regarding its future, which are discussed below.

Implementing Array Sensors. We have opted for sequential point-wise scanning using a galvo system as the beam diameter passing through our EOMs is limited by the EOM's small active area, 2.5 $\times$ 2.5 mm. An alternative implementation requiring further engineering efforts is the use of telescope optics to spatially expand the EOM-modulated light, hence exploiting that correlation ToF only mandates global intensity modulation instead of per-pixel intensity control, see also [Kim et al. 2019].

Flood-Illumination and Multi-Path Interference. As our prototype performs point-wise scanning, direct reflection dominates the measurement, which mitigates multi-path interference. When implementing the proposed system with flood illumination in the future, and using 2D array sensing with large-area EOMs, retraining the

network with flood-illumination might appear as an immediate solution to multi-path interference. We note that proposed high-frequency modulation may already provide sufficient robustness to the multi-path problem [Gupta et al. 2015].

Generalization to Complex Geometry and Reflectance. For scenes with simple shapes, moving planar targets in Fig. 13 and gage blocks and a slanted plane in Fig. 14, we demonstrate micron-scale depth resolution. In the future, we hope this approach can be extended to resolve micron-scale features in more complex scenes. While the proposed method outperforms previous methods for complex macroscopic scenes, capturing accurate depth still proves challenging; local geometries induce phase noise as the angular light beams are integrated over uneven depths. In the future, narrow beam sampling or flood-illuminated setups with array sensing might be a hardware solution to this challenge. In addition, more accurate ground truth sensing in the fine-tuning step might also overcome this domain gap issue in the neural network reconstruction.

Phase Unwrapping and Denoising. The proposed neural unwrapping method exploits the ordinal nature of the wrap counts and segmentation-based image semantics to recover dozens of wrap counts, while existing methods fail for more than a handful. While this approach shares some similarities with denoising in that we want to recover clean phase measurements from noisy readings, it does so in a *joint* manner. Rather than performing denoising and unwrapping sequentially, the proposed network ingests both correlation measurements simultaneously and can use their joint information – and independent noise distributions – to inform unwrapping. In this way we avoid accidentally denoising phase measurements into the wrong wrap count bin.

10 CONCLUSION

We propose a computational ToF imaging method that correlates light all-optically at centimeter-wave frequencies, without fiber coupling or photon-conversion – enabling high SNR sensing with more than 10 GHz modulation frequency. To this end, we solve two technical challenges: modulating without large signal losses at GHz rates, and unwrapping phase at theses rates which render conventional phase unwrapping methods ineffective. Specifically, we propose a two-pass intensity modulation with free-space EOMs and polarizing optics, which works in tandem with a neural phase unwrapping method to handle high wrapping counts in GHz-frequency measurements, on the order of dozens of wraps. The resulting imaging method achieves ToF imaging with centimeter intensity modulation for macroscopic scenes, robust to materials of low reflectance, highly-specular materials, and ambient light. We demonstrate accurate depth reconstructions, outperforming existing phase-unwrapping and post-photo-conversion ToF methods for *all* synthetic and real-world experiments. Our approach makes a step towards the goal of filling the gap between interferometric and correlation ToF. Our method performs computation optically that traditionally has been done after or during the sensing process. As such, in the future, we envision that the proposed approach could serve as an optical compute block for a diverse array of tasks, including velocity imaging, transient imaging, non-line-of-sight imaging, and imaging in scattering media, with the potential for fueling imaging of ultrafast phenomena across disciplines.

ACKNOWLEDGMENTS

Ilya Chugunov was supported by an NSF Graduate Research Fellowship. Felix Heide was supported by an NSF CAREER Award (2047359), a Sony Young Faculty Award, and a Project X Innovation Award.

REFERENCES

Supreeth Achar, Joseph R Bartels, William L'Red' Whittaker, Kiriakos N Kutulakos, and Srinivasa G Narasimhan. 2017. Epipolar time-of-flight imaging. *ACM Transactions on Graphics (ToG)* 36, 4 (2017), 1–8.

Narendra Ahuja and A. Lynn Abbott. 1993. Active stereo: integrating disparity, vergence, focus, aperture and calibration for surface estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 15, 10 (1993), 1007–1029.

M-C Amann. 1992. Phase noise limited resolution of coherent LIDAR using widely tunable laser diodes. *Electronics Letters* 28, 18 (1992), 1694–1696.

Yatong An, Jae-Sang Hyun, and Song Zhang. 2016. Pixel-wise absolute phase unwrapping using geometric constraints of structured light system. *Optics express* 24, 16 (2016), 18445–18459.

Brian F Aull, Andrew H Loomis, Douglas J Young, Richard M Heinrichs, Bradley J Felton, Peter J Daniels, and Deborah J Landers. 2002. Geiger-mode avalanche photodiodes for three-dimensional imaging. *Lincoln laboratory journal* 13, 2 (2002), 335–349.

Seung-Hwan Baek, Diego Gutierrez, and Min H Kim. 2016. Birefractive stereo imaging for single-shot depth acquisition. *ACM Transactions on Graphics (TOG)* 35, 6 (2016), 1–11.

Seung-Hwan Baek and Felix Heide. 2021. Polka Lines: Learning Structured Illumination and Reconstruction for Active Stereo. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 5757–5767.

Cyrus S Bamji, Swati Mehta, Barry Thompson, Tamer Elkhatib, Stefan Wurster, Onur Akkaya, Andrew Payne, John Godbaz, Mike Fenton, Vijay Rajasekaran, et al. 2018. IMpixel 65nm BSI 320MHz demodulated TOF Image sensor with 3μm global shutter pixels and analog binning. In *2018 IEEE International Solid-State Circuits Conference-(ISSCC)*. IEEE, 94–96.

Sankhyabrata Bandyopadhyay, Li-yang Shao, Wang Chao, Zhijun Yan, Fei Hong, Guoqing Wang, Jiahao Jiang, Ping Shum, Xiaoping Hong, and Weizhi Wang. 2020. Highly efficient free-space fiber coupler with 45° tilted fiber grating to access remotely placed optical fiber sensors. *Optics express* 28, 11 (2020), 16569–16578.

Behnam Behroozpour, Phillip AM Sandborn, Niels Quack, Tae-Joon Seok, Yasuhiro Matsui, Ming C Wu, and Bernhard E Boser. 2016. Electronic-photonic integrated circuit for 3D microimaging. *IEEE Journal of Solid-State Circuits* 52, 1 (2016), 161–172.

Behnam Behroozpour, Phillip AM Sandborn, Ming C Wu, and Bernhard E Boser. 2017. Lidar system architectures and circuits. *IEEE Communications Magazine* 55, 10 (2017), 135–142.

Ayush Bhandari, Achuta Kadambi, Refael Whyte, Christopher Barsi, Micha Feigin, Adrian Dorrington, and Ramesh Raskar. 2014. Resolving multipath interference in time-of-flight imaging via modulation frequency diversity and sparse regularization. *Optics letters* 39, 6 (2014), 1705–1708.

Shariq Farooq Bhat, Ibraheem Alhashim, and Peter Wonka. 2021. Adabins: Depth estimation using adaptive bins. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 4009–4018.

Jose Bioucas-Dias, Vladimir Katkovnik, Jaakko Astola, and Karen Egiazarian. 2008. Absolute phase estimation: adaptive local denoising and global unwrapping. *Applied optics* 47, 29 (2008), 5358–5369.

José Bioucas-Dias, Vladimir Katkovnik, Jaakko Astola, and Karen Egiazarian. 2009. Multi-frequency phase unwrapping from noisy data: adaptive local maximum likelihood approach. In *Scandinavian Conference on Image Analysis*. Springer, 310–320.

Jos M Bioucas-Dias and Gonalo Valadao. 2007. Phase unwrapping via graph cuts. *IEEE Transactions on Image processing* 16, 3 (2007), 698–709.

Danilo Bronzi, Yu Zou, Federica Villa, Simone Tisa, Alberto Tosi, and Franco Zappa. 2015. Automotive three-dimensional vision through a single-photon counting SPAD camera. *IEEE Transactions on Intelligent Transportation Systems* 17, 3 (2015), 782–795.

Clara Callenberg, Zheng Shi, Felix Heide, and Matthias B Hullin. 2021. Low-cost SPAD sensing for non-line-of-sight tracking, material classification and depth imaging. *ACM Transactions on Graphics (TOG)* 40, 4 (2021), 1–12.

Angel X Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, et al. 2015. Shapenet: An information-rich 3d model repository. *arXiv preprint arXiv:1512.03012* (2015).

Zhen Chen, BO Liu, Shengjie Wang, and Enhai Liu. 2018. Polarization-modulated three-dimensional imaging using a large-aperture electro-optic modulator. *Applied optics* 57, 27 (2018), 7750–7757.

Edward Collett. 2005. Field guide to polarization. Spie Bellingham, WA.

Sergio Cova, Massimo Ghioni, Andrea Lacaita, Carlo Samori, and Franco Zappa. 1996. Avalanche photodiodes and quenching circuits for single-photon detection. *Applied optics* 35, 12 (1996), 1956–1976.

Ryan Crabb and Roberto Manduchi. 2015. Fast single-frequency time-of-flight range imaging. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 58–65.

Angela Dai, Angel X Chang, Manolis Savva, Maciej Halber, Thomas Funkhouser, and Matthias Nießner. 2017. Scannet: Richly-annotated 3d reconstructions of indoor scenes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 5828–5839.

Angela Dai, Daniel Ritchie, Martin Bokeloh, Scott Reed, Jürgen Sturm, and Matthias Nießner. 2018. Scancomplete: Large-scale scene completion and semantic segmentation for 3d scans. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 4578–4587.

Adrian A Dorrington, John P Godbaz, Michael J Cree, Andrew D Payne, and Lee V Streeter. 2011. Separating true range measurements from multi-path and scattering interference in commercial range cameras. In *Three-Dimensional Imaging, Interaction, and Measurement*, Vol. 7864. International Society for Optics and Photonics, 786404.

David Droeschel, Dirk Holz, and Sven Behnke. 2010. Multi-frequency phase unwrapping for time-of-flight cameras. In *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 1463–1469.

Sergi Foix, Guillem Alenya, and Carme Torras. 2011. Lock-in time-of-flight (ToF) cameras: A survey. *IEEE Sensors Journal* 11, 9 (2011), 1917–1926.

Daniel Freedman, Yoni Smolin, Eyal Krupka, Ido Leichter, and Mirko Schmidt. 2014. SRA: Fast removal of general multipath for ToF sensors. In *European Conference on Computer Vision*. Springer, 234–249.

KD Froome and RH Bradsell. 1961. Distance measurement by means of a light ray modulated at a microwave frequency. *Journal of scientific instruments* 38, 12 (1961), 458.

Stefan Fuchs. 2010. Multipath interference compensation in time-of-flight camera images. In *2010 20th International Conference on Pattern Recognition*. IEEE, 3583–3586.

James Fujimoto and Eric Swanson. 2016. The development, commercialization, and impact of optical coherence tomography. *Investigative ophthalmology & visual science* 57, 9 (2016), OCT1–OCT13.

Shuang Gao and Rongqing Hui. 2012. Frequency-modulated continuous-wave lidar using I/Q modulator for simplified heterodyne detection. *Optics letters* 37, 11 (2012), 2022–2024.

Rahul Garg, Neal Wadhwa, Sameer Ansari, and Jonathan T Barron. 2019. Learning single camera depth estimation using dual-pixels. In *Proceedings of the IEEE International Conference on Computer Vision*. 7628–7637.

Dennis C Ghiglia and Mark D Pritt. 1998. *Two-dimensional phase unwrapping: theory, algorithms, and software*. Vol. 4. Wiley New York.

Ioannis Gkioulekas, Anat Levin, Frédo Durand, and Todd Zickler. 2015. Micron-scale light transport decomposition using interferometry. *ACM Transactions on Graphics (ToG)* 34, 4 (2015), 1–14.

Anant Gupta, Atul Ingle, and Mohit Gupta. 2019a. Asynchronous single-photon 3D imaging. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 7909–7918.

Anant Gupta, Atul Ingle, Andreas Velten, and Mohit Gupta. 2019b. Photon-flooded single-photon 3D cameras. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 6770–6779.

Mohit Gupta, Shree K Nayar, Matthias B Hullin, and Jaime Martin. 2015. Phasor imaging: A generalization of correlation-based time-of-flight imaging. *ACM Transactions on Graphics (ToG)* 34, 5 (2015), 1–18.

Mohit Gupta, Andreas Velten, Shree K. Nayar, and Eric Breitbach. 2018. What Are Optimal Coding Functions for Time-of-Flight Imaging? *ACM Trans. Graph.* 37, 2, Article 13 (Feb. 2018), 18 pages. https://doi.org/10.1145/3152155

Felipe Gutierrez-Barragan, Syed Azer Reza, Andreas Velten, and Mohit Gupta. 2019. Practical coding function design for time-of-flight imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 1566–1574.

Dipl-Ing Bianca Hagebeuker and Product Marketing. 2007. A 3D time of flight camera for object detection. *PMD Technologies GmbH, Siegen* (2007).

Miles Hansard, Seungkyu Lee, Ouk Choi, and Radu Patrice Horaud. 2012. *Time-of-flight cameras: principles, methods and applications*. Springer Science & Business Media.

Parameswaran Hariharan. 2003. *Optical Interferometry, 2e*. Elsevier.

Felix Heide, Steven Diamond, David B Lindell, and Gordon Wetzstein. 2018. Sub-picosecond photon-efficient 3D imaging using single-photon sensors. *Scientific reports* 8, 1 (2018), 1–8.

Felix Heide, Matthias B Hullin, James Gregson, and Wolfgang Heidrich. 2013. Low-budget transient imaging using photonic mixer devices. *ACM Transactions on Graphics (ToG)* 32, 4 (2013), 1–10.

Felix Heide, Matthew O'Toole, Kai Zang, David B Lindell, Steven Diamond, and Gordon Wetzstein. 2019. Non-line-of-sight imaging with partial occluders and surface normals. *ACM Transactions on Graphics (ToG)* 38, 3 (2019), 1–10.

Felix Heide, Lei Xiao, Andreas Kolb, Matthias B Hullin, and Wolfgang Heidrich. 2014. Imaging in scattering media using correlation image sensors and sparse convolutional coding. *Optics express* 22, 21 (2014), 26338–26350.

Steven Hickson, Stan Birchfield, Irfan Essa, and Henrik Christensen. 2014. Efficient hierarchical graph-based segmentation of RGBD videos. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 344–351.

Heiko Hirschmuller. 2005. Accurate and efficient stereo processing by semi-global matching and mutual information. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, Vol. 2. IEEE, 807–814.

David Huang, Eric A Swanson, Charles P Lin, Joel S Schuman, William G Stinson, Warren Chang, Michael R Hee, Thomas Flotte, Kenton Gregory, Carmen A Puliafito, et al. 1991. Optical coherence tomography. *science* 254, 5035 (1991), 1178–1181.

Darren D Hudson, Kevin W Holman, R Jason Jones, Steven T Cundiff, Jun Ye, and David J Jones. 2005. Mode-locked fiber laser frequency-controlled with an intracavity electro-optic modulator. *Optics letters* 30, 21 (2005), 2948–2950.

Shahram Izadi, David Kim, Otmar Hilliges, David Molyneaux, Richard Newcombe, Pushmeet Kohli, Jamie Shotton, Steve Hodges, Dustin Freeman, Andrew Davison, et al. 2011. KinectFusion: real-time 3D reconstruction and interaction using a moving depth camera. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*. 559–568.

David Jiménez, Daniel Pizarro, Manuel Mazo, and Sira Palazuelos. 2014. Modeling and correction of multipath interference in time of flight cameras. *Image and Vision Computing* 32, 1 (2014), 1–13.

Achuta Kadambi and Ramesh Raskar. 2017. Rethinking Machine Vision Time of Flight With GHz Heterodyning. *IEEE Access* 5 (2017), 26211–26223. https://doi.org/10.1109/ACCESS.2017.2775138

Achuta Kadambi, Refael Whyte, Ayush Bhandari, Lee Streeter, Christopher Barsi, Adrian Dorrington, and Ramesh Raskar. 2013. Coded Time of Flight Cameras: Sparse Deconvolution to Address Multipath Interference and Recover Time Profiles. *ACM Trans. Graph.* 32, 6, Article 167 (Nov. 2013), 10 pages. https://doi.org/10.1145/2508363.2508428

Achuta Kadambi, Hang Zhao, Boxin Shi, and Ramesh Raskar. 2016. Occluded Imaging with Time-of-Flight Sensors. *ACM Trans. Graph.* 35, 2, Article 15 (March 2016), 12 pages. https://doi.org/10.1145/2836164

Daehee Kim, Yang Lu, Jiyong Park, Byunggi Kim, Liping Yan, Liandong Yu, Ki-Nam Joo, and Seung-Woo Kim. 2019. Rigorous single pulse imaging for ultrafast interferometric observation. *Optics express* 27, 14 (2019), 19758–19767.

Ahmed Kirmani, Arrigo Benedetti, and Philip A Chou. 2013. Spumic: Simultaneous phase unwrapping and multipath interference cancellation in time-of-flight cameras using spectral methods. In *2013 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 1–6.

Alankar Kotwal, Anat Levin, and Ioannis Gkioulekas. 2020. Interferometric transmission probing with coded mutual intensity. *ACM Transactions on Graphics (TOG)* 39, 4 (2020), 74–1.

Robert Lange. 2000. 3D time-of-flight distance measurement with custom solid-state image sensors in CMOS/CCD-technology. (2000).

Robert Lange and Peter Seitz. 2001. Solid-state time-of-flight range camera. *IEEE Journal of quantum electronics* 37, 3 (2001), 390–397.

Felix Järemo Lawin, Per-Erik Forssén, and Hannes Ovrén. 2016. Efficient multi-frequency phase unwrapping using kernel density estimation. In *European Conference on Computer Vision*. Springer, 170–185.

Nalpantidis Lazaros, Georgios Christou Sirakoulis, and Antonios Gasteratos. 2008. Review of stereo vision algorithms: from software to hardware. *International Journal of Optomechatronics* 2, 4 (2008), 435–462.

R Leitgeb, CK Hitzenberger, and Adolf F Fercher. 2003. Performance of fourier domain vs. time domain optical coherence tomography. *Optics express* 11, 8 (2003), 889–894.

Fengqiang Li, Florian Willomitzer, Prasanna Rangarajan, Mohit Gupta, Andreas Velten, and Oliver Cossairt. 2018. SH-ToF: Micro Resolution Time-of-Flight Imaging with Superheterodyne Interferometry. In *Computational Photography (ICCP), 2018 IEEE International Conference on*. IEEE.

Reza Mahjourian, Martin Wicke, and Anelia Angelova. 2018. Unsupervised learning of depth and ego-motion from monocular video using 3d geometric constraints. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 5667–5675.

Riccardo Marchetti, Cosimo Lacava, Ali Khokhar, Xia Chen, Ilaria Cristiani, David J Richardson, Graham T Reed, Periklis Petropoulos, and Paolo Minzioni. 2017. High-efficiency grating-couplers: demonstration of a new design strategy. *Scientific reports* 7, 1 (2017), 1–8.

Aongus McCarthy, Robert J Collins, Nils J Krichel, Verónica Fernández, Andrew M Wallace, and Gerald S Buller. 2009. Long-range time-of-flight scanning sensor based on high-speed time-correlated single-photon counting. *Applied optics* 48, 32 (2009), 6241–6251.

Andreas Meuleman, Seung-Hwan Baek, Felix Heide, and Min H Kim. 2020. Single-Shot Monocular RGB-D Imaging Using Uneven Double Refraction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2465–2474.

Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. 2020. Nerf: Representing scenes as neural radiance fields for view synthesis. *arXiv preprint arXiv:2003.08934* (2020).

U Minoni, E Sardini, E Gelmini, F Docchio, and D Marioli. 1991. A high-frequency sinusoidal phase-modulation interferometer using an electro-optic modulator: Development and evaluation. *Review of scientific instruments* 62, 11 (1991), 2579–2583.

Nikhil Naik, Achuta Kadambi, Christoph Rhemann, Shahram Izadi, Ramesh Raskar, and Sing Bing Kang. 2015. A Light Transport Model for Mitigating Multipath Interference in Time-of-Flight Sensors. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Cristiano Niclass, Alexis Rochas, P-A Besse, and Edoardo Charbon. 2005. Design and characterization of a CMOS 3-D image sensor based on single photon avalanche diodes. *IEEE Journal of Solid-State Circuits* 40, 9 (2005), 1847–1854.

Matthew O'Toole, David B Lindell, and Gordon Wetzstein. 2018. Confocal non-line-of-sight imaging based on the light-cone transform. *Nature* 555, 7696 (2018), 338.

Gaurav Pandey, James R McBride, and Ryan M Eustice. 2011. Ford campus vision and lidar data set. *The International Journal of Robotics Research* 30, 13 (2011), 1543–1552.

Dingyi Pei, Arto Salomaa, and Cunsheng Ding. 1996. *Chinese remainder theorem: applications in computing, coding, cryptography*. World Scientific.

Christopher T Phare, Yoon-Ho Daniel Lee, Jaime Cardenas, and Michal Lipson. 2015. Graphene electro-optic modulator with 30 GHz bandwidth. *Nature Photonics* 9, 8 (2015), 511–514.

Rudra PK Poudel, Stephan Liwicki, and Roberto Cipolla. 2019. Fast-scnn: Fast semantic segmentation network. *arXiv preprint arXiv:1902.04502* (2019).

Fabio Remondino and David Stoppa. 2013. *TOF range-imaging cameras*. Vol. 2. Springer.

Mike Roberts and Nathan Paczan. 2020. Hypersim: A Photorealistic Synthetic Dataset for Holistic Indoor Scene Understanding. arXiv 2020.

Alexis Rochas, Michael Gösch, Alexandre Serov, Pierre-André Besse, Rade S. Popovic, Theo Lasser, and Rudolf Rigler. 2003. First fully integrated 2-D array of single-photon detectors in standard CMOS technology. *IEEE Photonics Technology Letters* 15, 7 (2003), 963–965.

Christopher Rogers, Alexander Y Piggott, David J Thomson, Robert F Wiser, Ion E Opris, Steven A Fortune, Andrew J Compston, Alexander Gondarenko, Fanfan Meng, Xia Chen, et al. 2021. A universal 3D imaging sensor on a silicon photonics platform. *Nature* 590, 7845 (2021), 256–261.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*. Springer, 234–241.

Phillip AM Sandborn, Noriaki Kaneda, Young-Kai Chen, and Ming C Wu. 2016. Dual-sideband linear FMCW lidar with homodyne detection for application in 3d imaging. In *2016 Conference on Lasers and Electro-Optics (CLEO)*. IEEE, 1–2.

Ashutosh Saxena, Sung H Chung, Andrew Y Ng, et al. 2005. Learning depth from single monocular images. In *NIPS*, Vol. 18. 1–8.

Daniel Scharstein and Richard Szeliski. 2003. High-accuracy stereo depth maps using structured light. In *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.*, Vol. 1. IEEE, I–I.

Brent Schwarz. 2010. LIDAR: Mapping the world in 3D. *Nature Photonics* 4, 7 (2010), 429.

Shikhar Shrestha, Felix Heide, Wolfgang Heidrich, and Gordon Wetzstein. 2016. Computational imaging with multi-camera time-of-flight systems. *ACM Transactions on Graphics (ToG)* 35, 4 (2016), 1–11.

Nathan Silberman, Derek Hoiem, Pushmeet Kohli, and Rob Fergus. 2012. Indoor segmentation and support inference from rgbd images. In *European conference on computer vision*. Springer, 746–760.

Nikolai Smolyanskiy, Alexey Kamenev, and Stan Birchfield. 2018. On the importance of stereo for accurate depth estimation: An efficient semi-supervised deep neural network approach. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 1007–1015.

Shuran Song, Samuel P Lichtenberg, and Jianxiong Xiao. 2015. Sun rgb-d: A rgb-d scene understanding benchmark suite. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 567–576.

Shuochen Su, Felix Heide, Robin Swanson, Jonathan Klein, Clara Callenberg, Matthias Hullin, and Wolfgang Heidrich. 2016. Material classification using raw time-of-flight measurements. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 3503–3511.

Shuochen Su, Felix Heide, Gordon Wetzstein, and Wolfgang Heidrich. 2018. Deep end-to-end time-of-flight imaging. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 6383–6392.

Murali Subbarao and Gopal Surya. 1994. Depth from defocus: A spatial domain approach. *International Journal of Computer Vision* 13, 3 (1994), 271–294.

Zhaoyang Tai, Lulu Yan, Yanyan Zhang, Xiaofei Zhang, Wenge Guo, Shougang Zhang, and Haifeng Jiang. 2016. Electro-optic modulator with ultra-low residual amplitude modulation for frequency modulation and laser stabilization. *Optics letters* 41, 23 (2016), 5584–5587.

Matthew Tancik, Pratul P Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan T Barron, and Ren Ng. 2020. Fourier features let networks learn high frequency functions in low dimensional domains. *arXiv preprint arXiv:2006.10739* (2020).

Michal Tölgyessy, Martin Dekan, L'ubo✓s Chovanec, and Peter Hubinskỳ. 2021. Evaluation of the azure Kinect and its comparison to Kinect V1 and Kinect V2. *Sensors* 21, 2 (2021), 413.

Shubham Tulsiani, Alexei A Efros, and Jitendra Malik. 2018. Multi-view consistency as supervisory signal for learning shape and pose prediction. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2897–2905.

Kaiqiang Wang, Ying Li, Qian Kemao, Jianglei Di, and Jianlin Zhao. 2019. One-step robust deep learning phase unwrapping. *Optics express* 27, 10 (2019), 15100–15115.

Xiang-Gen Xia and Genyuan Wang. 2007. Phase unwrapping and a robust Chinese remainder theorem. *IEEE Signal Processing Letters* 14, 4 (2007), 247–250.

Amnon Yariv. 1967. *Quantum electronics*. Wiley.

Amnon Yariv and Pochi Yeh. 2007. *Photonics: optical electronics in modern communications*. Oxford University Press.

Teng Zhang, Shaowei Jiang, Zixin Zhao, Krishna Dixit, Xiaofei Zhou, Jia Hou, Yongbing Zhang, and Chenggang Yan. 2019. Rapid and robust two-dimensional phase unwrapping via deep learning. *Optics express* 27, 16 (2019), 23173–23185.