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:

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;
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 20 cm 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 modulation 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 [Okkoukas 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 tables [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 FMCS 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.

$$\text{Phase Unwrapping.}$$ 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 tables [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 FMCS 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.

$$\text{Phase Unwrapping.}$$ 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 tables [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 FMCS 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.

$$\text{Phase Unwrapping.}$$ 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 tables [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 FMCS 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.
correlation measurement \( C_\psi \) does not directly give us access to the true phase \( \phi \), by sampling this function for multiple demodulation phase offsets \( \psi \) we can make use of Fourier analysis to discern the true phase \( \phi \). 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 \( 2\pi \) range as \( \hat{\phi} = \arctan \left( \frac{c_x - C_{\pi/2}}{c_z - C_{3\pi/2}} \right) \). Phase unwrapping amounts to estimating the integer factor \( n \) to recover the unwrapped phase \( \hat{\phi} + 2\pi n \). If successful, we can convert this phase estimate \( \hat{\phi} \) to depth as \( z = \frac{\hat{\phi} c}{4\pi \omega} \), 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 \( \phi \) phase shift. Setting \( t = 0, \beta = 0 \) and \( \omega = 100 \text{MHz} \) (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 frequencies, the same experiment repeated for \( \omega = 8 \text{GHz} \) 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 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 \( \phi \) 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 \( \pm 10 \text{MHz} \), 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
We calculate our final estimate 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.

This differentiable argmax allows for back-propagation through the ToF signal. 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 \( \phi' \) with

\[
\phi' = \sum_{n=0}^{N-1} \frac{\gamma \phi_n}{\sum_{m=0}^{N-1} \gamma \phi_m},
\]

where \( z \) and \( \phi \) are ground truth measurements. The cross-entropy loss allows us to train the network as a classifier, while the smooth \( \ell_1 \) 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 Paczak 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

\[
\phi = \frac{(z \omega \pi) c}{c}, \quad \omega \in \{7.15\text{GHz}, 14.32\text{GHz}\}, \quad c = \text{the speed of light.}
\]

We simulate ToF correlation images \( C_{\psi} \) for \( \psi \in \{0, \pi/2, 3\pi/2, \pi\} \) as

\[
C_{\psi} = G T_p (0.5 + \cos(\phi + \psi) / \pi) + \eta_p + \eta_G,
\]

where \( G \) is sensor gain, \( T \) is integration time, and \( T_p \) 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 \( \eta_p \) and Gaussian noise \( \eta_G \) with mean \( \mu \) and standard deviation \( \sigma \). 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 \( C_{\phi} \) obtained either from the training dataset or our GHz ToF imaging system, we recover the wrapped phase \( \phi \) and amplitude \( \hat{a} \) using a per-pixel Fourier transform

\[
\phi = \text{angle}(\mathcal{F}_2(C_{\phi})), \quad \hat{a} = 2|\mathcal{F}_2(C_{\phi})|,
\]

where \( \text{angle}(\cdot) \) 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, \( \phi \in [0, 2\pi] \) is phase wrapped, and is passed into our phase-unwrapping network.
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 $2 \times 1$ 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 $2 \times 2$ 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

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

where $V$ is a time-varying voltage function, $\eta$ is the modulation power, $\omega$ is the voltage modulation frequency, and $\phi$ 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 $\omega = 7.15\,\text{GHz}$. 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

$$ E_0 = A \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad (11) $$

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 an EOM at a given voltage $V$ as

$$ E_1 = B(V)Q(\theta_q)H(\theta_h)E_0, \quad (12) $$

ACM Trans. Graph., Vol. 39, No. 4, Article 1. Publication date: July 2021.
where \( H(\theta_h) \) and \( Q(\theta_q) \) are the Jones matrices of the HWP and the QWP oriented at angles \( \theta_h \) and \( \theta_q \). The light then propagates in free-space for a modulation wavelength \( c/\omega \) 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 \( \theta_q = 11.25^\circ \) \( \theta_q = 45^\circ \), we obtain the output light

\[
E_3 = L_\| H(-\theta_h)Q(-\theta_q)B(V)M(B)Q(\theta_q)H(\theta_h)E_0
\]

\[
= A \begin{bmatrix} \frac{i\cos V - \sin V}{\sqrt{2}} \\ 0 \end{bmatrix}. \tag{14}
\]

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

\[
I(t) = \frac{A^2}{2} (1 - \sin(2\eta \sin(\omega t - \phi))) \approx \frac{A^2}{2} (1 - 2\eta \sin(\omega t - \phi)). \tag{16}
\]

The last approximation is based on the Taylor expansion given that the modulation power \( \eta \) of our EOM is small. The applied voltage to the EOM has GHz modulation frequency \( \omega = 7.15 \text{ GHz} \), 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 polarization 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 \( \omega \) is limited to a narrow modulation band 20 MHz in our resonant EOM, we can modulate at the double frequency of \( 2\omega \) by adjusting the angle of the HWP, \( \theta_h \), 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 \( \theta_h \) as 11.25° resulting in the intensity modulation at the original frequency \( \omega \). For frequency doubling, we rotate the HWP to \( \theta_h = 22.5^\circ \). To derive the modulation behavior, we rely on the same Jones calculus from above. Specifically, changing the HWP angle \( \theta_h \) 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

\[
I(t) = \frac{A^2}{2} (1 - \cos(2\eta \sin(\omega t - \phi)))
\]

\[
= \frac{A^2}{2} \eta^2 \sin^2(\omega t - \phi)
\]

\[
= \frac{A^2}{4} \eta^2 (1 - \cos(2\omega t - 2\phi)). \tag{19}
\]

We use the same Taylor expansion. Eq. (19) shows that we can obtain the doubled frequency \( 2\omega \) at 1/4th amplitude compared to the single-frequency mode at \( \omega \) 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 \( \omega = 7.15 \text{GHz} \). (c) Changing the HWP angle to 22.5° at the EOM base frequency \( \omega = 7.16 \text{GHz} \) enables frequency doubling; \( 2\omega = 14.32 \text{GHz} \).

6.2.3 Validation of GHz Intensity Modulation. We validate our GHz intensity modulation module consisting of a PBS, a HWP, a QWP, an 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 \( \omega = 7.15 \text{GHz} \) and \( 2\omega = 14.32 \text{GHz} \).

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 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 14-bit quantization. This results in the digital read of

\[
C_\psi = Q \left( \frac{\alpha}{2} \cos(\psi - \phi) + TK \right),
\]

where \( Q \) is the 14-bit ADC quantization, \( \phi \) is the illumination phase, and \( \psi \) is the phase of the reference \( r \) shown in Eq. (4).

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 2.5x 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 \( \omega \). The light is reflected by a mirror (Thorlabs PF10-05-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 1.6x 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 $t_{RC} = 100\text{ms}$ 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 DDS 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 $\phi$. 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 photodetector 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 $\ell_1$ 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;
We compare the performance of the proposed neural unwrapping method and baseline methods including CRT [Xia and Wang 2007], KDE [Lawin et al. 2016], and the Phasor [2015] method to 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 comparison results for the full rank 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 to 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

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

<table>
<thead>
<tr>
<th>Input</th>
<th>Loss</th>
<th>Performance (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\phi$</td>
<td>$y(\phi)$</td>
<td>$L_{CE}$</td>
</tr>
<tr>
<td>Proposed</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>$L_{CE}$ Only</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>F. Features</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Phase Only</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

### 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.

<table>
<thead>
<tr>
<th>Method</th>
<th>$\uparrow \delta &lt; 1$</th>
<th>$\uparrow \delta &lt; 2$</th>
<th>$\uparrow \delta &lt; 3$</th>
<th>$\downarrow \delta \geq 3$</th>
<th>$\downarrow \delta \geq 10$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phasor [2015]</td>
<td>0.74%</td>
<td>1.66%</td>
<td>3.50%</td>
<td>96.5%</td>
<td>84.4%</td>
</tr>
<tr>
<td>CRT [2007]</td>
<td>9.29%</td>
<td>14.7%</td>
<td>19.7%</td>
<td>80.3%</td>
<td>56.0%</td>
</tr>
<tr>
<td>KDE [2016]</td>
<td>9.46%</td>
<td>18.56%</td>
<td>27.0%</td>
<td>73.0%</td>
<td>8.93%</td>
</tr>
<tr>
<td>One-Step [2019]</td>
<td>19.9%</td>
<td>37.6%</td>
<td>52.2%</td>
<td>47.8%</td>
<td>14.6%</td>
</tr>
<tr>
<td>U-Net [2015]</td>
<td>21.8%</td>
<td>45.6%</td>
<td>64.4%</td>
<td>35.6%</td>
<td>10.0%</td>
</tr>
<tr>
<td>Deep-ToF [2018]</td>
<td>20.3%</td>
<td>47.5%</td>
<td>67.6%</td>
<td>32.4%</td>
<td>8.40%</td>
</tr>
<tr>
<td>Rapid. [2019]</td>
<td>23.1%</td>
<td>45.4%</td>
<td>61.1%</td>
<td>38.9%</td>
<td>9.74%</td>
</tr>
<tr>
<td>Proposed</td>
<td>51.6%</td>
<td>69.1%</td>
<td>77.0%</td>
<td>23.0%</td>
<td>7.59%</td>
</tr>
</tbody>
</table>

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 simulate a high noise level \( \sigma \) with two different measurement distortions. First, we simulate global This is confirmed qualitatively in Fig. 10 with spatially-consistent ACM Trans. Graph., Vol. 39, No. 4, Article 1. Publication date: July 2021.

7.15GHz signal and a shifted plus frequency doubled (7.15GHz + 10MHz) \( \times 2 = 14.32\)GHz 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

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 \( \pm 0.01 \) 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) \( \times 2 = 14.32\)GHz 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

<table>
<thead>
<tr>
<th>Environment</th>
<th>( \delta &lt; 1 )</th>
<th>( \delta &lt; 2 )</th>
<th>( \delta &lt; 3 )</th>
<th>( \delta \geq 3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conventional</td>
<td>51.6%</td>
<td>69.1%</td>
<td>77.0%</td>
<td>23.0%</td>
</tr>
<tr>
<td>With Phase Offset</td>
<td>51.0%</td>
<td>68.9%</td>
<td>77.0%</td>
<td>23.0%</td>
</tr>
<tr>
<td>With Ambient Light</td>
<td>42.7%</td>
<td>60.4%</td>
<td>70.0%</td>
<td>30.0%</td>
</tr>
</tbody>
</table>

Table 3. Quantitative results for our phase unwrapping method against phase offset error and higher noise level due to the ambient light.
We note that we use the optimal frequency settings for the phasor with low albedo components. We evaluate the impact of our neural phase unwrapping method robustly performing 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 specular object corresponding to the measurements in Fig. 13.

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

<table>
<thead>
<tr>
<th>Method</th>
<th>RMSE Mirror</th>
<th>MAE Mirror</th>
<th>RMSE Diffuse</th>
<th>MAE Diffuse</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interferometry</td>
<td>20 μm</td>
<td>20 μm</td>
<td>14 μm</td>
<td>14 μm</td>
</tr>
<tr>
<td>RF</td>
<td>49.5 μm</td>
<td>48.8 μm</td>
<td>11800 μm</td>
<td>11800 μm</td>
</tr>
<tr>
<td>Proposed</td>
<td>33.5 μm</td>
<td>32.5 μm</td>
<td>34.6 μm</td>
<td>32.9 μm</td>
</tr>
</tbody>
</table>

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 μm with a standard deviation of 7.5 μm, 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.

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-ZB) 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 μm and 32.9 μm 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 DIN61 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 high-speed 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 2πi whenever the phase exceeds 2πi.

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-photonconversion 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 RMSE 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 × 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 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**


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


