What Are Optimal Coding Functions for Time-of-Flight Imaging?

MOHIT GUPTA and ANDREAS VELTEN, University of Wisconsin-Madison
SHREE K. NAYAR, Columbia University
ERIC BREITBACH, University of Wisconsin-Madison

The depth resolution achieved by a continuous wave time-of-flight (C-ToF) imaging system is determined by the coding (modulation and demodulation) functions that it uses. Almost all current C-ToF systems use sinusoid or square coding functions, resulting in a limited depth resolution. In this article, we present a mathematical framework for exploring and characterizing the space of C-ToF coding functions in a geometrically intuitive space. Using this framework, we design families of novel coding functions that are based on Hamiltonian cycles on hypercube graphs. Given a fixed total source power and acquisition time, the new Hamiltonian coding scheme can achieve up to an order of magnitude higher resolution as compared to the current state-of-the-art methods, especially in low signal-to-noise ratio (SNR) settings. We also develop a comprehensive physically-motivated simulator for C-ToF cameras that can be used to evaluate various coding schemes prior to a real hardware implementation. Since most off-the-shelf C-ToF sensors use sinusoid or square functions, we develop a hardware prototype that can implement a wide range of coding functions. Using this prototype and our software simulator, we demonstrate the performance advantages of the proposed Hamiltonian coding functions in a wide range of imaging settings.

CCS Concepts: • **Computing methodologies** → **Computational photography**; **3D imaging**;

Additional Key Words and Phrases: Computational time-of-flight imaging, low-power 3D cameras, Hamiltonian cycle, gray codes

ACM Reference format:
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 (February 2018), 18 pages.
https://doi.org/10.1145/3152155

1 INTRODUCTION

Time-of-flight (ToF) cameras have fast emerged as the preferred 3D imaging technique in several scientific and consumer applications, including robot navigation, motion capture, human computer interfaces, and 3D mapping. Especially popular are continuous-wave ToF (C-ToF) or indirect ToF imaging systems (Lange 2000; Payne 1973), which consist of temporally modulated light sources and sensors, as shown in Figure 2. These sensors need only low-cost and low-power components, do not require a large baseline for measuring depth, and thus, can potentially measure accurate 3D shapes over a large range of standoff distances. Due to these inherent advantages, C-ToF imaging systems are increasingly being commercialized as low-cost commodity sensors, such as Microsoft Kinect, PMD, and SoftKinectic.

One of the key limitations of current C-ToF cameras, however, is the limited depth resolution, especially in low signal-to-noise ratio (SNR) scenarios. Imagine a user wearing an augmented reality headset equipped with a low-power ToF depth camera. For her to achieve a realistic immersive experience, the camera must measure the 3D structure of the surroundings with very high resolution. Imagine an interplanetary rover navigating a rough terrain. It is critical for the camera to resolve fine details such as bumps on the terrain to navigate safely. Although the spatial resolution continues to rise with advances in sensor technology, the depth resolution is fundamentally limited by noise, especially photon noise. An example is shown in Figure 1(c) and (d). An object is imaged using a C-ToF system with a low-power source, resulting in low SNR and large depth errors. One way to increase the SNR is to use more light power or to increase the capture time. Unfortunately, this is not always possible. Most devices, especially in consumer and outdoor settings, often operate on a tight power and time budget.

Given a fixed energy and time budget, the depth resolution achieved by a C-ToF imaging system is determined by the coding (modulation and demodulation) functions that it uses. This raises a natural question: *What are the optimal coding functions for C-ToF imaging?*[1] Despite a lot of theoretical and engineering advances in C-ToF imaging over the past four decades, this fundamental question has received surprisingly little attention. Almost all current C-ToF systems use sinusoid or square codes, which are sub-optimal in their ability to achieve high depth resolution (Figure 1(d)).

In this article, we address the above question by establishing a theoretical foundation for analysis and design of C-ToF imaging systems. We develop a coding space theory of C-ToF imaging and define optimality of coding functions in terms of a mean depth error metric. Although this optimality criteria can be defined concisely, determining the optimal solution requires solving a high dimensional and computationally intractable global optimization problem. Our *key theoretical contribution* is to derive a novel depth precision metric based on a first-order differential analysis of the C-ToF image formation equation. This new metric allows us to explore the space of C-ToF coding functions (including non-analytic functions) in an intuitive, geometric space. This serves two purposes. First, this results in conceptual unification of

This research was supported in part by the DARPA REVEAL program and by ONR grant # N000141612995.
Authors' addresses: M. Gupta, 1210 W. Dayton Street, Department of Computer Sciences, UW-Madison, Madison, WI, 53706; email: mohitg@cs.wisc.edu; A. Velten and E. Breitbach, 1300 University Ave., Medical Sciences Building, UW-Madison, Madison, WI 53706; emails: velten@wisc.edu, eric.breitbach@gmail.com; S. K. Nayar, 500 W. 120th Street, Computer Science Department, Columbia University, New York, NY 10027; email: nayar@cs.columbia.edu.
Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
© 2018 ACM 0730-0301/2018/02-ART13 $15.00
https://doi.org/10.1145/3152155

[1] In this article, we consider optimality in the context of maximizing the depth resolution and the signal-to-noise ratio. We restrict the analysis to imaging systems with a single source and sensor, without any multi-path interference.

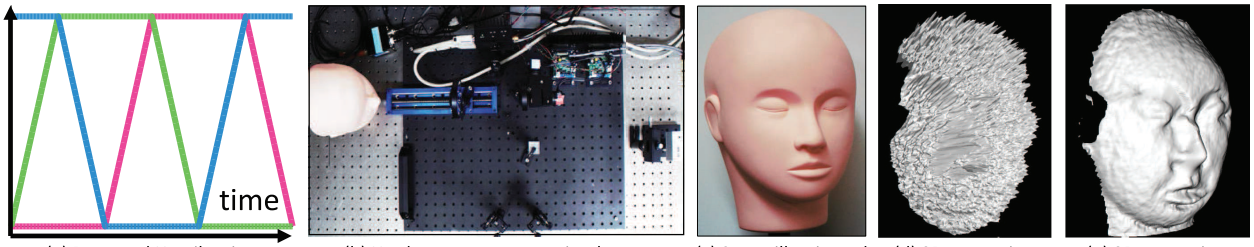


Fig. 1. **Hamiltonian time-of-flight (ToF) imaging.** (a) We propose a novel family of Hamiltonian ToF coding functions, which achieve up to 10 times higher depth resolution as compared to existing methods such as sinusoid coding, especially in low SNR settings. (b) We have developed a prototype ToF system for implementing the Hamiltonian coding functions. (c) A mannequin face was imaged using a low-power light source. (d) 3D imaging results using conventional sinusoid coding has large errors because of low-source power. (e) With the same total source power and capture time, Hamiltonian codes can accurately recover the overall shape, including facial features such as nose, lips, and eyes.

seemingly disparate C-ToF methods. Previously, different C-ToF coding schemes have been analyzed separately, with concrete performance metrics available only for specific schemes (e.g., sinusoid homodyne coding (Lange and Seitz 2001)). Using our framework, we can evaluate the performance of seemingly disparate coding techniques (e.g., sinusoid, square, triangle) on a unified platform.

Second, perhaps more importantly, our framework enables designing families of novel coding functions, which are derived from Hamiltonian cycles on hypercube graphs. These Hamiltonian coding functions, as shown in Figure 1(a), are tightly related to the theory of Gray codes (Gray 1953), and achieve substantially higher depth resolution as compared to existing schemes, given the same total capture time, total power, and depth range. For instance, while current schemes achieve a resolution of approximately 1 centimeter at a distance of 5–10 meters with commodity hardware, Hamiltonian coding scheme can achieve a resolution of $\sim$1 millimeter. An example result is shown in Figure 1(e).

We also develop a comprehensive physically-motivated simulator that can be used to evaluate the performance of C-ToF coding schemes, prior to a hardware implementation. This simulator can facilitate research in the fast growing field of ToF imaging where availability of appropriate hardware is often a bottleneck. Finally, since most commodity C-ToF sensors use sinusoid or square coding functions, we develop a hardware prototype system (Figure 1(b)) that can implement a wide range of C-ToF coding functions, including the proposed Hamiltonian scheme. We demonstrate the performance gains of the Hamiltonian scheme using exhaustive simulations as well as real experiments in a variety of imaging settings.

2 RELATED WORK

Impulse Time-of-Flight Imaging: Impulse (or direct) ToF systems (Goldstein and Dalrymple 1967; Koechner 1968) estimate scene depths by emitting a short light pulse into the scene and directly measuring the travel time of the reflected pulse. Impulse ToF method formed the basis of the first LIDAR systems nearly 50 years ago. Several current commercial range estimation systems (e.g., Velodyne sensor) are based on the impulse ToF method as well. Although conceptually simple, the main limitation of impulse ToF techniques is the high cost of its components (e.g., high-speed sensors) and large bandwidth requirements. Consequently, impulse ToF systems are not practical for most consumer applications. *The focus of this article is on C-ToF systems*, which require only low cost components and are fast becoming the method of choice for 3D imaging in a wide range of applications.

Code design for C-ToF imaging: Most research toward optimizing the accuracy of C-ToF imaging methods has been limited to sinusoid coding, the most widely used C-ToF coding technique (Lange 2000), especially in consumer ToF devices (e.g., Microsoft Kinect). Recently, (Payne et al. 2010) designed methods for mitigating depth errors in sinusoid coding-based systems if the modulation functions are not perfectly sinusoid (e.g., due to the presence of higher-order harmonics).

In parallel, techniques based on a few other specific modulation functions have been proposed, for example, square functions (Grootjans et al. 2006), triangular functions (Ferriere et al. 2008), ramp functions (Kolb et al. 2010) (used in commercial sensors from 3DV Systems), and pseudo random binary sequences (Grootjans et al. 2006; Kadambi et al. 2013). The goal of this article is to develop a principled framework for exploring the *complete space of ToF coding functions* in order to enable design of novel, high performance C-ToF coding schemes.

Simulation of C-ToF Sensors: Simulation engines based on physical models of C-ToF cameras are valuable tools for evaluating novel computational algorithms or hardware architectures for C-ToF imaging. Such simulation engines could be used to characterize the artifacts and predict the performance of specific sensors (e.g., PMD (Schmidt and Jähne 2009)), or model the effect of phenomena such as scene motion (Keller and Kolb 2009), and global illumination (Gupta et al. 2015; Meister et al. 2013). We use both simulations and real experiments to evaluate the performance of various C-ToF coding schemes, including those proposed in this article. In addition to a proof-of-concept hardware prototype, we have developed a physically-motivated simulator for C-ToF imaging, which can emulate different coding schemes, under a wide range of scene and sensor configurations. Although the simulator currently assumes that scene points receive light only directly

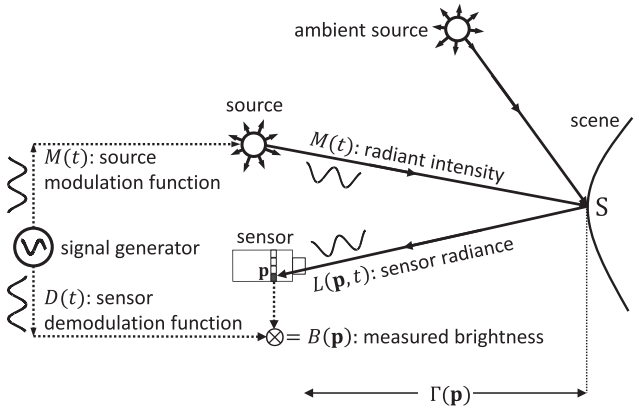


Fig. 2. **Image formation model of C-ToF imaging.** A C-ToF imaging system consists of a light source that illuminates the scene with time varying intensity $M(t)$. The light reflected from the scene is captured by a sensor, whose exposure is modulated during the integration time according to a function $D(t)$, called the demodulation function. The intensity $B(\mathbf{p})$ measured at a pixel $\mathbf{p}$ is given by the temporal correlation between the radiance $L(\mathbf{p}, t)$ incident at $\mathbf{p}$, and the demodulation function $D(t)$. The scene depth $\Gamma(\mathbf{p})$ at $\mathbf{p}$ can be estimated from three or more intensity measurements.

from the light source, it can be extended to incorporate global illumination effects such as interreflections (Gupta et al. 2015; Meister et al. 2013).

3 MATHEMATICAL PRELIMINARIES

A C-ToF or an indirect ToF imaging system consists of a temporally modulated light source, and a sensor whose exposure can be temporally modulated during integration time, as illustrated in Figure 2. Let the radiant intensity of the source at time t be $M(t)$, $(0 \leq M(t))$. The function $M(t)$, called the *source modulation function*, could be a continuous function such as a sinusoid (Lange 2000; Payne 1973), or even an impulse train function (Kolb et al. 2010).[2]

Consider a sensor pixel $\mathbf{p}$ that images a scene point S. Let $\Gamma(\mathbf{p})$ be the scene distance at pixel $\mathbf{p}$, i.e., the distance of scene point S from the sensor and the source.[3] We assume that there is no indirect or multi-bounce light component (Freedman et al. 2014; Heide et al. 2013; Kadambi et al. 2013; O'Toole et al. 2014), i.e., light from the source bounces only once at a scene point and returns to the sensor. Then, the radiance $L(\mathbf{p}, t)$ incident at pixel $\mathbf{p}$ at time t is given as:

$$L(\mathbf{p}, t) = \beta(\mathbf{p})\, M\left(t - \frac{2\,\Gamma(\mathbf{p})}{c}\right) + L_a(\mathbf{p}), \tag{1}$$

where c is the speed of light. $\beta(\mathbf{p})$ encapsulates the scene point's albedo and reflectance and light fall-off. $L_a(\mathbf{p})$ is the radiance component due to an ambient light source (e.g., sun in outdoor settings). Intuitively, the light emitted from the source $M(t)$ is time-shifted by $\frac{2\,\Gamma(\mathbf{p})}{c}$ due to propagation along the path: source → scene → sensor; the intensity is scaled by $\beta(\mathbf{p})$, and a constant offset $L_a(\mathbf{p})$ is added due to ambient illumination.

The sensor exposure is temporally modulated according to the demodulation function $D(t)$, $(0 \leq D(t) \leq 1)$. Both modulation and demodulation functions are periodic. The demodulation function (also called the sensor exposure function) can be physically realized either by on-chip gain modulation (Schwarte et al. 1997) or by external optical shutters (Carnegie et al. 2011).

The brightness $B(\mathbf{p})$ measured at pixel $\mathbf{p}$ is given by the correlation between the incoming radiance and the exposure function:

$$B(\mathbf{p}) = \int_0^\tau D(t)\, L(\mathbf{p}, t) dt, \tag{2}$$

where τ is the sensor integration time. Substituting Equation (1) into Equation (2), we get:

$$B(\mathbf{p}) = \beta(\mathbf{p}) \int_0^\tau D(t)\, M\left(t - \frac{2\,\Gamma(\mathbf{p})}{c}\right) dt + A(\mathbf{p}), \tag{3}$$

where $A(\mathbf{p}) = L_a(\mathbf{p}) \int_0^\tau D(t) dt$ is the ambient component of the measured brightness. Next, we define $\mathcal{F}(\Gamma)$ as the normalized correlation function between modulation and demodulation functions:

$$\mathcal{F}(\Gamma) = \frac{\int_0^\tau D(t)\, M\left(t - \frac{2\,\Gamma(\mathbf{p})}{c}\right) dt}{M_{total}}, \tag{4}$$

where the normalization factor $M_{total} = \int_0^\tau M(t)\, dt$ is the total flux emitted by the light source toward point S during the sensor integration time τ. Note that $0 \leq \mathcal{F}(\Gamma) \leq 1$. Substituting Equation (4) into Equation (3), and slightly abusing the notation by absorbing the constant M_{total} within the scale factor $\beta(\mathbf{p})$, we get:

$$\underbrace{\boxed{B(\mathbf{p}) = \beta(\mathbf{p})\, \mathcal{F}(\Gamma) + A(\mathbf{p}).}}_{\text{Image Formation Equation for C-ToF Imaging}} \tag{5}$$

Equation (5) is the *image formation equation for C-ToF imaging*. It expresses the image intensity B measured at a sensor pixel in terms of the three unknowns: (1) scene distance Γ, (2) albedo factor β, and (3) ambient brightness component A.

Significance of the normalized correlation functions: Equation (5) indicates that given an imaging scenario (defined in terms of albedos β and ambient component A), a C-ToF imaging system can be completely characterized by its normalized correlation functions $\mathcal{F}(\Gamma)$, which in turn depend *only* on the modulation and demodulation functions. $\mathcal{F}(\Gamma)$ is independent of the scene properties (albedos, ambient light) and device characteristics (sensor integration time, light source power). Our goal in this article is to (a) formalize a C-ToF system's performance in terms of its correlation functions, and (b) design families of correlation functions that lead to novel, high performance C-ToF systems.

4 CODING SPACE THEORY OF TOF IMAGING

The space of all possible values of unknowns Γ, β, and A, can be represented as a 3D *unknown space*, as shown in Figure 3.

[2] A C-ToF system using an impulse train source modulation function is different from impulse or direct ToF systems. In a direct ToF system, the time delay between the emitted and received pulse is measured directly, requiring sensors with a large bandwidth. In contrast, in a C-ToF system, the sensor measures the time delay indirectly by measuring the temporal correlation, thus requiring a small number of measurements.
[3] We assume that the source and the sensor are co-located.

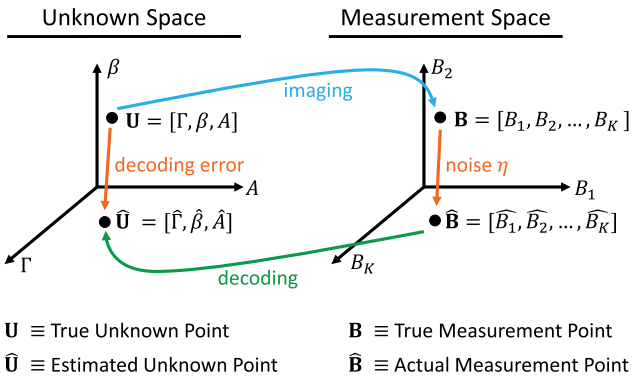


Fig. 3. **Coding space theory of C-ToF imaging.** The 3D space of all possible values of unknowns Γ (depths), β (albedo factors), and A (ambient illumination component) is called the *unknown space*. The K-dimensional ($K \geq 3$) space of measured intensities is called the *measurement space*. A C-ToF coding scheme maps every unknown point $\mathbf{U}$ to a measurement point $\mathbf{B}$. However, due to noise, the actual measurement point $\widehat{\mathbf{B}}$ may be different from the true measurement point $\mathbf{B}$. The decoding function maps $\widehat{\mathbf{B}}$ to the estimated unknown point $\widehat{\mathbf{U}}$, which results in depth errors $|\Gamma - \widehat{\Gamma}|$.

Definition 1 (Unknown Point). A point $\mathbf{U} = [\Gamma, \beta, A]$ in the unknown space is called an unknown point. Each unknown point represents a three-element *unknown vector*.

Since there are three unknowns, $K \geq 3$ intensity measurements $[B_1, B_2, \ldots, B_K]$ need to be captured using K different correlation functions $[\mathcal{F}_1, \mathcal{F}_2, \ldots, \mathcal{F}_K]$, corresponding to K different pairs of modulation and demodulation functions M_i and D_i:[4]

$$B_i(\mathbf{p}) = \beta(\mathbf{p})\, \mathcal{F}_i(\Gamma) + A(\mathbf{p})$$

The K-dimensional space of measured intensities $[B_1, \ldots, B_K]$ is called the *measurement space*, as shown in Figure 3.

Definition 2 (Measurement Point). A point $\mathbf{B} = [B_1, \ldots, B_K]$ in the measurement space is a valid measurement point if $[B_1, \ldots, B_K]$ are the intensity measurements corresponding to an unknown point $\mathbf{U} = [\Gamma, \beta, A]$.

Definition 3 (Coding Scheme). A C-ToF coding scheme is defined as a function $C : \mathbf{U} \rightarrow \mathbf{B}$ from the set of unknown points $\mathbf{U}$ to measurement points $\mathbf{B}$. A coding scheme C is completely characterized by a set of correlation functions $[\mathcal{F}_1, \ldots, \mathcal{F}_K]$, or equivalently, by sets of modulation and demodulation functions $[M_1, \ldots, M_K]$ and $[D_1, \ldots, D_K]$, respectively.

The process of recovering depths can be modeled as a decoding function $\mathcal{D} : \mathbf{B} \rightarrow \mathbf{U}$ from the measurement space to the unknown space, as shown in Figure 3. The decoding function could be an analytic expression, a table look-up, or a statistical procedure (e.g., maximum likelihood estimation). If there were no image noise, then given a measurement $\mathbf{B}$, the decoding algorithm could accurately recover the true unknown, and hence, the correct depth.

[4]For example, in conventional sinusoid coding (Lange 2000), both modulation and demodulation functions are sinusoids of the same frequency. We consider a general formulation where the modulation and demodulation functions could have arbitrary functional forms, as long as they adhere to the physical (e.g., non-negativity) constraints.

Effect of Image Noise on Depth Recovery

Consider a coding scheme C that maps an unknown point $\mathbf{U}$ to the true (noise-free) measurement point $\mathbf{B} = [B_1, \ldots, B_K]$. Let the actual measurement $\widehat{B}_i$ be given as:

$$\widehat{B}_i = B_i + \eta_i, \tag{6}$$

where $\eta_i = \triangle B_i = \widehat{B}_i - B_i$ is the noise in the intensity measurement B_i. We assume the affine noise model including both read noise and photon noise (Hasinoff et al. 2010). η_i is modeled as a Gaussian random variable with standard deviation σ_i, i.e., $\eta_i \sim \mathcal{N}(0, \sigma_i)$. Due to noise, the decoding function $\mathcal{D}$ estimates an incorrect unknown point $\widehat{\mathbf{U}} = [\widehat{\Gamma}, \widehat{\beta}, \widehat{A}]$, which results in depth errors $\triangle\Gamma = |\widehat{\Gamma} - \Gamma|$. This is illustrated in Figure 3. The expected depth error $\overline{\triangle\Gamma_C}(\mathbf{U})$ for an unknown $\mathbf{U}$ and a coding scheme C, is given as:

$$\overline{\triangle\Gamma_C}(\mathbf{U}) = \int_{\widehat{\mathbf{B}}} |\widehat{\Gamma} - \Gamma|\; p\left(\widehat{\mathbf{B}}|\mathbf{B}\right) d\widehat{\mathbf{B}}, \tag{7}$$

where $p\left(\widehat{\mathbf{B}}|\mathbf{B}\right) \sim \mathcal{N}(\mathbf{B}, \Sigma)$ is the probability distribution function of the actual measured intensity $\widehat{\mathbf{B}} = [\widehat{B_1}, \ldots, \widehat{B_K}]$. The covariance matrix Σ is given by the amount of image noise. The integral is taken over the K-dimensional measurement space.

Equation (7) gives the expected depth error $\overline{\triangle\Gamma_C}(\mathbf{U})$ for *a single unknown point* $\mathbf{U}$. The mean expected depth error $\overline{\triangle\Gamma_C}$ is given by averaging $\overline{\triangle\Gamma_C}(\mathbf{U})$ over the space of all unknowns:

$$\overline{\triangle\Gamma_C} = \frac{1}{\mathbf{V}_\mathbf{U}} \int_{\mathbf{U}} \overline{\triangle\Gamma_C}(\mathbf{U})\; d\mathbf{U}, \tag{8}$$

where $\mathbf{V}_\mathbf{U} = (\Gamma_{max} - \Gamma_{min})\,(\beta_{max} - \beta_{min})\,(A_{max} - A_{min})$ is the volume of the space of unknowns ($\Gamma_{min} \leq \Gamma \leq \Gamma_{max}$, $\beta_{min} \leq \beta \leq \beta_{max}$, $A_{min} \leq A \leq A_{max}$). Substituting Equation (7) into Equation (8), we get:

$$\boxed{\overline{\triangle\Gamma_C} = \frac{1}{\mathbf{V}_\mathbf{U}} \int_{\mathbf{U}} \int_{\widehat{\mathbf{B}}} \left|\widehat{\Gamma} - \Gamma\right| p\left(\widehat{\mathbf{B}}|\mathbf{B}\right) d\widehat{\mathbf{B}}\, d\mathbf{U}}$$

Mean Expected Depth Error of a C-ToF Coding Scheme (9)

The mean expected error is a global performance metric of a C-ToF coding scheme. It is an intrinsic property of a C-ToF coding scheme and is independent of choice of reconstruction method. We define the optimal coding scheme as the one that minimizes the mean expected depth error over a given unknown space:

$$C_{opt} = \arg\min_{C} \overline{\triangle\Gamma_C}. \tag{10}$$

Although the optimality criteria can be expressed concisely, finding the optimal codes is a formidable constrained global optimization problem, with no known analytical or closed-form solutions. It is computationally intensive to even evaluate this metric numerically as it involves computing a double integral over high-dimensional unknown and measurement spaces. This precludes the development of an efficient numerical or an exhaustive search-based optimization procedure as well due to prohibitively

high computational costs (Adam et al. 2016). We take a different approach. We derive a novel depth-precision metric based on first-order differential analysis of the image formation equation. This metric has an intuitive geometric interpretation, and can be computed extremely fast, leading to design of novel high performance coding schemes.

5 A NOVEL DEPTH PRECISION METRIC

The analysis in the previous section is based on the relationship between depth error $\triangle\Gamma$ and intensity noise $\triangle\mathbf{B} = \widehat{\mathbf{B}} - \mathbf{B} = \eta$. In this section, we analyze the relationship between the corresponding differential quantities $\partial\Gamma$ and $\partial\mathbf{B}$. Based on this analysis, we derive a novel depth precision metric. We start by taking the partial derivative of the intensity B_i (in Equation (5)) with respect to Γ:

$$\frac{\partial B_i}{\partial \Gamma} = \beta\, \mathcal{F}_i'(\Gamma),$$

where $\mathcal{F}_i'(\Gamma)$ is the derivative of correlation function $\mathcal{F}_i(\Gamma)$. For ease of exposition, we assume that the albedo factor β is independent of the scene depths. In practice, β depends on scene depths due to distance fall-off. Such depth dependence can be absorbed into the correlation function $\mathcal{F}_i(\Gamma)$. The partial derivative of the measurement vector $\mathbf{B} = [B_1, B_2, \ldots, B_K]$ is given as:

$$\left|\frac{\partial \mathbf{B}}{\partial \Gamma}\right| = \sqrt{\sum_{i=1}^{K}\left(\frac{\partial B_i}{\partial \Gamma}\right)^2} = \beta\sqrt{\sum_{i=1}^{K}\mathcal{F}_i'(\Gamma)^2}.$$

By rearranging terms, we get:

$$|\partial\Gamma| = \frac{|\partial\mathbf{B}|}{\beta\sqrt{\sum_{i=1}^{K}\mathcal{F}_i'(\Gamma)^2}}. \tag{11}$$

This equation expresses the differential depth error $|\partial\Gamma| = |\widehat{\Gamma} - \Gamma|$ due to small measurement noise $|\partial\mathbf{B}| = \sqrt{\sum_{i=1}^{K}\partial B_i^2}$, where $\partial B_i = \widehat{B_i} - B_i$. Since noise is random, both $|\partial\mathbf{B}|$ and $|\partial\Gamma|$ are random variables, with standard deviations Ω and σ_Γ, respectively. Then, from Equation (11), it follows that Ω and σ_Γ are related as:

$$\sigma_\Gamma = \frac{\Omega}{\beta\sqrt{\sum_{i=1}^{K}\mathcal{F}_i'(\Gamma)^2}}. \tag{12}$$

See Appendix for a derivation. Next, we introduce $\chi = \frac{1}{\sigma_\Gamma}$ as a *depth precision measure*:

$$\chi(\mathbf{U}) = \frac{1}{\sigma_\Gamma} = \frac{\beta\sqrt{\sum_{i=1}^{K}\mathcal{F}_i'(\Gamma)^2}}{\Omega}, \tag{13}$$

where $\chi(\mathbf{U})$ denotes the dependence of χ on the unknown point $\mathbf{U}$. Since χ *is the inverse of depth standard deviation* σ_Γ*, intuitively, larger the value of* χ*, lower is the overall depth error.* $\chi(\mathbf{U})$, as defined above, is the depth precision corresponding to a single unknown point $\mathbf{U}$. The mean depth precision $\overline{\chi_C}$ of a coding scheme C is given by averaging $\chi(\mathbf{U})$ over the space of all unknowns $\Gamma_{min} \le \Gamma \le \Gamma_{max}$, $\beta_{min} \le \beta \le \beta_{max}$ and $A_{min} \le A \le A_{max}$:

$$\overline{\chi_C} = \frac{1}{\mathbf{V}_\mathbf{U}}\int_{\mathbf{U}}\chi(\mathbf{U})\; d\mathbf{U} = \frac{1}{\mathbf{V}_\mathbf{U}}\int_{\Gamma}\int_{\beta}\int_{A}\chi(\mathbf{U})\; dA\, d\beta\, d\Gamma, \tag{14}$$

where $\mathbf{V}_\mathbf{U} = (\Gamma_{max} - \Gamma_{min})\,(\beta_{max} - \beta_{min})\,(A_{max} - A_{min})$ is the volume of the unknown space. Substituting Equation (13) into Equation (14), and simplifying with the assumption that the noise standard deviation $\Omega = \sqrt{\sum_{i=1}^{K}\sigma_i^2}$ is constant,[5] we get:

$$\overline{\chi_C} = \frac{\beta_{mean}\displaystyle\int_{\Gamma}\sqrt{\sum_{i=1}^{K}\mathcal{F}_i'(\Gamma)^2}\; d\Gamma}{\Omega\;\Gamma_{range}}, \tag{15}$$

where $\Gamma_{range} = \Gamma_{max} - \Gamma_{min}$ is the unambiguous depth range, and $\beta_{mean} = \frac{\beta_{min}+\beta_{max}}{2}$ is defined as the mean albedo factor. In order to understand the term inside the integral in the above equation, we define a new geometric construct called the *coding curve*.

Definition 4 (Coding Curve). Consider a coding scheme C defined by the correlation functions $[\mathcal{F}_1, \ldots, \mathcal{F}_K]$. The coding curve is achieved by plotting the values of the correlation functions $[\mathcal{F}_1(\Gamma), \ldots, \mathcal{F}_K(\Gamma)]$ in the K-dimensional space, as the depth Γ is varied. Formally, the coding curve Ψ_C is defined as the following set of points residing in K-dimensional space:

$$\Psi_C = \{[\mathcal{F}_1(\Gamma), \mathcal{F}_2(\Gamma), \ldots, \mathcal{F}_K(\Gamma)] \;\mid\; \Gamma_{min} \le \Gamma \le \Gamma_{max}\}.$$

The coding curve is a *geometric representation* of a coding scheme; given a scheme's correlation functions, we can determine its coding curve, and vice versa. For example, the coding curve of sinusoid coding is a circle in K-dim space, as shown in Figure 4(a), for $K = 3$. A proof is given in the supplementary technical report.

Length of the coding curve: The length $\mathcal{L}_{curve}$ of the coding curve is given in terms of partial derivatives $\mathcal{F}_i'$:

$$\mathcal{L}_{curve} = \int_{\Gamma_{min}}^{\Gamma_{max}}\sqrt{\sum_{i=1}^{K}\mathcal{F}_i'(\Gamma)^2}\; d\Gamma. \tag{16}$$

Substituting Equation (16) in Equation (15), we get the *main theoretical result of the article*:

$$\underbrace{\boxed{\overline{\chi_C} = \frac{\beta_{mean}\;\mathcal{L}_{curve}}{\Omega\;\Gamma_{range}}.}}_{\textbf{Depth Precision of a C-ToF Coding Scheme}} \tag{17}$$

Interpreting the Depth Precision Equation

Equation (17) states that the depth precision $\overline{\chi_C}$ of a coding scheme is directly proportional to its coding curve length. Intuitively, given a set of unknown points, a longer coding curve spreads the measurement points further apart in the measurement space, resulting in lower depth errors due to noise during decoding. Thus, in general, *the coding curve length (and hence, the depth precision* $\overline{\chi_C}$*) is inversely proportional to the mean depth error* $\overline{\triangle\Gamma_C}$. The longer the coding curve of a coding scheme, the larger is the value of $\overline{\chi_C}$ and the lower the mean depth error $\overline{\triangle\Gamma_C}$.[6] The coding curve and

[5] Strictly speaking, Ω is a function of image intensity, and varies for different unknown points. However, for simplicity, we assume Ω to be a constant, the upper bound of noise standard deviation over all possible measurement points.

[6] There are exceptions to this. If a coding curve is self-intersecting or has sharp turns, the differential analysis, which assumes that the derivatives $\frac{\partial B_i}{\partial \Gamma}$ are continuous, does not hold. Such a curve, despite being long, may result in large depth errors. See

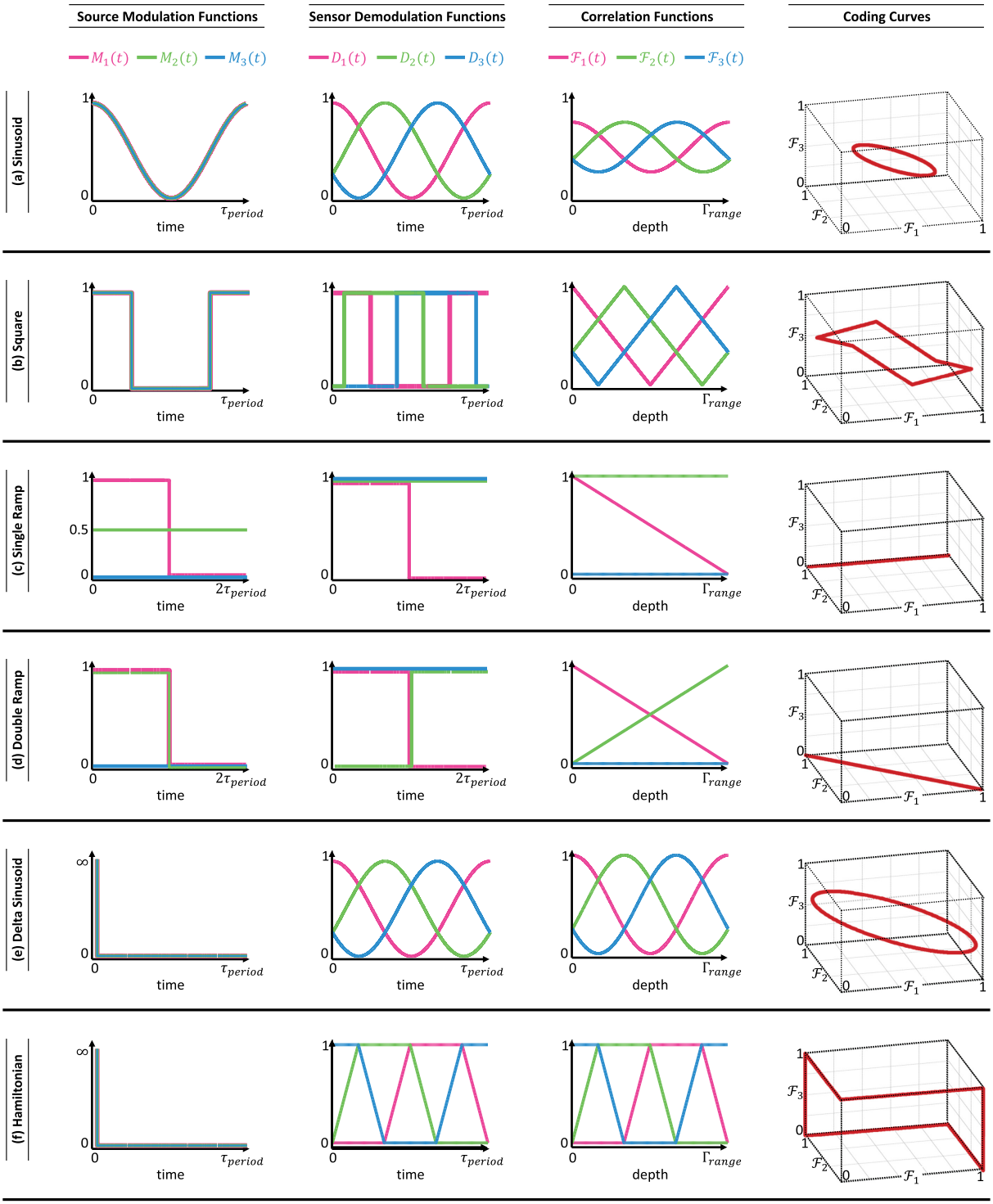


Fig. 4. **Coding curve representations of different C-ToF coding schemes for $K = 3$.** (a) Conventional homodyne sinusoid coding, (b) square coding, (c) ramp or "light wall" coding, (d) double ramp coding, (e) delta sinusoid coding, and (f) the proposed Hamiltonian coding.

its length are simple and fast to compute geometric properties of coding schemes (even non-analytic schemes), thus making them a compact representation that can be used as a powerful design tool for developing novel, high performance C-ToF coding schemes.

The depth precision $\overline{\chi_C}$ is directly proportional to the mean albedo factor β_{mean} and inversely proportional to measurement noise Ω.[7] This is not surprising, as the higher the signal (proportional to β_{mean}) received by the sensor and the lower the noise, the higher is the achieved precision. Perhaps more interesting is the inverse relationship between the depth precision and the depth range Γ_{range}. A small unambiguous depth range Γ_{range} results in higher depth precision (lower depth errors), and vice versa. Similar tradeoff between the depth range and depth error has been derived previously for sinusoid coding (Lange 2000; Lange and Seitz 2001; Lange et al. 2000). The depth precision equation derived above generalizes the range vs. precision relationship for any valid C-ToF coding scheme, and thus, allows *comparing the performance of a general class of coding schemes on a common ground by fixing the depth range* Γ_{range},[8] *scene albedo* β_{mean} *and sensor noise* Ω.

6 CODING CURVES OF C-TOF CODING SCHEMES

In this section, we derive the coding curve lengths of several C-ToF coding schemes, starting from their mathematical definitions in terms of their correlation functions. Figure 4 and Table 1 summarizes the results derived in this section. For fair comparisons, we assume that the total source power M_{total} (area under the curve of the modulation functions), exposure time, and the unambiguous depth range is the same for all coding schemes.

6.1 K-tap Conventional Sinusoid Coding

This is one of the most widely used coding schemes in commercial C-ToF–based depth cameras. Both the modulation and demodulation functions are sinusoids of the same frequency (Figure 4(a)):

$$M_i(t) = 0.5 + 0.5\cos(\omega t)$$
$$D_i(t) = 0.5 + 0.5\cos\left(\omega t - \frac{2i\pi}{K}\right), \quad 1 \le i \le K$$

K-tap denotes that K measurements are captured, with the demodulation function phase-shifted (typically by an equal amount $\frac{2\pi}{K}$) between successive measurements, while the modulation function remains the same. Note that the modulation function is normalized so that the area under the curve (for every period) is equal to π. After substituting these in Equation (4), and simplifying, we get the expression for the normalized correlation functions:

$$\mathcal{F}_i(\Gamma) = 0.5 + 0.25\cos\left(\frac{2\omega\Gamma}{c} - \frac{2i\pi}{K}\right). \tag{18}$$

Algebraic derivation of the coding curve length: Recall from Equation (16) that the coding curve length is given in terms of the derivatives of the correlation functions. Taking the derivative of Equation (18) with respect to depths Γ, we get:

$$\mathcal{F}_i'(\Gamma) = -0.25\left(\frac{2\omega}{c}\right)\sin\left(\frac{2\omega\Gamma}{c} - \frac{2i\pi}{K}\right)$$

After substituting the above into Equation (16), and simplifying, we get the coding curve length of sinusoid coding:

$$\boxed{\mathcal{L}_{curve}^{sin} = \frac{\pi}{2}\sqrt{\frac{K}{2}}} \tag{19}$$

Geometric derivation of the coding curve length: *The coding curve for sinusoid coding, i.e., the locus of points* $[\mathcal{F}_1(\Gamma), \ldots, \mathcal{F}_K(\Gamma)]$, *as* Γ *is varied, is a circle in* K*-dimensional space.* Please see the supplementary technical report for a proof. The center of the circle is the point $C = (0.5, 0.5, \ldots, 0.5)$. The radius of the circle is $rad(K) = \frac{\sqrt{K}}{4\sqrt{2}}$. The coding curve length is $\mathcal{L}_{curve}^{sin} = 2\pi\, rad(K) = \frac{\pi}{2}\sqrt{\frac{K}{2}}$. This is consistent with the algebraic derivation.

6.2 K-tap Square Coding

This coding scheme uses square waves instead of sinusoids, both for modulation and demodulation functions:

$$M_i(t) = 0.5 + 0.5\, sqr(\omega t)$$
$$D_i(t) = 0.5 + 0.5\, sqr\left(\omega t - \frac{2i\pi}{K}\right), \quad 1 \le i \le K$$

where $sqr(t)$ function is the binary version of the $\sin(t)$ function:

$$sqr(t) = \begin{cases} 1 & \text{if } sin(t) > 0 \\ -1 & \text{if } sin(t) < 0 \end{cases}$$

The correlation of two square functions is a triangle function. Thus, the normalized correlation function for square coding is given as:

$$\mathcal{F}_i(\Gamma) = 0.5 + 0.5\, tri\left(\frac{2\omega\Gamma}{c} - \frac{2i\pi}{K}\right),$$

where the $tri(t)$ function is shown in Figure 4(b).

Geometric derivation of the coding curve length: Due to the piecewise linear nature of the correlation functions, *the coding curve for square coding is a non-planar polygon in* K*-dimensional space*, as shown in Figure 4(b). The centroid of the polygon is the point $(0.5, \ldots, 0.5)$. The polygon has $2K$ sides, each with length $sideLength(K) = \frac{1}{\sqrt{K}}$ (see technical report for a proof). The total length of the coding curve is:

$$\boxed{\mathcal{L}_{curve}^{sqr} = 2\sqrt{K}} \tag{20}$$

Observation: The coding curve length of *square coding is* $\frac{4\sqrt{2}}{\pi} \approx 1.8$ *times that of conventional sinusoid coding*, for any K. This suggests that given the same scene and imaging system characteristics, square coding should achieve approximately 1.8 times high precision (lower error) as compared to sinusoid coding. We validate this result using simulations and experiments later in the article.

Section 10 for an example and discussion. We will design coding curves that, in addition to being long, are non self-intersecting and have continuous derivatives.

[7] Since noise Ω is approximately proportional to $\sqrt{\beta_{mean}}$ in a photon-noise limited regime, the depth precision $\overline{\chi_C}$ is effectively proportional to $\sqrt{\beta_{mean}}$.

[8] The depth range of a C-ToF system is given as $\Gamma_{range} = \frac{c\tau_{period}}{2}$, where τ_{period} is the period of the modulation and demodulation functions.

Table 1. Coding Curve Lengths of Various C-ToF Coding Schemes

Coding Scheme	Coding Curve	Curve Length	Curve Length ($K = 3$)	Curve Length ($K = 4$)	Curve Length ($K = 5$)
Single ramp	Edge of 3D unit cube	1	1.00	–	–
Double ramp	Face diagonal of 3D unit cube	$\sqrt{2}$	1.41	–	–
Sinusoid	Circle of radius $\frac{\sqrt{K}}{4\sqrt{2}}$	$\frac{\pi}{2}\sqrt{\frac{K}{2}}$	1.92	2.22	2.48
Square	Regular non-planar $(2K)$-gon	$2\sqrt{K}$	3.46	4.00	4.47
Impulse sinusoid	Circle of radius $\frac{\sqrt{K}}{2\sqrt{2}}$	$\pi\sqrt{\frac{K}{2}}$	3.84	4.44	4.96
Hamiltonian	Hamiltonian cycle	$2^K - 2$ (K is odd)	6.00	12.00	30.00
	on unit hypercube	$2^K - 4$ (K is even)			

Coding Curve Lengths of Existing Schemes Such as Sinusoid and Square are Proportional to $\sqrt{K}$, Where K is the Number of Measurements. In Contrast, Coding Curve Length of the Proposed Hamiltonian Scheme Increases Exponentially as a Function of K. As a Result, Hamiltonian Coding Achieves Significantly Higher Depth Precision as Compared to Existing Schemes, Especially as K Increases. Please Refer to Sections 6 and 7 for Detailed Derivations of Coding Curve Lengths.

6.3 Other Coding Schemes

The coding curve representation can be used to analyze the entire space of valid C-ToF coding schemes, such as ramp coding (also known as the "light wall" scheme (Kolb et al. 2010)) where one of the correlation functions is a ramp and the other two are constant (Figure 4(c)), double-ramp coding where two correlation functions are opposing ramps (Figure 4(d)), and impulse sinusoid coding where the demodulation functions are sinusoids (same as conventional sinusoid coding), but the modulation function is an impulse train function (Figure 4(e)). The coding curve lengths for these schemes are given in Table 1. For detailed derivations, refer to the supplementary technical report.

7 HAMILTONIAN TOF CODING

In this section, we use the coding curve representation to design a family of novel high-performance C-ToF coding schemes. There are three important desirable properties of a coding curve. First, since the mean depth precision is inversely proportional to the coding curve length, the curve should be *long*. Second, in order to ensure a unique mapping between the unknown points $\mathbf{U}$ and measurement points $\mathbf{B}$, the coding curve should *not be self-intersecting*. Third, the coding curve should preserve *locality*, i.e., the distance of points measured along the curve should be proportional to the Euclidean distance between two points. This property ensures that the intensity partial derivatives $\frac{\partial B}{\partial \Gamma}$ are continuous, and thus, small image noise does not result in large depth errors during decoding.

A family of curves that has all these desirable properties is *Hamiltonian cycles on hypercube graphs*. The hypercube graph Q_K is the graph formed from the vertices and edges of the K-dimensional hypercube. For example, the graph Q_3 is formed on a 3-D cube, and has 8 vertices (one for every cube corner) and 12 edges. A Hamiltonian cycle is a cycle (i.e., a closed loop) through a graph that visits every vertex exactly once while traversing the edges of the graph. Hamiltonian cycles are long, non-self intersecting and have provably good locality preserving properties (Gotsman and Lindenbaum 1996). This makes them an ideal candidate as coding curves for C-ToF coding schemes.

Design of Hamiltonian C-ToF coding scheme: We propose a family of C-ToF coding schemes called Hamiltonian coding which have *Hamiltonian cycles on hypercube graphs as coding curves*. Intuitively, in the context of C-ToF imaging, these curves can be thought of as maximal in a sense (although not provably optimal) because they follow the edges of the coding cube, and cannot be expanded further. The correlation functions are Hamiltonian functions, where the i^{th} Hamiltonian function $hamilt_{K,i}(t)$ is defined as the value of the i^{th} coordinate of points on a Hamiltonian cycle of the K-dimensional unit hypercube. Figure 4 (f) shows the plots of Hamiltonian functions $hamilt_{K,i}(t)$ for $K = 3$.

While constructing the Hamiltonian cycle–based coding curve, we exclude the origin $\mathbf{0_K} = [0, \ldots, 0]$ and the diagonally opposite vertex $\mathbf{1_K} = [1, \ldots, 1]$ so that the coding curve does not pass through these two vertices. This ensures that for every depth value Γ, at least two of the correlation function values are different, i.e., $\mathcal{F}_i(\Gamma) \neq \mathcal{F}_j(\Gamma)$ for some index pair i, j. It has been shown that a Hamiltonian cycle on this reduced set of vertices can be constructed.[9] In general, there are several possible Hamiltonian cycles on a hypercube graph, each inducing a different coding curve and coding scheme. One example of a Hamiltonian cycle on the reduced set of vertices for $K = 3$ is shown in Figure 4(f).

Geometric derivation of the coding curve length: The coding curve of a Hamiltonian scheme is a Hamiltonian cycle on the reduced hypercube graph $Q_K - [\mathbf{0_K}, \mathbf{1_K}]$, whose length is equal to the number of cube vertices that the cycle traverses ($2^K - 2$ if K is odd, and $2^K - 4$ if K is even). Thus, the coding curve length is:

$$\mathcal{L}_{curve}^{hamilt} = \begin{cases} 2^K - 2 & \text{if } K \text{is odd} \\ 2^K - 4 & \text{if } K \text{is even} \end{cases} \tag{21}$$

The coding curve length for the Hamiltonian scheme increases *exponentially* as a function of K, whereas for existing schemes such as sinusoid and square coding, the curve length is proportional to $\sqrt{K}$. As a result, even for relatively small K, such as $K = 5$, the curve length for Hamiltonian coding is an order of magnitude more than conventional sinusoid coding (see Table 1 for a comparison of coding curve lengths of various schemes).

Relationship to Gray codes and robustness to noise: Gray codes (Gray 1953) are a sequence of binary codes so that two

[9] This problem is NP-complete with no polynomial time algorithms. For graphs of small sizes, we can find solutions using depth-first search.

successive codes differ only in a single bit location, i.e., the Hamming distance between two adjacent codes is 1. This property makes Gray codes highly robust to noise: if a sequence of Gray codes is used to transmit information and a small number of bits gets corrupted (e.g., due to noise in a communication channel), the information can still be recovered with minimal errors. Consequently, Gray codes find applications in a diverse set of domains, including digital communication (Gray 1953), circuit design (Wakerly 2005), and robust structured light 3D scanning (Inokuchi et al. 1984).

Hamiltonian cycles on hypercube graphs are tightly related to the theory of Gray codes. There is a one-to-one correspondence between the set of K-bit Gray codes and the set of Hamiltonian cycles on the hypercube graphs Q_K (Mills 1963). The Hamiltonian coding curves that we use to design our Hamiltonian coding scheme can be considered a *continuous version of discrete Gray codes, and thus, inherit the strong robustness-to-noise properties of Gray codes.*

7.1 Modulation and Demodulation Functions

In order to physically implement the Hamiltonian coding scheme, we need to determine the source modulation functions $M_i(t)$, $(0 \leq M_i(t))$ and demodulation functions $D_i(t)$, $(0 \leq D_i(t) \leq 1)$ such that their normalized correlation (Equation (4)) is equal to the Hamiltonian functions $hamilt_{K,i}(\Gamma)$:

$$\frac{\int_0^\tau D_i(t) M_i\left(t - \frac{2\Gamma}{c}\right) dt}{M_{total}} = hamilt_{K,i}(\Gamma), \quad 1 \leq i \leq K, \tag{22}$$

where $M_{total} = \int_0^\tau M_i(t)\, dt$ is the total energy emitted by the source during integration time τ.

Let $\mathbf{h}_{K,i}[j], 1 \leq j \leq N$ be a vector representing the discrete sampled version (with N equi-spaced samples) of the function $hamilt_{K,i}(\Gamma)$. Similarly, let $\mathbf{m}_i[j]$ and $\mathbf{d}_i[j]$ be discrete representations of the functions $M_i(t)$ and $D_i(t)$. Then, the above system of equations and inequations can be written in matrix form as:

$$\begin{aligned} &\text{find} && (\mathbf{m_i},\ \mathbf{d_i}) \\ &\text{such that} && \frac{1}{M_{total}} \mathbf{C}_{\mathbf{m_i}} \mathbf{d_i} = \mathbf{h_{K,i}}, \qquad \sum_{j=1}^{N} \mathbf{m_i}[j] \leq M_{total} \\ & && 0 \leq \mathbf{m_i}[k], \qquad \leq \mathbf{d_i}[k] \leq 1, \quad 1 \leq k \leq N, \end{aligned}$$

where $\mathbf{C_m}$ is the $N \times N$ circulant matrix constructed from the vector $\mathbf{m}$. Since the equation contains product of the unknowns $\mathbf{m_i}$, $\mathbf{d_i}$, this is a non-linear problem and, in general, may not have a feasible solution. However, in the special case of light sources with large peak instantaneous power $\mathbf{m_i}[k] \geq M_{total}$, the above problem has a simple solution: $\mathbf{m_i} = [M_{total}, 0, 0, \ldots, 0]$ (so that $\mathbf{C}_{\mathbf{m_i}} = M_{total}\mathbf{I}$, where $\mathbf{I}$ is an identity matrix), and $\mathbf{d_i} = \mathbf{h_{K,i}}$. This solution corresponds to using a normalized impulse train function $\Lambda(\omega t)$ (as shown in Figure 4(e) and (f)) as the modulation function, and the correlation function itself as the demodulation function:

$$\begin{aligned} M_i(t) &= \Lambda(\omega t) \\ D_i(t) &= hamilt_{K,i}(\omega t), \quad 1 \leq i \leq K \end{aligned}$$

Incorporating peak power constraints: The above solution assumes that the light source has an infinite peak power and can emit an ideal impulse (delta) train function. For such ideal sources, the correlation function is simply factorized into a delta modulation function, and a demodulation function equal to the correlation function. However, this factorization is an idealized theoretical description, meant only for exposition.

In practice, sources have a finite peak power, and we approximate the impulse modulation function by a short (height equal to the peak power) but wider pulse (e.g., a Gaussian or a square) so that the area under the modulation functions (total emitted energy) is the same as that of modulation functions for other schemes (e.g., sine and square). The resulting correlation function (using a shorter and wider pulse) approximates the theoretical correlation function. This is shown in Figure 15, which compares the theoretical vs. measured correlation functions in our experiments. This approximation results in lower performance than theoretically predicted.

Ultimately, the performance will depend on the peak power of the light source. Lasers and didoes such as those increasingly being used in ToF systems can emit short pulses with high peak power (Adam et al. 2016; Kolb et al. 2010; Tadmor et al. 2014), but low average power due to energy consumption constraints and eye safety. Such sources can closely approximate an impulse modulation function, and thus, achieve high performance. For sources with low peak power (e.g., low-cost LEDs), the performance gains will be accordingly lower. For such sources, the performance could be improved by solving the factorization as a constrained optimization problem, where the goal will be to find $(\mathbf{m_i}, \mathbf{d_i})$ that satisfy all the physical constraints imposed by the hardware (including system bandwidth constraints), while minimizing the error $||\frac{1}{M_{total}} \mathbf{C}_{\mathbf{m_i}} \mathbf{d_i} - \mathbf{h_{K,i}}||$. Such *device-aware* factorization is an interesting future research direction, but different from the correlation function theory and design, which is the main focus of this article.

7.2 Depth Recovery Algorithm for Hamiltonian Coding

So far, we have discussed the coding function design aspect of the Hamiltonian coding scheme. In this section, we provide an overview of the depth recovery (decoding) algorithm. For any coding scheme, points on the coding curve $\mathbf{F}(\Gamma) = [\mathcal{F}_1(\Gamma), \mathcal{F}_2(\Gamma), \ldots, \mathcal{F}_K(\Gamma)]$ are parameterized by the depth value Γ, meaning there is a one-to-one mapping between Γ and points $\mathbf{F}(\Gamma)$. Thus, given the measurement point $\mathbf{B} = [B_1, \ldots, B_K]$, we can estimate depth Γ by determining the corresponding coding curve point $\mathbf{F}(\Gamma)$. From the image formation equation (Equation (5)):

$$\mathbf{B} = \beta\, \mathbf{F}(\Gamma) + A \quad \rightarrow \mathbf{F}(\Gamma) = \frac{\mathbf{B} - A}{\beta}, \tag{23}$$

where β and A are also unknown. In order to determine $\mathbf{F}(\Gamma)$ from measurements $\mathbf{B}$, we first estimate β and A.

Estimating unknowns β and A: The Hamiltonian coding curve follows the edges of a unit cube. Suppose the coding curve point $\mathbf{F}(\Gamma)$ lies on an edge between cube vertices $\mathbf{V_l} = [V_{l,1}, \ldots, V_{l,K}]$ and $\mathbf{V_r} = [V_{r,1}, \ldots, V_{r,K}]$. The coordinates of $\mathbf{V_l}$ and $\mathbf{V_r}$ are binary (0 or 1), and differ along only one index μ, i.e., $V_{l,\mu} \neq V_{r,\mu}$. Then, the coding curve point is given as $\mathbf{F}(\Gamma) = \alpha \mathbf{V_l} + (1 - \alpha)\mathbf{V_r}$, where $0 \leq \alpha \leq 1$ encodes the location of $\mathbf{F}(\Gamma)$ along the edge between $\mathbf{V_l}$ and $\mathbf{V_r}$. Note that for every Γ, $\mathcal{F}_{min}$ and $\mathcal{F}_{max}$, the minimum and maximum coordinates of the coding point $\mathbf{F}(\Gamma)$, are 0

and 1, respectively. Let $B_{min} = \beta \mathcal{F}_{min} + A$ and $B_{max} = \beta \mathcal{F}_{max} + A$ be the minimum and maximum measured intensities, respectively. Since $\mathcal{F}_{min} = 0$ and $\mathcal{F}_{max} = 1$, it follows that $B_{min} = A$ and $B_{max} = \beta + A$. Thus, we can estimate β and A as $A = B_{min}$ and $\beta = B_{max} - B_{min}$, where $B_{min} = \min([B_1, \ldots, B_K])$ and $B_{max} = \max([B_1, \ldots, B_K])$ are computed from the measurements B.[10] Once β and A are estimated, we determine the coding point $\mathbf{F}(\Gamma)$ using Equation (23), from which, as discussed above, depth can be estimated. We will release our MATLAB code of the depth recovery algorithms for Hamiltonian and other coding schemes (sinusoid, square, ramp, and double ramp), once the article is accepted.

8 VALIDATION AND SIMULATIONS

In this section, we use numerical methods and simulations to compare the relative performance of various C-ToF coding schemes, with the *same total capture time and emitted power.*

8.1 Comparisons of Mean Expected Depth Error

The mean expected depth error $\overline{\triangle\Gamma_C}$ (Equation (9)) is a global measure of the performance of a C-ToF coding scheme C. In order to compare various coding schemes, we numerically computed their mean expected depth errors by discretizing the unknown and measurement spaces. We implemented an adaptive grid method to ensure that this otherwise prohibitively expensive computation remains tractable. We used the following parameters for our computations. **Scene depth range:** $\Gamma_{min} = 0$, $\Gamma_{max} = 10$ meters. **Image noise:** Gaussian noise, with affine noise model including photon noise and sensor read noise of 20 electrons. **Sensor integration time:** 10 milliseconds. **Scene reflectivity:** Lambertian BRDF, albedo of $0 - 1$. All the coding schemes used the same fundamental frequency of 15MHz, and ideal waveforms were used for all coding functions.

Figure 5 shows the mean depth error for various coding schemes for $K = 3, 4, 5$, as a function of the light source and ambient illumination strengths. We considered a point light source with average strengths in the range of $\sim 10 - 1,000$ lumens. Ambient light was assumed to be uniform over the scene, with strengths in the range of $10 - 10,000$ lux. As expected, for every coding scheme, the depth error decreases rapidly as the source strength increases, and increases as ambient illumination increases. Hamiltonian coding scheme significantly outperforms existing methods across all settings, especially as K increases.

Correlation between coding curve lengths and mean expected depth errors: Figure 6 compares the coding curve lengths ($\mathcal{L}_{curve}$) and the *inverse* mean expected depth errors ($\frac{1}{\overline{\triangle\Gamma}}$) for various coding schemes. The inverse mean expected depth errors are plotted for all the source strength and ambient illumination values as used in Figure 5 (one solid colored plot for every source strength and ambient illumination combination). Each plot is normalized by dividing by the minimum value along the plot. For example, for $K = 3$, each plot is divided by its values at the *single ramp* coding scheme, and for $K = 5$, each plot is divided by its values at the *sinusoid* coding scheme.

There are two main observations. First, the ratio of the mean expected depth errors between two schemes is approximately consistent over a wide range of source and ambient illumination strengths. Second, perhaps more importantly, the normalized coding curve lengths of different schemes are *highly correlated* with the corresponding inverse mean expected depth errors. This shows that, for these coding schemes, *the coding curve length (and hence, the depth precision $\overline{\chi_C}$) is inversely proportional to the mean expected depth error $\overline{\triangle\Gamma}$.* As a result, the coding curve length can be used as an intuitive and fast-to-compute surrogate metric for optimization and design of novel, high performance C-ToF coding schemes.

8.2 Simulator for C-ToF Imaging

We have developed a physically motivated simulator for C-ToF imaging, which can emulate any valid coding scheme, under a wide range of scene configurations (different geometries and reflectance properties) and sensor-source parameters (quantization, pixel size, focal length, zoom, source brightness, and modulation frequencies). The simulator models different steps of the image formation process, including light emission and propagation, reflection, sensor demodulation and radiometric transformations (gain, saturation, quantization), and noise (photon and read noise).

These steps are illustrated in Figure 7. This simulator can help further research in the growing field of ToF imaging by facilitating rapid evaluation of different coding schemes. We will release the code of our simulator upon acceptance of the article. Figure 8 shows depth recovery results for three imaging geometries, and various coding schemes in a low noise setting ($K = 5$, source strength = 250 lumens, ambient strength = 250 lux), and a high noise setting due to *strong ambient illumination* ($K = 5$, source strength = 80 lumens, ambient strength = 2,500 lux). The remaining imaging parameters are the same as described in the previous section. Square coding achieves an error ≈1.6 times lower than conventional sinusoid (the ratio of coding curve lengths is ≈1.8). The proposed Hamiltonian coding achieves a mean depth error of about an order of magnitude lower than conventional sinusoid coding in both noise settings, consistent with the ratio of coding curve lengths. Figure 9 shows similar comparisons between various coding schemes for $K = 3$. The relative performance of different schemes is consistent with the respective coding curve lengths.

Figure 10 shows the 3D imaging simulations for a face 3D model in an outdoor setting with strong ambient light, resulting in large photon noise. The following parameters were used for these simulations: $K = 5$, source strength = 250 lumens, ambient strength = 2,500 lux. Although the large noise in sinusoid and square reconstructions can be reduced by smoothing, the 3D details cannot be recovered. In contrast, Hamiltonian coding recovers sufficient details to potentially allow recognizing the identity of the face.

Comparisons with bandlimited Hamiltonian coding functions: In practice, the Hamiltonian coding functions may be low-pass filtered due to bandwidth limitations of the hardware devices. For instance, Figure 15 shows the Hamiltonian coding functions as implemented on our hardware prototype. Due to bandwidth constraints, the higher-order Hamiltonian coding functions are

[10] This simple approach for estimating β and A may not be optimal. It may be possible to design a decoding algorithm with better performance, perhaps by jointly estimating β, A, and Γ. We leave the design of a better decoding algorithm as a topic of future study.

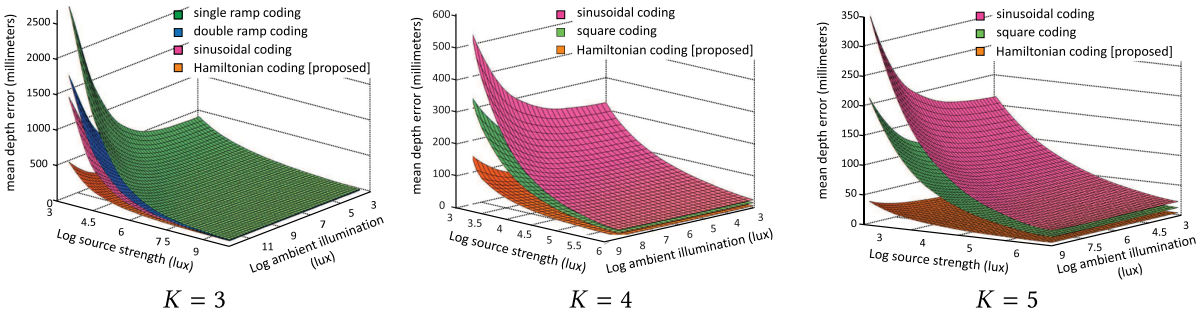


Fig. 5. **Mean expected depth errors of various C-ToF coding schemes.** The mean expected depth error (Equation (9)) of various C-ToF coding schemes was numerically computed using finite element methods over a depth range of [0−10] meters. The same fundamental frequency of 15MHz was used for all coding schemes. For every coding scheme, the depth error decreases rapidly as the source strength increases, and also increases as ambient illumination increases. The proposed Hamiltonian scheme significantly outperforms existing coding methods across a wide range of imaging settings.

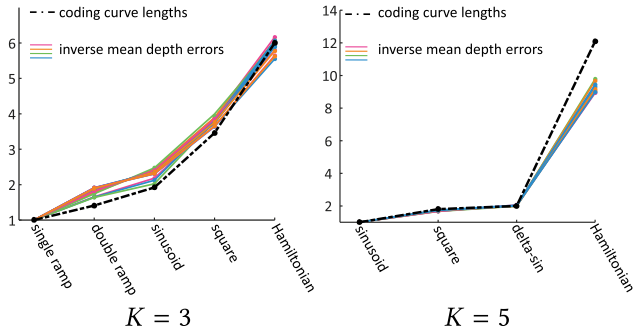


Fig. 6. **Coding curve length vs. mean expected depth error.** Comparison of the coding curve lengths and the *inverse* mean expected depth errors ($\frac{1}{\Delta\Gamma}$) for various coding schemes, for $K = 3$ and $K = 5$. The inverse mean expected depth errors are plotted for several source strength and ambient illumination values (one solid, colored line plot for every source strength and ambient illumination combination). Each individual line plot is normalized by dividing by the minimum value along the plot. This is for better visualization, since different line plots have different minimum values. The normalized coding curve lengths of different schemes are highly correlated with the corresponding inverse mean expected depth errors: the longer the coding curve of a scheme, the lower the mean depth error.

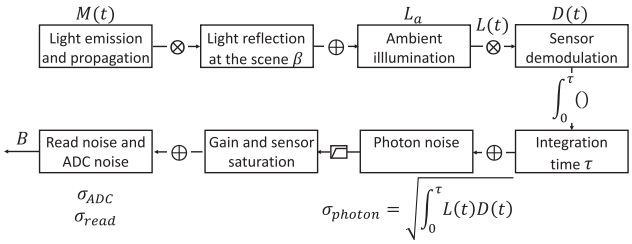


Fig. 7. **Chain of steps for simulation of C-ToF imaging.** We have developed a detailed simulator that simulates various steps of the C-ToF imaging process, including light transport (light emission, propagation, reflection and shading) and sensor physics (demodulation, gain, saturation, ADC noise, quantization). It uses a physically accurate affine noise model, including both photon noise and sensor read noise. This simulator can be used to emulate C-ToF imaging under a wide range of scene and sensor parameters.

low-pass filtered, which may result in loss of performance. We have incorporated bandwidth constraints in our simulator in order to evaluate the performance of various ToF coding schemes on real-world practical devices. Figure 11 shows depth recovery comparisons between ideal Hamiltonian coding functions, and bandlimited Hamiltonian functions (as output by our hardware prototype). The performance degrades by a factor of approximately 1.5 − 3, depending on the SNR level. This performance drop is consistent with our experimental results, which are achieved on an un-optimized, proof-of-concept hardware prototype.

8.3 Frequency Bandwidth Considerations

In the comparisons shown so far in Figures 8, 9, and 10, we used the same fundamental frequency for all coding schemes (10MHz, corresponding to an unambiguous depth range of 15 meters). Although the fundamental frequency is 10MHz, coding functions for non-sinusoid schemes such as square and Hamiltonian contain higher frequency components. It is well known that the depth precision achieved by sinusoid coding schemes is directly proportional to the modulation frequency (Lange 2000). However, high-frequency sinusoids also limit the unambiguous depth range due to phase wrapping (Lange 2000). For example, a 120MHz sinusoid can measure depths only in a small (~1 meter) range. Due to this fundamental tradeoff between depth precision and range, several commercial TOF systems (e.g., Microsoft Kinect v2 (Microsoft-Kinect 2014)) use multiple (at least two) frequency sinusoid coding (Droeschel et al. 2010), where a combination of low and high frequencies are used to simultaneously achieve high depth precision, and large unambiguous depth range. Specifically, phases from the high-frequency sinusoids provide high-precision, albeit *wrapped* (ambiguous), depth estimates. The phases corresponding to the low-frequency sinusoids are used to *unwrap* the high-frequency phases, thus providing unambiguous depth estimates over a large depth range. We call this the *multi-frequency sinusoid* coding scheme.

A related approach is to use multiple high-frequency sinusoids, wherein each high-frequency sinusoid's phase individually provides precise but ambiguous depth information, but the phases are combined to provide unambiguous depth information (Jongenelen et al. 2011). This method is used commercially in

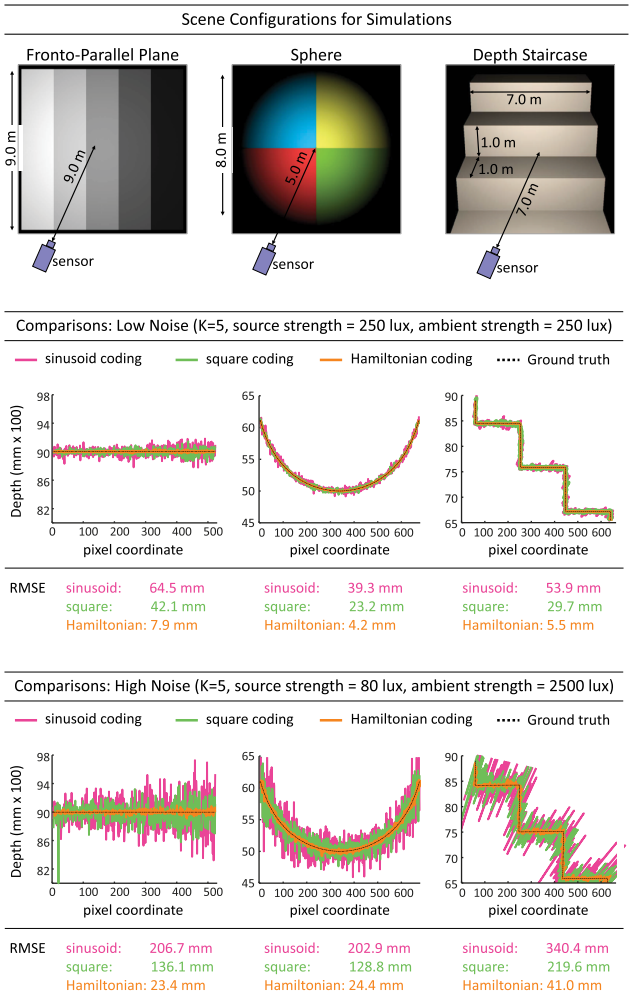


Fig. 8. **Comparisons between various schemes using simulations.** (Top row) We used our simulator to evaluate the performance of various schemes on different scene configurations (fronto-parallel plane with an intensity staircase, a textured hemisphere, and a depth staircase). Middle and bottom rows show depth recovery comparisons in low-noise and high-noise settings, respectively. Depths recovered by conventional sinusoid and square coding have large root mean square errors (RMSE), with respect to the ground truth. In contrast, the proposed Hamiltonian coding can recover accurate shape despite high noise levels, using the **same total light source power and capture time as conventional schemes.**

Texas Instrument's TI OPT8241 ToF sensor (Texas-Instruments 2017). We call this the *high-frequency sinusoid* coding scheme (also called Micro ToF coding in Gupta et al. (2015)).

Comparisons with multi-frequency sinusoid coding: Since the proposed Hamiltonian coding functions have high-frequency components, we perform comparisons with the above-mentioned multi-frequency and high-frequency sinusoid coding schemes. Most of the energy (>99%) of the Hamiltonian coding functions is contained within the first 12 harmonics (e.g., within the frequency range 10−120MHz, for a fundamental frequency of 10MHz). So, we compare with multiple frequency sinusoid schemes with frequencies in the same range, i.e., 10−120Mhz. Specifically, we

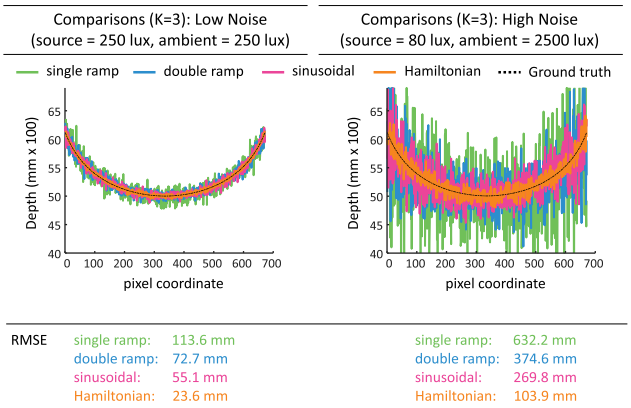


Fig. 9. **Comparisons between various schemes using simulations for $K = 3$.** The same noise and imaging parameters are used as in Figure 8. The relative performance of different schemes in terms of the root mean square depth error is consistent with the respective coding curve lengths (Table 1).

compare with a multi-frequency sinusoid scheme that uses two frequencies (10MHz and 120MHz), and a high-frequency sinusoid scheme that uses both high frequencies (110MHz and 120MHz). For both schemes, we consider $K = 5$ measurements, with three measurements for the lower frequency (phase-shifted by $\frac{2\pi}{3}$) and two measurements for the higher frequency (phase-shifted by $\frac{\pi}{2}$). The precise but ambiguous depths estimated from higher frequency sinusoids was unwrapped using standard phase unwrapping techniques. In order to ensure fair comparisons, the total source energy is also kept the same for all coding schemes.

Figure 12 shows depth recovery results for the two multiple frequency sinusoid coding schemes. As before, we consider a low noise setting ($K = 5$, source strength = 250 lumens, ambient strength = 250 lux), and a high noise setting ($K = 5$, source strength = 80 lumens, ambient strength = 2,500 lux). At low-noise settings, both multiple frequency sinusoid schemes have similar performance as that of the Hamiltonian coding. However, at high noise (low SNR), multiple frequency sinusoid schemes suffer from large depth errors due to inaccurate unwrapping. While it may be possible to reduce the unwrapping errors by using sophisticated algorithms based on spatial smoothness priors (Droeschel et al. 2010), this usually comes at the cost of reduced spatial resolution and high computational costs.

Figure 13 shows the root mean square (RMS) depth error of various schemes (for the plane-with-intensity-staircase scene) as a function of noise. At low noise, the performance of both the multiple frequency sinusoid schemes (multi-frequency sinusoid and high-frequency sinusoid) is comparable to that of the Hamiltonian coding. However, as noise increases, their performance degrades rapidly due to large phase unwrapping errors. On the other hand, the performance of Hamiltonian coding degrades gracefully.

9 HARDWARE PROTOTYPE AND RESULTS

Most commercial C-ToF systems (e.g., PMD, Microsoft Kinect) use sinusoid or square modulation. In order to evaluate different C-ToF coding schemes, we developed a hardware prototype that can implement a wide range of C-ToF coding functions, including the

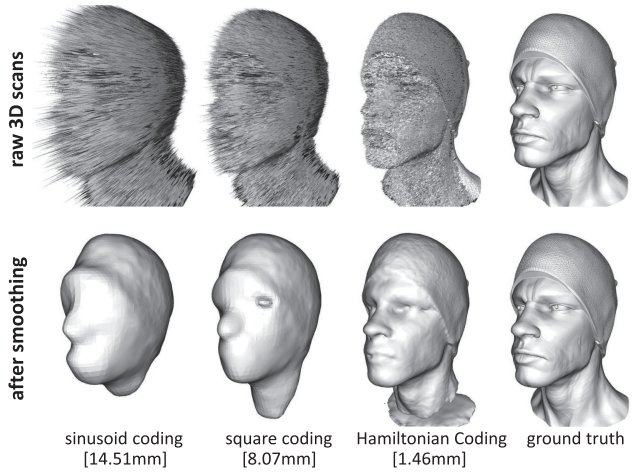


Fig. 10. **Shape recovery comparisons using simulations.** (First row) Comparison of raw 3D shape measurements in strong ambient light, resulting in large photon noise. (Second row) Comparisons after smoothing. The same total source power, capture time, and smoothing parameters were used for all schemes. The numbers in the square parentheses are the mean depth errors, with respect to the ground truth.

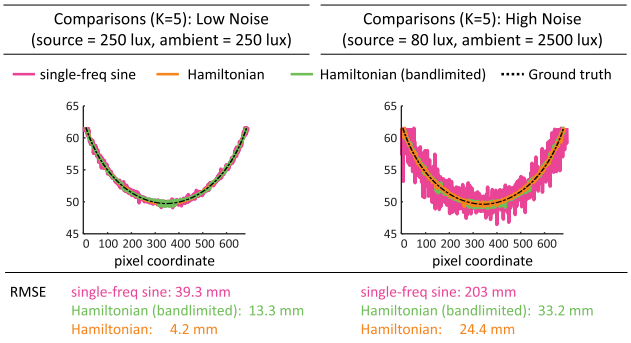


Fig. 11. **Performance of bandlimited Hamiltonian coding scheme.** Comparisons between ideal Hamiltonian coding and bandlimited Hamiltonian functions (as output by our hardware prototype). The performance degrades by a factor of approximately $1.5 - 3$, depending on the SNR level (low noise on left and high noise on right). This performance drop is consistent with our experimental results.

proposed Hamiltonian coding scheme. Our prototype, shown in Figure 14, is based on a single pixel point scanning setup. Our light source is an 830nm laser diode (Thorlabs $L830P200$) that can be modulated with arbitrary waveforms by an external signal of up to 500MHz bandwidth. The modulation signal is provided by a waveform generator, and amplified by an RF amplifier before being applied to the diode. The diode is operated at ~15mW average power, with approximately 40 times peak power (600mW) available. Such high peak power allows us to implement the impulse train modulation functions. The light emitted by the diode is collimated by an aspheric lens, and passes through a pair of galvanometer steered mirrors (Thorlabs $GVS012$), a scan lens, and an objective lens before reaching the scene. The reflected light returns to the objective along the same path as the illumination and is focused onto a photodiode (Menlo Systems, $APD210$) with a bandwidth

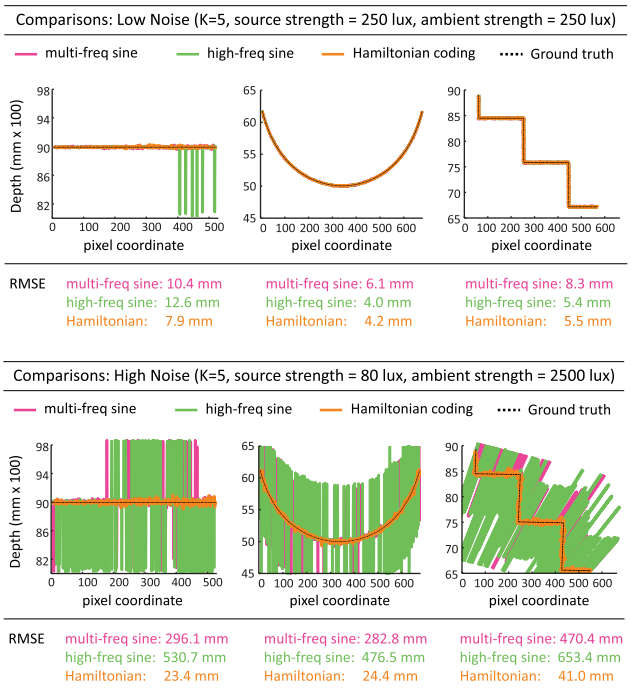


Fig. 12. **Comparisons with multiple frequency sinusoid coding schemes.** We compare Hamiltonian coding with a multi-frequency sinusoid scheme that uses one low and one high frequency (10MHz and 120MHz), and a high-frequency sinusoid scheme that uses two high frequencies (110MHz and 120MHz). The fundamental frequency of Hamiltonian coding functions is 10MHz. (Top row) At low noise settings, both sinusoid-based schemes have similar performance as that of the Hamiltonian coding. (Bottom row) However, at high noise (low SNR), both multiple frequency sinusoid schemes suffer from large depth errors due to inaccurate phase unwrapping, while Hamiltonian coding still achieves high depth precision.

up to 1GHz.[11] The electronic signal is multiplied inside an RF frequency mixer with a local oscillator signal (sensor demodulation function) generated by the second channel of the waveform generator, and then integrated using a low pass filter. Finally, the signal is digitized by a National Instruments $USB - 6000DAQ$.

To scan the beam, we use a second function generator to generate driving voltages for the galvanometer actuated mirrors. The galvanometer drivers provide feedback signals with voltages proportional to the position of the x and y mirrors. These signals are digitized by the data acquisition (DAQ) unit along with each data sample. From this data, an image is created by binning the collected data samples into a two dimensional histogram with the x and y dimensions in the scene forming the histogram axes. In our experiments, we use 400 bins along both axes, resulting in a spatial resolution of 400×400.

[11] It is possible to develop a full-frame prototype system by using sensors based on image intensifier tubes that can be gain-modulated by arbitrary functions (Kawakita et al. 2004), or by using an array of photo-diodes (Shcherbakova et al. 2013). We developed a single photo-diode scanning setup due to its low cost and ease of implementation. Our design specifications can lead to a future integrated solid state device (e.g., PMD, Kinect) that can achieve real-time performance.

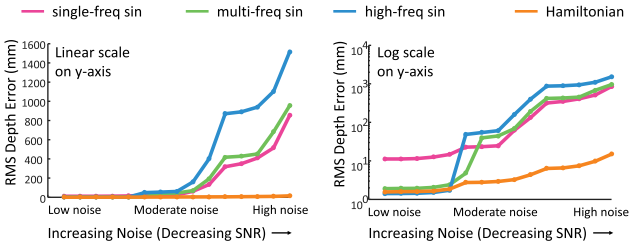
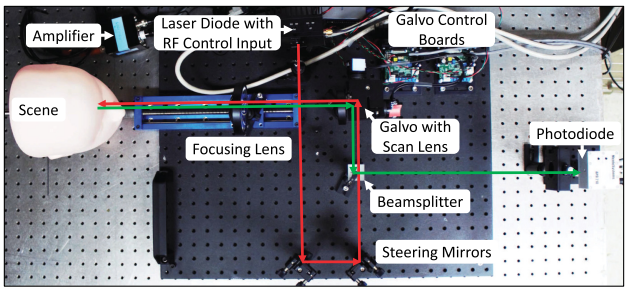
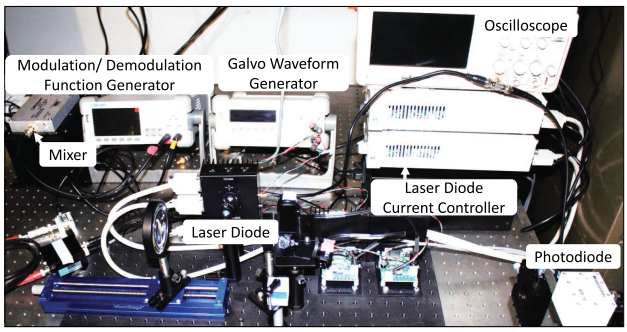


Fig. 13. **Comparisons of coding schemes as a function of SNR.** RMS depth error of various schemes (for the intensity-staircase scene) as a function of noise. (Left) Errors on a linear scale. (Right) Same plots on a log-scale. At low noise levels, the performance of multiple frequency sinusoid schemes is comparable to that of the Hamiltonian coding. However, as noise increases (SNR decreases), their performance degrades rapidly due to large phase unwrapping errors. In contrast, the performance of Hamiltonian coding degrades gracefully.



(a) Top view



(b) Side view

Fig. 14. Hardware prototype. (a) Top view of the setup illustrating the optical path from source to the scene (red arrows), and from the scene to the sensor (green arrows), (b) Side view showing non-optical components such as function generators and mixer.

Hardware limitations and future outlook: There are several challenges associated with developing a prototype with only low-cost off-the-shelf components. The system bandwidth is currently limited to 120MHz by the waveform generator. *We use the same fundamental frequency for all our codes (10MHz, corresponding to an unambiguous depth range of 15 meters). Also, we ensure that the total source power is the same for all coding schemes by using a light meter.* Figure 15 shows the correlation functions for various coding schemes as implemented on our prototype. Due to bandwidth constraints, the light source cannot emit a perfect impulse function, and the higher-order Hamiltonian coding functions are low-pass filtered, resulting in loss of performance. In future implementations, we envision dedicated nonlinear circuits designed to generate specific codes at high speeds, which will potentially achieve close to the theoretical performance.

Experimental Results

Depth recovery of a single scene point: The imaging setup consists of a single planar diffuse patch whose depth is estimated by the sensor, as shown in Figure 16. The patch is placed on a translation stage so that its depth can be varied between 2 meters and 3 meters. We estimated the depth of the patch at several locations along the stage; at every location, the depth was estimated 10^4 times, and the RMS error was computed. Figure 16(b) to (c) show plots of RMS depth errors for sinusoid, square, and Hamiltonian coding, for $K = 4$ and $K = 5$, respectively, as a function of the depth of the patch. Hamiltonian coding achieves considerably lower depth errors as compared to existing methods. The improvements are lower than theoretical prediction because of our un-optimized hardware prototype. As discussed earlier, the hardware limitations are not fundamental, and larger improvements can be achieved with optimized hardware implementation in the future.

Comparisons with multi-frequency sinusoid coding: The imaging setup is similar to that of Figure 16, consisting of a single planar diffuse patch whose depth is estimated by the sensor, as shown in Figure 17(a). For this experiment, the patch was placed at a *fixed* distance of approximately 2 meters. We estimated the depth of the patch at several SNR settings; the SNR was varied by changing the effective light source strength, which was modulated by a rotating neutral density filter wheel placed in front of the light source. At every SNR setting, the depth was estimated 10^4 times, and the RMS error was computed. Figure 17(b) plots the RMS depth errors as a function of source strength (SNR), for single-frequency sinusoid (10MHz), multi-frequency sinusoid (10MHz and 120MHz), high-frequency sinusoid (110MHz and 120Mhz), and the Hamiltonian coding schemes. For both multiple frequency sinusoud schemes, we captured three measurements for the lower frequency (phase-shifted by $\frac{2\pi}{3}$) and two measurements for the higher frequency (phase-shifted by $\frac{\pi}{2}$). As expected (and shown previously via simulations in Figures 12 and 13), as source strength decreases (SNR decreases), the performance of multiple frequency sinusoid schemes degrades rapidly due to large phase unwrapping errors. On the other hand, the performance of Hamiltonian coding degrades gracefully.

Results for 3D scanning: In our first 3D scanning experiment, the scene consists of three planar patches at different depths so that they form a "depth staircase," as shown in Figure 18. The distance between the patches is 2.5cm. Figure 18(c) shows comparison of 3D reconstructions recovered using sinusoid and Hamiltonian coding. The same total exposure time and light source power was used for both schemes. The sinusoid reconstruction has a mean depth error of 1.14cm. Consequently, the underlying geometry is barely discernible. In contrast, the staircase structure is clearly visible in the Hamiltonian reconstruction, which achieves approximately five times lower error as compared to sinusoid coding.

Figure 19 shows 3D scanning comparisons for multiple frequency sinusoid schemes. The scene consists of a single planar

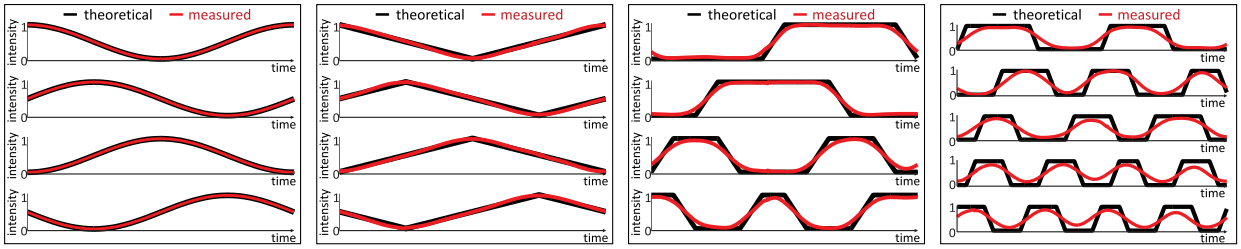
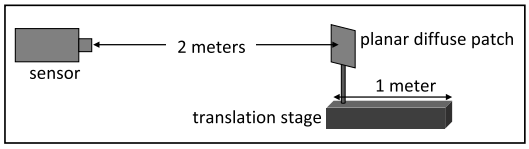
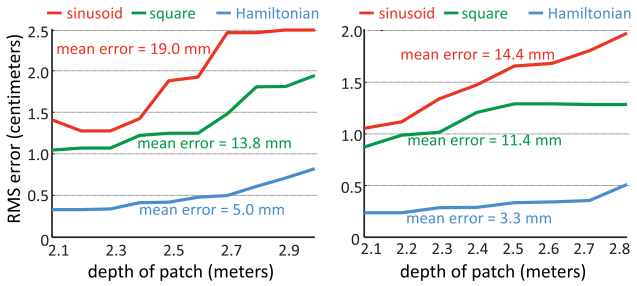


Fig. 15. **Measured vs. theoretical correlation functions.** Correlation functions (shown in red) for various coding schemes as implemented and measured using our hardware prototype. The modulation functions are normalized so that the total emitted energy is the same for all schemes. Due to bandwidth constraints of our current hardware prototype, the higher-order Hamiltonian coding functions are low-pass filtered, resulting in loss of performance. This is not a fundamental limitation and can be mitigated with improved hardware implementation.

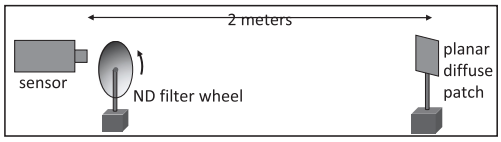


(a) Experimental setup

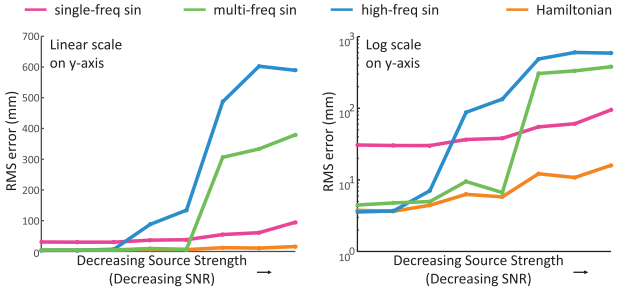


(b) Error comparisons ($K = 4$) (c) Error comparisons ($K = 5$)

Fig. 16. **Depth estimation of a single scene point.** (a) The sensor is used to estimate the depth of a **single scene point** (center of a planar diffuse patch). The patch is placed on a translation stage so that its depth can be varied in the range of 2–3 meters. We estimated the depth of the patch at several locations along the stage; at every location, the depth was estimated 10^4 times, and the RMS error was computed. (b-c) Plots of RMS depth errors for sinusoid, square, and Hamiltonian coding, as a function of the depth of the patch. Hamiltonian coding achieves considerably lower depth errors as compared to existing methods.



(a) Experimental setup



(b) Depth errors vs. source strength

Fig. 17. **Single point depth recovery for multiple frequency sinusoid schemes.** (a) The sensor is used to estimate the depth of a single scene point, at a fixed distance of approximately 2 meters, at several SNR settings. The SNR was varied by changing the effective light source strength via a rotating neutral density filter wheel placed in front of the light source. At every SNR setting, the depth was estimated 10^4 times, and the RMS error was computed. (b) RMS depth errors (linear scale on left, log-scale on right) as a function of source strength, for single-frequency sinusoid, multi-frequency sinusoid, high-frequency sinusoid, and the Hamiltonian coding schemes. As source strength decreases (SNR decreases), the performance of multiple frequency sinusoid schemes degrades rapidly due to large phase unwrapping errors.

diffuse wall, approximately 3 meters away from the sensor. The 3D shape of the scene was measured at three different noise levels (by varying the source strength). At low noise levels, the multiple frequency sinusoid schemes can estimate depths with high precision. However, as noise increases, phase unwrapping errors due to noise result in large depth errors, as indicated in the numbers in parentheses below each image. In comparison, Hamiltonian coding achieves precise depth estimates even at high noise levels.

Figures 1 and 20 show 3D scanning comparisons in two relatively low SNR scenarios: a face scanned with a low-power source, and a low albedo dark object. Both objects were placed approximately 2 meters from the sensor. In both cases, sinusoid reconstructions have strong noise, resulting in large depth errors. Hamiltonian coding is able to recover the overall structure as well as fine details, such as lips and eyes on the face, and the vertical ridge on the vase.

10 LIMITATIONS AND FUTURE WORK

Designing coding schemes with higher performance: Hamiltonian coding achieves substantial improvement over existing methods, but is not provably optimal. The general framework proposed in the article can be used for designing novel schemes in the future that potentially achieve better performance than

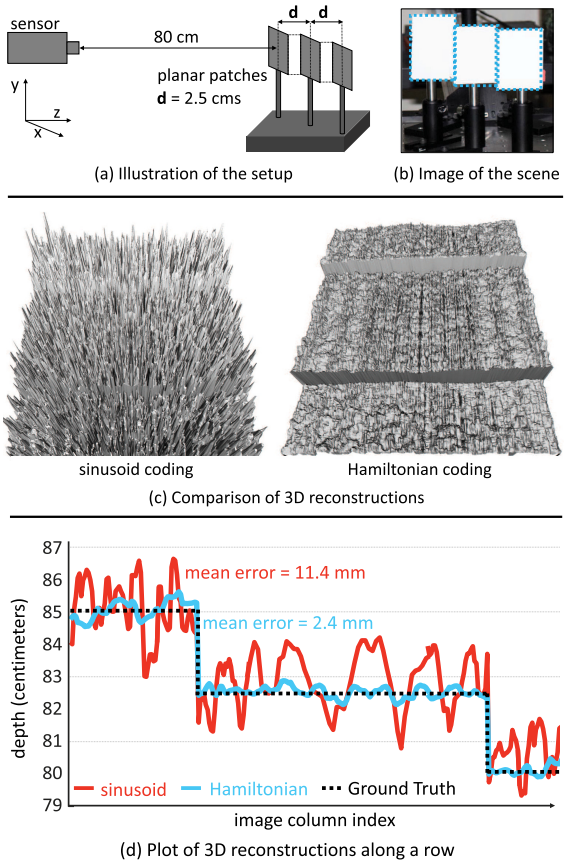


Fig. 18. **Experimental results for 3D scanning of a depth staircase.** (a) The scene consists of three planar patches at different depths so that they form a "depth staircase." The distance between the patches is 2.5cm. (b) An image of the scene from the viewpoint of the ToF sensor. (c) Comparison of 3D reconstructions recovered using sinusoid and Hamiltonian coding. The sinusoid reconstruction has large noise; the underlying geometry is barely discernible. In contrast, the depth staircase structure is clearly visible in the Hamiltonian reconstruction. (d) A plot of the 3D reconstructions along one image scan-line. Hamiltonian coding achieves approximately five times lower mean depth error as compared to sinusoid coding, with the same total capture time and source power.

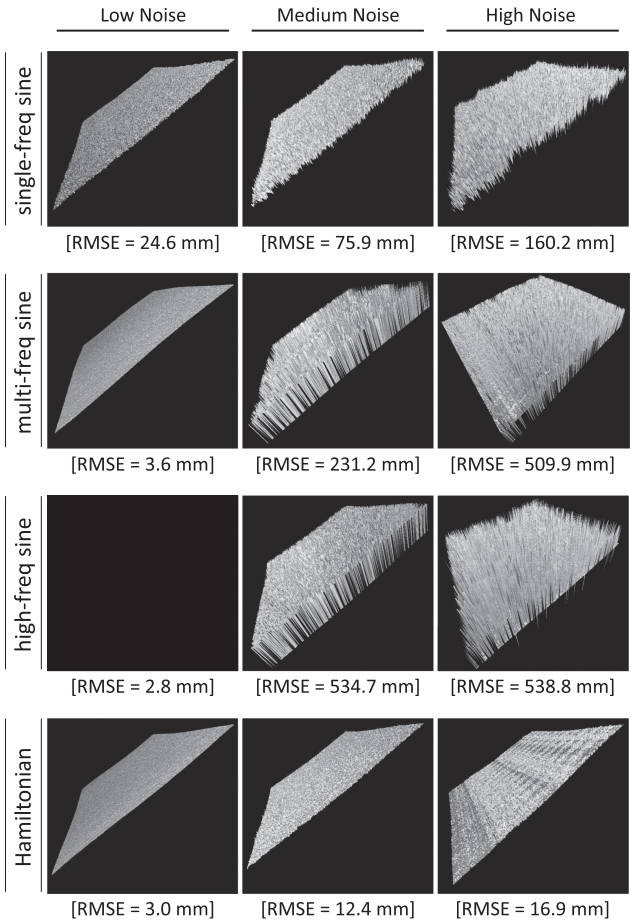


Fig. 19. **Experimental results for 3D scanning of a plane.** The 3D shape of the scene (a planar surface, approximately 3 meters from the sensor) was measured at three different noise levels by varying the source strength. At low noise, the multiple frequency sinusoid schemes can estimate depths with high precision. However, at higher noise, phase unwrapping errors due to noise result in large RMS errors, as indicated in the parentheses. In comparison, Hamiltonian coding achieves precise depth estimates even at high noise levels.

Hamiltonian coding. For example, we have explored coding schemes based on the *family of space filling Hilbert curves (Horn and Kiryati 1997; Sagan 1994)*. Due to their space-filling properties, Hilbert curves can have infinite length. However, Hilbert curves don't have the other desirable properties of C-ToF coding curves. These curves have sharp turns that manifest in poor locality preserving properties. Consequently, even small image noise can result in large depth errors. While a detailed study of Hilbert curves–based coding is beyond the scope of this article, we have performed simulations to evaluate Hilbert and Hamiltonian coding schemes. Our preliminary results (provided in the supplementary technical report) suggest that overall, Hamiltonian coding, given its strong locality properties, outperforms Hilbert coding in most real-world scenarios. In specific situations (very low noise), Hilbert coding can theoretically achieve better performance. An interesting direction of future research is to design *scene-adaptive* ToF systems, which use a different family of codes depending on the noise levels.

Hardware constraints: The coding functions designed in the article so far do not explicitly account for hardware constraints such as limited bandwidth. Increasingly, high-frequency components (Buxbaum et al. 2002) and sources that can admit large peak power (low-cost laser diodes) are being used in C-ToF imaging systems. However, for systems with limited bandwidth and low peak power, the codes designed in this article may require low-pass filtering and clamping, thereby resulting in lower performance. One of the next steps is to incorporate bandwidth and peak power constraints while estimating the modulation and demodulation functions (Section 7.1) so that they can maintain high performance while respecting the practical hardware constraints.

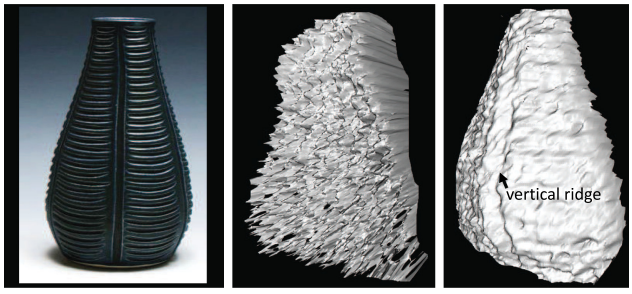


Fig. 20. **Comparison of 3D scanning in low SNR scenarios.** A dark vase with low albedo is scanned using sinusoid and Hamiltonian schemes. The sinusoid reconstruction has strong noise due to low SNR. Hamiltonian coding is able to recover the overall structure as well as fine details, such as the vertical ridge on the vase. See Figure 1 for another comparison.

Multi-path propagation: The image formation model in this article assumes that there is no multi-path light component (inter-reflections). Multi-path propagation is an important problem for almost all active imaging systems, including continuous-wave ToF. Several methods have been proposed to mitigate the (often large) depth errors due to multi-path propagation, including epipolar imaging (O'Toole et al. 2015) and high-frequency coding (Gupta et al. 2015; O'Toole et al. 2014). Since dealing with multi-path is an orthogonal problem to increasing depth resolution, these methods for dealing with multi-path can be used in a complementary manner to the proposed coding schemes. For example, with the same multi-path mitigation technique applied to both Hamiltonian and sinusoid coding, Hamiltonian coding can potentially achieve similar performance gains (as without multi-path) over sinusoidal codes. This represents an exciting research direction. Another potential direction is to extend the geometric code design framework to include global illumination, and to multi-camera arrays (Shrestha et al. 2016). With such a framework, it may become possible to analyze and design C-ToF coding schemes that are robust to multi-path interference (Godbaz et al. 2013; Gupta et al. 2015; Kadambi et al. 2013), as well as coding schemes for multi-source/multi-sensor systems.

Incorporating scene priors in code design algorithms: The depth error measures derived in this article are scene agnostic; they implicitly assume a uniform distribution of scene depths, albedos, and ambient illumination. In general, these distributions may not be uniform. For example, scene depths may have a bi-modal distribution (foreground and background objects). Furthermore, we assumed that the scene albedo factor is independent of scene depths. Incorporating scene priors, such as statistics of natural images (Torralba and Oliva 2003), and designing *scene-adaptive* code optimization algorithms that account for the intensity fall-off is an interesting line of future research.

APPENDIX

A DERIVATION OF DEPTH STANDARD DEVIATION

Consider an unknown point $\mathbf{U} = [\Gamma, \beta, A]$. Given a coding scheme, let the true intensity vector corresponding to $\mathbf{U}$ be $\mathbf{B} = [B_1, B_2, \ldots, B_K]$. Let the noise in intensity measurement B_i be η_i, so that the actual measurement $\widehat{B_i}$ is given as:

$$\widehat{B_i} = B_i + \eta_i. \tag{24}$$

Let the depth estimate corresponding to intensity vectors $\widehat{\mathbf{B}}$ be $\widehat{\Gamma}$. Since $\widehat{\mathbf{B}}$ is a random variable, the estimated depth $\widehat{\Gamma}$ is also a random variable. In order to compute the standard deviation σ_Γ of $\widehat{\Gamma}$, suppose we capture N intensity measurement vectors $\widehat{\mathbf{B}_\mathbf{j}} = [B_{1j}, B_{2j}, \ldots, B_{Kj}]$, $1 \le j \le N$, corresponding to the same unknown point $\mathbf{U} = [\Gamma, \beta, A]$. For each intensity vector measurement, we estimate the depth. Let $\widehat{\Gamma_j}$ be the depth estimate corresponding to the intensity vector $\widehat{\mathbf{B}_\mathbf{j}}$. Then, the standard deviation σ_Γ of the depth estimates $\widehat{\Gamma}$ is given as:

$$\sigma_\Gamma = \sqrt{\frac{1}{N}\sum_{j=1}^{N}\left(\widehat{\Gamma_j} - \mathbf{E}\left[\widehat{\Gamma}\right]\right)^2}, \tag{25}$$

where $\mathbf{E}[\widehat{\Gamma}]$ is the expected value or the mean of $\widehat{\Gamma}$. Since the image noise is assumed to be a zero mean random variable, we assume that $\mathbf{E}[\widehat{\Gamma}] = \Gamma$, where Γ is the true depth. Then, from Equation (25) we get:

$$\sigma_\Gamma = \sqrt{\frac{1}{N}\sum_{j=1}^{N}\left(\widehat{\Gamma_j} - \Gamma\right)^2} = \sqrt{\frac{1}{N}\sum_{j=1}^{N}\left(\triangle\Gamma_j\right)^2}, \tag{26}$$

where $\triangle\Gamma_j = \widehat{\Gamma_j} - \Gamma$ is the depth error corresponding to the j^{th} measurement. Next, we approximate $\triangle\Gamma_j$ by the differential quantity $\partial\Gamma_j$ (first-order approximation), and substituting Equation (11) into the above equation:

$$\sigma_\Gamma = \sqrt{\frac{1}{N}\sum_{j=1}^{N}|\partial\mathbf{B}_j|^2} \times \frac{1}{\beta\sqrt{\sum_{i=1}^{K}\mathcal{F}_i'(\Gamma)^2}}. \tag{27}$$

By substituting $|\partial\mathbf{B}_j| = \sqrt{\sum_{i=1}^{K}\partial B_{ij}^2}$ in the above Equation (27), we get:

$$\sigma_\Gamma = \sqrt{\frac{1}{N}\sum_{j=1}^{N}\sum_{i=1}^{K}\partial B_{ij}^2} \times \frac{1}{\beta\sqrt{\sum_{i=1}^{K}\mathcal{F}_i'(\Gamma)^2}}. \tag{28}$$

Let Ω be the standard deviation of the norm of the intensity vector $|\partial\mathbf{B}_j| = \sqrt{\sum_{i=1}^{K}\partial B_{ij}^2}$. Ω is given as $\Omega = \sqrt{\frac{1}{N}\sum_{j=1}^{N}\sum_{i=1}^{K}\partial B_{ij}^2}$. Substituting in the above equation, we get Equation (12), which expresses σ_Γ in terms of the noise standard deviations Ω.

$$\sigma_\Gamma = \frac{\Omega}{\beta\sqrt{\sum_{i=1}^{K}\mathcal{F}_i'(\Gamma)^2}}. \tag{29}$$

REFERENCES

Amit Adam, Christoph Dann, Omer Yair, Shai Mazor, and Sebastian Nowozin. 2016. Bayesian time-of-flight for realtime shape, illumination and Albedo. *IEEE PAMI* 38, 5 (2016), 851–864.

B. Buxbaum, R. Schwarte, T. Ringbeck, M. Grothof, and X. Luan. 2002. MSM-PMD as correlation receiver in a new 3D-ranging system. *Proc. SPIE* 4546 (2002).

D. A. Carnegie, J. R. K. McClymont, A. P. P. Jongenelen, A. A. Dorrington B. Drayto and, and A. D. Payne. 2011. Design and construction of a configurable full-field range imaging system for mobile robotic applications. *Lecture Notes in Electrical Engineering* 83 (2011).

David Droeschel, Dirk Holz, and Sven Behnke. 2010. Multi-frequency phase unwrapping for time-of-flight cameras. In *Proc. IROS*.

R. Ferriere, J. Cussey, and J. Dudley. 2008. Time-of-flight range detection using low frequency intensity modulation of a CW laser diode: Application to fiber length measurement. *Optical Engineering* 47, 9 (2008), 93602–1 to 93602–6.

D. Freedman, E. Krupka, Y. Smolin, I. Leichter, and M. Schmidt. 2014. SRA: Fast removal of general multipath for ToF sensors. In *Proc. ECCV*.

J. P. Godbaz, A. A. Dorrington, and M. J. Cree. 2013. Understanding and ameliorating mixed pixels and multipath interference in AMCW lidar. In *TOF Range-Imaging Cameras*, Fabio Remondino and D. Stoppa (Eds.). Springer Berlin.

B. S. Goldstein and G. F. Dalrymple. 1967. Gallium arsenide injection laser radar. *Proc. of the IEEE* 55, 2 (1967).

C. Gotsman and M. Lindenbaum. 1996. On the metric properties of discrete space-filling curves. *IEEE TIP* 5, 5 (1996), 794–797.

Frank Gray. 1953. Pulse code communication. *US Patent 2,632,058* (1953).

R. Grootjans, W. Van der Tempel, D. Van Nieuwenhovec, C. de Tandt, and M. Kuijk. 2006. Improved modulation techniques for time-of-flight ranging cameras using pseudo random binary sequences. In *Proc. IEEE LEOS Benelux Chapter*.

M. Gupta, S. K. Nayar, M. Hullin, and J. Martin. 2015. Phasor imaging: A generalization of correlation-based time-of-flight imaging. *ACM Trans. Graphics* 34, 5 (2015), 156:1–156:18.

S. W. Hasinoff, F. Durand, and W. T. Freeman. 2010. Noise-optimal capture for high dynamic range photography. In *Proc. IEEE CVPR*.

F. Heide, M. B. Hullin, J. Gregson, and W. Heidrich. 2013. Low-budget transient imaging using photonic mixer devices. *ACM SIGGRAPH* 32, 4 (2013), 45:1–45:10.

E. Horn and N. Kiryati. 1997. Toward optimal structured light patterns. In *International Conference on Recent Advances in 3D Digital Imaging and Modeling*.

S. Inokuchi, K. Sato, and F. Matsuda. 1984. Range imaging system for 3-D object recognition. In *Proc. IEEE ICPR*.

A. P. P. Jongenelen, D. G. Bailey, A. D. Payne, A. A. Dorrington, and D. A. Carnegie. 2011. Analysis of errors in ToF range imaging with dual-frequency modulation. *IEEE Transactions on Instrumentation and Measurement* 60, 5 (2011), 1861–1868.

A. Kadambi, R. Whyte, A. Bhandari, L. Streeter, C. Barsi, A. Dorrington, and R. Raskar. 2013. Coded ToF cameras: Sparse deconvolution to address multipath interference and recover time profiles. *ACM SIGGRAPH Asia* 32, 6 (2013), 167:1–167:10.

M. Kawakita, K. Iizuka, R. Iwama, K. Takizawa, H. Kikuchi, and F. Sato. 2004. Gain-modulated Axi-vision camera. *Opt. Express* 12, 22 (2004), 5336–5344.

Maik Keller and Andreas Kolb. 2009. Real-time simulation of time-of-flight sensors. *Simulation Modelling Practice and Theory* 17, 5 (2009), 967–978. DOI: https://doi.org/10.1016/j.simpat.2009.03.004

W. Koechner. 1968. Optical ranging system employing a high power injection laser diode. *IEEE Trans. AES* 4, 1 (1968), 81–91.

A. Kolb, E. Barth, R. Koch, and R. Larsen. 2010. Time-of-flight cameras in computer graphics. *Eurographics* (2010).

R. Lange. 2000. 3D ToF distance measurement with custom solid-state image sensors in CMOS-CCD-technology. *Ph.D. Thesis* (2000).

R. Lange and P. Seitz. 2001. Solid state time-of-flight range camera. *IEEE J. Quantum Electronics* 37, 3 (2001), 390–397.

Robert Lange, Peter Seitz, Alice Biber, and Stefan C. Lauxtermann. 2000. Demodulation pixels in CCD and CMOS technologies for time-of-flight ranging. *Proc. SPIE* 3965 (2000).

Stephan Meister, Rahul Nair, and Daniel Kondermann. 2013. Simulation of time-of-flight sensors using global illumination. In *Vision, Modeling and Visualization*, Michael Bronstein, Jean Favre, and Kai Hormann (Eds.). The Eurographics Association. DOI: https://doi.org/10.2312/PE.VMV.VMV13.033-040

Microsoft-Kinect. 2014. March NPD Results: Titanfall on Xbox One is Number One Selling Game. Retrieved from http://news.xbox.com/2014/04/xbox-one-march-npd.

W. H. Mills. 1963. Some complete cycles on the n-cube. *Proc. Amer. Math. Soc.* 14, 4 (1963), 640–643.

M. O'Toole, S. Achar, S. G. Narasimhan, and K. N. Kutulakos. 2015. Homogeneous codes for energy-efficient illumination and imaging. *ACM SIGGRAPH* 34, 4 (2015), 35:1–35:13.

M. O'Toole, F. Heide, L. Xiao, M. B. Hullin, W. Heidrich, and K. N. Kutulakos. 2014. Temporal frequency probing for 5D transient analysis of global light transport. *ACM SIGGRAPH* 33, 4 (2014), 87:1–87:11.

A. D. Payne, A. A. Dorrington, and M. J. Cree. 2010. Illumination waveform optimization for time-of-flight range imaging cameras. In *Proc. SPIE 8085*.

J. M. Payne. 1973. An optical distance measuring instrument. *Review of Scientific Instruments* 44, 3 (1973).

H. Sagan. 1994. *Space Filling Curves*. Springer, New York.

Mirko Schmidt and Bernd Jähne. 2009. *A Physical Model of Time-of-Flight 3D Imaging Systems, Including Suppression of Ambient Light*. Springer Berlin, Berlin, Germany, 1–15. DOI: https://doi.org/10.1007/978-3-642-03778-8_1

R. Schwarte, Z. Xu, H. Heinol, J. Olk, R. Klein, B. Buxbaum, H. Fischer, and J. Schulte. 1997. New electro-optical mixing and correlating sensor: Facilities and applications of the photonic mixer device. In *Proc. SPIE (3100)*.

O. Shcherbakova, L. Pancheri, G.-F. Dalla Betta, N. Massari, and D. Stoppa. 2013. 3D camera based on linear-mode gain-modulated photodiodes. In *Proc. IEEE ISSCC*.

Shikhar Shrestha, Felix Heide, Wolfgang Heidrich, and Gordon Wetzstein. 2016. Computational imaging with multi-camera time-of-flight systems. *ACM SIGGRAPH* 35, 4 (2016), 33:1–33:11.

E. Tadmor, I. Bakish, S. Felzenshtein, E. Larry, G. Yahav, and D. Cohen. 2014. A fast global shutter image sensor based on the VOD mechanism. In *IEEE SENSORS*. 618–621.

Texas-Instruments. 2017. OPT8241 3D Time-of-Flight Sensor. Retrieved from http://www.ti.com/lit/ds/sbas704b/sbas704b.pdf.

A. Torralba and A. Oliva. 2003. Statistics of natural image categories. *Network* 14, 3 (2003), 391–412.

John F. Wakerly. 2005. *Digital Design: Principles and Practices*.

Received June 2017; revised October 2017; accepted October 2017