Single-Photon Structured Light

Varun Sundar [†]   Sizhuo Ma [†]   Aswin C. Sankaranarayanan [‡]   Mohit Gupta [†]

[†] University of Wisconsin-Madison   [‡] Carnegie Mellon University

Abstract

We present a novel structured light technique that uses Single Photon Avalanche Diode (SPAD) arrays to enable 3D scanning at high-frame rates and low-light levels. This technique, called "Single-Photon Structured Light", works by sensing binary images that indicates the presence or absence of photon arrivals during each exposure; the SPAD array is used in conjunction with a high-speed binary projector, with both devices operated at speeds as high as 20 kHz. The binary images that we acquire are heavily influenced by photon noise and are easily corrupted by ambient sources of light. To address this, we develop novel temporal sequences using error correction codes that are designed to be robust to short-range effects like projector and camera defocus as well as resolution mismatch between the two devices. Our lab prototype is capable of 3D imaging in challenging scenarios involving objects with extremely low albedo or undergoing fast motion, as well as scenes under strong ambient illumination.

1. Introduction

Structured light (SL) 3D imaging systems have inherent tradeoffs that balance the precision of the 3D scan against its acquisition time. For instance, temporally-multiplexed SL techniques [21, 22, 25] achieve high depth resolution by projecting multiple patterns, thereby precluding high-speed capture. At the other extreme, SL based on spatially-modulated patterns [27, 62] can facilitate single-shot scans, but require assumptions of spatial-smoothness that invariably result in loss of detail.

The tradeoffs inherent to SL systems are exacerbated when operating in challenging regimes with low signal-to-noise ratios (SNR) arising from either low-albedo objects, dynamic scenes, or strong ambient illumination. In these scenarios, using longer temporal codes can offer robustness and precision, but at the cost of lowered acquisition speed. One way to mitigate the loss in time resolution is to use high-speed cameras and projectors. However, this approach

is limited by large bandwidth requirements and, more fundamentally, the presence of read noise. Each image has a constant amount of read noise, immaterial of the exposure time; this can dominate the received signal as the exposure times and, consequently, the image intensities are reduced.

This paper envisions a class of *Single-Photon Structured Light* systems that are based on single-photon detectors, such as Single Photon Avalanche Diodes (SPADs). SPADs can be operated at very high speeds when detecting photons and not their time-of-arrivals. In this 'photon detection' mode, the measurements are binary-valued—indicating whether or not a photon arrival occurred during a given acquisition time. For instance, a recently developed SPAD array [12] can capture $\sim 10^5$ binary frames at $1/8$-th megapixel resolution. Our key observation is that the binary measurements, normally considered a limitation due to limited information, are sufficient for a large family of SL coding schemes [53] that are binary as well. Since SPADs count photon arrivals, they are not corrupted by read noise. Finally, the use of SPADs for SL finds a natural coupling in high-speed projectors that use digital micromirror devices (DMDs) for displaying binary patterns.

Coding and decoding for Single-Photon SL. Due to the probabilistic nature of photon arrivals, the binary-valued measurements captured by SPADs are prone to strong photon noise. For instance, in the presence of strong ambient light, the SPAD could detect a photon even when the corresponding projector pixel is dark. Traditional SL coding schemes are designed for regimes where the image measurements are not binary-valued, and hence are not suitable for Single-Photon SL. We formulate novel SL encoding strategies using error-correction codes that enable robust decoding for Single-Photon SL even under large photon noise.

Beyond achieving robustness to photon noise, SL coding schemes must account for various imaging phenomena such as projector and camera defocus. Naïve error-correcting codes do not consider these practical effects, and thus cannot be used in a real-world SL system. We design a new class of hierarchical codes using error correction and binary phase shifting that guarantee a minimum stripe width, which provides robustness to such non-idealities. Finally, we design a high-throughput decoding scheme for the proposed codes to enable real-time decoding of the measurements. Our implementation can decode a disparity map for

Figure 1. **Single-Photon Structured Light.** *(left)* Our proposed system comprises of a SPAD array and a DMD-based projector used to project and acquire binary patterns at extremely high frame rates. We devise coding schemes that can obtain depth-maps from these photon noise limited captures. *(right)* We demonstrate several practical capabilities of Single-Photon SL, including: high-speed scanning of a rapidly moving hand at 250 FPS, sub-millimeter precise depth-imaging at a range 50 cm range and at 80 FPS, and reconstructing the tread pattern of a tire, a low-albedo object, at 40 FPS. We include RGB frames captured at 30 FPS to depict the high speeds involved.

a 512×256 array in 100 ms on a CPU and 3 ms on a GPU.

Implications. Single-photon SL has the potential to enable *extreme 3D imaging capabilities*, including high-speed 3D scanning and robust 3D imaging in low-SNR conditions while respecting low-power and latency budgets. Figure 1 demonstrates several unique practical capabilities of our prototype Single-Photon SL system, including scanning scenes with low albedo (a tyre) and at high frame rate (fast hand movements) with little loss in the spatial resolution.

Limitations. Single-Photon SL inherits limitations endemic to many SL systems. While we mitigate short-range effects such as defocus, long-range effects such as inter-reflections remain to be addressed, possibly by incorporating existing work addressing global illumination. Current SPAD technology is still nascent compared to its CMOS counterparts; the low-resolution of SPAD arrays and their poor fill factors constrains the reconstruction quality of our approach. Fortunately, the capabilities of single-photon sensors continue to improve with higher resolution arrays featuring increased fill factors [47, 48] on the horizon.

2. Related Work

Structured light 3D imaging. Active triangulation techniques have a rich history with early techniques including stripe scanning [3, 38, 57], shadow scanning [7, 8], binary patterns [53] and sinusoid patterns based phase-shifting [58]. Many methods achieve fast single-shot acquisition by projecting statistical patterns [34, 55, 63] or via Fourier Transform Profilometry (FTP) [4, 32, 33]. Such techniques require spatial-smoothness assumptions and have low accuracy for strongly textured surfaces.

Fast binary projectors in SL. Several SL systems achieve high-speed 3D scans [27,35,36,59,62,64] using the projection capabilities of DMDs. However, all of these techniques use sensors based on traditional photodiodes which, unlike SPADs, are fundamentally limited by read noise.

Event-based 3D imaging. Event-based cameras are bio-inspired devices [17] that are triggered asynchronously by intensity changes (or "events") typically from a pulsed laser [9, 43, 45] or by a DMD projecting multiple patterns [26, 37, 42]. In contrast to SPADs, event-cameras have 1-2 orders of magnitude lower event density ($\sim 10^6$ events/s [39]). Since each event recovers at most a single 3D point, the limited event density lowers the density and quality of the reconstruction especially in presence of scene-wide motion. While event-based cameras can achieve high dynamic range, their low-light sensitivity remains poor, precluding reliable 3D imaging in low albedo and low SNR scenarios.

Single-Photon imaging. Only recently have the capabilities of SPADs, operating without any temporal synchronization, been explored, with applications in high-dynamic range imaging [28, 29] and burst photography [41]. Our method operates the SPAD array similar to Ma et al. [41], using a sequence of binary frames. Although we focus on SPADs due to their superior frame-rate, the proposed techniques are applicable to other single-photon imaging technologies such as Jots [14,15], which feature high-resolution arrays with smaller pixels and increased photon-efficiency [40], albeit at lower frame-rates and higher read noise.

3. Image Formation in Single-Photon SL

Consider a SPAD pixel array observing a scene. The number of photons N arriving at a pixel $\mathbf{x}$ during an expo-

Figure 2. **Single-Photon SL features an asymmetric noise model.** Bit-flip probabilities as determined by Eqs. (3) and (4) are evaluated across a grid of (Φ_a, Φ_p) flux values. The plot parameters are $t_{\text{exp}} = 10^{-4}s$ and dark current rate $r_q = 0$.

sure time t_{exp} is modelled as a Poisson random variable:

$$\Pr\{N = k\} = \frac{(\Phi(\mathbf{x})\, t_{\text{exp}})^k \; e^{-\Phi(\mathbf{x})\, t_{\text{exp}}}}{k!}, \quad (1)$$

where $\Phi(\mathbf{x})$ is the flux[1]. During each exposure, a pixel detects at most one photon, returning a binary value $B(\mathbf{x})$ such that $B(\mathbf{x}) = 1$ if the pixel detects one or more photons. Hence, $B(\mathbf{x})$ is a Bernoulli random variable [61] with

$$\Pr\{B(\mathbf{x}) = 0\} = e^{-(\Phi(\mathbf{x})+r_q)t_{\text{exp}}}, \quad (2)$$

where r_q is the dark current rate—the rate of spurious counts unrelated to incident photons.

In a typical SL scan, the scene is illuminated with a sequence of 2D binary patterns from a projector. The SPAD captures a binary frame for each pattern. Each SPAD pixel receives a binary code over time, from which we estimate the projector column observed at the pixel—an operation that is critical for the success of any SL technique.

We now derive the probability that a projected temporal sequence will be decoded incorrectly. For a given binary pattern, consider a SPAD pixel $\mathbf{x}$ that observes a scene point illuminated by an ON pixel. Suppose the incident photoelectron arrival rate at $\mathbf{x}$ due to projector and ambient illumination are $\Phi_p(\mathbf{x})$ and Φ_a, respectively. Then, the probability of a bit-flip error, i.e., the probability of the SPAD pixel not detecting a photon is given as

$$\text{P}_{\text{flip, bright}} = \Pr\{B(\mathbf{x}) = 0 \,|\, \Phi(\mathbf{x}) = \Phi_a + \Phi_p(\mathbf{x})\}. \quad (3)$$

Similarly, the probability of a bit-flip error when the projector pixel is OFF, i.e., the probability of detecting a photon in spite of not illuminating the corresponding projector pixel is

$$\text{P}_{\text{flip, dark}} = \Pr\{B(\mathbf{x}) = 1 \,|\, \Phi(\mathbf{x}) = \Phi_a\}. \quad (4)$$

The bit-flip probabilities for "bright" and "dark" pixels, visualized in Fig. 2 for varying Φ_a and Φ_p, are not equal due to the asymmetric role played by the ambient photons.

We can now compute the probability of incorrectly decoding of an L-length Gray code. Since it is equally likely

---

[1]For simplicity, we assume a 100% quantum efficiency and use the term "flux" interchangeably with the arrival rate of photo-electrons.

	$\text{P}_{\text{flip, dark}}$	$\text{P}_{\text{flip, bright}}$	$\Pr\{\text{error}\}$
dark room	0.021	0.22	0.72
indoor lamp	0.23	0.19	0.9
spot lamp	0.75	0.06	0.99

(a) Bit-flip probabilties

(b) Binary frames captured by Single-Photon Camera

Figure 3. **Typical bit-flip probabilities** observed in our lab prototype under different ambient illumination conditions at an exposure of 10^{-4}s. (a) Using these probabilities, we can compute the decoding error probability for a code sequence of length $L = 10$. (b) We project a pattern where the left half is dark and the right half is bright. At higher ambient intensities, more photons are detected in the "dark" region—which makes reliable decoding a challenge.

to observe any L-bit binary code, the average probability of erroneous decoding over all codewords is:

$$\Pr\{\text{error}\} = 1 - \left(1 - \left(\frac{\text{P}_{\text{flip, bright}} + \text{P}_{\text{flip, dark}}}{2}\right)\right)^L. \quad (5)$$

A detailed derivation is provided in Suppl. Sec. 1.2.

Typical decoding error probabilities. Figure 3 shows the decoding error probability for a 10-bit Gray code across ambient light levels. Increasing ambient light levels drastically increases $\text{P}_{\text{flip, dark}}$, resulting in a near-certain decoding failure. In the next section, we propose coding strategies for Single-Photon SL that enable accurate decoding even in highly challenging conditions.

4. Coding for Single-Photon SL

We now describe temporal coding schemes for Single-Photon SL with the goal of achieving robustness to random bit-flips using error correction mechanisms, and incorporating practical considerations in code design. The overall coding and decoding pipeline is illustrated in Fig. 4.

One simple strategy to improve the reliability of any scheme is to repeat the projected patterns and perform a majority vote. This seems to be a viable option since the high-speed projection and capture of DMDs and SPADs, respectively, affords high temporal redundancy. For example, given a code sequence of length $L = 10$, we could simply repeat the patterns 25 times (called the redundancy factor), and still maintain a high overall frame rate for 3D capture. Such a majority vote will improve the decoding performance, provided the probability of bit flips is less than 0.5. However, in extreme conditions (say, low SNR) a large

Figure 4. Overview of Single-Photon SL. (a) Our coding scheme assigns to each projector column a unique binary message. The most significant bits (MSBs) of this message are transformed using error-correcting encoders (Sec. 4.1), while the least significant bits (LSBs) are encoded using circularly shifted binary patterns (Sec. 4.2). (b) This *hybrid* strategy offers robustness to bit-flips arising from photon noise while guaranteeing a minimum stripe-width, thereby retaining its effectiveness even in the presence of short-range non-idealities. (c) Captured frames, acquired here in 25ms, feature a large amount of photon noise. (d) Similar to encoding, the MSBs and LSBs portions are decoded separately, using minimum distance decoding and a matched filter respectively. In Sec. 4.3, we present a high-throughput decoding procedure that outputs correspondences in real-time. (e) Correspondence maps and the final 3D reconstruction illustrate the coarse-to-precision approach of the Hybrid strategy. Zoomed-in insets of correspondence maps are plotted with a different colormap.

redundancy factor may be needed to achieve even a modest improvement.

Can we design coding schemes for Single-Photon SL that perform better than simple repetition? It is well-known in coding theory that repetition is a sub-optimal error-correcting mechanism [11]. We, instead, turn to a popular family of binary error-correcting codes, the Bose–Chaudhuri–Hocquenghem (BCH) codes [6], used in applications ranging from QR codes [12] to satellite communication [16]. Our choice of BCH is motivated by their ability to correct stochastic errors and their flexibility in designing codes with varying redundancy factors—which permits a graceful tradeoff between speed and robustness.

4.1. BCH Codes

A BCH(n, k, d) : $\{0,1\}^k \to \{0,1\}^n$ encoder takes input messages of length k and produces output codewords of length n that are at least d-bits apart. Hence, such a coding scheme provides error correction capabilities up to $\lfloor \frac{d-1}{2} \rfloor$ bit flips, in the worst case or adversarial sense. This worst-case error-correcting capacity of BCH codes is significantly higher than of the repetition code. For instance, BCH$(63, 10, 27)$—which uses 63-length codewords to encode messages of length 10—can correct up to 13 worst-case errors, while the corresponding capability for Repetition$(60, 10, 6)$, where the message pattern is repeated 6 times, can correct only 2 errors. Going further, in our problem setting, the main source of bit flips is photon noise,

which is stochastic and non-adversarial, and thus we can expect error correction beyond the worst-case regime.

Designing BCH-encoded patterns. Consider a projector with C columns. We aim to design projector patterns so that each column is assigned a unique binary code with in-built BCH error correction. To produce the projector patterns, we start with a base binary coding scheme that uniquely represents each projector column, for example, with Gray codes [30]. Given a set of message codes $\{m_i\} \subseteq \{0,1\}^L$, where $L = \lceil \log_2 C \rceil$, we choose a BCH encoder $\mathcal{E}_{\text{BCH}(n,k,d)}$ that is capable of encoding at least C messages (i.e., $k \geq L$) and output column-wise codes $\{c_i = \mathcal{E}_{\text{BCH}(n,k,d)}(m_i)\}$.

As an example, Fig. 5 illustrates the BCH$(31, 11, 11)$ encoding of 10-bit Gray code messages. Since $L < k$ here, we use *shortening*, i.e., we prepend the message by $(k - L)$ zeros, but do not transmit them, thereby reducing the projected code length from n to $n - (k - L)$. We also use *systematic encoding*, i.e., the first L-bits of each code is the message itself—hence each sequence comprises of message patterns appended by parity patterns. The choice of n, the length of the BCH code, simultaneously determines the robustness of the code as well as the loss in time resolution. Longer codes have better error-correction capabilities; but since we need to acquire a larger temporal sequence, this reduces our ability to handle fast(er) moving objects. With this in mind, in the rest of the paper, we present results at two operating points—$n = \{63, 255\}$—to cover two distinct scenarios.

Evaluating BCH encoding for Single-Photon SL. To un-

Figure 5. **BCH(31, 11, 11) encoding of a 10-bit Conventional Gray message.** We use *systematic* encoding, where message patterns are appended by parity patterns, providing tolerance to random bit-flips caused by photon noise. The complete code lookup table describing these patterns is shown in Suppl. Fig. 3.

Figure 6. **Monte-Carlo evaluation of BCH and repetition strategies.** We empirically evaluate the probability of decoding error upon receiving a codeword randomly corrupted by bit-flips. The ambient flux (Φ_a) and projector flux (Φ_p) values at a pixel location determine the bit-flip probability. We use BCH encoders with $n = \{63, 255\}$. Both repetition and BCH strategies improve the robustness of conventional Gray codes to photon noise. Additionally, BCH outperforms repetition at all (Φ_a, Φ_p) with a pronounced difference at higher redundancies.

derstand the benefits of BCH encoding, we use Monte-Carlo simulation of decoding error probability across a grid of (Φ_a, Φ_p) values, and compare conventional Gray codes, repetition codes and BCH codes for 10-bit binary messages. The performance of these schemes is presented in Fig. 6. At most operating points, the decoding error probability of BCH codes is either close to zero or presents an order of magnitude improvement over repetition.

4.2. Code Design under Practical Considerations

Beyond achieving robustness to photon noise, SL coding schemes must account for various imaging non-idealities. A majority of the BCH encoded patterns, as seen in Fig. 5, comprise of high-spatial frequency patterns that do not perform well under projector/camera defocus and resolution mismatch between the devices. Therefore, in spite of achieving low errors in theory, BCH codes as described so far will simply be inadequate in a practical SL system.

A common approach to mitigate such short-range effects is to use long-run Gray codes, which are a subset of Gray codes that maximize the shortest stripe width [18, 19, 23] across all the projected patterns. However, applying BCH encoding on long-run Gray codes also results in a majority of patterns containing high-spatial frequencies (see Suppl. Fig. 4). Finding binary messages $\{m_i\}$ that maximize the minimum stripe-width of BCH patterns $\{c_i\}$ is an intractable combinatorial problem with an exorbitant solution space (1024! candidate solutions).

Hybrid codes. Our key idea is to design hierarchical codes where BCH encoding is performed only on the more significant bits (MSBs) of the base Gray code pattern. This ensures that all the BCH-encoded frames have large minimum stripe widths, making them robust to defocus effects. The remaining lower significant bits (LSBs) are resolved using circularly-shifted binary patterns, where we shift the pattern one-pixel-at-a-time to represent columns sequentially. We term this as "binary shifting".

Specifically, given a L-bit message, we encode its L_{BCH} MSBs and the remaining $L_{\text{shift}} = L - L_{\text{BCH}}$ LSBs in dif-

ferent ways. The MSBs are coded using BCH as described earlier; since the message codes corresponding to a specific MSB pattern remains unchanged for all values of the LSBs, the resulting BCH codes have a stripe width of at least $2^{L_{\text{shift}}}$. The L_{shift} LSBs are coded by a temporal sequence of length $2^{L_{\text{shift}}+1}$ featuring a burst of $2^{L_{\text{shift}}}$ ones—whose starting position (or phase) encodes the message. Figure 8 illustrates the codewords arising for this hybrid construction, which are guaranteed to have an overall stripe width of at least $2^{L_{\text{shift}}}$ pixels. In our implementation, for a $L = 10$-bit message, we set $L_{\text{BCH}} = 7$ and $L_{\text{shift}} = 3$, featuring a minimum stripe width of $2^3 = 8$ pixels. Additionally, we utilize BCH encoders with $n \in \{63, 255\}$.

We note that this coding scheme is similar in spirit to hybrid SL techniques [11, 20, 64, 65] where Gray codes provide global disambiguation and Phase Shifting [21] resolves LSBs, providing precise correspondences. However, binary shifting has a key difference compared to phase shifting in that intensity information cannot be inferred from a binary measurement. Consequently, unlike phase shifting, where a single measurement can determine the unwrapped phase, binary shifting requires projecting multiple patterns.

Binary shifted patterns are decoded using a matched filter approach, by autocorrelating the received sequence with an unshifted stripe sequence. In Suppl. Sec. 1.5, we show that binary shifted patterns offer significant robustness to random bit-flips by deriving the expected decoding error. Finally, to illustrate the hybrid codes' overall error-correcting capability, we compare them to repeated long-run Gray codes of similar codelength. We characterize performance using root mean squared error (RMSE) in decoded correspondence as the error metric. As we observe in Fig. 8, hybrid codes outperform repeated Gray codes, and the performance gap increases at higher redundancy factors.

Figure 7. **Hybrid encoding of a 10-bit conventional Graycode message**. We use BCH encoding for the first 7 bits of the message and binary shift encoding for the last 3 bits, resulting in a minimum stripe width of 8 pixels. Hence, these patterns do not possess high spatial-frequency, and are more robust to short-range effects.

BCH Message (7 patterns) +

BCH Parity (20 patterns) +

Binary Shifting (16 patterns)

Low Redundancy n = 63   High Redundancy n = 255

Figure 8. **Hybrid codes are more robust to photon noise than repeated Gray codes, evaluated across a grid of (Φ_a, Φ_p) photon fluxes.** We are interested in patterns robust to projector defocus and resolution-mismatch effects and hence consider repetition with long-run Gray codes. For a 10-bit message, both repeated long-run Gray and Hybrid codes have a minimum stripe width of 8 pixels. We use RMSE (in correspondence) as the evaluation metric to account for the locality in decoding error.

4.3. High-throughput Decoding Algorithms

In applications such as robotic navigation, augmented reality and high-throughput industrial inspection, it is imperative to obtain depth maps at speeds comparable to frame acquisition. For conventional codes, like Gray and repetition codes, decoding can be performed via fast analytical algorithms. While analytical decoding methods (e.g., the Berlekamp Massey algorithm [9]) with polynomial runtime exist for BCH codes as well, these methods can only correct up to the worst-case error (Hamming) limit, which is insufficient due to a potentially large number of bit-flips caused by photon noise. In this section, we discuss fast decoding techniques for BCH and, as an extension, the hybrid codes proposed in Sec. 4. The goal is to design techniques that can achieve real-time decoding, while also being able to handle a large number of individual bit-flip errors.

Minimum distance decoding for Single-Photon SL. One simple decoding approach is Minimum Distance Decoding (MDD), where the measured codewords are compared against every projected code word. MDD, while conceptually simple, can correct errors beyond the worst-case limit[2]. However, a brute-force implementation of MDD can often be unviable, owing to its exorbitant run-time and/or memory requirements. Fortunately, Single-Photon SL has certain favourable properties that lead to a fast, high-throughput MDD procedure. First, the space of messages (number of projector columns, $\sim 2^{10}$) is significantly smaller than space of codewords (2^n, $n \in \{63, 255\}$). Second, the number of queries for decoding, which is the number of pixels in the SPAD sensor, exceeds the number of messages.

These circumstances permit us to leverage the recent progress in similarity search [5, 56], which has lead to ef-

ficient nearest-neighbour algorithms for batched queries. Based on empirical comparisons (presented in Suppl. Fig. 7), we find that FAISS [8] offers the highest throughput, decoding a $1/8$th MP array in 100 ms on CPU and 3 ms on GPU. Such methods also scale to larger arrays, requiring 12 ms and 30 ms for one and four megapixels respectively.

5. Experimental Results

We now describe a range of experiments to demonstrate the performance of Single-Photon SL. Our lab prototype was constructed with the SwissSPAD2 array [12], which is a 512×256 SPAD array. The array has a pixel pitch of 16.38 μm, and can capture binary frames at speeds up to 100 kHz. In Suppl. Sec. 4, we provide additional details regarding the setup including the calibration procedure used.

5.1. Single-Photon SL on Static Scenes

To characterize the performance of Single-Photon SL, we image static scenes of varying albedo and ambient light levels. We use these case studies to compare the performance of different error correction schemes and to show the effectiveness of the proposed hybrid codes. To obtain ground truth scans, we operate a DMD projector at a low frame-rate of 2 Hz, while running the SPAD at 10240 Hz, thereby obtaining 5120 SPAD frames per projected pattern. The average of 5120 frames has minimal photon noise, and is considered as a ground truth measurement.

Performance of proposed codes. Figure 9 compares our proposed Hybrid and BCH strategies to repeated Gray codes and repeated long-run Gray codes. We report overall RMSE and RMSE among inliers, thereby measuring both accuracy and consistency. As seen in Sec. 4.2, the proposed hybrid codes are considerably more consistent and accurate across the two redundancy factors used. Whereas, naïve

---

[2]The exact number of correctable errors depends on (Φ_a, Φ_p); with MDD being the Maximum Likelihood Decoder when $P_{\text{flip, bright}} \approx P_{\text{flip, dark}}$.

Figure 9. **Strategy comparison for Single-Photon SL on porcelian bust.** *(left)* We obtain ground truth by averaging the burst of 5120 binary frames captured for each projected pattern. To illustrate the challenge of photon noise, we include reconstruction using Gray code without repetition which has severe artifacts. *(right)* Comparison between our proposed Hybrid strategy and other baseline methods across operating points $n = \{63, 255\}$. BCH codes having several high spatial-frequency frames, are easily distorted by short-range effects. We report three metrics: RMSE in estimated depth, percentage of inliers (absolute depth error < 5mm), and RMSE among these inliers.

Figure 10. **Effect of ambient light intensity.** Hybrid codes, shown here for $n = 255$, are fairly robust to indoor ambient light and recover coarse shapes even under bright ambient light. We report the average Photons incident Per Pixel (PPP) during an exposure window, as a measure of ambient illumination.

BCH codes are heavily distorted due to defocus. Since each strategy has access only to a single binary frame per projected pattern, this emulates a *3D capture speed* of 40 FPS in Hybrid ($n = 255$) and 130 FPS in Hybrid ($n = 63$).

Low SNR regimes. Figure 10 examines reconstruction quality across various ambient light sources, including indoor lighting and a bright work lamp. The reconstructions, shown for Hybrid ($n = 255$), are robust to ambient light, albeit with a drop in performance under the work lamp. These results can potentially be improved by judicious use of light redistribution schemes [24, 45, 49] and exposure control.

Next, we consider low-albedo scenes by imaging objects covered by highly absorptive materials, such as 3M Black Matte [1] and Acktar Velvet [2]; the latter absorbs upto 99.9% of incident light. As Fig. 12 shows, Single-Photon SL can recover the 3D geometry of these dark objects, even when visually imperceptible. As a practical example, Fig. 1 shows the reconstruction of tire treads scanned at 40 FPS.

5.2. Dynamic Scenes with a High-Speed Projector

For dynamic scenes, we used a projector based on the Texas Instruments DLP6500 DMD, capable of projecting binary images with a resolution of 1024×768 pixels at 20 kHz. For simplicity, we operate the SPAD at the same speed as the projector. The projector uses a broadband white LED (SugarCUBE Ultra White LED) as the illumination source.

Figure 13 shows high-speed 3D imaging for a sequence of fast hand movements. A commercial 3D scanner (Kinect-v2 camera [63]) operating at 30 FPS fails to recover the fingers of the rapidly moving hand, while Single-Photon SL continues to recover fine details. Finally, in Fig. 11, we reconstruct the deforming folds of a cloth as it is waved in front of the camera. For both sequences, we use Hybrid ($n = 63$) operated at 250 FPS. These demonstrate the ability of proposed Single-Photon SL techniques to recover detailed 3D geometry of high-speed deformable objects.

| (a) Averaged SPAD frames | (b) Binary Frame | (c) Single-Photon SL 3D Reconstruction |

Figure 11. **Non-rigid deforming object captured by Single-Photon SL** using Hybrid ($n = 63$) at 250 FPS. We include (a) a reference image captured by the SPAD camera using a long integration time, (b) a single binary frame and (c) the reconstructed meshes clearly showing the folds of the cloth. Capturing a non-rigid object is particularly challenging—unless we operate at high speeds, excessive motion blur is induced. **We include a high-speed depth video of this sequence in the supplementary material.**

Reference image   Binary Frame   3D Reconstruction

(a) 3M Black Matte (absorbance 99%)

(b) Acktar Metal Velvet (absorbance 99.9%)

Figure 12. **3D reconstruction in low-albedo scenes** of (a) a white cylinder covered by *3M Black Matte* and (b) an inverted V-groove covered by *Acktar Metal Velvet*; the materials used have extremely high absorbance of 99% and 99.9%, respectively. Both reconstructions are obtained using Hybrid ($n = 255$) at 40 FPS.

In summary, our results on these challenging scenes—both static and dynamic—illustrates the practical capabilities of the Single-Photon SL modality and its ability to simultaneously achieve high speed, precision and robustness.

6. Conclusion and Discussion

This paper shows that many of the tradeoffs inherent to SL systems can be addressed via the use of SPAD sensors. Single-Photon SL, the system that we propose, exploits the single photon detection capabilities of SPAD sensors, along with its lack of read noise. The proposed ideas are capable of detecting objects with high absorbance, and scenes with dramatic high-speed motion. The enabling techniques underlying these are a set of error correction codes, that are designed to be resilient to aberrations commonly present in SL systems. As with many recent efforts in this space, Single-Photon SL provides yet another case study for the wider adoption of SPAD sensors in the imaging community.

Optimal coding and decoding. Despite their effectiveness, the proposed coding and decoding strategies are not

t = 0s   t = 0.31s   t = 0.62s

(a) RGB frames

(b) Kinect-v2

(c) Single-Photon SL

Depth (in mm)

Figure 13. **Comparison to Kinect-v2 on fast hand movements.** We use Hybrid codes ($n = 63$) at 250 FPS to capture these rapid movements. Only the palm, and no fingers, is visible in the Kinect's depth maps. The motion blur distorted RGB frames, acquired by the Kinect at 30 FPS, convey the speed involved.

provably optimal. Our MDD implementation is agnostic to the asymmetry of bit-flips, a defining feature of Single-Photon SL. While maximum likelihood decoders remain to be constructed for the general asymmetric case, optimal decoders have been derived for edge cases [13], e.g., when $P_{flip, bright} >> P_{flip, dark}$. Further, leveraging an optimization framework [46] such as optical SGD [10] can lead to improved coding and decoding schemes for Single-Photon SL.

Handling ambient illumination. In strong ambient illumination, bit-flips arising from $P_{flip, dark}$ are predominant. An important next step is to exploit such asymmetry of bit-flips to design Single-Photon SL codes that are optimized for extreme ambient illumination. Finally, another promising research direction is to explore Single-Photon SL with complementary modalities such as light concentration [24, 50] and epipolar structured light [51, 52] to gain further robustness to extreme ambient and global illumination.

References

[1] 3M Wrap Film Series dead matte black. https://www.3m.com/3M/en_US/p/dc/v100534150/. Accessed: 2021-11-15. 7

[2] Acktar Metal Velvet optical black coating. https://www.acktar.com/product/metal-velvet-2/. Accessed: 2021-11-15. 7

[3] Gerald J Agin and Thomas O Binford. Computer description of curved objects. *IEEE Transactions on Computers*, 25(04):439–449, 1976. 2

[4] Yatong An, Jae-Sang Hyun, and Song Zhang. Pixel-wise absolute phase unwrapping using geometric constraints of structured light system. *Opt. Express*, 24(16):18445–18459, Aug 2016. 2

[5] Martin Aumüller, Erik Bernhardsson, and Alexander Faithfull. Ann-benchmarks: A benchmarking tool for approximate nearest neighbor algorithms. In *International Conference on Similarity Search and Applications*, 2017. 6

[6] Raj Chandra Bose and Dwijendra K Ray-Chaudhuri. On a class of error correcting binary group codes. *Information and control*, 3(1):68–79, 1960. 4

[7] J-Y Bouguet and Pietro Perona. 3d photography on your desk. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, December 1998. 2

[8] Jean-Yves Bouguet and Pietro Perona. 3d photography using shadows in dual-space geometry. *International Journal of Computer Vision*, 35(2):129–149, 1999. 2

[9] Christian Brandli, Thomas Mantel, Marco Hutter, Markus Höpflinger, Raphael Berner, Roland Siegwart, and Tobi Delbruck. Adaptive pulsed laser line extraction for terrain reconstruction using a dynamic vision sensor. *Frontiers in Neuroscience*, 7:275, 2014. 2

[10] Wenzheng Chen, Parsa Mirdehghan, Sanja Fidler, and Kiriakos N. Kutulakos. Auto-tuning structured light by optical stochastic gradient descent. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020. 8

[11] Xiangcheng Chen, Shunping Chen, Jie Luo, Mengchao Ma, Yuwei Wang, Yajun Wang, and Lei Chen. Modified Gray-Level Coding Method for Absolute Phase Retrieval. *Sensors (Basel, Switzerland)*, 17(10):2383, Oct. 2017. 5

[12] Denso Wave Corporation. Quick response code, 1994. 4

[13] Pal Ellingsen, Susanna Spinsante, O Ytrehus, and Angela Barbero. Maximum likelihood decoding of codes on the asymmetric z-channel. In *the 10th IMA International Conference on Cryptography and Coding 2005*, 2005. 8

[14] Eric R Fossum. What to do with sub-diffraction-limit (sdl) pixels?—a proposal for a gigapixel digital film sensor (dfs). In *IEEE Workshop on Charge-Coupled Devices and Advanced Image Sensors*, pages 214–217, 2005. 2

[15] Eric R Fossum, Jiaju Ma, Saleh Masoodian, Leo Anzagira, and Rachel Zizza. The quanta image sensor: Every photon counts. *Sensors*, 16(8):1260, 2016. 2

[16] A. Gabay, P. Duhamel, and O. Rioul. Real bch codes as joint source channel codes for satellite images coding. In *Globecom IEEE Global Telecommunications Conference*, 2000. 4

[17] Guillermo Gallego, Tobi Delbrück, Garrick Orchard, Chiara Bartolozzi, Brian Taba, Andrea Censi, Stefan Leutenegger, Andrew J. Davison, Jörg Conradt, Kostas Daniilidis, and Davide Scaramuzza. Event-based vision: A survey. *arXiv preprint*, 2019. abs/1904.08405. 2

[18] Luis Goddyn and Pavol Gvozdjak. Binary Gray Codes with Long Bit Runs. *The Electronic Journal of Combinatorics*, 10(1):R27, June 2003. 5

[19] Luis Goddyn, George M Lawrence, and Evi Nemeth. Gray codes with optimized run lengths. *Utilitas Mathematica*, 34:179–192, 1988. 5

[20] Sai Siva Gorthi and Pramod Rastogi. Fringe projection techniques: Whither we are? *Optics and lasers in engineering*, 48(ARTICLE):133–140, 2010. 5

[21] Jens Gühring. Dense 3d surface acquisition by structured light using off-the-shelf components. In *Videometrics and Optical Methods for 3D Shape Measurement*, volume 4309, pages 220–231. International Society for Optics and Photonics, 2000. 1, 5

[22] Yue Guo, Yanmei Li, Haitao Song, Wenhao He, and Kui Yuan. Towards a micron-level line structured light. In *2021 IEEE International Conference on Mechatronics and Automation (ICMA)*, pages 750–755, 2021. 1

[23] Mohit Gupta, Amit Agrawal, Ashok Veeraraghavan, and Srinivasa G. Narasimhan. Structured light 3D scanning in the presence of global illumination. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011. 5

[24] Mohit Gupta, Qi Yin, and Shree K. Nayar. Structured light in sunlight. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, December 2013. 7, 8

[25] Hyowon Ha, Tae-Hyun Oh, and In So Kweon. A multi-view structured-light system for highly accurate 3d modeling. In *2015 International Conference on 3D Vision*, pages 118–126. IEEE, 2015. 1

[26] Xueyan Huang, Yueyi Zhang, and Zhiwei Xiong. High-speed structured light based 3d scanning using an event camera. *Opt. Express*, 29(22):35864–35876, Oct 2021. 2

[27] Jae-Sang Hyun, George T.-C. Chiu, and Song Zhang. High-speed and high-accuracy 3d surface measurement using a mechanical projector. *Opt. Express*, 26(2):1474–1487, Jan 2018. 1, 2

[28] Atul Ingle, Trevor Seets, Mauro Buttafava, Shantanu Gupta, Alberto Tosi, Mohit Gupta, and Andreas Velten. Passive inter-photon imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2021. 2

[29] Atul Ingle, Andreas Velten, and Mohit Gupta. High Flux Passive Imaging With Single-Photon Sensors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019. 2

[30] S. Inokuchi, K. Sato, and F. Matsuda. Range imaging system for 3-d object recognition. In *International Conference Pattern Recognition*, 1984. 4

[31] Jeff Johnson, Matthijs Douze, and Hervé Jégou. Billion-scale similarity search with gpus. *IEEE Transactions on Big Data*, 2019. 6, 12

[32] Qian Kemao. Windowed fourier transform for fringe pattern analysis. *Appl. Opt.*, 43(13):2695–2702, May 2004. 2

[33] Qian Kemao. Two-dimensional windowed fourier transform for fringe pattern analysis: Principles, applications and implementations. *Optics and Lasers in Engineering*, 45(2):304–317, 2007. Phase Measurement Techniques and their applications. 2

[34] Leonid Keselman, John Iselin Woodfill, Anders Grunnet-Jepsen, and Achintya Bhowmik. Intel realsense stereoscopic depth cameras. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, July 2017. 2

[35] Sanjeev J Koppal, Shuntaro Yamazaki, and Srinivasa G Narasimhan. Exploiting dlp illumination dithering for reconstruction and photography of high-speed scenes. *International journal of computer vision*, 96(1):125–144, 2012. 2

[36] Shuangyan Lei and Song Zhang. Flexible 3-d shape measurement using projector defocusing. *Opt. Lett.*, 34(20):3080–3082, Oct 2009. 2

[37] T Leroux, S-H Ieng, and Ryad Benosman. Event-based structured light for depth reconstruction using frequency tagged light patterns. *arXiv preprint arXiv:1811.10771*, 2018. 2

[38] Marc Levoy, Kari Pulli, Brian Curless, Szymon Rusinkiewicz, David Koller, Lucas Pereira, Matt Ginzton, Sean Anderson, James Davis, Jeremy Ginsberg, et al. The digital michelangelo project: 3d scanning of large statues. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, 2000. 2

[39] P. Lichtsteiner, C. Posch, and T. Delbruck. A 128×128 120 dB 15 μs Latency Asynchronous Temporal Contrast Vision Sensor. *IEEE J. Solid-State Circuits*, 43(2), 2008. 2

[40] Jiaju Ma, Saleh Masoodian, Dakota A. Starkey, and Eric R. Fossum. Photon-number-resolving megapixel image sensor at room temperature without avalanche gain. *Optica*, 4(12):1474–1481, Dec 2017. 2

[41] Sizhuo Ma, Shantanu Gupta, Arin C Ulku, Claudio Bruschini, Edoardo Charbon, and Mohit Gupta. Quanta burst photography. *ACM Transactions on Graphics (TOG)*, 39(4):79–1, 2020. 2

[42] Ashish Rao Mangalore, Chandra Sekhar Seelamantula, and Chetan Singh Thakur. Neuromorphic fringe projection profilometry. *IEEE Signal Processing Letters*, 27:1510–1514, 2020. 2

[43] Julien NP Martel, Jonathan Müller, Jörg Conradt, and Yulia Sandamirskaya. An active approach to solving the stereo matching problem using event-based sensors. In *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 1–5. IEEE, 2018. 2

[44] J. Massey. Shift-register synthesis and bch decoding. *IEEE Transactions on Information Theory*, 15(1):122–127, 1969. 6, 3, 12

[45] Nathan Matsuda, Oliver Cossairt, and Mohit Gupta. MC3D: Motion Contrast 3D Scanning. In *2015 IEEE International Conference on Computational Photography (ICCP)*, 2015. 2, 7

[46] Parsa Mirdehghan, Wenzheng Chen, and Kiriakos N. Kutulakos. Optimal structured light a la carte. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 8

[47] Kazuhiro Morimoto, Andrei Ardelean, Ming-Lo Wu, Arin Can Ulku, Ivan Michel Antolovic, Claudio Bruschini, and Edoardo Charbon. Megapixel time-gated SPAD image sensor for 2D and 3D imaging applications. *Optica*, 7(4):346–354, Apr. 2020. 2

[48] K. Morimoto and J. Iwata. 3.2 megapixel 3d-stacked charge focusing spad for low-light imaging and depth sensing. In *IEEE International Electron Devices Meeting (IEDM)*, 2021. https://btbmarketing.egnyte.com/dl/BhiUjgx9HQ. 2

[49] Matthew O'Toole, Supreeth Achar, Srinivasa G Narasimhan, and Kiriakos N Kutulakos. Homogeneous codes for energy-efficient illumination and imaging. *ACM Transactions on Graphics (ToG)*, 34(4):1–13, 2015. 7

[50] Matthew O'Toole, Supreeth Achar, Srinivasa G. Narasimhan, and Kiriakos N. Kutulakos. Homogeneous codes for energy-efficient illumination and imaging. *ACM SIGGRAPH*, (0), 2015. 8

[51] Matthew O'Toole, John Mather, and Kiriakos N Kutulakos. 3d shape and indirect appearance by structured light transport. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014. 8

[52] Matthew O'Toole, Ramesh Raskar, and Kiriakos N Kutulakos. Primal-dual coding to probe light transport. *ACM Trans. Graph.*, 31(4):39–1, 2012. 8

[53] Jeffrey L Posdamer and MD Altschuler. Surface measurement by space-encoded projected beam systems. *Computer graphics and image processing*, 18(1):1–17, 1982. 1, 2

[54] Ron M Roth. Introduction to coding theory. *IET Communications*, 47:18–19, 2006. 4, 3

[55] Sean Ryan Fanello, Christoph Rhemann, Vladimir Tankovich, Adarsh Kowdle, Sergio Orts Escolano, David Kim, and Shahram Izadi. Hyperdepth: Learning depth from structured light without matching. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016. 2

[56] Larissa C Shimomura, Rafael Seidi Oyamada, Marcos R Vieira, and Daniel S Kaster. A survey on graph-based methods for similarity searches in metric spaces. *Information Systems*, 95:101507, 2021. 6

[57] Yoshiaki Shirai and Motoi Suwa. Recognition of polyhedrons with a range finder. In *IJCAI*, pages 80–87. Citeseer, 1971. 2

[58] V. Srinivasan, H. C. Liu, and Maurice Halioua. Automated phase-measuring profilometry: a phase mapping approach. *Appl. Opt.*, 24(2):185–188, 1985. 2

[59] Xian-Yu Su, Wen-Sen Zhou, G Von Bally, and D Vukicevic. Automated phase-measuring profilometry using defocused projection of a ronchi grating. *Optics Communications*, 94(6):561–573, 1992. 2

[60] Arin Can Ulku, Claudio Bruschini, Ivan Michel Antolovic, Yung Kuo, Rinat Ankri, Shimon Weiss, Xavier Michalet, and Edoardo Charbon. A 512 × 512 SPAD Image Sensor With

Integrated Gating for Widefield FLIM. *IEEE Journal of Selected Topics in Quantum Electronics*, 25(1):1–12, Jan. 2019. 1, 6, 11

[61] Feng Yang, Yue M. Lu, Luciano Sbaiz, and Martin Vetterli. Bits from photons: Oversampled image acquisition using binary poisson statistics. *IEEE Transactions on Image Processing*, 21(4):1421–1436, 2012. 3

[62] Song Zhang, Daniel Van Der Weide, and James Oliver. Superfast phase-shifting method for 3-d shape measurement. *Optics express*, 18(9):9684–9689, 2010. 1, 2

[63] Zhengyou Zhang. Microsoft kinect sensor and its effect. *IEEE MultiMedia*, 19:4–12, April 2012. 2, 7

[64] Dongliang Zheng, Feipeng Da, Qian Kemao, and Hock Soon Seah. Phase error analysis and compensation for phase shifting profilometry with projector defocusing. *Applied Optics*, 55(21):5721, July 2016. 2, 5

[65] Dongliang Zheng, Feipeng Da, Qian Kemao, and Hock Soon Seah. Phase-shifting profilometry combined with Gray-code patterns projection: Unwrapping error removal by an adaptive median filter. *Optics Express*, 25(5):4700, Mar. 2017. 5

Supplementary Material for Single-Photon Structured Light

1. Deriving Expected Error

In this supplementary note, we derive the decoding error probability of Gray codes and repetition codes subject to the bit-flip noise model (Sec. 3) of Single-Photon SL. We provide an analytic expression for error probability in BCH codes for a given decoding function and for the particular case of a bounded-error decoder. Additionally, we derive an upper-bound on the average decoding error (in terms of absolute deviation) for binary shifted patterns. Finally, we outline the Monte-Carlo simulation procedure to compare various strategies—such as BCH and repetition strategies in Fig. 6, and Hybrid and repetition strategies in Fig. 8.

1.1. Notation

Let $\mathcal{M} = \{\mathbf{m} \mid \mathbf{m} \in \{0,1\}^L\}$ denote the set of L-bit binary messages, e.g., Gray codes, that represent columns of a projector. For simplicity, we shall assume that the projector has 2^L columns, but the following analysis extends mutatis mutandis to the more general case. Further, let $\mathbf{1_m}$ and $\mathbf{0_m}$ denote the number of 1's and 0's in L-bit message $\mathbf{m}$ respectively.

1.2. Gray Codes

When transmitting a Gray code $\mathbf{m}$ an error occurs if any of the bits are flipped due to photon-noise, since no error-correcting strategy is employed. Assuming $\mathbf{y}$ denotes the received (possibly corrupted) codeword when $\mathbf{m}$ is projected, the expected error probability over the set of messages $\mathcal{M}$ is given as

$$
\begin{aligned}
\Pr\left\{\text{error}\right\} &= \sum_{\mathbf{m}\in\mathcal{M}} \Pr\left\{\mathbf{m}\right\}\Pr\left\{\text{error}\,|\,\mathbf{m}\right\} \\
&= \frac{1}{2^L}\sum_{\mathbf{m}\in\mathcal{M}} \Pr\left\{\mathbf{y}\neq\mathbf{m}\,|\,\mathbf{m}\right\} \qquad\qquad \text{(Each message is equally likely to be transmitted)} \\
&= \frac{1}{2^L}\sum_{\mathbf{m}\in\mathcal{M}} 1 - \Pr\left\{\mathbf{y}=\mathbf{m}\,|\,\mathbf{m}\right\} \\
&= \frac{1}{2^L}\sum_{\mathbf{m}\in\mathcal{M}} 1 - \prod_{i=1}^{L}\Pr\left\{\mathbf{y}_i=\mathbf{m}_i\right\} \\
&= 1 - \frac{1}{2^L}\sum_{\mathbf{m}\in\mathcal{M}} \Pr\left\{\mathbf{y}_i=1\,|\,\mathbf{m}_i=1\right\}^{\mathbf{1_m}} \Pr\left\{\mathbf{y}_i=0\,|\,\mathbf{m}_i=0\right\}^{\mathbf{0_m}} \\
&= 1 - \frac{1}{2^L}\sum_{i=0}^{L} \binom{L}{i}\Pr\left\{\mathbf{y}_i=1\,|\,\mathbf{m}_i=1\right\}^{i} \Pr\left\{\mathbf{y}_i=0\,|\,\mathbf{m}_i=0\right\}^{L-i} \\
&= 1 - \frac{1}{2^L}\left(\Pr\left\{\mathbf{y}_i=1\,|\,\mathbf{m}_i=1\right\} + \Pr\left\{\mathbf{y}_i=0\,|\,\mathbf{m}_i=0\right\}\right)^{L} \\
&= 1 - \left(1 - \frac{\left(\mathrm{P_{flip,\,bright}} + \mathrm{P_{flip,\,dark}}\right)}{2}\right)^{L} \tag{S1}
\end{aligned}
$$

where, we make use of Eq. (3) and Eq. (4) in the last step. Note that the decoding error is zero when $\mathrm{P_{flip,\,bright}}, \mathrm{P_{flip,\,dark}} \to 0$.

1.3. Repeated Gray Codes

Assume we repeat $\mathbf{m}$, the L-bit message, r times and decode it with a majority vote. Let $\mathbf{x}$ denote the received codeword and $\mathbf{y}$ the decoded message by using a majority vote. Then each bit y_i is decoded incorrectly if more than or equal $\lceil \frac{r-1}{2} \rceil$ bit-flips occur, with probability

$$
\begin{aligned}
\Pr\{\mathbf{y}_i \neq \mathbf{m}_i \mid \mathbf{m}_i = 1\} &= \prod_{j=\lceil \frac{r-1}{2} \rceil}^{r} \binom{r}{j} \mathrm{P}_{\text{flip, bright}}^{j}(1 - \mathrm{P}_{\text{flip, bright}})^{r-j} \\
&= g_r(\mathrm{P}_{\text{flip, bright}})
\end{aligned}
\tag{S2}
$$

where $g_r(p) = \prod_{j=\lfloor \frac{r-1}{2} \rfloor}^{r} p^r (1-p)^{r-j}$. Similarly, we can show that

$$
\Pr\{\mathbf{y}_i \neq \mathbf{m}_i \mid \mathbf{m}_i = 0\} = g_r(\mathrm{P}_{\text{flip, dark}})
$$

With this, we can evaluate decoding error probability for the repetition strategy by extending Eq. (S1) as

$$
\Pr\{\text{error}\} = 1 - \left(1 - \left(\frac{g_r\left(\mathrm{P}_{\text{flip, bright}}\right) + g_r\left(\mathrm{P}_{\text{flip, dark}}\right)}{2}\right)\right)^{L}
\tag{S3}
$$

Note that, $g_r(p) \geq p \; \forall \, p \in [0, 1]$, so repetition always improves decoding reliability when $\mathrm{P}_{\text{flip, dark}}, \mathrm{P}_{\text{flip, bright}} < 0.5$.

1.4. BCH Codes

Construction. A $\text{BCH}(n, k, d) : \{0, 1\}^k \to \{0, 1\}^n$ encoder takes input messages of length k and produces output codewords of length n that are at least d-bits apart. In this paper, we consider primitive, narrow-sense BCH codes over the finite field $\text{GF}(2)$. We present a brief summary of their construction here and refer readers to Roth [11] for a more detailed explanation.

1. Pick α a primitive element of $\text{GF}(2^n)$.

2. Compute the generator polynomial $g(x) = \text{lcm}\left(q_1(x), ..., q_{d-1}(x)\right)$, where $q_i(x)$ are minimal polynomials of α^i with coefficients in $\text{GF}(2)$.

3. Each codeword is obtained by multiplying a polynomial representing the message $p(x)$ and the generator polynomial $g(x)$.

4. For systematic encoding, we set $p(x) = m(x)x^{n-k} - r(x)$, where $m(x)$ is the polynomial with the symbols of message $\mathbf{m}$ as coeffecients and $r(x) = m(x)x^{n-k} \bmod g(x)$.

From the construction, it can be seen that BCH codes are a subset of Reed-Solomon codes over the finite field $\text{GF}(2^n)$ [6]. Specifically, $\mathcal{C}_{\text{BCH}}(n, k, d) = \mathcal{C}_{\text{RS}(n,k,d)} \cap \{0, 1\}^n$, where $\text{RS}(n, k, d)$ represents the Reed-Solomon encoder with parameters (n, k, d) as before.

Decoding Error. Let $\mathbf{c}$ denote a codeword corresponding to BCH encoded message $\mathbf{m}$. As before, assume we receive (possibly corrupted) bits $\mathbf{x}$ which we decode to obtain $\mathbf{y}$. Depending on whether decoder gives up beyond the worst-case error limit (e.g. Berlekamp Massey [9]) or not (e.g. minimum distance decoding), we can derive two expressions. For decoders that work up to the worst-case error limit of $\lfloor \frac{d-1}{2} \rfloor$ bit-flips,

$$
\Pr\{\mathbf{y} \neq \mathbf{m}\} = \Pr\left\{ d_H(\mathbf{x}, \mathbf{c}) > \frac{d-1}{2} \right\} \qquad \text{(where } d_H \text{ is the Hamming distance)}
$$

$$
= \Pr\left\{ \#\text{bit-flips bright} + \#\text{bit-flips dark} > \frac{d-1}{2} \right\}
$$

$$
= \sum_{j=0}^{\frac{d-1}{2}} \Pr\{\#\text{bit-flips bright} = j\} \Pr\left\{ \#\text{bit-flips dark} > \frac{d-1}{2} - j \right\}
$$

$$
= \sum_{j=0}^{\frac{d-1}{2}} \sum_{k=\frac{d-1}{2}-j}^{\mathbf{1}_c} \binom{\mathbf{0}_c}{j}\binom{\mathbf{1}_c}{k} \mathbf{P}_{\text{flip, dark}}^j (1 - \mathbf{P}_{\text{flip, dark}})^{\mathbf{0}_c - j}
$$

$$
\mathbf{P}_{\text{flip, bright}}^k (1 - \mathbf{P}_{\text{flip, bright}})^{\mathbf{1}_c - k} \tag{S4}
$$

Whereas, if we use a minimum distance decoder,

$$
\Pr\{\mathbf{y} \neq \mathbf{m}\} = \Pr\left\{ \min_{\mathbf{z} \in \mathcal{C}} d_H(x, z) \neq c \right\} \tag{S5}
$$

$$
\text{, where } \mathcal{C} = \{\, \mathbf{c} \mid \mathbf{c} = \mathcal{E}_{\text{BCH}(n,k,d)}(\mathbf{m}) \, \forall \, \mathbf{m} \in \mathcal{M} \,\}
$$

Evaluating this probability expression in closed-form is not straightforward, but can be done by enumerating each codeword in an exhaustive manner. For this reason, we opt to use Monte-Carlo simulations when comparing BCH and repetition strategies.

1.5. Binary Shifting

Decoding binary shifted patterns is achieved by using a matched filter, where the template corresponds to a pattern with a burst of $2^{L_{\text{shift}}}$ ones followed by $2^{L_{\text{shift}}}$ zeros. In this section, we present the error analysis for binary shifting with $L_{\text{shift}} = 3$, corresponding to a temporal pattern length of $2^{L_{\text{shift}}+1} = 16$ frames. Without loss of generality, we consider the temporal sequence representing pixel location 0—comprising of 8 ones followed by 8 zeros. The expected error of other circularly shifted sequences are identical to this. Let s_t denote the value of the received signal (possibly corrupted) at the t-th time instant ($0 \leq t \leq 15$). Then, the probability of the absolute decoding error being exactly 1 pixel can be derived as

$$
\begin{aligned}
\Pr\{|\text{error}| = 1\} = \Pr &\left\{ \{\sum_{t=1}^{8} s_t \geq \sum_{t=j}^{7+j} s_t \ \forall \ j\} \bigcup \{\sum_{t=-1}^{6} s_t \geq \sum_{i=j}^{7+j} s_i \ \forall \ j\} \right\} \\
&\leq 2\Pr\left\{ \{\sum_{t=1}^{8} s_t \geq \sum_{t=0}^{7} s_t\} \right\} \qquad \text{(Union bound, equally likely events)} \\
&= 2\Pr\{s_8 \geq s_0\} \\
&= 2\Pr\{X \geq Y\} \qquad (X \sim \text{Ber}(\text{P}_{\text{flip, dark}}), \ Y \sim \text{Ber}(1 - \text{P}_{\text{flip, bright}})) \\
& \hspace{10cm} \text{(S6)} \\
&= 2\left(\text{P}_{\text{flip, dark}} + (1 - \text{P}_{\text{flip, dark}})\text{P}_{\text{flip, bright}}\right) \qquad \text{(S7)}
\end{aligned}
$$

where, $\text{Ber}(p)$ denotes the Bernoulli probability distribution. In the above expression, we consider time instances modulo 16, i.e., $t = -1$ is interpreted as $t = 15$. Next, the probability of the absolute decoding error being exactly 2 pixels can computed as

$$
\begin{aligned}
\Pr\{|\text{error}| = 2\} = \Pr &\left\{ \{\sum_{t=2}^{9} s_t \geq \sum_{t=j}^{7+j} s_t \ \forall \ j\} \bigcup \{\sum_{t=-2}^{5} s_t \geq \sum_{t=j}^{7+j} s_t \ \forall \ j\} \right\} \\
&\leq 2\Pr\left\{ \{\sum_{t=2}^{9} s_t \geq \sum_{t=0}^{7} s_t\} \right\} \qquad \text{(Union bound)} \\
&= 2\Pr\{s_8 + s_9 \geq s_0 + s_1\} \qquad \text{(S8)} \\
&= 2\Pr\{X \geq Y\} \qquad (X \sim \text{Bin}(2, \text{P}_{\text{flip, dark}}), \ Y \sim \text{Bin}(2, 1 - \text{P}_{\text{flip, bright}}))
\end{aligned}
$$

where, $\text{Bin}(n, p)$ denotes the binomial probability distribution. Similarly, we can upper bound the probability of the absolute decoding error being $1 \leq r \leq 15$ pixels as $2\Pr\{X_r \geq Y_r\}$, where $X_r \sim \text{Bin}(r, \text{P}_{\text{flip, dark}})$, $Y_r \sim \text{Bin}(r, 1 - \text{P}_{\text{flip, bright}})$. With this, the expected absolute decoding error is given by:

$$
\begin{aligned}
\mathbb{E}[|\text{error}|] &= \sum_{r=0}^{15} r\Pr\{|\text{error}| = r\} \\
&\leq \sum_{r=1}^{15} 2r\Pr\{X_r \geq Y_r\} \qquad \text{(S9)}
\end{aligned}
$$

Using this expression, the upper bound on the expected decoding error for the dark room condition (using flux values from Fig. 3(a)) is 1.2 pixels.

1.6. Monte-Carlo Simulation Procedure

Algorithm 1 Monte-Carlo Simulation Procedure

Require: Bit-flips probabilities $P_{\text{flip, bright}}, P_{\text{flip, dark}}$
  Message set $\mathcal{M}$ with corresponding codeword set $\mathcal{C}$
  error metric $\mathcal{L}$
  Monte-Carlo iterations n_{iter}
  procedure MONTE-CARLO-SIMULATION($P_{\text{flip, bright}}, P_{\text{flip, dark}}, \mathcal{M}, \mathcal{C}, \mathcal{L}\ n_{\text{iter}}$)
      error $\leftarrow 0$
      for $\mathbf{m} \in \mathcal{M}$ **do**
          Pick corresponding $\mathbf{c} \in \mathcal{M}$
          error$_{\mathbf{m}} \leftarrow 0$
          for $1 \leq i \leq n_{\text{iter}}$ **do**
              Randomly flip $\mathbf{1_c}$ and $\mathbf{0_c}$ with probability $P_{\text{flip, bright}}$ and $P_{\text{flip, dark}}$ respectively to obtain $\mathbf{x}$
              Decode $\mathbf{x}$ to obtain $\mathbf{y}$
              Compute $\mathcal{L}(\mathbf{m}, \mathbf{y})$
              error$_{\mathbf{m}} = $ error$_{\mathbf{m}} + \mathcal{L}(\mathbf{m}, \mathbf{y})$
          end for
          error$_{\mathbf{m}} = $ error$_{\mathbf{m}}/n_{\text{iter}}$
          error $= $ error $+$ error$_{\mathbf{m}}$
      end for
      error $= $ error$/|\mathcal{M}|$
  return error
  end procedure

We describe the simulation procedure in Algorithm 1. The error metrics we consider are exact error ($\sum_{i=1}^{N} \frac{1}{N} \mathbb{I}(x_i, y_i)$, where $\mathbb{I}$ denotes the indicator function) and root mean square error (RMSE, $\sqrt{\frac{1}{N} \sum_{i=1}^{N} (x_i - y_i)^2}$). Figure 6 and Figure 8 both use $n_{\text{iter}} = 100$ (Monte-Carlo iterations).

1.7. Empirical Comparison of BCH and Repetition Strategies at Various Φ_a

(a) Redundancy $n = 63$

(a) Redundancy $n = 255$

Figure 1. **Monte-Carlo evaluation of BCH and repetition strategies,** plotted at various ambient flux Φ_a. $\Phi_a = 0$ photons/sec (*left column*) represents a dark room, while $\Phi_a = 100$ photons/sec and $\Phi_a = 5000$ photons/sec *(middle and right columns)* represent indoor illumination and bright intensity matching the projector's flux respectively. Across ambient illumination levels and redundancy factors, BCH codes outperform repetition codes. Particularly at $n = 255$, BCH codes result in almost zero error everywhere.

2. Comparison to High-Speed Cameras

Read-noise of existing high-speed cameras is quite large and worsens with increasing read-out rates [2]. For instance, the Phantom v2640 [1] has a read noise of 18.8e- when capturing 640×480 frames at 28 kHz. This leads to an order of magnitude lower SNR than SPADs (Tab. 1), and hence extremely noisy, practically unusable, reconstructions (Fig. 2). This is because high-speed cameras must capture sufficient photons in each frame to beat the high read noise floor, which is not possible especially in high-speed motion scenarios considered in this paper.

Device	Dark room	Indoor lamp	Spot lamp
Phantom-v2640	0.07	0.08	0.13
SwissSPAD	0.81	0.84	1.10

Table 1. **SNR of a high-speed camera (Phantom-v2640) vs a SPAD camera (this work),** when observing a single bright pattern at 20kHz across various ambient conditions. No new experiment was conducted—incident flux data is obtained from Fig 3.

SPAD camera	High-speed camera	Groundtruth	SPAD camera	High-speed camera
Sample Frame		RMSE:	**1.0 mm**	322.4 mm
		RMSE (inlier %):	**0.9 mm (99.8 %)**	2.7 mm (6.4 %)

Figure 2. **3D reconstruction from the Phantom-v2640 vs the SPAD camera.** *(left)* We simulate high-speed capture using photon arrival rates. *(right)* Read-noise dominated frames of the high-speed camera lead to poor reconstructions. For a fair comparison, we use Hybrid-255 and Hybrid-127 with complementary frames as the coding scheme for the SPAD camera and the high-speed camera respectively—requiring similar acquisition times.

3. Code Look-Up-Tables

In this note, we describe the coding strategies developed such as BCH and Hybrid encoding, by means of their code look-up-tables (LUTs)—which depict the projected patterns row-wise across time.

3.1. BCH(31, 11, 11) Encoding of a 10-bit Conventional Gray Message

Figure 3. **Code Look-Up-Table describing the BCH(31, 11, 11) encoding of a 10-bit Conventional Gray message.** Following systematic encoding, the code LUT upto the first 10 frames is identical to a Gray Code LUT. Most of the parity frames feature high spatial-frequency, which is easily distorted by short-range effects in a SL system (Sec. 4.2). Zoom-in to see high-frequency details.

3.2. BCH(31, 11, 11) Encoding of a 10-bit Long-run Gray Message

Figure 4. **Code Look-Up-Table describing the BCH(31, 11, 11) encoding of a 10-bit Long-run Gray message.** Owing to the maximal minimum stripe-widths of Long-run Gray codes, the first 10 frames of the code LUT are robust to short-range distortions such as projection defocus and/or resolution mismatch. However, the parity frames continue to comprise of high-spatial frequency patterns and are thus, easily distorted. Zoom-in to see high-frequency details.

9

3.3. Hybrid $(n = 31)$ Encoding of a 10-bit Gray Message

Figure 5. **Code Look-Up-Table describing the Hybrid$(n = 31)$ encoding of a 10-bit Conventional Gray message.** We encode $L_{\text{BCH}} = 7$ MSBs using the BCH(31, 11, 11) encoder and $L_{\text{shift}} = 3$ LSBs using binary shifting. Unlike the code LUTs of BCH encoded patterns (Fig. 3 and Fig. 4) which comprise of many high spatial-frequency patterns (typically featuring a minimum stripe-width of 1 or 2 pixels), the minimum stripe-width of these Hybrid $(n = 31)$ codes are at least 8 pixels by construction. This results in Hybrid codes offering robustness to both bit-flips arising from photon noise and other short-range effects arising from a combination of projector defocus and camera-projector resolution mismatch.

10

4. Single-Photon SL Hardware Prototype

(a) Single-Photon SL Prototype   (b) SwissSPAD2 array   (c) Vialux-7000 high-speed module   (d) SugarCUBE LED source

Figure 6. **Single-Photon SL hardware prototype.** The setup consists of a SwissSPAD2 array [12] and a Vialux-7000 development kit which is based on the Texas Instruments DLP6500 DMD. For an illumination source we use a SugarCUBE Ultra, a broadband white LED, featuring a light output of 4000 lumens.

Supp. Fig. 6 shows the various components of our Single-Photon SL prototype. The resolution of the SPAD array is 512×256 pixels, while the DLP projector has a resolution of 1024×768 pixels. We used a baseline of 14 cm between the camera and projector—with the intention of maximizing the field of view of the camera occupied by the objects of interest.

4.1. Calibration Procedure

Our calibration procedure is similar to Zhang et al. [13], where the projector is treated as an inverse camera. Specifically, we capture correspondence maps of a chessboard, repeated across 30 different positions. We then perform a calibration step mimicking a stereo setup based on the detected corners of the chessboard and their row and column correspondences.

To obtain correspondences, we use projected patterns based on Gray codes and binary shifting—similar to our proposed Hybrid strategy, but with no added redundancy—to encode both column and row indices. This provides us precise row and column correspondence while simultaneously dealing with short-range effects arising from projector defocus / resolution-mismatch. To suppress the effects of photon-noise, we capture 5120 SPAD frames for each projected pattern—which are subsequently averaged and compared against acquired complementary frames to produce binary outputs.

5. Benchmarking Minimum Distance Decoding Methods

Supp. Fig. 7 shows that Minimum Distance Decoding (MDD) can be significantly faster than algebraic decoding in Single-Photon SL. Supp. Fig. 8 further shows that the most performant MDD method—FAISS using a GPU runtime—also scales reasonably to larger sized arrays including 1MPixel and 4MPixel sensors.

Figure 7. **Benchmark of various Minimum Distance Decoding (MDD) methods.** We query a 512×256 array of codewords each encoding a 10-bit message. We compare an algebraic method (Berkelamp-Massey [9]) to several MDD methods including those using exhaustive search (Numpy [7], CuPy [10]), graph-based search (BallTree [3], NNDescent [5]) symbolic-tensors (KeOps [4]) and batched queries (FAISS [8]). Benchmarks reported across various BCH encoders using a Intel-i7 8700K CPU and a NVIDIA 1080Ti GPU.

(a) FAISS CPU

(a) FAISS GPU

Figure 8. **MDD methods scale to larger resolution arrays.** Our implementation based on FAISS requires about 4s and 15s on a CPU (Intel-i7); 12 ms and 30 ms for one and four megapixels respectively on a GPU device (GTX 1080Ti).

6. More Results

6.1. Low-albedo Objects

Supp. Fig. 9 includes additional results on low-albedo objects; demonstrating the capabilities of Single-Photon SL in 3D imaging commonplace dark objects—such as the teflon pan, coffee mug and car headrest.

Reference image　　Binary Frame　　3D Reconstruction

(a) Teflon pan

(b) Car headrest

(c) Coffee mug

Figure 9. **More 3D reconstructions of low-albedo scenes,** using Hybrid ($n = 255$) at 40 FPS.

6.2. Dynamic Motion Sequences

We include further results on dynamic scenes in Supp. Fig. 10. The curtain took an estimated 70 milliseconds to fall, while the Jack-in-the-box toy took around 40 milliseconds to spring up.

Going faster with pipelining. While we have decoded disjoint bursts of acquired frames to output depth-maps in Fig. 11 and Fig. 13, we can also use overlapping sequences. Assuming N patterns are projected, we can recover depth across time by decoding sequences $\{I_i\}_{i=1}^{N}, \{I_i\}_{i=N+1}^{2N}, \{I_i\}_{i=2N+1}^{3N}, ...$, or by considering sequences $\{I_i\}_{i=1}^{N}, \{I_i\}_{i=s+1}^{N+s}, \{I_i\}_{i=s+2}^{N+s+1}, ...$ i.e., using a sliding window of N patterns with stride $0 < s < N$. This strided or pipelined approach contains more temporal information than disjoint bursts. We utilize pipelining based frame-interpolation for the jack-in-the-box sequence to produce a high-speed 3D reconstruction video at 2000 FPS.

High-speed depth videos for the deforming cloth sequence and the Jack-in-the-box toy can be found on the project webpage. By using pipelining-based decoding, we output videos at 1000 FPS for the waving cloth and 2000 FPS for the Jack-in-the-box with a stride of 9 and 20 patterns respectively. We captured the Jack-in-the-box by placing it horizontally to maximise its occupied field-of-view. Both videos are also included in the teaser video which provides an overview of our paper.

(a) Hand Gesture Sequence

(b) Jack-in-the-box Sequence

(c) Falling Curtain Sequence

Figure 10. **Additional dynamic scenes imaged by Single-Photon SL.** The hand gesture sequence was captured using Hybrid ($n = 255$) at 75 FPS, while the Jack-in-the-box and falling curtain sequences were both acquired using Hybrid ($n = 31$) at 450 FPS.

References

[1] Phantom-v2640. `phantomhighspeed.com/products/cameras/ultrahighspeed/v2640/`. 7

[2] Assim Boukhayma, Arnaud Peizerat, and Christian Enz. A sub-0.5 electron read noise vga image sensor in a standard cmos process. *IEEE Journal of Solid-State Circuits*, 2016. 7

[3] Leonid Boytsov and Bilegsaikhan Naidan. Engineering efficient and effective non-metric space library. In *International Conference on Similarity Search and Applications*, 2013. 12

[4] Benjamin Charlier, Jean Feydy, Joan Alexis Glaunès, François-David Collin, and Ghislain Durif. Kernel operations on the gpu, with autodiff, without memory overflows. *Journal of Machine Learning Research*, 22(74):1–6, 2021. 12

[5] Wei Dong, Charikar Moses, and Kai Li. Efficient k-nearest neighbor graph construction for generic similarity measures. In *Proceedings of the International Conference on World wide web*, 2011. 12

[6] Venkatesan Guruswami, Atri Rudra, and Madhu Sudan. Essential coding theory. *Draft available at http://www. cse. buffalo. edu/ atri/courses/coding-theory/book*, 2012. 3

[7] Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. Array programming with NumPy. *Nature*, 585(7825):357–362, Sept. 2020. 12

[8] Jeff Johnson, Matthijs Douze, and Hervé Jégou. Billion-scale similarity search with gpus. *IEEE Transactions on Big Data*, 2019. 6, 12

[9] J. Massey. Shift-register synthesis and bch decoding. *IEEE Transactions on Information Theory*, 15(1):122–127, 1969. 6, 3, 12

[10] Ryosuke Okuta, Yuya Unno, Daisuke Nishino, Shohei Hido, and Crissman. Cupy : A numpy-compatible library for nvidia gpu calculations. 2017. 12

[11] Ron M Roth. Introduction to coding theory. *IET Communications*, 47:18–19, 2006. 4, 3

[12] Arin Can Ulku, Claudio Bruschini, Ivan Michel Antolovic, Yung Kuo, Rinat Ankri, Shimon Weiss, Xavier Michalet, and Edoardo Charbon. A 512×512 SPAD Image Sensor With Integrated Gating for Widefield FLIM. *IEEE Journal of Selected Topics in Quantum Electronics*, 25(1):1–12, Jan. 2019. 1, 6, 11

[13] Song Zhang and Peisen S Huang. Novel method for structured light system calibration. *Optical Engineering*, 45(8):083601, 2006. 11