Compressive Single-Photon 3D Cameras

Felipe Gutierrez-Barragan, Atul Ingle, Trevor Seets, Mohit Gupta, Andreas Velten
University of Wisconsin-Madison
{fgutierrez3, ingle, seets}@wisc.edu, mohitg@cs.wisc.edu, velten@wisc.edu

Abstract

Single-photon avalanche diodes (SPADs) are an emerging pixel technology for time-of-flight (ToF) 3D cameras that can capture the time-of-arrival of individual photons at picosecond resolution. To estimate depths, current SPAD-based 3D cameras measure the round-trip time of a laser pulse by building a per-pixel histogram of photon timestamps. As the spatial and timestamp resolution of SPAD-based cameras increase, their output data rates far exceed the capacity of existing data transfer technologies. One major reason for SPAD's bandwidth-intensive operation is the tight coupling that exists between depth resolution and histogram resolution. To weaken this coupling, we propose compressive single-photon histograms (CSPH). CSPHs are a per-pixel compressive representation of the high-resolution histogram, that is built on-the-fly, as each photon is detected. They are based on a family of linear coding schemes that can be expressed as a simple matrix operation. We design different CSPH coding schemes for 3D imaging and evaluate them under different signal and background levels, laser waveforms, and illumination setups. Our results show that a well-designed CSPH can consistently reduce data rates by 1-2 orders of magnitude without compromising depth precision.

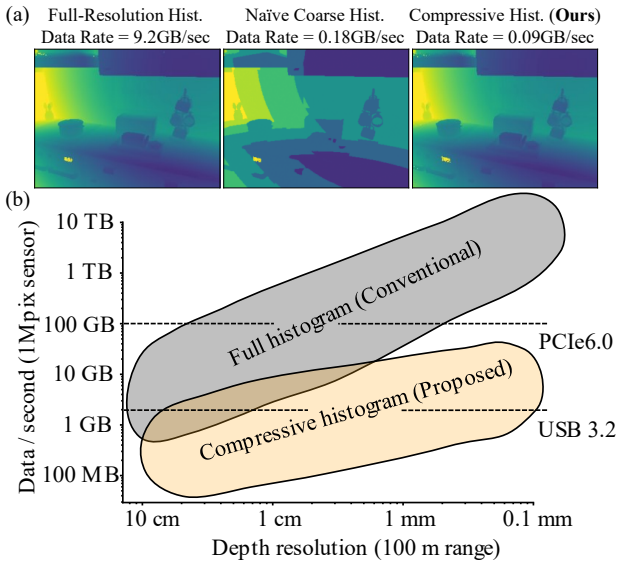


Figure 1. **Compressive Single-Photon 3D Imaging.** (a) Example depth maps with conventional (full histogram) capture, coarse resolution capture and our method with compressive capture. In this simulation, our method generates $100\times$ lower data, yet generates depth maps that are visually indistinguishable from the conventional method. (b) With conventional acquisition schemes, data bandwidth requirements scale linearly with the desired depth resolution. Our proposed compressive acquisition does not scale as strongly with depth resolution, keeping the output data rates manageable with existing data transfer standards like USB and PCIe.

1. Introduction

Single-photon cameras (SPC) are an emerging sensor technology with ultra-high sensitivity down to individual photons [8, 10]. In addition to their extreme sensitivity, SPCs based on single-photon avalanche diodes (SPADs) can also record photon-arrival timestamps with extremely high (sub-nanosecond) time resolution [45]. Moreover, SPAD-based SPCs are compatible with the complementary metal-oxide semiconductor (CMOS) photolithography process which can enable fabrication of kilo-to-mega-pixel resolution SPAD arrays [13, 41] at low costs. Due to these capabilities, SPAD-based SPCs are gaining popularity in various imaging applications including 3D imaging [15, 51, 52, 60], passive low-light imaging [3, 39, 55], HDR imaging [30, 31], non-line-of-sight (NLOS) imaging [5, 37, 65], fluorescence lifetime imaging (FLIM) microscopy [68], and diffuse optical tomography [38, 67].

Unlike a conventional camera pixel that outputs a single intensity value integrated over micro-to-millisecond timescales, a SPAD pixel generates an electrical pulse for each photon detection event. A time-to-digital conversion circuit converts each pulse into a timestamp recording the time-of-arrival of each photon. Under normal illumination conditions, a SPAD pixel can generate millions of photon timestamps per second. The photon timestamps are often captured with respect to a periodic synchronization signal generated by a pulsed laser source. To make this large volume of timestamp data more manageable, SPAD-based SPCs build a *timing histogram* in-sensor instead of transferring the raw photon timestamps to the processing chip. The histogram records the number of photons as a function of the time delay with respect to the synchronization pulse.

Consider a megapixel SPAD-based 3D camera. For short range indoor applications (up to tens of meters), a millimeter depth resolution would be desirable. For longer range

outdoor applications (hundreds of meters), centimeter level depth resolution would be desirable. Assuming state-of-the-art sub-bin processing techniques [23], this corresponds to histograms with thousands of bins per pixel. Moreover, the rate at which these histograms are acquired can vary from tens of frames per second (fps) for low speed applications to hundreds of fps for, say, an automotive application where objects may be moving at high speeds. Even a conservative estimate of a 30 fps megapixel camera leads to a large data-rate of 10^6 pixels/frame $\times$ 1000 bins/pixel $\times$ 2 bytes/bin $\times$ 30 fps $=$ 60 GB/sec. As shown in Fig. 1(b), the amount of data generated by this conventional full histogram capture method varies linearly with the desired depth resolution and exceeds the bandwidth of state-of-the-art data-transfer busses (like USB and PCIe) by orders of magnitude.

In this paper we propose a bandwidth-efficient acquisition strategy called *compressive single-photon histograms* (CSPH). Instead of capturing the full timing histogram in each pixel, a CSPH is constructed by mapping the time bins of the full histogram onto multiple "compressive bins" through an encoding step. We consider a family of compressive encoders that are linear, which means they can be represented as a simple matrix operation. Therefore, they can be implemented efficiently using vector addition operations that can be computed on-the-fly, as each photon arrives, without the need to store large arrays of photon timestamps in-sensor. CSPHs decouple the dependence of output data rate on the desired depth resolution. While a full histogram would require more time bins to achieve higher depth resolution, a CSPH can represent them using (almost) the same number of compressive bins. As illustrated in Fig. 1(a), CSPHs can reduce the required data rate by 1-2 orders of magnitude compared to the full histogram case.

We design and evaluate various CSPH coding schemes for SPAD-based 3D cameras. We propose a general decoding algorithm that directly estimates per-pixel scene depths from a CSPH, based on recent work in structured light [40]. Our method also accounts for different laser pulse shapes and pixel response characteristics, i.e., the system's impulse response function (IRF). We perform extensive simulations over a wide range of background and signal powers, laser waveform shapes and illumination schemes. We also evaluate depth reconstruction accuracy of our compressive acquisition method with real-world data captured using a single-pixel raster-scanning hardware prototype.

2. Related Work

Coarse in-pixel histogramming is one common strategy to reduce data rates in SPAD-based 3D cameras [9, 23, 28, 29, 53]. Despite the low time resolution in coarse histograms, it is possible to achieve high depth resolution by using wide pulses [23], pulse dithering [49], or with coarse-to-fine histogram architectures [66]. In this paper, we show that coarse histogramming is sub-optimal compared to other compressive histogramming strategies. Additional data reduction strategies, such as motion-driven operation [9] or multi-photon triggering [29], have also been proposed. Moreover, in the context of scanning-based systems, adaptive sampling methods have been proposed to reduce sampling rates and consequently data transfers [4, 25, 48]. These more complex methods can be used in a complementary manner with CSPHs to further reduce data rates.

Recently, *Fourier-domain histograms* (FDHs), were proposed for fast NLOS reconstruction [36, 42] and for single-photon 3D imaging [56, 57]. FDHs are one type of CSPHs that can achieve significant compression over regular histogramming [56]. In this paper, we present CSPH strategies that are not only more efficient than FDH for 3D imaging, but are also robust to diffuse indirect reflections commonly found in flash illumination systems.

Coding matrix design for 3D imaging has been studied in the context of correlation-based ToF (C-ToF) [19, 20, 22, 32, 33, 61] and structured light (SL) [6, 17, 18, 40]. In particular, there are interesting similarities between CSPH and C-ToF coding matrices, since both 3D cameras are based on the time-of-flight principle. Nonetheless, C-ToF coding is fundamentally different. In C-ToF, each coded *measurement* is captured sequentially with a coded light signal and sensor, making the noise statistics across measurements independent. In CSPHs, all coded *projections* are performed simultaneously on the same signal, making the noise statistics across projections dependent.

3. Single-Photon 3D Image Formation

Single-photon 3D cameras consist of a SPAD sensor and a periodic pulsed laser that illuminates the scene. Assuming direct-only reflections, the returning photon flux signal that will be captured by a SPAD pixel can be written as:

$$\Phi(t) = ah(t - t_z) + \Phi^{\text{bkg}} = \Phi^{\text{sig}}(t) + \Phi^{\text{bkg}} \qquad (1)$$

where $h(t)$ is the system's IRF which accounts for the pulse waveform and sensor IRF, a represents the returning signal photon flux, t_z is a time shift proportional to distance, and Φ^{bkg} corresponds to the background photon flux. Although simple, Eq. 1's model is a valid approximation in a wide variety of active illumination scenarios, in particular, for scanning systems [1].

SPAD-based 3D cameras sample $\Phi(t)$ using time-correlated single-photon counting (TCSPC) [45, 66]. The SPAD pixel, once triggered, starts acquiring photons. After detecting one photon, the photon timestamp is recorded, and the SPAD is inactive for a time period called the dead time ($\sim$50ns). As shown in Fig. 2, the above process is repeated for M cycles, and a histogram of the timestamps is constructed which approximates $\Phi(t)$. If the photons are time-tagged with a resolution, Δ, we can write the mean

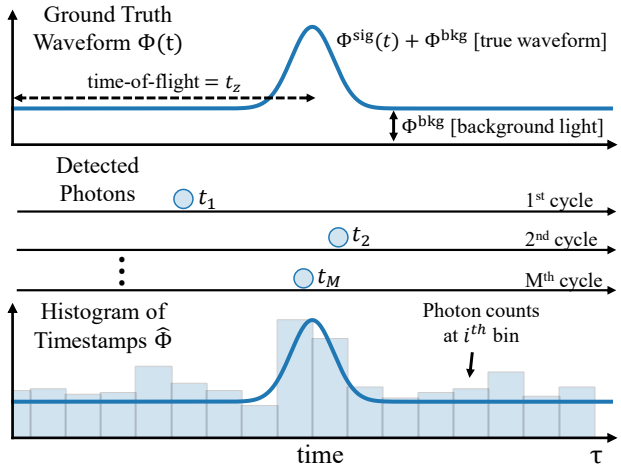


Figure 2. **Single-Photon Histogram Formation.** SPAD-based 3D cameras estimate distances by building a per-pixel histogram of the detected photons time-of-arrival. The histogram is a discrete approximation of the photon flux waveform incident on the pixel, which encodes distances in the time shift (t_z) of the pulse.

photon flux at histogram bin i as:

$$\Phi_i = \Phi_i^{\text{sig}} + \Delta\Phi^{\text{bkg}} \tag{2}$$

The vector, $\Phi = (\Phi_i)_{i=0}^{N-1}$, is the photon flux waveform histogram, where $N = \tau/\Delta$, and τ is the timestamp range which often equals the laser pulse repetition period. Here we assume that the SPAD sensor is being operated in asynchronous mode [15] or is capable of multi-event timestamp collection [23], which minimizes pile-up distortions [16, 26, 44], and guarantees that Φ_i is an appropriate approximation of $\Phi(t)$.

The histogram formation process generates a 3D histogram image, one histogram per pixel. In emerging megapixel SPAD arrays with picosecond time resolutions, building the histogram image off-sensor requires transferring thousands of timestamps per-pixel, leading to TB/s data rates. Moreover, building in-pixel histograms, would still require transferring the 3D data volume off-sensor for processing, which still results in impractical data rates of tens of GB/s. Overall, data bandwidth is an important practical challenge for emerging single-photon 3D cameras.

4. Compressive Single-Photon Histograms

In general, we *could* compress the 3D histogram image effectively *if* we had the entire histogram image. However, building and transferring the histogram image off the sensor is expensive. This raises the question, can we compress a histogram without ever explicitly constructing it? Recall that these histograms are created one photon at a time, raising the follow-up question: Can we compress the histogram in an online fashion where we see a photon (and its timing information) only once? This is challenging because compression schemes often require having access to the entire data before performing compression.

To answer the above question, we make two key observations. First, there is a class of linear compression techniques which can be expressed as a simple matrix-vector multiplication. Specifically, the compressed representation is the product of a $K \times N$ *coding matrix*, C, and the $N \times 1$ histogram Φ. The effectiveness of C can be measured by the compression ratio (N/K) that is achieved, while preserving down-stream task (e.g., depth estimation) performance.

Second, we observe that the entire histogram can be written as the sum of several one-hot encoding vectors, each vector representing one timestamp. Formally, let $t_j = (t_{j,i})_{i=0}^{N-1}$ be the one-hot encoding vector of the jth photon timestamp (T_j) detected, where all elements are 0 except for $t_{j,l} = 1$, in which $l = \lfloor \frac{T_j \mod \tau}{\Delta} \rfloor$. As illustrated in Fig. 3, the measured histogram, $\widehat{\Phi}$, can be written as:

$$\widehat{\Phi}_i = \sum_{j=0}^{M-1} t_{j,i} \tag{3}$$

where M is the total number of detected photons.

Given these observations, we can design an online histogram compression algorithm by simply multiplying the coding matrix with the one-hot encoding timestamp vector:

$$\widehat{B}_k = \sum_{i=0}^{N-1} C_{k,i}\widehat{\Phi}_i = \sum_{i=0}^{N-1}\sum_{j=0}^{M-1} C_{k,i}t_{j,i} \tag{4}$$

$\widehat{B}$ is the compressive single-photon histogram (CSPH), whose elements are coded projections of $\widehat{\Phi}$.

Although, Eq. 4 is expressed as a matrix-vector multiplication, a practical CSPH implementation would perform low-compute operations, per-photon, like conventional histograms. One possible implementation is to store C as a *lookup table* shared across pixels. For each new t_j with $t_{j,l} = 1$, the lth column of C is added to the per-pixel CSPH ($\widehat{B} = \widehat{B} + C_{:,l}$). Note that we never need to store the timing information explicitly, nor need to create the histogram. The only data stored and output by the compressive SPAD pixel is $\widehat{B}$. Given this on-the-fly compression method, a natural question is, what are good coding matrices for compressive single-photon 3D cameras?

4.1. How to design a coding matrix for 3D Imaging?

In theory, C can be chosen to be any set of K linear projections. In practice, however, we can define certain properties C should have to achieve high compression rates while preserving 3D imaging performance.

The ith column of C can be interpreted as a code word of length K that represents the ith time bin. We can view this code word vector as a point in a K-dimensional space. Moreover, consider the curve that is traced by the N points

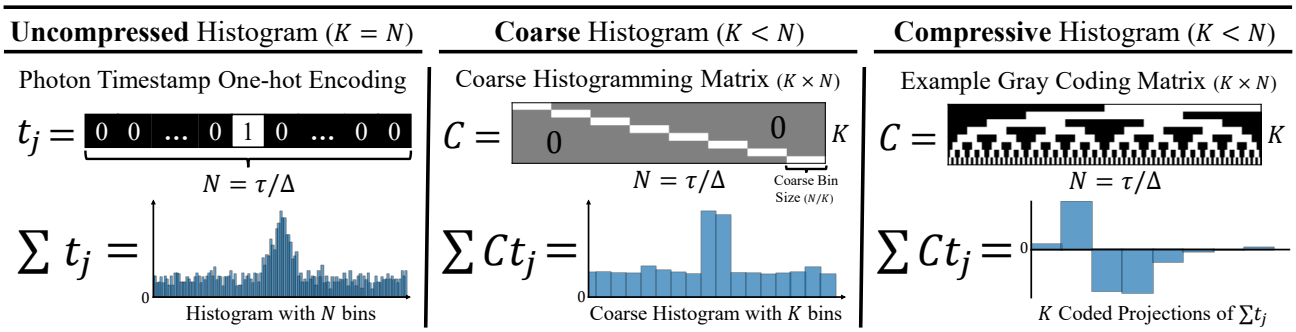


Figure 3. **On-the-fly Histogram Formation.** Single-photon timestamp histograms are often generated on-the-fly, as each photon timestamp comes in. The left column shows how a histogram, whose bin width matches the timestamp resolution (Δ), is formed as the sum of timestamps represented as one-hot encoded vectors. Transferring such a large histogram for every pixel can be impractical. By multiplying each timestamp with a down-sampling matrix to group timestamps into coarser bins, the size of the histogram can be reduced at the cost of resolution (middle column). Alternatively, as shown in the right column, a compressive histogram can be created by multiplying each timestamp with a coding matrix and adding them up as each photon timestamp comes in. A well-designed coding matrix can efficiently encode the location of the peak from which distance can be computed.

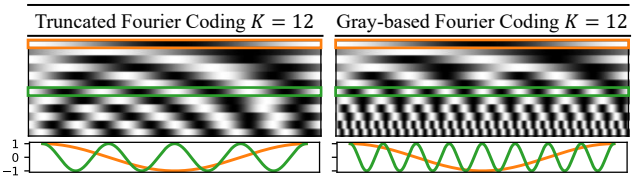


Figure 4. **Example Coding Matrices.** Truncated Fourier and Gray-based Fourier matrices with 12 codes (rows). The odd and even rows of the truncated Fourier matrix are given by $\cos(\frac{2\pi f i}{N})$ and $\sin(\frac{2\pi f i}{N})$, respectively, where $f = \lceil 0.5k \rceil$, $1 < k < K$, and i is the column. The odd and even rows of the Gray-based Fourier matrix are given by $\cos(\frac{2\pi 2^{f-1} i}{N})$ and $\sin(\frac{2\pi 2^{f-1} i}{N})$, respectively, and for $k > \log_2(N)$ the rows are generated using codes from the truncated Fourier matrix that have not been used.

(columns of C), denoted as the coding curve $\mathcal{C}$. In this section, we make use of the coding curve concept to define desirable properties of C. We would like to note that the coding curve construct has been previously used for other active imaging systems including: C-ToF [20, 22], SL [18], and FLIM [34]. A "good" compressive single-photon 3D imaging coding matrix will have the following properties:

Uniqueness Property: Each point in $\mathcal{C}$ is unique, i.e., $\mathcal{C}$ should be non self-intersecting. This guarantees that two different bins are not represented by the same code word.

Robustness Property: $\mathcal{C}$ should be locality preserving. Meaning that if we add a small perturbation to a point on the curve, it should map to neighboring points along the curve, which correspond to code words with similar time bins.

Indirect Reflections Property: The coding functions (rows of C) encode information of the photon flux waveform in a similar way as correlation functions do in C-ToF imaging [19, 20]. In the presence of diffuse indirect light reflections, the direct-only model from Eq. 1 becomes invalid. Diffuse indirect reflections arise when imaging concave geometries or in the presence of volumetric scattering, and is particularly problematic in flash illumination systems. Gupta et al. [19] showed that diffuse indirect reflections appear in $\Phi(t)$ as smooth band-limited signals. Therefore, the direct-only model will still apply for coding functions with frequencies above a scene dependent threshold.

Band-limit Property: The coding functions should not all be high-frequency functions when the goal is to encode a photon flux waveform with a smooth system IRF ($h(t)$). This is because if a given coding function is composed of frequencies that are above the bandwidth of $h(t)$, then the expected encoded value will be 0 (see supplement). The orange and green lines of Fig. 5 show example of smooth IRF's. This means that, although, high frequencies can mitigate indirect reflections, very high frequency codes may not be useful if they are outside of $h(t)$ bandwidth.

4.2. CSPH Coding Schemes

In this paper, we analyze and evaluate the following CSPH coding matrices:

1. **Coarse Histogram**: C is a downsampling matrix, where each row is a rectangular window with length N/K, as shown in Fig. 3. This matrix does not fulfill the uniqueness property because each window maps multiple bins to the same code word vector. This type of CSPH is equivalent to the commonly used coarse in-pixel histograms [9, 28, 29, 53].
2. **Truncated Fourier** [56]: C is made up of the first K rows of the discrete Fourier transform matrix, skipping the zeroth harmonic. This matrix fulfills the uniqueness property, and at higher K it may contain high frequency codes that can mitigate indirect reflections.
3. **Continuous Gray (Proposed)** : Also known as Hamiltonian codes [20], the Cont. Gray coding curve is a Hamiltonian cycle on a K-dimensional hypercube, which is provably locality preserving (robustness prop-

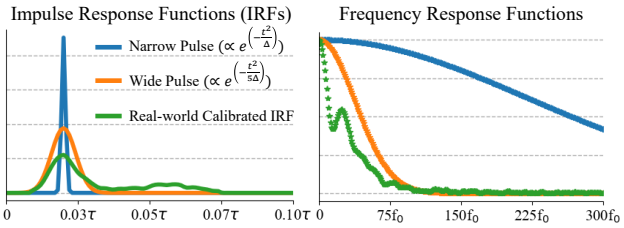


Figure 5. **System Impulse Response Functions ($h(t)$).** Gaussian pulses are commonly used approximations used for the system IRF because the single-photon 3D cameras use pulsed illumination signals. Although this approximation may be valid for some systems, the experimental system used for the results in Sec. 6.1 exhibited an IRF with a long-tail (green line). Analyzing the frequency-domain representation of the IRF is a useful step when designing the coding matrix of a compressive single-photon 3D camera.

erty) [12]. The rows of C are generated by constructing K-bit Gray code [14], where each code will have length 2^K, and linearly interpolating them to have length N. For a histogram of length $N = 2^K$ this coding matrix becomes fully binary. Fig. 3 shows a Gray C with $K = 8$.

4. **Gray-based Fourier (Proposed)**: For N histogram bins, Gray coding is only valid for $K \leq \log_2(N)$ because the higher-order coding functions start aliasing. However, in lower SNR scenarios it is sometimes desired to increase K to preserve depth precision. To this end, we design a new coding scheme that combines properties of Gray and Fourier coding. For the first $k \leq 2\log_2(N)$ rows, Fourier components are sampled using the frequency doubling pattern observed in the Gray coding matrix in Fig. 3. For the remainder $2\log_2(N) < k < K$ codes, we revert back to a truncated Fourier sampling using the remaining frequencies.

In the supplement we present further analysis on the properties of these coding matrices, and evaluate additional coding schemes including: Hadamard and short-time Fourier.

5. When is isometric compression achieved?

A CSPH achieves isometric compression when its performance is within a specified margin of the uncompressed full-resolution histogram (FRH). Specifically,

$$\text{IsometricCompression}(\epsilon) := \varepsilon_{\text{diff}} \leq \epsilon \tag{5}$$

where $\varepsilon_{\text{diff}} = |\varepsilon_{\text{FRH}} - \varepsilon_{\text{CSPH}}|$, ε_{FRH} and $\varepsilon_{\text{CSPH}}$ are the performance metrics for a CSPH and a FRH, and ϵ is the desired performance difference margin.

To quantify 3D imaging performance, we compute the relative mean depth errors (MDE) over the depth range through Monte Carlo simulations using Eq. 1's direct-only model, as in [22] (see supplement). Consequently, to identify isometric compression we take the difference between the relative MDE of a FRH (ε_{FRH}) and the CSPH ($\varepsilon_{\text{CSPH}}$) and classify the difference into different margins. For context, in a 3D imaging scenario with a 10m depth range, a relative MDE difference of $\varepsilon_{\text{diff}} = 0.1\%$ corresponds to the CSPHs MDEs being within 1cm of the FRH MDEs.

Depth Estimation: To decode depths from a CSPH, we compute the zero-mean normalized cross-correlation (ZNCC) [40] between $\widehat{B}$ and C:

$$\widehat{t}_z \propto \arg\max_i \frac{C^h_{:,i} - \text{mean}(C^h_{:,i})}{\left\|C^h_{:,i} - \text{mean}(C^h_{:,i})\right\|} \cdot \frac{\widehat{B} - \text{mean}(\widehat{B})}{\left\|\widehat{B} - \text{mean}(\widehat{B})\right\|} \tag{6}$$

where $C^h_{:,i}$ is the ith column of C^h, and C^h is the coding matrix, C, with each row convolved with the system IRF h. To compute depths for FRHs we use matched filtering [63]. See supplement for more details on ZNCC depth estimation.

5.1. Isometric Compression Analysis

A high-performance single-photon 3D imaging system will match the laser pulse width with the SPAD sensor time resolution (Δ) [23,24]. In this section, we analyze isometric compression at a wide range of signal-to-background ratio (SBR) and photon count levels[1], for the case of a FRH with $N = 1024$ bins that records a Gaussian pulse of width Δ, i.e., $h(t) \propto \exp^{-(t)^2/\Delta}$ (blue line in Fig. 5). In addition to the CSPH coding schemes from Sec. 4.2, we also evaluate the following two baselines:

- **Coarse Hist. (Wide Pulse)** [23]: A coarse histogram C paired with a wide Gaussian pulse width that matches the window length, which enables sub-bin precision.
- **Truncated Timestamps**: A FRH constructed with *at most* K timestamps, even if the number of detected photons is $> K$. Although, not a CSPH, this is an important baseline corresponding to the simplest SPAD pixel that transfers the same amount of data as a size K CSPH.

Extreme Compression: Fig. 6a shows the relative mean and median depth errors for different CSPH with $K = 8$, resulting in a compression ratio of 128x. The ideal FRH, obtains near 0 error in the visualized SBR and photon count levels. In this extreme compression regime, Gray coding is the only CSPH that achieves an isometric compression where, $\varepsilon_{\text{diff}} \leq 0.01\%$, at various SBR and photon count levels, essentially matching FRH performance (Fig. 6b). The difference in the mean and median error trends indicates that at low SBR and low photon counts, Gray coding produces either high or near zero errors. On the other hand, the error magnitudes of truncated Fourier coding are similar across SBR and photon count levels. Unfortunately, even at high SBR levels, truncated Fourier still does not each isometric compression with $\varepsilon_{\text{diff}} \leq 0.01\%$. Gray-based Fourier coding consistently outperform truncated Fourier, and is more robust to outliers than Gray coding at low SBR. As expected, the coarse histogramming methods are either, quantization-limited due to low time resolution, or noise-limited when using a wide pulse. Finally, only transferring

[1] $\text{SBR} = \frac{\sum_{i=0}^{N-1} \Phi_i^{\text{sig}}}{N\Delta\Phi^{\text{bkg}}}$, Photon Counts $= \sum_{i=0}^{N-1} \Phi_i$

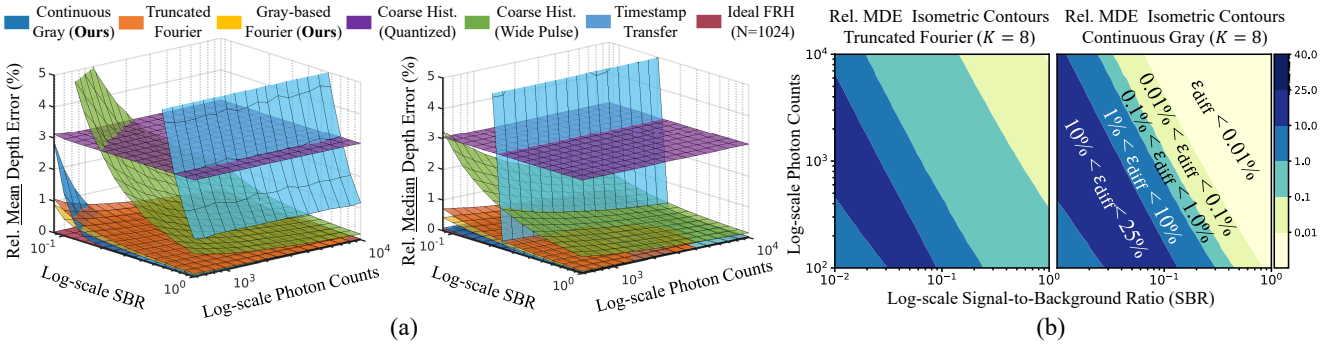


Figure 6. **Isometric Compression Analysis at 128x Compression.** (a) Shows the relative mean and median depth errors computed as described in Sec. 5. At this high compression level, where CSPH methods use $K = 8$ codes, only the Gray-based and Fourier-based CSPHs achieve low errors at a wide range of SBR and photon count levels. For a fixed SBR and photon counts, the mean and median errors of Gray coding and timestamp transfer can differ significantly, due to outliers. (b) Shows the isometric contours generated by the relative MDE difference of truncated Fourier and Gray coding with the ideal FRH. At SBR ≥ 0.1 and photon counts ≥ 1000 these CSPHs depth errors are consistently within 1% of the FRH. However, at SBR ≤ 0.05 and photon counts ≥ 1000 their performance significantly degrades relative to that of the ideal FRH. Finally, at low SBR and photon-starved settings (bottom left), the relative performance difference decreases again because FRH also starts performing poorly.

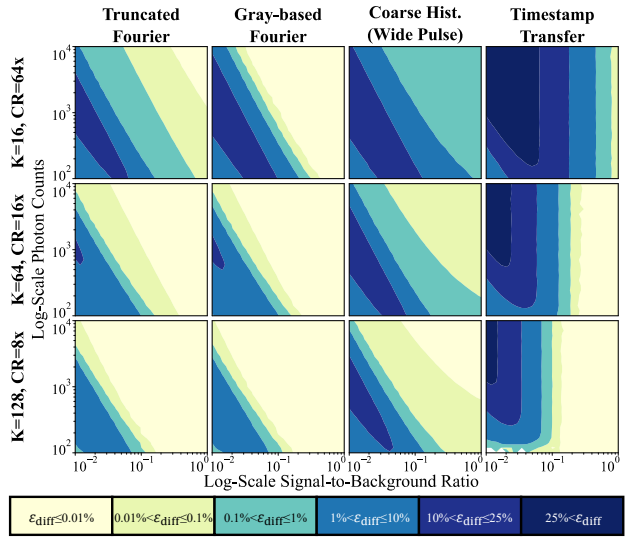


Figure 7. **Compression vs. Performance.** As we increase K the isometric compression regions with $\varepsilon_{\text{diff}} \leq 0.01\%$ increase for all methods. For SBR and photon count levels greater than ~ 0.1 and ~ 1000, Gray-based Fourier coding with $K \geq 16$ performs as well as a FRH with $N = 1024$. Truncated Fourier Coding, on the other hand, requires higher K to reach a $\varepsilon_{\text{diff}} \leq 0.01\%$ at those SBR and photon count levels. Overall, carefully designed coding matrices such as Fourier and Gray-based, consistently outperform standard approaches (coarse histograms and timestamp transfer).

8 timestamps, although practical, leads to poor performance at low SBR and to many outliers at high SBR.

Compression vs. Performance: Fig. 7 shows how the isometric compression contours change as we increase K for different coding schemes. As K increases and compression decreases, the performance of all methods improves and approaches FRH performance, in particular for carefully designed CSPH coding schemes like Gray-based and truncated Fourier. At $K \leq 64$ the benefits of Gray-based Fourier coding are more evident as the isometric contours where its performance matches FRH covers the biggest range of SBR and photon count levels (i.e., $\varepsilon_{\text{diff}} \leq 0.01\%$). As K continues to increase, the truncated and Gray-based Fourier coding matrices become similar, making their performance nearly identical, as observed in the $K = 128$ case. Overall, Gray-based and Fourier-based CSPH coding, consistently outperform current photon timestamp storage and transfer approaches, i.e., coarse histograms and direct transfer of timestamps. Interestingly, transferring only K timestamps significantly outperforms a coarse histogram at many SBR and photon count levels.

Isometric Compression with Wide Pulses: Here we analyzed CSPHs for a Gaussian pulse of width Δ. As the pulse width increases, the effective time resolution of the system decreases, impacting the performance of the FRH baseline, which makes the isometric compression regions with low relative differences become larger at lower K. Moreover, as seen in Fig. 5, slightly widening the pulse decreases the frequency content of the signal significantly, making compressive coding strategies more efficient because they only need to sample the non-zero frequencies. Please refer to the supplement for results and analysis using wider pulses.

6. Compressive Single-Photon 3D Imaging

In this section, we evaluate CSPH coding approaches on real-world data from a scanning-based system [15], and simulated data from a flash-illuminated system.

6.1. Real-world Scanning-based System Results

To evaluate the effectiveness of CSPHs on real SPAD timestamp data we downloaded and pre-processed the data acquired with a scanning-based system [15]. The pre-processed raw histograms have $\Delta = 8$ps and $N = 832$

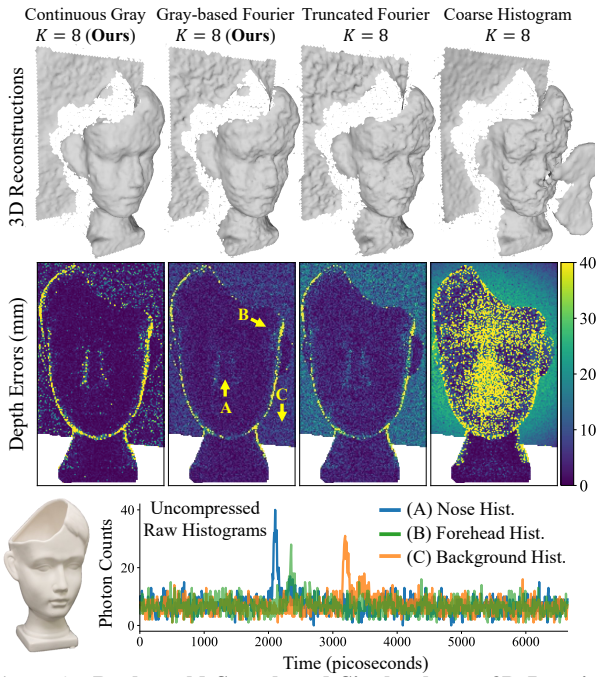


Figure 8. **Real-world Scan-based Single-photon 3D Imaging.** The depth and depth error images for different CSPH with $K = 8$ codes. The mean and median absolute errors (in mm) achieved by each method from left to right are: [7, 1], [6, 4], [9, 6], [23, 13].

(e.g., histograms in Fig. 8). For depth estimation we extract the center pixel histogram, denoise it, and use that as the system IRF (green line in Fig. 5). Ground truth depths are obtained from the FRH with light Gaussian denoising, and we mask the pixels where even FRH had too low of an SBR to estimate reliable depths (white regions in Fig. 8 images).

Fig. 8 shows the recovered 3D reconstructions using different CSPH at a $CR = 104$x ($K = 8$). Similar to our simulations, we find that Gray coding can essentially achieve 0 errors for pixels with sufficient signal, while sometimes making large errors (outliers). In contrast, truncated Fourier and Gray-based Fourier are robust to outliers, but make many small and medium sized errors leading to lower quality 3D reconstructions in this example. Moreover, we found that the background wall histograms exhibited a longer tail than the foreground face histograms due to indirect reflections (see supplement). Indirect reflections cause systematic errors in truncated Fourier, while Gray-based Fourier and Gray coding are more robust to these errors since their C have higher frequency coding functions, as discussed in Sec. 4.1. Finally, the wide system IRF allows a coarse histogram to achieve sub-bin precision using ZNCC decoding. Nonetheless, its performance is significantly worse than the other CSPH methods. Additional results at different K and another scan can be found in the supplement.

6.2. Simulated Flash Illumination System Results

SPAD arrays are often used in flash illumination systems to achieve a fully solid-state single-photon LiDAR [23]. To evaluate CSPHs in a flash illumination system, we used physically accurate histogram images rendered with MitsubaToF [44] ($\Delta = 50$ps, $N = 2000$) obtained from [21]. To simulate the FRHs we set the mean photon count and mean SBR levels for the scene, and scale the histogram and background image (R channel of RGB) accordingly.

Fig. 9 shows the resulting depth images and depth errors for two different scenes. In addition to the edges where true depth is ambiguous, both scenes have regions with very low SBR where even the FRH has some depth errors (stove in kitchen, and the mat in bathroom). Due to indirect reflections, truncated Fourier makes significant systematic errors even when using $K = 40$ codes. On the other hand, Gray-based Fourier recovers highly accurate depths like an FRH, while using 50-100x less data. Please refer to the supplement for additional comparisons.

	In-Pixel Memory	In-Pixel Compute (per-photon)	In-Pixel Compute (per-depth)	Per-pixel Output Data Rate	Depth Precision
Full-Resolution Hist. (In-sensor Depth Est.)	High $\mathcal{O}(N)$	Medium $\mathcal{O}(1)$	High $\mathcal{O}(\geq N)$	Low $\mathcal{O}(1)$	High
Full-Resolution Hist. (Off-sensor Depth Est.)	High $\mathcal{O}(N)$	Medium $\mathcal{O}(1)$	Low (None)	High $\mathcal{O}(N)$	High
Timestamp Transfer (Off-sensor Depth Est.)	Low $\mathcal{O}(1)$	Low (None)	Low (None)	High $\mathcal{O}(M)$	High
Coarse Hist. (Off-sensor Depth Est.)	Medium $\mathcal{O}(K)$	Medium $\mathcal{O}(1)$	Low (None)	Medium $\mathcal{O}(K)$	Low
CSPH (Proposed) (Off-sensor Depth Est.)	Medium $\mathcal{O}(K)$	High $\mathcal{O}(K)$	Low (None)	Medium $\mathcal{O}(K)$	Medium

Table 1. **SPAD-based 3D Camera Design Trade-offs**. Qualitative comparison of the memory, compute, and data rate requirements for different SPAD-based 3D camera designs.

7. Discussion and Limitations

SPAD-based 3D cameras with high spatio-temporal resolution can produce unmanageable data rates. To reduce their data bandwidth, we proposed to capture a compressive representation (CSPH) of the high-resolution timing histogram, from which depths can be computed. The CSPH is built in an online manner by projecting each photon timestamp with a coding matrix and aggregating them. By designing the coding matrix appropriately, a CSPH can match the depth precision of a full-resolution histogram in a wide range of scenarios, while outputting significantly less data.

In-sensor Memory and Compute Considerations: Although, CSPHs can reduce the in-pixel memory and sensor data transmission rates, this comes at the expense of higher per-photon computations, as summarized in Tab. 1. While conventional histograms require a single addition per-photon (increment histogram bin), CSPHs perform K additions per-photon. Nonetheless, it may be possible to implement these K additions efficiently using SIMD processing. The next step in this line of work is to further analyze these memory, computation, and data transmission trade-offs from a hardware perspective.

Why not compute depths in-sensor? One way to reduce the data rates would be to compute the per-pixel depths in-pixel. However, due to the non-linearity of depth estima-

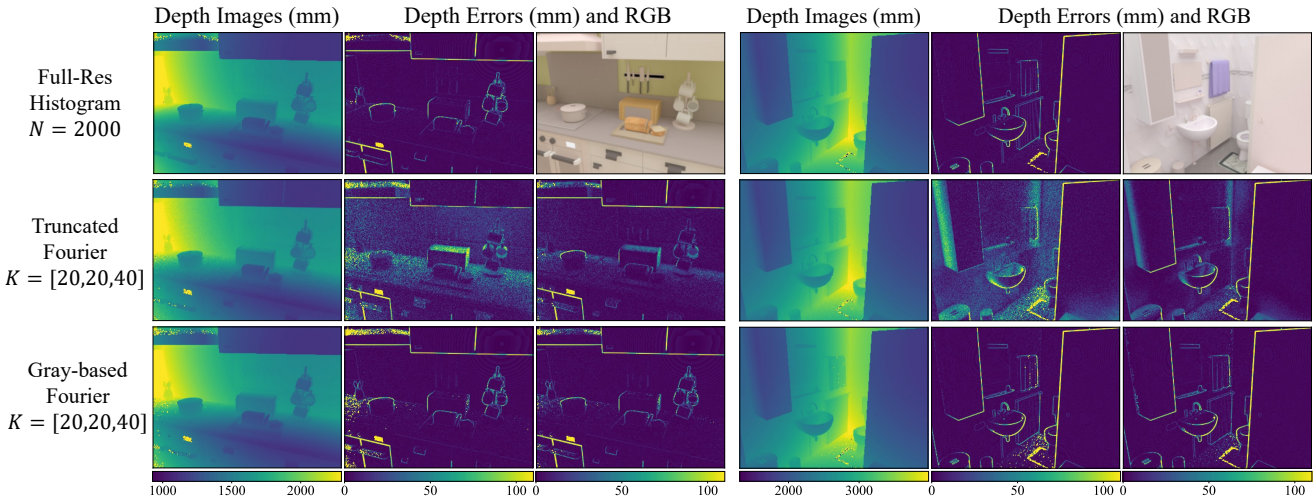


Figure 9. **Flash Illumination Compressive Single-Photon 3D Imaging.** Depth images and depth errors for CSPH methods applied to histograms of flash illuminated scenes. The kitchen and bathroom scenes were simulated with a mean photon count and mean SBR of (1000, 0.25) and (1000, 0.5), respectively. The top row has the recovered depths and depth errors of an FRH with 2000 bins, and the RGB image of the scene. The second and third rows have the CSPH depths ($K = 20$), and the depth errors for $K = 20$ (middle column) and $K = 40$ (right column). The mean and median absolute errors (in mm) for the kitchen scene were: FRH-2000: (14, 3), truncated Fourier-20: (26, 10), Gray Fourier-20: (22, 3), truncated Fourier-40: (16, 4), Gray Fourier-40: (14, 3). Similarly, for the bathroom scene: FRH-2000: (10, 3), truncated Fourier-20: (24, 11), Gray Fourier-20: (12, 3), truncated Fourier-40: (14, 4), Gray Fourier-40: (11, 3).

tion (e.g., via peak finding), it requires building and storing the full-resolution histograms in the SPAD sensor chip. Although, this approach could provide near-optimal compression, it requires significant in-pixel memory. Moreover, advanced depth estimation algorithms, such as matched filtering, can have computational complexities larger than the length of the histogram (N).

Hardware Implementation: CSPHs are designed to reduce the per-pixel output data rate, which requires in-pixel implementation. Recent advances in 3D-stacking CMOS technology has enabled the in-pixel implementation of one type of CSPH, namely coarse histograms [9, 28, 29, 53]. The next step in this line of work is to explore SPAD pixel architectures that implement different CSPHs, which may impose interesting constraints on the structure of C. For instance, the architecture for a binary C (e.g., coarse histograms or Gray coding with $K = \log_2(N)$) may be simpler than for a C with continuous values.

Code Optimization: Instead of designing a coding matrix based on the heuristics discussed in Sec. 4.1, C could be optimized. We attempted to optimize C by setting it to the PCA basis learned over a dictionary of shifted Gaussian pulses. Unfortunately, this method simply leads to a Fourier matrix (see supplement). One direction for future work could explore optimizing C through gradient descent [40].

Photon-starved regime: When less than 20 photons are recorded, timestamp transfer can outperform a CSPH at $SBR > 1$. At lower SBR levels all methods begin to fail and perform comparably. We present this analysis in the supplement. Ultimately, low SBR and photon-starved scenarios require denoising to recover reliable depths [35, 46]. Denoising a CSPH instead of the full 3D histogram image may provide some interesting computational benefits.

Non-linear Compression: It may be possible to design a non-linear on-the-fly compression method. One challenge in the design of such method is that it may not be able to leverage the priors that enabled the design of coding matrices robust to noise and indirect reflections. Nonetheless, exploring non-linear online compression algorithms remains an interesting direction for future work.

Efficient Depth Estimation: ZNCC depth estimation allows comparing different coding matrices under a single framework. One limitation of our ZNCC implementation is its linear computational and memory complexity. Given that ZNCC is a template matching algorithm, efficient coarse-to-fine implementations may be possible [2]. Alternatively, algorithms tailored for a particular C may provide further computational benefits. For instance, Fourier coding has different analytical [11,47], optimization-based [27,56], and data-driven [59] depth decoding algorithms.

Social Implications: The work in this paper may contribute towards the deployment of SPAD-based 3D cameras. The emergence of SPAD-based NLOS imaging, raises privacy concerns as these systems become available. Additionally, 3D applications such as autonomous vehicles, may have unintended socio-economic and environmental implications.

Acknowledgments: This work was supported by the Department of Energy and National Nuclear Security Administration (DE-NA0003921), National Science Foundation (1846884, 1943149, 2107060, 2003129), and a Draper TIF grant from UW-Madison. U.S. DOE full legal disclaimer: `https://www.osti.gov/stip/about/disclaimer`.

Supplementary Document for "Compressive Single-Photon 3D Cameras"

Felipe Gutierrez-Barragan, Atul Ingle, Trevor Seets, Mohit Gupta, Andreas Velten
University of Wisconsin-Madison
{fgutierrez3, ingle, seets}@wisc.edu, mohitg@cs.wisc.edu, velten@wisc.edu

S. 1. Coding Matrix Properties

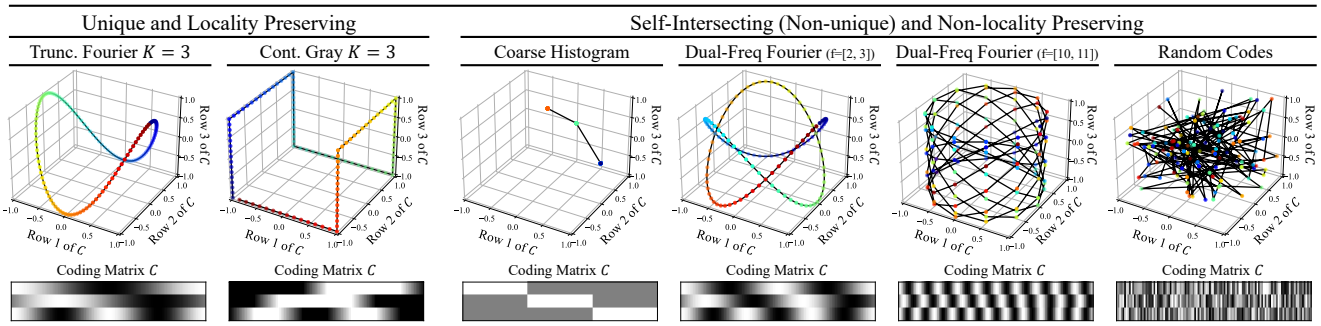
In this section we provide additional intuition for the coding matrix properties discussed in Section 4.1 of the main paper.

S. 1.1. A Geometric Perspective

Recall that the ith column of the coding matrix, C, can be interpreted as a code word of length K that represents the ith time bin. We can view this code word vector as a point in a K-dimensional space. Moreover, consider the curve that is traced by the N points (columns of C), denoted as the coding curve $\mathcal{C}$. In other words, the rows of C parameterize the coding curve $\mathcal{C}$, as illustrated in Suppl. Fig. 1.

Uniqueness Property (non self-intersecting): Suppl. Fig. 1 shows different coding curves parameterized by the coding matrices below them. Assuming a direct-only reflection model, i.e., $\Phi(t) = a\delta(t - t_z) + \Phi^{\text{bkg}}$, each point in the curve ($K \times 1$ code word vector) will correspond to the projection of a different time-shift t_z [20]. Therefore, the uniqueness property guarantees that each time-shift will map to a different point. If two or more different time-shifts map to the same point in the coding curve, that means the coding scheme will have ambiguities when estimating depths. For instance, as observed in the coarse histogram and dual-freq Fourier coding curves in Suppl. Fig. 1, the curves are self-intersecting at multiple points, which means that those points correspond to more than one true depth value.

Robustness Property (locality preserving): If we add a small perturbation to a point on a locality preserving coding curve, the perturbed point may map to a neighboring point *along* the curve. In Suppl. Fig. 1, neighboring points along the curve will have similar colors. For instance, in the Truncated Fourier curve, if we add a perturbation to a "reddish" point, then its nearest neighbor is likely to be another "reddish" point. On the other hand, a similar perturbation added to certain red points in the Dual-Freq Fourier curves may cause them to map to a blue point. These perturbations are similar to adding noise to the coded projection of the ground truth histogram, and mapping to non-neighboring points is similar to estimating a very different depth from the true depth. Hence, coding matrices with non-locality preserving curves are less robust to noise.



Supplementary Figure 1. **Example Coding Curves.** A coding curve is formed by plotting each column of the $K \times N$ coding matrix in K dimensional space. This figure shows examples with $K = 3$. Observe that unlike dual-frequency Fourier coding and random coding, truncated Fourier and Gray coding provide non-self-intersecting and locality preserving coding curves. This implies that the latter methods have a unique one-to-one mapping between the true (unknown) depth values and the coded measurements, and small perturbations (due to noise) do not cause large jumps in the compressed measurements.

S. 1.2. Coding Function Bandwidth Selection

The band-limit and indirect reflection properties, discussed in Sec. 4.1 of the main paper, give us a heuristic for choosing the frequency content of our coding functions (i.e., rows of C). We begin by analyzing the band-limit property.

Band-limit Property: Consider the discrete Fourier series of the system IRF, $h = (h_i)_{i=0}^{N-1}$, whose maximum harmonic is $b < N$.

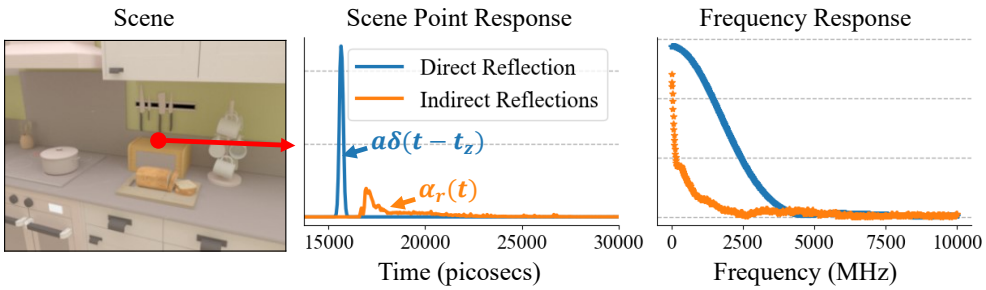
$$h_i = \sum_{k=0}^{b} H_k \exp\left(\frac{\mathrm{j}\,2\pi k i}{N}\right) \tag{7}$$

where $\mathrm{j} = \sqrt{-1}$, $0 \leq i < N-1$, and H_k are the Fourier coefficients. Moreover, let $c = (c_i)_{i=0}^{N-1}$ be one coding function of C. The band-limit property says that c should not only be composed of frequencies larger than the system's band-limit (i.e., b). If c is only composed of frequencies that are not contained in b, the inner product of c and h will be zero, due to the orthogonality of sinusoids [54]. Hence, the coded measurement captured by c will not capture any information from the histogram photon flux waveform.

Indirect Reflection Property: Let $\alpha(t)$ be the impulse response of a point in the scene. Under the direct-only assumption, $\alpha(t) = a\delta(t - t_z)$, the impulse response is just the time-shifted direct reflection of that point. In the presence of diffuse indirect reflections, $\alpha(t)$ can be written as the sum of the direct reflection and the diffuse indirect reflections:

$$\alpha(t) = a\delta(t - t_z) + \alpha_r(t) \tag{8}$$

where $\alpha_r(t)$ are the diffuse indirect reflections. Suppl. Fig. 2 shows an example of a scene point impulse response with diffuse indirect reflections (orange line). As observed in the frequency response, the indirect reflections have a much smaller bandwidth than the direct reflection. Beyond 1 GHz, the direct reflection is the dominant component. Therefore, for coding functions mainly composed of frequencies above 1 GHz, the direct-only assumption continues to apply because the indirect component for frequencies above the threshold will be near 0. In general, the threshold is scene dependent [19], but scenes with relatively similar scales may have similar thresholds.



Supplementary Figure 2. **Diffuse Indirect Reflections Example.** Observe that the direct reflection has a narrow support in time and hence has much wider spectrum compared to the indirect response which, in this scene, does not go beyond 1 GHz.

S. 2. ZNCC Depth Estimation

Given a CSPH, $\widehat{B}$, acquired with a coding matrix C, we want to estimate the depth encoded by $\widehat{B}$. There exists a general depth decoding algorithm based on the zero-mean normalized cross-correlation (ZNCC) between $\widehat{B}$ and C, originally proposed for structured light [6, 40], and later applied to iToF imaging [21, 22]. Assuming a direct-only model, ZNCC decoding becomes optimal as noise goes to zero [40]. Moreover, even when the direct-only assumption does not hold, ZNCC continues to provide robust depth estimates [6, 21, 40].

Band-limited ZNCC Decoder [40]: Let C^h be the band-limited coding matrix, whose kth row, is defined as:

$$C^h_{k,i} = (h * C_{k,:})_i \tag{9}$$

where h is the system IRF, and $C_{k,:}$ is the kth row of the coding matrix. Let $C^h_{:,i}$ be the ith column of C^h. We can estimate depths, by finding $C^h_{:,i}$ that produces the maximum ZNCC with $\widehat{B}$ as follows:

$$\widehat{t_z} \propto \arg\max_i \frac{C^h_{:,i} - \text{mean}(C^h_{:,i})}{\left\|C^h_{:,i} - \text{mean}(C^h_{:,i})\right\|} \cdot \frac{\widehat{B} - \text{mean}(\widehat{B})}{\left\|\widehat{B} - \text{mean}(\widehat{B})\right\|} \tag{10}$$

Depths can be estimated from the time-shift t_z. Note that, although, we capture the CSPH using the input C matrix, we estimate distances using the band-limited coding matrix C^h. Therefore, Suppl. Eq. 10 assumes the system's IRF, h, is obtained in a prior one-time calibration step. Furthermore, if all coding functions in C are zero-mean, like the Fourier codes in Suppl. Fig. 1, then the above algorithm can be implemented as a normalized cross-correlation (NCC), where we do not need to subtract the mean in the numerator or denominator [40].

Suppl. Eq. 10 accounts for the system IRF (h) when estimating depths by using C^h instead of C. Not accounting for the IRF is equivalent to assuming that the system IRF is a dirac-delta function. In our experiments we find that not accounting for h can reduce the performance and sometimes introduce systematic depth errors in coding schemes that contain coding functions with frequency content outside of the band-limit of h.

Depth Estimation for Full-Resolution Histogram: To compute depths for FRHs we used matched filtering [63]. Interestingly, matched filtering is similar to applying Eq. 6 when C is an $N \times N$ identity matrix.

S. 3. Monte Carlo Simulations and Isometric Compression Analysis

In this section we provide further details on the Monte Carlo simulations performed to quantify the performance of each coding scheme. We also present additional mean and median depth errors results and isometric compression results for a wide range of SBR and photon count levels at various compression ratios. Finally, we present a similar analysis but for the case of a wide Gaussian pulse.

S. 3.1. Quantifying Coding Scheme Performance

A high-performance coding scheme should be able to provide high depth precision for the full depth range and at a wide range of signal and background light levels. One way to quantify the performance of a coding scheme is to compute the *mean expected absolute depth error* (MDE) [20, 22] at a wide range of signal and background light levels. In this paper we report the relative MDE, which is the MDE metric divided by the maximum possible depth value. Moreover, since depths are proportional to time shifts, the relative depth errors will be equivalent to relative time shift errors, hence we use these interchangeably when computing the relative MDE.

Monte Carlo MDE Computation: For a given coding scheme, fixed average photon counts ($\sum_{i=0}^{N} \Phi_i$), signal-to-background ratio (SBR $= \frac{\sum_{i=0}^{N} \Phi_i^{\text{sig}}}{N \Delta \Phi^{\text{bkg}}}$), histogram length ($N$), and system IRF ($h(t)$), we compute the MDE as follows. First, we generate $h(t)$ at D equispaced time shifts over N, and scale and vertically shift the noiseless $h(t)$ according to the input SBR and mean photon counts. Second, we draw a poisson sample at each time bin (i.e., add noise). Third, we encode the noisy histogram as in Eq. 4 of the main paper. Then, we compute the D time shifts as in Suppl. Eq. 10, and take the absolute difference with the ground truth. Finally, the above process is repeated S times, and the expected errors are computed for all $S \cdot D$ time shifts, averaged to obtain the MDE, and divide by N to obtain the relative MDE. Moreover, we also report the median expected depth error which is in the same way as the MDE, but instead of averaging the $S \cdot D$ time shifts, we take the median. For all the simulations in this paper we use $N = 1024$, $S = 1000$, and $D = 64$.

S. 3.2. Isometric Compression Analysis

In this section we present the mean and median depth error results and the isometric compression analysis for all coding schemes at different compression ratios. We compare the relative MDE of the FRH (ε_{FRH}) and the CSPH ($\varepsilon_{\text{CSPH}}$) to determine isometric compression. Since Continuous Gray coding is only defined up to $K \leq \log_2(N) = 10$ for $N = 1024$, we defined a coding scheme that extends Gray coding higher K values using a similar intuition as Gray-based Fourier coding in the main paper, and we denote it as Fourier-based Gray coding:

- **Fourier-based Gray Coding:** Suppl. Fig. 5 shows a Fourier-based Gray coding matrix for $K = 16$. The first $\lfloor \log_2(N) \rfloor$ rows of this matrix are the same as Gray coding. The second $\lfloor \log_2(N) \rfloor - 2$ codes are the 90 degree shifted version of each Gray code, excluding the first two gray codes. For the remainder codes we sample binary square functions at increasing frequencies that have not been sampled, in the same way the Gray-based Fourier coding samples Fourier components. The Fourier-based Gray coding matrix at $K > 2\lfloor \log_2(N) \rfloor - 2$, becomes a binary form of the Gray-based Fourier matrix with the initial rows in a different order.

Main Observations: Suppl. Fig. 3 and 4 show the isometric compression results for compression ratios of 64 ($K = 16$) and 16 ($K = 64$). At high compression ratios of 64x only Gray-based Fourier and Fourier-based Gray coding are the only coding schemes that are able to achieve isometric compression with $\varepsilon_{\text{diff}} < 0.01$ for a wide range of SBR and photon count levels. As we use a greater number of codes (reduce compression), Truncated Fourier coding starts approaching the performance of the proposed coding schemes in the isometric compression regions with low $\varepsilon_{\text{diff}} < 0.01$. Overall, commonly used approaches such as coarse histograms and timestamp transfer, consistently exhibit lower performance than a CSPH.

- **Gray-based Fourier vs. Fourier-based Gray:** Overall, the performance of these two coding schemes is very similar at $K > \log_2 N$ and across SBR and photon count levels. In the main paper we focus our discussion on Gray-based Fourier coding due to its marginally better performance at low SBR and its interpretability due to its use of Fourier functions. Nonetheless, Fourier-based Gray coding is an interesting coding scheme because at $K > \log_2 N$ it becomes fully binary, which may have a simple hardware implementation.

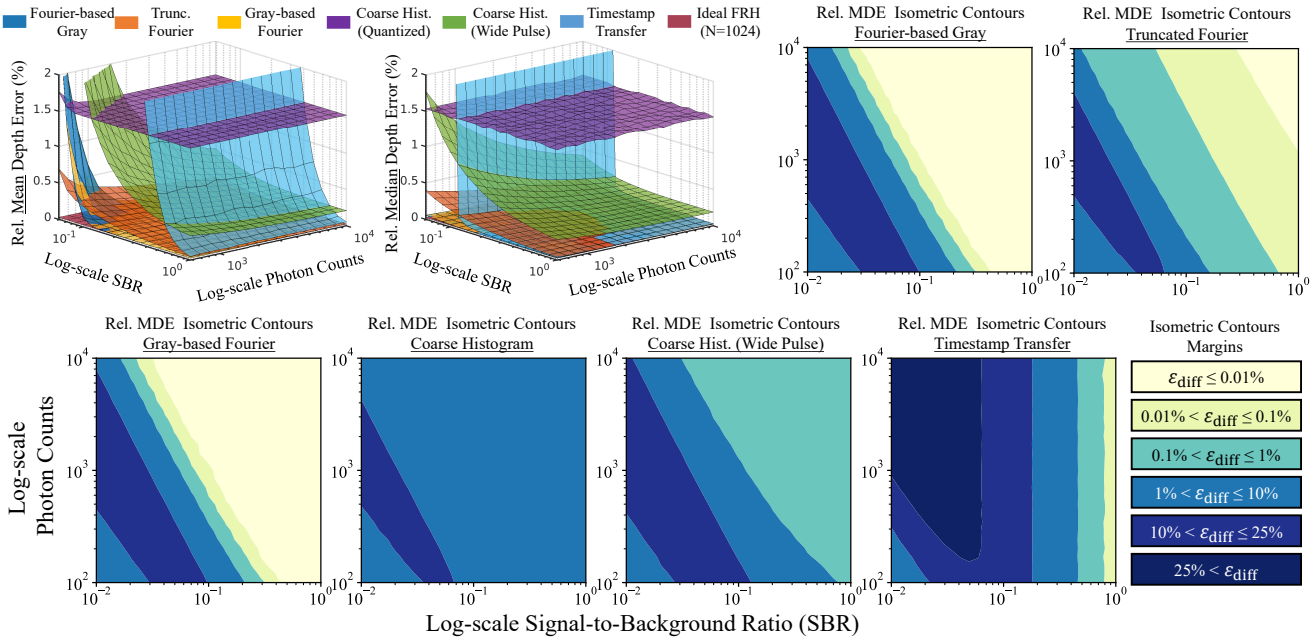
- **Rate of Performance Decay:** As we decrease SBR and the photon count levels (approach lower left corner of isometric countour plots), the performance of all methods begins to degrade at different rates. Gray-based Fourier coding is able to maintain low relative MDE up to a certain SBR and photon count level, and then its relative MDE quickly degrades faster than the ideal FRH, creating small isometric contours for $0.01\% < \varepsilon_{\text{diff}} < 1\%$. On the other hand, the performance of

Truncated Fourier coding progressively degrades, creating larger isometric contour regions for $0.01 < \varepsilon_{\text{diff}} < 1$, albeit at higher SBR and photon count levels. At $K = 16$ the progressive performance decay of Truncated Fourier, allows it to achieve isometric compression with lower $\varepsilon_{\text{diff}}$ for some SBR (< 0.1) and photon count levels (< 1000). Unfortunately, this advantage of Truncated Fourier coding is encountered around $\varepsilon_{\text{diff}} \approx 1\%$, which for a 10m depth range is an relative MDE difference of 10cm. Therefore, the better choice at those challenging SBR and photon count levels is to increase K for Gray-based Fourier Coding.

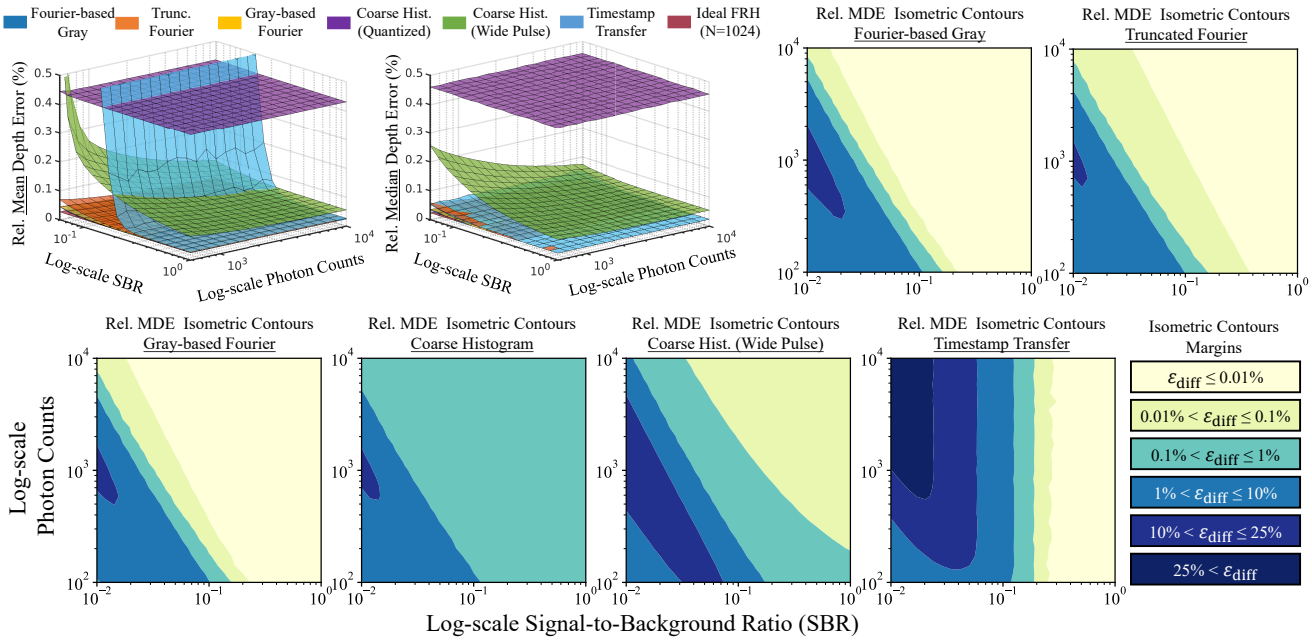
- **Mean vs. Median Depth Errors:** Comparing the relative mean and median errors surface plot trends allow us to learn useful things about each coding scheme. Coding approaches where the trends are similar produce depth errors of similar magnitudes. In the case of Truncated Fourier coding or a Coarse Hist (Wide Pulse), this means that they will make a lot of small errors, even at high SBR. On the other hand, approaches such as Timestamp transfer and Gray-based Fourier coding, can produce high relative mean depth errors while maintaining near-zero median depth errors. This is because, as we decrease SBR and photon count levels, these approaches either estimate the depth with near-zero error or produce a very high depth error (outliers).

In Appendix S. 8 we show additional results for compression ratios of 128, 32, and 8.

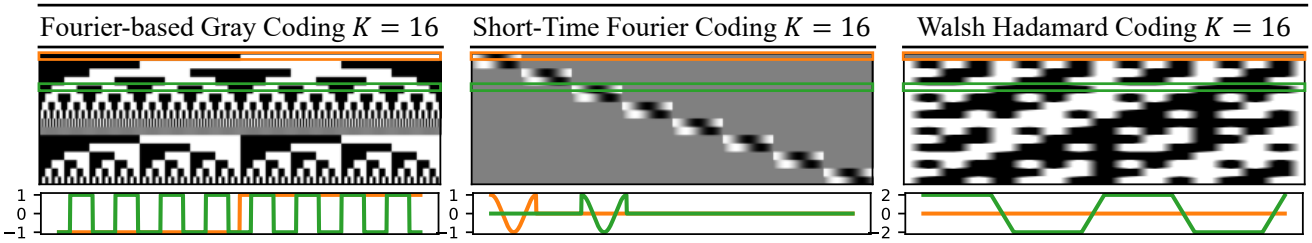
Summary: Our extensive Monte Carlo simulations and isometric compression analysis provide a guide for selecting the appropriate CSPH coding scheme when we have prior knowledge on SBR and photon count levels. Our results suggest that at $K < 32$, Gray-based Fourier coding is a good choice that can provide high compression without sacrificing depth precision in a wide range of scenarios. At higher K, Truncated Fourier becomes a competitive choice that can perform comparably to Gray-based Fourier. Overall, carefully designed coding matrices can significantly outperform a coarse histogram or transferring timestamps.



Supplementary Figure 3. **Isometric Compression Analysis at 64x Compression ($K = 16$) .** The surface plots show the relative mean and median depth errors computed as described in Sec. S. 3.1 for coding schemes with $K = 16$. The isometric compression contour plots are created by taking the element-wise absolute difference of each coding scheme relative MDE with the ideal FRH, and then classifying the difference into various margins ($\varepsilon_{\text{diff}}$). As we decrease SBR and photon count levels, the isometric compression margin achieved by each method increases. Finally, as we approach the lowest SBR and photon count levels, even the FRH starts failing too which makes $\varepsilon_{\text{diff}}$ smaller again. Although, the range of SBR and photon count levels for the isometric contour plots are 0.01-1 and 100-10000, the surface plots show SBR and photon count levels of 0.05-1 and 500-10000, for visualization purposes.



Supplementary Figure 4. **Isometric Compression Analysis at 16x Compression** ($K = 64$) **.** The surface plots show the relative mean and median depth errors computed as described in Sec. S. 3.1 for coding schemes with $K = 64$. The isometric compression contour plots are created by taking the element-wise absolute difference of each coding scheme relative MDE with the ideal FRH, and then classifying the difference into various margins ($\varepsilon_{\text{diff}}$). As we decrease SBR and photon count levels, the isometric compression margin achieved by each method increases. Finally, as we approach the lowest SBR and photon count levels, even the FRH starts failing too which makes $\varepsilon_{\text{diff}}$ smaller again. Although, the range of SBR and photon count levels for the isometric contour plots are 0.01-1 and 100-10000, the surface plots show SBR and photon count levels of 0.05-1 and 500-10000, for visualization purposes.



Supplementary Figure 5. **Additional Coding Matrices.** In addition to the coding schemes presented in the main paper we also evaluate these coding schemes. Fourier-based Gray coding is evaluated in Sec. S. 3.2. Short-time Fourier and Walsh Hadamard coding are evaluated in Sec. S. 3.3.

S. 3.3. Comparisons with Additional Coding Schemes

In this section we evaluate two additional CSPH coding schemes and compare them with the coding schemes presented in the main paper:

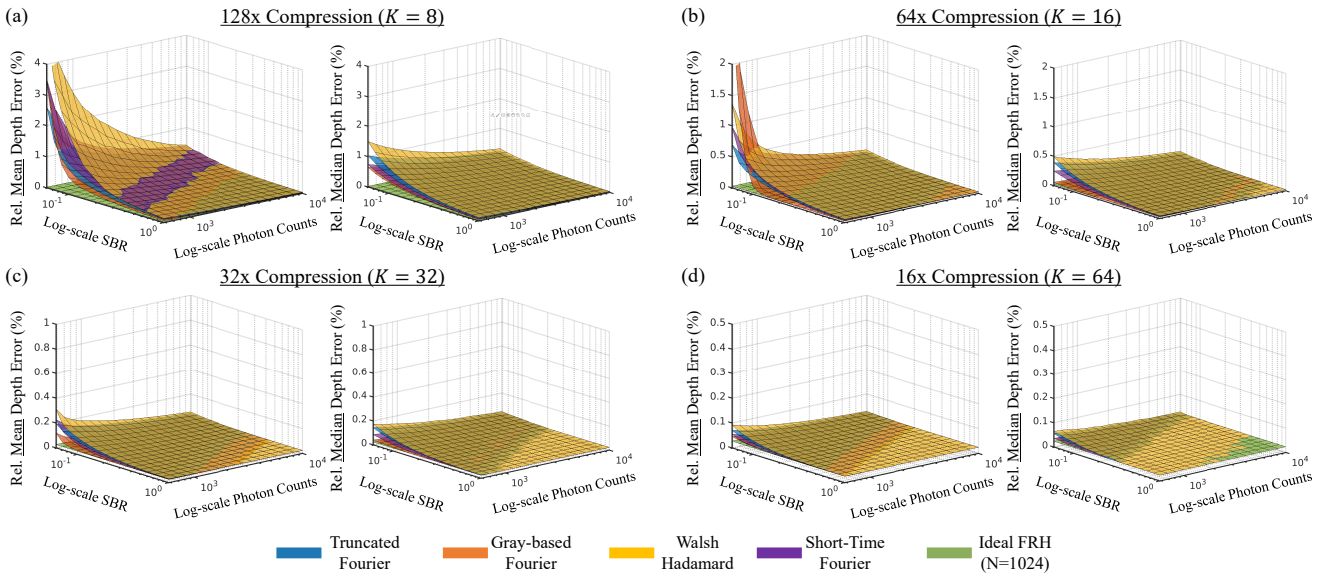
- **Short-time Fourier:** The coding matrix is similar as the coarse histogram, but instead of setting the rectangular window at each row to all 1's, we set the values inside the window to a single period of a sinusoidal function. As illustrated in Suppl. Fig. 5, the odd and even rows use a cosine and sine function respectively. We refer to this method a short-time Fourier, due to its similarity to a short-time Fourier transform.

- **Walsh Hadamard:** To generate C, we first generate a $K \times K$ Hadamard matrix using scipy's Hadamard function [64], transpose it, and then linearly interpolate each row to make it length N.

To evaluate the coding schemes we use the relative mean and median depth errors metrics described in Sec. S. 3.1.

Main Observations: Suppl. Fig. 6 compares the performance of Short-time Fourier and Walsh Hadamard coding with the coding schemes used in the main paper, at multiple compression ratios. Walsh Hadamard coding consistently achieves the lowest performance. Short-time Fourier, achieves lower relative median depth errors than Truncated Fourier but higher relative median depth errors, while having similar performance trends. Gray-based Fourier coding achieves outperforms all coding schemes at most SBR and photon count levels.

Summary: Although, the new coding schemes evaluated in this section were not able to outperform Gray-based Fourier coding for the most part, their performance was competitive and well above a coarse histogram method.



Supplementary Figure 6. **Performance of Additional Coding Schemes.** Relative mean and median depth errors at different compression ratios. As we decrease compression, all coding schemes quickly approach the performance of the ideal FRH (note the z-axis is different at each compression ratio), and the performance difference across coding schemes becomes small. However, at high compression ratios, the choice of coding scheme is crucial at most SBR and photon count levels.

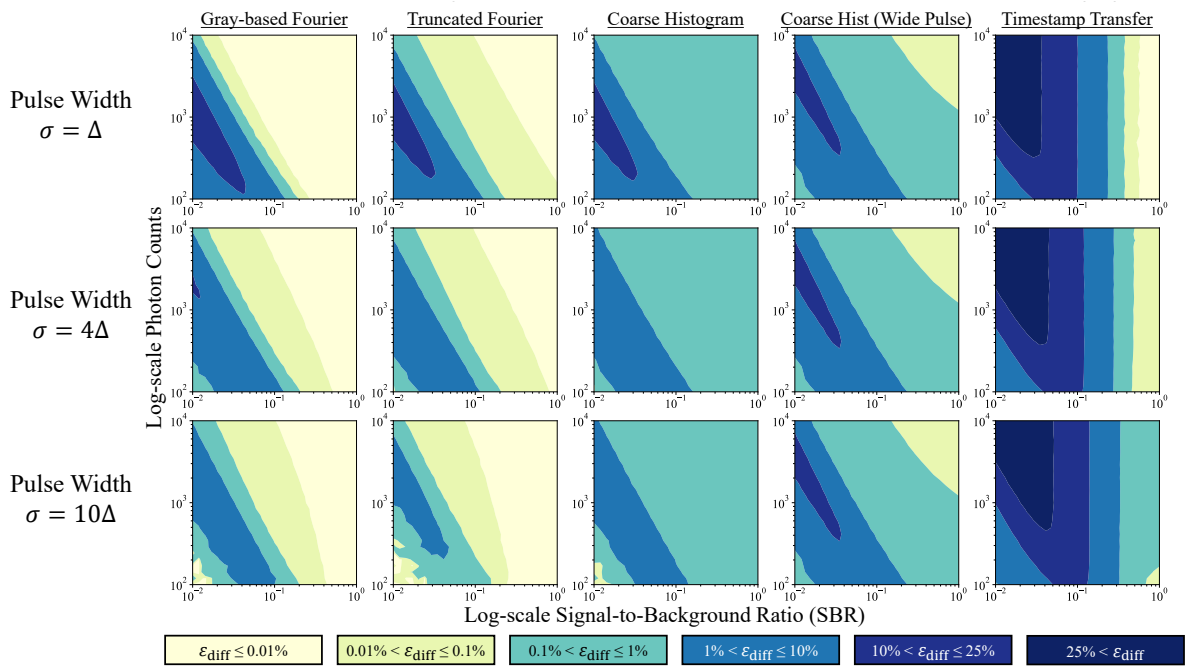
S. 3.4. Isometric Compression with Wide Gaussian Pulses

In this section we analyze how isometric compression trends change as we increase the system IRF pulse width.

Suppl. Fig. 7 shows the isometric compression contour plots for multiple coding schemes with system IRFs set to Gaussian pulses with varying widths. The relative MDE isometric contour plots are generated using similar MDE Monte Carlo simulations described in Sec. S. 3.1.

Main Observations: As we increase the pulse width used, the isometric contours with low $\varepsilon_{\text{diff}} \leq 1\%$ increase for Truncated Fourier, Gray-based Fourier, and coarse histograms. In particular, the overall performance difference between Truncated Fourier and the ideal FRH shrinks. This is because increasing the pulse width, decreases the bandwidth of the photon flux waveform, which makes Truncated Fourier more efficient. For medium SBR and photon count levels, the performance of Gray-based Fourier coding decreases, as the pulse width increases, because some of the high-frequency coding functions get zero-ed out (band-limit property). Moreover, the performance of an FRH with a wider pulse also decreases at a faster rate as we decrease SBR and photon counts. Therefore, at the lowest SBR and photon count levels, the FRH performs as poorly as the CSPH methods, leading to isometric compression with small $\varepsilon_{\text{diff}}$ again. Finally, the timestamp transfer baseline is one of the methods that is impacted the most as the pulse width increases. In Suppl. Sec. S. 9 we show additional contour plots for compression ratios of 64 and 16.

Summary: Truncated Fourier coding is the better CSPH for a SPAD-based 3D camera that has a pulse width significantly wider than the SPAD's time resolution (Δ). At narrower pulse widths, Gray-based Fourier coding performs better because these narrow pulses have useful high-frequency content that Gray-based Fourier coding samples. Overall, the system IRF and consequently the laser pulse waveform, play an important role in the choice of the CSPH coding scheme.



Supplementary Figure 7. **Wide Pulse Isometric Compression at 32x Compression ($K = 32$).** The Gaussian pulses used as $h(t)$ are proportional to: $\exp -\frac{t^2}{\Delta}$ (1st row), $\exp -\frac{t^2}{4\Delta}$ (2nd row), $\exp -\frac{t^2}{10\Delta}$ (3rd row). The jagged edges observed in the isometric contours at low SBR and low photon counts are caused by the variance of the Monte Carlo simulations which adds small variance to the relative MDE of all methods.

S. 4. Low Photon Count Regime

In this section we analyze the performance of different coding schemes in the low photon count regimes where less than 100 photons are detected per pixel, on average. In practice, this is important for scenarios such as low scene albedo, long standoff distances, extremely short acquisition times, or in case of constrained laser power budget.

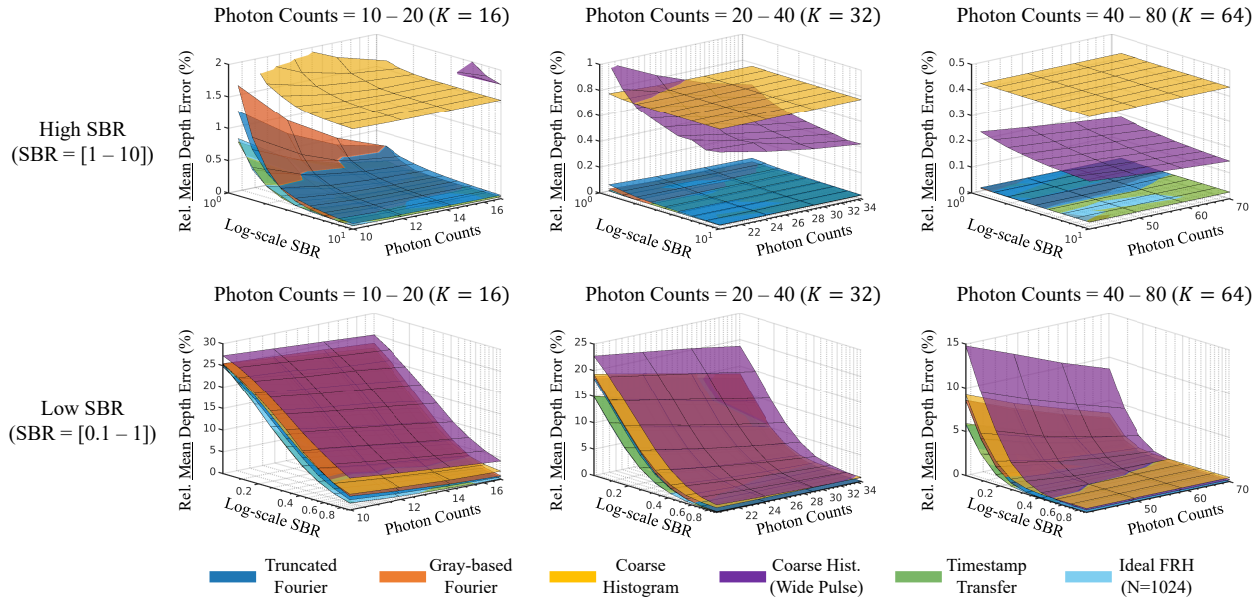
Suppl. Fig. 8 shows the mean depth errors for different coding schemes as a function of SBR and total number of photons. The total number of photons includes both signal (laser) and background (ambient) photons. The relative MDE Monte Carlo simulations were performed with the same parameters as in Sec. S. 3.2. The number of coding functions used for each photon count range is chosen to roughly match the expected number of photons (e.g., for a photon count range of 10-20, we use 16 coding functions). Note that in some cases the compressive algorithm may require *more* data bandwidth than timestamp transfer (*e.g.* when there are only 10 photons captured and the compressive method uses $K = 16$ Fourier coefficients.) Moreover, in the low count regime, timestamp transfer is nearly equivalent to the FRH that transfers all time bins.

Main Observations:

- In the high SBR regime the proposed Fourier-coding based methods provide some benefits over näive coarse histogramming approaches. When the number of photons captured is > 20, compressive methods perform almost as well as FRH and timestamp transfer and provide almost perfect depth reconstruction. The extent of compression achieved *vis a vis* transferring raw timestamps is limited due to the low number of photons captured in each pixel.

- In the low SBR regime most of the photons captured by the pixel are ambient photons, hence, the overall relative MDE remains high. In this regime, neither compressive approaches nor the baseline approaches (coarse histogramming, FRH and timestamp transfer) provide usable depth maps, with relative MDEs as high as 25% due to Poisson noise. In such scenarios it may become necessary to either increase the laser power, increase the total acquisition time, or resort to other denoising approaches (*e.g.* spatio-temporal filtering) [35, 50, 58, 60] that can complement the compressive capture strategies discussed in this paper.

Summary: In the low photon count regime it is necessary to combine the compressive acquisition methods described in this paper with denoising approaches that operate on raw timestamp data or on compressed histogram measurements.



Supplementary Figure 8. **Photon-Starved Mean Depth Error Results.** (Top Row) In the high SBR regime, compressive approaches provide better relative MDE as compared to coarse histogramming or raw timestamp transfer. (Bottom Row) In the low SBR regime compressive acquisition alone is not sufficient to achieve low MDE and other spatio-temporal denoising methods may be necessary to combat Poisson noise in these photon-limited regimes.

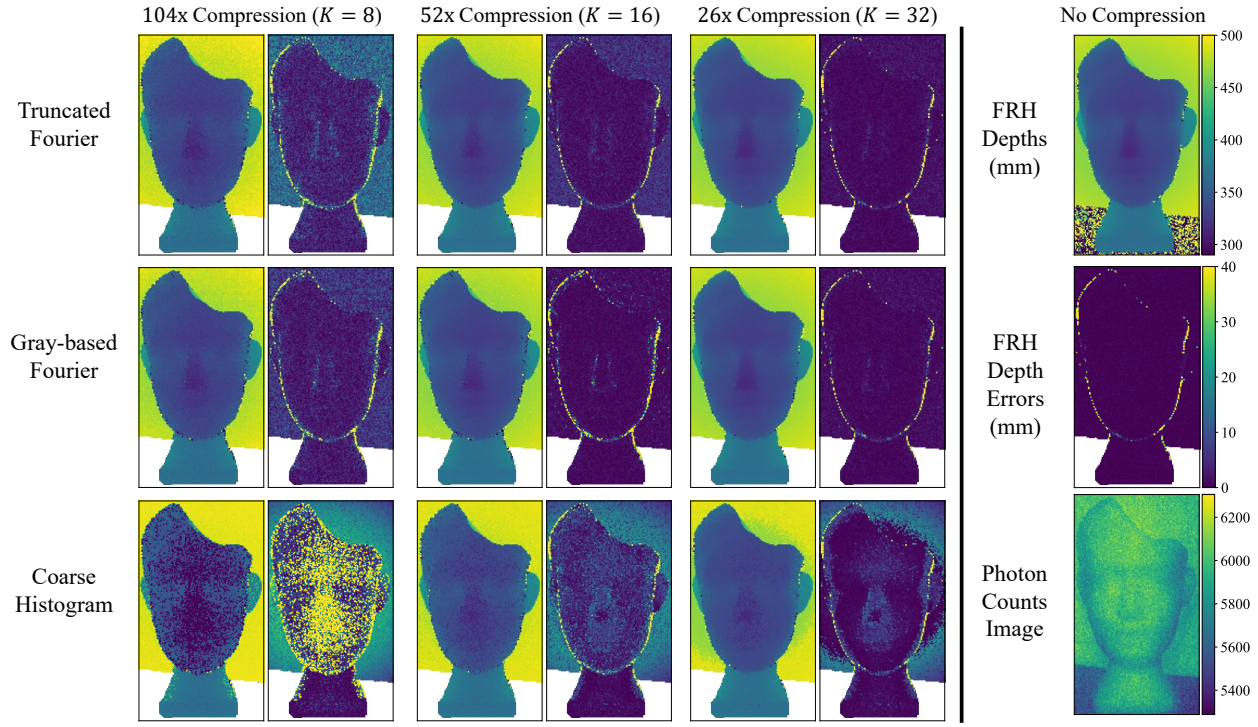
S. 5. Additional Real-world Scanning-based System Results

In this section we present additional results from the real-data acquired with the scanning-based system from [15].

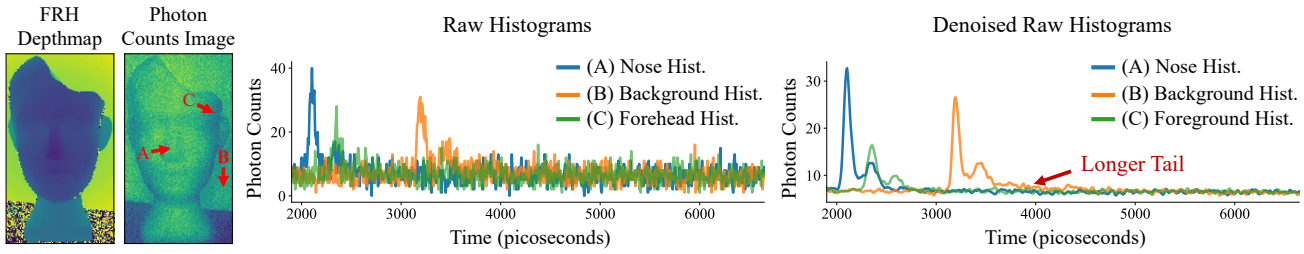
Suppl. Fig. 9 and 12 show the two scans that we downloaded and pre-processed from [15]. The pre-processed per-pixel full-res histograms have $\Delta = 8$ps and $N = 832$. To obtain ground truth depths we applied Gaussian denoising ($\sigma = 0.75$) to the full-resolution histogram images of each scene, and computed depths using matched filtering. As observed in the FRH depth images in Suppl. Fig. 9 and 12, there are some points whose SNR is too low to obtain a reliable depth even after denoising, therefore we manually mask those points when visualizing the depths and depth errors (white regions). The system IRF, $h(t)$, is extracted from on of the points near the center of the denoised histogram, and then is further pre-processed to remove most background photons and center it around $t = 0$.

Main Observations: As observed in Suppl. Fig. 9 and 12, at $K = 8$, all CSPH methods make systematic depth errors in the background points. These errors are caused by the longer tail observed in the background points histogram shown in Suppl. Fig. 10. At $K = 16$, Gray-based Fourier coding contains sufficient coding functions with high-frequency content that mitigate the effects of the longer tail and estimate reliable depths in the background points. Finally, at $K = 32$ Gray-based Fourier and Truncated Fourier are able to essentially match the performance of the FRH. Coarse histograms can achieve sub-bin precision because $h(t)$ spreads over multiple coarse time bins. Nonetheless, the depth precision is lower, and the systematic errors in the background points continue to affect it, even at $K = 32$.

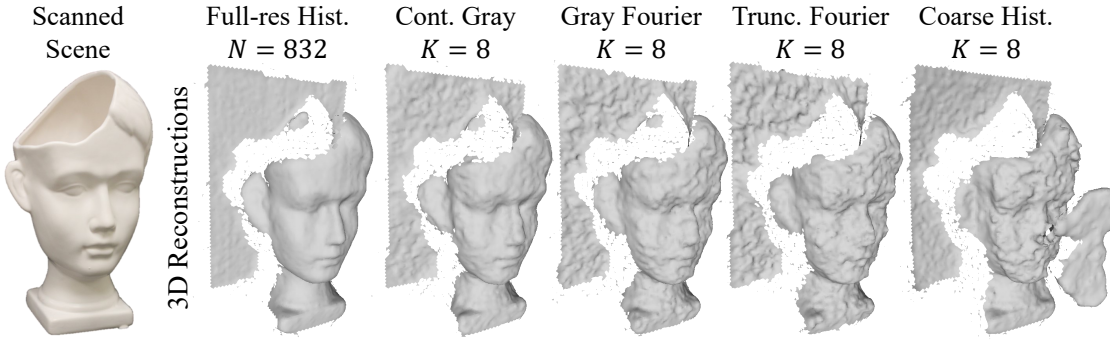
Summary: Despite the non-idealities encountered in real data (long-tailed $h(t)$ and indirect reflections), CSPH coding approaches, such as Truncated Fourier and Gray-based Fourier, can provide significant compression without sacrificing depth precision.



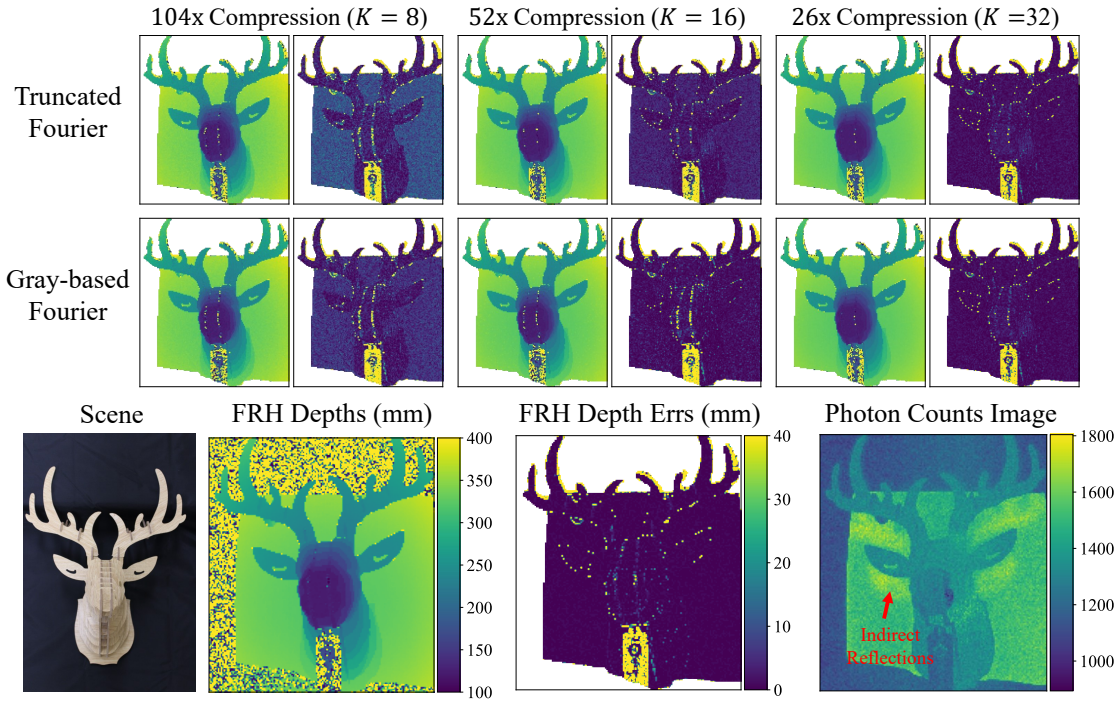
Supplementary Figure 9. **Scan Data Results at Different Compression Ratios.** As we decrease the compression (increase K) the errors for all coding schemes decrease. Truncated Fourier at $K \leq 16$ is susceptible to the systematic errors caused by the longer tail that appears in the background histograms (Suppl. Fig. 10). At $K = 32$ both Gray-based Fourier and Truncated Fourier essentially match the performance of using the full-res histogram. Comparatively, coarse histogram have low depth precision and are quite susceptible to the longer tail of the background points histogram, even at $K = 32$.



Supplementary Figure 10. **Denoised Raw Histograms.** We apply a 3D Gaussian filter to the histogram image of the face scan and visualize histograms at different scene points. Note that the histogram extracted from the background point (orange line) exhibits a longer tail than the histograms obtained from points on the face. This longer tail is likely due to indirect reflections.



Supplementary Figure 11. **3D Reconstruction Visualizations.** To generate the 3D reconstructions in this paper we apply a 3×3 median filter to the recovered depth images, generate point clouds, and use MeshLab to estimate normals and perform Poisson surface reconstruction [7].



Supplementary Figure 12. **Scan Data Results for Deer Scene.** The first and second row show the depth and depth error image pairs obtained from each CSPH coding method. As we decrease the compression (increase K) the errors for all coding schemes decrease. Truncated Fourier at $K = 8$ is susceptible to the systematic errors caused by the longer tail that appears in the background histograms. Increasing K resolved these errors are mitigated by using coding functions with higher frequencies.

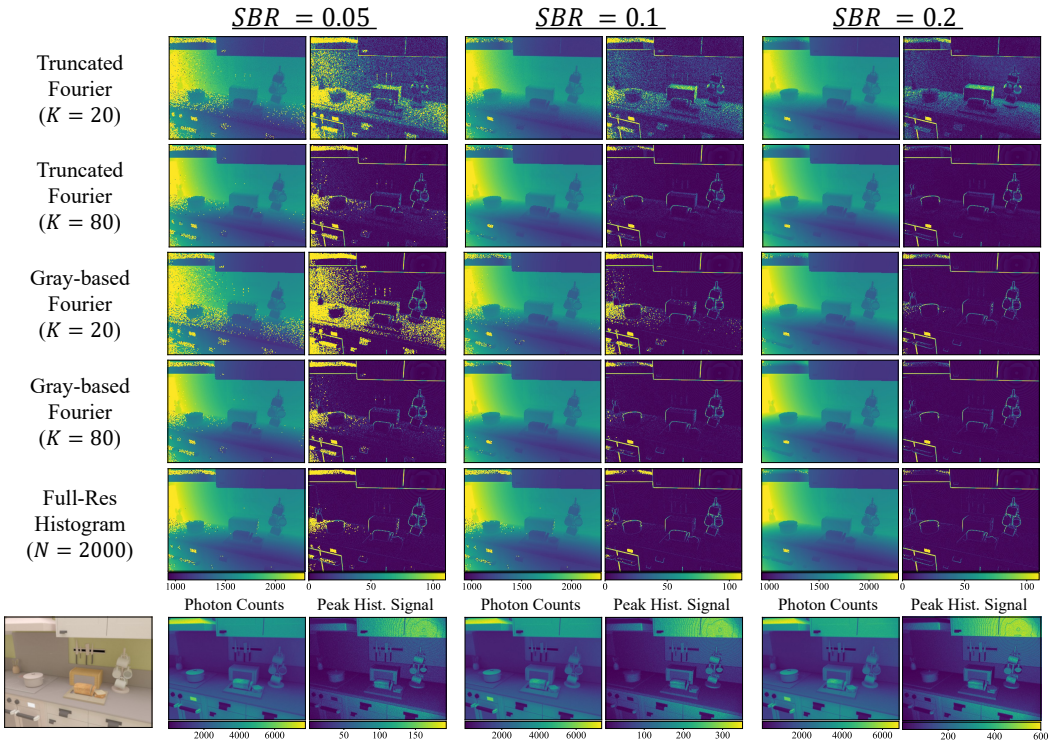
S. 6. Additional Simulated Flash Illumination System Results

In this section we evaluate different CSPH strategies applied on a flash illumination scenario at varying SBR levels. The flash illumination system data is simulated from physically accurate histogram images rendered with MitsubaToF [43]. The scenes input to MitsubaToF are obtained from [21]. The histogram image parameters were $\Delta = 50$ps and $N = 2000$ time bins. As described in the main paper, we use the R channel of the corresponding RGB image of the scene as an approximation of the per-pixel background photons. Then to simulate the FRHs we set the mean SBR and mean photon counts of the overall scene, and use those input values to scale and the histogram image and vertically shift it (add background). Similar to the relative MDE simulation results, we use an $h(t) \propto \exp -\frac{t^2}{\Delta}$ which is convolved with the rendered ground truth per-pixel histograms.

Main Observations:

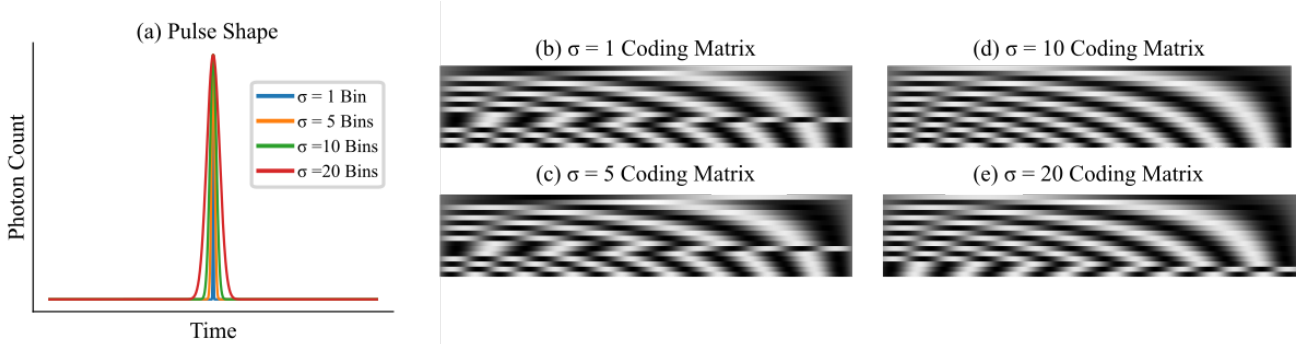
- **Indirect Reflections Errors:** At $K = 20$ Truncated Fourier is susceptible to systematic depth errors caused by diffuse indirect reflections. For Truncated Fourier coding, the only way to overcome these errors is to increase K and use higher frequencies. Gray-based Fourier coding can mitigate these errors with only 20 coding functions.
- **Outlier Depth Estimates:** At $K = 20$, in low SBR regions, Gray-based Fourier coding produces depth estimates with either near-zero or very high (outlier) errors. This is consistent with the observations made in Sec. S. 3.2, where the mean depth errors were significantly higher than the median (due to outliers). Increasing K helps Gray-based Fourier coding mitigate these outliers and perform comparably to a FRH.

Summary: Truncated Fourier coding at low K is susceptible to errors due to indirect reflection, regardless of SBR or photon count levels. This means that a Truncated Fourier CSPH on a flash illumination system may require higher K (lower compression) to avoid trading-off depth precision. Gray-based Fourier coding, on the other hand can use lower K (high compression) and still achieve high depth precision, as long as the SBR level is sufficiently high. Overall, having prior knowledge on the SBR and photon count levels can help select the correct K to use in a CSPH.



Supplementary Figure 13. **Flash Illumination Results at Multiple SBR.** We simulate the same scene with different mean SBR levels, and fixed mean photon counts to 2000. A mean scene SBR of 0.05 means that if we look at the per-pixel SBR and average them, we will obtain approximately 0.05. As observed in the bottom row, as we increase the SBR, the photon count image remains unchanged (because mean scene photon counts is fixed at 2000), but the peak of the per-pixel histogram increases.

S. 7. PCA Codes



Supplementary Figure 14. **PCA Codes at Different Pulse Widths.** In (a) two Gaussian pulses at different pulse widths (σ) are shown. (b-e) shows PCA coding matrix ($K = 16$) for $\sigma = 1, 5, 10, 20$. All four pulse widths yield similar PCA codes that look like Fourier components. Note that some PCA codes in (b,c,e) are multiplied by negative one when compared to (d) or the Fourier codes.

This section will describe how we found the principal component analysis codes used in the main paper and why they relate to Fourier codes. To numerically find the PCA codes, we generate many Gaussian LiDAR pulses at different signal-to-background ratios and with different shifts representing the binned return signal at the camera sensor. Let $(\Phi_{i,s,b})_{i=0}^{N-1}$ be the binned photon flux waveform from Eq. 2 in the main text when $\Phi^{\text{sig}}(t)$ is a Gaussian pulse centered at time s and $\Phi^{\text{bkg}} = b$. Therefore, $(\Phi_{i,s,b})_{i=0}^{N-1}$ is the binned return signal from a Gaussian laser pulse, from an object at distance $\frac{cs}{2}$, and signal-to-background ratio of $\frac{1}{b}$.

In order to run PCA we generate $(\Phi_{i,s,b})_{i=0}^{N-1}$ for many s and b values and stack them into a large matrix. Specifically, we use 1000 s values in increments of $\Delta_s = \frac{\tau}{1000}$ from 0 to τ, that is $s \in \{0, \Delta_s, 2\Delta_s, \dots \tau\}$. At each s value we use 50 b values with log spacing with increments of $\Delta_b = \frac{2}{50}$ from 10^{-2} to 1, that is $b \in \{10^{-2}, 10^{-2+\Delta_b}, 10^{-2+2\Delta_b}, \dots 1\}$. The final matrix, M, is given by,

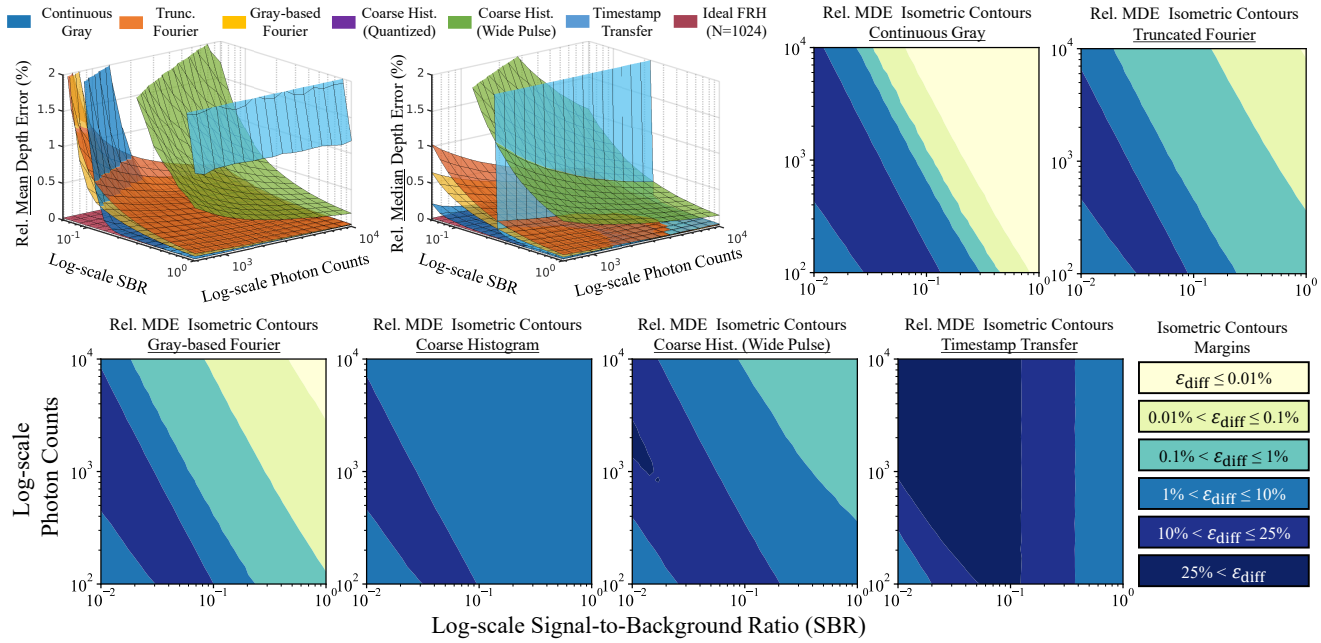
$$M(b) = \frac{1}{1+b} \begin{bmatrix} \Phi_{0,0,b} & \Phi_{1,0,b} & \dots & \Phi_{N-1,0,b} \\ \Phi_{0,\Delta_s,b} & \Phi_{1,\Delta_s,b} & \dots & \Phi_{N-1,\Delta_s,b} \\ \Phi_{0,2\Delta_s,b} & \Phi_{1,2\Delta_s,b} & \dots & \Phi_{N-1,2\Delta_s,b} \\ \vdots & \vdots & & \vdots \\ \Phi_{0,\tau,b} & \Phi_{1,\tau,b} & \dots & \Phi_{N-1,\tau,b} \end{bmatrix} \tag{11}$$

$$M = \begin{bmatrix} M(10^{-2}) \\ M(10^{-2+\Delta_b}) \\ \vdots \\ M(1) \end{bmatrix} \tag{12}$$

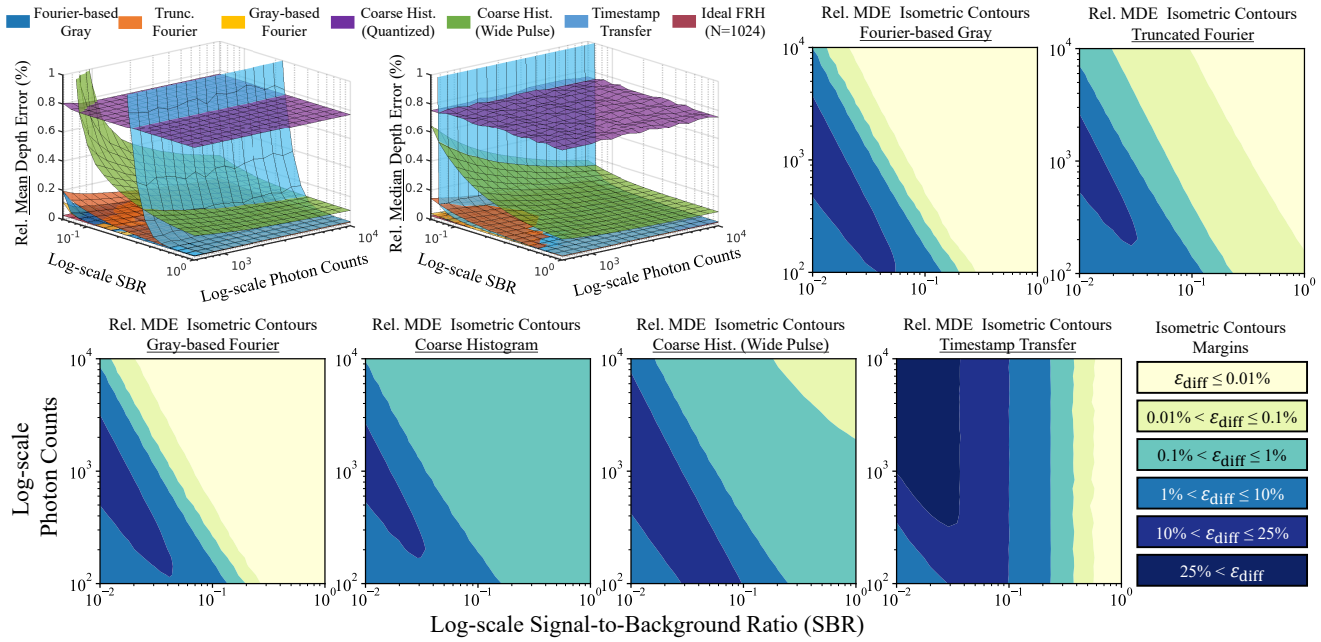
Where the factor of $\frac{1}{1+b}$ ensures that the rows of $M(b)$ sum to 1. We then use PCA on M and use the found PCA components as the PCA codes. Some of our found PCA codes for four different pulse widths are shown in Suppl. Fig. 14. Interestingly, the resulting PCA codes where very similar to Fourier codes.

To understand why PCA codes are similar to Fourier codes consider a single b value and set Δ_s to the length of a single time bin, then $M(b)$ will be a circulant matrix. It turns out that eigenvectors and thus PCA components of circulant matrices are given by Fourier components [62]. To understand why recall that circulant matrices implement convolutions. Convolutions are simply multiplications in the Fourier domain which implies that the eigenvectors of a convolution and thus circulant matrices are the Fourier components. Although, we do not use Δ_s equal to the length of a time bin, the $M(b)$'s we use are close to a circulant matrix so our resulting PCA components are very close to Fourier components.

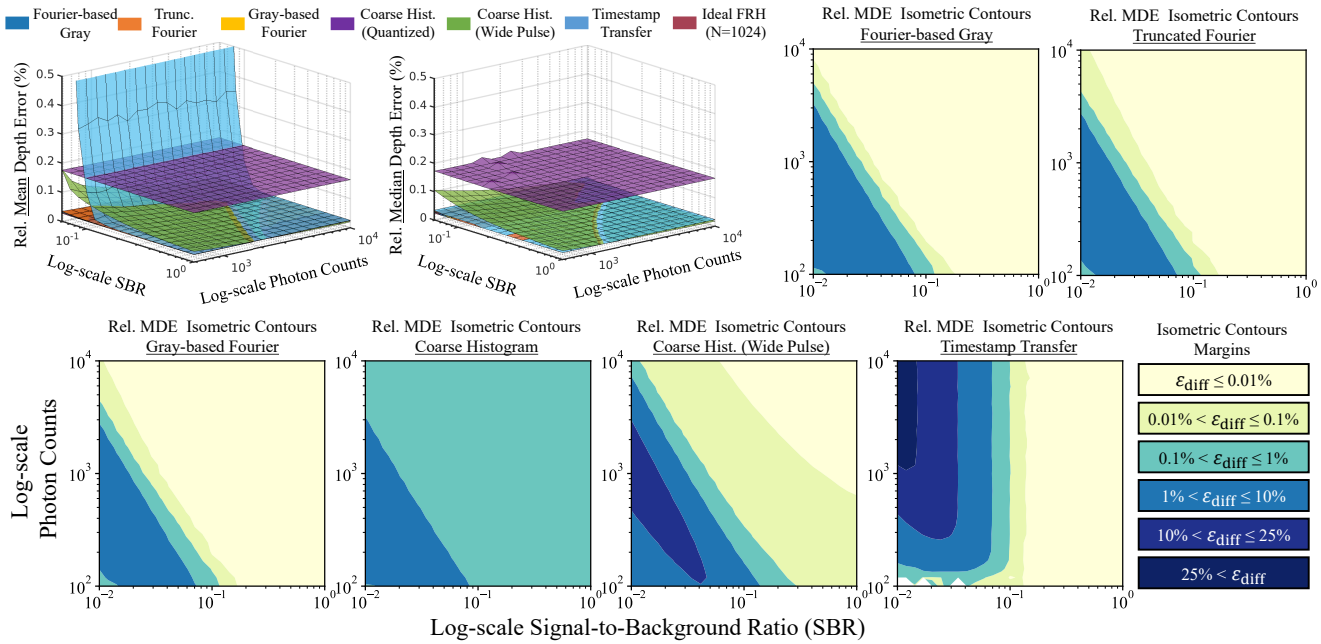
S. 8. Appendix A: Additional Isometric Compression Plots



Supplementary Figure 15. **Isometric Compression Analysis at 128x Compression** ($K = 8$) **.**

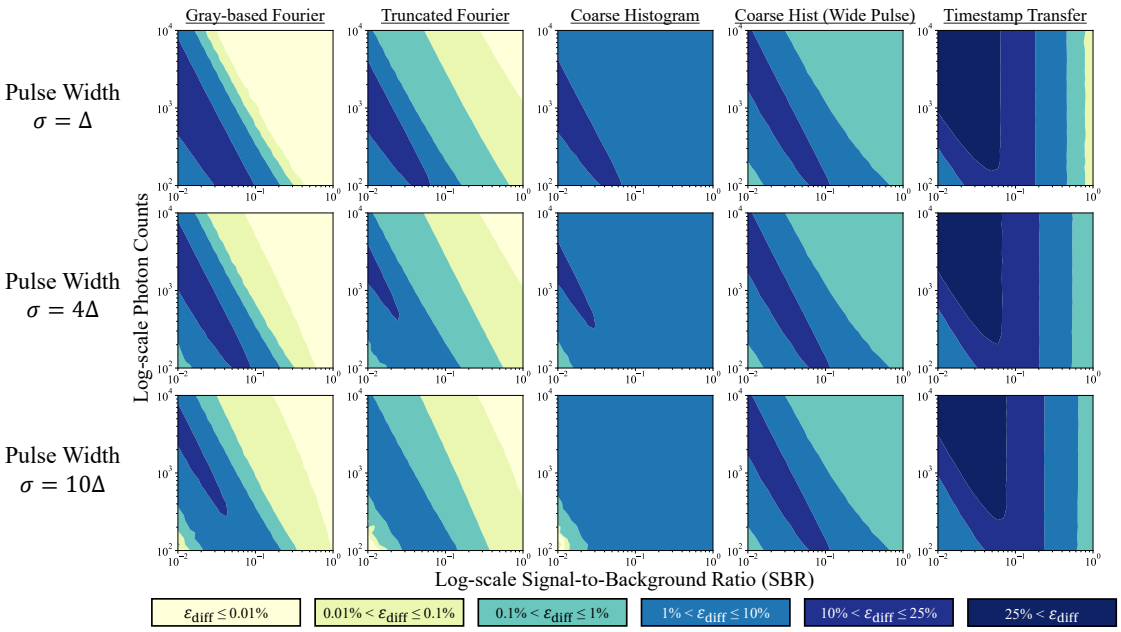


Supplementary Figure 16. **Isometric Compression Analysis at 64x Compression** ($K = 32$) **.**

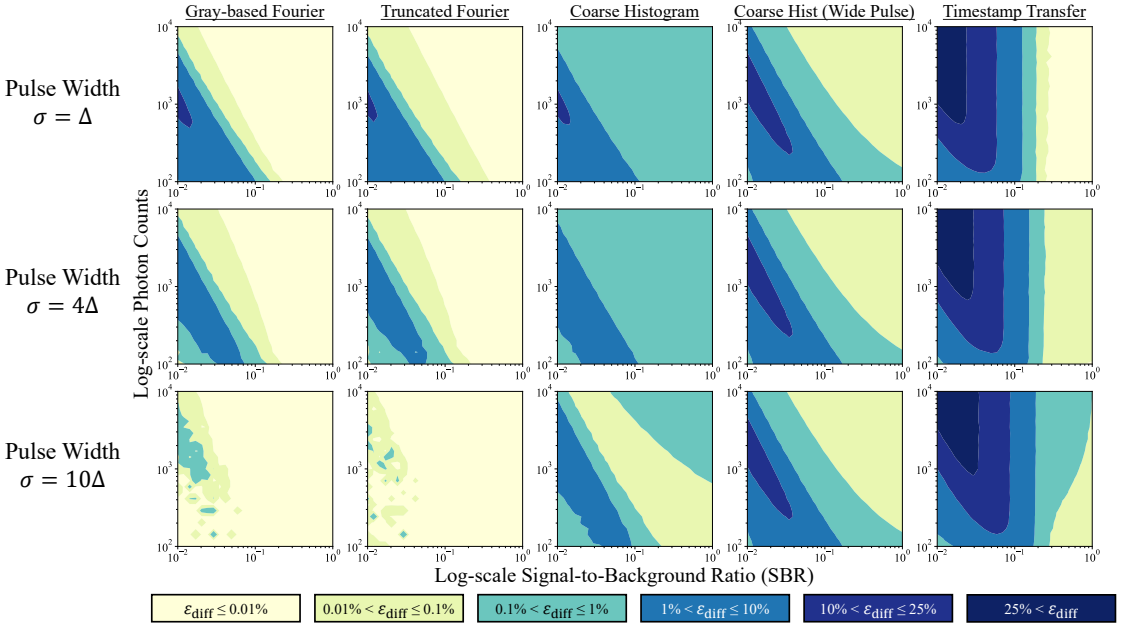


Supplementary Figure 17. **Isometric Compression Analysis at 8x Compression** ($K = 128$) **.**

S. 9. Appendix B: Additional Isometric Compression Results with Wide Pulses



Supplementary Figure 18. **Wide Pulse Isometric Compression at 64x Compression ($K = 16$).** The Gaussian pulses used as $h(t)$ are proportional to: $\exp -\frac{t^2}{\Delta}$ (1st row), $\exp -\frac{t^2}{4\Delta}$ (2nd row), $\exp -\frac{t^2}{10\Delta}$ (3rd row).



Supplementary Figure 19. **Wide Pulse Isometric Compression at 16x Compression ($K = 64$).** For $\sigma = 10$ the coarse histogram method follows a different trend from what is seen before. This is because at high SBR and high photon counts, the coarse histogram is quantization limited, while the FRH can achieves low depth errors. As SBR and photon counts decrease the FRH performance decreases, while the coarse histogram continues to be quantization limited. As SBR and photon counts continue to decrease the performance of the coarse histogram starts degrading more rapidly than the FRH. And finally, at the lowest SBR and photon count levels both methods are achieving very high depth errors (almost estimating depths at random), making their $\varepsilon_{\text{diff}}$ small again.

References

[1] Supreeth Achar, Joseph R Bartels, William L Whittaker, Kiriakos N Kutulakos, and Srinivasa G Narasimhan. Epipolar time-of-flight imaging. *ACM Transactions on Graphics (TOG)*, 36(4):37, 2017. 2

[2] Edward H Adelson, Charles H Anderson, James R Bergen, Peter J Burt, and Joan M Ogden. Pyramid methods in image processing. *RCA engineer*, 29(6):33–41, 1984. 8

[3] Yoann Altmann, Stephen McLaughlin, Miles J Padgett, Vivek K Goyal, Alfred O Hero, and Daniele Faccio. Quantum-inspired computational imaging. *Science*, 361(6403), 2018. 1

[4] Alexander W Bergman, David B Lindell, and Gordon Wetzstein. Deep adaptive lidar: End-to-end optimization of sampling and depth completion at low sampling rates. In *2020 IEEE International Conference on Computational Photography (ICCP)*, pages 1–11. IEEE, 2020. 2

[5] Mauro Buttafava, Jessica Zeman, Alberto Tosi, Kevin Eliceiri, and Andreas Velten. Non-line-of-sight imaging using a time-gated single photon avalanche diode. *Optics express*, 23(16):20997–21011, 2015. 1

[6] 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*, pages 5970–5980, 2020. 2, 3

[7] Paolo Cignoni, Marco Callieri, Massimiliano Corsini, Matteo Dellepiane, Fabio Ganovelli, Guido Ranzuglia, et al. Meshlab: an open-source mesh processing tool. In *Eurographics Italian chapter conference*, volume 2008, pages 129–136. Salerno, Italy, 2008. 11

[8] Sergio Cova, Massimo Ghioni, Andrea Lacaita, Carlo Samori, and Franco Zappa. Avalanche photodiodes and quenching circuits for single-photon detection. *Applied optics*, 35(12):1956–1976, 1996. 1

[9] Francesco Mattioli Della Rocca, Hanning Mai, Sam W Hutchings, Tarek Al Abbas, Kasper Buckbee, Andreas Tsiamis, Peter Lomax, Istvan Gyongy, Neale AW Dutton, and Robert K Henderson. A 128× 128 spad motion-triggered time-of-flight image sensor with in-pixel histogram and column-parallel vision processor. *IEEE Journal of Solid-State Circuits*, 55(7):1762–1775, 2020. 2, 4, 8

[10] Eric R Fossum, Jiaju Ma, Saleh Masoodian, Leo Anzagira, and Rachel Zizza. The quanta image sensor: Every photon counts. *Sensors*, 16(8):1260, 2016. 1

[11] John P Godbaz, Michael J Cree, and Adrian A Dorrington. Closed-form inverses for the mixed pixel/multipath interference problem in amcw lidar. In *Computational Imaging X*, volume 8296, page 829618. International Society for Optics and Photonics, 2012. 8

[12] Craig Gotsman and Michael Lindenbaum. On the metric properties of discrete space-filling curves. *IEEE Transactions on Image Processing*, 5(5):794–797, 1996. 5

[13] Francesco Gramuglia, Ming-Lo Wu, Claudio Bruschini, Myung-Jae Lee, and Edoardo Charbon. A low-noise cmos spad pixel with 12.1 ps sptr and 3 ns dead time. *IEEE Journal of Selected Topics in Quantum Electronics*, 2021. 1

[14] Frank Gray. Pulse code communication. *United States Patent Number 2632058*, 1953. 5

[15] Anant Gupta, Atul Ingle, and Mohit Gupta. Asynchronous single-photon 3d imaging. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 7909–7918, 2019. 1, 3, 6, 10

[16] Anant Gupta, Atul Ingle, Andreas Velten, and Mohit Gupta. Photon-flooded single-photon 3d cameras. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6770–6779, 2019. 3

[17] Mohit Gupta, Amit Agrawal, Ashok Veeraraghavan, and Srinivasa G Narasimhan. Structured light 3d scanning in the presence of global illumination. In *CVPR 2011*, pages 713–720. IEEE, 2011. 2

[18] Mohit Gupta and Nikhil Nakhate. A geometric perspective on structured light coding. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 87–102, 2018. 2, 4

[19] Mohit Gupta, Shree K Nayar, Matthias B Hullin, and Jaime Martin. Phasor imaging: A generalization of correlation-based time-of-flight imaging. *ACM Transactions on Graphics (ToG)*, 34(5):156, 2015. 2, 4

[20] Mohit Gupta, Andreas Velten, Shree K Nayar, and Eric Breitbach. What are optimal coding functions for time-of-flight imaging? *ACM Transactions on Graphics (TOG)*, 37(2):13, 2018. 2, 4, 1

[21] Felipe Gutierrez-Barragan, Huaijin Chen, Mohit Gupta, Andreas Velten, and Jinwei Gu. itof2dtof: A robust and flexible representation for data-driven time-of-flight imaging. *arXiv preprint arXiv:2103.07087*, 2021. 7, 3, 12

[22] Felipe Gutierrez-Barragan, Syed Azer Reza, Andreas Velten, and Mohit Gupta. Practical coding function design for time-of-flight imaging. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1566–1574, 2019. 2, 4, 5, 3

[23] Istvan Gyongy, Sam W Hutchings, Abderrahim Halimi, Max Tyler, Susan Chan, Feng Zhu, Stephen McLaughlin, Robert K Henderson, and Jonathan Leach. High-speed 3d sensing via hybrid-mode imaging and guided upsampling. *Optica*, 7(10):1253–1260, 2020. 2, 3, 5, 7

[24] Nathan Hagen, Matthew Kupinski, and Eustace L Dereniak. Gaussian profile estimation in one dimension. *Applied optics*, 46(22):5374–5383, 2007. 5

[25] Abderrahim Halimi, Philippe Ciuciu, Aongus Mccarthy, Stephen Mclaughlin, and Gerald S Buller. Fast adaptive scene sampling for single-photon 3d lidar images. In *2019 IEEE 8th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pages 196–200. IEEE, 2019. 2

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

[27] Felix Heide, Matthias B Hullin, James Gregson, and Wolfgang Heidrich. Low-budget transient imaging using photonic mixer devices. *ACM Transactions on Graphics (ToG)*, 32(4):45, 2013. 8

[28] Robert K Henderson, Nick Johnston, Sam W Hutchings, Istvan Gyongy, Tarek Al Abbas, Neale Dutton, Max Tyler, Susan Chan, and Jonathan Leach. 5.7 a 256× 256 40nm/90nm cmos 3d-stacked 120db dynamic-range reconfigurable time-resolved spad imager. In *2019 IEEE International Solid-State Circuits Conference-(ISSCC)*, pages 106–108. IEEE, 2019. 2, 4, 8

[29] Sam W Hutchings, Nick Johnston, Istvan Gyongy, Tarek Al Abbas, Neale AW Dutton, Max Tyler, Susan Chan, Jonathan Leach, and Robert K Henderson. A reconfigurable 3-d-stacked spad imager with in-pixel histogramming for flash lidar or high-speed time-of-flight imaging. *IEEE Journal of Solid-State Circuits*, 54(11):2947–2956, 2019. 2, 4, 8

[30] 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*, pages 8585–8595, 2021. 1

[31] Atul Ingle, Andreas Velten, and Mohit Gupta. High flux passive imaging with single-photon sensors. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6760–6769, 2019. 1

[32] Achuta Kadambi, Refael Whyte, Ayush Bhandari, Lee Streeter, Christopher Barsi, Adrian Dorrington, and Ramesh Raskar. Coded time of flight cameras: sparse deconvolution to address multipath interference and recover time profiles. *ACM Transactions on Graphics (ToG)*, 32(6):167, 2013. 2

[33] Martin Laurenzis and Emmanuel Bacher. Three-dimensional laser-gated viewing with error-free coding. *Optical Engineering*, 57(5):053103, 2018. 2

[34] Jongho Lee, Jenu Varghese Chacko, Bing Dai, Syed Azer Reza, Abdul Kader Sagar, Kevin W Eliceiri, Andreas Velten, and Mohit Gupta. Coding scheme optimization for fast fluorescence lifetime imaging. *ACM Transactions on Graphics (TOG)*, 38(3):1–16, 2019. 4

[35] David B Lindell, Matthew O'Toole, and Gordon Wetzstein. Single-photon 3d imaging with deep sensor fusion. *ACM Trans. Graph.*, 37(4):113–1, 2018. 8, 9

[36] Xiaochun Liu, Sebastian Bauer, and Andreas Velten. Phasor field diffraction based reconstruction for fast non-line-of-sight imaging systems. *Nature communications*, 11(1):1–13, 2020. 2

[37] Xiaochun Liu, Ibón Guillén, Marco La Manna, Ji Hyun Nam, Syed Azer Reza, Toan Huu Le, Adrian Jarabo, Diego Gutierrez, and Andreas Velten. Non-line-of-sight imaging using phasor-field virtual wave optics. *Nature*, 572(7771):620–623, 2019. 1

[38] Ashley Lyons, Francesco Tonolini, Alessandro Boccolini, Audrey Repetti, Robert Henderson, Yves Wiaux, and Daniele Faccio. Computational time-of-flight diffuse optical tomography. *Nature Photonics*, 13(8):575–579, 2019. 1

[39] 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. 1

[40] Parsa Mirdehghan, Wenzheng Chen, and Kiriakos N Kutulakos. Optimal structured light à la carte. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6248–6257, 2018. 2, 5, 8, 3

[41] 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, 2020. 1

[42] Ji Hyun Nam, Eric Brandt, Sebastian Bauer, Xiaochun Liu, Eftychios Sifakis, and Andreas Velten. Real-time non-line-of-sight imaging of dynamic scenes. *arXiv preprint arXiv:2010.12737*, 2020. 2

[43] Adithya Pediredla, Ashok Veeraraghavan, and Ioannis Gkioulekas. Ellipsoidal path connections for time-gated rendering. *ACM Transactions on Graphics (TOG)*, 38(4):1–12, 2019. 12

[44] Adithya K Pediredla, Aswin C Sankaranarayanan, Mauro Buttafava, Alberto Tosi, and Ashok Veeraraghavan. Signal processing based pile-up compensation for gated single-photon avalanche diodes. *arXiv preprint arXiv:1806.07437*, 2018. 3, 7

[45] Sara Pellegrini, Gerald S Buller, Jason M Smith, Andrew M Wallace, and Sergio Cova. Laser-based distance measurement using picosecond resolution time-correlated single-photon counting. *Measurement Science and Technology*, 11(6):712, 2000. 1, 2

[46] Jiayong Peng, Zhiwei Xiong, Xin Huang, Zheng-Ping Li, Dong Liu, and Feihu Xu. Photon-efficient 3d imaging with a non-local neural network. In *European Conference on Computer Vision*, pages 225–241. Springer, 2020. 8

[47] Christoph Peters, Jonathan Klein, Matthias B Hullin, and Reinhard Klein. Solving trigonometric moment problems for fast transient imaging. *ACM Transactions on Graphics (TOG)*, 34(6):1–11, 2015. 8

[48] Francesco Pittaluga, Zaid Tasneem, Justin Folden, Brevin Tilmon, Ayan Chakrabarti, and Sanjeev J Koppal. Towards a mems-based adaptive lidar. In *2020 International Conference on 3D Vision (3DV)*, pages 1216–1226. IEEE, 2020. 2

[49] Joshua Rapp, Robin MA Dawson, and Vivek K Goyal. Dithered depth imaging. *Optics Express*, 28(23):35143–35157, 2020. 2

[50] Joshua Rapp and Vivek K Goyal. A few photons among many: Unmixing signal and noise for photon-efficient active imaging. *IEEE Transactions on Computational Imaging*, 3(3):445–459, 2017. 9

[51] Joshua Rapp, Yanting Ma, Robin MA Dawson, and Vivek K Goyal. High-flux single-photon lidar. *Optica*, 8(1):30–39, 2021. 1

[52] Joshua Rapp, Julian Tachella, Yoann Altmann, Stephen McLaughlin, and Vivek K Goyal. Advances in single-photon lidar for autonomous vehicles: Working principles, challenges, and recent advances. *IEEE Signal Processing Magazine*, 37(4):62–71, 2020. 1

[53] Ximing Ren, Peter WR Connolly, Abderrahim Halimi, Yoann Altmann, Stephen McLaughlin, Istvan Gyongy, Robert K Henderson, and Gerald S Buller. High-resolution depth profiling using a range-gated cmos spad quanta image sensor. *Optics express*, 26(5):5541–5557, 2018. 2, 4, 8

[54] Walter Rudin. *Real and complex analysis*. McGraw-HiII Book Company, 1987. 2

[55] Trevor Seets, Atul Ingle, Martin Laurenzis, and Andreas Velten. Motion adaptive deblurring with single-photon cameras. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1945–1954, 2021. 1

[56] Michael P. Sheehan, Julián Tachella, and Mike E. Davies. A sketching framework for reduced data transfer in photon counting lidar. *IEEE Transactions on Computational Imaging*, 7:989–1004, 2021. 2, 4, 8

[57] Michael P Sheehan, Julián Tachella, and Mike E Davies. Surface detection for sketched single photon lidar. *arXiv preprint arXiv:2105.06920*, 2021. 2

[58] Dongeek Shin, Feihu Xu, Dheera Venkatraman, Rudi Lussana, Federica Villa, Franco Zappa, Vivek K Goyal, Franco NC Wong, and Jeffrey H Shapiro. Photon-efficient imaging with a single-photon camera. *Nature communications*, 7(1):1–8, 2016. 9

[59] Shuochen Su, Felix Heide, Gordon Wetzstein, and Wolfgang Heidrich. Deep end-to-end time-of-flight imaging. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6383–6392, 2018. 8

[60] Julián Tachella, Yoann Altmann, Nicolas Mellado, Aongus McCarthy, Rachael Tobin, Gerald S Buller, Jean-Yves Tourneret, and Stephen McLaughlin. Real-time 3d reconstruction from single-photon lidar data using plug-and-play point cloud denoisers. *Nature communications*, 10(1):1–6, 2019. 1, 9

[61] Ryuichi Tadano, Adithya Kumar Pediredla, and Ashok Veeraraghavan. Depth selective camera: A direct, on-chip, programmable technique for depth selectivity in photography. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3595–3603, 2015. 2

[62] Garry J. Tee. Eigenvectors of block circulant and alternating circulant matrices. 2005. 13

[63] George Turin. An introduction to matched filters. *IRE transactions on Information theory*, 6(3):311–329, 1960. 5, 3

[64] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020. 7

[65] Shumian Xin, Sotiris Nousias, Kiriakos N Kutulakos, Aswin C Sankaranarayanan, Srinivasa G Narasimhan, and Ioannis Gkioulekas. A theory of fermat paths for non-line-of-sight shape reconstruction. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6800–6809, 2019. 1

[66] Chao Zhang, Scott Lindner, Ivan Michel Antolović, Juan Mata Pavia, Martin Wolf, and Edoardo Charbon. A 30-frames/s, 252×144 spad flash lidar with 1728 dual-clock 48.8-ps tdcs, and pixel-wise integrated histogramming. *IEEE Journal of Solid-State Circuits*, 54(4):1137–1151, 2018. 2

[67] Yongyi Zhao, Ankit Raghuram, Hyun Kim, Andreas Hielscher, Jacob T Robinson, and Ashok Narayanan Veeraraghavan. High resolution, deep imaging using confocal time-of-flight diffuse optical tomography. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021. 1

[68] Vytautas Zickus, Ming-Lo Wu, Kazuhiro Morimoto, Valentin Kapitany, Areeba Fatima, Alex Turpin, Robert Insall, Jamie Whitelaw, Laura Machesky, Claudio Bruschini, et al. Fluorescence lifetime imaging with a megapixel spad camera and neural network lifetime estimation. *Scientific Reports*, 10(1):1–10, 2020. 1