Inter-Photon-Limited Videography

Andrew Xie[1, 2] Dongyu Du[1, 2] Sotiris Nousias[3] David B. Lindell[1, 2] Kiriakos N. Kutulakos[1, 2]

[1]University of Toronto [2]Vector Institute [3]Purdue University

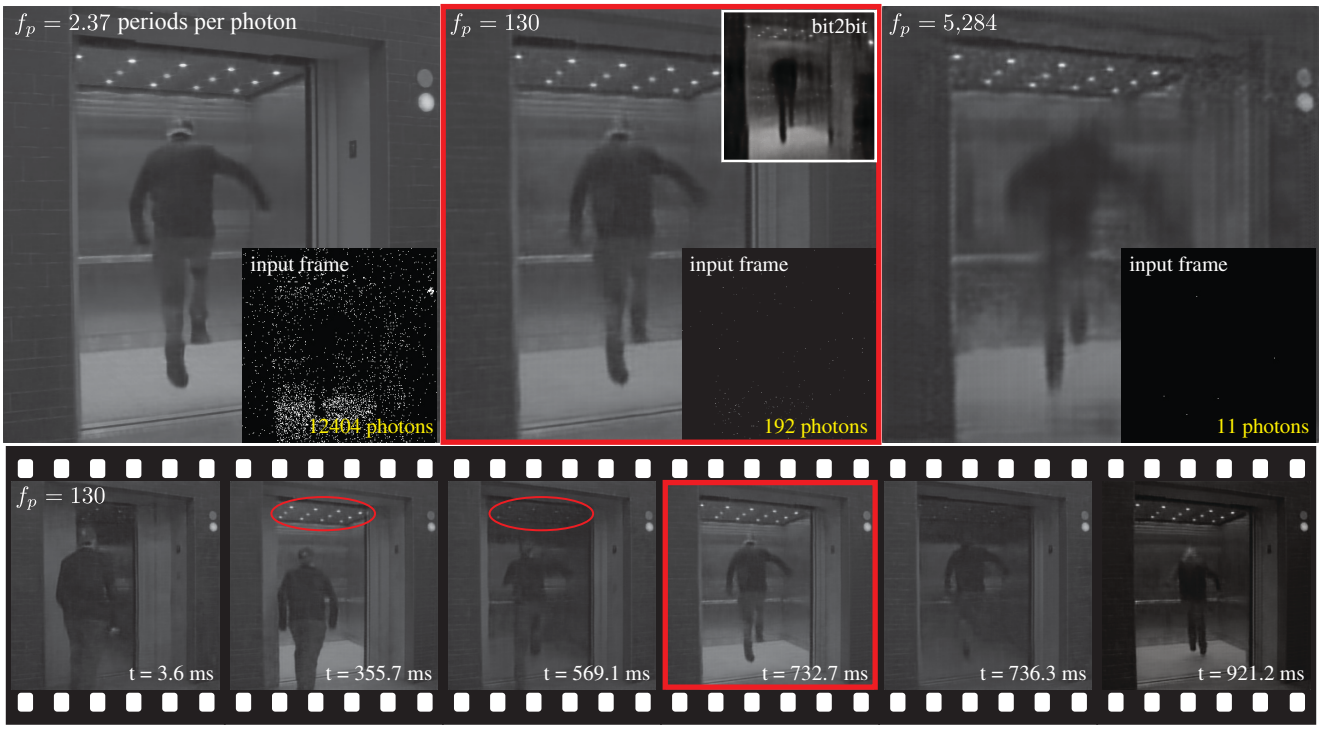

Figure 1. Videography in the inter-photon-limited regime. Our method reconstructs high-quality video from binary photon detections or timestamp data when scene intensity fluctuates faster than photon arrivals. We characterize this regime using the inter-photon frequency f_p in periods per photon, which directly couples the speed of photon arrivals and scene variations in a timescale-invariant manner. **Bottom:** Our method reconstructs both periodic (light flicker) and non-periodic scene variations (elevator door, human motion) simultaneously.

Abstract

We consider the problem of imaging a dynamic scene when scene appearance variations can outpace photon arrivals. Under such conditions, a pixel is effectively "blind" to changes in appearance that occur within the timespan separating the photons it detects, and so the inter-photon interval presents a significant speed barrier to video acquisition systems. To analyze and advance imaging capabilities at the inter-photon limit, we introduce a novel reparameterization of time-varying flux that reveals the intrinsic difficulty of signal reconstruction by relating the Fourier decomposition of a flux function to the number of photons arriving within each oscillation period. We find that inter-photon-limited videography of general scenes is underexplored and beyond the reach of existing reconstruction techniques. To this end, we introduce Neural Flux Fields (NFFs)—a technique that combines statistical modeling of photon arrivals with intrinsic priors of a neural network to achieve robust videography at the inter-photon limit. Using this approach, we demonstrate never-before-seen capabilities in video reconstruction across a range of captured single-photon video datasets spanning the inter-photon-limited regime.

1. Introduction

All forms of video acquisition—whether in bright or dim conditions—rely on the implicit assumption that the available light is sufficient to sustain the chosen frame rate and recording timespan. From recording speeding bullets in bright sunlight with million-frame-per-second cameras [39], to visualizing light propagation at billion-frame-per-second rates with pulsed lasers [42, 43], to imaging moving scenes in the dark with smartphones [13, 21] and single-photon cameras [22, 37]—this assumption underlies acquisition across diverse illumination and speed regimes.

But what speed can be achieved for a given light level? A general answer to this question has proved elusive be-

Corresponding authors: {axie, dongyu}@cs.toronto.edu
Project website: https://www.dgp.toronto.edu/inter-photon

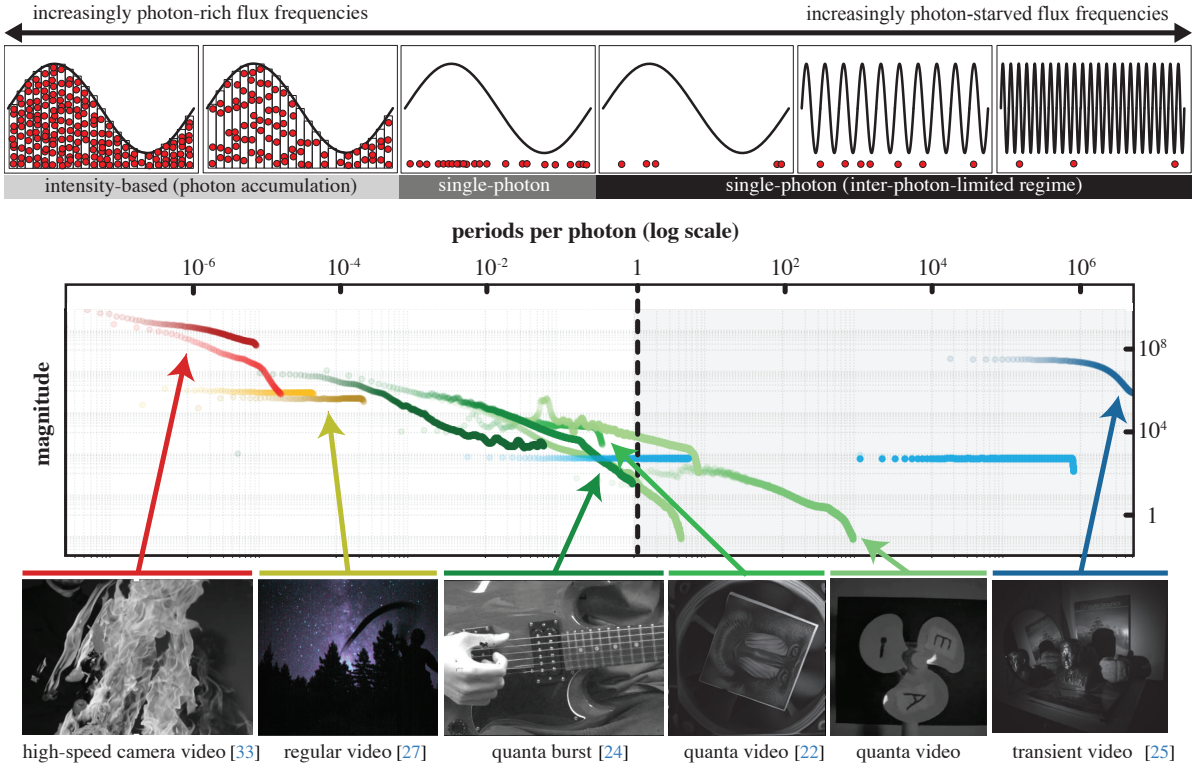

Figure 2. Top row: As the overall light level in a scene decreases, photons arrive farther apart on average and become individually resolvable (middle). As the light level decreases even further, pixels become "blind" to variations that unfold over progressively longer timespans. Our focus is imaging under inter-photon-limited conditions, where scene variations may outpace photon arrivals (right). **Middle and bottom rows:** Comparison of videography regimes across nearly 14 orders of magnitude in inter-photon frequency. We show inter-photon frequency spectra for 14 video datasets spanning diverse imaging conditions: from extremely bright to very dim, passive to active, slow to ultra-fast. Spectra in the far left (red) represent comparatively easy regimes where photons arrive much faster than variations in the scene; the far right (blue) represents extremely inter-photon-starved settings, where videography has only been possible with stroboscopic methods [25, 43] or by utilizing strong periodicity priors [45]. The one-period-per-photon regime represents a (soft) speed barrier for passive imaging of general scenes, with video datasets—even from single-photon cameras—falling mainly to the left.

cause of the diversity of illumination conditions and camera hardware—both active and passive—used for imaging time-varying scenes. Yet, understanding the relation between light level and attainable imaging speed is crucial for pushing the limits of video acquisition to even higher speeds, lower light levels, or both.

Toward this goal, we consider video acquisition under *inter-photon-limited conditions*—where scene variations outpace photon arrivals, making the time between consecutive photons the primary barrier to imaging speed (see Figure 1). Under such conditions, a pixel is effectively "blind" to changes in appearance that occur within the timespan separating the photons it detects, regardless of the camera's capabilities.

Inter-photon-limited conditions arise whenever the light level is too low—or scene dynamics too fast—for multiple photons to arrive at the timescale of appearance variation. They occur across a broad spectrum of timescales: from fractions of a second in extreme low-flux imaging [28, 45], to fractions of a millisecond in passive night-time videography [5, 22], down to sub-nanosecond timescales in bright, high-speed settings [23]. Many of these timescales already fall within the detection capabilities of modern single-photon avalanche-diode (SPAD) sensors [3], making the inter-photon interval a significant speed barrier in a growing range of video-acquisition systems.

To analyze and advance imaging capabilities under inter-photon-limited conditions, we propose two innovations: (1) a novel reparameterization of time-varying flux that offers new insights into the intrinsic difficulty of reconstructing a signal and (2) a method that enables robust videography at the inter-photon limit. Our parameterization is based on a simple idea: instead of representing flux variations in absolute units of time, frequency, frame rate, and power—which ties them to a very specific imaging setup and obscures the relation between photon availability and attainable speed—we represent them in relative units of *periods per photon detection*. In other words, we jointly consider the Fourier decomposition of time-varying flux and the rate at which photons arrive relative to the signal's frequency content. The only requirement is that pixels can time-stamp individual photons with a timing jitter, dead time, and dark count rate that are negligible relative to the inter-photon interval.

We leverage this representation to analyze the opera-

tional range of existing imaging systems, datasets, and video reconstruction algorithms relative to the inter-photon-limited regime (see Figure 2). Interestingly, we find that nearly all imaging systems—including those designed for single-photon sensing—operate in a regime where thousands or even millions of photons typically arrive within every period of a time-varying optical signal. As the rate of photon arrivals slows relative to variations in the flux function, we find (1) that few techniques operate at the inter-photon limit, and (2) existing methods fail to solve the challenge of inter-photon-limited videography.

To overcome this barrier, we introduce a new method that enables robust recovery of time-varying flux in the inter-photon-limited regime. Our approach—Neural Flux Fields—achieves this capability by combining statistical modeling of photon arrivals with spatiotemporal priors encoded by an untrained neural field. By optimizing the Neural Flux Field to find a representation of time-varying flux that is consistent with the arrival times of individual photons, we enable videography deep into the inter-photon-limited regime where even the best-performing previous reconstruction methods fail (Figure 1). Similar to previous self-supervised reconstruction techniques [8], our approach operates on a set of photon arrival times captured in a single video, and no large-scale training is required.

Overall, our work identifies and analyzes the regime of inter-photon-limited videography, establishes its connections to various imaging modalities—including low-light photography, quanta video reconstruction, inter-photon imaging, and stroboscopy—and reveals it to be both underexplored and largely inaccessible to existing reconstruction techniques. Further, we use our Neural Flux Fields approach to demonstrate never-before-seen capabilities in photon-starved video reconstruction, across a range of captured single-photon video datasets spanning the inter-photon-limited regime.

1.1. Related Work

Although the inter-photon interval is well-recognized as an intensity cue for imaging scenes that change very slowly at that timescale [16], its role as a speed limit for imaging is far less understood. In particular, existing models of time-varying image formation [1, 11, 15, 34] do not capture both the discrete nature of incident light and the speed constraints imposed by the "blind" timespan between photons. Crucially, existing research conflates photon-limited imaging—where light levels are low enough in *absolute terms* to detect and process individual photons [24]—from the inter-photon-limited regime, where light levels are low *relative to a scene's time-varying appearance*. This is a far more challenging condition where existing methods break down for general scenes (Figure 1).

The only prior methods that successfully surpass the inter-photon limit are restricted to periodic signals, where the underlying structure can be inferred by observing photon arrivals across many repeated cycles [28, 45]. Extending this capability to arbitrary, non-periodic scenes requires stronger priors that integrate information from sparse photon arrivals over both space and time. Self-supervised image and video reconstruction techniques [8, 19, 36, 47] have demonstrated such priors, but they operate on dense pixel arrays and are not directly applicable to photon arrivals. Consequently, it remains unclear how to scale them to handle the potentially millions of asynchronous photon events acquired over long temporal windows.

Previous approaches for single-photon videography leverage hand-crafted or learned spatiotemporal priors to constrain the reconstruction task. For instance, burst imaging methods [5, 24] align and merge photon detections to enhance signal-to-noise ratio, but this requires that flux variations occur much more slowly than photon arrivals—an assumption that fails in the inter-photon-limited regime. More recent approaches employ convolutional neural networks [22] or diffusion models [6] to recover dynamic flux directly from photon timestamps; however, they can only process a limited number of photon arrivals in a single forward pass, which prevents exploiting long-range dependencies across space and time to break through the inter-photon limit. See supplement Section A for further discussion of related techniques in neural video representation and restoration based on conventional and quanta sensors.

2. Inter-Photon-Limited Imaging

Videography in the inter-photon-limited regime depends on the relation between three characteristic timescales: (1) the timescale of *significant appearance variations* in a scene, (2) the timescale of *consecutive photon arrivals* at a pixel, and (3) the timescale of *individual video frames*. We consider the first two below and the third in Section 3.

Photon flux. Incident light on the image plane can be expressed as a time-varying function $\phi(\mathbf{x}, t)$ that represents instantaneous photon flux at pixel $\mathbf{x}$ and time t [2]. Formally, $\phi(\mathbf{x}, t)$ is the rate function of an inhomogeneous Poisson process that governs photon arrivals at each pixel [26]. We assume that $\phi(\mathbf{x}, t)$ is defined over a discrete pixel grid, is continuous over an acquisition interval $[0, t_{\text{acq}}]$, and is measured in units of photons per second.

The mean inter-photon interval. The mean of $\phi(\mathbf{x}, t)$ at pixel $\mathbf{x}$ describes the average number of photon detections per second at $\mathbf{x}$. Its inverse, measured in units of seconds per photon, is the *mean inter-photon interval*:[1]

$$\tau(\mathbf{x}) = \left(\frac{1}{t_{\text{acq}}} \int_0^{t_{\text{acq}}} \phi(\mathbf{x}, t) dt \right)^{-1}. \tag{1}$$

Spectral support of photon flux at a pixel. The temporal frequency spectrum of $\phi(\mathbf{x}, t)$, which captures the timescales of significant appearance variation, is largely determined by a scene's illumination and dynamics. Natural videos captured under steady (DC) illumination exhibit an approximate $1/f$ behavior in the temporal frequency domain [9], corresponding to an aperiodic flux signal with

[1] Strictly speaking, $\tau(x)$ is the mean inter-photon interval of the equivalent homogeneous Poisson process that yields the same expected number of photon detections over $[0, t_{\text{acq}}]$ as the inhomogeneous flux $\phi(x, t)$.

no dominant timescale. In most real-world scenes, spectral support under DC lighting is confined to a relatively narrow band, from sub-hertz to tens of kilohertz, reflecting physical limits on scene and camera motion[2] and on their vibrational modes [7]. In contrast, when one or more light sources in a scene are temporally modulated—due to light bulb flicker [35], strobe lighting [41], projectors [32], continuous time-of-flight sensors [12], or pulsed lasers [14, 37]—the flux may contain periodic components with harmonics extending into the gigahertz range [28]. Temporal variations in flux therefore occur at two broad timescales: a "long" timescale of seconds to microseconds, where flux is generally aperiodic, and a "short" timescale of nanoseconds to picoseconds, where it is mainly periodic. Our approach is applicable to both settings and enables passive acquisition of aperiodic as well as periodic appearance variations occurring in a scene (see Figures 1 and 6).

Inter-photon-limited camera model. SPAD-based cameras span a wide range of architectures (1D/2D arrays [40], scan-based single-pixel systems [20]); detection characteristics (quantum efficiency, timing precision, dead time, dark count rate); and data formats (binary photon detection frames [46] or asynchronous photon timestamp streams [10]). Since our specific focus is a regime where the fundamental speed barrier to imaging is not the camera but the mean inter-photon interval, we assume that the incident flux lies within the camera's bandwidth,[3] that the detector's dead time is much shorter than $\tau(\mathbf{x})$, and that the probability of spurious photon detections is low enough to be ignored. In this case, photon detection timestamps can be modeled as realizations of the flux function $\phi(\mathbf{x}, t)$.

Our approach is agnostic to the specific representation of photon detection data. In the following, we represent the camera's output as a sequence of tuples (x_i, y_i, t_i), where t_i is either the timestamp of the i-th photon detection or the time assigned to the binary frame containing that detection. In all cases, we assume that timestamps t_i track real time.

2.1. Inter-Photon Flux Parameterization

While the above model captures the photon-detection process, it does not make explicit how reconstruction uncertainty scales with flux level or temporal frequency. To analyze and compare imaging performance across different light levels and timescales, we therefore seek a parameterization of photon flux that is invariant to absolute time units and unifies measurements across imaging conditions of widely differing brightness and speed.

Timescale-invariant uncertainty property. Our approach builds on a key observation. Consider two sinusoidal flux signals—one bright and fast, the other dim and slow—that produce, on average, the same number of photon detections within each of their periods (Figure 3, top). Because the photon timestamps generated by these two signals follow identical distributions over their respective periods, an inter-photon-limited camera observing the same number of periods from each will yield parameter estimates with identical statistical uncertainty. Thus, despite differing in absolute frequency and amplitude, both signals appear equally uncertain from the camera's perspective.

Inter-photon flux & inter-photon frequency. The above observation suggests that absolute time is not the natural variable for describing flux uncertainty in the inter-photon-limited regime. To exploit it, we redefine time in units of mean photon detections $p = t/\tau(\mathbf{x})$, which is timescale-invariant. Intuitively, each pixel $\mathbf{x}$ operates on its own "photon clock" that advances at the average rate of photon detections. This substitution yields a representation of incident flux that is mathematically equivalent to $\phi(\mathbf{x}, t)$ but independent of any absolute timescale or illumination level.

In this representation, flux is expressed in units of *photons per period*, implicitly coupling the mean inter-photon interval and temporal variation of flux into a single frequency. Conversely, inter-photon frequency is expressed in units of periods per photon, with higher frequencies corresponding to increasingly photon-starved flux signals (Figure 2). We refer the reader to Figure 3 (middle) for full expressions related to this parameterization.

Impact of various acquisition strategies on f_p. Shortening the exposure time and reducing light level affect the conventional Fourier magnitude similarly, but they differ fundamentally in the inter-photon domain (Figure 4): lower light levels increase the inter-photon interval and the inter-photon frequency, whereas shortening the exposure does not. In stroboscopic imaging [41], synchronization to a modulation frequency f_{sync} effectively folds time at the sync period, downscaling all frequencies by f_{sync}. This can reduce inter-photon frequency and estimation uncertainty considerably.

Discussion. The inter-photon parameterization has several important ramifications both within and beyond the context of modeling and understanding single-photon cameras. First, it exposes imaging constraints that are not directly captured by conventional representation of flux (Figure 3, bottom). Second, by jointly coupling light level and temporal variation, the parameterization provides a timescale-invariant basis for performance assessment—one that replaces absolute metrics such as illuminance (lux) or frame rate with an *inter-photon frequency response*. Third, it reveals a unified imaging landscape that spans more than fourteen orders of magnitude in inter-photon frequency, encompassing both passive and active imaging systems (Figure 2).

The unit inter-photon frequency. Frequency $f_p = 1$ represents a soft but physically meaningful boundary beyond which pixels are blind over timespans longer than a period (Figure 2, top). Under such photon-starved conditions, performance of both passive and active videography systems requires use of spatiotemporal priors on scene illumination, scene appearance, or both. We turn to this problem next.

[2] MHz-scale variations in incident flux can occur in highly specialized conditions, and cameras supporting passive video recording up to a few Mfps under bright illumination are commercially available [17].

[3] Typical bandwidths range from ∼50 kHz for quanta-like sensors [40] to >10 GHz for time-stamping SPADs [30], with dead times from ∼10 μs to a few nanoseconds, respectively.

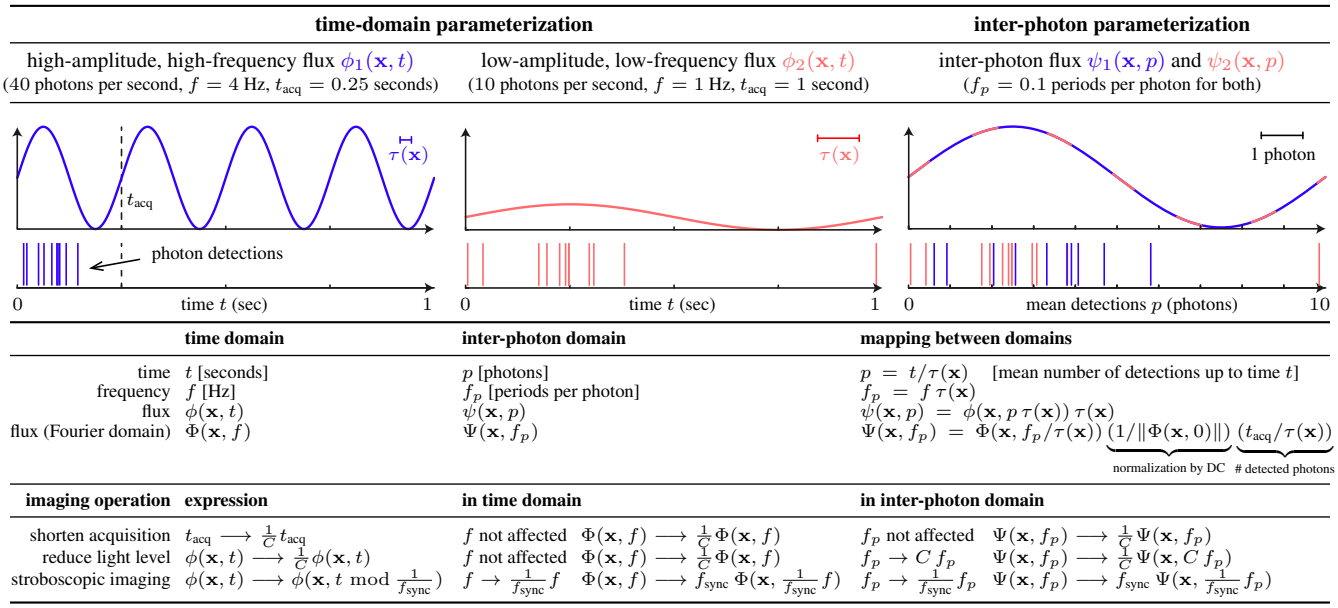

Figure 3. Inter-photon flux parameterization. **Top row:** Illustration of the time-invariant uncertainty property. **Middle row:** Key relations governing time-domain and inter-photon-domain flux parameterizations. As the mean inter-photon interval increases (*i.e.*, lower light level) the corresponding inter-photon frequency increases, reflecting greater uncertainty in flux estimation. **Bottom row:** Comparison of basic imaging operations. Shortening exposure time and reducing light level affect the conventional Fourier magnitude in the same way, whereas they differ considerably in the inter-photon domain: lower light levels increase the inter-photon interval and thus the apparent inter-photon frequency, whereas shortening the exposure does not. In stroboscopic imaging, synchronization to a modulation frequency f_{sync} effectively folds time at the sync period, downscaling all flux frequencies by f_{sync} and reducing inter-photon frequency. See Figure 4 for an illustration.

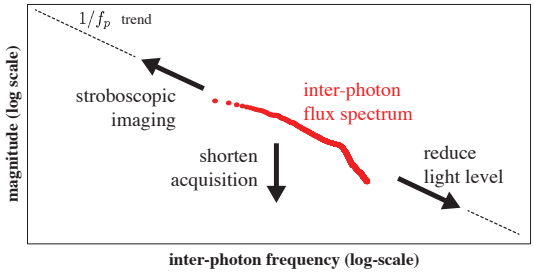

Figure 4. Relationship between the inter-photon flux spectrum and the operations described in Figure 3. Shortening the acquisition time reduces the magnitude of the inter-photon flux, while reducing the light level increases inter-photon frequency as well. Stroboscopic imaging has the same effect as boosting a scene's light level, shifting the inter-photon spectrum to the upper left.

3. Neural Flux Fields

We address the problem of inter-photon-limited videography by training a Neural Flux Field (NFF)—a neural network that estimates time integrals of the flux function from photon detection data across a sensor.

Neural field representations are particularly well suited to this problem because they allow flux integrals over a wide range of temporal integration domains to be represented and optimized differentiably. In particular, videography in all inter-photon-limited regimes shown in Figure 2—from quanta videography and transient imaging to passive ultra-wideband imaging—can be modeled with the same neural network; the only difference is the temporal integration domain associated with each video frame.

Crucially, because the same temporal frequency may correspond to very different inter-photon frequencies depending on a pixel's mean inter-photon interval, the network architecture can be designed so that variations at "brighter" pixels (low f_p) inform recovery of similar variations at much "dimmer" ones (same f, much higher f_p). In our NFF, this joint inference is enabled by a frequency-encoded temporal latent space that is independent of pixel position.

Flux integral of a video frame. Flux is an instantaneous quantity that must be integrated over time to form frame v:

$$v(\mathbf{x}) = \int_{\mathcal{V}} \phi(\mathbf{x}, t) dt \ . \tag{2}$$

In conventional video acquisition, frames are formed by integrating light over consecutive time intervals. This form of integration is appropriate for general, temporally aperiodic variations. When flux has a periodic component, however, substantial reductions in inter-photon frequency can be achieved by synchronizing frame integration to its period (Figure 3, bottom). Intuitively, this allows each frame to integrate light over multiple repeated cycles while preserving temporal structure within a period. To enable such stroboscopic integration schemes, we allow $\mathcal{V}$ to be a finite collection of non-consecutive intervals that distribute a frame's exposure over one or more timespans in $[0, t_{\text{acq}}]$.

Video reconstruction task. Given a sequence of photon detections $\mathcal{P} = \{(x_i, y_i, t_i)\}_{i=1}^{P}$, we seek to compute the integral in Eq. (2) for an ordered sequence $\mathcal{V}_1, \ldots, \mathcal{V}_N$

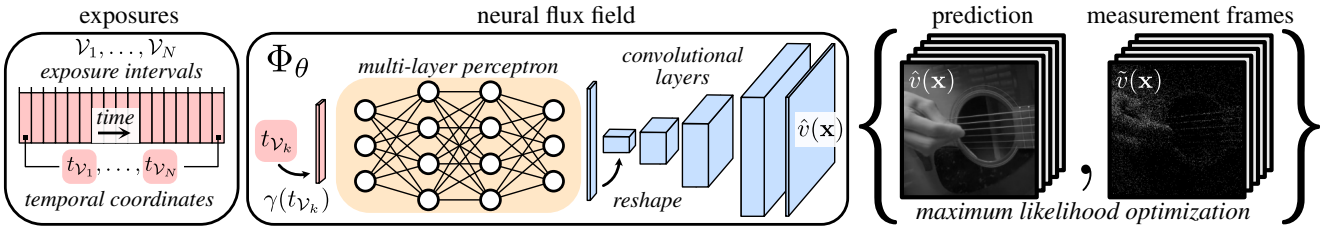

Figure 5. Overview of neural flux fields. We associate a temporal coordinate $t_{\mathcal{V}_k}$ with each exposure $\mathcal{V}_k$. The neural flux field Φ_θ takes the temporal coordinate as input and passes it through a temporal frequency encoding function γ. This encoding is processed by a multi-layer perceptron, whose output is shaped into a spatial tensor and processed by convolutional layers. Finally, the resulting predicted video frame $\hat{v}(\mathbf{x})$ is compared to the measurement frame $\tilde{v}(\mathbf{x})$, and we optimize Φ_θ to maximize the likelihood of the measurements.

of equal-length, frame-specific integration domains corresponding to video frames $(v_1, \ldots, v_N)$.

Photon aggregation. To create the measurements used to optimize the network, we aggregate photon detections corresponding to each frame's exposure. Specifically, we define a measurement frame $\tilde{v}(\mathbf{x})$ as the count of photons detected at each pixel within the exposure $\mathcal{V}$:

$$\tilde{v}(\mathbf{x}) = |\{(x_i, y_i, t_i) \mid (x_i, y_i) = \mathbf{x} \text{ and } t_i \in \mathcal{V}\}| \ . \tag{3}$$

We seek to train a neural flux field to predict video frames that are consistent with these measurement frames.

3.1. Architecture

The neural flux field is parameterized by a neural network Φ_θ, which maps a temporal coordinate $t_\mathcal{V}$ associated with an exposure $\mathcal{V}$ to a predicted video frame $\hat{v}$, or

$$\hat{v} = \Phi_\theta(t_\mathcal{V}) \ . \tag{4}$$

The network architecture consists of three main components (Figure 5)—a temporal frequency encoding layer, a multi-layer perceptron (MLP), and a set of convolution layers. We train the network to minimize a negative log-likelihood loss based on the Poisson statistics of photon detections within each measurement frame.

Temporal frequency encoding. We set the input temporal coordinate $t_\mathcal{V}$ to the start time of the first interval in $\mathcal{V}$. Then, we normalize the resulting coordinate values such that $t_\mathcal{V} \in [-1, 1]$. We apply a temporal frequency embedding to $t_\mathcal{V}$ to improve the ability of the model to represent high-frequency variations and to facilitate recovery of frequencies present in the photon data [38]:

$$\boldsymbol{\gamma}(t_\mathcal{V}) = \left[t_\mathcal{V}, \sin(\boldsymbol{\omega} t_\mathcal{V}), \cos(\boldsymbol{\omega} t_\mathcal{V})\right]^T \ , \tag{5}$$

where $\boldsymbol{\omega} = [2, 2^2, \ldots, 2^L]^T$ and we use $L = 16$.

Pixel-independent latent space representation. By making the network's initial layers independent of pixel position $\mathbf{x}$, we force it to learn a temporal latent space that is shared by all pixels. In particular, the embedding $\boldsymbol{\gamma}(t_\mathcal{V})$ is processed by a four-layer MLP, producing a high-dimensional latent vector that we reshape into a spatial tensor of size $16 \times 16 \times 256$. This tensor is then passed through a convolutional network adapted from a recent implicit video representation [4], consisting of four residual blocks interleaved with $2\times$ bilinear upsampling. Finally, we apply a softplus activation to obtain the predicted video frame $\tilde{v}$, normalized to $[0, 1]$. See supplement Section C for more details.

Loss function. We optimize the network parameters θ by comparing the predicted video frame $\hat{v} = \Phi_\theta(t_\mathcal{V})$ with the corresponding measurement frame $\tilde{v}$. Assuming that photon detections at pixels $\mathbf{x}$ follow an inhomogeneous Poisson process with rate $\hat{v}(\mathbf{x})$, we minimize the Poisson negative log-likelihood (NLL) across all exposures $\{\mathcal{V}_k\}_{k=1}^N$:

$$\mathcal{L}(\theta) = \sum_{k=1}^{N} \sum_{\mathbf{x}} \left[\hat{v}_k(\mathbf{x}) - \tilde{v}_k(\mathbf{x}) \log \hat{v}_k(\mathbf{x})\right] \ . \tag{6}$$

3.2. Computational Stroboscopy

For periodic flux signals, the number of detected photons per measurement frame can be greatly increased—and inter-photon frequencies reduced—using a computational analog to stroboscopy. Specifically, we identify the signal's fundamental frequencies and design interleaved exposures $\mathcal{V}$ that integrate light over multiple cycles, with intervals separated by the period T.

Fundamental frequency estimation. We identify the fundamental frequency of the periodic signal using harmonic probing [28] on the raw photon data. The procedure locates strong peaks in the temporal Fourier spectrum of the signal and iteratively localizes them to obtain a precise estimate of the fundamental frequency.

Creating the periodic exposures. For a scene with a single detected fundamental frequency f (with period $T = 1/f$), we define a set of periodic exposures $\{\mathcal{V}_k\}_{k=1}^N$. Here, each exposure $\mathcal{V}_k$ spans multiple, disjoint intervals that are separated by the period T:

$$\mathcal{V}_k = \bigcup_{m=1}^{M} \left[t_{\mathcal{V}_k} + mT, t_{\mathcal{V}_k} + mT + \Delta t\right] \ , \tag{7}$$

where $t_{\mathcal{V}_k}$ is the start time of the first such interval, Δt is the duration of each disjoint exposure interval, and we create $N = \lfloor T/\Delta t \rfloor$ exposures to cover the full period T. See supplement Section C.2 for more details.

3.3. Implementation Details

We implement the model in PyTorch [29] and train end-to-end using the Adam optimizer [18] with a learning rate

of 10^{-3} and a batch size of 128 temporal coordinates. For a typical dataset of 100k–130k frames, we train with early stopping for a fixed 30 epochs. This is sufficient for convergence and avoids overfitting to the noisy photon data. On an NVIDIA RTX A6000 Ada GPU, training for a single scene completes in approximately 3.5 hours.

4. Experimental Results

We evaluate our approach on both captured and synthetic datasets spanning a wide range of inter-photon spectra (Figure 6, row 1); inter-photon frequencies f_p ranging from less than one to over a million; and photon data in the form of quanta image sequences as well as picosecond-resolution timestamp streams. See the supplemental webpage for videos and supplement Section D for more results.

Inter-photon-limited videography with quanta cameras. Figure 1 shows reconstruction of 100 kfps video of a person jumping into an elevator with fast flickering lights. The one-second input sequence from a SPAD512 camera [40] yielded 2000 photons per pixel per second on average, corresponding to $f_p = 2.37$, *i.e.*, slightly above the soft $f_p = 1$ threshold. To simulate settings that are even more photon-starved, we thin photon detections by up to three orders of magnitude ($f_p = 5284$). Despite this, our method recovers the scene's time-varying appearance even for the highest f_p.

Figure 6 (row 5) shows another very challenging case of a foam bullet shot from a toy gun: the bullet is small and fast-moving, the gun's motion is aperiodic, and the low light levels produce extremely sparse photon detections in both space and time. Our method successfully reconstructs the trajectory of the bullet (Figure 6, row 5, middle). As predicted by our inter-photon frequency parameterization, the hand's slower motion and larger size generate more photons, making it easier to recover (Figure 6, row 1). Even when we increase f_p by $1000\times$ from 0.447 to 309 via thinning—rendering the bullet unobservable—the hand is accurately reconstructed (Figure 6, row 5, right).

We also consider inter-photon-limited videography using simulated photon streams from high-speed video datasets. Figure 6 (row 5) shows results on a highly dynamic scene of milk splashing over cereal. To achieve acceptable quality, conventional high-speed cameras operate in the range of $f_p = 10^{-6}$ to $f_p = 10^{-4}$ (Figure 2) whereas our method recovers video of comparable quality at inter-photon frequencies that are 5 orders of magnitude higher.

The NFF advantage. Figure 6 (row 4) shows videos of a fan spinning at 54 Hz in a dark room reconstructed from a quanta image sequence with just 40 photons detected per pixel per second ($f_p = 44.7$). We compare two methods: (1) our NFF approach *without* employing computational stroboscopy and (2) passive ultra-wideband imaging (UWB) [45]. The inter-photon frequency parameterization reveals the fundamental limit of single-pixel methods such as UWB: in these extreme inter-photon-limited settings, the fan's harmonics fall below the noise floor, causing UWB, which is well-suited for periodic flux signals but neglects spatial correlations, to produce blurred results. In contrast, NFF successfully reconstructs the fan's time-varying appearance even after significant thinning ($f_p = 5188$).

Computational stroboscopy with free-running SPADs. We use the method of Section 3.2 and train the same NFF network as above to recover videos from UWB's *fan* photon timestamp dataset—an inter-photon-limited regime that is far beyond the capabilities of quanta-based methods ($f_p > 10^6$). As shown in Figure 6 (row 7), we recover superior slow-motion videos compared to UWB, and transient videos of higher fidelity.

Comparison to Quanta Burst Photography (QBP) [24]. Figure 6 (row 2) shows results on two thinned versions of QBP's *guitar* sequence. In particular, since the original sequence is not inter-photon limited, we synthetically thin it to increase f_p to 4.21 and 21.07, respectively. With this photon sparsity, QBP's temporal aggregation cannot improve intensity estimates; there are not enough photons to accumulate, causing its reconstructions to collapse to binary patterns. NFF, on the other hand, recovers fine guitar string motions at both f_p levels.

Comparison to bit2bit [22]. Lastly, we show results on two thinned versions of the *drill* sequence from [22] (Figure 6, row 3). While bit2bit works well on the original sequence ($f_p = 2.82$), its performance under aggressive thinning ($f_p = 1995$, and $f_p = 32359$) reveals the limitations of its small temporal context window: only photon detections within its temporal window can be exploited, and as f_p increases those correlations are no longer present. As a result, it yields motion-blurred intensity estimates.

5. Concluding Remarks

We believe that the inter-photon frequency representation provides a useful lens through which the intrinsic difficulty of photon-starved imaging regimes and datasets can be characterized. It remains an open question, however, whether the inter-photon re-parameterization should be used directly in the reconstruction process. Neural Flux Fields, while enabling videography well inside the inter-photon-limited regime, also have limitations. Like other deep-image-prior-like methods [8, 19], they require early stopping to avoid overfitting the flux function to stochastic photon arrivals. Because NFFs are trained on frames with short integration times, repeatedly querying the flux field for longer exposures can be computationally expensive. Reconstruction quality can also degrade when interpolating between training time indices.

There are many avenues for future work. Beyond further improvements to the architecture, our approach could be combined with strong statistical priors over natural videos learned by generative models [31, 44]. There are also opportunities in active imaging and computational stroboscopy, where the interplay between illumination and a scene's time-varying appearance could further improve inter-photon-limited video reconstruction.

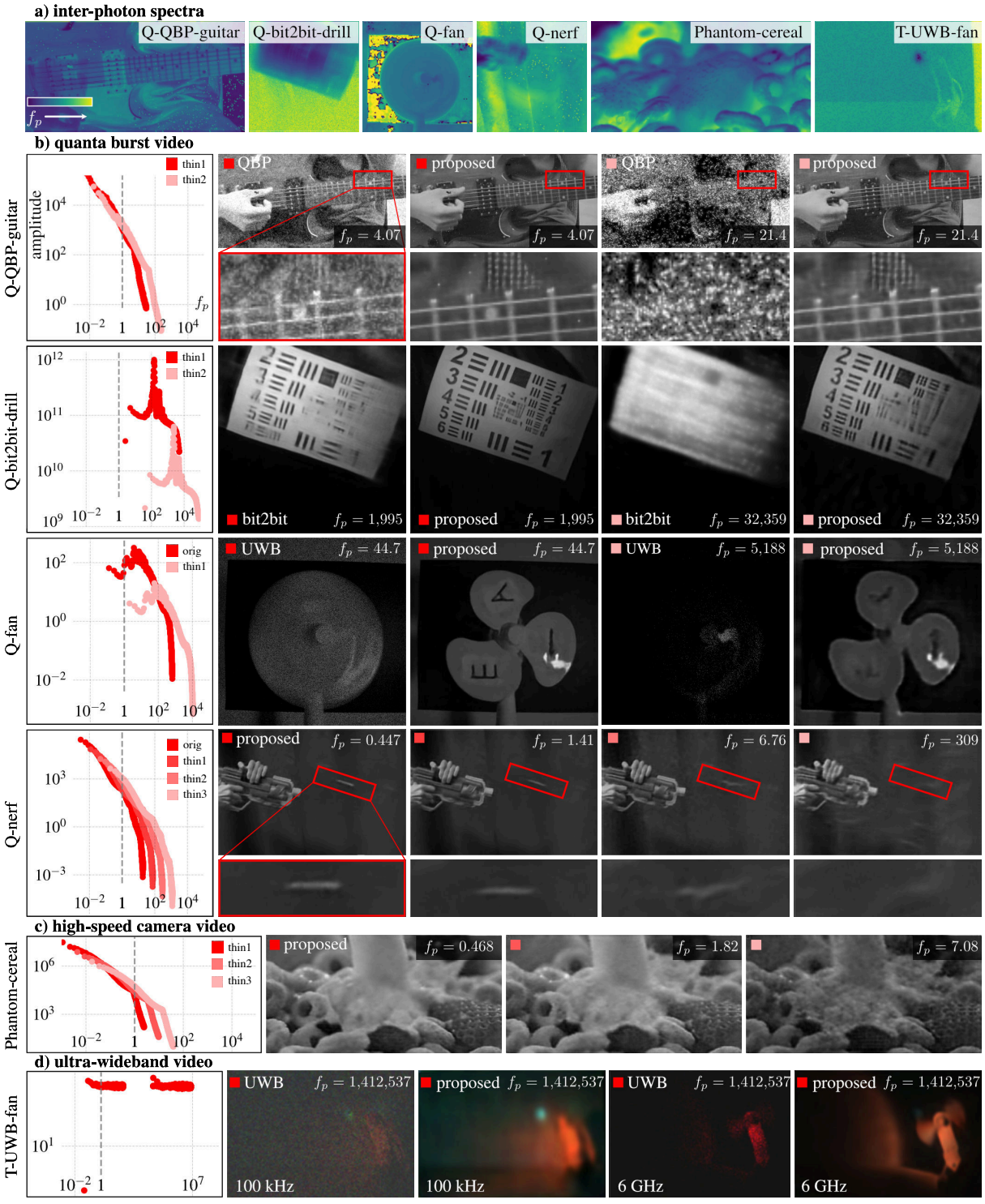

Figure 6. Inter-photon-limited videography results. **Row 1:** We show the pixel-resolved inter-photon frequency for five different scenes captured by a range of cameras, including quanta sensors, a high-speed phantom camera, and a picosecond-resolution single-photon avalanche diode (yellow means higher frequency). The inter-photon spectrum of each sequence is shown on the leftmost column. **Rows 2–5:** We compare inter-photon-limited video reconstructions from Neural Flux Fields to those from quanta burst photography (QBP) [24], bit2bit [22], and ultra-wideband imaging [45] at two different inter-photon frequencies (f_p). Our approach successfully reconstructs videos where other methods fail. **Row 6:** We also compare to a scene with complex motion, where photon arrivals are simulated from videos captured from a conventional high-speed camera. **Row 7:** We compare to ultra-wideband imaging and show that our approach can recover cleaner videos at two different video timescales (100 kHz and 6 GHz).

6. Acknowledgements

DBL, KNK, and AX acknowledge the support of NSERC under the RGPIN, RTI, and CGRS-M programs. DBL also acknowledges the support from the Canada Foundation for Innovation and the Ontario Research Fund. DD acknowledges the support of the Arts & Science Postdoctoral Fellowship Award.

References

[1] Israel Bar-David. Communication under the Poisson regime. *IEEE Trans. Inf. Theory*, 15(1):31–37, 2003. 3

[2] Robert W Boyd. *Radiometry and the detection of optical radiation*. John Wiley & Sons, 1983. 3

[3] Francesco Ceccarelli, Giulia Acconcia, Angelo Gulinatti, Massimo Ghioni, Ivan Rech, and Roberto Osellame. Recent advances and future perspectives of single-photon avalanche diodes for quantum photonics applications. *Adv. Quantum Technol.*, 4(2):2000102, 2021. 2

[4] Hao Chen, Bo He, Hanyu Wang, Yixuan Ren, Ser Nam Lim, and Abhinav Shrivastava. Nerv: Neural representations for videos. In *Adv. Neural Inform. Process. Syst.*, pages 21557–21568, 2021. 6

[5] Prateek Chennuri, Yiheng Chi, Enze Jiang, GM Dilshan Godaliyadda, Abhiram Gnanasambandam, Hamid R Sheikh, Istvan Gyongy, and Stanley H Chan. Quanta video restoration. In *Eur. Conf. Comput. Vis.*, pages 152–171, 2024. 2, 3

[6] Prateek Chennuri, Dongdong Fu, and Stanley H Chan. Quanta diffusion. In *IEEE Int. Conf. Image Process.*, pages 229–234, 2025. 3

[7] Abe Davis, Katherine L Bouman, Justin G Chen, Michael Rubinstein, Fredo Durand, and William T Freeman. Visual vibrometry: Estimating material properties from small motion in video. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 5335–5343, 2015. 4

[8] Ulyanov Dmitry, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. *Int. J. Comput. Vis.*, 128(7):1867–1888, 2020. 3, 7

[9] Dawei W Dong and Joseph J Atick. Statistics of natural time-varying images. *Network: computation in neural systems*, 6 (3):345, 1995. 3

[10] Anant Gupta, Atul Ingle, and Mohit Gupta. Asynchronous single-photon 3d imaging. In *IEEE/CVF Int. Conf. Comput. Vis.*, pages 7909–7918, 2019. 4

[11] Anant Gupta, Atul Ingle, Andreas Velten, and Mohit Gupta. Photon-flooded single-photon 3d cameras. In *IEEE/CVF Conf. Comput. Vis. Pattern Recog.*, pages 6770–6779, 2019. 3

[12] Miles Hansard, Seungkyu Lee, Ouk Choi, and Radu Patrice Horaud. *Time-of-flight cameras: principles, methods and applications*. Springer Science & Business Media, 2012. 4

[13] Samuel W Hasinoff, Dillon Sharlet, Ryan Geiss, Andrew Adams, Jonathan T Barron, Florian Kainz, Jiawen Chen, and Marc Levoy. Burst photography for high dynamic range and low-light imaging on mobile cameras. *ACM Trans. Graph.*, 35(6):1–12, 2016. 1

[14] Felix Heide, Steven Diamond, David B Lindell, and Gordon Wetzstein. Sub-picosecond photon-efficient 3d imaging using single-photon sensors. *Sci. Rep.*, 8(1):17726, 2018. 4

[15] Atul Ingle, Andreas Velten, and Mohit Gupta. High flux passive imaging with single-photon sensors. In *IEEE/CVF Conf. Comput. Vis. Pattern Recog.*, pages 6760–6769, 2019. 3

[16] Atul Ingle, Trevor Seets, Mauro Buttafava, Shantanu Gupta, Alberto Tosi, Mohit Gupta, and Andreas Velten. Passive inter-photon imaging. In *IEEE/CVF Conf. Comput. Vis. Pattern Recog.*, pages 8585–8595, 2021. 3

[17] Hyun-Ha Kim, Jong-Ho Kim, and Atsushi Ogata. Time-resolved high-speed camera observation of electrospray. *J. Aerosol Sci.*, 42(4):249–263, 2011. 4

[18] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *Int. Conf. Learn. Represent.*, 2015. 6

[19] Chenyang Lei, Yazhou Xing, Hao Ouyang, and Qifeng Chen. Deep video prior for video consistency and propagation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 45(1):356–371, 2022. 3, 7

[20] Zheng-Ping Li, Jun-Tian Ye, Xin Huang, Peng-Yu Jiang, Yuan Cao, Yu Hong, Chao Yu, Jun Zhang, Qiang Zhang, Cheng-Zhi Peng, et al. Single-photon imaging over 200 km. *Optica*, 8(3):344–349, 2021. 4

[21] Orly Liba, Kiran Murthy, Yun-Ta Tsai, Tim Brooks, Tianfan Xue, Nikhil Karnad, Qiurui He, Jonathan T Barron, Dillon Sharlet, Ryan Geiss, et al. Handheld mobile photography in very low light. *ACM Trans. Graph.*, 38(6):164–1, 2019. 1

[22] Yehe Liu, Alexander Krull, Hector Basevi, Ales Leonardis, and Michael Jenkins. bit2bit: 1-bit quanta video reconstruction via self-supervised photon prediction. *Adv. Neural Inform. Process. Syst.*, pages 88443–88485, 2024. 1, 2, 3, 7, 8

[23] Specialised Imaging Ltd. Kirana ultra high-speed video camera. https://www.specialised-imaging.com/products/video-cameras/kirana/, 2025. 2

[24] Sizhuo Ma, Shantanu Gupta, Arin C Ulku, Claudio Bruschini, Edoardo Charbon, and Mohit Gupta. Quanta burst photography. *ACM Trans. Graph.*, 39(4):79–1, 2020. 2, 3, 7, 8

[25] Anagh Malik, Noah Juravsky, Ryan Po, Gordon Wetzstein, Kiriakos N Kutulakos, and David B Lindell. Flying with photons: Rendering novel views of propagating light. In *Eur. Conf. Comput. Vis.*, pages 333–351, 2024. 2

[26] James E Mazo and Jack Salz. On optical data communication via direct detection of light pulses. *Bell Syst. Tech. J.*, 55(3): 347–369, 1976. 3

[27] Kristina Monakhova, Stephan R Richter, Laura Waller, and Vladlen Koltun. Dancing under the stars: video denoising in starlight. In *IEEE/CVF Conf. Comput. Vis. Pattern Recog.*, pages 16241–16251, 2022. 2

[28] Sotiris Nousias, Mian Wei, Howard Xiao, Maxx Wu, Shahmeer Athar, Kevin J Wang, Anagh Malik, David A Barmherzig, David B Lindell, and Kiriakos N Kutulakos. Opportunistic single-photon time of flight. In *IEEE/CVF Conf. Comput. Vis. Pattern Recog.*, pages 15852–15862, 2025. 2, 3, 4, 6

[29] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. In *Adv. Neural Inform. Process. Syst.*, pages 8026–8037, 2019. 6

[30] Simone Riccardo, Enrico Conca, Vincenzo Sesta, Andreas Velten, and Alberto Tosi. Fast-gated 16× 16 spad array with

16 on-chip 6 ps time-to-digital converters for non-line-of-sight imaging. *IEEE Sensors J.*, 22(17):16874–16885, 2022. 4

[31] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *IEEE/CVF Conf. Comput. Vis. Pattern Recog.*, pages 10684–10695, 2022. 7

[32] Joaquim Salvi, Jordi Pages, and Joan Batlle. Pattern codification strategies in structured light systems. *Pattern Recognition*, 37(4):827–849, 2004. 4

[33] Bruno Saravia Vega. The world in slow motion in 8K ultra HD. [Online Video]. Available: https://youtu.be/A2eNVDGesYY, 2021. 2

[34] Trevor Seets, Atul Ingle, Martin Laurenzis, and Andreas Velten. Motion adaptive deblurring with single-photon cameras. In *IEEE/CVF Winter Conf. Appl. Comput. Vis.*, pages 1945–1954, 2021. 3

[35] Mark Sheinin, Yoav Y Schechner, and Kiriakos N Kutulakos. Computational imaging on the electric grid. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 6437–6446, 2017. 4

[36] Dev Yashpal Sheth, Sreyas Mohan, Joshua L Vincent, Ramon Manzorro, Peter A Crozier, Mitesh M Khapra, Eero P Simoncelli, and Carlos Fernandez-Granda. Unsupervised deep video denoising. In *IEEE/CVF Conf. Comput. Vis. Pattern Recog.*, pages 1759–1768, 2021. 3

[37] 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 Commun.*, 7 (1):12046, 2016. 1, 4

[38] Matthew Tancik, Pratul Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan Barron, and Ren Ng. Fourier features let networks learn high frequency functions in low dimensional domains. *Adv. Neural Inform. Process. Syst.*, pages 7537–7547, 2020. 6

[39] Sigurdur T Thoroddsen, Takeharu Goji Etoh, and Kohsei Takehara. High-speed imaging of drops and bubbles. *Annu. Rev. Fluid Mech.*, 40(1):257–285, 2008. 1

[40] Arin Can Ulku, Claudio Bruschini, Ivan Michel Antolović, Yung Kuo, Rinat Ankri, Shimon Weiss, Xavier Michalet, and Edoardo Charbon. A 512×512 spad image sensor with integrated gating for widefield flim. *IEEE J. Sel. Top. Quantum Electron.*, 25(1):1–12, 2018. 4, 7

[41] Ashok Veeraraghavan, Dikpal Reddy, and Ramesh Raskar. Coded strobing photography: Compressive sensing of high speed periodic videos. *IEEE Trans. Pattern Anal. Mach. Intell.*, 33(4):671–686, 2010. 4

[42] Andreas Velten, Thomas Willwacher, Otkrist Gupta, Ashok Veeraraghavan, Moungi G Bawendi, and Ramesh Raskar. Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging. *Nature Commun.*, 3(1):745, 2012. 1

[43] Andreas Velten, Di Wu, Adrian Jarabo, Belen Masia, Christopher Barsi, Chinmaya Joshi, Everett Lawson, Moungi Bawendi, Diego Gutierrez, and Ramesh Raskar. Femto-photography: capturing and visualizing the propagation of light. *ACM Trans. Graph.*, 32(4):1–8, 2013. 1, 2

[44] Team Wan, Ang Wang, Baole Ai, Bin Wen, Chaojie Mao, Chen-Wei Xie, Di Chen, Feiwu Yu, Haiming Zhao, Jianxiao Yang, et al. Wan: Open and advanced large-scale video generative models. *arXiv:2503.20314*, 2025. 7

[45] Mian Wei, Sotiris Nousias, Rahul Gulve, David B Lindell, and Kiriakos N Kutulakos. Passive ultra-wideband single-photon imaging. In *IEEE/CVF Int. Conf. Comput. Vis.*, pages 8135–8146, 2023. 2, 3, 7, 8

[46] Feng Yang, Yue M Lu, Luciano Sbaiz, and Martin Vetterli. Bits from photons: Oversampled image acquisition using binary poisson statistics. *IEEE Trans. Image Process.*, 21(4): 1421–1436, 2011. 4

[47] Huan Zheng, Tongyao Pang, and Hui Ji. Unsupervised deep video denoising with untrained network. In *AAAI Conf. Artif. Intell.*, pages 3651–3659, 2023. 3