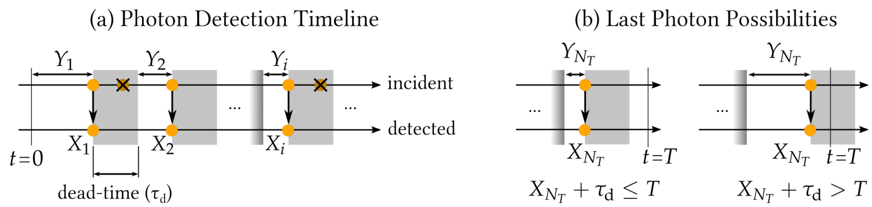
Supplementary Document for "Passive Inter-Photon Imaging"

Atul Ingle*, Trevor Seets, Mauro Buttafava, Shantanu Gupta,
Alberto Tosi, Mohit Gupta†, Andreas Velten†

(CVPR 2021)

Supplementary Note 1. Image Formation



Suppl. Fig. 1: **Photon Detection Timeline:** (a) The photon timeline shows the random variables used in the derivation of our photon flux estimator. X_i's denote the photon arrival time with respect to the start of the exposure at $t = 0$ and Y_i's denote the i^{th} time-of-darkness. (b) There are two possibilities for the final dead-time window after the last photon detection. In high photon flux scenarios, the final dead-time ends after the end of the exposure time $(X_{N_T} + \tau_\mathrm{d} > T)$ with high probability.

Consider an IP-SPAD sensor pixel with quantum efficiency q exposed to a photon flux of Φ photons/second. Photon arrivals follow a Poisson process, so inter-photon times follow an exponential distribution with rate $q\Phi$. After a detection event the IP-SPAD is unable to detect photons for a period of τ_d. Because of the memoryless property of a Poisson process, the arrival time of a photon after the end of a dead-time window (called the *time-of-darkness*), follows an exponential distribution with rate $q\Phi$. The probability that no photons are detected in the exposure time T is equal to $\int_T^\infty q\Phi e^{-q\Phi t} dt = e^{-q\Phi T}$.

Let the first time-of-darkness be denoted by Y_1. If no photons are detected, we define $Y_1 := T$. If $Y_1 < T$ it follows an exponential distribution. Therefore, the probability density function of Y_1 can be written as:

$$Y_1 \sim f_{Y_1}(t) = \begin{cases} q\Phi e^{-q\Phi t} & \text{for } 0 < t < T \\ e^{-q\Phi T}\delta(t-T) & \text{for } t = T \\ 0 & \text{otherwise,} \end{cases} \tag{S1}$$

where δ is the Dirac delta function.

Now consider Y_2, the second time-of-darkness. Y_2 is non-zero if and only if $Y_1 \neq T$ (to be the second, there must be a first). If the second photon is detected, Y_2 will be exponentially distributed. But the exposure time interval has shrunk because a time interval of $Y_1 + \tau_\mathrm{d}$ has elapsed due to the first photon detection. We define the remaining exposure time $T_2 = \max(0, T - Y_1 - \tau_d)$, where the $\max()$ function ensures T_2 is non-negative. Then the probability distribution of Y_2 conditioned on Y_1 will be given by replacing T for T_2 in Eq. (S1). More generally, the conditional distribution of Y_i can be written as:

$$Y_i \sim f_{Y_i}(Y_i | Y_1 \ldots Y_{i-1}) = \begin{cases} q\Phi e^{-q\Phi Y_i} & \text{for } 0 < Y_i < T_i \\ e^{-q\Phi T_i}\delta(t-T_i) & \text{for } Y_i = T_i \\ 0 & \text{otherwise,} \end{cases} \tag{S2}$$

†Equal contribution.
*Email: ingle@uwalumni.com

where,

$$
\begin{aligned}
T_1 &= T \\
T_i &= \max(0, T_{i-1} - Y_{i-1} - \tau_\text{d})) \\
&= \max\left(0, T - \sum_{j=1}^{i-1} (Y_j + \tau_\text{d})\right).
\end{aligned}
\tag{S3}
$$

The T_i's model the fact that the effective exposure time for the i^{th} photon shrinks due to preceding photon detections. Note if $Y_i = T_i$ then no i^th photon is detected and $Y_{i+1} = T_{i+1} = 0$. Note that the X_i in the main text are related to Y_i by $X_i - X_{i-1} - \tau_\text{d} =: Y_i$ for $i \geq 2$ and $X_1 = Y_1$. Suppl. Fig. 1(a) shows X_i and Y_i on a photon timeline.

Maximum Likelihood Flux Estimator

For a fixed exposure time T, the maximum number of possible photon detections is $L = \lceil \frac{T}{\tau_\text{d}} \rceil$. Let N be the number of detected photons, then Y_{N+1} will be the last possibly non-zero time-of-darkness, and $Y_{N+2} \ldots Y_L = 0$ with probability 1. The log-likelihood of the unknown flux value given the set of time-of-darkness measurements $Y_1 \ldots Y_L$ is given by:

$$
\begin{aligned}
\log l(q\Phi; Y_1, \ldots, Y_L) &= \log\left(\prod_{i=1}^{L} f_{Y_i}(Y_i | Y_1 \ldots Y_{i-1})\right) \\
&= \log\left(f_{Y_{N+1}}(Y_{N+1} | Y_1 \ldots Y_N) \prod_{i=1}^{N} f_{Y_i}(Y_i | Y_1 \ldots Y_{i-1})\right) \\
&= \log\left(e^{-q\Phi T_{N+1}} \prod_{i=1}^{N} q\Phi\, e^{-q\Phi Y_i}\right) \\
&= -q\Phi\left(T_{N+1} + \sum_{i=1}^{N} Y_i\right) + N \log q\Phi \\
&= -q\Phi\left(\max\left(0, T - \sum_{i=1}^{N} Y_i - \tau_\text{d}\right) + \sum_{i=1}^{N} Y_i\right) + N \log q\Phi \\
&= -q\Phi\, \max\left(\sum_{i=1}^{N} Y_i, T - N\tau_\text{d}\right) + N \log q\Phi.
\end{aligned}
\tag{S4}
$$

We find the maximum likelihood estimate, $\widehat{\Phi}$, by setting the derivative of Eq.(S4) to zero and solving for Φ:

$$
-q \max\left(\sum_{i=1}^{N} Y_i, T - N\tau_\text{d}\right) + \frac{N}{\widehat{\Phi}} = 0,
\tag{S5}
$$

which gives:

$$
\widehat{\Phi} = \frac{N}{q \max\left(\sum_{n=1}^{N} Y_i, T - N\tau_\text{d}\right)}.
\tag{S6}
$$

The $\max()$ function can be thought of as selecting the time-of-darkness based on whether or not the final dead-time window ends after $t = T$, see Suppl. Fig. 1(b). In practice the beginning and end of the exposure time may not be known precisely, introducing uncertainty in X_1 and T. Because of this we instead use an approximation:

$$
\widehat{\Phi} = \frac{N-1}{q \sum_{n=2}^{N} Y_i}.
\tag{S7}
$$

Plugging in $Y_i = X_i - X_{i-1} - \tau_\text{d}$ gives Eq. (2) in the main text.

Flux Estimator Variance

Let N be the number of photons detected in an exposure time T. Using the law of large numbers for renewal processes we find the expectation and the variance of N to be:

$$\mathrm{E}[N] = \frac{q\Phi(T+\tau_\mathrm{d})}{1+q\Phi\tau_\mathrm{d}} \tag{S8}$$

$$\mathrm{Var}[N] = \frac{q\Phi(T+\tau_\mathrm{d})}{(1+q\Phi\tau_\mathrm{d})^3} \tag{S9}$$

In the following derivation we will assume N is large enough that it can be assumed to be constant for a given T. This holds, for example, when $\Phi \gg \frac{1}{T}$. This assumption also allows us to approximate Y_i's as i.i.d. shifted exponential random variables. We will consider the estimator in Eq. (S7) where the sum in the denominator is given by $S_{N_T} = Y_2 + Y_3 + ...Y_N$ and letting $N_T = N - 1$. The final photon timestamp S_{N_T} is the sum of exponential random variables and follows a gamma distribution:

$$S_{N_T} \sim f_{S_{N_T}}(t) = \begin{cases} \frac{(q\Phi)^{N_T} t^{N_T-1} e^{-q\Phi t}}{(N_T-1)!} & \text{for } t \geq 0 \\ 0 & \text{otherwise.} \end{cases} \tag{S10}$$

The mean of $\widehat{\Phi}$ can be computed as:

$$\begin{aligned}
\mathrm{E}\left[\frac{N_T}{qS_{N_T}}\right] &= \frac{N_T}{q}\mathrm{E}\left[\frac{1}{S_{N_T}}\right] \\
&= \frac{N_T}{q}\int_0^\infty \frac{(q\Phi)^{N_T} t^{N_T-1} e^{-q\Phi t}}{t(N_T-1)!}\,dt \\
&= \frac{N_T}{q}\frac{q\phi}{N_T-1}\int_0^\infty \frac{(q\Phi)^{N_T-1} t^{N_T-2} e^{-q\Phi t}}{(N_T-2)!}\,dt \\
&= \frac{N_T}{N_T-1}\Phi
\end{aligned} \tag{S11}$$

where the last line comes from the recognizing the argument of the integral as the p.d.f. for a gamma distribution and for large N_T, $\frac{N_T}{N_T-1} \approx 1$.

The second moment of $\widehat{\Phi}$ is given by:

$$\begin{aligned}
\mathrm{E}\left[\left(\frac{N_T}{qS_{N_T}}\right)^2\right] &= \frac{N_T^2}{q^2}\mathrm{E}\left[\frac{1}{S_{N_T}^2}\right] \\
&= \frac{N_T^2}{q^2}\int_0^\infty \frac{(q\Phi)^{N_T} t^{N_T-1} e^{-q\Phi t}}{t^2(N_T-1)!}dt \\
&= \frac{(q\Phi)^2 N_T^2}{q^2(N_T-1)(N_T-2)}\int_0^\infty \frac{(q\Phi)^{N_T-2} t^{N_T-3} e^{-q\Phi t}}{(N_T-3)!}dt \\
&= \frac{\Phi^2 N_T^2}{(N_T-1)(N_T-2)}
\end{aligned} \tag{S12}$$

This expression is valid for $N_T > 2$. The variance of $\widehat{\Phi}$ is given by:

$$\begin{aligned}
\mathrm{Var}\left[\frac{N_T}{qS_{N_T}}\right] &= \Phi^2 \frac{N_T^2}{(N_T-2)(N_T-1)} - \frac{N_T^2}{(N_T-1)^2}\Phi^2 \\
&= \Phi^2 \frac{N_T^2}{(N_T-2)(N_T-1)^2} \\
&\approx \Phi^2 \frac{1}{N_T} && \text{(S13)} \\
&= \Phi^2 \frac{q\Phi\tau_\mathrm{d}+1}{q\Phi(T+\tau_\mathrm{d})} && \text{(S14)} \\
&\approx \Phi \frac{q\Phi\tau_\mathrm{d}+1}{qT} && \text{(S15)}
\end{aligned}$$

where we replace N_T with its mean value. The last line follows if we assume $T \gg \tau_\mathrm{d}$. Finally, the SNR is given by:

$$\begin{aligned}
\mathrm{SNR} &= 20\log_{10} \frac{\Phi}{\sqrt{\mathrm{Var}[\frac{N_T}{qS_{N_T}}]}} \\
&= 10\log_{10} \frac{q\Phi T}{q\Phi\tau_\mathrm{d}+1} && \text{(S16)}
\end{aligned}$$

We make the following observations about our estimator $\widehat{\Phi}$:

- At high flux, when N_T is large enough, Eq. (S11) reduces to $\mathrm{E}[\widehat{\Phi}] = \Phi$, i.e. our estimator is unbiased.
- Unlike [25] which only uses photon counts N_T, our derivation explicitly accounts for individual inter-photon timing information captured in S_{N_T}.
- As $\Phi \to \infty$, SNR $\to 10\log_{10}(\frac{T}{\tau_\mathrm{d}})$. So at high flux the SNR will flatten out to a constant independent of the true flux Φ. In practice, the SNR drops at high flux due to time quantization, discussed next in Supplementary Note 2.

Supplementary Note 2. Time Quantization

Consider an IP-SPAD with quantum efficiency q, dead time τ_{d}, and time quantization Δ that detects photons over exposure time T. To match our hardware prototype, the start of the dead time window is not quantized and time stamps are quantized by Δ. The quantization noise variance term derived in previous work [25, 4] that relies on a counts-only measurement model is given by:

$$V_{\text{count-quantization}} = \frac{(1 + q\Phi\tau_{\mathrm{d}})^4}{12q^2T^2}. \tag{S17}$$

We derive a modified quantization noise variance expression by modifying this counts-only expression to account for two key insights gained from extensive simulations of SNR plots for our timing-based IP-SPAD flux estimator. First, we note that the timing-based IP-SPAD flux estimator follows a similar SNR curve as the counts-based PF-SPAD flux estimator when $\Delta = \tau_{\mathrm{d}}$. Second, the rate at which the SNR drop off moves slows after Δ exceeds τ_{d}. In this way we propose a new time quantization term:

$$V_{\text{time-quantization}} = \frac{(1 + q\Phi\tau_{\mathrm{d}} + q\Phi\Delta)^2(1 + q\phi\Delta)^2}{12q^2T^2}. \tag{S18}$$

Note we break the quartic term from Eq. (S17) into two quadratic terms. The two quadratic terms strike a balance between quantization due to counts and timing. If $\Delta = 0$ then $V_{\text{time quantization}}$ is an order 2 polynomial with respect to Φ which leads to a constant SNR at high flux. Also note if $\Delta = \tau_{\mathrm{d}}$ the time quantization term is roughly equal to the counts quantization term. We found this expression matches simulated IP-SPAD SNR curves for a range of dead-times and exposure times.

Supplementary Note 3. IP-SPAD Imaging with Low Photon Counts

The scene brightness estimator (Eq. (2)) requires the IP-SPAD pixel to capture at least two photons; It does not make sense to talk about "inter-photon" times with only one photon. The situation where an IP-SPAD pixel captures only one incident photon timestamp can be thought of as an extreme limiting case of passive inter-photon imaging under low illumination.

Intuitively, we can reconstruct an image from a single photon timestamp per pixel by simply computing the reciprocal of the first photon timestamp at each pixel. Brighter scene points should have a smaller first-photon timestamp (on average) because, with high probability, a photon will be detected almost immediately after the pixel starts collecting light. In this supplementary note we show that the conditional distribution of this first photon timestamp (conditioned on there being at least one photon detection) is a uniform random variable:

$$\{Y_1|Y_1 \le T\} \sim \mathcal{U}[0,T].$$

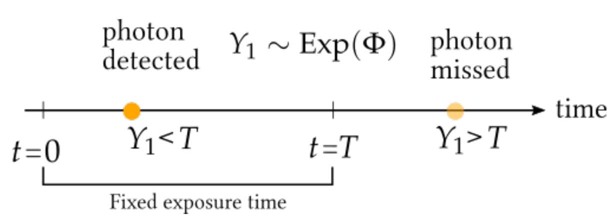
when operating under low incident photon flux. This implies that timestamps provide no additional information beyond merely the fact that at least one photon was detected. We must, therefore, relax the requirement of a constant exposure time and allow each pixel to capture at least one photon by allowing variable exposure times per pixel. When operated this way, first-photon timestamps do carry useful information about the scene brightness. The estimate of the scene pixel brightness is given by $\widehat{\Phi} = {}^1\!/_{q\,Y_i}$.

When the total number of photons is extremely small, the information contained in the timestamp data is extremely noisy. We leverage spatial-priors-based image denoising techniques that have been developed for conventional images, and adapt them denoising these noisy IP-SPAD images. Coupled with the inherent sensitivity of SPADs, this enables us to reconstruct intensity images with just a single photon per pixel [26].

In this section we show that for passive imaging in the low photon flux regime with a constant exposure time per pixel, the timestamp of the first arriving photon is a uniform random variable and hence, carries no useful information about the true photon flux. If we drop the constant exposure time constraint and instead operate in a regime where each pixel is allowed to wait until the first photon is captured (random exposure time per pixel), then the first-photon timestamps carry useful information about the flux, albeit noisy.

Supplementary Note 3.1. When Do Timestamps Carry Useful Information?

Let us assume an IP-SPAD pixel operating with a fixed exposure time T is observing a scene point with photon flux Φ. We assume that the photon flux is low enough so that the pixel captures at most one photon during this exposure time. The (random) first-photon arrival time is denoted by Y_1 as shown in Suppl. Fig. 2. We would like to know if the first photon time-of-arrival carries useful information about Φ.



Suppl. Fig. 2: We capture the first arriving photon and record its arrival time in a fixed exposure time T. Note that in the low photon flux regime $\Phi T \ll 1$, so there is a high probability that zero photons are detected in the time interval $[0,T]$.

We derive the probability distribution of Y_1, conditioning on $Y_1 \le T$. For any $t > 0$,

$$P(Y_1 \le t | Y_1 \le T) = \frac{P(Y_1 \le t \cap Y_1 \le T)}{P(Y_1 \le T)} \tag{S19}$$

$$= \frac{P(Y_1 \le t)}{P(Y_1 \le T)} \tag{S20}$$

$$= \frac{1 - e^{-\Phi t}}{1 - e^{-\Phi T}} \tag{S21}$$

where Eq. (S19) follows from Bayes's rule, Eq. (S20) assumes $t \le T$ (otherwise the answer is 1, trivially) and Eq. (S21) is obtained by plugging in the c.d.f. of $Y_1 \sim \text{Exp}(\Phi)$.

Due to the low flux assumption, $\Phi \ll \frac{1}{T}$. Then $\Phi t \leq \Phi T \ll 1$ and we can approximate $1 - e^{-\Phi T} \approx \Phi T$ and $1 - e^{-\Phi t} \approx \Phi t$. This gives

$$P(Y_1 \leq t | Y_1 \leq T) = \frac{t}{T} \tag{S22}$$

which is the c.d.f. of a uniform random variable. This implies that, in the low photon flux regime the arrival time distribution converges weakly to a uniform random variable:

$$\{Y_1 | Y_1 \leq T\} \xrightarrow{D} \mathcal{U}[0, T].$$

For low illumination conditions, we drop the requirement of a fixed exposure time and allow the IP-SPAD pixel to wait until the first photon timestamp is captured.

Supplementary Note 3.2. KPN-based Denoising Network for Low Light IP-SPAD Imaging

In principle, any standard neighborhood-based image denoising algorithm (e.g., bilateral filtering [42] and BM3D [15]) can be applied to the IP-SPAD images captured in a low photon count regime. But the heavy-tailed nature of the timestamps poses problems to off-the-shelf denoising algorithms as they usually assume a light-tailed distribution of pixel intensities (e.g., Gaussian distribution). A solution to this issue is the use of a variance-stabilizing Anscombe transform [3] to make the noise variance uniform across the whole image. For photon timestamp data, the variance-stabilizing transform is the logarithm. See Supplementary Note 3.3 for a proof. We design an image denoising deep neural network (DNN) that operates on log-transformed first-photon timestamp images.

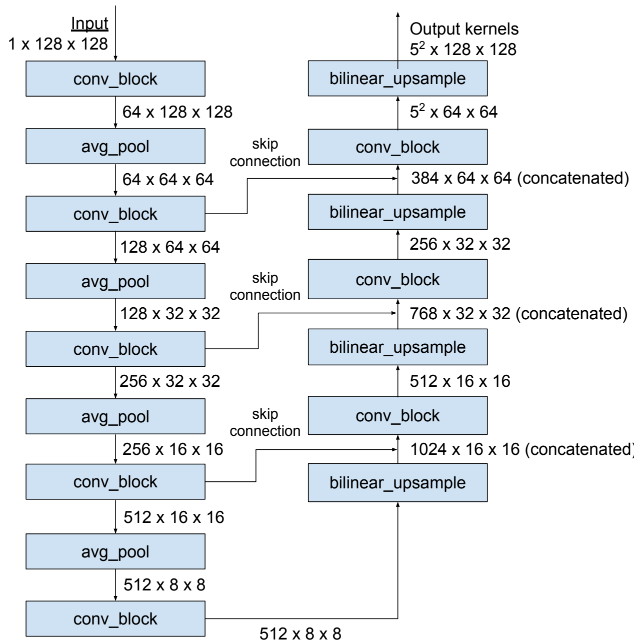
We use a kernel prediction network (KPN) architecture [5, 38]. Our network architecture is shown in Suppl. Fig. 3. The network produces 5×5 kernels for every pixel in the input image, which we apply to generate the denoised image. The only substantial post-processing step is to correct the bias introduced by using the log-timestamp instead of the timestamp itself (see Supplementary Note 3.3).

We train the network with timestamp images simulated from the DIV2K dataset [2, 48]. This dataset has 800 high-resolution images; we simulate four random timestamp images for each image in the dataset for a total of 3200 training images. The original 8-bit images are first converted to 16-bit linear luminance [47], before simulating the timestamps. The simulated timestamps are then log-transformed.

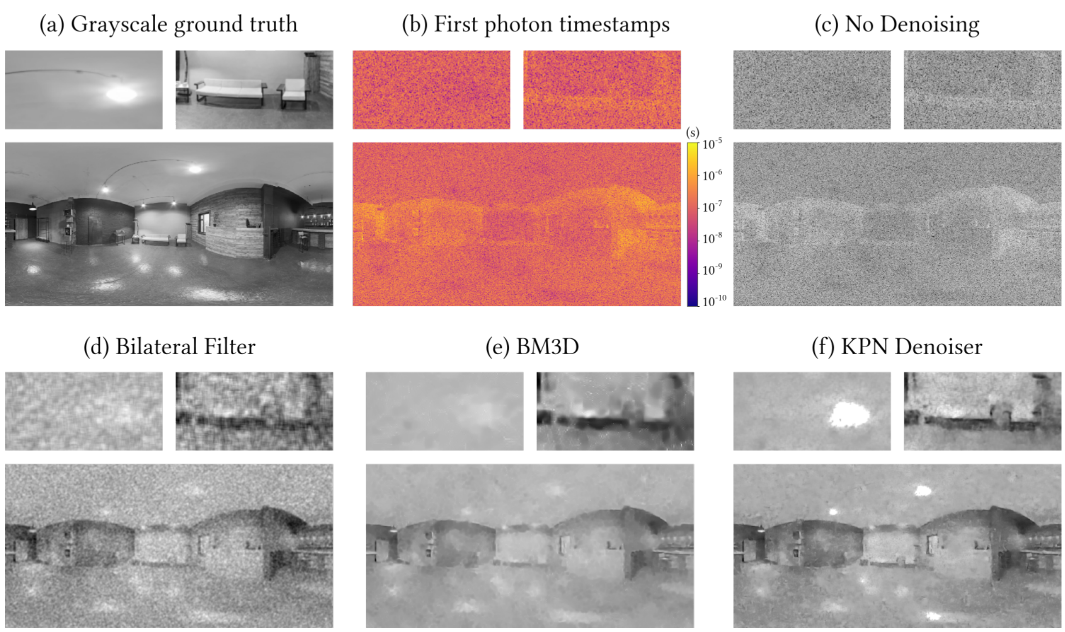
We use the Adam optimizer [28] with a learning rate of 10^{-4}. The loss function is a sum of squared errors in the pixel intensities and absolute errors in the pixel-wise image gradients, both with respect to the original image from which the timestamps are simulated [38]. Training runs for 1920 iterations with a batch size of 5 images, for a total of 3 epochs. Images are randomly cropped into 128×128 patches before passing into the network when training. However, since the network is fully convolutional, it can handle arbitrary input image sizes at test time.

The architecture of our kernel prediction network-based denoising DNN is shown in Suppl. Fig. 3.

Suppl. Fig. 4 shows simulated denoising results comparing our KPN-based denoiser with two standard denoising methods: bilateral filtering and BM3D.



Suppl. Fig. 3: The kernel prediction network (KPN) architecture we have used to estimate per-pixel kernels of size 5×5, which is adapted from the architectures used for burst denoising in [38] and for denoising Monte-Carlo renderings in [5]. The input image size is 128×128 when training the network, but any image size can be used at the inference stage.



Suppl. Fig. 4: **Denoising IP-SPAD Images with Low Photon Counts:** (a-b) We simulate the extreme case of IP-SPAD imaging by sampling at most one photon time stamp per pixel of a ground truth image. (c) Simply inverting the each time stamp is not enough due to extreme noise, (d-f) so it is necessary to combine time stamps spatially. (d) We apply a bilateral filter ($\sigma = 7$), which incorporates some spatial information, but still remains quite noisy. (e) BM3D [14] may over smooth, and it seems to have particular trouble in bright regions. (f) Our KPN denoiser trained on photon timestamp data preserves some object shapes like the bright ceiling light and the couches.

Supplementary Note 3.3. Homoskedasticity of log-timestamps

In this section we show that the log-transformation is a variance stabilizing transformation for first-photon timestamp data. Let $Y \sim \text{Exp}(\Phi)$ be the arrival time of the first incident photon (we drop the subscript in Y_1 for simplicity). Our goal is to show that the variance of $\log(Y)$ is constant (homoskedasticity).

From the properties of the exponential distribution,

$$P(Y \leq t) = 1 - e^{-\Phi t} \tag{S23}$$

$$\implies P(\log Y \leq \log t) = 1 - e^{-\Phi t}. \tag{S24}$$

Defining $\tilde{Y} = \log(Y)$ and $y = \log(t)$,

$$P(\tilde{Y} \leq y) = 1 - e^{-\Phi e^y} \tag{S25}$$

$$\implies p_{\tilde{Y}}(y) = \Phi e^{-\Phi e^y} e^y \tag{S26}$$

where $p_{\tilde{Y}}(y)$ is the p.d.f. of $\tilde{Y}$.

$$\implies E[\tilde{Y}] = \int_{-\infty}^{\infty} y \Phi e^{-\Phi e^y} e^y \, dy \tag{S27}$$

$$= \int_{-\infty}^{\infty} \Phi y e^y e^{-\Phi e^y} \, dy. \tag{S28}$$

Take $e^y = u$

$$E[\tilde{Y}] = \int_0^{\infty} \Phi \log(u) e^{-\Phi u} \, du \tag{S29}$$

Take $\Phi u = v$

$$E[\tilde{Y}] = \int_0^{\infty} (\log(v) - \log(\Phi)) e^{-v} dv \tag{S30}$$

$$= -\log(\Phi) + \int_0^{\infty} \log(v) e^{-v} dv \tag{S31}$$

$$= -\log(\Phi) - \gamma, \tag{S32}$$

where the second expression is an integral known to evaluate to $-\gamma$ ($\gamma \approx 0.577$ is the Euler-Mascheroni constant). We can see that the log-timestamp only has a constant bias away from the true log-timestamp ($= \log(1/\Phi)$), which can be removed separately.

We repeat the same ideas for calculating the variance:

$$E[\tilde{Y}^2] = \int_{-\infty}^{\infty} y^2 \Phi e^{-\Phi e^y} e^y \, dy \tag{S33}$$

Take $e^y = u$ again:

$$E[\tilde{Y}^2] = \int_0^{\infty} \Phi (\log^2 u) e^{-\Phi u} \, du \tag{S34}$$

and $v = \Phi u \implies \log u = \log v - \log \Phi$. Then

$$\begin{aligned} E[\tilde{Y}^2] = &\int_0^\infty \log^2(v) e^{-v}\, dv \\ &- 2\log(\Phi) \int_0^\infty \log(v) e^{-v}\, dv \\ &+ \log^2(\Phi) \\ =&\gamma^2 + {}^{\pi^2}\!/_6 + 2\gamma \log(\Phi) + \log^2(\Phi), \end{aligned} \tag{S35, S36}$$

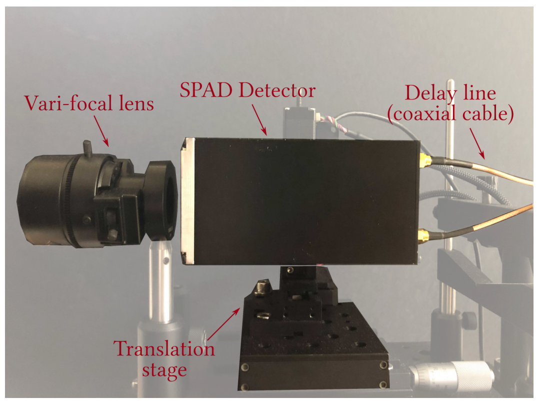
where the first term on the right-hand side is also a known standard integral. Finally we have

$$\mathrm{Var}(\tilde{Y}) = E[\tilde{Y}^2] - E[\tilde{Y}]^2 = {}^{\pi^2}\!/_6, \tag{S37}$$

which proves the homoskedasticity of $\tilde{Y} = \log(Y)$.

Supplementary Note 4. Hardware Prototype

Our hardware prototype (5) consists of a single SPAD pixel mounted on two translation stages. Dead-time is controlled using a long cable that produces analog delay. After each photon detection event the SPAD pixel has to be kept disabled for few tens of ns to lower the probability of afterpulses [12] and reset its original bias condition. Usually this dead-time is set either using the discharge time of an R-C network or employing a digital timing circuit, since the dead-time accuracy is not a limiting factor in conventional SPAD applications. In case of dead-time defined using digital timing circuits, there are implementations where its accuracy depends on the period of an uncorrelated (with respect to photon arrival times) digital clock. For example, a $100\ \mathrm{MHz}$ clock frequency will limit the accuracy of the dead-time to about $10\ \mathrm{ns}$, which is too coarse to get reliable photon flux estimates. This is true especially at extremely high photon flux values where photons get detected almost immediately after each dead-time duration ends. As described in the main text, we rely on low-jitter voltage comparators and analog delays introduced by long coaxial cables to obtain precisely controlled dead-time durations with low jitter.



Suppl. Fig. 5: The IP-SPAD hardware prototype consists of a single SPAD pixel mounted on translation stages to scan the image plane of a vari-focal lens (Fujinon DV3.4x3.8SA-1). Part of a $20\,\mathrm{m}$ long co-axial cable used for generating the dead-time delay is also shown.

Supplementary Note 5. Pixel Non-idealities

When conducting experiments with our hardware prototype we found two non-idealities: dead-time drift and non-zero gate rise time.

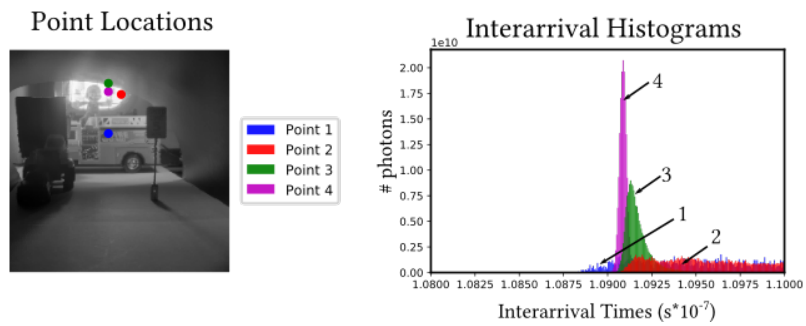
Dead-time Drift

When imaging high flux regions for extended periods of time our hardware prototype's dead-time increases; we call this *dead-time drift*. This is due to heating of the SPAD front-end. We calibrated each pixel position individually by constructing an inter-photon timing histogram and using the first non-zero bin of this histogram as an estimate of the true dead-time for that pixel position. Experimentally, we observed that the dead-time drift is slower than the 5ms exposure times used so this method should approximate the true dead-time well for each pixel position. Without this correction the error introduced by the drift dominates the denominator in Eq. (2) at high flux values, limiting the dynamic range.

When our single-pixel IP-SPAD stays active for long periods of time dead-time drift becomes a problem. Suppl. Fig. 6 shows inter-photon timing histograms of four different scene points with increasing flux levels ($1 \to 4$). Notice that the histograms are not aligned on the left edges indicating a difference in dead-times at these points. We correct for the dead-time drift in by using a dead-time estimate for each pixel in the final image. We set the dead-time estimate in a pixel to the smallest inter-arrival time from that pixel, this has the effect of shifting each pixel's inter-arrival histogram to zero. In the tunnel scene the difference between the longest used and shortest used dead-time is $887\,\mathrm{ps}$, a variation of about 0.8%.

Gate Rise Time

When the SPAD enters and exits the dead-time phase, its bias voltage has to be quickly changed from above to below the breakdown value, and vice-versa [8]. The duration of these transitions is as critical as the dead-time duration itself, and has



Suppl. Fig. 6: This figure shows inter-photon histograms for 4 points from the tunnel scene. Notice that the histograms are not aligned on the left edge, indicating a drift in dead-time. We correct for this drift by taking the time of the first non-zero bin as the dead-time for that pixel.

to be short (in order to swiftly restore the SPAD bias for the next detection) and precise (to prevent variations in the overall dead-time duration). In our system the rise times are on the order of 200 ps: it translates into non-exponentially shaped inter-photon timing histograms, especially in high flux regions. We did not find that this behavior detrimentally effected our results; however, it has an effect similar to slightly tone mapping bright regions downward.

Unlike dead-time drift, the rise time behavior seems to be independent of how long the SPAD was exposed to a high flux source. Fig. 7 shows inter-photon timestamp histograms for increasing photon flux levels. Rise time causes these to deviate from an exponential shape at high flux levels.

We found that this behavior made it virtually impossible to fully saturate the SPAD pixel, that is increasing the incident flux would lead to a non-linear increase in photons counted. We performed an experiment where a laser was directly pointed into the SPAD active region and the power of the laser was increased. We found that the photon counts did not saturate before the SPAD overheated and shut itself off.

The rise-time behaviour can by incorporated into the flux estimator derived in Supplementary Note 1 using a time-varying quantum efficiency $q(t)$. For $t < 0$, $q(t) = 0$ and $\int_0^\infty q(t)dt \to \infty$. When the dead time ends, the IP-SPAD pixel's $q(t)$ ramps up to its peak value. The probability distribution of time-of-darkness, Y_i, can be written as:

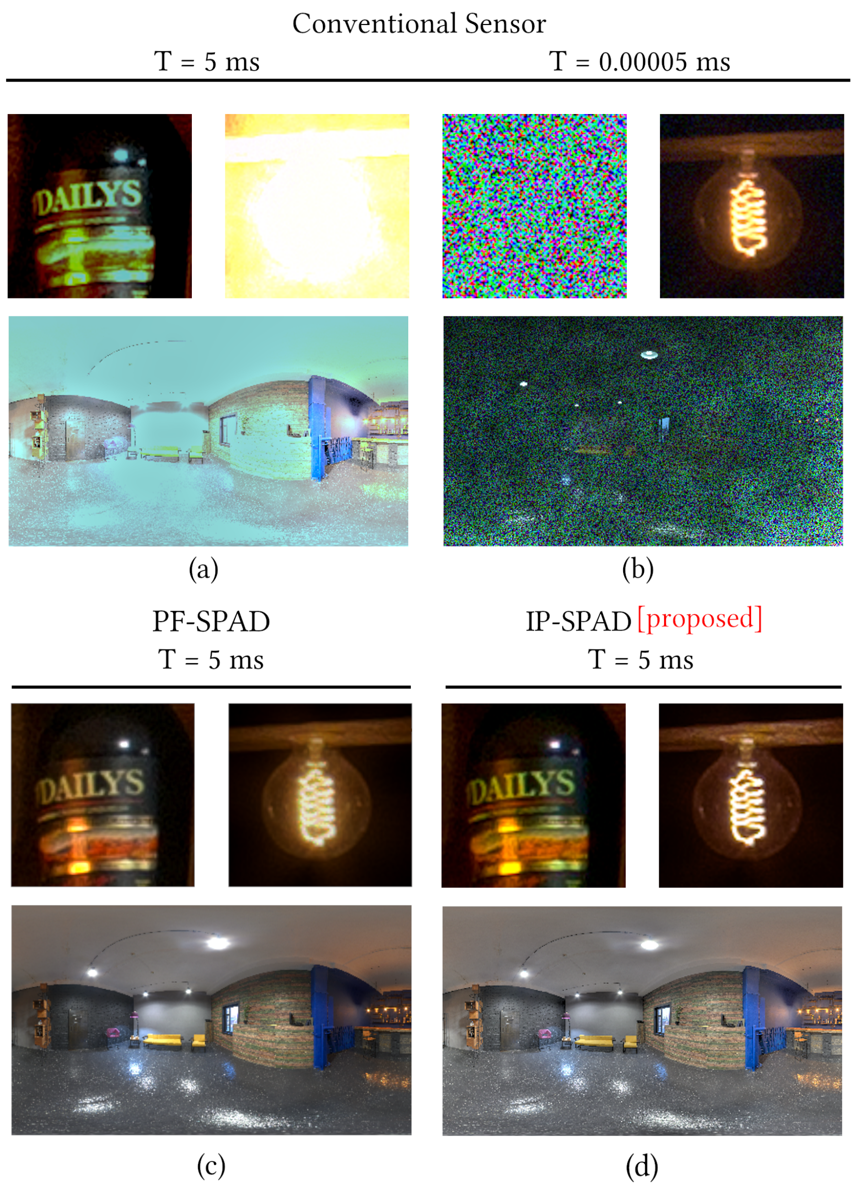
$$Y_i \sim f_{Y_i}(Y_i|Y_1 \ldots Y_{i-1}) = \begin{cases} q(Y_i)\Phi e^{-\Phi \int_0^{Y_i} q(l)dl} & \text{for } 0 < Y_i < T_i \\ e^{-\Phi \int_0^{T_i} q(l)dl} \cdot \delta(t - T_i) & \text{for } Y_i = T_i \\ 0 & \text{otherwise.} \end{cases} \tag{S38}$$

where T_i is defined in Supplementary Note 1. For a series of N timestamps with times-of-darkness given by $Y_1 \ldots Y_N$, we use a similar derivation to Supplementary Note 1 to find the maximum likelihood estimator (MLE):

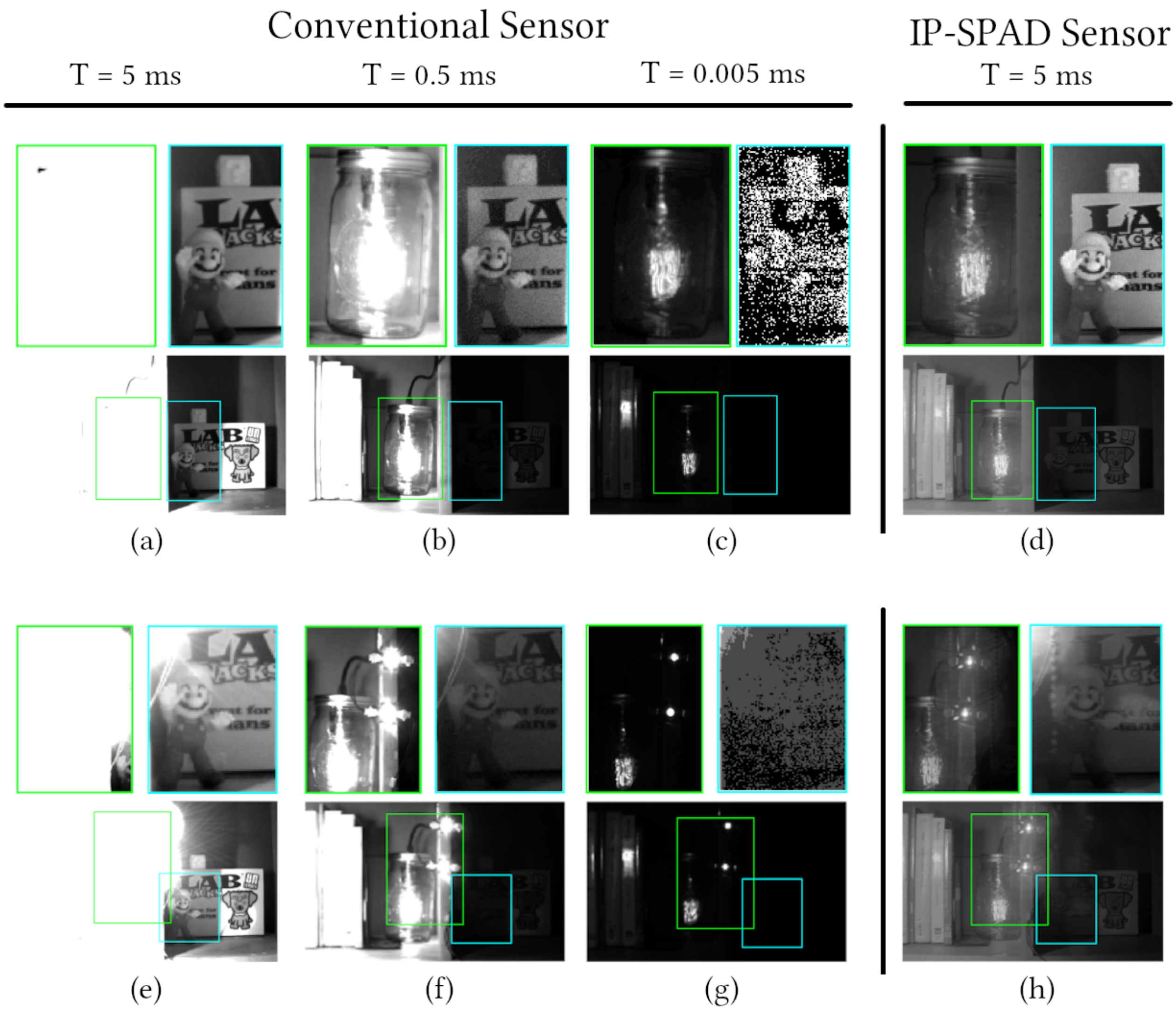
$$\widehat{\Phi} = \frac{N}{\int_0^{T_{N+1}} q(t)dt + \sum_{i=1}^{N} \int_0^{Y_i} q(t)dt}. \tag{S39}$$

Eq. (S39) reduces to Eq. (S6) if $q(t)$ is an ideal step function. For the experimental results shown in the main text, the IP-SPAD pixel's $q(t)$ was not precisely known so we could not apply this correction. Future work will look at estimating $q(t)$ from inter-photon histograms and quantifying SNR improvements from such a correction.

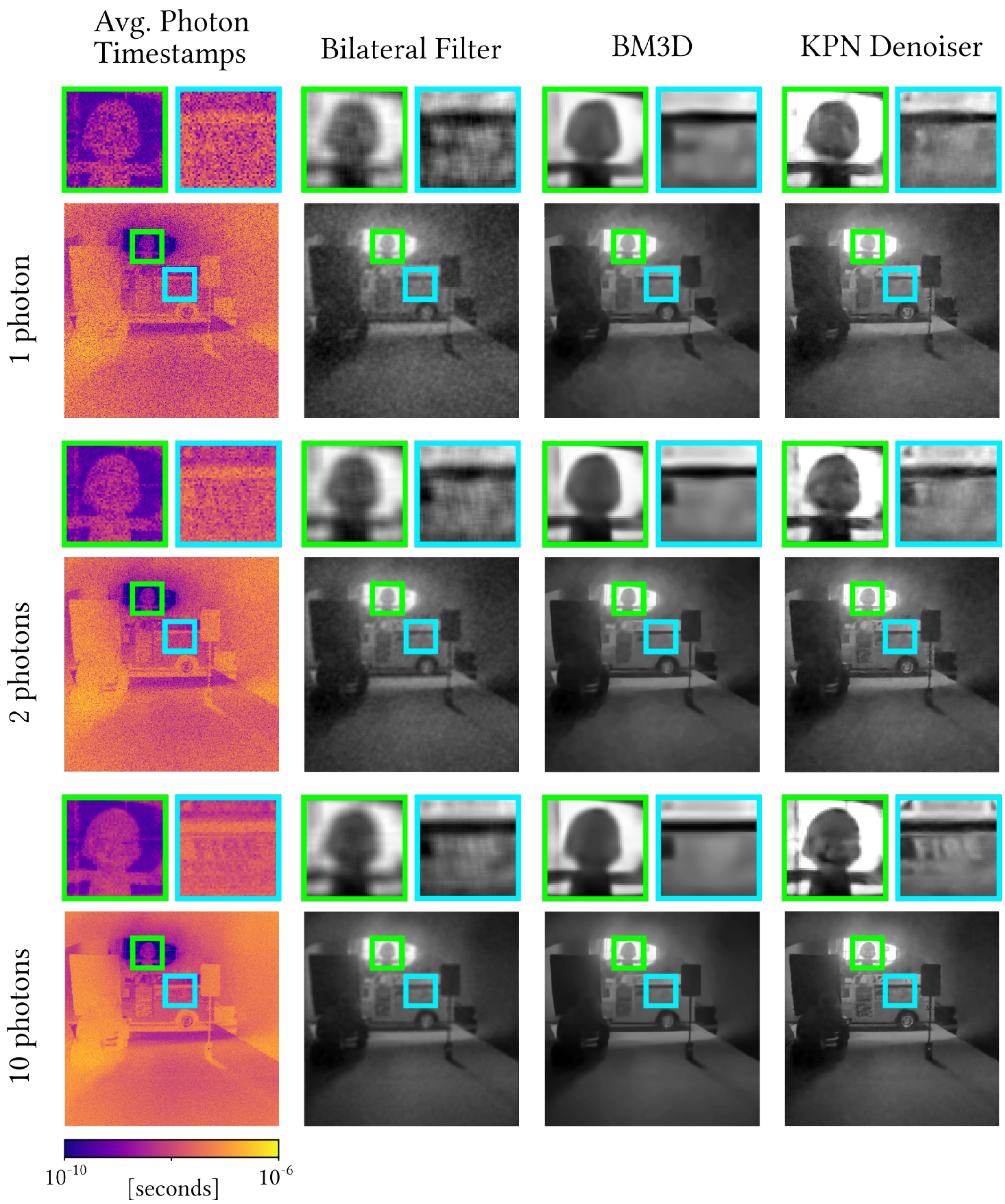
Supplementary Note 6. Additional Results



Suppl. Fig. 7: **Simulated Extreme Dynamic Range Scene:** This figure shows simulated extreme dynamic range images using an IP-SPAD camera compared with a conventional camera with different exposure settings. (a) A $5\,\mathrm{ms}$ exposure image with a conventional camera (full-well capacity 34,000 and read noise $5e^{-}$ has many saturated pixels. Observe that the bright bulb region is washed out. (b) A short exposure image is dominated by shot noise in darker parts of the scene. It becomes visible only at a much lower exposure setting. Since this is a simulation we were able to reduce the exposure time down to $5 \times 10^{-5}\,\mathrm{ms}$ which may be impossible to achieve with a conventional camera. In practice, this exposure can be achieved by, say, reducing the shutter speed to 1/16,000 and adding a 10-stop ND filter. (c) A PF-SPAD camera is able to capture both dark and bright regions in a single exposure, but the bright bulb filament still suffers from noise due to the soft-saturation phenomenon. (d) Our proposed IP-SPAD method estimates scene brightness using high-resolution timestamps to capture both extremely dark and extremely bright pixels, beyond the soft-saturation limit of a counts-based PF-SPAD. (Original image from HDRIHaven.com)

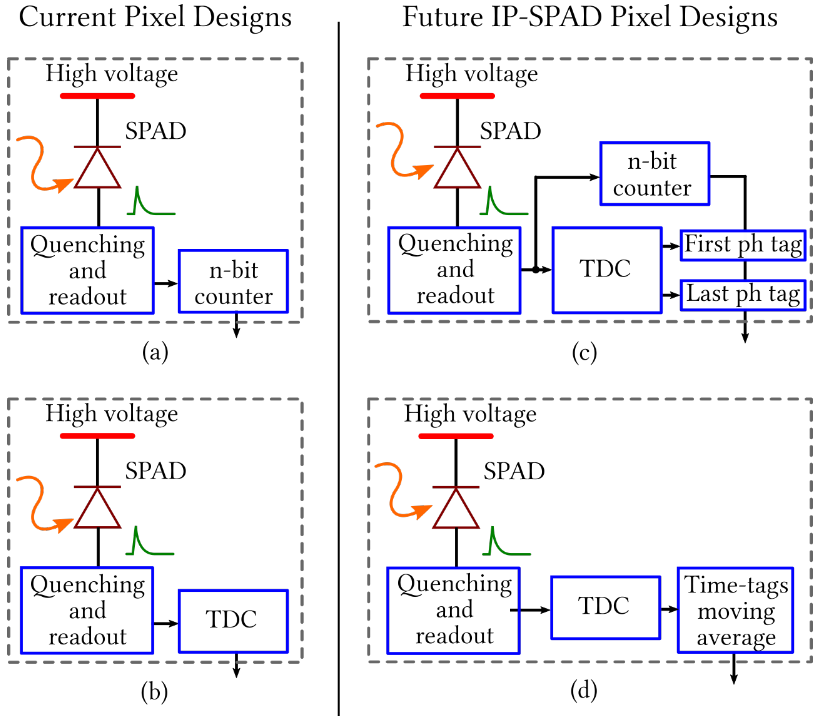


Suppl. Fig. 8: **Experimental Extreme Dynamic Range "Shelf" Scene:** This "Shelf" scene shows extreme dynamic range, with a bright bulb filament in one of the shelves and text in the neighboring shelf which is dark and not directly illuminated by the light source. The bottom row of images uses a similar setup as the top row but also includes two bright LED lights in addition to the filament bulb. The conventional camera requires three exposures to cover the dynamic range of this scene. The proposed IP-SPAD flux estimator captures the scene in a single exposure.



Suppl. Fig. 9: **Effect of Increasing Number of Photons on Denoising.** Some image details start appearing with as few as 10 photon timestamps per pixel. For example, the text on the fire-truck is visible with images denoised with the bilateral filter and our KPN-based denoiser. BM3D appears to give less noisy results in this example but finer details are lost.

Supplementary Note 7. Pixel Designs for Passive SPAD Imaging



Suppl. Fig. 10: **IP-SPAD pixel designs for passive imaging:** (a) and (b) are existing SPAD pixel designs with counts and in-pixel timing circuits. (c) and (d) are hypothetical future pixel designs for passive IP-SPAD cameras that store individual photon timestamps or compute summary statistics on the fly.

Passive SPAD Pixel Architectures Many current SPAD pixel designs are targeted towards specialized active imaging applications that operate the detector in synchronization with a light source, such as pulsed laser. The most common data processing task is to generate a *timing histogram* which counts the number of photons detected by the SPAD pixel as a function of the (discretized) time delay since the transmission of the most recent laser pulse. The requirements for the passive imaging technique shown in this paper are different: there is no pulsed light source to provide a timing reference. Instead, it is important to precisely control (1) the dead time duration (2) rise and fall times of the SPAD bias circuitry, and (3) the duration of the global exposure time.

Our single-pixel IP-SPAD hardware prototype, although acceptable as a proof-of-concept, is not a scalable solution for sensor arrays. Delay-locked loops circuits suitable for multi-pixel implementation can be used in the future to precisely control the dead-time duration. A large array of IP-SPAD pixels will generate an unreasonably large volume of raw photon timestamp data that cannot be transferred off the sensor chip for post-processing. A megapixel SPAD array has been recently demonstrated using a 180 nm CMOS technology [39], but the in-pixel electronics is currently limited to gating circuitry and a 1-bit data register. The trade-off between SPAD performance and pixel number can be overcome by recently-developed 3D-stacking approaches where SPAD arrays are fabricated in a dedicated technology, the high density data-processing electronics are developed in scaled technology, and then the two chips/wafers are mounted one on top of the other [23, 10].

Fig. 10(a) shows the simplest single-pixel architecture currently used as a building-block in large SPAD arrays. It comprises the photodetector, its readout and quenching circuits and a digital counter, for storing the number of detected photons. While this architecture is widely used [6], it does not exploit photon arrival times to increase dynamic range. As shown in Fig. 10(b), adding an in-pixel time-to-digital converter (TDC) able to acquire and store individual photon time-stamps (with respect to the exposure time synchronization signal) can solve this limitation. Also this approach is nowadays quite common when designing SPAD arrays [24, 44], however, increasing the array dimension and considering a very high incident photon flux, it will be impractical to acquire and transfer timestamps for each photon and each pixel, because it will lead to intractable volume of data to be processed. Instead, a more efficient way of storing and transmitting photon time-stamp data for passive imaging can rely on simply storing the first and last photon time-stamps within a single exposure time, together with the total photon counts. The corresponding pixel design is shown in Fig. 10(c). While this increases pixel complexity over the previous SPAD pixel design examples, it only requires two additional data registers. The disadvantage of this scheme is that, depending on the total exposure time, the TDC may require a large full-scale range. For example, using an exposure time in the millisecond range and the timestamp resolution in picoseconds, the TDC data depth will be $\log_2(10^{-3}/10^{-12}) \approx 30$ bits.

Note that our brightness estimator keeps track of the average time-of-darkness between photon detections over a fixed exposure time. An alternative to storing first and last timestamps may be to instead store a running average of the inter-photon times, as shown in Fig. 10(d). This can be implemented in-pixel using basic digital signal processing circuits. At high photon flux levels, the expected inter-photon times will be short enough that a TDC with smaller full scale range could be used. Although the inter-photon times may still be quite long for low flux levels, the flux estimator can fall back to using photon counts only, instead of timestamps.

SPAD Array Designs for Passive Imaging The theoretical analysis and experimental results in this paper were restricted to a single SPAD pixel. For most passive imaging applications, in practice, there will be a need to scale this method to large form factor SPAD arrays with thousands of pixels. This will introduce additional design challenges and noise sources not discussed in this work. A large form-factor SPAD array of free-running SPAD pixels will generate an unreasonably large volume of raw photon time-stamp data that cannot be simply transferred off the sensor chip for post-processing. For instance, consider a hypothetical 1 megapixel SPAD array consisting of pixels shown in Fig. 10(b), with dead time of 100 ns. Assume an average photon flux of 10^5 photons/s over the pixel array and the pixels generate 32-bit IEEE floating-point timestamps for each detected photon. This corresponds to 400 GiB/s of data generated from the chip. A megapixel SPAD array has been recently demonstrated using a 180nm CMOS technology [39] , but the in-pixel electronics is currently limited to gating circuitry and 1-bit data register (photon detected or not).

One possible solution to overcome this problem could include the design of large arrays using a combination of pixel architectures sketched in Fig. 10, i.e. where only a fraction of pixels would include high resolution TDCs while the rest of the pixels only use photon counters. This will still enable capturing extremely high flux values albeit with reduced spatial resolution. In another solution TDCs are shared among more pixels, while counters are integrated in each pixel. This will reduce the maximum count rate, but each detected photon is counted and time tagged.

SPAD performance (i.e. detection efficiency, dark count noise, temporal resolution, afterpulsing probability) in developing multi-pixel arrays is usually better when using "legacy" fabrication technologies, like 350 nm and 180 nm CMOS, or even "custom" technologies (which, however, do not allow the on-chip integration of ancillary electronics) [17]. With such technologies, the relatively large minimum feature size prevents in-pixel integration of sophisticated electronics like high-resolution (few ps) TDCs, data processing circuits and memories (unless without accepting an extremely low fill-factor). The trade-off between SPAD performance and pixel number can be overcome by recently-developed 3D-stacking approaches: SPAD array is fabricated in a dedicated technology, the high density data-processing electronics is developed in scaled technology, and then the two chips/wafers are mounted one on top of the other [23, 10].

Passive IP-SPAD arrays may also require pixel-wise calibration. The non-linear pixel response curve may make this more challenging than conventional CMOS camera pixels. It will be necessary to characterize non-uniformity in terms of dead-time durations and timing jitter and account for these for removing any fixed pattern noise.

Another important practical consideration is power requirement, especially when operating in high flux conditions where a large number of avalanches will be created causing huge power requirement for processing these in real-time and reading out the counts. There is also a significant heat dissipation issue which can exacerbate pixel calibration due to the strong temperature dependence of various pixel parameters like dark count rate and dead-time drifts. Such power issues may be mitigated with scaled technologies operating at lower supply voltage.