Quanta Burst Photography

SIZHUO MA and SHANTANU GUPTA, University of Wisconsin-Madison, USA
ARIN C. ULKU, CLAUDIO BRUSCHINI, and EDOARDO CHARBON, EPFL, Switzerland
MOHIT GUPTA, University of Wisconsin-Madison, USA

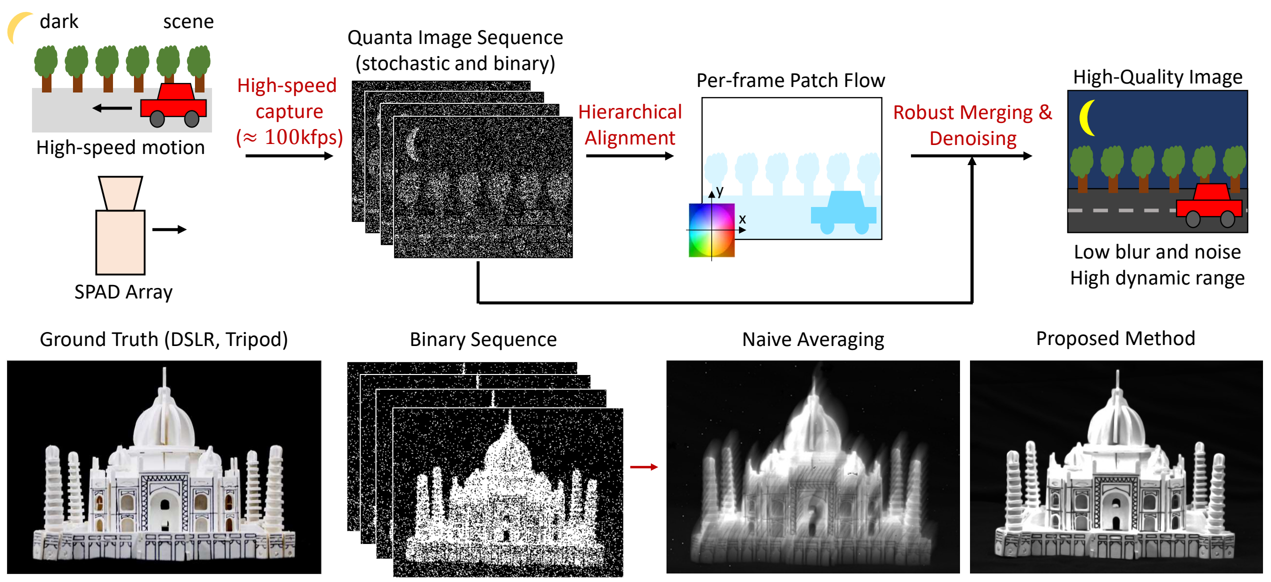

Fig. 1. **Quanta burst photography.** (Top) Single-photon image sensors capture stochastic, binary images at high speeds (~ 100 kfps). Such high-speed image sequences can be aligned to compensate for scene/camera motion using a spatial-temporal hierarchical alignment algorithm. By merging the aligned sequence robustly, a high-quality image can be reconstructed, with minimal motion blur and noise, and high dynamic range, even in challenging photography conditions. (Bottom, from left to right) An example low-light scene captured by a DSLR camera on a tripod; binary image sequence captured by a handheld single-photon camera; image reconstructed by naive averaging of the binary sequence (shown to illustrate the amount of motion during capture); super-resolved image reconstructed using the proposed techniques has low blur and noise. Figure reproduced with permission from [1].

Abstract - Single-photon avalanche diodes (SPADs) are an emerging sensor technology capable of detecting individual incident photons, and capturing their time-of-arrival with high timing precision. While these sensors were limited to single-pixel or low-resolution devices in the past, recently, large (up to 1 MPixel) SPAD arrays have been developed [2]. These single-photon cameras (SPCs) are capable of capturing high-speed sequences of binary single-photon images with virtually no read noise. We present quanta burst photography, a computational photography technique that leverages SPCs as *passive imaging devices* for photography in challenging conditions, including ultra low-light and fast motion. Inspired by recent success of conventional burst photography, we design algorithms that align and merge binary sequences captured by SPCs into intensity images with minimal motion blur and artifacts, high signal-to-noise ratio (SNR), and high dynamic range. We theoretically analyze the SNR and dynamic range of quanta burst photography, and identify the imaging regimes where it provides significant benefits. We demonstrate, via a recently developed SPAD array, that the proposed method is able to generate high-quality images for scenes with challenging lighting, complex geometries, high dynamic range and moving objects. With the ongoing development of SPAD arrays, we envision quanta burst photography finding applications in both consumer and scientific photography.

1 THE SINGLE-PHOTON REVOLUTION

A conventional camera typically captures hundreds to thousands of photons per pixel to create an image. An emerging class of sensors, called single-photon avalanche diodes (SPADs) [3, 4], can record *individual* photons, and precisely measure their time-of-arrival. Due to their sensitivity and picosecond time resolution, SPADs are driving an imaging revolution. A new generation of devices is emerging, with novel functionalities such as imaging at trillion fps [5], non-line-of-sight (NLOS) imaging [6, 7], and microscopic imaging of nano time-scale bio-phenomena [8].

Passive single-photon imaging: So far, most SPAD-based imaging systems are active, where the SPAD is used in precise temporal synchronization with an active light source (e.g., a pulsed laser). This includes applications such as NLOS imaging, LiDAR [9], and microscopy. Can SPADs be used not just with controlled and precisely synchronized active light sources as is the norm, but more generally under passive, uncontrolled illumination (e.g., sunlight, moonlight)? Such passive SPAD-based imaging systems have the potential to expand the scope of SPADs to a considerably larger set of applications, including machine vision and photography.

Consider a SPAD sensor (an array of SPAD pixels) imaging a scene illuminated by passive lighting. Since photons arrive at the sensor randomly according to Poisson statistics, photon detection events are also random, and can be visualized as a spatio-temporal *photon-cube* [10]. A SPAD camera can capture a sequence of thin, equally-sized temporal slices of the photon-cube, where each slice is a binary (1-bit) image, as shown in in Fig. 1. Each pixel location records a 1 if it receives one or more photons during the temporal extent of the slice, and 0 otherwise. For example, a recent SPAD camera [11] can capture $\sim 10^5$ binary frames per second, at 1/4 MPixel resolution. Due to the random nature of photon arrivals, the binary images are stochastic. The light intensity at each pixel can then be recovered from the statistics of photon arrivals throughout a sequence of binary images. Denoising techniques are often applied such as total variation, BM3D [12, 13], and end-to-end neural networks [14, 15].

Authors' addresses: Sizhuo Ma, sizhuoma@cs.wisc.edu; Shantanu Gupta, sgupta226@wisc.edu, University of Wisconsin-Madison, USA; Arin C. Ulku, arin.ulku@epfl.ch; Claudio Bruschini, claudio.bruschini@epfl.ch; Edoardo Charbon, edoardo.charbon@epfl.ch, EPFL, Switzerland; Mohit Gupta, mohitg@cs.wisc.edu, University of Wisconsin-Madison, USA.

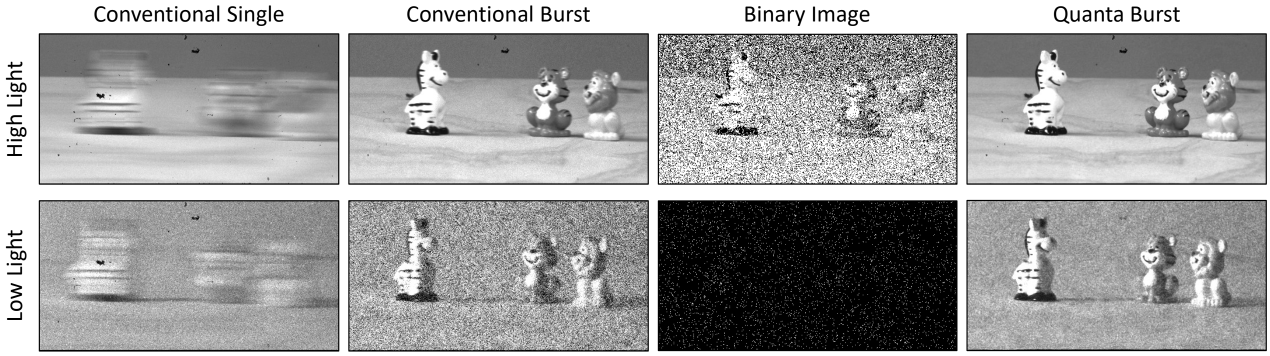

Fig. 2. **Performance under different lighting conditions.** We capture two 2000-frame binary sequences for the same scene under different lighting conditions. A sample binary image from each sequence is shown in the third column. The binary images become sparser as the light level decreases. For conventional cameras, there is a trade-off between motion blur and noise, which makes it difficult to generate a high-quality image in low-light environments, either with a single long exposure (first column) or with a burst (second column). For quanta burst photography, it is possible to resolve fast motion without sacrificing the SNR (fourth column). Even in very low light, a reasonable image is reconstructed by aligning and merging the sparse and noisy binary frames. Figure adapted with permission from [1].

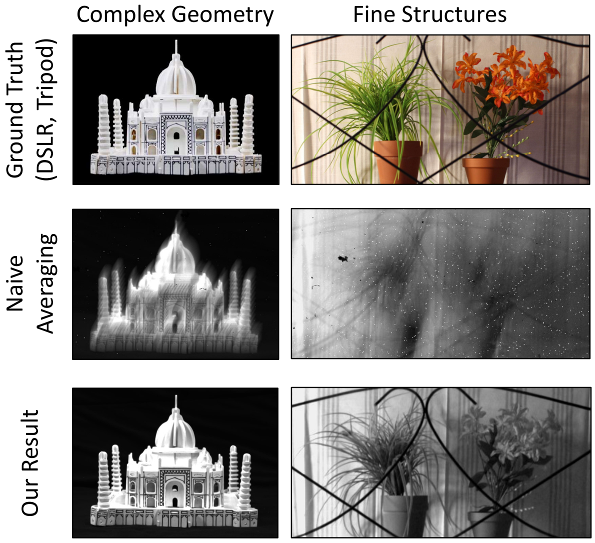

Fig. 3. **Challenging scenes.** Reconstruction results of the proposed method for challenging scenes involving complex scene geometry and fine structures. The camera is handheld and follows a random 6DoF motion. Images are reconstructed from 10000 binary frames. In all cases, the proposed method is able to create a blur-free image with high SNR. Figure adapted with permission from [1].

Photon-cubes and single-photon binary image sequences were first considered in the context of jots [10, 16], another emerging single-photon sensing technology. By avoiding avalanche, jots achieve smaller pixel pitch, higher quantum efficiency and lower dark current, but have lower temporal resolution [13, 17]. In this paper, we primarily focus on SPADs due to their high frame rate. However, since both jots and SPADs have similar imaging model and data format, the analysis and techniques presented here are applicable to jots as well.

Passive single-photon imaging under motion: How does motion manifest in a stochastic binary image sequence? If the scene (or camera) moves during acquisition, the photons emitted by a scene point get spread over multiple SPC pixels, resulting in motion blur. In the presence of motion, Fossum [18] suggested shifting the binary images to compensate for motion and achieve blur-free image reconstruction. This idea has been implemented recently [19–21], albeit for simplistic motion models (*e.g.*, planar objects with in-plane motion and no occlusions).

In this paper, we propose *quanta burst photography*, a computational photography technique that computationally re-aligns the photons along motion trajectories, achieving high-quality images in challenging scenarios including low-light and high-speed motion (Fig. 1). We develop algorithms that align the binary frames by estimating the patch-wise 2D translation between frames. To obtain reliable motion estimates, we first group an entire sequence into temporal blocks and compute a coarse-level motion between the blocks. The motion is then interpolated temporally to get a fine-level motion, which is used to warp and merge the binary frames into intensity images with minimal motion blur and ghosting artifacts, high SNR, and high dynamic range, at up to 2x resolution. Details on the proposed algorithm can be found in [1].

The proposed method is similar in spirit to conventional burst photography where a burst of short-exposure images ($5-10$) are aligned and merged into a single high-quality image [22, 23]. Quanta burst photography can be considered a limiting case because each binary image captures at most one photon per pixel, thus extremely noisy and quantized. Recently, [24] proposed a learning-based approach which takes as input a short sequence of quanta images (8 frames) at photon levels of about 1 photon-per-pixel (ppp). Due to the fast capture of SPADs, quanta burst photography is designed to work with longer binary sequences (10^2-10^5 frames, depending on light level, dynamic range and motion), with light levels down to the order of 0.001 ppp.

2 RESULTS

We capture real data using a SwissSPAD2 camera [11] for the experiments, which takes binary images at a spatial resolution of 512×256 pixels. The maximum frame rate of the camera is 96.8kfps. The camera does not have microlenses and has a native fill factor of about 13% (later built systems do have microlenses equipped). Currently the sensor is not equipped with Bayer filters, so only grayscale images are reconstructed.

Performance for different lighting conditions. Fig. 2 shows the performance of quanta burst photography for different lighting conditions. We choose the same still scene for all sequences. The camera was moved horizontally to ensure the motion is controllable and reproducible across different sequences. The conventional camera images are emulated from the captured binary images by first reconstructing the intensity and then adding the read noise and quantization error according to sensor parameters of a high-end machine vision camera [1]. Quanta burst photography generates images with higher quality than conventional single and burst images. Even in very low light, where the individual binary frames are sufficiently sparse to make it nearly impossible to

---

[1] https://www.flir.com/products/grasshopper3-usb3/?model=GS3-U3-123S6C-C

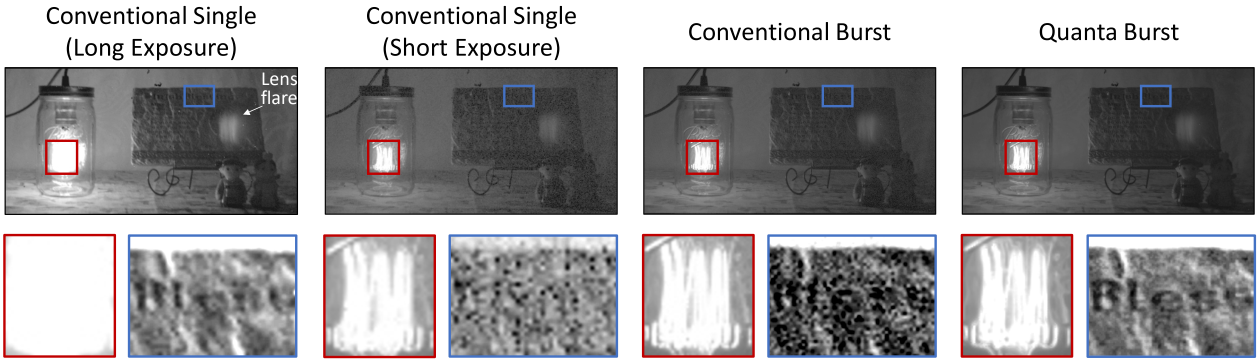

Fig. 4. **Reconstructing high dynamic range scenes.** We capture a scene with high dynamic range where the light source (the lamp) is directly visible in the image. A single conventional image either gets saturated (long exposure) or fails to capture the details in the dark regions (short exposure). Conventional burst photography improves the dynamic range, but remains noisy in the dark regions due to read noise. Quanta burst photography achieves very high dynamic range and is able to recover the details of the filament and the text on the plaque at the same time. 100000 frames are captured to reconstruct the full dynamic range. All images are processed using the same tone-mapping algorithm [25]. Figure reproduced with permission from [1].

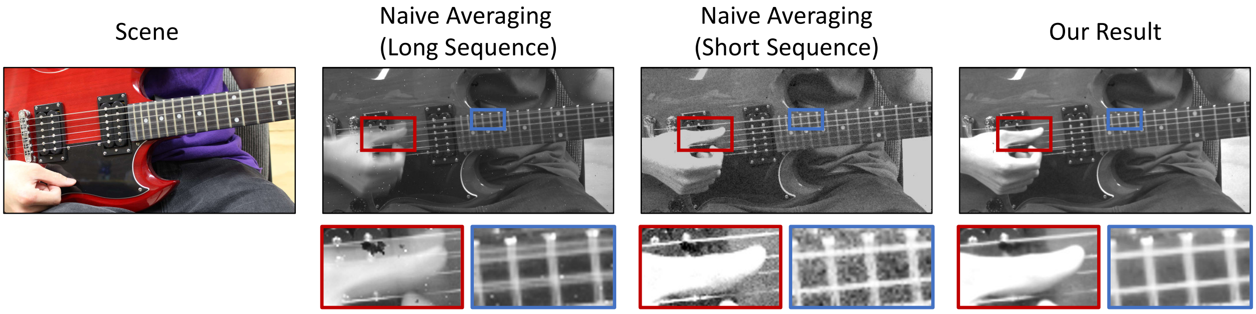

Fig. 5. **Resolving scene motion.** A person plucking the lowest two strings of a guitar. Averaging the captured binary sequence results in either ghosting artifacts (long sequence with 2000 binary frames) or a low SNR (short sequence with 100 binary frames). Our method is able to reconstruct a high-quality image from 2000 frames despite fast and non-rigid scene motion. Figure reproduced with permission from [1].

make out the scene structure, a reasonable image is reconstructed by aligning and merging the sparse and noisy binary frames.

The purpose of this experiment is not to compare a conventional sensor and SPAD directly. In fact, due to the low resolution and low quantum efficiency of current SPAD sensors, SPADs will almost always generate worse-quality images than commercial CMOS sensors. Here we simulate the conventional images by assuming a conventional sensor with the same resolution and quantum efficiency as the SPAD array. Due to the blur-noise trade-off, conventional sensor struggles in reconstructing high-quality images, while SPAD has the potential of super-sampling in time and mitigate motion blur even for low-light and fast-moving scenes.

Reconstructing challenging scenes. Fig. 3 shows challenging scenes involving complex geometry and fine structures. Such scenes are usually challenging for optical flow and block matching algorithms. The camera was handheld, and underwent a random 6DoF motion when capturing the images. Since a long-focus lens is used, even natural hand tremor causes a large apparent motion in the image space. Despite these challenges, the proposed method is able to reconstruct blur-free images with high SNR.

Reconstructing high dynamic range scenes. Fig. 4 shows a high dynamic range scene captured by the SPAD array. The only light source in the scene, the lamp (red box), is directly visible in the image, which is about 2000 times brighter than the text on the plaque (blue box), which does not receive any direct light. Similar as in Fig. 2, we simulate the conventional images by adding read noise and quantization error. With a single capture, the conventional image is either saturated around the lamp, or cannot recover the texts on the plaque. Conventional burst photography improves the dynamic range, but the text is still indiscernible due to read noise. Quanta burst photography is able to recover both the filament and the text at the same time.

Resolving scene motion. Since the proposed method only computes patch-wise motion and does not assume any global motion model, it is capable of resolving scene motion. Fig. 5 shows a person plucking the lowest two strings on the guitar. Simple averaging of binary frames creates ghosting artifacts or strong noise. Our method is able to resolve the plucking motion of the thumb and the vibration of the strings with lower noise.

3 OUTLOOK ON SINGLE-PHOTON SENSORS

In this section, we discuss the current state and future outlook of SPAD sensor arrays, in terms of their key characteristics: spatial resolution, temporal frame rate, photon detection efficiency (PDE), and the dark count rate (DCR).

Spatial resolution. Due to their compatibility with mainstream CMOS fabrication lines, it was predicted in 2008 that SPAD image sensors could reach large resolutions within one decade [26, 27]. In recent years, significant effort has been devoted to achieve this goal, with the world's first 1 MPixel SPAD array reported recently [2]. With the same fabrication process, it is possible to go up to 5-10 MPixel, not far from their counterparts in CMOS imagers in several cell-phone cameras. Can we go even higher (e.g., 50 MPixel) in the long term? The key factor that limits the spatial resolution is the minimum pixel pitch, which in turn is limited by the necessity of placing a *guard ring* [2] around each

---

[2] A SPAD pixel detects single photons by creating an avalanche of photo-electrons (large current) when a photon is incident, and sensing the avalanche current via a

SPAD pixel. In current CMOS technologies, due to the guard ring, SPAD pitch cannot be reduced below 1μm [28]. At that pitch, the guard ring occupies a large portion of the pixel, thus reducing the fill factor to a minimum. This limitation could be addressed via 3D-stacking [29], a potentially effective way to reduce SPAD pixel pitch by moving all the active and passive components associated with a SPAD pixel to the bottom tier of the sensor.

Frame rate and power consumption. The frame rate of a SPAD sensor array is limited by the bit-rate the chip can deliver and by the number of communication channels it can host. For example, a 1 Mpixel camera with a frame rate of 1kfps will generate 1Gbps of data, which can be handled by a single LVDS (low-voltage differential signalling) communication channel. Typically, this kind of channel requires about 10mW of power at full speed. If one wants to increase the frame rate by, say, $100X$, the data rate will increase to 100Gbps, with 1W of power required, which may be prohibitive for consumer devices.[3] Recent SerDes can work at 16Gbps with less than 50mW, which means 128Gbps can be achieved with a dissipation of 400mW. The communication power consumption can be mitigated by performing on-chip image processing operations, and designing more efficient motion computation and image alignment operations that are amenable to on-chip processing. Furthermore, it is possible to exploit the spatio-temporal sparsity in the photon-cube raw data in low-light scenarios. Depending on the light-level in the scene, one could achieve a considerable data rate reduction by compressing raw photon-cube data [30].

Photon detection efficiency (PDE). PDE is defined as the product of the pixel fill factor, and the photon detection probability (PDP), which is the probability that an impinging photon generates a detectable signal. PDP is the product of quantum efficiency and the probability of triggering an avalanche. PDP is dependent on the wavelength of photons; for current devices, the PDP is typically $50-70\%$ at $450-550$ nm. Due to low fill factors, earlier SPAD arrays had PDEs as low as 1% making them highly inefficient due to significant light loss. However, the PDE in recent arrays has increased to approximately 40% by using microlens arrays, which increase PDE by effectively increasing the fill factor. While still lagging the quantum efficiency of conventional sensors (approximately $60-90\%$), the PDE of SPAD arrays will likely improve due to improving fabrication processes, including 3D stacking.

Dark count rate (DCR). DCR is the rate of avalanche counts unrelated to photons, measured in counts-per-second (cps). Earlier SPAD devices were largely considered impractical due to high DCR, up to several tens of cps at cryogenic temperatures, and tens of kcps at room temperature. Fortunately, for current devices, DCR has been drastically reduced to 2 cps [2], even at room temperature. Since SPADs do not have read noise, this DCR is sufficiently low to achieve nearly shot-noise-limited SNR, even in ultra low-light. Since DCR is proportional to the active area, as the pixels become smaller, DCR could be further reduced.

REFERENCES

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

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

[3] Alexis Rochas. *Single Photon Avalanche Diodes in CMOS Technology.* PhD thesis, EPFL, 2003.

[4] Cristiano Niclass, Alexis Rochas, Pierre-André Besse, and Edoardo Charbon. Design and characterization of a CMOS 3-D image sensor based on single photon avalanche diodes. *IEEE Journal of Solid-State Circuits*, 40(9):1847–1854, September 2005.

[5] Matthew O'Toole, Felix Heide, David B Lindell, Kai Zang, Steven Diamond, and Gordon Wetzstein. Reconstructing transient images from single-photon sensors. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1539–1547, 2017.

[6] 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.

[7] Matthew O'Toole, David B. Lindell, and Gordon Wetzstein. Confocal non-line-of-sight imaging based on the light-cone transform. *Nature*, 555:338–341, Mar 2018.

[8] Claudio Bruschini, Harald Homulle, Ivan Michel Antolovic, Samuel Burri, and Edoardo Charbon. Single-photon avalanche diode imagers in biophotonics: Review and outlook. *Light: Science & Applications*, 8(1):87, December 2019.

[9] Dongeek Shin, Feihu Xu, Dheera Venkatraman, Rudi Lussana, Federica Villa, Franco Zappa, Vivek K. Goyal, Franco N. C. Wong, and Jeffrey H. Shapiro. Photon-efficient imaging with a single-photon camera. *Nature Communications*, 7:12046, Jun 2016.

[10] Eric R. Fossum. The quanta image sensor (QIS): Concepts and challenges. In *OSA Topical Mtg on Computational Optical Sensing and Imaging*, Toronto, Canada, 2011.

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

[12] Stanley Chan, Omar Elgendy, and Xiran Wang. Images from Bits: Non-Iterative Image Reconstruction for Quanta Image Sensors. *Sensors*, 16(11):1961, November 2016.

[13] Abhiram Gnanasambandam, Omar Elgendy, Jiaju Ma, and Stanley H. Chan. Megapixel photon-counting color imaging using quanta image sensor. *Optics Express*, 27(12):17298, June 2019.

[14] Joon Hee Choi, Omar A. Elgendy, and Stanley H. Chan. Image Reconstruction for Quanta Image Sensors Using Deep Neural Networks. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6543–6547, Calgary, AB, April 2018. IEEE.

[15] Paramanand Chandramouli, Samuel Burri, Claudio Bruschini, Edoardo Charbon, and Andreas Kolb. A Bit Too Much? High Speed Imaging from Sparse Photon Counts. In *IEEE International Conference on Computational Photography (ICCP)*, pages 1–9, Tokyo, Japan, May 2019.

[16] Eric R. Fossum. What To Do With Sub-Diffraction Limit (SDL) Pixels?—A proposal for a gigapixel digital film sensor (DFS). In *IEEE Workshop on Charge-Coupled Devices and Advanced Image Sensors*, pages 214–217, 2005.

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

[18] Eric R. Fossum. Modeling the Performance of Single-Bit and Multi-Bit Quanta Image Sensors. *IEEE Journal of the Electron Devices Society*, 1(9):166–174, September 2013.

[19] Kiyotaka Iwabuchi, Tomohiro Yamazaki, and Takayuki Hamamoto. Iterative Image Reconstruction for Quanta Image Sensor by using Variance-based Motion Estimation. In *International Image Sensor Workshop (IISW)*, 2019.

[20] Istvan Gyongy, Neale Dutton, and Robert Henderson. Single-Photon Tracking for High-Speed Vision. *Sensors*, 18(2):323, January 2018.

[21] Istvan Gyongy, Tarek Al Abbas, Neale A W Dutton, and Robert K Henderson. Object Tracking and Reconstruction with a Quanta Image Sensor. In *International Image Sensor Workshop (IISW)*, page 5, 2017.

[22] 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 Transactions on Graphics*, 35(6):1–12, November 2016.

[23] Orly Liba, Ryan Geiss, Samuel W. Hasinoff, Yael Pritch, Marc Levoy, Kiran Murthy, Yun-Ta Tsai, Tim Brooks, Tianfan Xue, Nikhil Karnad, Qiurui He, Jonathan T. Barron, and Dillon Sharlet. Handheld mobile photography in very low light. *ACM Transactions on Graphics*, 38(6):1–16, November 2019.

[24] Yiheng Chi, Abhiram Gnanasambandam, Vladlen Koltun, and Stanley H. Chan. Dynamic low-light imaging with quanta image sensors. In Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm, editors, *European Conference on Computer Vision (ECCV)*, pages 122–138, Cham, 2020. Springer International Publishing.

[25] Michael Ashikhmin. A Tone Mapping Algorithm for High Contrast Images. In *Eurographics Workshop on Rendering*, pages 145–156, 2002.

[26] Edoardo Charbon. Will Avalanche Photodiode Arrays Ever Reach 1 Megapixel? In *International Image Sensor Workshop (IISW)*, 2007.

[27] Edoardo Charbon. Towards large scale cmos single-photon detector arrays for lab-on-chip applications. *Journal of physics D: Applied physics*, 41(9):094010, 2008.

[28] Kazuhiro Morimoto and Edoardo Charbon. High fill-factor miniaturized SPAD arrays with a guard-ring-sharing technique. *Optics Express*, 28(9):13068, April 2020.

[29] Juan Mata Pavia, Mario Scandini, Scott Lindner, Martin Wolf, and Edoardo Charbon. A 1 × 400 Backside-Illuminated SPAD Sensor With 49.7 ps Resolution, 30 pJ/Sample TDCs Fabricated in 3D CMOS Technology for Near-Infrared Optical Tomography. *IEEE Journal of Solid-State Circuits*, 50(10):2406–2418, October 2015.

[30] Chao Zhang, Scott Lindner, Ivan Michel Antolović, Juan Mata Pavia, Martin Wolf, and Edoardo Charbon. A 30-frames/s,252x144 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, 2019.

---

comparator or a high-gain amplifier. A guard ring is a region around each SPAD pixel that forces the avalanche to be confined in the region, in order to prevent edge breakdown. Guard rings are implemented via geometric structures that are not sensitive to light.

[3] This assumes that the internal power dissipation due to SPADs and chip operation is negligible, and that the readout speed of the pixels internally is not the bottleneck.