3-2021

A Comparative Evaluation of the Fast Optical Pulse Response of Event-based Cameras

Tyler J. Brewer

A COMPARATIVE EVALUATION OF THE FAST OPTICAL PULSE RESPONSE OF EVENT-BASED CAMERAS

THESIS

Tyler J. Brewer, Capt, USAF

AFIT-ENP-MS-21-M-103

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT-ENP-MS-21-M-103

A COMPARATIVE EVALUATION OF THE FAST OPTICAL PULSE RESPONSE OF EVENT-BASED CAMERAS

THESIS

Presented to the Faculty

Department of Engineering Physics

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

in Partial Fulfillment of the Requirements for the

Degree of Master of Science in Applied Physics

Tyler J. Brewer, B.S.

Capt, USAF

March 2021

AFIT-ENP-MS-21-M-103

A COMPARATIVE EVALUATION OF THE FAST OPTICAL PULSE RESPONSE OF EVENT-BASED CAMERAS

THESIS

Tyler J. Brewer, B.S.
Capt, USAF

Committee Membership:

Michael Hawks, Ph.D.
Chair

Peter McMahon-Crabtree, Ph.D.
Member

Anthony Franz, Ph.D.
Member

Brian McReynolds, M.S.
Member

Abstract

Event cameras use biologically inspired readout circuit architecture to offer a faster and more efficient method of imaging than traditional frame-based detectors. The asynchronous event reporting circuit timestamps events to 1 microsecond resolution, but latency increases when many pixels are stimulated simultaneously. To characterize this variability, the DAVIS240, DAVIS346, DVXPlorer, and Prophesee Gen3M VGA-CD 1.1 cameras were exposed to single step-function flashes with amplitudes from 9.3-771cd/m^2, stimulating from 0.0042-100% of pixels. The Median Absolute Deviation of pixel response times ranged between 0 and 6086μs, increasing with the percent of pixels stimulated (PSP). The number of events generated per pixel generally decreased with increasing PSP, with all cameras producing fewer than 59 events per pixel. Surprisingly, as stimulus amplitude increased, the DVXPlorer generated fewer events, to as low as 0.32 events per stimulus. Short-term throughput exceeded advertised limits in 3 of 4 cameras. While individual pixels may be able to accurately detect microsecond-scale change, data bottlenecks can cause missed events or erroneous timestamps.

Acknowledgements

I would like to thank Dr. Hawks for his guidance throughout this effort. I particularly appreciate his encouragement of autonomy, granting the opportunity to try my own ideas and allowing me to find what doesn't work along the way to finding what does. His leading questions and encouragement to find my own way helped me grow as a scientist and officer. I would also like to thank my committee and the rest of the neuromorphic collaboration group for their availability and willingness to discuss progress throughout the project. The interesting discussions and brainstorming hopefully advanced Air Force interests across the country; I know they had a significant impact on the quality of this research. Finally I would like to thank my family for supporting and encouraging my lifelong interest in science. I hope you enjoy the following light reading.

Tyler J. Brewer

Contents

Page

Page

List of Figures

List of Tables

List of Acronyms

AER Address Event Representation

CCD Charge-Coupled Device

CMOS Complementary Metal-Oxide-Semiconductor

DAVIS Dynamic and Active Pixel Vision Sensor

DN Digital Number

DVS Dynamic Vision Sensor

EPS Events per Stimulus (per pixel)

fL foot-Lambert

FPA Focal Plane Array

FPD fast photodiode

MAD Median Absolute Deviation

Meps Million Events per Second

PSP Pixel Stimulation Percent

ROI Region of Interest

ROIC Readout Integrated Circuit

A COMPARATIVE EVALUATION OF THE FAST OPTICAL PULSE RESPONSE OF EVENT-BASED CAMERAS

I. Introduction

1.1 Motivation

Soon after 19th Century engineers first invented the camera, the military saw its utility. Early photographers documented battle damage and aftermath, though few actual photographs of combat were taken due to technical limitations [1]. These limitations likely included the long duration necessary to expose photographic plates, leaving the equipment and photographer vulnerable to enemy fire. Early cameras were also large and heavy, likely making transportation through combat arduous. As camera technology developed, cameras became smaller and more sensitive, leading to the wide-spread adoption of aerial imaging during World War I. This new capability allowed collection and dissemination of intelligence such as troop locations and trench maps, changing the way commanders approached battle [2].

Further technological development through World War II brought high-altitude imaging and the dawn of space-based remote sensing, as camera weight decreased and image quality improved. The introduction of digital camera sensors, arrays of Charge-Coupled Devices (CCDs), in the second half of the 20th century eliminated the need for film as images could be digitized and transmitted wirelessly. The industry continues to improve image quality by decreasing pixel area, increasing sensor size, improving reliability, and advancing data processing algorithms. However, certain fundamental challenges remain, and conventional frame-based imaging systems may

not offer the best solution in all situations.

Event-based, or neuromorphic, cameras record activity in a scene using a fundamentally different sensing paradigm than conventional cameras. When recording video, digital frame-based cameras record the value of every pixel in every frame, and each new frame is recorded at a preset interval. Event cameras, on the other hand, only records pixels when they detect a change in illumination. When the irradiance on a pixel increases or decreases by a preset amount, the pixel records an "event": a data point including the pixel's location on the sensor, the time at which the event occurred, and the polarity of the change (an increase or decrease). Because pixels are not controlled by an external clock as in conventional cameras, they are considered an "asynchronous" imaging system. This novel method of detecting and recording information has several significant benefits.

In order to capture high speed motion, a conventional camera must have a high frame rate, consuming significant electrical power and producing a very large amount of data. Event cameras are able to record events seperated by only microseconds, meaning very fast motion can be recorded accurately, without recording thousands of redundant pictures. This reduces data processing and storage requirements by several orders of magnitude. Additionally, high-speed cameras are very impractical for recording low-speed activity due to their excessive power consumption and data generation rates, while low speed cameras cannot accurately record high speed activity. This requires multiple cameras to be used if a wide variety of activity is expected. Event cameras, due to their asynchronous readout, are able to accurately record high-speed and low-speed activity simultaneously.

Additionally, if a camera is in a location that requires wireless transmission, such as flying on a satellite, every reduction in data also reduces the bandwidth required to transmit information to the user. Modern video processing and compression algo-

rithms calculate differences between frames and delete the information which does not change between images, in order to minimize redundant data. While this has been an effective and popular technique, it requires additional computational resources after the images have been recorded and moved off-chip. By only detecting and recording changes in a scene, event-based cameras eliminate the need for compression. Finally, rather than storing charge in an electron well like conventional cameras, leading to the possibility of under- and over-exposure, event cameras indirectly measure the photocurrent from each pixel. This increases the dynamic range of event cameras by a significant amount, enabling users to record activity in very dark and very bright conditions simultaneously. Event-based cameras provide a powerful new capability to the imaging community.

1.2 Industry Activity

The benefits introduced above make this new technology appealing to a variety of industries and use cases. As such, several companies are developing sensors. iniVation is a Swiss company that released their first sensor, a 128x128 pixel sensor in 2008. In 2014 they released a 240x180 sensor called the DAVIS240, which was capable of recording traditional frames in addition to events. Their DAVIS346 was introduced in 2017, containing a 346x260 pixel sensor. [3] Their most recent release was the DVX-Plorer and DVXPlorer Lite in early 2020 [4], with a VGA-resolution 640x480 array developed in partnership with Samsung [5]. Samsung also independently released an event camera for direct-to-consumer sales in limited markets in 2020, targeting the home security market [6]. They also recently announced a 1280x960 pixel Gen4 sensor which offers an expanded event-holding process[7]. Prophesee is a French company that released their first sensor, a 304x240 array, in 2011. A Gen3 640x480 sensor was released in 2017, followed by a Gen4 camera featuring a 1024x720 sensor in early 2020

[8]. CelePixel is a Chinese company that has released two sensors, in 2017 and 2019. Insightness is a Swiss company which has released one sensor in 2018, and has since partnered with Sony [3, 9]. Notably, while the foundational technology was developed at CalTech[10], no American companies have marketed or sold any event-based cameras to date. This may change in the next several years as government funding becomes available for domestic development.

1.3 Problem Statement

The number of manufacturers and available camera models is expanding, and defense and commercial applications will continue to grow as the technology becomes more mature. For users to make informed decisions when seeking to buy an event camera for their application, they need reliable and consistent metrics by which to compare competing options. Conventional metrics for comparing frame-based camera sensors do not apply due to fundamental differences in circuit operation[3], and relevant event-based metrics are still inconsistent across the industry, likely due to the newness of the commercial market. Additionally, physics-based models of pixel behavior are being developed, but sensor packages are more than just a linear combination of pixels: readout circuits controlling the flow of events are a critical and still-developmental component of asynchronous architecture. Metrics used to describe these sensors must therefore take sensor-level behavior into account, in addition to the characteristics of individual pixels. To generate a sensor-level model, the entire sensor needs to be evaluated, and metrics describing that sensor need to apply to the whole system.

This thesis starts the process of characterizing devices and generating full-sensor models. Using an optomechanical setup adapted from recent work by others [11, 12], the proposed characterization process generates metrics which enable direct compar-

ison between multiple event cameras. The methodology is applied to three cameras from iniVation (the DAVIS240, DAVIS346, and DVXPlorer) and one from Prophesee (the Gen3M VGA-CD 1.1 with PPS3MVCD sensor).

1.4 Document Overview

Chapter two introduces several key technical concepts which are necessary for understanding the work accomplished in this thesis. It discusses the operating principle behind event-based pixel circuits, and describes how data will be represented in various ways to emphasize different aspects of the information contained therein. Chapter three describes the physical setup and procedures followed to gather data and the post-processing workflow is presented. Chapter four then presents the data in its various representations, which is interpreted in chapter five. Finally, chapter six reports conclusions and describes the significance of this work. A list of acronyms is provided at the beginning of the document for reference.

II. Background

2.1 Overview

This chapter begins by introducing radiometric and photometric terms. Section 2.3 introduces the operation of CCD and Complementary Metal-Oxide-Semiconductor (CMOS) pixels, the traditional digital imaging technologies. Pixels in event-based cameras use a fundamentally different architecture for transducing optical information into digital data, and the most important concepts are described in Section 2.4. A detailed explanation of individual circuits is beyond the scope of this document, as their complexities require a dedicated effort to understand. The purpose of this thesis is for comparing the overall behavior of different cameras, so the electronic details are left for other works.

Due to the distinct nature of data produced by event cameras compared to traditional cameras, new techniques for processing, viewing, and ultimately extracting useful information from that data are necessary. The specific techniques used in this work are introduced in Section 2.5. A few recent publications have been critical to developing a foundation and theoretical support for the experimental process, interpretation, and understanding of this effort. Those documents are discussed in Section 2.6 to provide a thorough jumping-off point for this research.

2.2 Radiance vs Luminance

Before camera operation can be discussed, it is important to clarify several terms which are sometimes used differently by various industries: radiance and luminance. Radiance is the optical power per emitting area per solid angle with the symbol L_e (for "energetic") and the units of $W/m^2/sr$ [13]. Spectral radiance is the radiant power per wavelength of emission, which produces total radiance when integrated

over all wavelengths. Luminance L_v is the spectral integral of spectral radiance $L_{(e,\lambda)}$ scaled by the spectral response of the human eye (the luminosity function) $\bar{y}(\lambda)$

$$L_v = 683.002lm/W * \int_0^\infty \bar{y}(\lambda)L_{(e,\lambda)}(\lambda)d\lambda \tag{1}$$

and is thus considered a "photometric" unit, as opposed to the power-based radiant units. Luminance effectively describes how bright something appears to the human eye. The candela (cd) is the SI unit of luminous intensity, which is the photometric equivalent of the more familiar radiant unit of watts/steradian. Thus, luminance can be expressed as cd/m^2 and can be imagined as a measure of the brightness being emitted by a differential surface area [14].

Radiance can be accurately converted to luminance through the evaluation of Equation 1 without any additional information, but luminance cannot be converted to radiance without knowing the exact spectral emission function $L_{(e,\lambda)}$ of the source. For a light source of fixed spectral emission such as an LED in the visible band, radiance and luminance scale proportionally, so changes in luminance also describe proportional changes in radiance. Assuming the green LEDs used in this experiment were monochromatic with $\lambda = 530nm$ and approximating $\bar{y}(530nm) \approx 1$, one can approximate Equation 1 as

$$L_v = 683.002lm/W * 1 * L_{(e)}. \tag{2}$$

However, because the true spectral radiance of the LEDs was not able to be measured due to laboratory hardware limitations, assuming monochromaticity reduces precision without adding significant value. Therefore this document reports luminance in cd/m^2.

Finally, radiance and luminance both refer to light leaving or passing through

some surface. Irradiance and illuminance both specifically refer to light incident onto a surface from some solid angle, such that illuminance $E_v = cd/m^2 * sr$. In this document, then, luminance is a description of the light leaving the light source, while irradiance and illuminance are the amount that actually falls on a pixel, thus enabling creation of an image.

2.3 Traditional Pixel Principles

In order to understand the way in which event-based cameras improve the state of imaging technology, it is important to understand how conventional camera sensors respond to irradiance to record images. The sensing device in a conventional camera consists of an array of detectors called pixels, which each convert light into electrical information. Each pixel contains a photodiode which generates electrical current proportional to the irradiance incident on that pixel. When a camera begins recording an image, the charge accumulated by the photocurrent is stored in some manner, typically a capacitor. At the end of what is technically referred to as the "integration time" or "exposure time", the storage system stops accumulating any additional charge. At this point, the camera begins a "readout" process, during which the accumulated charge in each pixel is transferred out of that storage area to an analog-to-digital converter and converted to a Digital Number (DN), between 0 and 255 for 8-bit systems [15]. The combination of the maximum amount of charge each pixel can contain, and the maximum DN produced by the converter both contribute to the dynamic range of the camera. Cameras with a higher dynamic range are able to accurately reproduce images with a larger difference between the brightest and darkest regions without saturation.

Different pixel circuit architectures conduct the readout and conversion process in different ways, differentiating CCD and CMOS sensors. Once the DN has been gen-

erated, it is transferred out of the sensor to some electronic storage system. A picture generated by the sensor is the grid of digital numbers translated into grayscale or RGB color, and a video is a collection of subsequently recorded images (or "frames"), each of which contains data for every pixel in that image.

This has proven to be an effective way of storing and communicating spatial information, as evidenced by the fact that the vast majority of consumer and industrial digital images produced have followed this format. There is a significant challenge that afflicts this architecture, however. When recording a video of a scene, there is often a primary subject which fills a small fraction of the total field of view. This subject can be expected to change between frames. The rest of the field of view often does not add significant useful information and does not change much between subsequent frames. An easily-imagined example of this would be an airplane flying from the left side of an image to the right side. The pixels that see the airplane at some point add information. The number of pixels between the front and back of the craft describe its size, while the difference in time between the nose crossing into one pixel then another pixel describes its velocity. However, any pixel that does not record the aircraft is only recording a video of the sky, which may not change significantly through the duration of the video. Thus, a pixel viewing only the sky reports a similar DN in every frame. This creates redundant data without adding useful information. In the case of modern high-definition sensors, relatively short videos can contain several gigabytes of data, most of which is redundant.

One solution to this problem of highly-redundant data is to compress the video files through post-processing. This is a popular and widely adopted practice, but increases the computational load, and by its nature, can only be done in post-processing. This is where event-based cameras add significant value to the industry. Event camera pixels use a fundamentally different process of converting light to digital informa-

tion, generating less redundant data, and recording scenes with a significantly higher dynamic range.

2.4 Event-Based Camera Principles

The event-based pixel was first developed and presented in 1992 by Mahowald in [10]. They were also referred to as "neuromorphic" sensors, due to their functional similarity to biological structures such as neurons and the retina in the eye. In 2008, the company iniVation published [16] which introduced a pixel architecture known as a Dynamic Vision Sensor (DVS). In 2014, [17] introduced a pixel which is capable of recording both events and traditionally-integrated CMOS-style images. This does so by incorporating an APS circuit with the DVS circuit in each pixel unit cell. This combination is referred to as a Dynamic and Active Pixel Vision Sensor (DAVIS) pixel. The DAVIS240 and DAVIS346 cameras both use DAVIS pixels, while the DVXPlorer and Prophesee VGA-CD cameras only contain DVS pixels.

2.4.1 Principles of Operation

In an event-based camera such as the DAVIS240, each pixel has a photodiode which generates current in response to incident light. This is where the similarity to conventional pixels ends. An arrangement of transistors generates a voltage proportional to the logarithm of the photocurrent, which schematically appears as V_p in Figure 1 [16].

Once the capacitors C_1 and C_2 in Figure 1(a) amplify the voltage, the comparators continuously compare it to the preexisting threshold voltage, and the difference is output as V_{diff}. As activity in the field of view changes the irradiance on the photodiode, the current and thus the voltage difference V_{diff} changes. Any time the voltage V_{diff} exceeds the positive or negative threshold shown in Figure 1(b), a com-

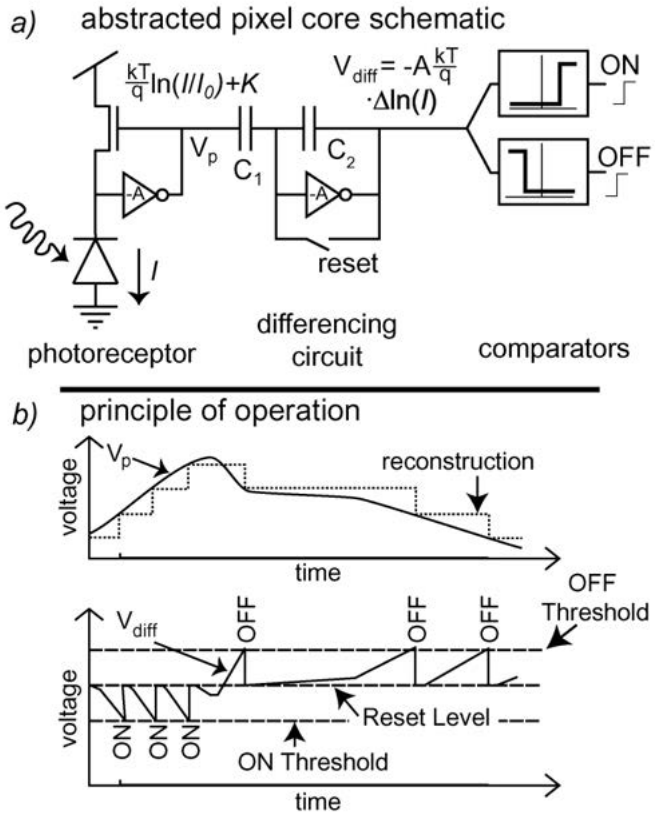


Figure 1. Abstracted Pixel Schematic ©2008 IEEE [16]. (a) shows a simplified event-based pixel circuit. Light is incident on the photodiode on the left. The transistor-amplifier combination outputs a voltage V_p proportional to the logarithm of photocurrent. The differencing circuit subtracts V_p from a reference to produce V_{diff}, which triggers a comparator and thus an event when it exceeds a threshold. (b) shows nominal circuit behavior in response to ongoing change in irradiance.

parator circuit triggers a digital process that records the pixel address (x,y), the the time at which the the address was reported (t), and polarity of the change (p). The collection of these 4 data points is referred to as an event, and is generally referred to as an Address Event Representation (AER) described by Equation 3

$$E = [x, y, t, p]. \tag{3}$$

The voltage V_{diff} is then reset to zero and the pixel is ready to detect subsequent

changes in the scene [16]. In this manner, only pixels which detect some change will record any data. To construct an image, events must be processed into some convenient visualization.

2.4.2 Pixel Arbitration

In the DAVIS240 camera developed by iniVation, the asynchronous process which records the pixel address relies on a series of interactions between the reporting pixel and address encoding circuits on the periphery. In this process, depicted in Figure 2, a pixel which has crossed a threshold generates a row-request (RR) signal. In the event that several rows send requests simultaneously, an arbitration circuit (or "arbiter") chooses the order in which to respond. The arbiter stores that row address then sends a row-acknowledge signal (RA). All pixels in that row which have crossed threshold then send a column-request-on (CRON) or column-request-off (CROFF), depending on which comparator was triggered. These column addresses are stored, and the arbiter generates a column-acknowledge (CA) signal. Timestamps are assigned to addresses in the order determined by the arbiter, after which the pixels are reset, and the now-complete events are transmitted out of the sensor [17]. The DAVIS240, DAVIS346, and Prophesee report events in this asynchronous method, though specifics almost certainly differ between models. The DVXPlorer differs from this by applying a sequential column scan readout [18]. The column scan process results in reduced timestamp resolution ($200\mu s$), but presumably enables more predictable event handling.

2.4.3 Latency

Latency is the delay between a real stimulus occurring and the timestamp of the event describing that activity. While there is a deterministic component due to the

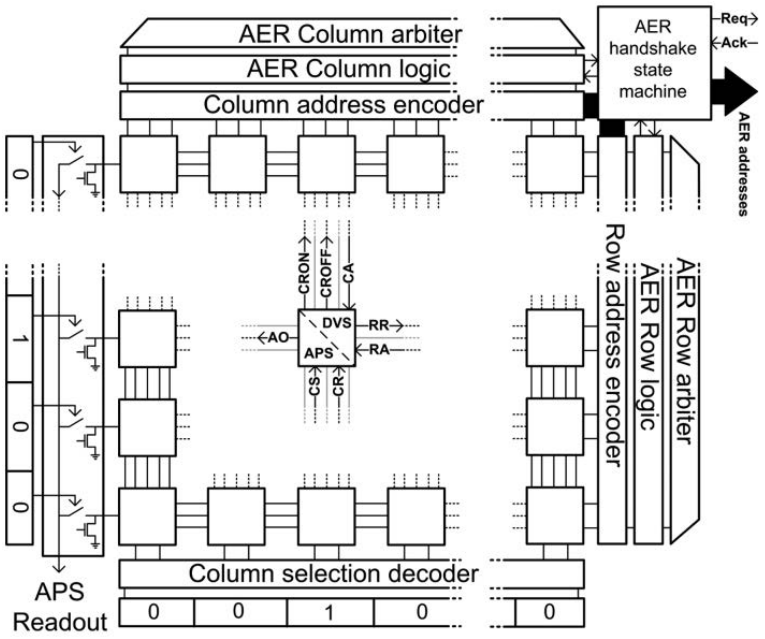


Figure 2. AER Block Diagram ©2014 IEEE[17]. When the comparators trigger an event, the DAVIS240 pixel begins a 4-phase handshake. A row request (RR) is sent to the row address encoder, which responds with a row acknowledge (RA) signal. All pixels above threshold in that row then send an ON or OFF column request (CRON or CROFF) to the column address encoder. The column acknowledge signal resets the pixels, and the addresses are moved to the handshake state machine for timestamping.

geometry of the circuit, noise adds an element of uncertainty to latency. Event-based cameras are plagued by the same sources of noise as traditional sensors, including photon shot noise, thermal noise within the circuit, and fixed pattern noise on the focal plane array [19]. This combination of noise sources can potentially increase uncertainty of event timing to as much as 100 microseconds[20]. Figure 3 shows how variation in voltage may cause deviations in the time at which an event is timestamped, thus contributing to overall latency.

In the white paper produced by iniVation[20], they recognize that while specification sheets may report latency as low as 1 microsecond from a device, that specification may be misleading due to the conditions required to report events with such precision. Generally speaking, as long as the rate at which events are being reported

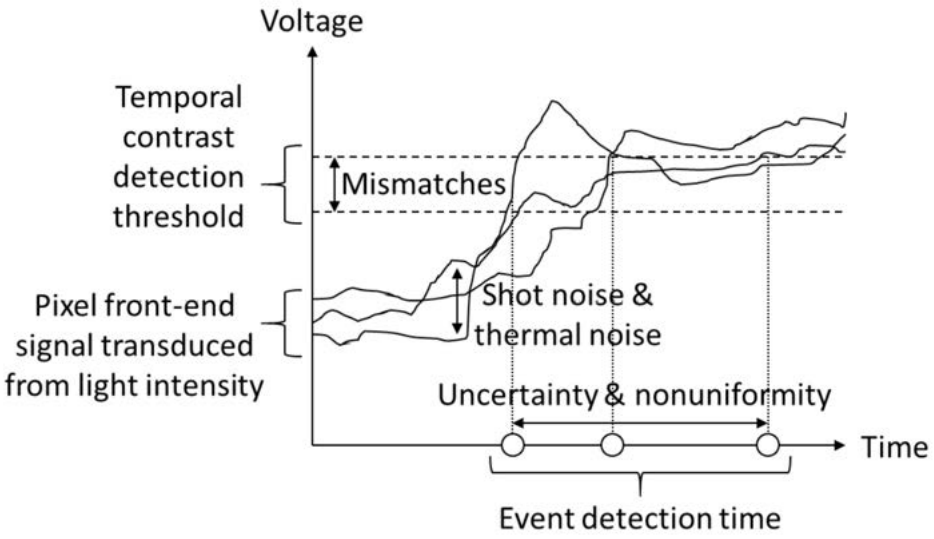


Figure 3. Event Time Uncertainty [20]. Threshold mismatch between pixels and random noise sources contribute to variation in latency, referred to as jitter.

is low, the latency will also be low. However, if even 10% of pixels are stimulated simultaneously with enough contrast to generate events, the readout latency could increase to as much as 1 millisecond. The spread of this latency is not discussed, aside from noting that the temporal distribution can be observed in the timestamps. One contribution of this thesis is to reveal correlations between optical inputs and latency variability.

2.4.4 Dynamic Range

Another central concept to event-based cameras is their larger dynamic range compared to traditional APS camera sensors. Traditional cameras can operate over 60dB [3] where the dynamic range in decibels is defined as

$$DR = 20 * log_{10}(E_{high}/E_{low}) \tag{4}$$

where E_{high} and E_{low} describe the maximum and minimum values of irradiance over which the sensor can operate. 60dB is roughly equivalent to 10 stops ($1 : 2^{10}$) or a ratio of 1:1,000. Their logarithmic response to photocurrent allows event-based cameras to record elements of a scene spanning more than 120dB, equivalent to roughly 20 stops or 1:1,000,000. It should be noted that there is currently no industry standard for determining responsiveness of event cameras, and manufacturer testing procedures and thresholds used to generate that specification may vary slightly. Additionally, event-based cameras respond to temporal contrast, or the relative change in irradiance. This causes a reduction in absolute sensitivity at the high end of the functional range. For example, if an event-based camera has a 10% contrast threshold, it would probably detect a change from $10cd/m^2$ to $11cd/m^2$. However, if the starting illuminance was $10,000cd/m^2$, the camera would be unlikely to detect any change smaller than $1000cd/m^2$.

2.4.5 Throughput

A key parameter used to describe event camera behavior is the number of events per second being recorded by a camera. This is often reported in Million Events per Second (Meps) and depends on the number of pixels being exposed to activity, and the number of events being generated by each pixel. Maximum reported throughput capabilities are 12Meps for the DAVIS240 and DAVIS346, 50Meps for the Prophesee, and 165Meps for the DV XPlorer. Note that this document will use EPS to mean "Events per Stimulus", referring to the testing procedure described later. When referring to events per second in the context of throughput, this document will use "Meps". The unique nature of the address event representation data known as events presents several challenges in processing, visualizing and understanding. Techniques for doing so are presented in the following section.

2.5 Data Comprehension

2.5.1 Time Surface

Event-based data presents a new way of understanding activity in a scene. Because the data presents information in a fundamentally different way than traditional images, it must also be visualized and analyzed in new ways. One method of visualizing sparse event data is through a plot known as a time surface. The first two dimensions correspond to the physical dimensions of the sensor, while the third dimension corresponds to time. In [21], an exponential kernel is applied to the timestamp of events at each pixel. The result of this is that the most recent events have a value of positive or negative 1, depending on event polarity, while older events decay exponentially toward zero. The figure resulting from this analysis indicates peaks or bright spots for the most recent activity, and nullity in regions without activity. This provides a historical context for each pixel and its most recent event compared to its neighborhood, without cluttering the image with excessive past events.

This technique is well suited for representing a long history of events in a compact plot, as the exponential kernel continually emphasizes most recent events and diminishes the visual impact of no-longer-relevant old events. An alternative way of producing a time surface is to leave time as represented linearly rather than applying an exponential decay. Because event cameras report events with microsecond resolution, expressing the time linearly invokes a very short limit on the depth of a surface plot. For example, a hypothetical 8-bit image for which each pixel can contain a value from 0-255 could only represent 256 microseconds of activity without compressing time divisions by binning events and thus lowering temporal resolution.

A strength of expressing time linearly in a time surface is that small differences between pixels are easy to detect for a sufficiently short period of interest. In the case of this research, where the field luminance is simultaneously changing in less than

one microsecond, the period of interest for comparing pixels is very short, and spatial patterns emerging from the camera circuits are expected to be small.

To understand the functionality of a time surface better, an idealized example is shown in Figure 4. Figure 4(a) shows the simulated field of view of an event camera. In this example, the dark gray background is not changing at all, while a light gray bar steadily expands to the right, triggering one column of events at a time as it expands from left to right. Figure 4(b) displays a time surface generated

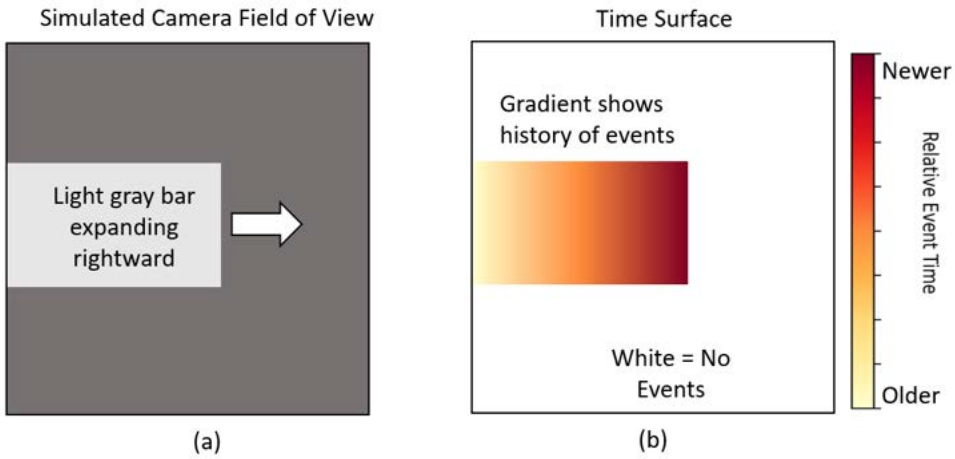


Figure 4. Simulated Time Surface. (a) shows a hypothetical simple scene as imaged by an event camera, in which a light gray bar is steadily expanding to the right, with a dark gray background. As the expanding bar crosses into each new column of pixels, the pixels in that column report an ON event, each column with steadily increasing timestamps. (b) shows a time surface describing those events, where the most recently-triggered pixels are colored darker, while the first pixels to trigger are colored lighter. Pixels that have not seen any transition do not produce any events, and are depicted as white. If the light gray bar were to shrink, pixels would generate OFF events and could be visualized in the same manner.

from the events produced by the expanding bar. The lightest region of the gradient shows oldest events, and the darker pixels to the right show newer events. The white space shows where no activity was detected so no events were generated. If the light gray bar were shrinking instead, the pixels would generate OFF events and could be visualized in the same manner.

Both expansion and contraction demonstrate spatial contrast, in that an object is moving through the scene. An example of temporal contrast would be if the entire frame was initially black, then suddenly turned to white. This would generate a flat and level time surface, suggesting all pixels recorded an event simultaneously. An ordinary camera flash is a good example of high temporal contrast. Time surfaces can be effective for qualitatively revealing spatial patterns, either from the scene or from the sensor behavior. A more quantitative understanding can be gained by analyzing the range of times at which pixels report events.

2.5.2 Median Absolute Deviation

Each time surface is a set of data which can be easily described by common statistics including the mean, median, standard deviation, and a less common statistic called the Median Absolute Deviation (MAD). Due to complexity in the scene and noise in the sensor, time surfaces are likely to have significant outliers in addition to any ideal patterns such as those depicted in Figure 4.

In general, datasets can be described by their magnitude and their spread. The magnitude is commonly described using the arithmetic mean, while spread is usually described by the standard deviation. Each metric, calculated from a sample of the population, is a statistic that estimates the population parameter, and tells nothing about the other metric. That is, one can learn nothing about the spread of the data based on just the mean, and nothing about the average magnitude of the data based on just the standard deviation.

The mean and standard deviation are not considered "robust" statistics, in that both are significantly affected by outliers in the data. For any arbitrary set of data, the standard deviation and mean can be made arbitrarily large by adjusting merely one data point. This suggests that even if a data set has a small number of outliers,

those outliers can significantly change the value of the statistics, and thus tell an inaccurate story regarding the source of the data.

The median of a data set is a measure of magnitude that is robust to outliers, as the actual value of any single outlier does not affect the value of the median. The Median Absolute Deviation (MAD) is a measure of spread based on the median, similar to how the standard deviation is based on the mean. The MAD is the median of the deviations of the data from the data median.

In terms of time surfaces, the MAD is a way of describing the time scale over which events are recorded following instantaneous stimulation. Figure 5 depicts how a series of 5 events might be read out from a sensor, and how the MAD of the recorded times would be calculated. In Figure 5, some stimulus around the reference time stimulates 5 pixels. After some latency period, each event gets timestamped and saved to memory. In post-processing, the algorithm finds the median event and calculates the deviation of every other event from that median. It then calculates the median magnitude of the deviations, resulting in the metric of spread referred to as the MAD.

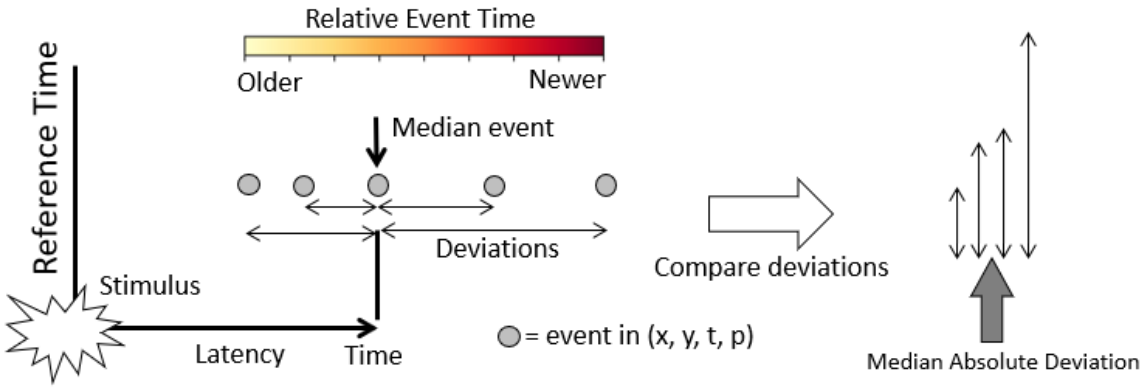


Figure 5. Graphical Depiction of MAD. The Median Absolute Deviation describes the spread of a group of event times. It is calculated by finding the median event, calculating the time difference (deviation) between the median event and every other event, then finding the median of those deviations.

Latency and temporal resolution are two often-reported metrics describing the performance of event cameras [3], and together they give the reader an impression of the timing accuracy of each sensor. A camera's precision, however, is limited by both the temporal resolution and the spread of the events being reported. Resolution is designed and reported by the manufacturer, and spread is reported in this thesis.

2.6 Recent Publications

Four recent publications are particularly relevant to this characterization effort. The first presents an experimental setup and process for uniformly exposing pixels. The second uses a simpler setup but develops a useful histogram technique for discussing event generation rates. The third uses a similar setup as the first to characterize the frequency response of a pixel, and develop an initial model of the DAVIS240. The fourth expands on that model and provides additional detail on the characteristics of event cameras.

2.6.1 Mechanical Setup

A characterization setup is proposed by Joubert et al.[11], in which LEDs were installed in an integrating sphere. LEDs were used to stimulate the cameras because they respond quickly to current drivers, enabling sub-microsecond radiance control. An integrating sphere provided a means to uniformly illuminate the sensor, providing direct comparison between pixels while minimizing features in the scene. Neutral density filters were installed to reduce the irradiance on the sensor, but high irradiances were challenging due to power and thermal concerns regarding the LEDs and drivers. They develop a way of quantifying initial and final irradiance, including a percent change which is the controlled variable in the experiment. Several characteristics of a DVS sensor were reported, but the document did not specify which camera was

actually tested.

A similar, albeit less sophisticated arrangement was used by Hollidt in [22]. In that work, a single white LED illuminated a brown cardboard target set on a table. The cardboard was arranged such that it only filled a small portion of the field of view. The motivation of that work did not require high throughput utilization so the author actively avoided exceeding the throughput capacity of the cameras. Hollidt processed the data using a form of temporal histogram, summarizing event generation rates by how many events were generated within subsequent millisecond periods. That procedure formed a foundation and starting point for the data processing workflow used in this thesis.

2.6.2 Model Development

McReynolds[12] identified key parameters necessary to describe the behavior of pixels in a DAVIS240. These parameters were contrast threshold, fixed pattern noise, noise equivalent contrast, effective time constant, refractory period, pixel threshold mismatch, and bandwidth. Each parameter was experimentally measured, and used to create a functioning model of the sensor. In doing so, it was reported that the pixels are able to reliably respond to repeated stimuli up to 3kHz, representing a refractory period of approximately $330\mu s$. McReynolds proposed that pixels can be approximated as RC lowpass filters, described by a time constant. Temporal noise behavior and fixed pattern noise were also replicated. This effort produced the first model of an event camera to incorporate physics-based frequency response and noise generation.

2.6.3 Model Application

Delbruck et al.[23] present a more comprehensive model which can convert traditional videos into simulated event streams. This converter incorporates a pixel model which expanded on pixel threshold mismatch, illumination-dependent bandwidth, and various noise sources. The paper describes sources and characteristics of latency and variability, or jitter. It also describes the non-instantaneous response of a pixel to a step-function stimulus using an RC lowpass filter approximation, such that when photocurrent changes rapidly in response to a stimulus, the voltage being applied to the comparators approaches its new value over some relatively long period, compared to the time of the photocurrent change. This non-instantaneous response causes pixels to generate several events over an extended time period. Thus, as a sharply-focused edge passes over a pixel, the pixel may still be generating new events after the edge has passed, reducing the spatio-temporal resolution of the image of the edge. This is effectively an event-based camera's version of motion blur.

Jourbet et al.[11] and Hollidt[22] presented opto-mechanical setups for controlling fast stimulation of event cameras. McReynolds[12] used a similar setup to classify event camera pixels as an RC lowpass filter, and Delbruck et al.[23] used a similar approximation to present a model for sensor operation. These papers have examined and discussed characteristics of individual event cameras. Aside from some discussion and analysis of fixed pattern noise, there has not been significant exploration of the spatial tendencies of sensors in response to stimuli. Few papers have applied the same test methodology to several different makes and models of cameras, resulting in few direct comparisons between cameras. From a scientific perspective, the development of models is highly useful. However from an application perspective, the lack of side-by-side comparison makes selection of a specific model challenging.

2.7 Readout Behavior

A significant amount of recent work has expanded the industry's understanding of pixel-level behavior in event-based cameras. While current models describe event generation, they do not describe event timestamping or storage processes, or potential performance limitations associated with those operations.

One question which has not received much attention in the literature regards the maximum rate at which a camera can generate and timestamp events. Specification sheets report maximum throughput, but do not specify if the maximum throughput is an instantaneous limit or a long-term average. If it's an instantaneous maximum, cameras in this study would be limited to 12, 50, or 165 events per microsecond. If it's an average limit, then instantaneous events/microsecond could potentially be larger than the limit, as long as there follows a period of low throughput to compensate. Regardless of the instantaneous vs average limit, one should expect performance to degrade above that limit, while performance should be in line with advertised specifications below that limit.

Simultaneously timestamping events from many pixels could easily exceed the throughput capabilities of these cameras, so based on the throughput capability of the camera there is necessarily a minimum time required to read out events from the entire array. This sets a minimum theoretical value for the Median Absolute Deviation obtained from a full-frame time surface. For example, the DAVIS240 has 43200 pixels. If every pixel is simultaneously stimulated and only generates one event, at the quoted rate of 12 Meps it would take $3600\mu s$ to timestamp and read out one event from every pixel in the array. Table 1 show minimum readout times calculated for all four cameras, defined as t_r.

To determine the theoretical minimum MAD, assume that the camera timestamps these events at a uniform rate. For every time surface generated, there will be an

equal number of events before and after the median event, with equal and opposite deviations from that median event. The absolute deviations from the median would then range linearly from 0 to $\pm t_r/2$. The median of these deviations (the MAD), is then simply the center of that set: $t_r/4$. This is the lowest MAD one can expect with 100% of pixels reporting. If a smaller percentage of the array is generating events, the minimum MAD would likely scale down accordingly.

Table 1. Theoretical Full-Array Readout Time (μs)

	DAVIS 240	DAVIS 346	DV XPlorer	Prophesee Gen3
Array Size (pixels)	43200	89960	307200	307200
Throughput ($events/\mu s$)	12	12	165	50
Readout Time t_r (μs)	3600	7497	1862	6144
Min MAD $t_r/4(\mu s)$	900	1874	465	1536

2.8 Summary

This chapter presented the operating principle behind dynamic vision sensor pixels. Readout architecture and metrics such as latency, dynamic range, and throughput were defined. Techniques for processing data were described, establishing a common vocabulary necessary for the rest of this document. Recent work toward characterizing new sensors was examined, which provides a foundation for the experimental methods used to compare the cameras in this research effort. Finally, manufacturer-provided specifications were used to predict the minimum MAD under high event loads. The following chapter presents the methodology used to investigate these parameters.

III. Methodology

3.1 Overview

This chapter begins with a description of the physical setup used to evaluate the four cameras: iniVation's DAVIS240, DAVIS346, and DVXPlorer, and Prophesee's Gen3M VGA-CD 1.1. Following the mechanical setup is a description of the optical stimulus and the electronic considerations associated with it. A brief experiment is described which was used to verify that the characterization process was not being influenced by external light sources, followed by the procedure implemented to conduct that characterization process. The chapter concludes by describing the data processing pipeline and production of metrics described in the previous chapter.

3.2 Physical Setup

3.2.1 Pixel Stimulation Percent

In order to understand each camera's behavior in response to different levels of activity, it was required to stimulate various fractions of the pixel array, ranging from a very small to a very large number of pixels. The fraction of the sensor being exposed to the stimulus is defined as the Pixel Stimulation Percent (PSP). PSP is directly related to the "load" being experienced by a camera and the associated consumption of communication bandwidth: more pixels being exposed results in more events being generated and transmitted to storage.

A consistent and repeatable stimulus was used to compare behaviors of different cameras and configurations. In order to uniformly illuminate a controllable number of pixels while blocking out all other light, an integrating sphere was placed inside a light-proof enclosure. An array of ten LEDs was installed as the sole light source in an access port previously used to hold one of the halogen light sources for the sphere.

A 1.5 inch iris diaphragm was installed on the exit port of the integrating sphere and opaque tape was applied between the iris and the sphere to seal the gap. The other access ports on the sphere were also taped closed to prevent light from passing through them. The camera being tested was positioned in front of the iris and sphere, with the lens focused on the iris. This created an image of the iris on the camera sensor. Black felt was placed over the integrating sphere with a circular hole cut out for the iris, creating a nearly-constant-reflectance background. Felt was also placed around the walls of the enclosure to minimize stray reflections.

To expose a minimum number of pixels and thus a minimal fraction of the camera sensor, otherwise referred to as a Focal Plane Array (FPA), an additional 1.5 inch (3.8 cm) diameter middle iris was installed 5 inches (12.7 cm) in front of the sphere iris as depicted in Figure 6. By leaving the sphere iris fully open and adjusting the diameter of the middle iris, stray light reflections into the camera were minimized. Each camera under test was installed 16 inches from the middle iris, as measured from the camera body mounting hole.

By closing the middle iris to its smallest aperture, a small number of pixels were exposed while all other pixels image the constant black background, remaining effectively unstimulated. This enabled the capability to produce a very small source of controlled brightness approximately 16 inches from the camera under test with a nearly-constant brightness background. This arrangement allowed for illumination of a very small fraction of the FPA, but its limited diameter did not allow for illumination of a large fraction of the FPA. To accomplish this, the middle iris was removed and the camera mounting location was relocated to 2.5 inches (5 cm) in front of the sphere iris as shown in Figure 7. In this configuration, the end of the lens was less than one inch from the iris. This close proximity was necessary for the iris to retract completely out of the field of view of the Prophesee camera, which had the largest

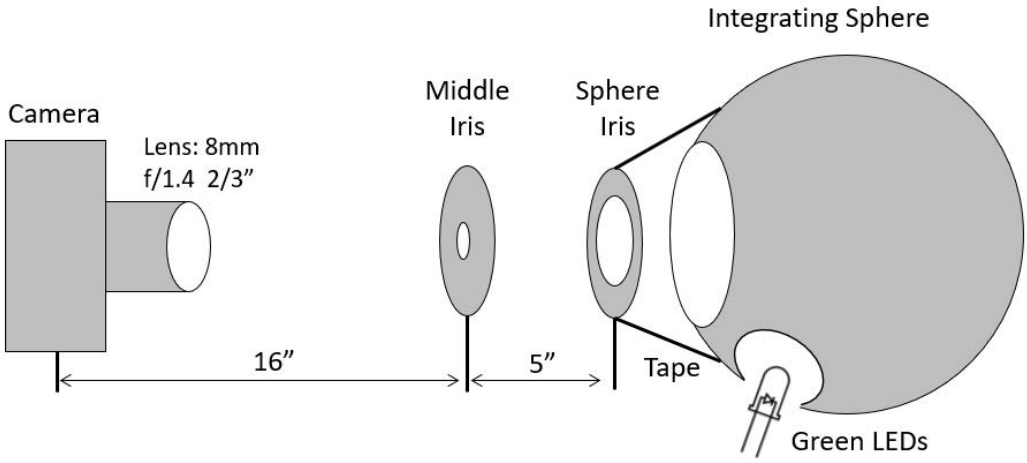


Figure 6. Two-Iris Setup. Each camera was located 16 inches (40cm) from the middle iris. The LEDs were mounted in a lamp access port, and the sphere iris and tape blocked stray light from the output aperture.

field of view. This was slightly closer than the minimum focal length of the lens used, so some slight focus error was present. Some focus error was determined to be acceptable in order to maximize the fraction of the field of view being filled by the source.

The integrating sphere was a 12 inch (30.48 cm) diameter Electro Optical Industries model. Due to the multiple-bounce nature of integrating sphere, the sphere's impulse response is of the form

$$e^{\frac{-t}{\tau}} \tag{5}$$

where t is the time after the impulse is applied. τ is a time constant calculated as

$$\tau = -\frac{2}{3} * \frac{D_s}{c} * \frac{1}{ln(p)} \tag{6}$$

where p = average wall reflectance, c= the speed of light, and D_s = sphere diameter [24]. For the sphere used in this research, the time constant τ was approximately 15 nanoseconds. The rise time of the sphere output is ultimately determined by

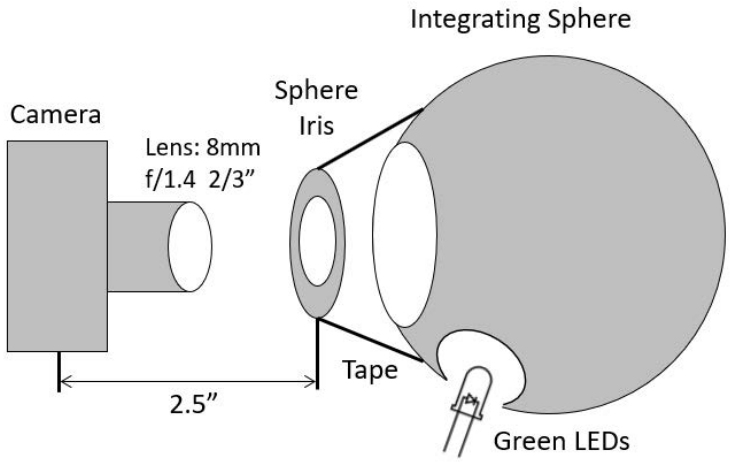


Figure 7. One-Iris Setup. A higher fraction of the field of view could be filled by moving each camera to only 1 inch (2.5cm) from the sphere iris.

convolving this impulse response with the timing profile of the LED light source.

3.2.2 LED Wiring

One of the factory-installed halogen bulbs was removed from one access port and an array of green LEDs was installed in its place. Hardware availability and circuit complexity drove the decision to use green LEDs rather than other options, in addition to the possible convenience of converting between radiometric and photometric units. These LEDs were mounted on a custom plastic cylinder which was installed into the integrating sphere in place of the 4th halogen lamp. This arrangement caused the light from the LEDs to scatter several times inside the integrating sphere before being captured by the camera under test, producing a uniformly-illuminated field of view for every pixel. Uniform illumination was critical so that any differences in response could be attributed to camera behavior rather than differences in stimulus.

A function generator drove the LED array in order to produce an “instantaneous” impulse with which to evaluate the cameras. One additional matching LED, wired in

parallel with the main array, was mounted outside the enclosure and aligned with a fast photodiode (FPD). The FPD had a typical rise time of 1ns[25] and was connected to an oscilloscope to enable real-time monitoring and recording of the radiance emitted by the LEDs.

3.2.3 Pulse Generation

The function generator was set to produce an increasing (OFF to ON) stimulus at approximately 10 Hz with a 50% duty cycle square wave, which left enough time between impulses for the camera under test to reach equilibrium. To adjust brightness, the function generator amplitude and offset were adjusted such that the top voltage was near zero, and the base voltage decreased, resulting in higher current through the LED array. Actual base voltage was recorded through the oscilloscope and the average and standard deviation reported over several cycles. The voltage amplitudes and offsets were chosen to drive the LEDs near their low-power and high-power limits, without under- or over-driving them. The indicator LED rise time, duration, and fall time were directly measured by the fast photodiode, and the timing of the array was correlated to the indicator timing, so effective array timing could be calculated.

3.2.4 LED Rise Time

It has been observed that several events may be generated in response to a single moment of high temporal contrast [23]. Each camera tested in this research had a minimum timestamp resolution of $1\mu s$, so any activity occurring in less than one μs would be effectively instantaneous. In order to confirm that the LED array was generating "instantaneous" (sub-microsecond) pulses, the rise time of each individual LED was directly measured using the FPD. The single LED left outside the enclosure (referred to as the "indicator LED") was also characterized.

Each LED was aligned with a ThorLabs Fast PhotoDiode such that the output voltage from the FPD was maximized without significantly distorting the square waveform, as recorded by a Rhode & Schwartz 300MHz oscilloscope. Rise time and fall time were automatically calculated by the oscilloscope by calculating the elapsed time between the voltage passing 20% and 80% of the maximum value attained that cycle. At least 1000 cycles were recorded and averaged to produce averages for each LED. The mean rise time of all 10 LEDs was calculated to be 73 nanoseconds with a standard deviation of 8 nanoseconds. This is assumed to be effectively "instantaneous" due to the 1 microsecond minimum timestamp resolution of all cameras tested.

Figure 8 shows the optical output of one representative LED as measured by the FPD. The upper line shows the output of the LED represented by volts generated by

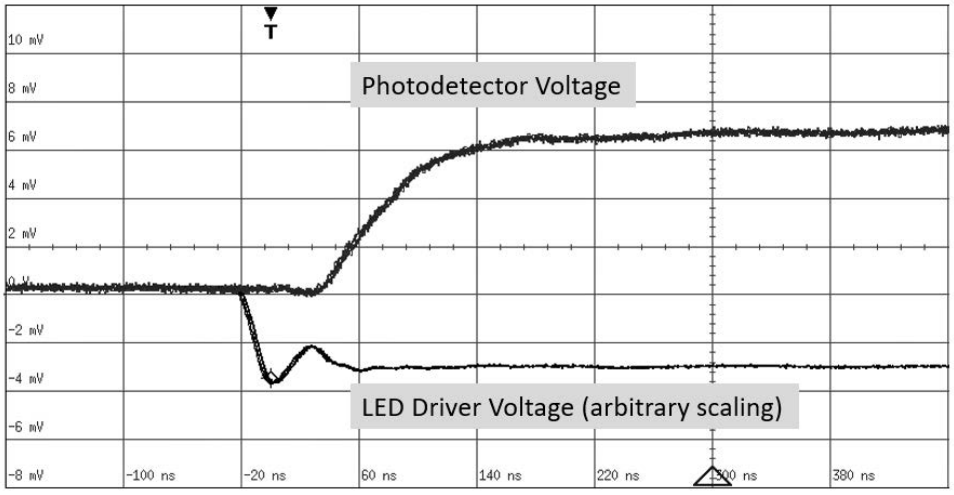


Figure 8. LED Stimulus Profile. This oscilloscope trace shows the output of a representative LED in the source array in response to the driver voltage. The vertical scale corresponds to the photodetector voltage used to measure the LED rise time, the horizontal scale is 80 ns/division. The driver voltage changed from 0 to roughly -5 volts to power the LED array, and is displayed scaled down for clarity. The photodetector voltage shows that LED rise time is less than approximately 100 nanoseconds. The LEDs remained on for 50 milliseconds per cycle, allowing sufficient time for the camera to reach a steady state.

the FPD, while the lower line shows the voltage being applied to the LED array. The overshoot and subsequent ringing were considered negligible because those effects did not appear in the observed FPD voltage. The 15ns time constant of the integrating sphere and the 73ns rise time of the LED array, when convolved, produce an optical impulse with a rise time significantly shorter than 1 microsecond, the minimum temporal resolution of every camera tested.

3.2.5 Sphere Luminance

Event cameras respond to change in the logarithm of the photocurrent being produced by the sensor[23] so

$$Contrast = \frac{L_{final} - L_{initial}}{L_{initial}} \tag{7}$$

is a relevant metric to describe scene activity. The integrating sphere displayed the current luminance as measured by a detector inside the sphere. However, the initial luminance in this experiment was lower than could be detected by the integrating sphere's detector, so was approximated as $L_{initial} \approx 0$. This causes a problem for calculating contrast. Thus instead of reporting a percentage-based change, this paper reports the magnitude of change by reporting the final luminance value

$$Contrast = L_{final} - L_{initial} = L_{final}. \tag{8}$$

One could calculate a contrast by comparing the photocurrent generated by noise versus the photocurrent generated by an illuminated pixel, but there is currently no way to directly access the photocurrent in these cameras. Those cameras which include an APS imaging mode could not be used to calculate contrast due to dynamic range limitations and an observed tendency to generate noise events upon readout of

integrated frames. However, in the case of a physical parameter such as luminance where negative values are nonphysical, describing the magnitude of the final value is equivalent to describing the magnitude of the change when the initial value was zero.

The integrating sphere controller reported value of luminance was in foot-Lamberts (fLs). During normal operation, the integrating sphere controller would use this signal in its closed-loop control system to adjust the output of the various halogen lamps to maintain a preset desired luminance. However, if the desired luminance is set to 0 fL, the controller turns off all 4 lamps, while continuing to measure and display the luminance of the sphere. When light is then injected into the sphere by the externally-controlled LED array, the internal sphere controller sensor displays the effective luminance generated by the LEDs. In this way, the luminance generated in the sphere by the LED array was measured. Simple dimensional analysis allowed conversion from foot-Lamberts to the metric equivalent, cd/m^2. By definition, $1fL = \frac{1}{\pi}\frac{candela}{ft^2}$. Converting to metric length gives

$$1\frac{cd}{m^2} = \frac{1}{\pi}\frac{candela}{ft^2} * \frac{(1ft)^2}{(12in)^2} * \frac{(1in)^2}{(2.54cm)^2} * \frac{(100cm)^2}{m^2} \tag{9}$$

By Equation 9, the luminance in fL reported by the sphere was converted to cd/m^2.

The function generator was not able to adjust the duty cycle of the square wave driving the LEDs, which was set to 50%. The sphere controller only updates the luminance display at roughly 1Hz, which meant it was averaging the luminance over a relatively long period (approximately 1 second). Thus when the function generator was flashing the LEDs at 10Hz during the data collection, the sphere controller was only reporting 50% of the real luminance from the LEDs. To measure the real luminance of the sphere, the function generator was set to run the LEDs in a DC steady-on mode. The voltage provided by the function generator was recorded with the luminance reported by the sphere controller. To reproduce the desired luminance

during subsequent trials, the square wave amplitude was adjusted to match the target voltage.

While some optical characterization efforts would consider on-pixel illuminance, measuring the flux incident on a unit surface from all solid-angles, this effort only discusses luminance propagating outward from an "object" or source surface. The purpose of this research is to directly compare the properties and behavior of the entire system including each camera's FPA, Readout Integrated Circuit (ROIC), and associated storage architecture rather than studying each sensor at an individual-pixel level. Therefore, reporting the illuminance on a single pixel is less informative than reporting the luminance of the scene the array of pixels is observing.

3.2.6 Lens Selection

A Computar 8mm f/1.4 2/3" format lens was used with all 4 cameras. Each camera's sensor had different physical dimensions while the image produced by the lens remained constant regardless of camera. The Prophesee camera had the largest physical sensor, thus defining the minimum lens format required to accurately fill the sensor. The three other cameras had smaller sensors which reduced their field of view when compared to the Prophesee camera.

Alternately, choosing a lens optimized for the iniVation camera would underfill the Prophesee sensor, which would result in the outer pixels experiencing lower irradiance than the central pixels, therefore inducing nonuniformity that would effect noise rates. This would be a contributing factor to behavior between the Prophesee and iniVation cameras. Therefore, it was concluded that it would be better to use a lens that adequately fills or overfills all sensors than to use one that underfills a sensor. The aperture was left fully open in all cases.

3.2.7 Dark Noise

In order to use absolute final radiance in place of contrast as discussed in Section 3.2.5, the initial scene radiance must be effectively zero. However, event cameras produce "noise events" in the absence of scene activity [23], even with a lens cap on. To confirm that the enclosure was dark-noise limited for every camera, the LED array was turned off and events were recorded for 10 seconds. Every camera's settings were left on their default values.

The lens cap was installed and events recorded for another 10 seconds. The total number of ON and OFF events recorded was divided by the length of the recording and the number of pixels in that camera's FPA. The result is the average events per pixel per second, values for which are displayed in Table 2. The Percent Difference takes the difference divided by the covered rate.

Table 2. Dark Noise Event Rate (Events/pixel/sec)

	Lens Uncovered	Lens Covered	Difference (%)
DAVIS 240	0.1251	0.1245	0.4819
DAVIS 346	1.9989	1.9826	0.8222
DV XPlorer	0.7412	0.7143	3.766
Prophesee Gen3M	0.0471	0.0480	1.875

The Percent Difference values represent the deviation of the enclosure or "scene" radiance from the lens cap radiance. Since there is negligible visible light being emitted by the lens cap, and no way for light to leak around the lens assembly into the sensor, it can be approximated that when the lens cap is installed, there is zero optical irradiance on the sensor. Based on the small difference observed when the cap was removed, it can be assumed that when the LEDs are off, there is effectively zero illumination on the sensor and the camera is dark noise limited.

3.3 Test Procedure

Cameras were tested in two phases. The first phase evaluated their response when a very small fraction of the array was exposed. In order to evaluate their response when a large fraction was exposed, the testbed had to be rearranged slightly before repeating a similar process in the second phase.

3.3.1 Camera Settings

Each camera's sensitivity and bias settings were left at their default values. Any software filters were turned off, and the interal Inertial Measurement Units and gyroscopes were turned off if applicable. On the cameras capable of recording APS frames in addition to events, the frame collection was turned off.

3.3.2 Low PSP

Cameras were installed and tested one at a time, with each camera mounted as described in Section 3.2.1. The lens was manually focused on the middle iris. The middle iris was set to 2mm, 4mm, then 6mm diameter for each camera. The sphere iris was opened fully, to 43mm diameter. For each iris diameter, the function generator output voltage was set to the predetermined amplitude and offset to generate between $9.3cd/m^2$ and $771cd/m^2$.

After installing and focusing each camera, the enclosure door was closed and the room lights were turned off. Minimal light from the controlling computer and oscilloscope remained present in the room but the enclosure prevented light leakage into the test space. The function generator driving the LED array was left running continuously, starting several minutes before the first set of data collection to allow for mechanical warmup and stabilization.

The function generator was set to 10Hz and various voltages as determined pre-

viously. The real frequency and amplitude of the output was measured directly by the oscilloscope, and the output of the indicator LED was measured by the FPD and displayed on the oscilloscope. The output from each camera was recorded for approximately 5 seconds per configuration, and the oscilloscope data for both the function generator and FPD channels from that same period was saved for later reference.

3.3.3 High PSP

After all four cameras had been tested at the 16" position, the middle iris was removed and the camera mounting location was moved up to 2.5" (6.4cm) from the sphere iris. At this location, the iris was slightly inside the nearest possible focus of the lens, but the corners of the Prophesee field of view were slightly blocked by the full-open iris, preventing 100% PSP. The location was chosen as a compromise, to allow slight focus error in exchange for maximizing the fraction of the sensor exposed to the integrating sphere.

For each camera, the sphere iris was set to 2mm, 5mm, 10mm, 20mm, and fully open (approximately 43mm). The LED array was controlled in the same manner as before. Since the DAVIS240 camera has the smallest sensor and corresponding field of view, setting the iris to 20mm completely filled the camera's field of view. This meant that the recording from the 43mm setting was redundant, and is not reported.

3.4 Data Processing

The data processing pipeline consisted of two primary stages. The first stage converted raw data from a proprietary format into an accessible and useful format where events could be individually examined, then filtered out the irrelevant events, leaving only those which could be attributed to the stimulus. The second stage analyzed the set of relevant events to generate time surfaces and statistics describing

each configuration.

3.4.1 Filtering

Each camera produced proprietary data formats. iniVation cameras output data using a .aedat4 filetype, while the Prophesee camera generated files using the .raw extension. Of note, this is different than the .raw files used as image files in the photography industry. Fortunately, the manufacturers published software toolboxes to convert data into the [x,y,t,p] AER format using the Python programming language. Many real-world applications of event cameras include activity where illuminance first increases then decreases later. For characterization then, it was appropriate to consider only ON events. Characterizing OFF events, generated by decreasing illuminance, would likely be informative but is beyond the scope of this research. As such, once the data was converted into the AER format, the OFF events were removed.

In an ideal event-based camera, every change in scene radiance would generate events at the corresponding pixels, and every event would correspond to real changes in scene radiance. Neither of these behaviors is accurate, however. As discussed in [12], pixels cannot always keep up with changes in scene radiance if those changes are on the order of kilohertz. As discussed in [19], pixels can also generate events even when no real changes in radiance occur.

Diffraction through the iris caused light to fall on pixels outside the ideal image, but at an unknown illuminance. Unsurprisingly, pixels exposed to this diffracted light responded differently than the pixels illuminated by the ideal image. To generate meaningful statistics, only the pixels exposed to a well-controlled stimulus could be included in the analysis, so pixels beyond the edge of the ideal image were removed from the datasets.

To determine which pixels to analyze, the total number of ON events generated

by each pixel for the duration of the recording was calculated and divided by the duration of the recording, establishing an average event rate for each pixel. When this average rate for each pixel is displayed as an image, it was often obvious which pixels correspond to the ideal image of the iris. An example of such an image is shown in Figure 9. Due to the reduced magnitude of diffracted light as discussed above, the dimmest stimulus produced the most clearly identifiable circle.

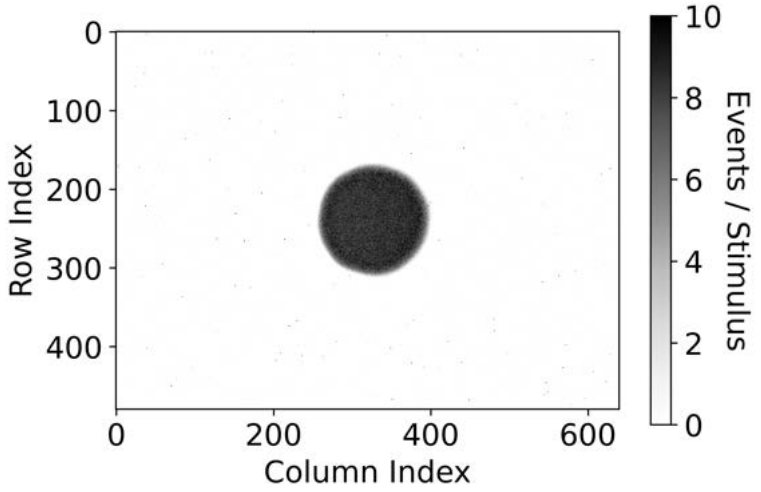


Figure 9. Pixel Event-Rate Image. When displayed as an image, the average number of events generated per stimulus for each pixel clearly shows which pixels were stimulated and which were not. Darker pixels suggest higher event rates. The center and radius of the exposed area were manually selected. This particular example shows the event rate image produced by the DVXPlorer when 3.4% of the array was exposed to a 9.3cd/m^2 stimulu. The black pixels scattered across the sensor show hot pixels.

For each iris diameter, the center and radius of this circle was manually recorded and used to generate a circular Region of Interest (ROI). All events from pixels outside the ROI were ignored, while all events from pixels inside the ROI were included in calculations. This algorithm was applied to the dimmest recording for each object, which reduced the number of pixels generating events from diffracted light, while keeping events from pixels corresponding to the ideal image projected onto the FPA.

Several individual pixels per camera generated significantly more events than their

nearest neighboring pixels and are referred to as "hot pixels". Events generated by hot pixels were included in throughput calculations, but further analysis of those defective pixels is beyond the scope of this document.

3.4.2 Analysis

As described in Section 2.5.1, time surfaces are a convenient and informative way to visualize a large quantity of event data while maintaining temporal resolution. By plotting the timestamps of events against the 2-D array corresponding to the FPA, they enable the viewer to see spatial and temporal patterns in the field of view. Pixels often generated several events in response to a single impulse. The time of the first recorded event was used in the time surface, and all subsequent events from that pixel were disregarded until the next stimulus period.

It was not possible with the hardware available to synchronize the timing of each light source transition with the clock in each camera, so it was not possible to calculate absolute delay between the real increase in radiance and the corresponding events. Instead, a datum time was defined between the first and second impulse by manually inspecting the event throughput histogram. For each pixel inside the ROI discussed in Section 3.4.1, the difference in time between the datum and the first event after that datum is calculated and stored as that pixel's value in that time surface. While producing each time surface, the time datum was continually advanced to be just before the impulse response in question. This generated approximately 50 independent time surfaces for each configuration of camera, iris diameter, and scene radiance, where each time surface depicts the response to a separate trial. In addition to generating time surfaces, the Median Absolute Deviation (MAD) and number of Events per Stimulus (EPS) were also calculated. Mathematical considerations for those statistics are discussed in Chapter 4.

3.5 Summary

Each camera was exposed to a series of flashes generated by LEDs in an integrating sphere. The function generator drove the LEDs at amplitudes ranging from $9.3cd/m^2$ to $771cd/m^2$. Iris diameters were adjusted to stimulate between 0.0042% and 100% of pixels. For each trial, the LEDs were initially off and the entire field of view of the camera was completely dark. The LEDs rose to the high-power state in less than 1 microsecond, which was fast enough to be considered "instantaneous" due to each camera's one-microsecond minimum temporal resolution. The LEDs remained on long enough for the event readout rate to normalize to a static-scene background noise level. For the four cameras tested, 50 milliseconds was sufficiently long. After this time, the LEDs were turned off for 50 milliseconds, to allow the event readout rate to normalize again. This pattern was repeated continuously for approximately 5 seconds, resulting in 21 independent trials per configuration.

Events produced by each camera in response to this repeated stimulus were recorded. In post-processing, time surfaces for each positive (increasing) stimulus were generated as described in Chapter 2 and the number of ON events produced by each pixel per stimulus was calculated.

IV. Results

4.1 Overview

Events generated in response to stimuli can be analyzed in many ways, depending on the application and required information. Four metrics are presented here to describe the behavior of the cameras. The first metric used to understand camera behavior is the time surface generated in response to a stimulus. Time surfaces show spatial-temporal relationships between pixels, revealing physical patterns and tendencies associated with each sensor. A flat or low-contrast time surface suggests a high degree of simultaneity in the temporal repsonse of a subset of pixels, while a higher contrast or more sloped time surface suggests larger deviation in event timing. The second metric used to describe camera behavior is throughput. This describes the rate at which a camera generates events, and provides insight into how the arbiter in each camera handles a large number of pixels reporting events in a short period of time.

The third metric presented here is the Median Absolute Deviation (MAD), as introduced in Section 2.5.2. The MAD quantifies the level of simultaneity between pixels responding to a stimulus. While the MAD removes all spatial information, it enables the reader to quickly compare camera responses to a variety of stimuli. Finally, the number of ON events generated by each pixel in response to each impulse is presented, referred to as the Events per Stimulus (per pixel) (EPS). This metric describes the pixel sensitivity and determines bandwidth needed by the camera to report activity. The number of pixels inside the stimulated ROI is the other driving element.

4.2 Uncertainty

Two values required a formal treatment of uncertainty. The luminance generated by the LEDs in the sphere was a directly measured value to which standard uncertainty applies. The number of pixels in the ROI was exact, but there was some error due to the indistinct border between the exposed and unexposed pixels.

Luminance was manually recorded at the time of data collection. The limit of precision for stimulus luminance was based on the readout of the integrating sphere controller. Standard uncertainty was used as defined in Equation 10

$$u = \frac{a}{\sqrt{3}} \tag{10}$$

where a is the half-width of the range of reported values, and u is the reported uncertainty [26].

For example, luminance measurements for the brightest stimulus ranged between 220fL and 230fL across configurations. The standard uncertainty associated with those measurements according to Equation 10 is 2.89fL. Converting each value to cd/m^2 gives measured values between $754 \pm 9.89 cd/m^2$ and $788 \pm 9.89 cd/m^2$. Thus the range of actual luminance is between $751 - 9.89 = 744 cd/m^2$ and $788 + 9.89 = 798 cd/m^2$. The center of this range is $771 cd/m^2$ with an uncertainty of $(798 - 771) = 27 cd/m^2$, which is 3.5% of 771. Thus the reported value is $771 cd/m^2 \pm 3.5\%$. The other three luminance uncertainties are calculated in the same manner and shown in Table 3. The lowest luminance shows a high degree of uncertainty because the instrument

Table 3. Luminance Uncertainty

$L_v(cd/m^2)$	$\pm\%$
9.3	19
204	4.7
511	5.3
771	3.5

was approaching the low end of its functional range, between 2.2 and 3.3fL, so even a slight variation in readout ultimately caused a large percent uncertainty.

Pixel Stimulation Percent (PSP) was manually determined by selecting a center and radius of a circle of responding pixels (see Figure 9). While the number of selected pixels could be counted exactly, selecting the correct pixels was less precise and invoked some degree of error. While the border of the stimulated region was not sharply defined, the region was easily definable to within 5% of its radius, so error bars are calculated based on the relative error of $radius = r \pm 5\%$ in pixels. The pixel count is the area $a = \pi r^2$. Because the relative error follows r which is squared, the error gets multiplied by 2 such that

$$a = \pi r^2 \pm 10\%, \tag{11}$$

where a is the number of pixels being stimulated[27]. To normalize for the varying sensor array size between cameras, a is divided by the number of pixels $N_{pix-tot}$ in the entire array and multiplied by 100% in Equation 12:

$$PSP = \frac{a}{N_{pix-tot}} * 100\%. \tag{12}$$

This term appears on the horizontal axes of Figures 15, 16, and 17. Error bars are included in those figures, but due to the compressive nature of log scales, most are smaller than the markers themselves.

4.3 Time surfaces

In this section, time surfaces for one scene luminance are presented for every PSP for each camera. The time surfaces for the largest stimulus showed the observed tendencies most clearly, so the surfaces for lower stimuli are excluded. Significant

trends for each camera are highlighted and discussed.

As introduced in Section 2.5.1, a time surface is one way of visualizing the events generated by a camera in response to activity in the scene. The time surfaces often have a slight positive skew, meaning that the last reporting pixel is farther from the median than the first reporting pixel. This is consistent with the discussion and Figure 3 in [20]. The result is that zero deviation is often not in the exact center of the scale for the time surfaces. One time surface was generated for each of 21 stimuli. Combining them to create an "average" time surface destroyed patterns that are obvious in individual surfaces, so individual surfaces were selected for presentation that are representative of observed behavior.

Every time surface appears slightly different though general trends are apparent. Only the time surfaces produced in response to the brightest stimulus, $771 cd/m^2$, are shown. Time surfaces from dimmer stimuli demonstrated similar but less obvious behavior. For the sake of brevity, they are excluded from the main discussion. In all time surfaces presented, earlier pixels are colored lighter while later pixels are colored darker.

4.3.1 DAVIS240

The DAVIS 240 camera sensor was subjected to 100% PSP by both the 20mm and 40mm iris diameters, making the 40mm data redundant. The 40mm data was therefore excluded from analysis. Time surfaces from the DAVIS240 are shown in Figure 10. Parts (a) through (e) do not show any apparent spatial patterns. Parts (f) and (g) show a gradient generally increasing from top to bottom. There is a row below which the timestamps appear to drop to near the global minimum within that surface, then continue increasing with lower rows. Detailed inspection of time surfaces from sets associated with (f) and (g) shows that 2-4 rows are often recorded

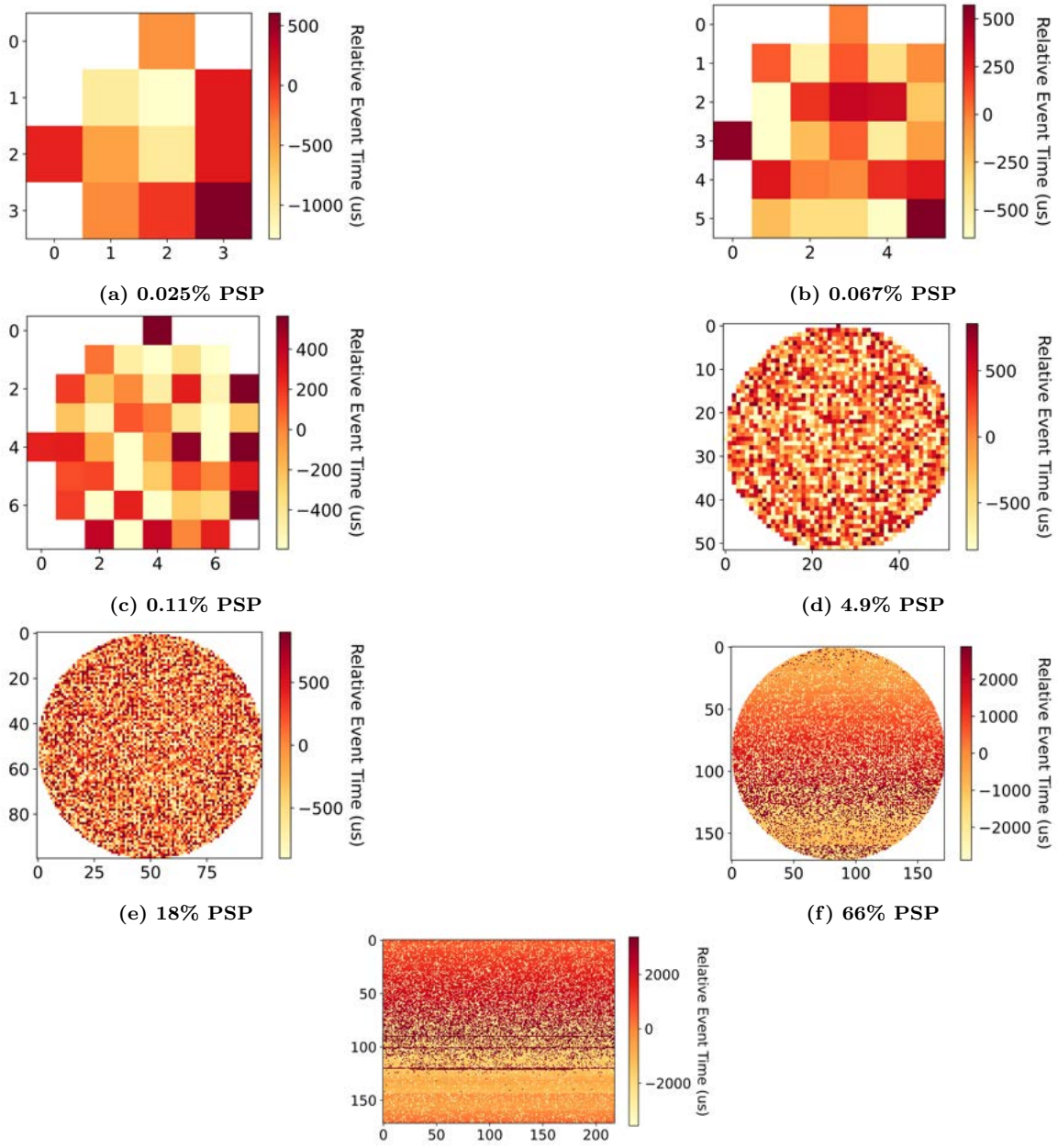
simultaneously, with most pixels in those rows receiving similar timestamps. While the trend is for lower (larger index) rows to report later, row timestamps do not increase monotonically, i.e. there are occasional rows or sets of rows which have slightly earlier timestamps than the row immediately above them. Occasionally pixels within each row report significantly earlier or later timestamps, with no apparent spatial pattern between outlying pixels.

4.3.2 DAVIS346

For low PSPs seen in Figure 11 parts (a) through (d), the DAVIS 346 appears to show a tendency for more central pixels to report events sooner than outer pixels. Beginning with 5.6% exposure in Part (e), columns begin to report events with similar timestamps. Parts (f) through (h) show a clear tendency for columns of pixels to report similar timestamps, non-monotonically increasing from right to left. Similarly to the DV240 timesurfaces, in parts (f), (g), and (h) there appears to be a small region of columns to the left of which the timestamps drop significantly then continue to increase. The location of this region of "reset" columns seems to randomly change between subsequent time surfaces, and is one reason that averaging several time surfaces was not useful. Though this reset region did not appear in every time surface, it was common enough to obscure the gradient in an average surface. Also similarly to the DV240, sets of 2-10 columns seem to report nearly simultaneously, with a subsequent set of columns sharing a later timestamp, though this behavior is not monotonic. Subsequent sets of columns differ in time by 10 to 100 microseconds.

4.3.3 XPlorer

The XPlorer shows a significant departure in behaviour from the DAVIS240 and DAVIS346. Figure 12 parts (a) through (d) show highly ordered columns of equal time



(a) 0.025% PSP

(b) 0.067% PSP

(c) 0.11% PSP

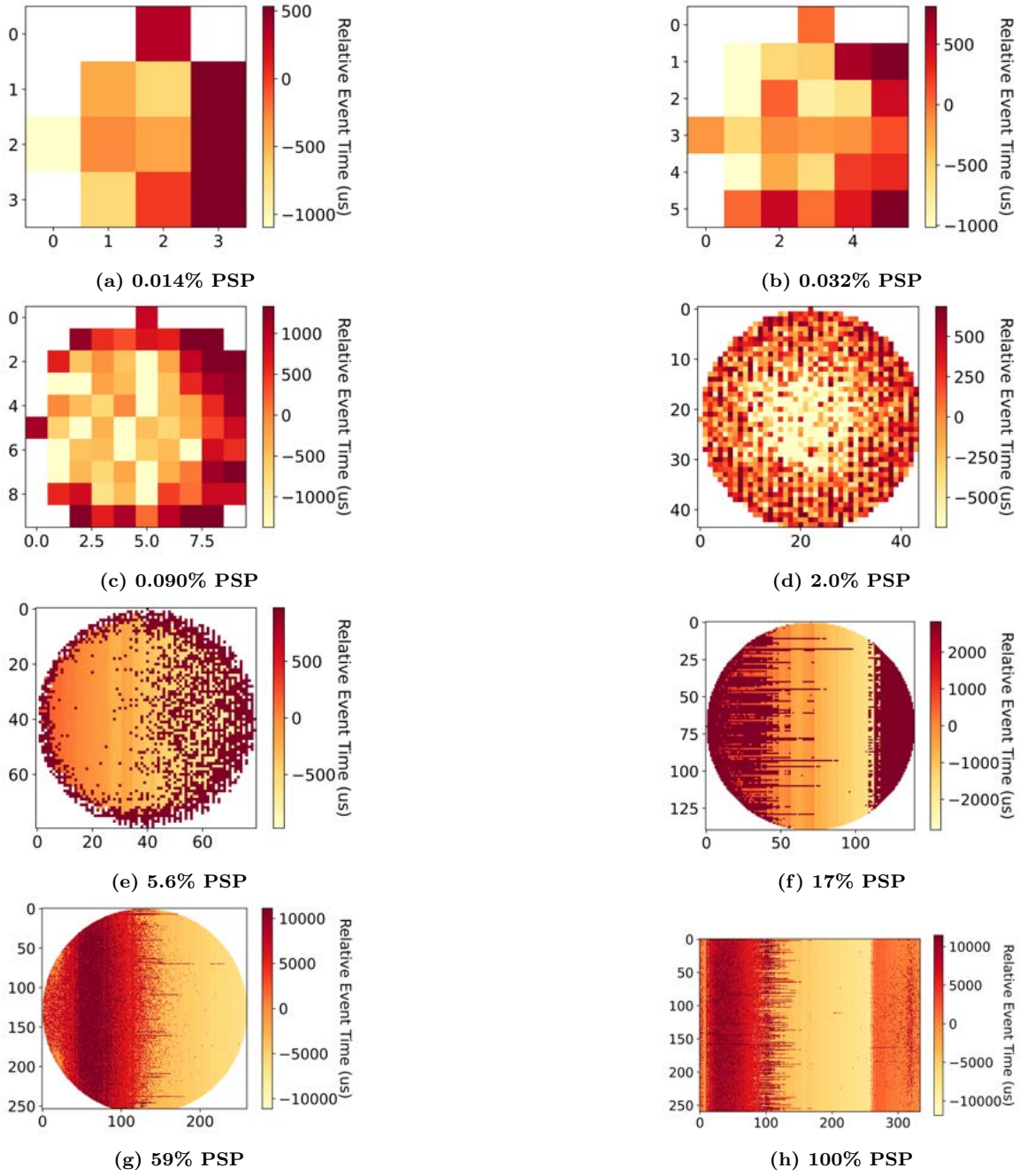
(d) 4.9% PSP

(e) 18% PSP

(f) 66% PSP

(g) 100% PSP

Figure 10. DAVIS240 Individual Time Surfaces. Time surfaces show the relative time at which stimulated pixels produce their first event in microseconds. Parts (a) through (g) show an increasing percentage of the array being exposed to the 771cd/m^2 stimulus. Note the appearance of a gradient in Parts (f) and (g), revealing a row-wise readout process. Each time surface shown is representative of the time surfaces generated from each configuration. The exposed area is circular, and the white space in the corners indicates pixels for which no data was recorded, which the exception of part (g)



(a) 0.014% PSP

(b) 0.032% PSP

(c) 0.090% PSP

(d) 2.0% PSP

(e) 5.6% PSP

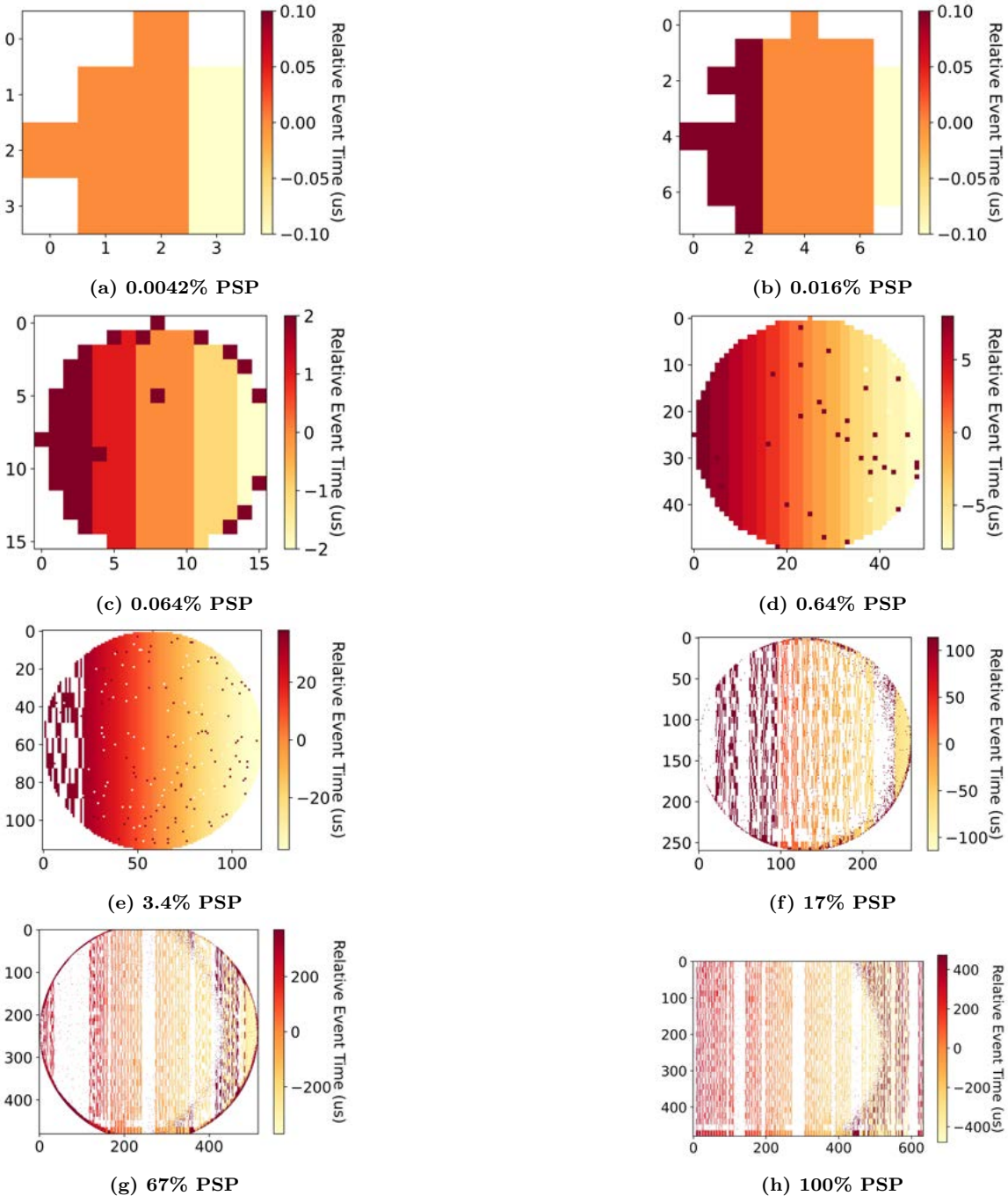
(f) 17% PSP

(g) 59% PSP

(h) 100% PSP

Figure 11. DAVIS346 Individual Time Surfaces. Time surfaces show the relative time at which stimulated pixels produce their first event in microseconds. Parts (a) through (h) show an increasing percentage of the array being exposed to the 771cd/m^2 stimulus. A gradient appears in part (e), suggesting that the camera transitions to an ordered readout process around 5.6% PSP. Each time surface shown is representative of the time surfaces generated from each configuration. The exposed area is circular, and the white space in the corners indicates pixels for which no data was recorded, which the exception of part (h).

increasing from right to left, including even the lowest PSP. In each case, sets of 3 to 5 columns report the exact same timestamp, separated by only 1 or 2 microseconds. Unlike the previous cameras, these sets of columns do increase monotonically in time. That is, each set of columns is greater than the set immediately to its right. There are occasional pixels scattered throughout the region which report earlier or later than the rest of their column. The deviation of timestamps between columns is significantly lower for the XPlorer than it was for either previous camera, and there are far fewer outlying pixels that do not follow the column-wise pattern. Another significant departure of behavior is seen beginning with the 3.4% exposure shown in Part (e) and increasing through Part (h): broken columns of seemingly missing data. The columns of broken white space indicate pixels which did not report an event within approximately 100 milliseconds of the stimulus. Each time surface in a set had seemingly random columns containing missed events, indicating there was no apparent preference for some columns to miss events over others.



(a) 0.0042% PSP

(b) 0.016% PSP

(c) 0.064% PSP

(d) 0.64% PSP

(e) 3.4% PSP

(f) 17% PSP

(g) 67% PSP

(h) 100% PSP

Figure 12. XPlorer Individual Time Surfaces. Time surfaces show the relative time at which stimulated pixels produce their first event in microseconds. Parts (a) through (h) show an increasing percentage of the array being exposed to the 771cd/m^2 stimulus. Each time surface shown is representative of the time surfaces generated from each configuration. The exposed area is circular, and the white space in the corners indicates pixels which were not stimulated and thus not included in the time surface. The white pixels inside the circular region, however, indicate pixels which did not generate events for that stimulus.

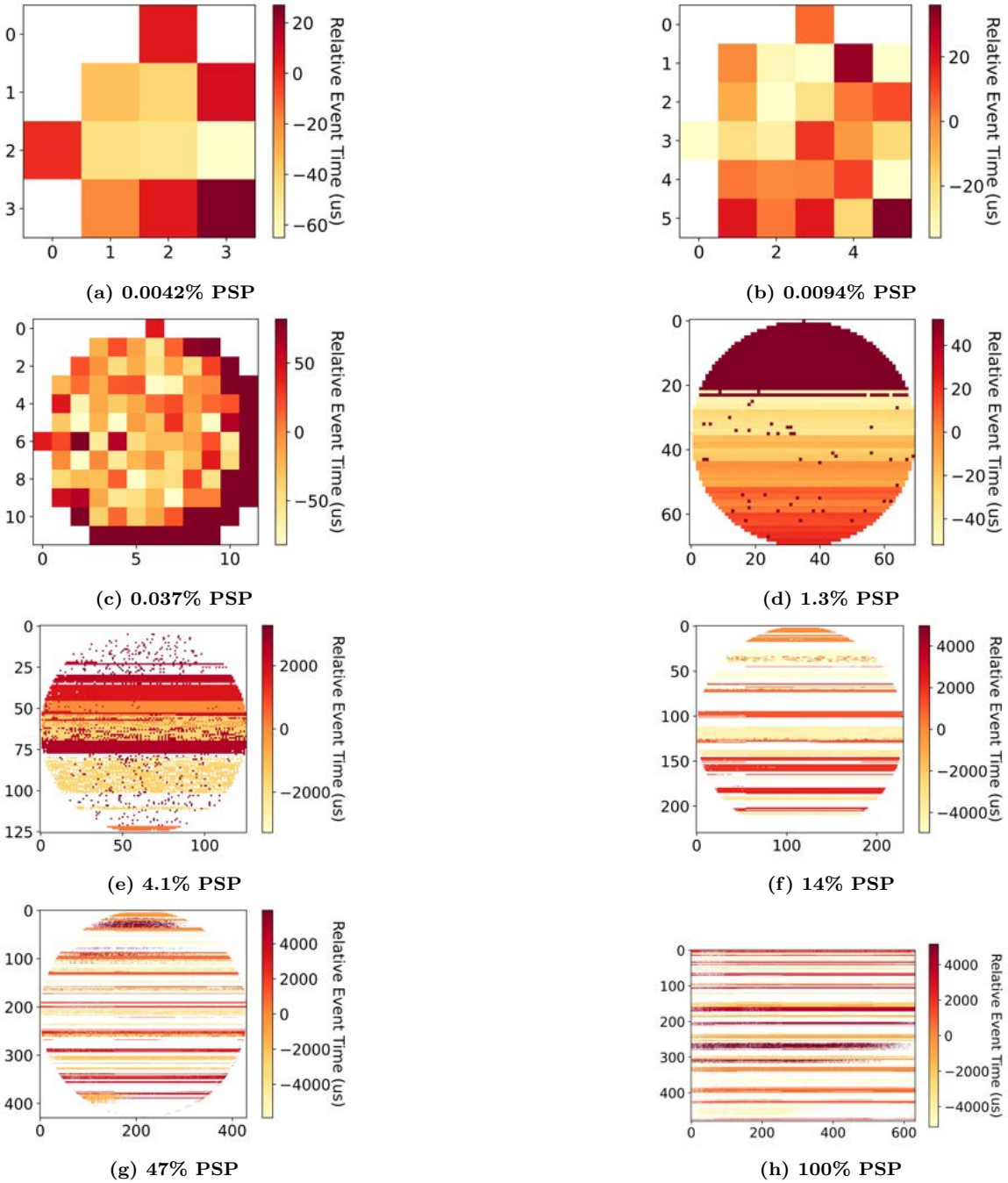
The total time difference between the first and last columns was as small as 2 microseconds for the smallest PSP (0.0042%) and increased to approximately 800 microseconds for the full array (100%). The time seemed to increase proportionally to the number of columns exposed to the stimulus.

Similarly to the DV346, in Parts (e) through (g), the outermost pixels tend to report events later than the majority of the center pixels, generating the thin dark ring around the periphery. These pixels may report events later due to lower illuminance. Unlike the previous cameras, there does not appear to be a tendency for timestamps to shift significantly at some boundary. For this reason, an average time surface is representative of the individual time surfaces, and columns typically have similar deviations between trials.

Parts (g) and (h) show a unique feature: an off-center circular region where fewer pixels report events, with a smaller radius than the defined ROI. This structure does not appear in time surfaces from any other camera. Since the same lens and physical setup was used for all four cameras, this feature is believed to be an artifact induced by the XPlorer camera itself. A series of back-reflections between the lens and the protective glass cover on the sensor may be responsible for this artifact.

4.3.4 Prophesee

The Prophesee camera produced time surfaces showing similarities to both the DAVIS346 and XPlorer sensors. As seen in Figure 13 parts (a) through (c), at low PSPs there do not appear to be spatial patterns in timestamps. Beginning at 1.3% exposure in Part (d) and continuing through Part (h), rows share timestamps. Part (d) shows a clear gradient increasing from top to bottom with the exception of the upper third.



(a) 0.0042% PSP

(b) 0.0094% PSP

(c) 0.037% PSP

(d) 1.3% PSP

(e) 4.1% PSP

(f) 14% PSP

(g) 47% PSP

(h) 100% PSP

Figure 13. Prophesee Individual Time Surfaces. Time surfaces show the relative time at which stimulated pixels exposed to the 771cd/m^2 stimulus produce their first event in microseconds. Parts (c) and (d) show the transition from asynchronous to ordered readout behavior. Moving to part (e) shows the rapid reduction in detection rate. Parts (f) through (h) show that rows tend to be either simultaneous or don't report events at all. Each time surface shown is representative of the time surfaces generated from each configuration. The exposed area is circular, and the white space in the corners indicates pixels which were not stimulated and thus not included in the time surface. The white pixels inside the circular region, however, indicate pixels which did not generate events for that stimulus.

Parts (e) through (h) show an increasing number of rows containing pixels which did not report events, in a similar manner to those columns from the XPlorer. Like the XPlorer's columns, there was not a preference toward specific rows regarding which pixels did or did not report an event.

4.4 Throughput

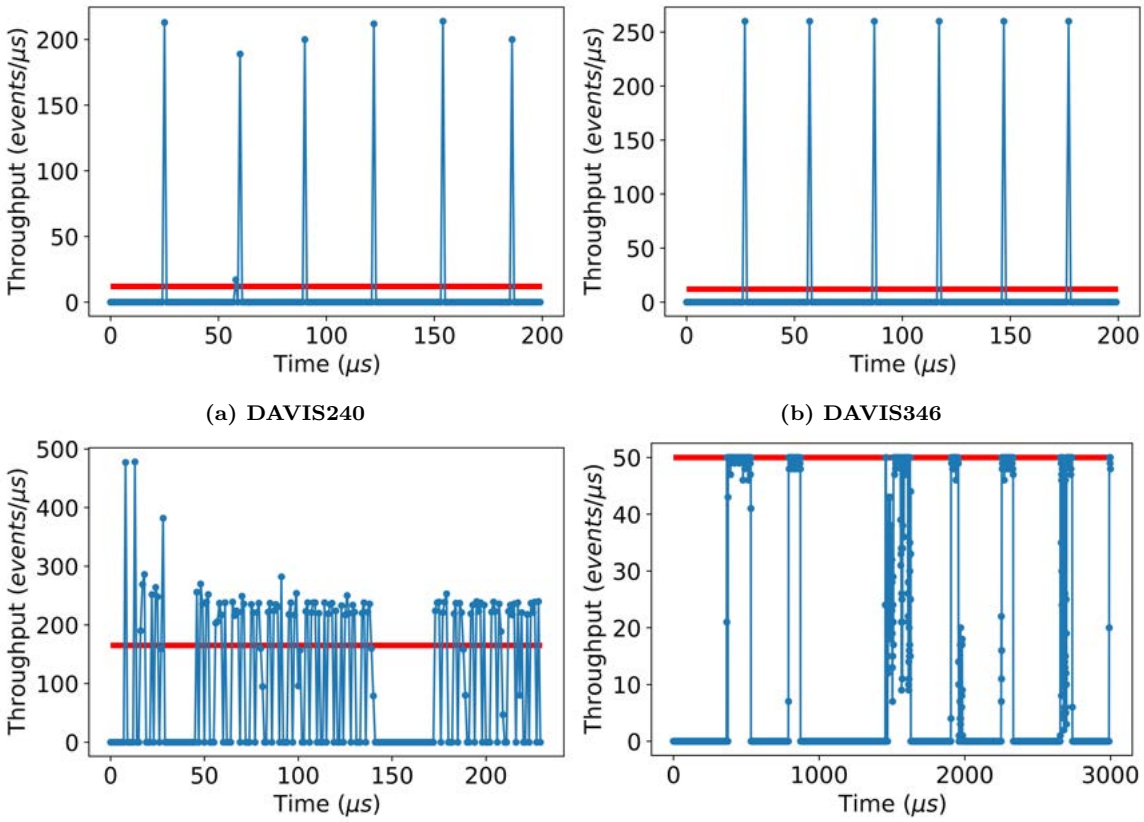
Event throughput describes the number of events being timestamped per microsecond. It is plotted here as a function of time, by creating histograms of number of events per microsecond on the vertical axis, versus time in microseconds on the horizontal axis. In order to explore sensor behavior under maximum load, Figure 14 presents data from $771 cd/m^2$ stimuli exposing 100% PSP. Every camera demonstrated non-uniform throughput, though the DAVIS240 and DAVIS346 did appear to operate in a consistent repeating pattern. The long-term (100-μs-scale) average throughput is reported in Table 4. Figure 15 shows the peak throughput with respect to the maximum specified throughput, experienced at each PSP and each stimulus amplitude.

Table 4. Effective Throughput (Meps)

	DAVIS 240	DAVIS 346	DV XPlorer	Prophesee Gen3
Advertised	12	12	165	50
Observed	6.25	8.67	113	10
Observed Peak	223	260	478	50

4.4.1 DAVIS240

The DAVIS240 demonstrated a spike of roughly 200 events every 32 microseconds, clearly visible in Figure 14(a). This is equivalent to a long-term average of 6.25 Million events per second (Meps), falling short of the advertised throughput capability by



(a) DAVIS240

(b) DAVIS346

(c) DVXPlorer

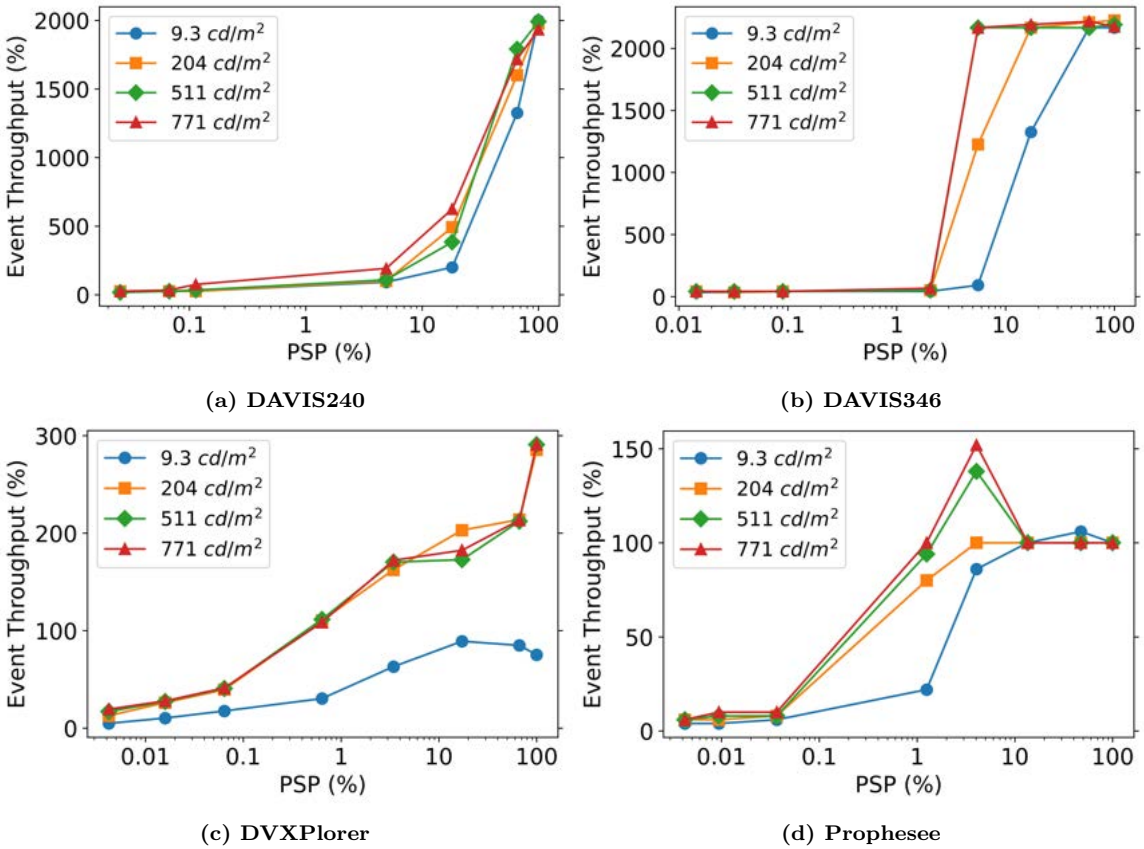
(d) Prophesee

Figure 14. Event Throughput. The histograms of events per microsecond versus time shows delays experienced by each camera when reading out events from the full array. The red horizontal line in each plot shows the maximum advertised event throughput for each camera (12, 12, 165, and 50 $events/\mu s$ respectively). The DAVIS240, DAVIS346, and DVXPlorer all exceed their advertised throughput on single-microsecond scales, while the Prophesee never exceeds its advertised throughput. In all cameras, the 100-μs-scale average throughput was lower than the advertised maximum throughput. These data were taken from the 771cd/m^2, 100% PSP trials.

roughly half. The interval between event spikes is very consistent, varying only by a few microseconds.

4.4.2 DAVIS346

The DAVIS346 produced a spike of precisely 260 events every 30 microseconds in Figure 14(b). Its similarity to the DAVIS240 is not surprising to due the extensive similarities in design between the two cameras. Its long term average event rate is



(a) DAVIS240

(b) DAVIS346

(c) DVXPlorer

(d) Prophesee

Figure 15. Event Throughput vs PSP. The peak throughput of each camera compared to the manufacturer-specified maximum throughput. Both the DAVIS240 and DAVIS346 exceeded their rated maxima by a factor of 20 at high PSP. The DVXPlorer exceeded its maximum rated throughput by a factor of 3. The Prophesee only exceeded its maximum throughput at 4.1% PSP. However, the actual time at which the throughput exceeded the limit could not be found, so it is suspected that it did not actually happen and an error in the processing code is responsible for the two points above 100%.

8.67 Meps, closer to but still short of the advertised throughput.

4.4.3 XPlorer

The DVXPlorer demonstrates a significant departure in behavior from the other iniVation cameras in Figure 14(c). This is not surprising due to its readout architecture being substantially different. The camera produced two spikes of nearly 500 events then quickly settled down to spikes of approximately 230 events each. While difficult to discern at the presented scale, close inspection of the cluster from the 50μs

to 140μs marks shows there are typically 2-4 microseconds with roughly 230 events each, followed by 1-3 microseconds of zero events, repeating several dozen times before relatively long period (10-50 microseconds) of zero events. In the particular sample shown in here, 15587 events occur before the 150μs mark, averaging to a long-term rate of 91 to 113 Meps, depending on if the 33μs gap at the 150μs mark is included.

4.4.4 Prophesee

Notably, the Prophesee displays significantly different behavior than the other cameras. Again this is unsurprising as it was designed by a separate company than the first three. The Prophesee never exceeded its advertised throughput of 50 events/microsecond, and only met that throughput limit irregularly. In Figure 14(d), the first cluster contains 8181 events, the second contains 4090, the third 6337, with 28482 total events shown in the frame, resulting in a long term average of roughly 11 Meps, only 22% of its advertised event throughput. This demonstrates the need for standardizing industry terms. While iniVation seems to define maximum throughput as a long-term average, Prophesee uses it to describe a limit on instantaneous rates. It would be beneficial to the industry to clarify these terms.

4.5 Median Absolute Deviation (MAD)

In this section, the Median Absolute Deviation is described in the context of event-based time surfaces. The MADs for all cameras and configurations tested are presented graphically in Figure 16 and general trends are highlighted. Then key observations regarding each camera are described. Specific data for each camera is included in tables for detailed comparison.

The Median Absolute Deviation condenses the temporal information from an entire time surface into a single statistic by calculating each pixels' deviation from the

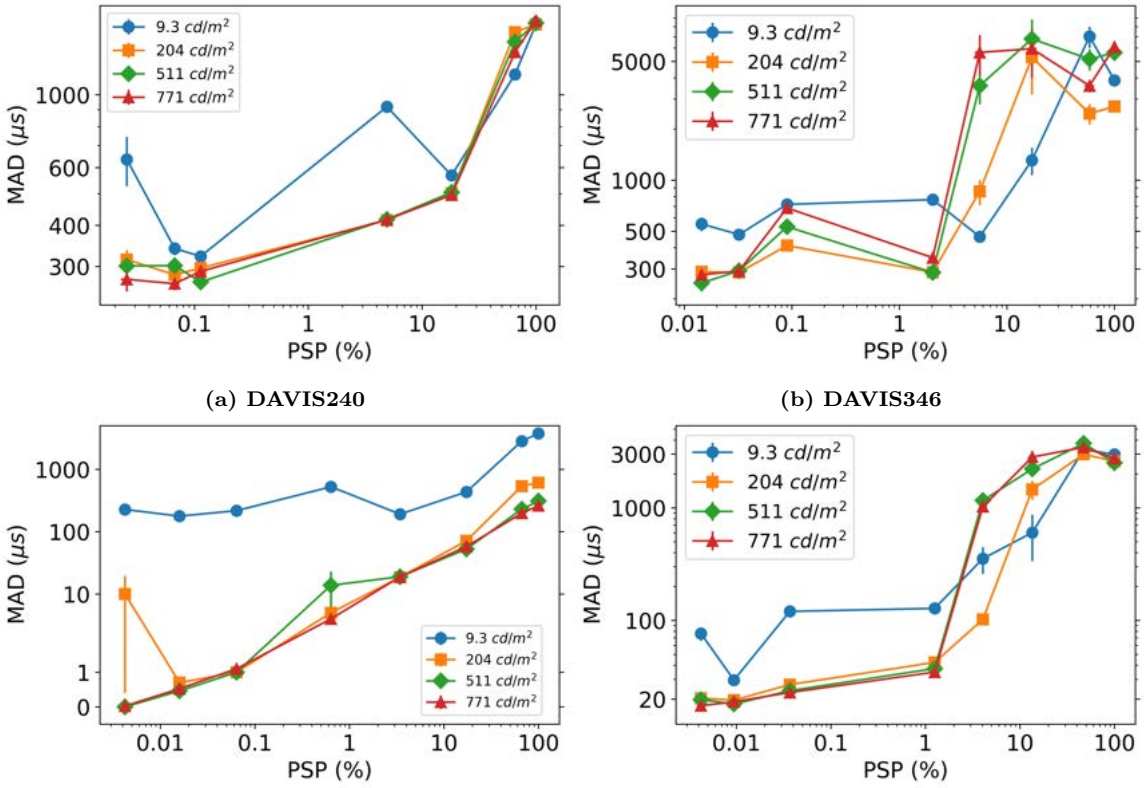
median timestamp, then calculating the median of those deviations as previously seen in Figure 5 of Section 2.5.2. The MAD is a measure of spread, much like the standard deviation. However, where the standard deviation relies on the arithmetic mean, the MAD relies on the median. Using the median makes the metric less sensitive to outliers. As can be observed in the time surfaces in the previous section, there are often pixels with timestamps significantly different than the majority of pixels in their neighborhood. This relatively large number of outliers would significantly skew the mean and standard deviations, and so those statistics would not accurately describe the behavior of the cameras.

Functionally, the MAD describes the variability in latency, or timing accuracy, of each camera when exposed to a spatially-extended simultaneous stimulus. The MAD is a single number which summarizes the camera's response to a stimulus. In order to describe the behavior of the cameras with high confidence that only real tendencies are reported, the MAD was calculated individually for $n = 21$ stimuli (trials) per configuration, and the mean of that set of 21 values is reported in this section. The standard deviation σ of the set was calculated and used to report the standard uncertainty u of the set as defined in Equation 13:

$$u = \frac{\sigma}{\sqrt{n}}. \tag{13}$$

The mean is used to describe this set of 21 MADs because there were no severe outliers that would skew the mean. Figure 16 shows the MAD calculated for each camera. The vertical axis shows the MAD in response to increasing PSP from left to right in each plot. All four levels of source luminance are depicted in each plot. Uncertainty bars are included for both horizontal and vertical axes based on the calculated uncertainty u, though in most cases the uncertainty bars are smaller than the markers themselves. Tables 5-8 give the precise values. When the uncertainty was smaller than 0.01, it

was listed as 0. The lower PSPs tend to have larger uncertainty in MAD, due to the small number of pixels illuminated and the corresponding increase in standard deviation σ. Figure 16 summarizes the following subsections. The deviation tends to increase with PSP for all cameras at all stimulus intensities.



(a) DAVIS240

(b) DAVIS346

(c) DVXPlorer

(d) Prophesee

Figure 16. Median Absolute Deviations. The MAD of every camera tends to increase with pixel stimulation percent (PSP). Error bars for both horizontal and vertical axes are calculated for each point, but are smaller than the markers in most cases. Parts (a), (b), and (d) are log-log scales, while Part (c) uses an adapted log-log scale in which the range between 0 and 1 on the vertical axis is linear. The deviation is precisely zero at the lowest PSP for the XPlorer sensor.

4.5.1 DAVIS240

Figure 16(a) shows that for low PSPs, the DAVIS240 deviation is highly dependent on scene illuminance. Table 5 contains the data shown in Figure 16. The DAVIS240's lowest deviation was $268\mu s$ at 0.11% exposure, when subjected to the $511cd/m^2$

stimulus. The highest deviation was 1655μs at 100% for the same luminance. This is nearly twice the predicted MAD for these conditions.

Table 5. DAVIS240 Median Absolute Deviations (μs)

	Pixel Stimulation Percent (PSP)						
L_v (cd/m^2)	0.025%	0.067%	0.11%	4.9%	18%	66%	100%
9.3	636±109	340±16	322±10	920±13	568±6	1154±17	1639±15
204	315±21	283±15	297±11	415±5	501±16	1552±16	1639±20
511	301±22	302±11	269±11	417±4	504±20	1449±32	1655±19
771	274±22	266±12	290±14	416±5	495±18	1348±32	1682±16

4.5.2 DAVIS346

The DAVIS346 generated the highest MAD of all cameras. Table 6 shows the highest deviation was 6780μs at 17% PSP to 511cd/m^2. The lowest deviation was similar to the DAVIS240, at 248μs with minimal (0.014%) exposure to 511cd/m^2. When 100% of the array was stimulated, the maximum deviation was 6086μs responding to the 771cd/m^2 stimulus, more than 3 times larger than the predicted MAD of 1874μs.

Table 6. DAVIS346 Median Absolute Deviations (μs)

	Pixel Stimulation Percent (PSP)			
L_v (cd/m^2)	0.014%	0.032%	0.09%	2.0%
9.3	552±51	478±20	720±29	767±4
204	289±23	286±14	412±12	287±2
511	248±27	293±21	532±26	286±2
771	277±23	292±23	686±25	349±2

L_v (cd/m^2)	5.6%	17%	59%	100%
9.3	464±2	1309±239	7009±998	3868±265
204	859±146	5287±2096	2463±337	2713±27
511	3600±805	6780±2055	5155±737	5634±97
771	5646±1507	5937±1930	3592±299	6086±58

4.5.3 XPlorer

The XPlorer demonstrated the lowest deviation overall, with zero deviation from the median at 0.0042% in response to both $511cd/m^2$ and $771cd/m^2$ stimuli. It maintained extremely low deviations for subsequent PSPs. The highest deviation demonstrated by the XPlorer was $3771\mu s$ when 100% of the array was exposed to $9.3cd/m^2$. The MAD was 609, 310, and $262\mu s$ for increasing luminance. The expected minimum MAD was $465\mu s$. However, Section 5.2 discusses why a direct comparison may not be appropriate. Additionally, the XPlorer is the only camera for which the lowest scene luminance always generated the highest deviation for all PSPs. Table 7 contains the specific datapoints.

Table 7. XPlorer Median Absolute Deviations (μs)

	Pixel Stimulation Percent (PSP)			
L_v (cd/m^2)	0.0042%	0.016%	0.064%	0.64%
9.3	226.9±16	177.6±13	217.3±13	524.8±39
204	10.0±10	0.7±0	1.0±0	5.0±0
511	0.0±0	0.5±0	1.0±0	13.8±9
771	0.0±0	0.5±0	1.1±0	4.0±0
L_v (cd/m^2)	3.4%	17%	67%	100%
9.3	190.9±1	434.2±8	2831.3±52	3771.1±54
204	18.6±0	71.2±0	536.9±15	609.4±12
511	19.0±0	52.6±0	230.0±5	310.4±8
771	18.9±0	57.8±2	197.6±6	262.4±7

4.5.4 Prophesee

The Prophesee sensor produced the lowest deviation ($17\mu s$) when the smallest fraction of the array (0.0042%) was exposed to the brightest stimulus ($771cd/m^2$). This was the second smallest deviation after the XPlorer. The Prophesee's largest deviation was $3734\mu s$ when 47% of the array was exposed to the $511cd/m^2$ stimulus.

The minimum predicted MAD was 1536μs, which the Prophesee exceeded by more than a factor of 2. Table 8 contains the specific datapoints.

Table 8. Prophesee Median Absolute Deviations (μs)

	Pixel Stimulation Percent (PSP)			
L_v (cd/m^2)	0.0042%	0.0094%	0.037%	1.3%
9.3	76±11	30±1	120±4	128±15
204	20±2	20±1	27±1	42±2
511	20±1	18±1	24±1	37±1
771	17±2	19±1	23±1	35±2
L_v (cd/m^2)	4.1%	14%	47%	100%
9.3	354±95	603±267	3286±81	2987±14
204	102±6	1458±278	2994±248	2587±52
511	1162±207	2216±273	3734±325	2512±67
771	1024±105	2825±352	3449±283	2748±39

4.6 Events Per Stimulus

In this section, the mean number of events generated by each pixel in response to a stimulus is described. The data is graphically summarized in Figure 17, and the data for each camera is broken out in Tables 9 though 12.

To calculate the number of events per pixel per stimulus (EPS), the total number of ON events generated by each pixel for the duration of the recording was divided by the number of stimuli to which the camera was exposed. This produced an EPS for each pixel in the entire ROI. The mean and standard deviation of that set was calculated. The mean is reported in the following section, and the standard deviation was used to calculate the standard error for the uncertainty bars as in Section 4.5. The PSP is plotted on a logarithmic axis, while the events per stimulus is plotted on a linear axis.

It is well established by previous work that such characteristics as refractory period

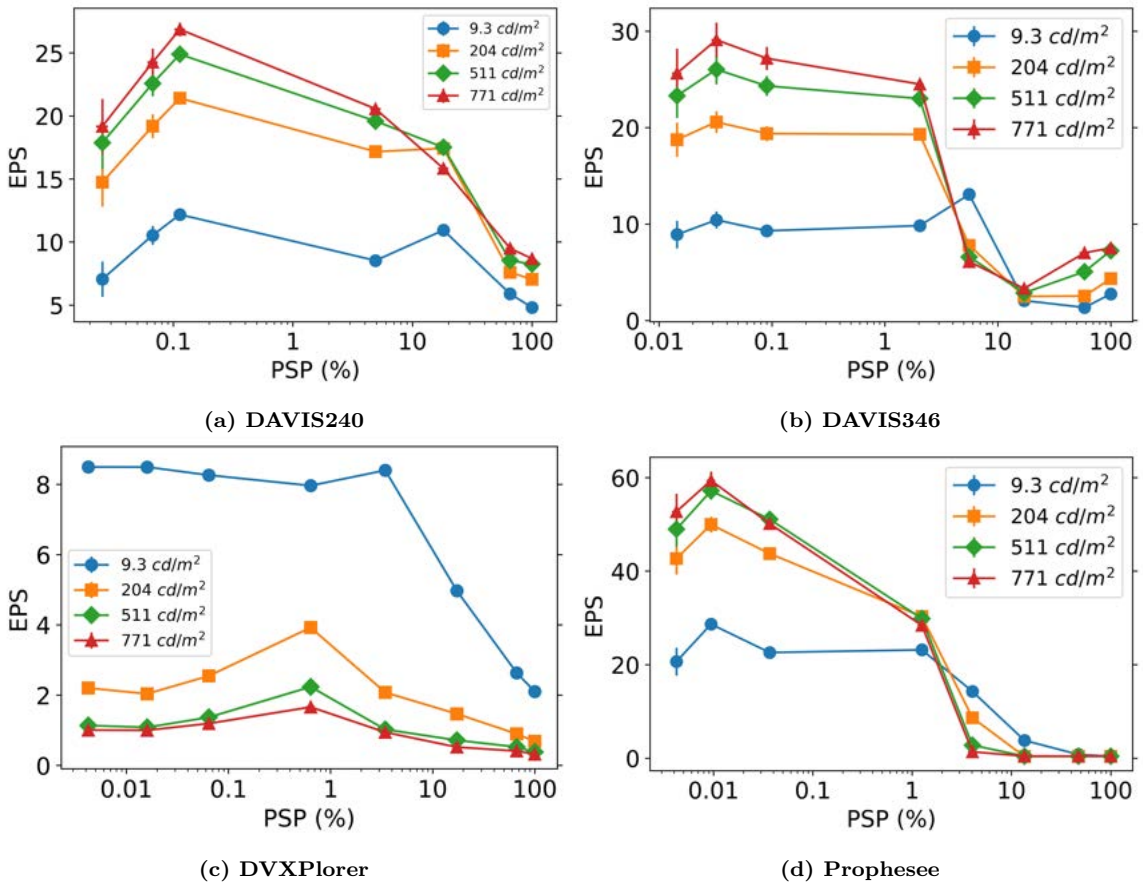
and sensitivity are controlled by various user-adjustable bias settings [23]. Since both of those parameters have a direct affect on the production of events in response to a stimulus, adjusting the bias settings of a camera will probably change the shape and magnitude of the curves shown in Figure 17. These data were recorded using the factory default settings for all parameters.

Every camera showed a decrease in reported events as PSP increased. Both the XPlorer and Prophesee dropped to less than 1 event per stimulus at the higher percentages, corresponding to the broken white row and/or columns seen in Figures 12 and 13, where many pixels did not produce an event for every impulse. The Prophesee camera produced the most events overall, nearly 8x more than the XPlorer, which produced the fewest events overall. The DAVIS346 produced slightly more events per stimulus than did the DAVIS240, in general. For all cameras, low PSPs increased the influence of scene luminance. Higher PSPs reduced the effect of luminance on event generation.

It's not possible to predict bandwidth utilization based on these numbers, because these only apply to the pixels in the ROI. Pixels outside the ROI produced highly varied numbers of events per stimulus. In the case of higher PSPs, for example, the pixels outside the illuminated ROI produced several events per stimulus while the pixels inside the ROI dropped to below 1 event per stimulus. In order to predict bandwidth requirements, the event generation rate of the entire array must be considered. In general though, higher PSP and higher illuminance both increase bandwidth utilization.

4.6.1 DAVIS240

The DAVIS240 generated between 4 and 12 events per stimulus at the lowest luminance amplitude of $9.3 cd/m^2$, and produced between 8 and 27 events per stimulus



(a) DAVIS240

(b) DAVIS346

(c) DVXPlorer

(d) Prophesee

Figure 17. Events per Stimulus (per pixel). The EPS of every camera tends to decrease with pixel stimulation percent (PSP). Error bars for both horizontal and vertical axes are calculated for each point, but are smaller than the markers in most cases.

at the highest luminance. For each scene luminance amplitude, the camera produced the most events when 0.11% of the array was exposed. Higher scene luminance caused a higher EPS for every PSP except 18%. Table 9 shows the mean event statistics.

4.6.2 DAVIS346

The DAVIS346 followed a similar trend as the DAVIS240. It generated between 1 and 11 events per stimulus at low luminance amplitude, and between 3 and 29 events per stimulus at the highest luminance amplitude. For all PSPs except 4.9%, an increased scene luminance corresponded to increased event generation. For every scene luminance step amplitude, an increasing PSP corresponded to generally reduced

Table 9. DAVIS240 Mean Events Per Stimulus

	Pixel Stimulation Percent (PSP)			
L_v (cd/m^2)	0.025%	0.067%	0.11%	4.9%
9.3	7.05±1.40	10.53±0.75	12.18±0.35	8.53±0.04
204	14.74±1.94	19.19±0.95	21.41±0.42	17.17±0.04
511	17.87±2.06	22.57±1.02	24.89±0.48	19.59±0.04
771	19.17±2.19	24.26±1.1	26.87±0.54	20.57±0.05
L_v (cd/m^2)	18%	66%	100%	
9.3	10.94±0.02	5.89±0	4.80±0	
204	17.42±0.02	7.59±0	7.04±0	
511	17.51±0.02	8.53±0	8.24±0	
771	15.84±0.02	9.49±0	8.68±0	

event generation. Table 10 shows the exact mean event statistics.

Table 10. DAVIS346 Mean Events Per Stimulus

	Pixel Stimulation Percent (PSP)			
L_v (cd/m^2)	0.014%	0.032%	0.090%	2.0%
9.3	8.91±1.42	10.43±0.89	9.30±0.62	9.84±0.04
204	18.75±1.75	20.58±1.13	19.38±0.77	19.30±0.06
511	23.31±2.29	26.04±1.55	24.33±1.04	23.02±0.07
771	25.64±2.56	29.10±1.80	27.18±1.21	24.51±0.09
L_v (cd/m^2)	5.6%	17%	59%	100%
9.3	13.07±0.03	2.07±0	1.35±0	2.75±0
204	7.80±0.03	2.50±0.01	2.54±0	4.35±0
511	6.59±0.05	2.86±0.01	5.07±0	7.24±0
771	6.11±0.04	3.32±0.01	7.02±0	7.52±0

4.6.3 XPlorer

The XPlorer produced the fewest events per stimulus of all the cameras. It was also unique in that scene luminance amplitude was negatively related to event generation. The lowest luminance generated the most events per stimulus (8.49 events per stimulus at 0.0042%), and for increasing luminance, the number of events per

stimulus monotonically decreased to 0.32 events per stimulus at 100% exposure. The XPlorer was also unique from the other cameras in that the rate of change of events per stimulus was lower with respect to PSP, particularly for higher scene luminance. Table 11 shows the exact mean event statistics.

Table 11. XPlorer Mean Events Per Stimulus

	Pixel Stimulation Percent (PSP)			
L_v (cd/m^2)	0.0042%	0.016%	0.064%	0.64%
9.3	8.49±0.17	8.49±0.08	8.26±0.06	7.96±0.03
204	2.20±0.08	2.04±0.03	2.54±0.08	3.92±0.04
511	1.13±0.05	1.08±0.02	1.37±0.05	2.23±0.02
771	1.00±0	1.00±0	1.19±0.03	1.66±0.02
L_v (cd/m^2)	3.4%	17%	67%	100%
9.3	8.40±0.01	4.97±0.01	2.63±0	2.10±0
204	2.08±0	1.47±0	0.90±0	0.68±0
511	1.02±0	0.71±0	0.52±0	0.38±0
771	0.94±0	0.52±0	0.41±0	0.32±0

4.6.4 Prophesee

The Prophesee camera generated the most events per stimulus of any camera at low PSPs, for all amplitudes of scene luminance. When 0.016% of the array was exposed to the $771cd/m^2$ stimulus, 59.38 events were generated. When 100% of the array was exposed to any stimulus, between 0.39 and 0.42 events were generated. The Prophesee camera tended to generate the fewest events at PSPs greater than 10%, with only the $9.3cd/m^2$ stimulus producing more than 1 event per stimulus with 17% of the array exposed. Table 12 shows the exact mean event statistics.

Table 12. Prophesee Mean Events Per Stimulus

	Pixel Stimulation Percent (PSP)			
L_v (cd/m^2)	0.0042%	0.016%	0.064%	0.64%
9.3	20.65±2.97	28.65±1.28	22.60±1.08	23.16±0.09
204	42.68±3.43	49.96±1.69	43.75±1.28	30.27±0.08
511	48.96±3.70	57.17±1.88	51.08±1.41	29.84±0.08
771	52.71±3.85	59.38±1.92	50.14±1.28	28.41±0.05
L_v (cd/m^2)	3.4%	17%	67%	100%
9.3	14.26±0.02	3.80±0	0.78±0	0.41±0
204	8.65±0	0.35±0	0.38±0	0.39±0
511	2.78±0	0.43±0	0.42±0	0.41±0
771	1.35±0	0.45±0	0.43±0	0.42±0

4.7 Summary

In this chapter, three key metrics of characterization were presented. Time surfaces for the highest scene luminance show spatial patterns associated with each sensor. It was noted that all four cameras demonstrated row- or column-wise readout tendencies when large fractions of the array were exposed. The Median Absolute Deviation described the spread of pixel response times for varying PSP and scene luminance. It was found that the MAD increased significantly for increasing PSP. Finally, the average number of events generated by each pixel in response to each stimulus was reported. The EPS for every camera dropped as larger fractions of each array was exposed. In the following chapter, relationships and trends will be presented and analyzed, resulting in several general conclusions.

V. Discussion

5.1 Overview

In the last chapter, visualizations of raw data in the form of time surfaces were presented. The information gleaned from these time surfaces was condensed into three metrics: throughput, the Median Absolute Deviation (MAD), and the Events per Stimulus per pixel (EPS). The MAD and EPS in response to configurations of varying PSP and scene luminance were reported. This chapter discusses key findings and trends arising from that data. The following sections propose possible reasons for these trendsand explore the implications for various applications. Finally, general conclusions are presented.

5.2 Throughput

Both the DAVIS240 and DAVIS346 showed spikes of high throughput when 100% of pixels were exposed to the stimulus. In the DAVIS240, roughly 200 events were reported simultaneously every time, corresponding to the simultaneous rows observed in the time surfaces. The DAVIS346 displayed the same tendency in columns. This suggests that when the event load on these cameras exceeds some threshold, the arbiter circuit transitions from the asynchronous process and behaves more like a rolling-shutter type system. The DVXPlorer follows a synchronous readout pattern, though that is not apparent in the throughput plot. It is possible that there is a form of on-chip memory that serves as a sort of buffer. The gaps in event transmission around the 40 and 150 μs marks may indicate where that buffer reached capacity, preventing the timestamping of any subsequent pixels until the queue was sufficiently reduced. However, because the long-term average event rate was well under the advertised limit, it is not clear why that capacity would have been reached. The

Prophesee showed similar gaps, suggesting that there is a maximum throughput that is lower than the advertised value. Overfilling a buffer is one possibility to describe the limited capability.

5.3 MAD

The median absolute deviation (MAD) describes the spread of time required for each pixel to respond to a stimulus and is one measure of timing precision. The MAD was more sensitive to PSP than to scene luminance. While this research provided sufficient data to describe relationships between these variables, confounding variables prevent formal statements of causality.

In every camera except the DVXPlorer, the MAD exceeded its predicted value under high throughput. It has been reported that if the arbiter circuit receives additional events while still processing prior events, those events queue up and lose their arrival order. This means secondary events from some Pixel A could be timestamped prior to the first event from some other Pixel B, if the sensor is experiencing a high throughput. McReynolds [12] described the pixel cutoff frequency as 3kHz.

Assuming this applies similarly to all cameras, this suggests a minimum event generation period of $333\mu s$. If a pixel continues generating events while the rest of the array is being read out, that pixel could generate between 5 and 20 additional events in the times shown in Table 1, before the sensor even finishes reading out the first events from other pixels. If the order of readout of secondary and primary events is random, this would cause some events to be timestamped much later than predicted, and force the MAD to be higher than the theoretical minimum.

While the DVXPlorer's MAD was lower than predicted, its pixels were only producing events in response to roughly 30% of stimuli. This means the event throughput was not as high as was assumed when making the prediction, so the comparison is

not meaningful. Because the camera demonstrated the ability to timestamp events at a higher instantaneous throughput than advertised, the reduction in detection rate cannot be attributed to that throughput limitation.

5.3.1 MAD vs Luminance

There is a weak correlation between MAD and luminance amplitude. At low PSP, the lowest scene luminance resulted in a noticeably higher MAD. However, at PSPs greater than 1%, and for scene luminance greater than $9.3cd/m^2$, there does not appear to be a significant correlation between luminance and MAD. This suggests that the luminance of a scene is not the most significant predictor of timing precision.

5.3.2 MAD vs PSP

Pixel Stimulation Percent has a far more significant affect on timestamp precision. There appears to be a correlation between median absolute deviation and PSP, such that MAD increases with PSP. It is suggested that this is because the timestamping mechanism is not specific to each pixel, so when many pixels report events nearly simultaneously, the timestamping circuit is unable to keep up and thus delays are incorporated into those events. To date, the timestamping function has been shared between pixels due to spatial limitations in pixel design. However, if each pixel were to independently generate its own timestamp, the timing accuracy with increasing PSP could be improved.

With each pixel conducting its own timestamping, the timing accuracy would be limited only by the time response of each pixel's analog circuit, which could then be tuned by the user for their specific application. This would result in consistent timing accuracy for all ranges of PSP, including up to 100% of pixels. A detailed cost-benefit analysis could be done by the designer or perhaps a sufficiently-funded customer to

compare the expense associated with increased complexity and explore what additional timing accuracy could be possible with that architectural modification.

Based on the right-to-left increase in timestamp seen on the XPlorer, it is predicted that the timing accuracy over the breadth of the stimulus would strongly depend on the orientation with respect to the sensor. For example, if the XPlorer imaged a lightning bolt stretching from top to bottom, there would likely be a microsecond-scale MAD, as a majority of events would fall within a few columns and thus get read out nearly simultaneously. On the other hand, if a lightning bolt stretched from side to side, the MAD would likely be much higher due to the image spanning the entire width of the sensor.

While a clear relationship exists between the MAD and the PSP, it cannot be said that PSP directly causes an increase in deviation. In general, declaring that X causes Y requires three elements: proving that X came before Y, that Y wouldn't have happened if not for X, and there is nothing else that accounts for the relationship between X and Y [28]. The challenge here lies in the third element. In this research, the PSP only describes a number of pixels without regard to number of rows/columns or the spatial distribution of those pixels, which may be significant. In other words, the number of pixels might not be the cause of increased MAD, the number of stimulated rows/columns might be. This is particularly easy to visualize with the XPlorer and Prophesee. As discussed above, it appears likely that an asymmetric object generating events would produce different deviations depending on whether was oriented parallel or perpendicular to the direction of column readout. Thus, the spatial arrangement of stimulated pixels could reasonably have an affect on MAD, which confounds the effect of the number of pixels and prevents any formal claim of causality.

5.4 EPS

The number of events generated by each pixel per stimulus was clearly affected by both scene luminance and array fraction. It was also observed that the XPlorer displayed a significantly higher number of "hot" pixels than the other cameras. Approximately 30 hot pixels each produced several hundred times more events than pixels exposed to the real image. Other cameras did not have nearly as many hot pixels of this nature, or filtered out those events through an intermediate filter. Each camera produced several events per stimulus under various configurations, though the XPlorer did not follow the same trend as the other cameras.

5.4.1 EPS vs Luminance

Illumination step size had a significant effect on the number of events generated per stimulus. The model proposed in [23] says that "If the change [in brightness] is multiple times the threshold, then multiple DVS events are generated". This appears to be consistent with the findings in this research, with the notable exception of the XPlorer. Each pixel in the XPlorer generated the most events in response to the dimmest stimulus and generated fewer events for each increase in stimulus magnitude, which seems to be directly counter to the model suggested in [23] and opposite the behavior of every other camera tested.

Not only was the XPlorer's strange behavior apparent when comparing different stimuli, it was also apparent within individual trials, confirming that it is a real phenomenon and not a error in data processing. In all stimuli throughout this research, the central region of the image on the sensor was evenly intense (spatially homogeneous), and the circular edges of the image dropped off in intensity according to the usual diffraction around a sharp edge, fading into the dark, unstimulated region on the periphery. Thus for every stimulus of every amplitude, there is a ring of decreasing

illumination with increasing radius.

Figure 18 shows the average number of events produced per stimulus for every pixel in the XPlorer's sensor. The example shown was for 3.4% PSP, $204cd/m^2$. The brightest central region caused pixels to produce roughly 4 events per stimulus. The ring of of reduced brightness produced more events, nearly 10 events per stimulus, while most unstimulated pixels produced no events. Detailed analysis of individual pixels in all three regions showed that the vast majority of events generated in these regions were concurrent with each other, suggesting that this behavior is a response to the stimuli and not from external stray light. The author is not aware of any pixel-level model which suggests a maximum event generation rate dependent on illumination. While it is beyond the scope of this document to analyze circuit-level behavior, this finding suggests the sensitivity of the DVXPlorer is highly dependent on the amplitude of the stimulus.

In Ryu's talk given in 2019 [18], he describes response time increasing under low illumination, resulting in more events being generated by moving edges. It is possible that the same circuit-level phenomenon causing that challenge is responsible for the negative relationship between luminance and EPS in the XPlorer camera. In his talk, Ryu suggests that implementing a global hold technique solved their problem, so it's likely that adjusting camera settings would have an affect on the EPS for the XPlorer.

5.4.2 EPS vs PSP

At low PSP, only the DAVIS240 and Prophesee demonstrated a noticeable relationship between PSP and the number of events generated per stimulus. However, above 1% exposure, every camera showed a significant drop in EPS with increasing exposure. At high exposures, both the XPlorer and the Prophesee showed a tendency for groups of rows or columns of pixels to miss stimuli, which is likely responsible for

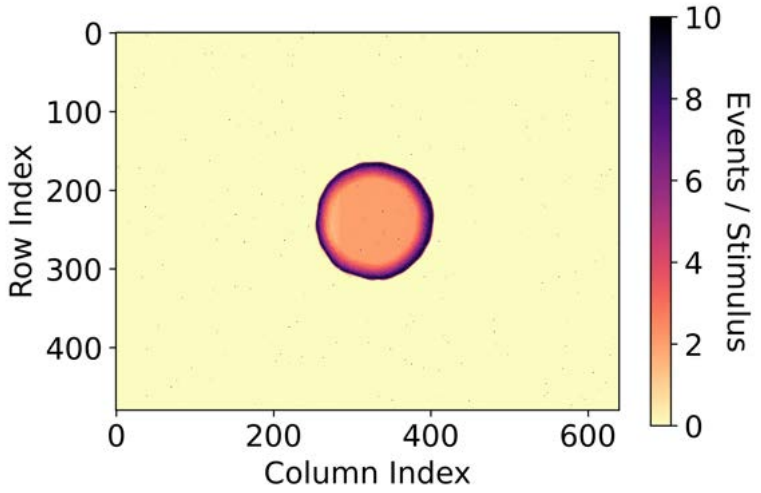


Figure 18. XPlorer Event-Rate Image. In this plot of the XPlorer's entire sensor, the average number of events produced by each pixel is represented by color. The luminance at the center of the stimulus was spatially homogeneous, and decreased at the edges due to diffraction until there was no detectable light beyond the edges of the circular region. The orange central region shows a low event generation rate, roughly 4 events per stimulus, when exposed to the relatively high luminance. The dark ring around the edge shows an increased event generation rate as luminance decreases, at nearly 10 events per stimulus. The light yellow majority of the image shows the "background" event generation rate, where there was effectively zero stimulation and zero events. This indicates that sensitivity is nonlinearly dependent on stimulus. This particular example shows the average event rate of the DVXPlorer when exposed to a 3.4% PSP, 204cd/m^2 stimulus.

the drop in EPS. Both of those cameras produced, on average, less than one event per pixel per stimulus. In several cases, the pixels produced fewer than 0.5 events per stimulus, suggesting that under those conditions, the probability of a pixel failing to respond is greater than 50%. This could be a significant concern for applications which require a consistent high quality image throughout the entire frame. On the other hand, that is still only representing individual pixels. As is clear in the time surfaces, even when many pixels do not respond, a large portion of the sensor does respond. If an application was attempting to detect a large and extremely fast event, then, it is likely that many pixels would respond, and the artifacts could potentially be used to suggest the nature of the detected activity.

Another key observation is that in most cases, each camera produced several events per stimulus. This is worth noting because each camera has a limited event throughput, and a major benefit of event-based cameras as a technological family is the inherent sparsity of data. If several events are generated from a single stimulus, those events produce excessive throughput and require additional storage space. Unless some information can be extracted from the number of events generated, such as the magnitude of change of the scene luminance or probability of real activity, the extra events are redundant and lower the overall efficiency of the system. Any application in which an event camera is used in orbit may be highly sensitive to this redundancy. Because strict downlink bandwidth limits may require maximum data-to-information efficiency, the size and contrast of anticipated targets should be predicted, and biases adjusted such that redundant event generation is minimized.

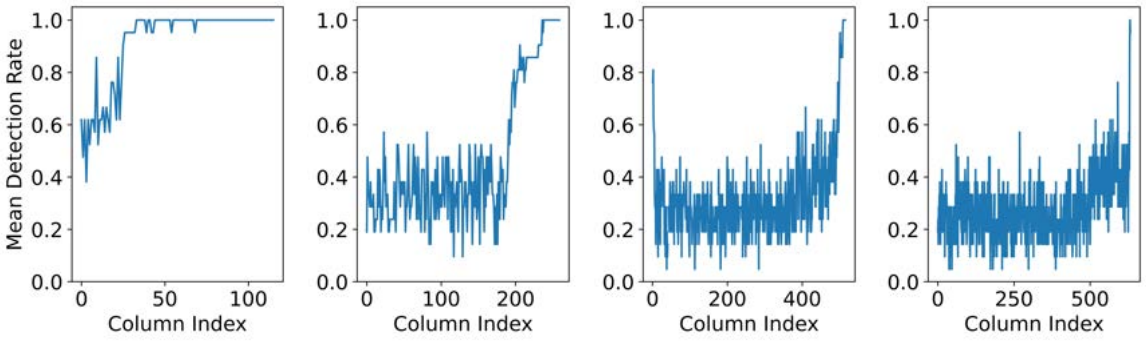
5.5 Detection Rate

When a large percentage of the DAVIS240 and DAVIS346 pixels were exposed to the stimulus, timing precision dropped but every pixel continued reporting events. In the XPlorer and Prophesee, when the PSP was high, a large and discontinuous fraction of the array stopped reporting events entirely. This effect appears in the time surfaces of Figures 12 and 13 as broken white rows or columns of no data.

The detection rate describes the fraction of stimuli to which a row or column of pixels generates events. Recall in Figures 12 and 13, where beginning at 3.4% and 4.1% respectively, individual rows or columns appeared not to generate events. To determine if each camera exhibits a preference toward dropping one region of rows or columns over another, the fraction of stimuli to which a row/column responded was calculated. This effect was most significant when cameras were exposed to the $771 cd/m^2$ stimulus, so further analysis focuses on that case. To calculate this metric,

each time surface's center row or column for the XPlorer or Prophesee, respectively, was assumed to be a cross-section of the array, representative of the behavior of all rows or columns. This can be verified by visual inspection of the time surfaces in Figures 12 and 13.

For each pixel in that representative row (column), the number of times it generated at least one event in response to a stimulus was counted and divided by the number of stimuli, producing a metric ranging from 0 to 1. For example, if a pixel's detection rate was 1, it generated at least one event for every stimulus. If a pixel's detection rate was 0, it generated no events for any stimulus (though this was never actually observed). This is similar to the Events per Stimulus metric used previously. The detection rate, however, provides spatial information regarding row (column) preference, whereas the EPS did not. This metric reveals a spatial tendency for a camera to stop reporting events from one region over another.



(a) 3.4% PSP (b) 17% PSP (c) 67% PSP (d) 100% PSP

Figure 19. XPlorer Column Detection Rates. This data for the DVXPlorer's response to the 771cd/m^2 stimulus shows that once 3.4% of columns were stimulated in (a), the left-most columns began responding to fewer stimuli. As a larger fraction of the array was stimulated, parts (b), (c), and (d) show fewer columns responding to every stimulus, and the diminished columns responding to fewer stimuli resulting in a progressively lower detection rate.

At 3.4% PSP, the XPlorer camera's left-most columns begin missing stimuli. This is shown by the dip in Figure 19(a). As PSP increases through Figure 19(b) through (d), columns further and further right begin to miss stimuli, until at 100% exposure,

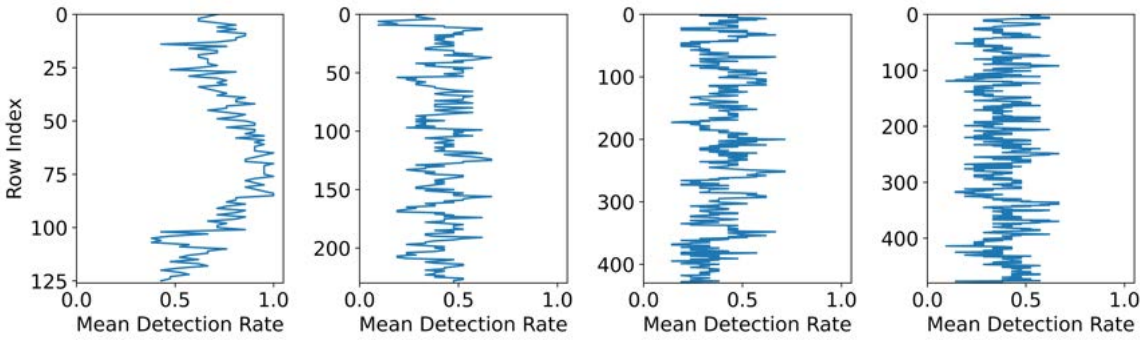
only a very small number of columns consistently respond to every stimulus.

Comparing detection rate to the throughput shown in Figure 15(c) and (d), it appears that detection rate begins to drop once the instantaneous throughput meets or exceeds the manufacturer's specified limits. This suggests that the maximum specified throughput limit describes a limit of reliability, rather than pure functionality. Careful inspection of the time surfaces in Figures 12 and 13 shows that while a majority of a skipped column may not get read out, several pixels within that column may, so the phenomenology is not as simple as whole columns or rows being skipped entirely.

The pattern in column dropping in the XPlorer is similar to the time surface patterns of increasing timestamp from right to left. This is further evidence to suggest the camera reads out events from right to left, until the throughput limit is exceeded. Once the limit has been exceeded, sections of columns appear to be skipped almost at random.

The Prophesee camera behaves differently under high throughput than the XPlorer. In Figure 20(a), the 4.1% exposure shows a slight tendency for rows on the upper and lower edges to detect fewer stimuli. Figures 20(b) through (d) show a variable but generally consistent detection rate across all rows. This suggests that the Prophesee camera does not have a preference toward which rows to drop when the throughput limit is exceeded. It was observed that groups of adjacent rows got read or skipped together; only rarely was a single row treated differently than its neighbors.

The arbiter circuit in each camera is likely responsible for determining which columns or rows get skipped at high PSPs. In the time surface in Figure 13(e), there are several dozen pixels scattered throughout whose events are much later than their neighbors. In the bottom half of the figure between rows 100 and 120, where the rows got skipped, those individual pixels still respond later than one would expect the rest of the row to respond, had it been recorded. Assuming the Prophesee follows the



(a) 4.1% PSP (b) 14% PSP (c) 47% PSP (d) 100% PSP

Figure 20. Prophesee Row Detection Rates. Because the Prophesee read out simultaneous rows instead of columns like the XPlorer, row detection rates are considered instead, thus causing the plot format to be rotated 90°. In contrast to the XPlorer, when exposed to the 771cd/m^2 stimulus, the Prophesee did not show a tendency for some rows to become less responsive. Rather, it appears that all rows lost responsiveness with some random variation.

same 4-phase request-acknowledge handshake described earlier, it's possible that the row request went unacknowledged, and yet all pixels in that row reset themselves. Approximately 3000μs later, a few late-coming pixels triggered their first events and sent a row request, which was acknowledged as the camera "caught up".

The drop in detection rate at high PSPs has a significant effect on the quality and completeness of an image recorded by these cameras. Current models describing sensor operation, such as that described in [23], do not include this behavior, as the models are primarily focused on individual pixel behavior, and the drop in detection rate is likely an arbiter-based phenomenon. It will be important to accurately describe and account for this behavior in future models. Real world activity which is likely to produce this response includes any scene which changes illumination rapidly. This could include turning lights on or off in an enclosed room, a camera flash illuminating a space, a streetlight turning on at dusk, lightning illuminating an outdoor scene at night, or an explosion or electric arc illuminating a space. Given the wide possibilities for real world activity to cause a reduction in detection rate, it will be beneficial to the industry to understand and model the behavior of the arbiter under a high load.

Specifically, determining the process by which some pixels do not get sampled could be influential the development of future sensors designed for specific tasks where this phenomena could detract from system performance.

5.6 Limitations

Several methodological limitations restrict the scope and accuracy of the work presented. Experimentally, the pixel selection process was suboptimal. Restricting analysis to exclusively ON events potentially hid useful information that could be gained from including OFF events. Finally, the step-function stimulus profile only enabled characterization of these nonlinear cameras in a limited range of their operational capability.

In order to control the number of pixels being stimulated, an iris was opened and closed, thereby creating an image on the sensor. While focus and stray light were well-controlled, diffracted light stimulated pixels beyond the edges of the ideal image. Uncertainty bars presented in Figures 16 and 17 describe this variability. One possible technique for controlling PSP more precisely could be to directly illuminate the focal plane array with a low-power laser, whose beam diameter is carefully determined.

Additionally, stimulated pixels generated both ON and OFF events. This research did not analyze the nature or behavior of OFF event generation, nor were OFF events included in the statistics. The bias settings play a significant role in determining pixel response and tendency toward generating ON or OFF events. Based on experience, it appears that positive and negative stimuli produce similar results, though this was not been systematically evaluated. In a linear system, a single impulse response can be used to describe the system response under all conditions. Event cameras are fundamentally nonlinear and therefore cannot be completely described by just one impulse response. This research only considered one form of stimulus: from

near-zero illuminance to some higher illuminance. Since event cameras respond to the logarithm of the change in illuminance, to fully characterize each camera would require applying stimuli which start at progressively higher levels of illuminance. Therefore, this research cannot be used to quantitatively predict the response to a stimulus beginning at a nonzero luminance, though it is likely that overall trends in behavior would persist. Despite these limitations, several broad conclusions can be made with high confidence.

5.7 Summary

The XPlorer and Prophesee demonstrated a lower MAD and thus higher timing precision compared to the DAVIS240 and DAVIS346, in all configurations. An advantage enjoyed by the DAVIS240 and DAVIS346 over the XPlorer and Prophesee is that they did not show spatial patterns associated with column-wise readout at low PSPs. However, their timing precision at those exposures was significantly worse.

One key advantage demonstrated by the DAVIS240 and DAVIS346 which was not matched by either the XPlorer or the Prophesee was their event reporting reliability at high PSPs. Both the XPlorer and Prophesee began to drop events at high exposure fractions, while the DAVIS240 and DAVIS346 continued to generate events from 100% of pixels that were stimulated. This presents a tradeoff: the increased timing accuracy with the XPlorer and Prophesee comes with the high probability of significant artifacts under high load. If high timing accuracy is a priority, the XPlorer and Prophesee may be most appropriate. However, their poor reliability responding to high load makes the DAVIS346 a better option if full-field reliability is a priority.

In addition to skipping pixels, more familiar rolling-shutter artifacts like distortion are likely to appear if scene activity occurs on the 100-microsecond time scale. Every camera tended to read out rows or columns in order, at high PSPs. If activity occurs

on a similar time-scale to the readout period of the arrays, on the order of 100-500μs, distortion is likely. Manufacturers have noted this and implemented various forms of global shutter controls in the software. Users should consider the expected scene activity and plan for distortion if that activity is on the order of hundreds of microseconds. However, for the DAVIS240 and DAVIS346, this behavior only began to apply at large PSP fractions. So if a scene is largely static with only a small object moving through the field of view, it is unlikely that the camera would switch into the ordered readout regime, so the small object's motion would get sampled at the maximum rate allowed by the camera. However, if the scene is largely dynamic such that a large fraction of the array is attempting to report an event simultaneously, the motion of the object through that scene would begin to be sampled according to the ordered process.

Additionally, event latency will depend on activity orientation with respect to the sensor. For the DVXPlorer and DAVIS346, broad objects producing high throughput with respect to pixel columns will have more latency variability, while narrower objects with respect to pixels columns will have less latency variability. The DAVIS240 and Prophesee would exhibit similar behavior with respect to rows. Arbiter behavior has not yet been given much attention in the neuromorphic community, but arbiters have the potential to significantly impact the quality of data being produced by event cameras. Models for arbiter behavior should be developed and published in order to ensure further adoption of event cameras into widespread use, and to ensure they can withstand the scenarios that test performance limits.

VI. Conclusion

The first two-dimensional images depicting 3-dimensional space and time were created in caves tens of thousands of years ago. Those individuals wishing to document activity in their surroundings progressed to more familiar formats such as painting on canvasses, recording light directly onto film, then eventually digital sensors. While the mechanics of imaging improved with time, the static-frame-based principle did not change in a significant way until the late 1800s, when several still pictures were combined to create the illusion of motion. Still today, visualizing motion relies on displaying a rapidly-updated series of static 2-D images. Event-based cameras offer a new technique for recording and displaying motion throughout a scene, by asynchronously recording only the elements of the scene that change. Departing from the 2-dimensional paradigm of imaging has required significant effort in technological development, but also in interpretation of AER data and application of this new capability.

6.1 Significance of Work

In this thesis, four event cameras from two manufacturers were exposed to a controlled, repeating, step-function light source. The various responses to this simple stimulus provide a significant amount of insight into the characteristics of the cameras, revealing strengths, weaknesses, and novel behaviors not yet described by leading scientific models. This research rigorously subjected four separate cameras to the same test conditions. This head-to-head testing accomplished three main tasks in pursuit of expanding the application of event based imaging.

First, time surfaces compiled from the first event from each exposed pixel revealed clues about the functional nature of each camera's readout arbiter. Spatial patterns

were apparent, showing that fixed pattern noise is not the only spatial consideration applicable to non-uniformity in event camera output, and that in all cases, arbiters prefer one side of each focal plane array when under a sufficiently high instantaneous load. Additionally, rolling shutter artifacts have been predicted by previous work. The time surfaces generated in this thesis showed over what speeds those artifacts may occur, and in which orientation they would occur. These time surfaces also revealed a new finding that has not previously been described in the literature: the detection rate of the DVXPlorer and Prophesee Gen3 drops significantly at high instantaneous load, potentially impacting image quality in certain high-speed applications. Current models focus on pixel-level behavior and do not describe this behavior.

Second, quantifying the duration over which a time surface extended described the deviation in latency, or jitter, of each camera. This effectively describes the timing precision of each camera. It was discovered that increasing the number of pixels subjected to a simultaneous stimulus significantly reduces timing precision. This effect became particularly apparent above roughly 1% PSP. Finally, measuring the number of events produced in response to a single stimulus described the level of redundancy exhibited by each camera. This presents an opportunity for impactful follow-on research, as the event generation rate is directly controlled by user-adjusted bias parameters.

Most naturally lit scenes would rarely change quickly enough for these behaviors to be a concern. However, artificial lighting or various natural phenomena could cause a rapid stimulation of the entire sensor and produce artifacts. Possible examples where an event camera could be vulnerable include dim rooms when lamps or strobe lights such as fire alarms turn on, an outdoor night scene that is illuminated by a bolt of lightning, or a suburban roadway night scene when the streetlights turn off. While these examples are not every-day occurrences, they are far from extraordinary.

For example, if a vehicle was using an event camera for navigation at night and the streetlights fail, causing a momentary degradation in image quality, it is easy to imagine how the situation could quickly become dangerous. Further research should be done to examine the susceptibility of event cameras to artifacts from occurrences such as these.

6.2 Recommendations for Future Work

Event cameras are becoming commercially available on a large scale. This availability is going to rapidly accelerate the growth of applications and in interest in this unique technology. In order to ensure maximum utilization of the novel capabilities provided by these cameras, significant research must still be done to understand their behavior, and to develop methods for using AER data efficiently and effectively.

6.2.1 Illumination Techniques

It was noted during processing that array statistics were sensitive to which pixels were discarded and which were kept. This suggests that the technique used in this thesis is not the most optimized method of describing the nature of these cameras in a realistic and accurate manner. If future characterization efforts require the stimulation of a small fraction of the array, it is recommended that the pixel exposure process be improved. The method presented in this thesis caused some amount of diffracted light to expose pixels beyond the edges of the reported areas, though the amount of exposure was not quantified. Other focal plane array characterization efforts have placed a physical mask on the sensor itself [29], which might work well in this application. In the event that 100% PSP is required, a strategy more consistent with standard industry practices would be to remove the lens and expose the sensor to a radiating disc 8x farther from the sensor than its diameter [30]. In order to

expose a small fraction of the array, a laser and careful beam conditioning could be used to precisely illuminate specific pixels while minimizing stray light incident on neighboring pixels.

6.2.2 Event Redundancy

The characterization of throughput redundancy through measuring the number of events generated per stimulus would be extremely beneficial to the industry. Event generation rates can be controlled through various bias configurations, but there is currently no rigorous discussion describing the detailed effects of the various parameters. This characterization process should be a high priority for near-term work in the neuromorphic imaging community.

While physics-based models exist for a few models of event cameras, significant effort must be put in to fully characterize their response to various stimuli. Several phenomena are reported in this work which are only introduced at the empirical level, let alone understood at a component engineering level. An accurate physics-based electrical circuit model should be developed, or the nature of various artifacts will remain difficult to understand.

6.2.3 Event Suppression

The DVXPlorer demonstrated the surprising tendency to produce fewer events as luminance increased. This could have significant impacts on behavior and image quality under a wide variety of applications. Currently there is no description of the camera that would motivate this behavior. Such a significant tendency ought to be explored before the camera is installed in significant applications.

6.2.4 Data Processing

The nature of AER data makes the data processing pipeline prone to using loops. This was a source of significant computational load, causing processing times to be exceedingly long. It is well established that vectorizing data formats can lead to significantly faster processing depending on the programming language used. Since AER data is initially produced in a list, it is very efficient in terms of bandwidth and storage space requirements, but it is resistant to vector-based processing. A possible solution for this is that once data is no longer size-critical, the events could be sorted into a 3-D array of (x,y,t) for further processing. This would reduce the necessity of looping through events and enable further vectorization and time optimization.

6.2.5 Arbiter Preference

All four cameras demonstrated a tendency to read columns or rows from one side to the other when under sufficient load. Previous publications have alluded to arbiter circuits fairly choosing the order in which to respond to requests, but there has not yet been a thorough description of the logic implemented, or at what point that logic is overwhelmed. It would be quite useful if the manufacturers could release more information regarding the nature of the arbiter circuits, enabling end users to predict and plan for the mode of failure of fair arbitration. In the absence of this data from the manufacturer, further characterization of these cameras under high load may be able to generate an empirical model which satisfies the same requirement.

6.3 Summary

This thesis considers three metrics of performance to create one reference which describes the span of behavior of event cameras. Four cameras were included in the study, and the procedure could easily be applied to other cameras in the future.

This head to head comparison enables end users to make informed decisions when choosing a sensor for their potential applications, based on observed performance in response to well-defined stimuli. It also enables researchers to target their efforts in prioritizing future work, such that the most valuable characteristics and applications of these novel cameras can be exploited.

Bibliography

1. Military photography. Encyclopedia.com. [Online]. Available: https://www.encyclopedia.com/history/united-states-and-canada/us-history/military-photography [Accessed: 2021-01-05]

2. D. Gettinger. The ultimate way of seeing: Aerial photography in wwi. The Center for the Study of the Drone at Bard College. [Online]. Available: https://dronecenter.bard.edu/wwi-photography/ [Accessed: 2021-01-08]

3. G. Gallego, T. Delbruck, G. M. Orchard, C. Bartolozzi, B. Taba, A. Censi, S. Leutenegger, A. Davison, J. Conradt, K. Daniilidis, and D. Scaramuzza, "Event-based vision: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1, 2020.

4. Dvxplorer. Consumer Technology Association. [Online]. Available: https://www.ces.tech/Articles/2020/DVxplorer.aspx [Accessed: 2021-01-05]

5. inivationxplorer. iniVation. [Online]. Available: https://inivation.com/inivation-wins-best-of-innovation-award-at-ces-2020/ [Accessed: 2021-01-05]

6. Smartthings vision specifications. Samsung. [Online]. Available: https://www.samsung.com/au/smartthings/camera/smartthings-vision-gp-u999gteeaac/ [Accessed: 2021-01-05]

7. Y. Suh, S. Choi, M. Ito, J. Kim, Y. Lee, J. Seo, H. Jung, D. H. Yeo, S. Namgung, J. Bong, S. Yoo, S. H. Shin, D. Kwon, P. Kang, S. Kim, H. Na, K. Hwang, C. Shin, J. S. Kim, P. K. J. Park, J. Kim, H. Ryu, and Y. Park, "A 1280×960 dynamic vision sensor with a 4.95-m pixel pitch and motion artifact minimization," in *2020 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2020, pp. 1–5.

8. T. Finateu, A. Niwa, D. Matolin, K. Tsuchimoto, A. Mascheroni, E. Reynaud, P. Mostafalu, F. Brady, L. Chotard, F. LeGoff, H. Takahashi, H. Wakabayashi, Y. Oike, and C. Posch, "5.10 a 1280×720 back-illuminated stacked temporal contrast event-based vision sensor with 4.86µm pixels, 1.066geps readout, programmable event-rate controller and compressive data-formatting pipeline," in *2020 IEEE International Solid- State Circuits Conference - (ISSCC)*, 2020, pp. 112–114.

9. Insightness news. Insightness. [Online]. Available: https://www.insightness.com/?p=484 [Accessed: 2021-01-05]

10. M. Mahowald, "Vlsi analogs of neuronal visual processing: A synthesis of form and function," Ph.D. dissertation, California Institute of Technology, 1992.

11. D. Jourbert, M. Hébert, H. Konik, and C. Lavergne, "Characterization setup for event-based imagers applied to modulated light signal detection," *Applied Optics*, vol. 58, p. 1305, 02 2019.

12. B. McReynolds, "A comprehensive test methodology and physics-based camera model for characterizing neuromorphic imagers," Master's thesis, Air Force Institute of Technology, 2019.

13. R. Paschotta. Radiance. RP Photonics. [Online]. Available: https://www.rp-photonics.com/radiance.html [Accessed: 2020-12-31]

14. Radiometric vs. photometric units. Thorlabs. [Online]. Available: https://www.thorlabs.com/images/Catalog/V19_05_LED.pdf [Accessed: 2020-12-31]

15. What is a ccd. Spectral Instruments. [Online]. Available: http://www.specinst.com/What_Is_A_CCD.html [Accessed: 2021-01-08]

16. P. Lichtsteiner, C. Posch, and T. Delbruck, "A 128× 128 120 db 15 μs latency asynchronous temporal contrast vision sensor," *IEEE Journal of Solid-State Circuits*, vol. 43, no. 2, pp. 566–576, 2008.

17. C. Brandli, R. Berner, M. Yang, S. Liu, and T. Delbruck, "A 240 × 180 130 db 3 µs latency global shutter spatiotemporal vision sensor," *IEEE Journal of Solid-State Circuits*, vol. 49, no. 10, pp. 2333–2341, 2014.

18. H. E. Ryu. Industrial dvs design: Key features and applications. Youtube. [Online]. Available: https://www.youtube.com/watch?v=7fAPckjQSGE [Accessed: 2020-12-24]

19. Y. Nozaki and T. Delbruck, "Temperature and parasitic photocurrent effects in dynamic vision sensors," *IEEE Transactions on Electron Devices*, vol. 64, no. 8, pp. 3239–3245, 2017.

20. Understanding the performance of neuromorphic event-based vision sensors. iniVation. [Online]. Available: https://inivation.com/white-papers/ [Accessed: 2021-01-03]

21. X. Lagorce, G. Orchard, F. Galluppi, B. E. Shi, and R. B. Benosman, "Hots: A hierarchy of event-based time-surfaces for pattern recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 7, pp. 1346–1359, 2017.

22. D. Hollidt, "Evaluation of camera resolution in optical flow estimation using event-based cameras," Ph.D. dissertation, KTH Royal Institute of Technology, 07 2020.

23. T. Delbruck, Y. Hu, and Z. He, "V2e: From video frames to realistic dvs event camera streams," 2020.

24. Labsphere, *Integrating Sphere Theory and Applications*, 2017. [Online]. Available: https://www.labsphere.com/site/assets/files/2551/integrating_sphere_theory_apps_tech_guide.pdf [Accessed: 2020-12-7]

25. *DET10A(/M) Si Biased Detector User Guide*, Thorlabs. [Online]. Available: https://www.thorlabs.com/drawings/b5a7363d0ebd0bd1-209F9EEF-0F18-8DAF-7E2300DA1C8DBBE4/DET10A_M-Manual.pdf [Accessed: 2021-2-28]

26. "Jcgm 100: Evaluation of measurement data - guide to the expression of uncertainty in measurement," Joint Committee for Guides in Metrology, Tech. Rep., 2008.

27. V. Lindberg. Uncertainties and error propagation. Louisiana State University. [Online]. Available: https://www.geol.lsu.edu/jlorenzo/geophysics/uncertainties/Uncertaintiespart2.html [Accessed: 2021-01-08]

28. D. A. Kenny, *Correlation and Causality*, 1 1979, vol. 1.

29. B. Wysocki and M. Marciniak, "Discrimination between electronic and optical blooming in an insb focal-plane array under high-intensity excitation," *Infrared Physics & Technology*, vol. 51, no. 3, pp. 137 – 145, 2008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1350449507000886 [Accessed: 2021-1-11]

30. *Standard for Characterization of Image Sensors and Cameras*, European Machine Vision Association, November 2010. [Online]. Available: https://www.emva.org/wp-content/uploads/EMVA1288-3.0.pdf [Accessed: 2020-1-11]

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704–0188

1. REPORT DATE *(DD–MM–YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From — To)*
08–03–2021	Master's Thesis	Sept 2019 — Mar 2021

4. TITLE AND SUBTITLE

A Comparative Evaluation of the Fast Optical Pulse Reponse of Event-Based Cameras

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Brewer, Tyler J., Capt, USAF

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/EN)
2950 Hobson Way
WPAFB OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER

AFIT-ENP-MS-21-M-103

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Air Force Research Laboratory
Space Vehicles Directorate
3550 Aberdeen Dr SE
Kirtland AFB, NM 87117

10. SPONSOR/MONITOR'S ACRONYM(S)

AFRL/RVSW

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

DISTRIBUTION STATEMENT A:
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Event cameras use biologically inspired readout circuit architecture to offer a faster and more efficient method of imaging than traditional frame-based detectors. The asynchronous event reporting circuit timestamps events to 1 microsecond resolution, but latency increases when many pixels are stimulated simultaneously. To characterize this variability, the DAVIS240, DAVIS346, DVXPlorer, and Prophesee Gen3M VGA-CD 1.1 cameras were exposed to single step-function flashes with amplitudes from 9.3-771cd/m^2, stimulating from 0.0042-100% of pixels. The Median Absolute Deviation of pixel response times ranged between 0 and 6086μs, increasing with the percent of pixels stimulated (PSP). The number of events generated per pixel generally decreased with increasing PSP, with all cameras producing fewer than 59 events per pixel. Surprisingly, as stimulus amplitude increased, the DVXPlorer generated fewer events, to as low as 0.32 events per stimulus. Short-term throughput exceeded advertised limits in 3 of 4 cameras. While individual pixels may be able to accurately detect microsecond-scale change, data bottlenecks can cause missed events or erroneous timestamps.

15. SUBJECT TERMS

Event Based Camera, Neuromorphic Camera, Asynchronous Vision Sensor, Address Event Representation (AER), Machine Vision, High Speed Imaging, Robot Vision System

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Dr. Michael Hawks, AFIT/ENP
U	U	U	U	102	19b. TELEPHONE NUMBER *(include area code)* (937) 255-3636 x4828