2017 23rd IEEE International Symposium on Asynchronous Circuits and Systems

A Frameless Imaging Sensor with Asynchronous Pixels: An Architectural Evaluation

Montek Singh, Pintian Zhang, Andrew Vitkus, Ketan Mayer-Patel and Leandra Vicci
Dept. of Computer Science, Univ. of North Carolina at Chapel Hill, NC 27599, USA
{montek, pintianz, vitkus, kmp, vicci}@cs.unc.edu

***Abstract*—The goal of this work is to develop a novel CMOS camera sensor that provides frameless capture, and has significantly higher dynamic range, finer color sensitivity, and lower noise as compared to the current state-of-the-art sensors. The strength of the approach lies not in developing new types of photodetectors or amplifiers, but in the manner in which information is extracted from the pixel sensor, transported to the processing logic, and processed to yield intensity values. At the heart of the sensor is an asynchronous network to transport events from the pixel sensors to the off-grid processing circuitry. The asynchronous nature of pixel communication is the key to achieving frameless image capture.**

Keywords—image sensing; HDR; frameless; asynchronous

I. Introduction

Imagine a camera that allows you to take a picture of a landscape in which the whitest clouds can be discerned against a bright sky while, at the same time, the bark and foliage on the trees can be clearly seen in the darkest shadows. Imagine being able to take a picture in a dark room without a flash yet avoiding noisy grains in the dark regions. Imagine taking a fast action shot of an animal in the wild without losing detail in the shadows. Imagine taking a studio portrait with gradations in color a thousand times finer than possible today. And now imagine doing all of this at a thousand captures per second while recording video. Even better, imagine doing so unencumbered by fixed frame rates: reading any regions or even pixels of interest as frequently or infrequently as needed. All of these scenarios could be a reality with a new high-quality image sensor that is the subject of this work.

Today's camera sensors are stuck with the notion of timed frames, an artifact from the days of the cathode ray tube (CRT) televisions, which imposes an unnecessary synchronizing constraint on all pixels (i.e., all pixels must start and stop their measurement at the same time). While fixed time frames do offer some simplicity—simpler sensor design, simpler encoding, and simpler transmission—the rigidity of frames imposes significant costs in terms of dynamic range, precision and sensitivity, and we claim that with advances in hardware and software the time is ripe for a paradigm shift.

A. Overview of Our Contribution

We introduce a new sensing approach that turns the measurement paradigm around. Existing sensors measure, in effect, the number of photons received at each pixel within a globally fixed frame time. Our approach is to instead measure the time each pixel takes to accumulate a certain fixed number of photons; once the accumulation has been achieved, the pixel emits a "tick" and starts over. Thus, our approach, in effect, uses intensity-to-frequency conversion, but actually measures the reciprocal of the light intensity (by measuring period instead of frequency).

The significance of this subtle change—measuring "time-to-fill" for each pixel instead of counting photons over a time interval—is enormous:

- *Frameless capture* since there is no need to count photons in a fixed time for all pixels. Each pixel's time-to-fill is independently measured.
- *Higher dynamic range* since the darkest pixels could, in effect, have a longer exposure time (say, 10^{-2} second) than the brightest pixels (say, 10^{-8} second).
- *Greater precision* since time is easily measured to a much finer granularity—e.g., to within 1 nanosecond within a typical 1 millisecond measurement interval (i.e., 20 bits precision)—than counting the number of events (typically one in thousands, or 10-12 bits precision).
- *Better low-light sensitivity:* Measuring time has the tremendous benefit of much greater sensitivity (i.e. smoother perceived gradations, or smaller $\Delta I/I$, where I is the intensity) in the darker regions because the time interval between ticks is actually longer (a benefit of reciprocal measurement).
- *Multiple regions of interest* (ROIs) can be read at arbitrary time granularity because each pixel can be read independently, enabling more judicious use of precious bandwidth and energy.

B. Rationale

With the small transistor size afforded by today's CMOS technologies, combined with backside illumination, which allows nearly the entire pixel area to be covered with computational circuitry, a tremendous opportunity has now opened up to make pixels "smarter". In particular, for a 50 square-micron pixel of a digital SLR camera, between 500 and 1000 transistors can be accommodated, assuming a 45-nm process [2] and backside illumination, which is sufficient to carry out our objectives. (More advanced technologies such as 22 nm offer even greater potential. And even without backside illumination, about half as many transistors can be still accommodated, adequate for our objectives.)

While the state-of-the-art is hobbled by limited dynamic range, coarse intensity gradations, significant low-light noise, and a rigid frame-oriented capture regime, we claim that by endowing pixels with a modest amount of smart circuitry, time is ripe for a new paradigm.

C. Paper Organization

This paper begins with a discussion in Section II of the design goals that are key to high-quality sensing. Section III reviews related work. Section IV reviews traditional CMOS

978-1-5386-2749-5/17 $31.00 © 2017 IEEE
DOI 10.1109/ASYNC.2017.19

IEEE computer society

Authorized licensed use limited to: University of North Carolina at Chapel Hill. Downloaded on November 20,2026 at 04:43:40 UTC from IEEE Xplore. Restrictions apply.



(a) (b) (c)

Fig. 1: (a) Poor shadow detail; (b) Blown highlights; (c) High-dynamic-range image (contrast reduced to fit print range). Picture credit [1].

sensor architectures. Section V presents our proposed sensor, including its architecture and operation. Section VI presents experimental results, and Section VII gives conclusions.

II. Background: Design Goals

Several metrics are key to determining sensor quality.

Dynamic Range: This metric is the ratio between the highest brightness a sensor can record (before reaching saturation) and the lowest brightness it can discern (before noise overwhelms the signal). A greater dynamic range is required in higher-contrast scenarios, i.e., where the variation in lighting is high. Example scenarios include sunny landscapes; backlit silhouettes; artificially lit emergency responder situations; etc. A sensor with inadequate dynamic range (e.g., 10-14 bits) will either result in blown highlights or lost shadow detail. Fig. 1 shows such an example. A high-dynamic-range (HDR) sensor with adequate range (e.g., 16-20 bits) could capture all the intensities in the scene. Fig. 1c shows such an HDR image, though with contrast reduction—via intensity scale compression—to fit within the range allowed by today's software and displays.

Limitations of State-of-the-Art: Fig. 2 shows that dynamic range is typically 10-14 bits for most consumer-level cameras costing under $1500. Even the most expensive professional cameras (costing upwards of $10K) have sensors with at most 14 bits of dynamic range. Thus, today's sensors can record scenes with a contrast ratio of at most 2^{10}:1 to 2^{14}:1. By comparison, the human eye is capable of 20 bits dynamic range (with time for full adaptation), or a contrast ratio of 10^6:1.

Intensity Gradation (Step Size): This metric characterizes the ability to distinguish between fine gradations of intensity. How fine a gradation can be captured is determined by the quantization step of the analog-to-digital converter (ADC), for a given ISO setting (i.e., pixel's internal amplifier's gain). A larger step size means the closest distinct values produced by the sensor are relatively far apart (i..e, large quantization error), resulting in unseemly banding or posterization [3]. A picture of a smooth ball lit from a distant source would show circular bands (Fig. 3a). A finer-gradation sensor would capture much smoother variations (Fig. 3b).

Limitations of State-of-the-Art: Current camera sensors perform rather poorly in terms of gradation in dark regions of images. This is because a typical sensor is digitized to a linear scale of gradations (i.e., nearby intensity values are approximately a fixed distance apart, or constant ΔI), whereas the human eye has a logarithmic response and is therefore much more sensitive to intensity changes at the lower end of the scale (cf. Weber-Fechner law [4]). Thus, the difference in the sensation in the eye-brain system (the *perceived* intensity difference) for two nearby intensity levels is proportional to $\Delta I / I$. For example, from intensity value 1 to value 2, there is a 100% change, and from 2 to 3, a 50% change, etc. An ideal sensor would need to have darker intensity values packed much more densely (smaller ΔI at lower I values) to achieve smoother gradations and avoid banding.

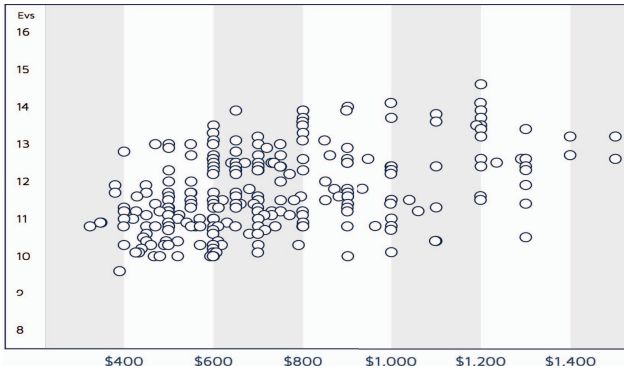


Fig. 2: Dynamic range of sensors under $1500. (Often reported in *Ev* units, where each *Ev* equals a doubling of intensity, or 1 bit.) [20]



(a) (b)

Fig. 3: (a) showing banding due to coarse perceived intensity gradation; (b) no banding because of finer digitization step size.

Noise: The signal-to-noise ratio (SNR) represents the noisiness in image capture. Noise becomes especially important in regions that receive less light, ultimately limiting how low a light intensity can be captured. This is critical in both low-lit scenarios (e.g., indoor flashless photography) and in fast action scenarios. There are four main source of noise [5, 6]: (1) *photon shot noise* due to the randomness inherent in the arrival times of photons; (2) *dark current noise* due to the presence of a non-zero current in the photo-sensor even when no light is received; (3) *sensor read noise* due the noise in the amplifiers and pixel read-out circuitry; and (4) *ADC noise* due to the noise inherent in the analog-to-digital converter. Although photon shot noise cannot be avoided, this noise is

Authorized licensed use limited to: University of North Carolina at Chapel Hill. Downloaded on November 20,2026 at 04:43:40 UTC from IEEE Xplore. Restrictions apply.

proportional to the square root $\sqrt{N}$ of the number N of photons collected and as such, SNR improves as more photons are collected. Consequently, maximizing the amount of light reaching the sensor and collected is paramount. The dark current noise in today's sensors is insignificant and is easily eliminated via calibration. The third source of noise, sensor read noise, is also quite low: about 1-2 photoelectrons in mid to upper-end cameras. The fourth noise source, ADC, however, exposes the drawbacks of the state-of-the-art.

Limitations of State-of-the-Art: The ADC noise, which is dominated by the quantization error, is a significant limitation in current sensors, because it introduces far greater measurement errors than all the other three noise sources just described, especially in low-light situations. In particular, in most top-end sensors (e.g., Canon 1DX, Nikon D3), each pixel can accumulate about 80,000 electrons before saturating (*"full well capacity"*). A 12-bit ADC is typical, which means the lowest intensity value recordable is 80,000 / 2^{12}, or about 20 electrons [5, 6]. Hence for the darkest parts of the image, where a pixel might receive, say, 4 electrons, the photon shot noise is 2 electrons, dark noise is negligible, and read noise is 1-2 electrons, which amounts to a total noise less than 3 electrons (noise sources combine as the square-root of sum of squares) prior to ADC. But the ADC's quantization step of 20 electrons overwhelms all other noise. While industry has tried increasing ADC bit-depth to 14 bits, that is quite difficult to achieve while retaining high-speed low-power operation. We propose a better approach without these limitations.

Circuit Size / Fill factor: This metric is the fraction of the pixel's total area that is devoted to photon collection. The remainder of the area has computational circuitry. Conventional CMOS sensors are of the type called *front illuminated.* Light arrives on the front side of the silicon wafer, the same side that computational circuitry (i.e., transistors and metal wires) are present. The presence of circuitry obstructs the path of light (Fig. 4, top). The fraction of the silicon area that is not obscured by the circuitry and can therefore gather photons is called the *fill factor.* Ideally, the fill factor should be close to 1, but typically it ranges from 30% to 70%. Lower fill factors mean fewer photons accumulated, which results in a loss of the darker pixel values, and greater overall noise.

Backside Illumination is a new technique that has been introduced (e.g., Sony Exmor, and iPhone 4) to overcome lower fill factors by using the entire back side of the wafer (which is not covered by computational circuitry) for illumination (Fig. 4, bottom). There are technological and fabrication challenges with such sensors, so they are currently expensive, but they are poised to become commonplace because of their better photon gathering capability. With such a sensor, substantially more computational circuitry can be put in a pixel, even covering the entire pixel area, yet offer the image quality of a sensor with 100% fill factor.

III. Related Work

CMOS Sensors. While there are a large variety of sensor designs, they all have significant limitations. In the vast majority of CMOS sensors on the market today, the analog value from each pixel is transferred to the column (one at a time), and converted to digital at the end of the column. This approach has the disadvantage of analog degradation. In some

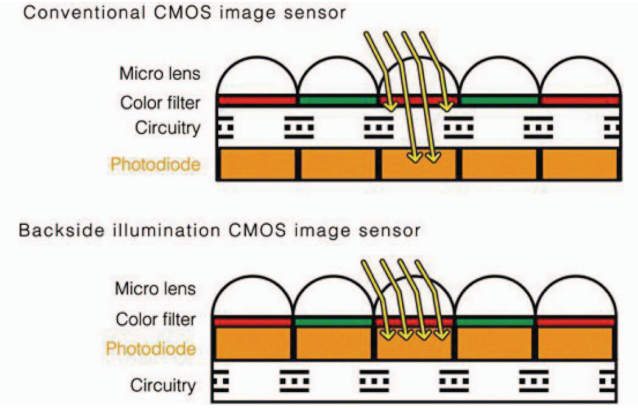


Fig. 4: Benefit of backside illumination: circuitry can occupy more area without obstructing light to the photodiode [7].

experimental designs, analog-to-digital conversion (ADC) is performed within the pixel [8]. While doing so alleviates the analog degradation, achieving high-precision in-pixel ADC is quite difficult: if the maximum analog value is 1 V, then a 14-bit ADC must be able to resolve down to 2^{-14} V, or 0.06 mV, which would require much area and long conversion times. Hence, in-pixel ADC in sensors is rare. Most conventional sensing approaches suffer from not only limited precision but also limited dynamic range because the analog value saturates once the photocurrent has fully drained the capacitor [5].

To overcome the limitation on dynamic range due to saturation, intensity-to-frequency conversion has been introduced [9–15]. The key idea is to recharge the capacitor once it has been fully drained by photoelectrons, record this event by emitting a "tick", and start over. Thus, higher light intensities result in more frequent events, and a pixel's light value is calculated by counting the number of ticks emitted per frame time. (One can regard such an approach as performing 1-bit delta-sigma ADC.) While this approach improves the dynamic range and avoids costly ADC, it has the drawback of coarse-grain readout: no partial ticks are generated, resulting in large errors in the dark regions.

Another limitation of existing sensors is that they typically perform rather poorly in terms of low-light sensitivity (i.e., large $\Delta I/I$, where I is the intensity), resulting in coarse gradations, especially in darker regions of images. This results in unseemly banding and noise grains in low-light regions.

Finally, today's sensors are stuck with the notion of timed frames, an artifact from the days of the cathode ray tube (CRT) televisions, which imposes an unnecessary synchronizing constraint on all pixels (i.e., all pixels must start and stop their measurement at the same time).

Neuromorphic Computing. This is a related field in which researchers are trying to mimic the activity of neurons in the human brain [16-17]. The goal is to emulate neurons directly in silicon, instead of merely simulating their activity in software. Within this field, a number of efforts have focused especially on creating silicon retinas [18]. There are similarities with the type of "spiking" pixels employed in our design: a neuron, similarly, emits spikes when excited by stimulus, and some approaches use an asynchronous network to transport these spikes. However, the key design challenge in neuromorphic computing is a little different from ours: the

Authorized licensed use limited to: University of North Carolina at Chapel Hill. Downloaded on November 20,2026 at 04:43:40 UTC from IEEE Xplore. Restrictions apply.

main focus is not on designing extremely accurate neurons, but instead on implementing the massive-scale connectivity between neurons. In camera sensors though, there are no such connectivity requirements—although pixels in a column share readout circuitry, there is no inter-pixel communication—instead, the key challenge is one of designing pixels with high-quality light capture and precise measurement. Our approach introduces several novel ideas, including in-pixel frequency decimation for greater dynamic range, configurable regions-of-interest, and direct floating-point capture.

IV. Background: Traditional Sensor Architecture

A typical sensor consists of an array of sensing elements ("sensels" or "pixels") and processing logic off to the side (Fig. 5). The sensor array is typically organized as multiple columns, with each column having multiple pixels. All pixels within a column share the communication or readout circuitry that connects the column to the processing logic.

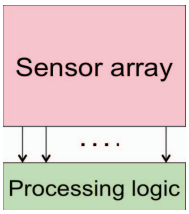


Fig. 5: A camera sensor

Fig. 6 shows the main circuit components inside a pixel. The photodetector generates current proportional to the intensity of light incident on it. For good quality sensing, the photodetector must be physically sized as large as possible to collect the most number of photons. The integrator is a capacitor that integrates the current from the photodetector—typically by discharging from a preset voltage—during the measuring interval. The in-pixel amplifier (if present) and column readout varies per implementation—depending on rolling vs. global shutter, capability to subtract zero errors, etc.—but typically consist of a small number of transistors (often as few as 4 to 6).

The pixel's value is transferred to the readout bus that is shared by all the pixels in a column, and all pixels in a row typically share the readout enable, the overall topology reminiscent of a memory organization. The rows are read out sequentially, and typically result in a rolling shutter artifact since each pixel row is effectively imaged at a slightly different time instant. But there are global shutter designs where all pixel values are sampled at the same instant and buffered internally. In some experimental work, per-pixel ADC has been proposed [9, 13, 15], which has the advantage of avoiding transmission of analog values on a bus, thereby eliminating analog degradation. However, the basic problems of ADC remain: large quantization noise, and an inability to convert to more than 12 bits because of speed and power limitations.

V. Our Sensor Architecture

We build upon the idea of intensity-to-frequency conversion that has been used before [9–15]: the pixel's output is a train of events whose frequency is proportional to the incident light intensity. This is accomplished by using a very low-cost "1-bit" ADC (essentially a comparator) and a reset circuit to restart the integrator. When the integrator reaches a threshold, an output event (or "spike") is generated, and the integration starts over. In prior spike-based approaches, the intensity value is determined by counting the number of events generated at a pixel during a globally fixed exposure interval.

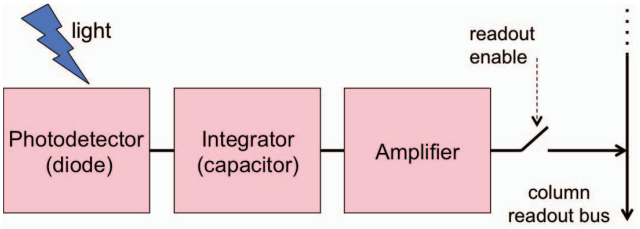


Fig. 6: Inside a pixel: photodiode, integrator and readout circuitry

Where our approach differs from prior ones is that the intensity value is now calculated by *measuring the time between events* in the train, instead of by counting the number of events during some fixed time interval. Our approach, in effect, measures the *reciprocal* of the incident light intensity.

The significance of this change is enormous:

- Time is easily measured to a much finer granularity—e.g., to within 1 nanosecond within a 1 millisecond interval, or 1 in 10^6— than counting the number of events (typically one in thousands).
- Measuring time has the tremendous benefit of delivering much finer gradations of intensity in the darker regions because the interval between events is actually *longer*.
- There is no need for a globally fixed exposure interval for all the pixels; each pixel can capture asynchronously, yielding frameless capture.

The key challenge is to accomplish this goal in a manner that is robust, low-noise and area-efficient.

A. Pixel Architecture

The pixel circuitry is shown in Fig. 7. When the integrator output reaches a threshold, the comparator triggers the reset/restart circuit, which re-initializes the integrator which starts over. At the same time, an event is generated. The train of events goes through a "prescaler" or "decimator" that divides the frequency of the event train by an appropriate factor (typically a power of two) before passing through to the column readout. The role of the decimator is to reduce the output frequency from the pixel so as not to overwhelm the traffic along a column. Decimation is key to obtaining a high dynamic range.

B. Communication Architecture

The architecture of each column is shown in Fig. 8. Instead of a bus, each column is a fully-registered pipeline (FIFO with merge nodes). The events in this pipeline travel downwards, where they are finally received and processed by the processing logic. Each pixel inserts its events into the stream through its associated merge node. The merge node receives events from its neighboring node above it, and from the pixel to its left. If the merge node receives events from both sides close in time, it performs arbitration to decide which one to insert first. The actual order of their insertion, however, is immaterial to the correct operation of the sensor.

Data Encoding. Each event carries two pieces of information: (1) *Row:* an integer specifying the pixel that generated this event, i.e., its row number; and (2) *Decimation Factor:* the prescaling factor being used by this pixel. The

Authorized licensed use limited to: University of North Carolina at Chapel Hill. Downloaded on November 20,2026 at 04:43:40 UTC from IEEE Xplore. Restrictions apply.

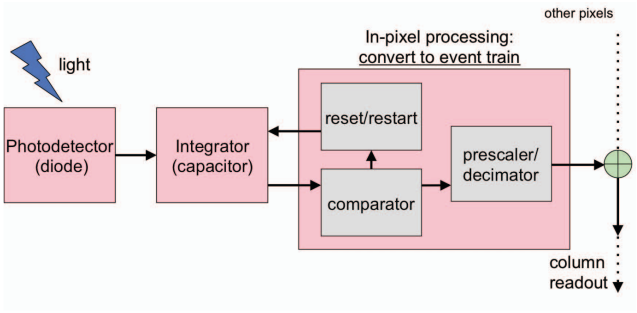


Fig. 7: Proposed pixel architecture

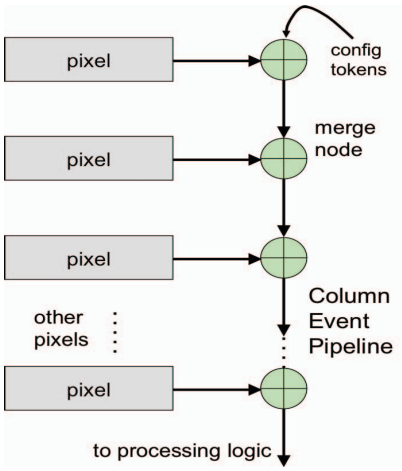


Fig. 8: Column architecture of proposed sensor

decimation factor will be conveniently set to be a power of two (2^D), but only D needs to be sent. This information is needed to properly re-scale the computation of the light intensity value in the processing logic. Thus: *Event* := <*Row*, *D*>.

Time Stamping. A key aspect of the design is that *events do not carry timestamps.* Time stamping is performed in the processing logic instead, at the times when the events are received from the bottom of the column pipeline. This feature was deliberately chosen so as to keep the in-pixel circuitry small, and also to avoid the wiring needed for transmitting timing information throughout the sensor array.

Transport Jitter. It is crucial that the column event pipeline have *low transport jitter* because events are time-stamped after they have been transported to the processing logic. It is not crucial, however, for the pipeline to have a low end-to-end propagation delay. In particular, the intensity value at each pixel is computed by measuring the time difference between events emanating from that pixel. Therefore, the propagation delays, however long they may be, will cancel each other as long as they are equal. However, variability in the propagation delay of events from a particular pixel to the processing logic will directly affect the time difference observed, and will therefore add noise to the computed value. *Therefore, while the magnitude of the propagation delay is immaterial, achieving low transport jitter is key to high-quality sensing in this approach.*

The dominant cause of transport jitter is congestion in the column pipeline. If there is high amount of traffic going vertically through a merge node, then new events generated by the associated pixel are more likely to "collide" with the events already propagating through the pipeline. Each merge node has arbitration capability, so the merging will take place safely. However, one of the two competing events will suffer a delay while it waits for the other to go through first. The network traffic steadily increases down the pipeline as more event trains merge into the common transport pipeline, so congestion is worst near the bottom end.

The delays due to network congestion have a random component: two different events from the same pixel may suffer different congestion-related network delays. Therefore, congestion leads to higher transport jitter, and therefore greater imaging noise. In our simulations, transport jitter is acceptably low (SNR of transport network is ~120 dB) when the pipeline is operating at no more than 80% bandwidth utilization.

Our approach achieves low transport jitter through two techniques: (i) using frequency decimation, and (ii) using asynchronous circuits for implementation. Each of these is discussed next.

C. Decimation and Dynamic Range Improvement

By pre-scaling or decimating the frequency of the event train within each pixel, the overall demand on the column pipeline's bandwidth is alleviated. In particular, a simple frequency divider is used, which transmits one event for every 2^D events produced. This *decimation factor* could be any positive integer, but in order to keep the implementation simple and low-area-cost, it is chosen to be a power of two, hence 2^D. In effect, this is a form of *rate control.*

Per-Pixel Configuration. For the best result, the decimation factor should be settable on a per-pixel basis. One approach is to have the processing logic assign the decimation factor for each pixel in the column. The decimation factor could be communicated, e.g., by feeding pairs of <*Row*, *D*> into the top of the column pipeline (see Fig. 8). Thus, the column pipeline now has a dual role: carrying events from pixels to the processing logic, and carrying decimation factors from the processing logic to the pixels, with both roles requiring the same pair of values <*Row*, *D*>, thereby allowing the same column wiring to be reused.

Pixel Self-Calibration. Another approach is to allow each pixel to determine its own decimation factor autonomously. In particular, the processing logic lets each pixel know the measurement interval using a reference signal (which is low-speed, e.g., 1 KHz for a 1 ms measurement interval). This reference interval is not the actual "exposure time"; as discussed later, the sensor actually operates in a frameless manner. Each pixel uses the reference signal to determine whether its currently set decimation factor 2^D should be increased, decreased or remain the same. If the pixel inserted, say, 4 or more events into the pipeline during a reference interval, then the decimation factor is increased. If the pixel inserted fewer than 2 events, then the decimation factor is decreased. The goal is generally to produce 2-3 events during the reference interval, so enough information is generated to calculate the intensity, and no more.

Decimation has two benefits. First, the available pipeline bandwidth is equitably distributed among the pixels in that column: a really bright subset of pixels cannot overwhelm the

Authorized licensed use limited to: University of North Carolina at Chapel Hill. Downloaded on November 20,2026 at 04:43:40 UTC from IEEE Xplore. Restrictions apply.

entire column pipeline. Therefore, the dynamic range is dramatically increased, limited mainly by the maximum speed at which the integrator can be reset and restarted. Initial simulation results show that a dynamic range of 20 bits is readily achieved even at a 1 millisecond reference interval (equivalent to a fast 1000 fps frame rate). Second, successive events emanating from a given pixel are spaced maximally apart within the reference interval. As a result, the accuracy of computation is enhanced because of the longer time base, even for events from the really bright pixels.

Since the decimation factor is a power of two in this approach, only the small number D needs to be stored within the pixel and communicated, instead of 2^D. Thus, in our design, with just 3 bits of storage within each pixel (D ranges from 0 to 7), a decimation factor of 2^0 to 2^7 can be applied.

D. Pipeline Implementation

We consider two types of implementation for the column pipeline: asynchronous and synchronous.

1) Asynchronous

An asynchronous logic style is particularly well-suited for this application for two reasons. First, an asynchronous implementation will consume significantly less energy than a synchronous one. In particular, a clocked implementation will consume energy in every node on every clock cycle whether or not there is an event to transport. In contrast, the asynchronous version consumes energy only where an event is transported. Second, distributing a high-speed clock throughout a multi-megapixel sensor array costs valuable area, consumes energy, and produces noise due to clock spikes. The asynchronous implementation will avoid these issues.

The communication pipeline was modeled structurally at the transistor and gate level in Verilog. The Mousetrap circuit style [19] was used because it can be realized using predominantly standard cells (except arbiters for merging).

2) Synchronous

A synchronous implementation of the pipeline was also undertaken for two reasons. First, a clocked implementation may fit more easily into established design flows. Second, it is used as a baseline against which the benefit of the asynchronous implementation is evaluated.

The synchronous version combined clocking along with validity information (but no backwards synchronization using acknowledges). Thus, the column pipeline is essentially a synchronous shift register, but with the additional capability of allowing new events to merge in from pixels. A simple arbitration scheme was used: a pixel may insert an event only onto a pipeline slot that is empty. Thus, existing traffic is favored, which simplifies the design by avoiding the need to introduce stalling and back pressure. While a more sophisticated implementation with full elasticity is possible, the simpler approach is sufficient to provide a base case for evaluating the asynchronous implementation.

E. Processing Logic

The processing logic for each column performs two key operations: time stamping of events, and storage of most recent events.

Time Stamping: A high-speed counter inside the processing logic timestamps each event as it arrives from the column event pipeline.

Storage: There is a special-purpose memory for storing information from each event as it is received. Each pixel has a unique address in this memory, addressed by the *Row* number of each event. The data is the tuple <*Timestamp*, *D*>. For each pixel, information about at least two events is needed in order to compute the event frequency. These two events could be successive ones inserted by a pixel into the pipeline, but because the event train had been decimated, they are actually apart by 2^D events. The storage unit is organized accordingly, to hold information about the most recent two (or more) events, in a first-in-first-out (FIFO) fashion.

F. Computation of Intensity Value

The value of the incident light intensity sensed by a pixel is computed from the time difference between events received by the processing logic from that pixel. Assume the following:

- t_1 and t_2 are the timestamps of the two events from a pixel
- 2^D is the decimation factor currently used by the pixel

The light intensity, I, at the pixel is then determined by calculating the event frequency $1/(t_2 - t_1)$ and multiplying it by the decimation factor (2^D). If the time separation is denoted by Δt, the formula can be written as:

$$I \propto \frac{2^D}{\Delta t}$$

Direct Floating-Point Representation. The formula above lends itself nicely to a floating-point representation, where: $1/\Delta t$ is the mantissa, and D is the exponent. As a result, computing the final pixel value is fairly trivial. In fact, simply reporting $(\Delta t, -D)$ without performing any computation on it suffices as the reciprocal of the computed intensity in floating-point representation. Or, ignoring the negative sign, each intensity value is simply the tuple $\langle \Delta t,\ D \rangle$.

Impact on Intensity Gradation. The approach for computing intensity outlined here has a dramatic impact on intensity gradation. In particular, if an intensity level measured for a pixel is $I_1 = 2^D/\Delta t$, then the next discrete intensity level that is measurable but distinct is $I_2 = 2^D/(\Delta t - \tau)$, where τ is the clock period at which the timestamp counter is incrementing. That is, the next higher intensity is one where the two events are closer together by the smallest measurable amount, τ. Since the eye-brain system perceives intensity variations proportional to the intensity, the perceived gradation step is $\Delta I/I \cong \tau/\Delta t$.

Assuming a 1 GHz timestamp counter, τ is 1 ns. Let us assume a reference interval of 1 ms (which would correspond to 1000 frames/second in a frame-oriented sensor). The decimation technique discussed above ensures that each pixel generates about two events during this reference interval. Hence, the inter-event time measured, Δt, is on the order of 1 ms. Therefore, the perceived intensity gradation is:

$$\text{Intensity Gradation } \frac{\Delta I}{I} \cong \frac{1\text{ ns}}{1\text{ ms}} = 10^{-6} \cong 2^{-20}$$

Comparison to Existing Approaches. Fig. 9 compares the intensity gradation of our sensor with prior approaches. Since our sensor uses the floating-point representation described above, it has constant gradation (i.e., *relative* step size) throughout the dynamic range. Thus, absolute increments between discernibly distinct intensity values are much smaller near the darker end of the scale; the increments increase in size

Authorized licensed use limited to: University of North Carolina at Chapel Hill. Downloaded on November 20,2026 at 04:43:40 UTC from IEEE Xplore. Restrictions apply.

as intensity increases. This is a very nice feature since the human eye has a similar sensitivity curve [4].

In contrast, existing approaches convert an analog value of accumulated charge to digital using an ADC, with typically fixed step size, resulting in drastically larger relative step sizes at lower intensities. For example, if value 1 is just above the noise floor, then the next digital pixel value will be 2, which represents a $\Delta I/I = 1$, or a 100% change. The next value, 3, represents a 50% change. These relative changes are too large, and therefore the perceived image shows unseemly banding in low-light scenarios (see Fig. 3).

G. Frameless Capture and Regions of Interest

Frameless Image Capture. A key novelty of the proposed sensor is that it demolishes the notion of a frame with start and end times that are synchronized across all pixels. In fact, each pixel is continuously sensing and its value is being continuously recorded by the processing logic. The measurement interval is used only as a reference to set the decimation factor so that the overall light intensity computed will be a running average over a sliding window of time of that duration. For example, if the reference interval is set to 1 ms, then each pixel will configure its decimation factor so as to produce approximately 2 events per millisecond. Any time this pixel's value is read from the processing logic, it will represent the average light intensity over the just-elapsed millisecond. Therefore, different pixels can be read from the processing logic at different times and even different rates, resulting in a truly frameless image capture system.

Multiple Regions of Interest (ROIs). Extending the notion of frameless capture, the proposed approach makes it easier to specify one or more regions of interest where higher quality capture may be desired. Examples include: faces in a family portrait; a traffic sign in a navigation application; an athlete in an action shot; etc. Once these regions have been identified by the camera's microprocessor, the processing logic can set lower decimation factors for the pixels in these regions. Thus, pixels within ROIs are able to update their values at a higher rate than those in the background. This approach effectively allows different "refresh rates" for different parts of the image. Another configuration parameter that can potentially be varied across pixels is the threshold at which the integrator is reset/restarted (ISO setting). The key to enabling multiple configurable parameters and multiple ROIs is our communication architecture, which allows configuration information to be transmitted from the processing logic through the column pipeline, and addressed to particular pixels or groups of pixels (see Fig. 8).

VI. Experimental Results

A Verilog model of the proposed sensor was developed and evaluated. The size of the sensor was 512x512 pixels, but this is easily scalable to higher pixel counts. The in-pixel circuitry consists of mostly analog circuits—the photodetector, integrator, comparator and restart components—except for the decimator; all of these were modeled behaviorally. Based on the internal frequency of events generated relative to the period of a slow timing reference (1 KHz), each pixel self-calibrates its decimation factor from 2^0 to 2^7.

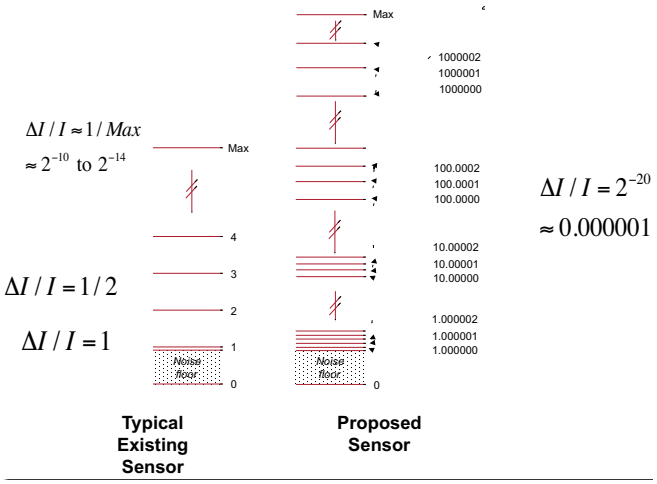


Fig. 9: Perceived intensity gradation: A conventional sensor exhibits rather coarse gradation at the darker intensity levels since the relative change between successive intensity levels is quite large. Our design, which employs a floating-point representation, has constant gradation throughout the dynamic range (about 2-20), which is far finer.

The entire communication architecture was modeled structurally at the gate and transistor level in Verilog, in two flavors: (i) an asynchronous one using Mousetrap pipeline primitives [19]; and (ii) a clocked one to compare energy efficiency. Gate delays were estimated using a 45 nm CMOS technology (45nm FreePDK from NCSU) and the NanGate Open Cell Library, with allowance for wiring loads, assuming $8\mu \times 8\mu$ pixels (typical of modern digital SLR cameras).

For comparison, an existing experimental intensity-to-frequency sensor [13] was also modeled as a base case. Note that this baseline sensor with spiking pixels is already superior in quality to today's commercial sensors, most of which use analog pixels with limited dynamic range and higher noise.

The sensor models were stimulated using a high-dynamic-range (HDR) scene synthesized via ray tracing, with a contrast ratio of 2^{32}:1, though 99.9% of all the intensity values lie within a ratio of 2^{22}:1 (or 22 bits). Each sensor's raw output was captured via simulation and then converted to images using software for viewing. Since the scene's dynamic range was far greater than is possible to show on paper, we provide two intensity-windowed versions of the results, focusing on the darkest and the brightest ends of the intensity range.

Fig. 10a shows that our approach can provide significant detail and smooth gradations in the dark regions, whereas the existing sensor offers much less information and shows significant banding. This is a direct result of our reciprocal measurement approach: time can be measured with much finer granularity than counting events in a fixed time interval.

Fig. 10b shows that in the brightest regions, our approach can provide significant detail, whereas the existing sensor has blown highlights. This is a direct result of two features: (i) intensity-to-frequency conversion, which avoid the problem of sensor saturation by repeatedly emptying and restarting; and (ii) in-pixel frequency decimation so that brighter intensities can be imaged without overwhelming the communication pipeline.

Authorized licensed use limited to: University of North Carolina at Chapel Hill. Downloaded on November 20,2026 at 04:43:40 UTC from IEEE Xplore. Restrictions apply.

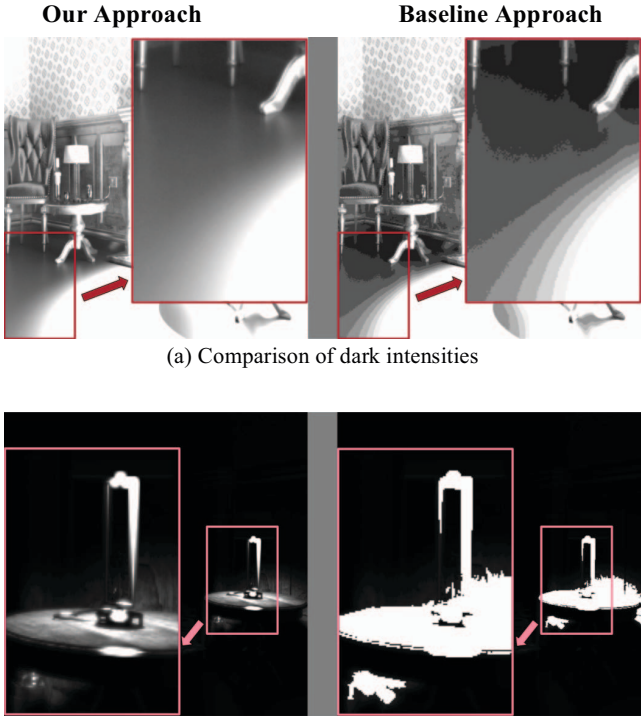
Overall, our sensor was able to capture a dynamic range of 22 bits, within which 99.9% of the image pixels lie. The measured power SNR between the scene and the image was 120 dB. While these results are for the asynchronous version of our sensor, the clocked version obtained results that were essentially identical.

They key difference between the asynchronous and clocked versions of our sensor, however, was observed in the switching activity in the communication network. The asynchronous version inserted 1.5 million pixel events into the communication network, aggregated over the entire sensor. The clocked version also had the same number of events carried by the network, but in 64 million clock cycles, giving it a relative efficiency of only 2.3%, compared to the asynchronous one. The reason is that the clock speed has to be fast enough to handle the worst case of a uniformly very bright scene; when illuminated by a scene of less overall intensity, only a fraction of the clock cycles carry meaningful information from the pixel array. In contrast, the asynchronous version only carries events when and where needed. Thus, for this example, the asynchronous version has a 43x (i.e., 64/1.5) energy advantage. Even if other more sophisticated clocking approaches, such as elastic clocking, were used, the design effort and area cost of running a 1 GHz clock to the entire pixel array is significant.

VII. Conclusions and Further Work

In this paper, we have introduced a novel sensing architecture that does away with the frame-synchronous paradigm and allows pixels to generate information asynchronously. The result is significant improvements in image quality (higher dynamic range, fine gradations), and the ability to vary how often updates are needed from different regions of the image, thereby allowing a more judicious use of precious communication bandwidth.

In ongoing work, we are undertaking detailed transistor-level design, and in the near future plan to develop a complete layout of the sensing array. Spice simulations on extracted layout will allow us to characterize noise, crosstalk, and other issues that Verilog simulations cannot capture. However, given that it is well established that the ADC quantization is the largest source of noise in current designs, and the dynamic range is limited because of analog saturation, our architectural evaluation gives us confidence that our proposed sensor has the potential to perform significantly better than existing practice.



(a) Comparison of dark intensities

(b) Comparison of bright intensities

Fig. 10. Comparison of our approach (left) and base approach [13] (right). The HDR image captured by each is shown by "windowing" into the dark and bright ends of the intensity range.

References

[1] Wikipedia entry on High-dynamic-range imaging. Retrieved June 10, 2014, from https://en.wikipedia.org/wiki/High-dynamic-range_imaging

[2] C. Auth, et al. "45nm High-k+Metal Gate Strain-Enhanced Transistors." Intel Technology Journal, Volume 12, Issue 2, 2008.

[3] Wikipedia entry on Posterization. Retrieved June 2, 2014, from http://en.wikipedia.org/wiki/Posterization

[4] Wikipedia entry on Weber–Fechner law. Retrieved June 2, 2014, from http://en.wikipedia.org/wiki/Weber–Fechner_law

[5] A. Darmont, and Society of Photo-Optical Instrumentation Engineers. *High Dynamic Range Imaging: Sensors and Architectures.* Publisher: SPIE, 2012.

[6] R.N. Clark. "Digital Camera Reviews and Sensor Performance Summary." Retrieved October 20, 2016, from http://www.clarkvision.com/articles/digital.sensor.performance.summary/#noise_sources

[7] Nikon Europe Inc. support article no. 64623. "What is a Backside Illumination CMOS sensor?" Retrieved June 10, 2016.

[8] D.Yang, B.Fowler and A. El Carnal. "A Nyquist-Rate Pixel-Level ADC." IEEE J. of Solid-State Circuits, vol.34, pp. 348-356, 1999.

[9] L. McIlrath. "A low-power low-noise ultrawide-dynamic-range CMOS imager with pixel-parallel A/D conversion." IEEE J. of Solid-State Circuits, pp. 846-853, 2001.

[10] A. Kitchen, A. Bermak, and A. Bouzerdoum. "PWM digital pixel sensor based on asynchronous self-resetting scheme." IEEE Trans. Electron Devices., vol. 25, no. 7, 2004.

[11] X.Wang, W. Wang, and R. Hornsey. " A high-dynamic-range CMOS image sensor with in-pixel light-to-frequency conversion." IEEE Trans. Electron Devices., Vol. 53, No. 12, pp. 2988-2992, 2006.

[12] Y. Chen, F. Yuan, and G. Khan,. "A New Wide Dynamic Range CMOS Pulse-frequency-modulation digital image sensor with in-pixel variable reference voltage." In *Proc. MWSCAS,* 2008.

[13] J. Doge, G. Schonfelder, G.T. Streil, and A. Konig. "An HDR CMOS image sensor with spiking pixels, pixel-level ADC, and linear characteristics." *IEEE Trans. Circuits Syst.,* vol. 49, no. 2, pp. 155158, 2002.

[14] A. Bermak. "VLSI implementation of a neuromorphic spiking pixel and investigation of various focal-plane excitation schemes." Int. J. Robot. Autom., vol. 19, no. 4, pp. 197205, 2004.

[15] B. Fowler, A. Gamal, and D. Yang. *A CMOS area image sensor with pixel-level A/D conversion.* Stanford University tech report, 1995.

[16] K. Boahen. "Neuromorphic microchips." Scientific American 292.5 (May 2005): 56-63.

[17] P.A. Merolla, et al. "A million spiking-neuron integrated circuit with a scalable communication network and interface." Science 345.6197 (Aug 2014): pp. 668-673.

[18] P. Lichtsteiner, C. Posch, and T. Delbruck. "A 128X128 120 dB 15 us Latency Asynchronous Temporal Contrast Vision Sensor." IEEE Journal of Solid-State Circuits, vol. 43, no. 2, pages 566-576, 2008.

[19] M. Singh and S.M. Nowick. "MOUSETRAP: High-Speed Transition-Signaling Asynchronous Pipelines." IEEE Transactions on VLSI Systems (TVLSI), vol. 15, no. 6, pages 684-698, June 2007.

[20] DxOLabs. DxOMark Sensor Scores. Retrieved June 1, 2014, from http://www.dxomark.com/About/Sensor-scores/Use-Case-Scores

Authorized licensed use limited to: University of North Carolina at Chapel Hill. Downloaded on November 20,2026 at 04:43:40 UTC from IEEE Xplore. Restrictions apply.