A Low-Power Low-Noise Ultrawide-Dynamic-Range CMOS Imager with Pixel-Parallel A/D Conversion

Lisa G. McIlrath, Member, IEEE

Abstract—A CMOS image sensor with pixel-parallel analog-to-digital (A/D) conversion fabricated with different array sizes and photodiode types in a three-metal 0.5-/m process is presented. Nominal power dissipation is 40 nW per pixel at \( V_{DD} = 3.3 \) V. A/D conversion results from sampling a free-running photocurrent-controlled oscillator to give a first-order \( \Sigma-\Delta \) sequence. The sensor displays dynamic range capability of greater than 150 000 : 1 and exhibits fixed pattern noise correctable to within 0.1% of signal.

Index Terms—Analog–digital conversion, image sensors, sigma–delta modulation, smart pixels.

I. INTRODUCTION

PIXEL-PARALLEL analog-to-digital (A/D) conversion has several advantages over column-parallel methods, now commonly used in most CMOS imagers [1], [2]. Pixel readout is not limited by bus settling time and can be processed as fast as the frame memory interface will allow. Maintaining all analog processing within the pixel can reduce overall system power and allows scaling to larger array sizes more easily than a design where analog values must be transmitted over long capacitive busses.

An A/D architecture which allows a great deal of flexibility for many application-specific needs is an oversampling \( \Sigma-\Delta \) converter, as its output resolution and frame rate can be easily adjusted to suit the user. Some previous designs based on implementation of a synchronous first-order \( \Sigma-\Delta \) pixel-parallel converter have been demonstrated [3], [4]. In this paper, it is shown that a free-running continuous oscillator sampled at fixed intervals can be used as a first-order \( \Sigma-\Delta \) converter in a low-power CMOS imager to give high-quality images with extremely wide dynamic range and low noise.

A sampled free-running oscillator pixel does not require any internal clocked components, is easy to build, and its bandwidth is limited only by the speed of its internal comparator—its only analog component. In addition to demonstrating the performance of two prototype arrays of these pixels fabricated in a three-metal 0.5-/m process, the focus of this paper is on some variation on the 3T voltage-mode pixel shown in Fig. 1 [1]. The photodiode node voltage is periodically reset to \( V_{rst} \), and the photocurrent from the reverse-biased diode is integrated on the node capacitance for a fixed exposure time. The final voltage is read by raising the select signal (SEL) to connect the source of M2 to the column bus, which is driven by a current source and acts as the output node of a source follower.

The primary problem with this design is the limited capacitance and voltage range available at the sensing node. With typical storage capacity of between 20 and 50 Ke-, photon shot noise by itself limits the resolution at the pixel to around 7 b. Less than unity gain of the source follower, if well ties are unavailable, can then consume another bit, leaving only 6 b/pixel. As normal scene illuminations may contain over five orders of magnitude (100 dB, or 17 b) variation from the darkest to brightest objects, the limited signal-to-noise ratio (SNR) of the voltage-mode pixel, as is, is unacceptable for photographic applications.

To overcome this problem, many designs have incorporated the use of time, in one form or another, as a control variable. Decker et al. used a time-dependent barrier voltage to control the electron capacity of the integration node [5] producing a nonlinear illumination-to-voltage transfer curve. In the locally autoadaptive (LARS) imager developed by Silicon Vision GmbH [6], the integration time for each pixel is determined by how long it takes the sense node voltage to equal or pass a given reference level. The integration time is measured in
intervals, and a time-stamp voltage is stored in the pixel to record the number of intervals required to cross the reference level. By allowing the integration times to range from 5 μs to 2.56 ms, or 54 dB, an additional 9 b of resolution are added to the inherent dynamic range of the pixel.

A third design, developed by Yang et al. [7], used multiple sampling combined with a pixel-parallel analog-to-digital converter (ADC) to achieve a wide-dynamic-range digital output imager. Their approach was to image the scene \( k \) times per frame at exponentially increasing exposure periods, \( T, 2T, \ldots, 2^kT \), digitizing the outputs at each exposure at low resolution (4–6 b). By combining the values from each exposure at which the pixel did not saturate, a higher resolution \( m \)-bit value \((m > 4 \rightarrow 6)\) could be obtained. This value, together with the exponent of the longest nonsaturating exposure period, then gave a floating-point digital representation of the pixel illumination.

While the preceding examples used either variable integration times or time-dependent well capacities to increase dynamic range, other designs have been based on directly measuring the time it takes the photocurrent to produce a given voltage change at the sense node. Similarly to the voltage-mode pixel, the canonical “integration-time”-mode pixel is depicted in Fig. 2. In this case, the sense node is connected to a comparator—which changes state when \( V_p \) goes below some reference value \( V_{ref} \). The state is reflected in the binary signal \( OUT \), which may be connected to an output bus and/or fed back to the reset transistor. If a global signal is used to reset the sense node to \( V_{ref} \), the pixel operates as a timer. Conversely, if the loop is closed to connect the comparator output to the reset transistor, the pixel becomes an oscillator which generates pulses on the \( OUT \) node at a frequency inversely related to the integration time. One of the first imagers based on direct integration-time measurement was the MAPP2200 sensor developed by Forchheimer et al. [8]. Later, Brajovic [9] developed a twist on the direct timer architecture by assigning indices to pixels based on their relative switching times. This design allowed inherent gain control and histogram-based quantization, but was limited by its use of analog values to represent global quantities.

Yang [10] presented perhaps the first implementation of an oscillating pixel sensor. By selecting the \( OUT \) signals from a given row onto column busses and then measuring (with a counter) the time between pulses, he was able to demonstrate a \( 32 \times 32 \) imaging array with very wide—over five orders of magnitude—dynamic range sensitivity. While Yang’s pixel design is similar to the one presented here, the possibility of sampling the output to obtain a \( \Sigma-\Delta \) sequence was never explored. The need to time each pulse resulted in long row readout times to capture dimly illuminated pixels. In addition, the pixel comparator, implemented with CMOS inverters, consumed significant power under normal room light illuminations.

Functionally, the designs most closely related to the one presented here are the pixel-parallel \( \Sigma-\Delta \) ADC imagers developed by Fowler [3] and D. Yang et al. [4]. The primary difference is that these designs are based on a synchronous first-order \( \Sigma-\Delta \) architecture containing a clocked comparator and a switched-capacitor circuit. In a sense to be clarified in the next section, the synchronous \( \Sigma-\Delta \) modulator is a discrete-time oscillator and fits into the canonical integration-time-mode pixel depiction by adding a clock input to the box in Fig. 2. As shown below, a synchronous modulator and a sampled oscillator driven by the same inputs will produce the same outputs. However, the former is more difficult to build. The synchronous modulator requires several analog components—a comparator, a 1-b D/A, and a voltage summing operational amplifier—all of which contribute to fixed pattern noise and design complexity. Furthermore, nonidealities in the analog circuits limit the maximum clock rate at which they may be operated, and in turn limit the achievable dynamic range. The design presented here, on the other hand, is a straightforward relaxation oscillator which has been optimized for low power. The pixel core contains a single comparator, its only analog component, and a pulse reset circuit. Instead of directly connecting the \( OUT \) signal, as it is called in Fig. 2, to a bus, the information that it has switched state (or not) is stored on a MOS gate. This value is then read out and reset at a frequency determined by an external sampling clock. As reading this bit does not affect the oscillator, only the diode photocurrent determines the pixel frequency. It is thus possible to realize the full dynamic range allowed by the process technology.

III. Theory

A. Sampled Oscillator as a First-Order \( \Sigma-\Delta \) Modulator

The equivalence between a synchronous first-order \( \Sigma-\Delta \) modulator and a sampled oscillator running asynchronously with respect to the sampling clock was first observed by Candy and Benjamin [11]. To illustrate this relation, we first consider the canonical form of the synchronous first-order \( \Sigma-\Delta \) modulator as shown in Fig. 3. An input \( x[m] \), generated by sampling the continuous quantity \( x(t) \) at time \( mT \), is fed into an accumulator. If the accumulator output is above some threshold value \( W_{ref} \) the output \( y[m] \) is set to 1 and a quantity \( x_{max} \) is subtracted from the next sample. If the accumulator output is
Fig. 4. Waveforms of a synchronous $\Sigma$–$\Delta$ modulator (solid staircase) and an equivalent asynchronous oscillator (dashed sawtooth) based on a continuous integration of the input.

Fig. 5. Periodic rectangular waveform generated by the asynchronous oscillator and the binary output from sampling at a rate $f_s = 1/\tau$.

below threshold, $y[m]$ is set to 0, and the next input sample passes unmodified to the accumulator. A typical waveform for a constant input $x(t) = x$ is shown as the solid staircase pattern in Fig. 4.

Next, consider a continuous-time circuit which integrates an input $x$ until the accumulated value reaches the reference level $W_{\text{ref}}$. At this point the accumulator is reset to zero, giving the sawtooth pattern shown in the diagram. After dimensional scaling to eliminate $t$, such that $w[m] = w(mt)$, it can be seen graphically that the continuous and discrete-time waveforms track each other. Furthermore, it can be observed that if the synchronous modulator crosses the threshold on a given clock edge $m$, then the equivalent asynchronous oscillator must have reset during the interval $((m - 1)\tau, m\tau]$. Suppose now that the asynchronous circuit generates a pulse of width $\tau$ each time it resets and that this pulse is sampled on the next clock edge to generate the output, as shown in Fig. 5. The width of the pulse being set equal to the clock period guarantees that it will be sampled exactly once. Clearly, the binary output streams from the sampled oscillator and the first-order modulator will be the same.

One can in fact generate such a pattern from any type of oscillator, whether internally it follows the sawtooth pattern of Fig. 4 or not, as long as it is possible to remember that a reset has occurred until the next clock edge. Sampling provides information only on the oscillator frequency. In the case of the sawtooth waveform shown in the diagram, the frequency is proportional to the input and is the inverse of the time it takes to integrate from 0 to $W_{\text{ref}}$, or $T = W_{\text{ref}}/x$. If $T \leq \tau$, the output will saturate (all 1’s) giving no further information. The maximum input which can be observed is thus $x_{\text{max}} = W_{\text{ref}}/\tau$, and the oscillator frequency can be expressed as

$$f = \frac{1}{T} = \frac{x}{x_{\text{max}} \cdot \tau}. \quad (1)$$

In the more general case, where the frequency is a monotonic increasing, but not necessarily linear, function of the input, one still measures $f$. Equation (1) then describes a linear mapping of frequency values onto the range $[0, 1]$ in units of $1/\tau$.

B. Decoding First-Order $\Sigma$–$\Delta$ Streams

The bit streams generated by each pixel must be decoded outside the imager array to produce digital numbers. Several techniques, ranging from convolution with finite-impulse response (FIR) filters to linear programming methods, have been investigated for this purpose in other work [3]. In the course of the present design, a new $O(N \log N)$ optimal algorithm was developed to decode first-order $\Sigma$–$\Delta$ sequences produced by a constant input $x \in [0, 1]$ [12]. While a full proof of the underlying theorem is outside the scope of this paper, in summary it was shown that for every such sequence generated by $x \in (0, 1/2]$, one can derive a bound on $[1/x]$, i.e., the greatest integer less than or equal to $1/x$, and generate a shorter $\Sigma$–$\Delta$ subsequence corresponding to the constant input $x' = \langle 1/x \rangle \equiv 1/x - [1/x]$. As it can be shown that every sequence generated by $x \in [1/2, 1]$ has a one-to-one mapping to a sequence generated by $\tilde{x} = 1 - x$, $\tilde{x} \in (0, 1/2]$, this theorem allows us to recursively decompose every constant input first-order sequence to determine the most probable digital estimate of its input. The algorithm can be performed with fixed point arithmetic and, unlike with $O(N)$ FIR filters, it is never necessary to recompute filter weights as a function of the sequence length.

A plot of the average SNR versus number of samples for three methods: the optimal decoder and two FIR filters, is shown in Fig. 6. These averages were computed over a large sample set...
for each input $x \in (0, 1)$ with randomly chosen initial states of the modulator. As can be seen, the optimal decoder not only produces a roughly 4.2-dB overall improvement in SNR, but has a slightly higher slope, 9.1 dB/octave, than the best FIR filter, which exhibits only 9 dB/octave.

The recursive algorithm is also robust with respect to input noise, which in the present case could be caused by shot noise, dark current bursts, or FET noise in the input comparator, all of which would result in jitter in the oscillator frequency. Because of the regularity in the first-order $\Sigma$–$\Delta$ sequence patterns, very simple error-correction techniques can be applied. Fig. 7 shows the degradation of output versus input SNR for the three methods and for two sequence lengths: 32 and 64 samples. Two observations may be derived from these graphs. First, the optimal decoder is still superior to FIR methods at least down to 25-dB input SNR. Second, the output SNR degradation with input noise is less for shorter length sequences.

The beauty of the sampled oscillator approach, however, is that one can change the sampling frequency without affecting the operation of the pixel oscillators. By sweeping through a set of binary weighted frequencies, $f_x = f_0, f_0/2, \ldots, f_0/2^k$, and recording only short-length sequences—e.g., 8–16 b—for each pixel, one can produce a high-resolution output image over a very wide dynamic range. The concept is identical to that applied by D. Yang et al., in their floating-point pixel-level ADC image sensor [7], although the basis of their design was a voltage-mode pixel whose output was digitized after repeated exposures at binary-weighted integration times.

In the case of the sampled oscillator, one can read out 8 samples/pixel at each frequency to obtain an average 25 dB, or 4–5 b, of resolution. In this case, the output bandwidth required is $2x$ that which would be needed if an on-chip 4-b ADC had been implemented. If $K$ is the smallest integer such that the ratio between the brightest and darkest pixels in the scene is bounded above by $2^K$, $[P_{\text{bright}}/P_{\text{dark}}] \leq 2^K$, then only $8K$ samples need be taken to give a total dynamic range of $6K + 25$ dB.

Certainly, other combinations of sampling frequencies and sequence lengths may be used for different application requirements. However, the point to be made is that wide-dynamic-range high-resolution data may be obtained from the sampled oscillator imager with only $2x$ the output bandwidth needed for an integrating voltage-mode sensor with pixel-parallel ADC. In exchange for the additional bandwidth, one obtains an extremely simple and robust architecture containing...
the absolute minimum number of analog components that any digitizing output sensor could have.

IV. CIRCUIT DESIGN

The schematic of the implemented asynchronous oscillator cell is shown in Fig. 9. The input signal is the photocurrent generated by the n⁺-p photodiode. The voltage on the integrating node decreases over time and is reset to $V_{DD}$ when it drops below the global reference level, $V_{low}$. The circuit is composed of four sections: a differential amplifier, which continuously compares the photodiode voltage to $V_{low}$; a bistable half-latch which triggers the reset; a regenerative section that switches the bistable latch and restarts the integration; and pulse capture logic that stores a bit upon reset. A row select signal $SEL$ gates the cell output onto a column bus in order to read the bit. Following the read, signal $QS$ resets the storage bit to GND.

The differential pair and the common-source amplifiers are biased in the weak inversion region to reduce power and maximize gain. Transistor dimensions were sized to optimize the balance between glitch energy and the on-time of resistive paths between $V_{DD}$ and GND in order to minimize dynamic power and to effectively eliminate coupling between adjacent cells. In the three-metal 0.5 $\mu$m implementation, the unit cell, including the photodiode, measured 30 $\mu$m x 30 $\mu$m.

The external control signals needed to operate the imager are very basic. Voltages $V_{bias}$ and $V_{low}$ are set externally by DAC's contained on the camera board. Data readout is performed by providing in sequence three pulsed signals for each row: bus precharge, row select ($SEL$), and bit reset ($QS$). In order to minimize the possibility of losing a bit, the pulse width of these signals is kept short (<100 ns) regardless of the row sampling rate. Rows may be selected using either a direct address decoder or a sequential shift register. Column bits are read out in parallel in groups, with the number of columns per group determined by the number of available output pins (in our case, 16). For normal image acquisition, each row is read in order. Hence, if $t_{row}$ is the time required to read one row, and there are $N$ rows in the array, the effective sampling frequency seen at each pixel is

$$f_s = \frac{1}{N \cdot t_{row}}.$$  \hspace{1cm} (2)

Providing binary weighted sampling frequencies is easily performed on the camera board by dividing down the clock which controls the row-pattern-generating programmable logic device (PLD).

V. EXPERIMENTAL RESULTS

Two imagers were fabricated in a three-metal 0.5-$\mu$m process through the MOSIS service: a 48x48 array with n-well photodiodes, and a 64x64 array with nonsilicided n⁺-diffusion photodiodes. Sample raw images from each are shown in Fig. 10.

The pixel bias current was nominally set at $\sim$12 nA, or 40 nW power dissipation per pixel at $V_{DD} = 3.3$ V by controlling the total average current drawn by the full array to be $\sim$50 $\mu$A. As almost all on-chip processing, other than the row and output select logic, is done in the pixel, this number essentially determines the total on-chip power dissipation. Extensive measurements were made to verify that the sensor performance was not significantly impacted by exact bias current levels. As seen by the plots in Fig. 11, only modest variations—less than 2%, and of the same order as fluctuations in the poorly controlled light source—were observed at fixed illumination as $V_{bias}$ varied from its nominal value to the point of strong inversion.

To determine the dynamic range of the sensor, the average output with respect to sampling rate was measured over a wide range of illuminations. The binary pixel output streams encode the ratio of the oscillator frequency to the sampling frequency

$$f_s = \frac{x}{x_{\max}} = \frac{f_{pix}}{f_s} = \frac{I_{ph}}{C\Delta V}.$$  \hspace{1cm} (3)

Due to charge injection from the reset switch, $V_{max}$ differs slightly from $V_{DD}$. For each of a set of clock frequencies from 305 Hz to 156 kHz, the light level was set to just saturate at the given frequency, i.e., $I_{ph}/C\Delta V = f_s$. The sensor output was then measured at this illumination at several higher sampling frequencies. Some of the results for the 48x48 imager are plotted in Fig. 12. Due to the range of illuminations measured,
all of the data cannot be shown on one graph. Nonetheless, it was clearly demonstrated that the pixel could oscillate at photocurrent-induced frequencies up to 156 kHz before bandlimiting of the internal differential amplifier became apparent. The output with respect to sampling period was linear at all frequencies, further confirming that sampling did not affect the pixel frequency. Measurement of the absolute light intensity to frequency transfer characteristic was more difficult because of the lack of calibrated light sources and photodetectors over the range of illuminations tested. Calibrated measurements of monochromatic 610-nm light at irradiances up to 20 μW/cm² did indicate a linear response, however. Having determined the minimum sampling rate (≈1 Hz) in the dark at which no response was measured due either to dark current or to transistor leakage, it could be ascertained that the effective dynamic range of the pixel oscillator was greater than 150,000:1.

The pixel frequency can also be controlled through the global \( V_{\text{low}} \) parameter. Let \( \alpha \equiv C(V_{\text{max}} - V_{\text{low}})/(I_{\text{ph}}\tau) \) and \( \dot{x} \equiv x/x_{\text{max}} \), then (3) can be rewritten as

\[
\frac{1}{\dot{x}} = -\alpha V_{\text{low}} + \alpha V_{\text{max}}. \tag{4}
\]

A least-squares linear fit to a plot of \( 1/\dot{x} \) versus \( V_{\text{low}} \), as shown on the left side of Fig. 13, allows us to determine both \( \alpha \) and \( V_{\text{max}} \).

Spatial and temporal noise were measured by acquiring 32 sample images at each of several parameter settings. Temporal noise, computed as the average of the standard deviations of each pixel value across the 32 samples, was measured to be between 0.35% and 0.45% of signal for each sensor type. Repeated measurements of temporal noise while varying \( V_{\text{low}} \) and \( V_{\text{bias}} \) confirmed that it was uncorrelated with either of these parameters. It should be noted that correlated double sampling is unnecessary for the sampled oscillator pixels. Repeated sampling of the integration time effectively reduces the noise due to the reset transistor by the square root of the number of resets in the sequence. Furthermore, error correction in the output stream decoder, as described in Section III-B also diminishes the effect of frequency jitter.
Fixed pattern noise, defined as the standard deviation of array values from the array mean at constant illumination, was computed after averaging the 32 sample images to eliminate temporal variations. The raw fixed pattern noise relative to signal was found to be essentially independent of either $V_{k_m}$ or of the illumination, but was clearly related to $V_{k_w}$, as seen by the plot on the right of Fig. 13. Referring to (4) and writing $\Delta V \equiv V_{\text{raw}} - V_{k_w}$ as $\Delta V = \Delta V + \delta$, where $\Delta V$ is the average value and $\delta$ the local variation, we obtain

\[
\hat{x} = \frac{1}{\alpha(\Delta V + \delta)} \approx \frac{1}{\alpha \Delta V} \left( 1 - \frac{\delta}{\Delta V} \right). \tag{5}
\]

The standard deviation in $\hat{x}$ is thus

\[
\sigma_{\hat{x}} \approx \frac{1}{\alpha \Delta V} \cdot \frac{\sigma_\delta}{\Delta V}, \tag{6}
\]

where $\sigma_\delta$ is the standard deviation of $\Delta V$ and the absolute error $\sigma_{\hat{x}}/\hat{x}$ is thus $\approx \sigma_\delta/\Delta V$. From a least-squares fit to the measured absolute error versus $V_{k_w}$, it can be estimated that $\sigma_\delta$ is approximately 35–40 mV across the array. While some of this variation may be accounted for by fixed differences in the average charge injected by the reset transistors, $V_f$ mismatch in the comparators is most likely the primary source of fixed pattern noise.

The encouraging conclusion of these measurements is that the fixed pattern noise sources, whether due to reset transistor capacitance mismatch or $V_f$ mismatch in the comparators, are relatively constant—i.e., temperature- and illumination-independent properties of each cell—and thus can be easily corrected. To make this point, the relative gain correction factors for the 64×64 array were computed for each pixel at a given $V_{k_w}$ setting from a single image acquired under flat-field illumination. The left side of Fig. 14 shows mesh plots of subsequent raw images taken at different illuminations, while on the right are shown the plots of the same images after multiplication with the computed correction factors. The residual relative fixed pattern noise after correction, measured over many illumination levels and $V_{k_w}$ values, has been shown to be approximately 0.1%.

VI. CONCLUSION

Pixel-parallel A/D conversion based on sampling a photocurrent-controlled relaxation oscillator at each pixel has been shown to be a viable approach for building high-quality CMOS imagers. Very wide-dynamic-range images (>150 000 : 1, or 104 dB) may be obtained with only a 2× increase in output bandwidth over that required for a conventional Nyquist-rate digitizing imager. Output resolution and/or dynamic range can be arbitrarily adjusted for any application by altering the control signal timing. In exchange for the extra output bandwidth needed for oversampling, one obtains a very simple architecture that contains a minimal number of analog components and that requires very little interface logic on the camera board. The imager operates at very low power,
\(\sim 40 \text{nW/pixel at } V_{DD} = 3.3 \text{ V}—\text{a } 1K \times 1K \text{ array of these pixel processors would consume less than } 50 \text{ mW at normal light levels—exhibits very low temporal noise, } \sim 0.4\% \text{ of signal, and its fixed pattern noise is correctable to within 0.1\% of signal. This performance was demonstrated through extensive measurements on two prototype imagers, a } 48 \times 48 \text{ array with n-well photodiodes and a } 64 \times 64 \text{ array with nonsilicided n}^+\text{-diffusion photodiodes, fabricated in a three-metal 0.5-} \mu \text{m process through the MOSIS service.}

ACKNOWLEDGMENT

The author would like to thank S. Lefian who designed the test apparatus and assisted with the experiments, and V. Lum who acquired many of the images.

REFERENCES


Lisa G. McIlrath (S’91–M’92) received the S.B. degree in physics from the Massachusetts Institute of Technology (MIT), Cambridge, in 1976, the M.S.E. degree from the University of Texas, Austin, in 1989, and the Ph.D. degree in electrical engineering and computer science from MIT in 1994.

From 1994 to 1998, she was Assistant Professor of electrical and computer engineering at Northeastern University, Boston, MA, and from 1998 through September 2000, a Visiting Professor at the MIT Artificial Intelligence Lab. Her research interests are in highly parallel VLSI computational sensors for machine vision and 3-D analog/digital circuit architectures. Since 1996, in collaboration with MIT Lincoln Laboratory, she has led the effort to develop a 3-D integrated circuit process technology. She is the Founder and President of 3D-IC, Inc., whose mission is to develop commercial products based on this technology.

Dr. McIlrath was the recipient of a 1996 National Science Foundation Career Development award.