A CMOS Array Sensor for Sub-800-ps Time-Resolved Fluorescence Detection

T. Huang*, S. Sorgenfrei*, K. L. Shepard*, P. Gong[#*], and R. Levicky[+#]
*Department of Electrical Engineering and [#]Department of Chemical Engineering. Columbia University, New York, NY 10027
[+]Department of Chemical and Biological Engineering. Polytechnic University, Brooklyn, NY 11201
E-mail: tdh2@columbia.edu, shepard@ee.columbia.edu

Abstract

This paper describes the design of an active CMOS sensor array for fluorescence applications which enables time-gated, time-resolved fluorescence spectroscopy. The 64 x 64 array is sensitive to photon densities as low as 8 x 10^6 photons/cm^2 with 64-point averaging and, through a differential pixel design, has a measured impulse response of better than 800 ps. Applications include both active microarrays and high-frame-rate imagers for fluorescence lifetime imaging microscopy.

1. Introduction

Fluorescence techniques find wide application in biology and are ideally suited for biomolecular detection both *in vivo* and *in vitro*. Targets molecules of interest are labeled with fluorophores. When optically excited, photons emitted from these labels have a longer wavelength than the excitation source (Stokes shift). Optical filtering (of typically more than 160 dB) is employed for background rejection, filtering out the excitation light. Sensitive photodetection approaches, including cooled CCD imagers and photomultiplier tubes (PMTs), measure the fluorescent light.

One of the most pervasive *in vitro* application of fluorescence is in surface-based biomolecular (DNA, protein) microarrays[1], which are increasingly popular and powerful tools in many medical and biological applications. In this case, fluorescently-labeled analyte targets from solution bind to appropriate probe molecules immobilized on a passive solid support. Microarray laser scanners are then employed to image these arrays. Fluorescent techniques are also widely employed *in vivo* to observe the location of molecules in cells. Both wide-field epifluorescent microscopes and confocal laser-scanning microscopes can be employed.

Fluorophores have associated with them a characteristic lifetime, which defines the exponential fluorescent decay transient after the removal of the excitation source. Fluorescent lifetimes are sensitive to excited-state reactions such as fluorescent resonant energy transfer (FRET)[2], enabling a growing field of fluorescent lifetime imaging microscopy (FLIM)[3]. Lifetimes, on the order of nanoseconds for organic dyes and longer for quantum dots[4], can be imaged spatially with detectors capable of either high-frequency modulation or fast time-gating. Such time-resolved fluorescence detection also provides additional advantages in improved background rejection.

Time-resolved applications require a detector that is "fast," as determined by the time-constant(s) of the impulse response, and sensitive, as determined by integrated emission photon flux (photons/cm^2). In most commercial time-resolved systems, time-correlated single photon counting (TCSPC) is employed with photomultiplier tube (PMT) detectors and laser scanning, with time resolution limited by jitter and sensitivity limited by dark count[5]. Frame rates are typically limited to fewer than 0.2 frames/s[6]. Alternate time-resolved fluorescence detectors rely on CCD imagers and gated intensifiers[7]. Frame rates of up to 100 frames/s have been achieved for a small imager (120-by-60) if only two points of the decaying exponential are sampled, which is suitable for some monoexponential lifetime characterization. In this work, we consider the design of a fast, time-resolved CMOS imager specifically designed for high-frame-rate FLIM applications with a timing resolution of better than 150 ps and frame rate of better than 150 frames/s for a 64-by-64 imager with two points per transient waveform extracted.

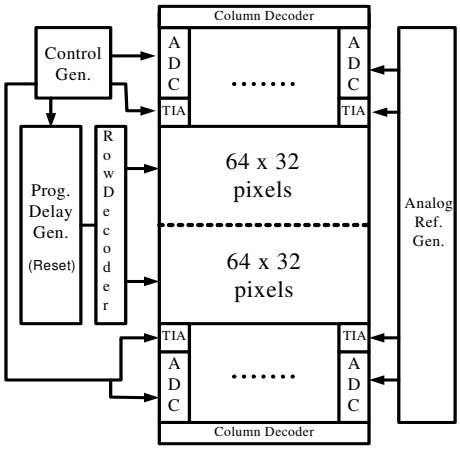

Fig. 1. Prototype chip top level block diagram.

Our previous efforts[8] at CMOS time-resolved imagers were specifically directed toward active microarray applications. This imager had quantization-noise limited sensitivities of more than 10^8 photons/cm^2 and impulse-response time-constants of more than 1.1 ns. Impulse response was further limited by a long "tail" in the photocurrent. The new design described here improves on our earlier efforts in several significant ways to enable high-frame-rate FLIM applications in addition to improved performance as an active microarray platform. Differential photodiodes (PD) are implemented to improve photocurrent impulse response and noise immunity while preserving fast time gating. Twelve-bit data conversion produces read-noise-limited sensitivities. Correlated double sampling (both digital and analog) reduces 1/f noise. Active reset is employed to reduce pixel reset noise. These combined techniques produce an imager with sensitivities approaching 8x10^6 photons/cm^2, timing resolutions of better than 150 ps for lifetime measurement, impulse-response time constants of better than 800 ps, and frame rates of better than 300 frames/s for a 64-by-64 imager with 64 points of averaging.

This work was supported in part by the NIH under grant HG003089, by the NSF under grant BES-428544, and by NYSTAR.

2. Biochip architecture

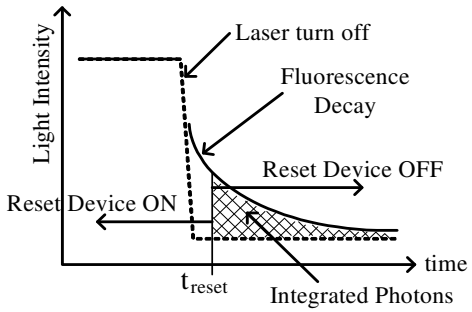

Fig. 2. Time-gated, time-resolved fluorescence detection.

Fig. 1 shows the top level block diagram of the prototype chip. The chip contains a 64-by-64 array of current-mode-output pixels, with per-column transimpedance amplifiers (TIAs) and per-column integrating analog-to-digital converters (ADCs) for row-multiplexed conversion. The pixels are divided into two blocks in order to mitigate the effect of wire loading on pixel response and to increase frame rate by a factor of two. The transimpedance amplifiers have a nominal gain of 1 kΩ and a 3-dB bandwidth of 350 MHz. Data conversion is pipelined, occurring concurrently with the digital outputs from the previous ADC cycle being column-multiplexed off the chip. In addition to selecting rows, the

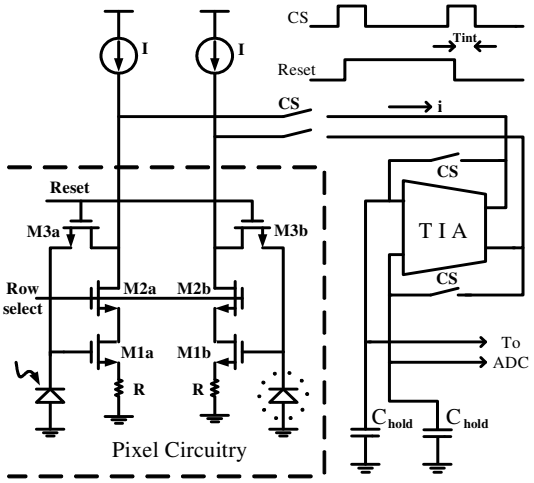

Fig. 3. Differential pixel circuitry with active reset.

row decoders gate the pixel reset signal for unselected pixels. Time-resolved measurement comes from repeated photocurrent integration with a successively delayed reset as shown in Fig. 2. The reset delay line has 511 steps of 150 ps, providing a range of 75 ns.

Differential photodiodes. The impulse response of the pixel is largely determined by the photocurrent response. In conventional CMOS PDs, minority carriers generated within a diffusion length of the depletion region of the nwell/psub PD create a long "tail" to the photocurrent impulse response. In this application, since the excitation source is much stronger than the fluorescent signal, the diffusive component of the photocurrent could potentially overwhelm the fluorescent photocurrent collected after time gating, depending on the effectiveness of optical filtering. To attenuate this diffusive component of the impulse response, a fingered differential PD is used in which alternating fingers are covered with metal[9], rendering the diffusive component of the photocurrent common-mode.

The size of these fingered pixels is limited by n-well spacing rules and the number of fingers required for sufficient common-mode rejection. In this design each pixel, measuring 50-μm on each side, contains a total of eight PD fingers. Each finger is 3.2 μm wide and 45 μm long. Finger periodicity, defined as the sum of finger width and spacing, is 4.8 μm. Smaller pixel sizes can be realized with n-diffusion/p-sub diodes, but photocurrent collection efficiency in the wavelength of interest would be reduced. The ability for such differential PDs to reject common-mode diffusive minority carriers will improve with technology as minimum n-well spacing decreases.

Pixel circuitry. Other high-sensitivity CMOS imagers[10] have benefited from the use of long integration times to improve SNR. This is not possible in this case, because, as shown in Fig. 2, an exponentially decaying photocurrent response is being integrated; as a result, longer integration times results in no increase in integrated photocurrent signal. This imager employs averaging of a number of repeated measurements alone to achieve the desired SNR. Read noise must be limited by other means in order to achieve high frame rates.

Reset noise, being physically limited by PD size, is the most dominant component of total read noise. However, by employing active reset techniques[11] in the pixel (through transistors M3a and M3b in Fig. 3), it is possible to reduce reset noise power by a factor equal to the feedback gain. The transconductor used during active reset is employed again for voltage-to-current conversion for the pixel. This differential current is routed to the transimpedance amplifier (TIA) during the integration period, during which the *reset* signal is low and the *cs* signal is high. At the end of the integration period, a single-slope (with dual-ramp option) integrating ADC converts the resulting sampled differential voltage on the C_{hold} capacitors. The response time of the TIA (C_{hold}/G_m= 1 ns, where $1/G_m$ is the input impedance of the TIA) will not limit system performance as long as this time constant is much smaller than the integration window. A sufficiently large C_{hold} of 1 pF is used to reduce its kT/C noise to below that of the pixel reset noise.

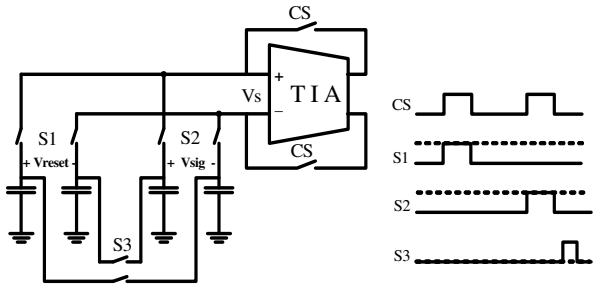

Fig. 4. Schematic to accommodate both analog and digital CDS.

Correlated double sampling (CDS). CDS is effective for canceling certain systematic offsets such as fixed pattern noise (FPN) and for reducing 1/f noise from pixel and transimpedance amplifiers. In this imager, CDS is implemented through both analog and digital methods.

To implement analog CDS, switched capacitors are used to store signals captured during reset (V_{reset}) and photocurrent integration (V_{sig}) as shown in Fig. 4. The difference of the two signals is taken when switch S3 closes, after the second sampling period has completed, yielding the differential voltage $V_{S,CDS}=(V_{sig}-V_{reset})/2$.

CDS is also performed digitally, in which both samples are converted right after the integration period and the difference taken after conversion. In this mode, S1 and S2 are always on and S3 is always off to further lower sampled kT/C noise.

Dual-ramp, single-slope integrating ADC. An integrating ADC is employed here due to its relatively small area, low power, and high resolution. In order to achieve sufficient frame rate for high-speed FLIM applications while facilitating sensitivity requirement through averaging, a dual-ramp, single-slope ADC architecture is employed as shown in Fig. 5.

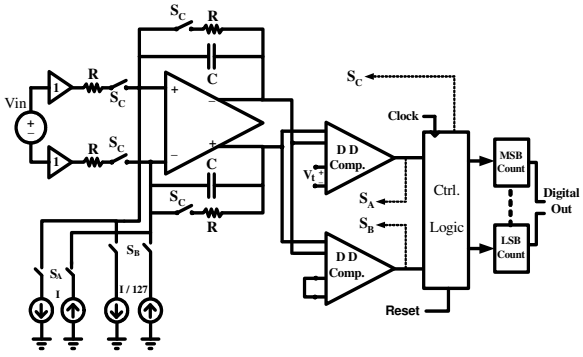

Fig. 5. Dual-ramp single-slope ADC

At the start of conversion, switches controlled by S_C are closed, sampling the input voltage at the output of the differential amplifier. At the start of discharge/integration phase, the S_C switches are open while both S_A and S_B are closed. The differential voltage drops until it falls below the threshold voltage V_t, at which time the differential difference comparator opens switch S_A. Fine integration continues until the differential voltage at the output of the amplifier drops below zero. The most-significant-bit (MSB) counter determines the number of clock cycle S_A stays closed (coarse integration) while the least-significant-bit (LSB) counter measures number of cycles required for fine integration, defined by the time between when S_A opens and when S_B opens.

This dual-ramp ADC architecture, which improves conversions rates by a factor of 64, adds an additional comparator and reference current source over what would be required for a single-ramp implementation. Accuracy is degraded by mismatch between coarse and fine discharge differential current sources. For higher accuracy (but lower conversion rate) the ADC can also be used as a conventional single-slope ADC by combining both counters as one and leaving S_A open during the integration phase.

3. Measurements

The prototype 5-mm-by-5-mm imager chip is fabricated in a standard mixed-signal 0.18 μm CMOS process. Fig. 6 shows measured external quantum efficiency (QE) using a standard measurement setup including a monochromator, integrating sphere, and calibrated photodiode. This quantum efficiency can be expected to be depressed by the differential

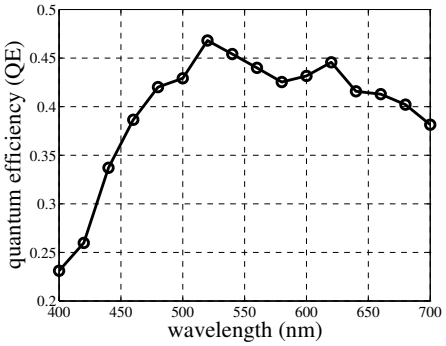

Fig. 6. Measured quantum efficiency of the differential PD.

nature of the pixel design since some of the integrated photocurrent will be common-mode and rejected. Peak quantum efficiencies of greater than 42% occur for wavelengths between 600 nm and 650 nm.

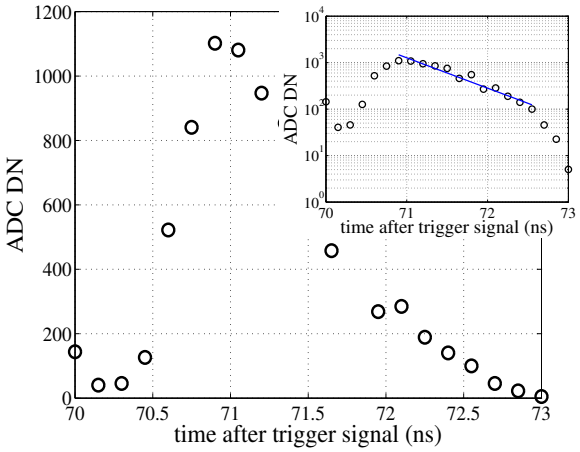

Fig. 7. Measured laser impulse response

Time-domain characterization. Fig. 7 is the measured impulse response of the system. A PiLas Picosecond Laser Diode System with a 406-nm laser diode head is used for this measurement. This gain-switched laser is capable of producing laser pulses of less than 50-ps FWHM with peak collimated beam power as high as 1500 mW. The laser pulses are synchronized with the chip and the entire laser pulse is integrated. Results come from averaging a 4096-point measurement dataset and numerically differentiating the resulting imager response. The data show that the system has an impulse response with a time constant of less than 800 ps. With fingered PD periodicity of 4.8 μm, this result agrees with that found in Reference [9] for similar finger periodicity. Preliminary results with quantum dots with lifetimes exceeding 15 ns

show that time-gating of the differential PDs provides 82 dB of background rejection without external optical filtering.

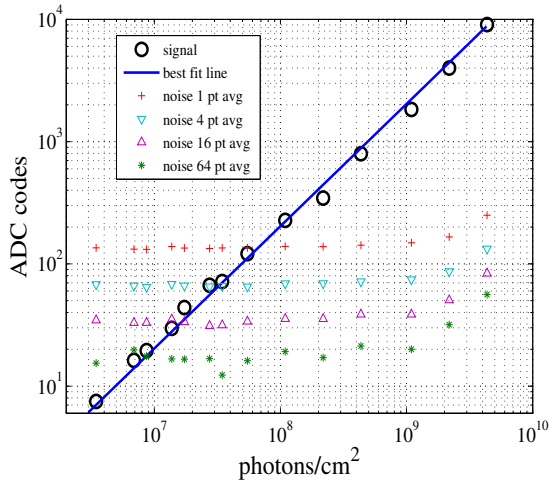

Fig. 8. Measured biosensor system sensitivity

Sensitivity characterization. Simulation predicts a reset-noise-limited sensitivity of approximately 180 e^- without any averaging, which corresponds to a photon density of 6.2 x 10^7 photons/cm^2, assuming 50% quantum efficiency. N-point averaging improves this sensitivity by a factor of $\sqrt{N}$; for N=4096, sensitivity should improve to less than 1 x 10^6 photon/cm^2. For typical ADC settings, the gain of the entire sensor path is approximately 2 x 10^{-6} DN/(photons/cm^2). Fig. 8 shows measured sensor sensitivity slightly higher than those predicted by simulation. Measurements are taken by integrating the entire laser pulse by positioning the reset before the pulse. Neutral density filters are added to vary the intensity of the laser signal. Each point is the average of 4096 measurements and the noise floors are calculated as standard deviations of different datasets, each defined as a set of averages of

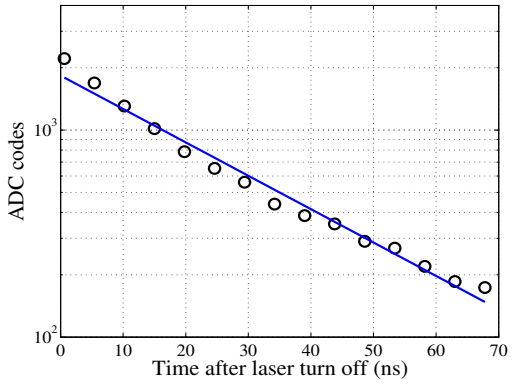

Fig. 9. Measured time-resolved response of Qdot signal after time gating

N consecutive measurements. Measured noise floors determine a sensitivity of approximately 8 x 10^6 photons/cm^2 (SNR of 0 dB) after 64-point averaging.

Flourescence lifetime measurement. Fig. 9 shows a measured lifetime response at a single pixel for Qdot-655 (Molecular Probes). This is imaged from stretavidin-labelled Qdots immobilized directly on the chip surface. The laser delivers approximately 8 x 10^7 photons/pulse. The peak emitted photon density is approximately 8 x 10^8 photons/cm^2. Assuming a molar extinction coefficient of 5.7 x 10^6 M^{-1} cm^{-1} (corresponding to a cross section of 2.18 x 10^{14} cm^2/molecule) and a quantum yield of 80%, we predict a Qdot surface concentration of approximately 5 x 10^{10} cm^{-2}. This result shows measured quantum dots exhibit a lifetime of approximately 19.6 ns.

Table 1 Chip specifications and summary

Item	Value
Technology	0.18-µm Mixed-Signal CMOS
Die Size	5 mm x 5 mm
Array Size	64 x 64
Pixel Size	50 µm x 50 µm
Active Photodiode Area	576 µm^2
Quantum Efficiency (at 650 nm)	0.45
Time resolving resolution	150 ps
Signal path gain	2 x 10^{-6} DN/(photon/cm^2)
Sensitivity (after 64-point averaging)	8 x 10^6 photons/cm^2
ADC conversion time (single-ramp)	102 µs
ADC conversion time (dual-ramp)	1.6 µs
Linearity (single/dual-ramp)	13 bits / 12 bits
Frame Rate (dual-ramp)	310 frames/s

With ADC counters running on a 80-MHz clock, 102 µs is required for single-ramp, single-slope conversion at 13-bit resolution. A single frame with 64-points averaging requires 0.21 s to complete. With the dual-ramp enabled, the frame rate increases to 310 frames/s (single sample point).

4. Conclusions and future work

In this work we have demonstrated a 64-by-64 CMOS fluorescence sensor with sensitivity of better than 10^7 photons/cm^2 after 64-point averaging. Impulse-response time constants of better than 800 ps with frame rates better than 150 frames/s are achieved for two sample points per exponential. Future work will further explore applications in both active CMOS microarrays and lifetime microscopy.

ACKNOWLEDGMENT

We gratefully acknowledge United Microelectronics Corporation (UMC) for chip fabrication.

REFERENCES

[1] M. Schena, et al, *Trends Biotechnology* **16**, pp. 301-306, Jul. 1998.
[2] Y. Chen, et al, *Differentiation* **71**, pp. 528-541, 2003.
[3] P. I. H. Bastiaens and A. Squire, *Trends in Cell Biology* **9**, pp. 48-52, Feb. 1999.
[4] X. Michalet, et al, *Science* **307**, pp. 825-827, June, 2001.
[5] R. A. Yotter and D. M Wilson, *IEEE Sensors J.* **3**, pp. 288-303, Jun, 2003.
[6] W. Becker, et al, *Microscopy Research and Techniques* **63**, pp. 58-66, 2004.
[7] A. V. Agronskaia, et al, *Journal of Biomedical Optics* **9**, pp. 1230-1237, Nov/Dec, 2004.
[8] G. Patounakis, K. L. Shepard, and R. Levicky, IEEE JSSC 41, pp. 2521-2530, Nov. 2006.
[9] J. Genoe, et al, *IEEE Trans. Elect. Dev.* **48,** pp 1892-1902, Sept. 2001.
[10] H. Eltoukhy, K. Salama, and A. El Gamal, *IEEE JSSC* **41**, No. 3, pp651-662, Mar. 2006.
[11] B. Fowler, M. Godfrey, J. Balicki, and J. Canfield, *Proc. SPIE* **3965**, pp. 126-135, May 2000.