Active CMOS Biochip for Time-Resolved Fluorescence Detection

George Patounakis[1], K. L. Shepard[1], and R. Levicky[2]

Columbia University, [1]Department of Electrical Engineering, [2]Department of Chemical Engineering
New York, New York USA

Abstract

This paper describes the design of an active CMOS biosensor array for fluorescence-based assays which enables time-gated, time-resolved fluorescence spectroscopy. The array is sensitive to photon densities as low as $1.15 \times 10^8/cm^2$, has a dynamic range of over 74 dB, and has sub-nanosecond timing resolution. High sensitivity is achieved through subsampling and averaging.

Introduction

Surface-based sensing assays have become commonplace for both environmental and biomedical diagnostics. Analyte "targets" from solution bind to appropriate "probe" molecules immobilized on a solid support, generally a passive glass substrate, as a result of strong and specific probe-target interactions. Multiplexed detection occurs by constructing an array of different probes on the surface. For DNA analysis, high density microarrays examine gene expression at the scale of entire genomes by simultaneously assaying target mixtures derived from expressed mRNA against thousands of array sites (locations), each bearing probes for a specific gene. Microarrays generally quantify target concentrations in relative terms; e.g. in the form of a ratio to hybridization signal obtained using a "reference" target sample. Other biosensing applications are calibrated to provide absolute target concentrations.

Fluorescent-based detection, in which the target is labeled with a fluorophore, has arguably become the standard for quantifying the extents of probe-target binding in surface-based sensing assays [1]. Traditional microarray scanners consist of an excitation source (usually a laser) with the fluorescent light emitted by bound target fluorophores focused and collected, through a generally lossy optical path, onto a cooled charge-coupled device (CCD) or photomultiplier tube (PMT). Optical filtering is used to improve the SNR by removing "background" light or reflected excitation light. Generally, arrays must be sensitive to fluorophore concentrations from 10^8 to 10^{11} cm^{-2}.

Fluorophores have associated with them a characteristic lifetime, which defines the exponential fluorescent decay transient after the removal of the excitation source. These lifetimes, on the order of nanoseconds, are characteristic of the dye and its environment and can be used in addition to color and intensity for multiplexed detection. Fluorescent lifetime detection, for example, has been employed for capillary electrophoresis in both the time and frequency domain [2].

In this paper, we describe the design and measurement of an active CMOS biosensor array for fluorescent-based assays which enables time-gated, time-resolved fluorescence spectroscopy without the need for optical filters and is sensitive to photon densities as low as $1.15 \times 10^8 photons/cm^2$ with a conventional nwell/p-sub photodiode. The system is designed for probe immobilization *directly* on the chip surface, eliminating losses in large, complex optics and allowing for efficient solid-angle collection. Time-gated operation allows for effective elimination of background excitation light. In addition, the ability to distinguish fluorophore lifetime offers the potential to detect the presence of two different fluorophores without the need for multiple optical filters.

This work was supported in part by NSF Grant BES-04-028544 and by a fellowship from the Intel Foundation.

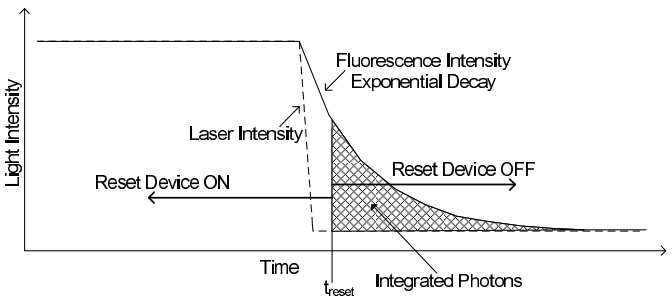

Fig. 1. Time-resolved, time-gated fluorescence detection by integrating photocurrent response.

Time-resolved fluorescence detection in CMOS

Most time-resolved fluorescence systems rely on real-time photodetection with a PMT, which provide high gain and high light sensitivity [2]. Photodiodes, which are the photosensitive devices compatible with a CMOS process, do not have gain and require averaging (in the form of integrating photocurrent onto a capacitor and averaging the results of multiple measurements) to achieve high signal-to-noise performance. Recent work has demonstrated the high sensitivity that photodiode-based CMOS imagers can achieve when long integration times are combined with averaging [3].

In our system, we seek to achieve high sensitivity in what is tantamount to a real-time detection application in which we must extract the *transient* fluorescent decay response, as shown in Fig. 1, that follows the rapid turn-off of a laser excitation. To preserve the sensitivity benefits of averaging and to reduce the bandwidth requirements on circuit components, we turn to subsampling to achieve this real-time detection. The transient response is repeated; each time the integral of the photodiode current ($i_{photo}(t)$) is taken from a different starting time t_{reset} relative to the laser turn-off time, as shown in Fig. 1, yielding output $\int_{t_{reset}}^{\infty} i_{photo}(t)dt$. The result for a single t_{reset} can also be repeated to improve the overall detection sensitivity. The $i_{photo}(t)$ transient, which is directly proportional to the instantaneous fluorescence, is generated by numerical differentiation.

System description

The die photo of the $5mm \times 5mm$ sensor chip as fabricated in a standard mixed-signal $0.25\mu m$ process is shown in Fig. 2. The 8×4 pixel array is time-multiplexed into four current-mode $\Sigma\Delta$ analog-to-digital converters. Digital results are stored in an on-chip SRAM. The arrival time of the reset signal to the pixels (t_{reset}) is programmably skewed from the timing of the on-chip laser drivers. An on-chip digital controller, configured externally with a serial bit stream, generates the clocks and control signals for the ADC, steps through the appropriate t_{reset} values, controls

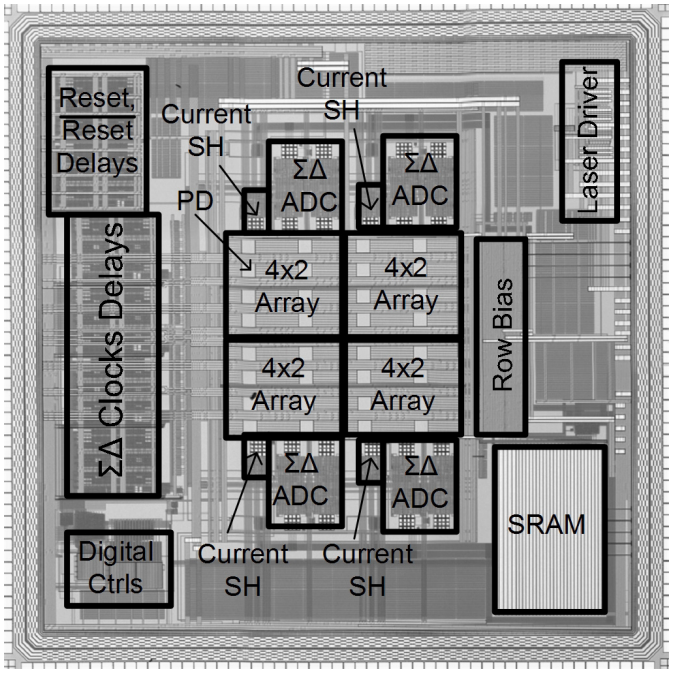

Fig. 2. Die photograph of CMOS sensor chip.

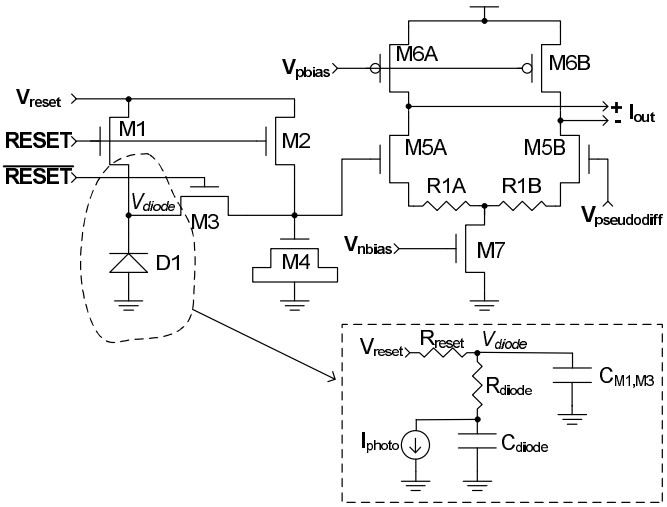

Fig. 3. Pixel schematic. Inset: Equivalent circuit for pixel front-end during reset phase.

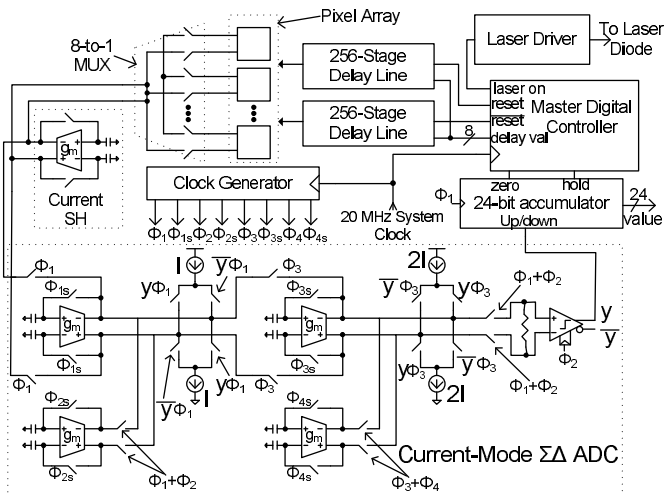

Fig. 4. Simplified top-level system schematic.

the storage of digital samples, and determines the laser pulse duration.

A. Pixel array

The pixel array is divided into four banks of eight pixels with a ΣΔ ADC for each bank. Each pixel, as shown in Fig. 3, contains two reset transistors ($M1$ and $M2$), an isolation device ($M3$), a storage capacitor ($M4$), a transconductor ($M5$-$M7$), and an nwell/p-sub photodiode ($D1$). Each photodiode contains an nwell guard ring to collect carriers generated by neighboring pixels. $R1A$ and $R1B$ are non-silicided polysilicon resistors which linearize the transconductor through source degeneration. The transconductor converts the voltage across the storage capacitor, which results from the integrated photocurrent, into a differential current for subsequent current-mode data conversion. The measured transconductance of these elements is 0.22 mS. Large input nFETs ($4.32mm/1\mu m$) are used for the transconductor to reduce 1/f noise and improve matching performance.

As shown in Fig. 4, eight differential pixel signal currents are time-multiplexed onto a single current-mode sample-and-hold (SH) element consisting of a differential transconductor with two feedback storage capacitors [4]. The output of the current-mode SH element is continuously sampled by the ΣΔ ADC. Using a sampled version of the pixel current rather than feeding the pixel current directly into the ADC reduces charge-injection and clock-feedthrough noise coupling back into the pixel array through the multiplexer.

During the reset phase, as determined by a high value on the $RESET$ signal, transistor $M3$ is off, isolating $M4$ from $D1$, reducing the capacitance on node V_{diode} to the reverse-biased capacitance of $D1$ ($C_{diode} \cong 0.9\ pF$) and the capacitances of $M1$ and $M3$ ($C_{M1,M3} \cong 10pF$). The $0.5\mu m$ $M1$ reset device is $3mm$ wide to provide a triode region resistance of $R_{reset} \cong 1.8\Omega$, which allows V_{diode} to be held within 20 mV of V_{reset} even for photodiode currents as large as 10 mA, which is the photocurrent associated with a 500 W/m^2 635-nm laser. The isolate transistor $M3$, also $0.5\mu m$ long, is $1.5mm$ wide and acts to mitigate some of the voltage offset associated with charge-injection from $M1$. The inset of Fig. 3 shows the equivalent circuit of the pixel in the reset phase. $R_{diode} \cong 1k\Omega$ is the parasitic resistance associated with the nwell bulk connection to the diode and is unfortunately larger than desired because of a design oversight. The large value of R_{diode} limits the maximum sustainable photocurrent to about 1 mA before blooming would occur in the diode. There are two time constants associated with this circuit, $\tau_{diode} = (R_{diode}+R_{reset})C_{diode} \cong 900ps$ and $\tau_{M1,M3} = R_{reset}C_{M1,M3} \cong 20ps$. The bandwidth-critical response of the pixel is determined by how quickly the internal diode voltage across C_{diode} can track the external diode voltage V_{diode}. The laser diode pulse fall-time should be greater than both of these time constants for the pixel to track the photocurrent up to t_{reset}. Transistor $M3$ acts to provide a larger capacitance for charge integration (50 pF when reset is low) while removing the bulk of this capacitance (that of transistor $M4$) from the performance-limiting time constants.

B. Current-mode analog-to-digital conversion

The current-mode ΣΔ ADC, as shown in Fig. 4, is a fully-differential, second-order, one-bit system [5] with a full-scale input level of $\pm 10\mu A$. The differential one-bit current-output DACs are comprised of two cascode current sources and a switch network as shown in Fig. 4. Pattern-dependent supply loading is mitigated with the current-switch design because there is always a fixed current ($I = 10\mu A$) flowing through each DAC. The design uses four non-overlapping clocks and operates with a $8\mu s$ period

while achieving over 12 bits of settling accuracy in the discrete-time current-copier integrators.

In a manner similar to the pixel transconductors, the transconductors in the ADC, as well as the transconductor in the SH, also use source-degenerating polysilicon resistors. These transconductors have a nominal transconductance of 0.1 *mS*. The transconductors in the ΣΔ ADC are further enhanced with active cascode topologies in the output stage to boost output resistance, minimizing gain error from current division.

The one-bit ΣΔ output is used as an "up" or "down" signal for a 24-bit accumulator, a simple low-pass digital filter. The 12-bit value generated by the accumulator after running the ΣΔ for 4096 cycles has a relative accuracy of approximately 11-bits, limited by idle tones in the ΣΔ. The measured detrimental effect of idle tones is less than what behavioral modeling of the ADC predicts because of the dithering effect of noise at the input of the ADC from the current-copier SH and other analog noise sources in the ΣΔ loop [6].

Accumulator results are cached into on-chip SRAMs, which eliminates the need for firing noisy off-chip drivers during repeated measurements. The outputs of the four accumulators are fed to the input of the SRAM controller that coordinates writing these to a single 2048×24 array. The address space of the SRAM is organized by ΣΔ sub-blocks and by which pixel within the sub-block is being written. The SRAM can be written in a "single-pixel" mode for a maximum of 2048 24-bit pixel values or in a "multiple-pixel" mode in which 64 values for each of the 32 pixels are stored. When measurements are complete, the entire contents of the SRAM can be loaded off-chip in under 310 μs.

A single measurement takes 4097 ADC cycles to complete; one additional cycle is consumed to allow the SH to settle. This means that one measurement on each of four pixels can be repeated every 33 ms.

C. Reset and laser-diode-driver control

The chip must drive both the array reset signal and the off-chip laser with the ability to vary the skew between these signals to achieve the time-resolved fluorescence detection shown in Fig. 1.

The laser driver consists of a variable width inverter with independent tunability of the pull-up and pull-down widths, selected digitally with two seven-bit control words. To accommodate laser diodes with operating voltages greater than 2.5V, thick oxide 3.3V I/O are used in the laser driver output circuitry. This also allows the diode to tolerate overshoot at the near-end, which sometimes occurs as a result of reflections against the highly nonlinear load resistance turn-on characteristic of the laser diode. The maximum current sourcing capability at 2.7V output is greater than 130 mA, which is sufficient to drive commercial laser diodes with 50 mW of optical output. Larger laser diodes with input capacitances of up to 40 pF can be driven by up to four off-chip 50-Ω transmission lines in parallel. Near-end fall times of about 500 ps are achieved. Pulse width and synchronization are determined by the master digital controller. Pulse width can be varied in 50 ns intervals up to 204.8 μs, but typically a pulse length of 300 ns is used.

A programmable, variable delay line is used to trigger the pixel reset predrivers. This delay can be multiples of $T_{cycle} = 50ns$ (the period of the 20 MHz system clock) combined with sub-clock-period delay generation with a simple 256-stage inverter chain delay line (with stage delay, $T_{delay} \cong 230ps$) as shown in Fig. 4: $t_{reset} = nT_{cycle} + mT_{delay}$, m and n positive integers. An eight-bit multiplexer is used to choose one of the phases of the delay line. The delay line and multiplexer are carefully laid out to limit mismatch between buffer stages due to layout parasitics.

Because large on-chip drivers for the reset and laser diode drivers are rapidly switching to achieve sufficient resolution for time-resolved detection, power-supply and substrate noise issues are a concern for the sensitive analog circuits of the array and ADC. The slew rate of the reset signal is limited to slightly less than 1 ns to control noise generation. The pixel array and ΣΔ ADC are isolated by a double guard ring. Supplies are carefully separated and decoupled on chip. All bias currents and voltages are fed in from one side of the chip while all digital signals interface from the other side.

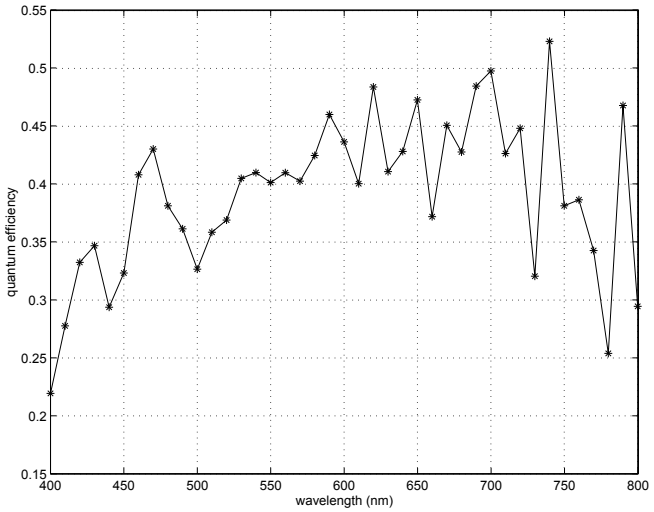

Fig. 5. Quantum efficiency of nwell/p-sub photodiode sensor.

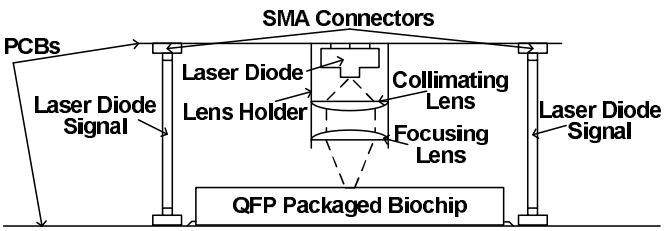

Fig. 6. Packaging of sensor chip with laser and simple focusing optics.

Results

We present experimental results quantifying the quantum efficiency, signal-to-noise (SNR) performance, dynamic range (DR), and time-domain response of the sensor chip. A summary of the measured performance of the chip is presented in Table I.

Quantum efficiency. Fig. 5 shows the measured external quantum efficiency of the nwell/p-sub photodiode using a fairly standard measurement set-up including a monochromator, integrating sphere, and calibrated photodetector. Peak quantum efficiencies in the range of $0.45 - 0.5\ e^{-}/ph$ occur at wavelengths between 600-700 nm for the relatively deep diode junctions, suggesting peak performance for dyes operating in this wavelength range, such as AlexaFluor 633 (with peak absorption at 630 nm). The structure in the curve, particularly evident at the longer wavelengths, is due to interference effects in the dielectric stack.

SNR and DR characterization Both the SNR/DR and time-domain characterization are performed with the sensor mounted with the laser diode and focusing optics as shown in Fig. 6. To keep the optics simple, the elliptical shape of the laser beam is not altered. The laser diode used for these measurements is a 635-nm, 5-mW AlGaInP diode packaged in a 9mm CAN style package. Mounted on a PCB that is positioned over the PCB containing the sensor chip, up to four cables connect the laser to the laser diode drivers on the sensor chip. The biochip sensor is packaged in a ceramic quad-flat-pack package.

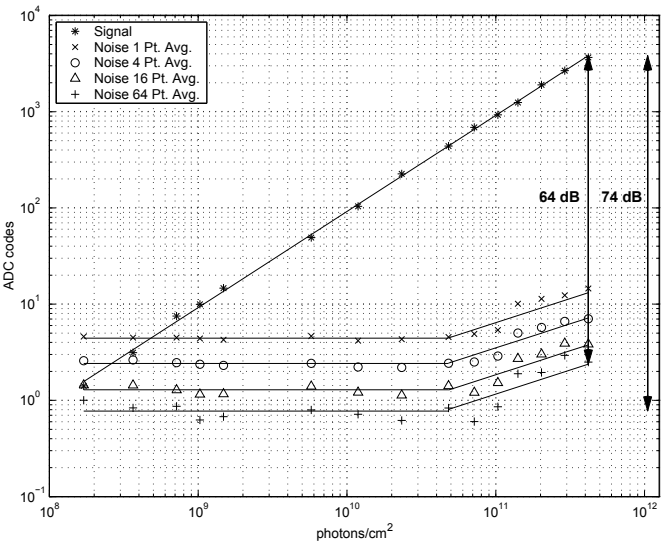

Fig. 7. Complete sensor system sensitivity.

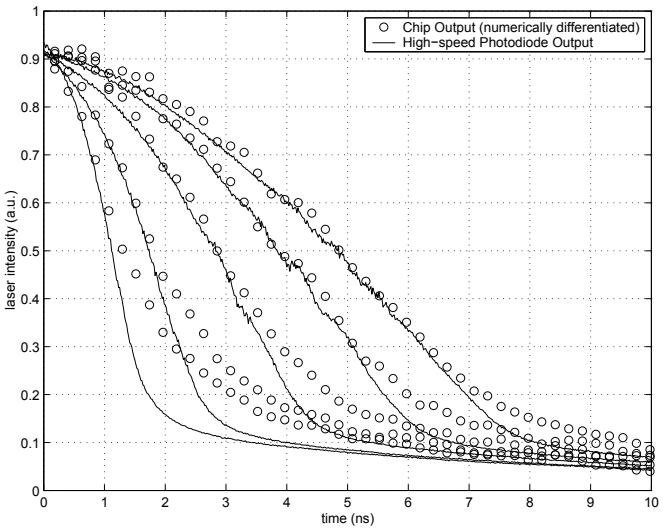

Fig. 8. Time-resolved detection of laser turn-off edge.

Fig. 7 characterizes the sensor sensitivity. In this case, t_{reset} is chosen to position the end of the reset phase 8.5 ns *before* the laser diode turn-off, allowing laser diode power to be directly integrated in the sensor. Neutral density filters are used to vary the laser signal strength. While the neutral density filters have transmission tolerances of only $\pm 10\%$, the transmission of individual neutral density filters is calibrated to remove this source of possible measurement error. Each point in the signal curve is the average of 2047 measurements. The signal curve shows good linearity over three orders of magnitude. The integral linearity is limited principally by the in-pixel transconductor to about 9 bits. This can be calibrated to further improve linearity, leaving the system to perform at the limit of the ADC (11 bits).

The noise curves of Fig. 7 are generated by taking the average of blocks of N consecutive measurements in the original 2047 measurement dataset. The standard deviation of the $floor(\frac{2047}{N})$ averaged blocks of data is recorded as the noise after averaging. The noise curves show the effect of averaging on reducing the noise floor of the system. Averaging 64 points yields a noise floor close to the quantization noise limit of the system. The system shows a peak SNR of 64 dB and dynamic range of 74 dB.

The dark current in the photodiode is approximately 2.31 pA, which corresponds to a dark current signal from the pixel transconductor of $10.4\mu A$ per second of integration. The corresponding dark current shot noise level is approximately 10^{-3} digital numbers (DN) for 12 bit conversion, well below the quantization noise limit.

Time-domain characterization. Fig. 8 shows the time-domain response of the sensor in which t_{reset} is varied to reconstruct the actual turn-off transient of the laser. The solid lines show the output from a 12 GHz-bandwidth, New Focus 1577-A photodetector as measured by a Agilent 86100B oscilloscope. The circles show the numerically differentiated output of the sensor. The laser fall-time is varied from approximately 1 ns to 5 ns. For fastest fall times, we find that the sensor output lags the high-speed detector output, due to the limitations of the τ_{diode} time constant described above.

Fluorescence lifetime measurements. These are on-going with fluorescently tagged DNA and DNA probes attached directly to the sensor surface at each pixel site. Results will be reported in a biology-related journal.

TABLE I
CHIP SPECIFICATIONS AND MEASURED PERFORMANCE

Item	Value
Chip	
Technology	TSMC $0.25\mu m$ Mixed-Signal CMOS
Die Size	5mm x 5mm
Clock Speed	20 MHz
SRAM Size	2048x24-bits
Sensitivity	$1.15 \times 10^8 photons/cm^2$
Linearity	9 bits (w/o calibration)
Pixel array	
Array size	8×4
Quantum efficiency (at 635 nm)	0.45
Pixel size	$160 \times 215\ \mu m^2$
Photodiode size	$100 \times 100\ \mu m^2$
Transconductor gain	0.22 mS
Dark signal	$10.4\mu A/s$
ADC	
Architecture	Current Mode $\Sigma\Delta$
Order	2
DAC	1-bit Differential
Cycle Time	$8\mu s$ (>12-bit settling)
Cycles per Sample	4096 (adjustable)
Input full scale	$\pm 10\mu A$

Conclusions

In conclusion, we have demonstrated a CMOS fluorescence array sensor with sensitivities of almost $10^8 photons/cm^2$, a dynamic range of 74 dB, and subnanosecond timing resolution. This substrate should enable powerful, low-cost devices for surface-based biomolecular assays.

References

[1] J. G. Hacia and F. S. Collins, J. Med. Genetics **36**, pp. 730-736, 1999.
[2] S. L. McIntosh, et al., Anal. Chem. **72**, pp. 5444-5449, 2000.
[3] H. Eltoukhy, et al., ISSCC 2004, pp. 222-223.
[4] S. J. Daubert, D. Vallancourt,Y. P. Tsividis, Electronics Letters **24**, pp. 1560-1562, Oct. 1988.
[5] S. J. Daubert, D. Vallancourt, IEEE JSSC **27**, pp. 821-830, May 1992.
[6] D. Hyun, G. Fischer, IEEE Trans. on CAS-I **49** pp. 646-656, May 2002.