This is an Accepted Manuscript, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this Accepted Manuscript with the edited and formatted Advance Article as soon as it is available.

You can find more information about Accepted Manuscripts in the Information for Authors.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal’s standard Terms & Conditions and the Ethical guidelines still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this Accepted Manuscript or any consequences arising from the use of any information it contains.
An integrated CMOS quantitative-polymerase-chain-reaction lab-on-chip for point-of-care diagnostics

Haig Norian, Ryan M. Field, Ioannis Kymissis, and Ken Shepard

Considerable effort has recently been directed toward the miniaturization of quantitative-polymerase-chain-reaction (qPCR) instrumentation in an effort to reduce both cost and form factor for point-of-care applications. Considerable gains have been made in shrinking the required volumes of PCR reagents, but resultant prototypes retain their bench-top form factor either due to heavy heating plates or cumbersome optical sensing instrumentation. In this paper, we describe the use of complementary-metal-oxide-semiconductor (CMOS) integrated circuit (IC) technology to produce a fully integrated qPCR lab-on-chip. Exploiting a 0.35-µm high-voltage CMOS process, the IC contains all of the key components for performing qPCR. Integrated resistive heaters and temperature sensors regulate the surface temperature of the chip to an accuracy of 0.45°C. Electrowetting-on-dielectric microfluidics are actively driven from the chip surface, allowing for droplet generation and transport down to volumes less than 1.2 nanoliter. Integrated single-photon avalanche diodes (SPADs) are used for fluorescent monitoring of the reaction, allowing for the quantification of target DNA with more than four-orders-of-magnitude of dynamic range and sensitivities down to a single copy per droplet. Using this device, reliable and sensitive real-time proof-of-concept detection of Staphylococcus aureus (S. aureus) is demonstrated.

Introduction

Quantitative polymerase chain reaction (qPCR) has become one of the most widely used methods for the quantification and detection of both DNA and messenger RNA (after reverse transcription) due to its accuracy, sensitivity and large dynamic range. As an example, infectious disease recognition through qPCR relies upon the selective amplification and subsequent detection of a DNA sequence that has been shown to be unique to a particular pathogen. The concentration of DNA present in a test volume after each iteration of a temperature cycling protocol is monitored in real time. As a result, the initial copy number of the target can be determined given the controlled, known amplification at each step. Commercial qPCR systems, which generally operate on standard volumes of 25 or 50 µL, require a heating block weighing anywhere from 2 to 10 kg to ensure a uniform temperature profile across a multiwell plate, leading to (at best) a bench-top form factor. While previous work has brought the volume of a single qPCR reaction down to volumes as low as 10pL, easing some of the temperature uniformity requirements, these systems have either relied on external optics for fluorescence detection or lacked integrated droplet formation. In the first case, the sensing of fluorescent dyes requires bulky and delicate optics to obtain a reliable measurement, and in the second case, any integrated microfluidics have to rely on pneumatic control, regulated by external pumps. Overall form factor remains that of a stationary, bench-top device.
Complementary metal-oxide-semiconductor (CMOS) integrated circuit (IC) technology provides a unique opportunity to further scale PCR hardware by providing for high-bandwidth, distributed temperature control across the chip surfaces and integrated detection. Tomazou et al. describe the first CMOS-based qPCR IC. Instead of using fluorescence detection, pH is monitored during PCR to gauge the progression of the reaction with heating and temperature control also performed on-chip. This device is still limited by use of microfluidic pumps, requiring reagent volume on the microliter scale. Sensitivity is limited by noise in the integrated ion-sensitive field-effect transistors to 10 copies per sample.

Electrowetting-on-dielectric (EWoD) droplet transport provides a means for scaling microfluidics by controlling droplets electrically without the need for external pumps. In the presence of an electric field, a droplet of aqueous reagent seeks to minimize its internal energy through translational motion toward the direction of positive voltage. Nearly all EWoD devices are passive; that is, they have no integrated means to control or generate controlling voltages. EWoD devices employed for qPCR applications have been limited to such passive devices with external fluorescence detection. The one active CMOS EWoD device that has been reported was applied to the movement of cells through microfluidic chamber at speeds up to 30 µm sec−1 with off-chip detection. The required voltage for EWoD transport (70-250 V, depending on the size of the droplet) is also generated off-chip in this design.

In the work described here, we use CMOS technology for the first time to perform all the required functions for qPCR -- temperature control, heating, microfluidics, and fluorescence detection. Manual pipetting has been replaced with EWoD-based droplet transport. Thermal cycling traditionally performed using an aluminum heating block and a thermocouple is instead performed using micron scale polysilicon heaters and resistive aluminum temperature sensors. The fluorescent detection necessary for determining the progression of the PCR reaction is performed using integrated Geiger-mode single-photon avalanche diodes (SPADs). All of the high voltage generation required to perform EWoD droplet control is integrated on-chip as well with droplet sizes on the order of a nanoliter (droplet diameters of approximately 200 µm). The chip currently supports three parallel measurement channels but future chip iterations aim to place a SPAD under every pixel. In addition, the ability to move droplets around allows 40 PCR reactions to be occurring in parallel by multiplexing the on-chip detectors. Since each electrowetting pixel is independent and isolated from its neighbours, each droplet can contain a unique combination of primers drawn from two fluid reservoirs. Sensitivity and dynamic range allow for the quantification of target DNA from 1 to 104 copies per 1.2-nanoliter droplet. Using this device, reliable and sensitive proof-of-concept detection of Staphylococcus aureus (S. aureus) is demonstrated.
2. Experimental

qPCR Chip overview

The qPCR chip is implemented in a 0.35-μm high-voltage CMOS process (austriamicrosystems, Unterpremstaetten, Austria). The 4-mm-by-4-mm chip, shown in Fig. 1, contains a 7-by-8 array of 200-μm-by-200-μm electrodes to enable electrowetting-on-dielectric (EWoD) droplet transport. The chip contains a total of 128 bondpads that are wirebond connected to a 272-pin ball-grid-array (BGA) package. A custom-developed SU-8 encapsulation scheme is employed to simultaneously encapsulate the wirebonds while patterning the active electrowetting array. The droplet array is covered with an indium-tin-oxide (ITO)-coated polyethylene naphthalate (PEN) coverslip mounted 100μm from the chip surface, supported by the SU-8 encapsulant.

To perform qPCR on the surface of the chip, primers are placed in one or two reservoirs, DNA target is placed into another reservoir, and the PCR reagents, including dNTPs, DNA polymerase and intercalator dye, are placed into the remaining reservoir. Droplets are electrostatically drawn from each of these constituent reservoirs and mixed together. The entire surface of the chip is uniformly thermally cycled according to the prescribed heating profile. After each qPCR cycle, each test droplet is brought over the fluorescence measurement region of the chip in order to determine the procession of the reaction. SPAD diode fluorescence readings are taken at the end of each elongation stage.

Electrowetting Droplet Transport

The pixel dimension of 200μm allows for a droplet size of approximately 1.2 nanoliter to be supported. The EWoD control circuitry is shown in Fig. 2. Each electrode is controlled to have a voltage of either $V_{DDHV}=90\,V$ or $0\,V$ relative to the voltage on the ITO-coated PEN coverslip. In each pixel, a control bit is latched to indicate this state, and this latched value is level-shifted from the 3.3-V core voltage level to $V_{DDHV}$ at the electrode. The pixel control bits are loaded in parallel from eight data lines on a row basis (determined from a decoded three-bit row address). The bits associated with Row 8 determine the potential for the four reservoirs and the electrodes connecting these to the 7-by-8 array (see Fig. 1). A row can be updated every 10 msec.

An on-chip charge pump steps up the 3.3-V core logic supply voltage to generate the 90-V $V_{DDHV}$ supply. Typically, each stage of a typical Dickson charge pump configuration consists of a diode-connected transistor and a capacitor. To improve efficiency, a three-transistor-based stage is used in our implementation (see Fig. 2). Transistors $M_1$, $M_2$, $M_3$ are high voltage devices capable of supporting drain-body and drain-source voltages of up to 120V. $M_{C3}$ turns on when the diode-connected transistor $M_{C3}$ is off, establishing zero bias across the drain-body diode of $M_{C3}$. By preventing the flow of current...
across this diode junction during operation, efficiency is increased. Transistor $M_{C1}$ turns on when $M_{C3}$ is on, eliminating body-effect enhancement of the threshold voltage of $M_{C3}$, reducing the drain-to-source voltage drop across this device.

The level shifter in each pixel makes use of high voltage transistors $M_{L1}$, $M_{L2}$, $M_{L3}$, $M_{R1}$, $M_{R2}$, and $M_{R3}$ that can tolerate drain-bulk voltages of 120V, allowing for the safe use of a $V_{DD}$ of 90 V. Furthermore, the use of cascode transistors $M_{L1}$ and $M_{R1}$ prevents the gate voltages of $M_{L3}$ and $M_{R3}$ from exceeding 25V.

**SPADs in CMOS**

The CMOS-integrated SPAD design here follows one previously used for a SPAD-based imager. This SPAD is located until Pixel (7, 3) in Fig. 1. An area of 100µm² is removed from the electrode area for this pixel to form a window for light to pass with no noticeable effects on EWoD transport. The SPAD’s shallow p+ region has a diameter of 7.5 µm in an n-well of diameter 15 µm, as shown in the inset of Fig. 3.

![SPAD avalanche events over an integration window of 1 second as a function of the delay between the laser turn-off and the on-gating of the SPAD. The curve saturates at short delays as the excitation source is collected. The curve saturates at long delays due to SPAD dark count. Inset: Basic geometry of the SPAD.](image1)

The diode is biased beyond its reverse breakdown voltage (approximately $V_{breakdown} = 20$V) by an overvoltage, $V_{ov}$. With careful engineering of the implants to avoid deep-level traps, the device does not draw any current until a photon triggers an avalanche. Geiger mode of operation of the SPAD requires a circuit which reduces the reverse bias across the diode to below $V_{breakdown}$ after triggering, quenching the avalanche. The simplest quenching circuit is a passive resistor, used here. When an avalanche is triggered, a current flows through the resistor. The resulting voltage drop across the resistor reduces the reverse bias across the SPAD below $V_{breakdown}$. The associated RC time constant to return to a reverse bias of $(V_{breakdown} + V_{OV})$ defines the dead-time for the SPAD. In our device, a quenching resistance of 67 kΩ is used, tuned to avalanche current levels of 10.3 µA, yielding a dead-time of 0.5 µs with $V_{ov}$ of 3.5 V. We observe no significant after-pulsing (events triggered by charges that do not clear the multiplication region before the SPAD is reset).

For on-chip fluorescence detection, the excitation source must be filtered or separated from the emitted fluorescence. This can be done spectrally using optical filters, but the required optical density (OD) for rejecting the excitation source is usually four or more, not achievable without a very thick integrated thin-film interference filter. As an alternative, a time-resolved approach can be used in which the excitation source is pulsed and the emitted photons collected after the excitation has been removed. Our current instrumentation uses a Fianium Supercontinuum laser (Fianium, Southampton, United Kingdom) with an acousto-optical tunable filter (AOTF) at 500 nm. The 1-MHz trigger signal from the laser drives a pulse generator to produce a synchronized gate signal for the SPAD. The average number of photons collected in an integration window after the excitation signal has been shut off determines the measured intensity. Timing of the gate signal is important to ensure that excitation light is not collected by the SPAD. As shown in Fig. 3, enabling the gate of the SPAD more than 100 psec after the laser signal gives an effective four-orders-of-magnitude rejection of the excitation light. Future revisions can replace the Fianium laser with a simpler gain-switched laser diode.

**Integrated heater design**

Temperature control is maintained on the surface of the chip with on-chip temperature sensors and heaters. Current is passed through high-resistance polysilicon resistors to heat the chip. Resistive temperature sensors consisting of serpentine interconnect metal are calibrated to monitor the temperature.

Three polysilicon heaters, separated by 400µm and located as shown in Fig. 1, provide even heating of the chip surface, allowing variation of less than ±1.5ºC within 3 seconds of turning on. Each heater consists of a serpentine trace 5300-µm long and 20-µm wide with a nominal effective resistance of 270kΩ. In contrast, the temperature sensors are approximately 80-mm long and 0.67-µm wide with a room temperature resistance of 4.3kΩ and a temperature coefficient of 3.3x10⁻³. The sensors are located 4 µm below the reagent-containing droplet and 6µm above the heaters as shown in Fig. 4. Temperature calibration is performed using the techniques described in Section 3.

![Cross section showing position of heaters and sensors below the pixels in Columns 1, 4, and 7.](image2)

---

**References:**

1. J. Name, 2012, 00, 1-3
2. Lab on a Chip
3. This journal is © The Royal Society of Chemistry 2012
Device Packaging

The chip is mounted on a BGA package. SU-8 2100 (Microchem, Newton, Massachusetts, USA) is used to encapsulate the wirebonds and define the active microfluidic area, as shown in Fig. 6. Following similar work encapsulating ICs onto printed-circuit boards, our method of encapsulation is generalizable to any package geometry using SU-8 as thick as 1 mm. In addition, the encapsulation creates a ledge to support the overlying ITO-coated PEN coverslip and hold this 100 µm from the chip surface as shown in Figs. 5a and 5c. Special care is taken during the SU-8 process to ensure suitable planarity to allow tilt-free placement of the coverslip.

The BGA package is heated to 70°C in order to decrease SU-8 viscosity upon contact with the heated substrate. Using a PEN reservoir, SU-8 is deposited to a depth of 800 µm. Cast SU-8 is not fully baked prior to photolithographic exposure to allow a degree of pliability. Photolithography is used to define the opening in the SU-8 to expose the active area of the chip. The soft-bake consists of slowly ramping the temperature from room temperature to 65°C, holding for 30 minutes at 65°C, increasing the temperature from 65°C to 95°C, and holding at 95°C for 8.5 hours. A longpass filter with a cutoff wavelength of 360 nm is used for exposure, resulting in sidewall gradients of 89.9°, as shown in Fig. 5b.

Following the SU-8 patterning, a conformal coating of 2 µm of Parlyene C is deposited on the chip surface to provide adequate dielectric isolation of the EWOD electrodes. Deposition is performed using the Labcoater2 Parlylene Deposition Chamber (Specialty Coating Services, Indianapolis, Indiana). To increase the hydrophobicity of the dielectric layer, a 100-nm-thick layer of Teflon AF (DuPont, New York, New York) is spin coated on top of the Parlylene-C. This dielectric surface is noted in the inset to Fig. 5c. The detailed process flow on the post-processing steps performed on the chip is given in the Supplementary Information. Following this post-processing, the electrowetting areas of the chip are planar to 0.2 µm.

3. Results

Integrated temperature sensor calibration using a ‘DNA thermometer’

The embedded heaters, calibrated with DNA melting, are used along with the temperature sensors and proportional-integral-differential (PID) control to produce the requisite qPCR heating profiles. All three heaters are run together with temperature measurement taken from the center sensor for control. Readings from the center resistive temperature sensor are shown in Fig. 6 during representative thermal cycling. With the use of a PID control loop, the surface heats from 72°C to 94°C in 11 sec, from 94°C to 60°C in 13 sec, and from 60°C to 72°C in 8 sec. The maximum ripple observed is 0.45°C. No active cooling is required; cooling at a rate of up to -3°C per second occurs by means of heat sinking through the very thermally conducting silicon substrate.

Fig. 6. qPCR thermal cycling profile. Temperature reading taken from center sensor.
In order to calibrate the temperature sensors to the droplet temperature on the chip surface, we developed a method using DNA contained in the droplet as an active temperature sensing material. Our approach builds on previous work and uses DNA melting, as tracked by an intercalator dye, to determine temperature. A single 12-base oligonucleotide can achieve a dynamic range of approximately 10°C. DINAMELT is used in the sequence design to simulate oligonucleotide melting and obtain the required melting temperature and profile. Special care is taken to avoid strands that exhibit self-folding tendencies. Many factors influence the melting temperature and slope. Longer sequences (>50bp) with a random assortment of nucleotides will typically have very high melting temperatures approaching 94°C. Shorter sequences will have lower melting temperatures with a slope highly dependent on the sequence. Sequences with high concentration of guanine-cytosine pairs will have broad melting curves due to strength of the inter-base bonds, but too many G-C pairs should be avoided to prevent excessive single-strand folding. For the calibration here, we use the sequence AAAGGAAAGGAAAAGGAAAAGG with a melting temperature of 72°C and calibration curve shown in Fig. 7a.

Heating across the surface is also relatively uniform. Fig. 7b shows the temperature distribution on the surface on the chip, as measured by DNA melting, as a function of time for a set point of 72°C for all three heaters running together. The center of chip surface ramps from 25°C to 72°C in 9.4 seconds. Steady-state is reached in approximately 10.8 seconds. At this point, roughly 75% of the chip’s surface is within 0.5°C of the desired set point. The remaining 25% is within 1.5°C.

**On-Chip SPAD performance**

The SPADs employed here have an avalanche breakdown voltage of approximately 20V. Operated at $V_{ov}=3.5$ V, they have an average dark count rate of 100 Hz and a photon detection probability (PDP) of better than 7.25% at 515 nm. Fig. 8a shows how the PDP varies with both wavelength and $V_{ov}$ at 25°C. PDP peaks at a wavelength of approximately 425 nm due to the shallow, heavily-doped junctions employed in the diode design. As shown in Fig. 8b, we also observe a decrease in the PDP of the SPADs with increasing temperature, which we attribute to temperature dependence in the impact ionization coefficients of the diode. As the temperature increases, the probability of carrier multiplication at the SPAD junction decreases and a decrease in sensitivity is observed. Fig. 8c shows a full calibration curve of SPAD counts with an integration time of 1 second as a function of dsDNA concentration with EvaGreen 20X intercalator dye (Biotium, Hayward, USA) for a few representative temperatures. Fluorescent data is collected following the elongation stage of each PCR cycle. SPAD detection limits are in the nanomolar regime, limited by SPAD dark count. The readings taken at...
50°C leave the noise floor at a double-stranded DNA concentration of approximately 1.7 µM. The curve exhibits approximately linear behavior at higher concentrations of DNA and thus higher incident fluorescent emission intensity. The exponential increase just as the curve leaves the noise floor can be attributed to a combination of the integration time and the non-linear scattering of low-intensity light at the Teflon/Parylene interface. Note that as the temperature increases, the curve shifts to the right, indicating a decrease in overall sensitivity.

**Droplet Evaporation**

Fig. 9a shows the rate of evaporation of 1.2 nanoliter droplets in air at 25°C. This evaporation rate of 20 picoliters per second makes this an impractical modality for qPCR. In our case, this evaporation rate is reduced to 6.3 femtoliter per second (Fig. 9b), resulting in a volume reduction of less than 1% during the 3600 sec that characterize a typical qPCR run.

**Performance of Electrowetting Droplet Transport**

![Fig. 10. Longterm electrowetting droplet transport testing. Increase in threshold voltage is due to decrease in surface hydrophobicity due to surface charge collection.](image)

The non-specific binding of biomolecules to the surface of Parylene-C/Teflon-AF dielectric has been shown to degrade the surface’s hydrophobicity. This factor was evaluated on our test chip using droplets containing the *S. aureus* PCR mixture. The threshold voltage necessary for droplet actuation is plotted as a function of number of actuations in Fig. 10. This test consisted of a ~1.2nL droplet being transported between set of adjacent electrowetting pixels. Supplementary Video 1 shows a representative droplet movement for this test. Note that a significant increase in the necessary electrowetting actuation voltage does not occur until after approximately 200 actuations, allowing significant use of the chip before this becomes a factor.

**Fig. 9.** a.) Droplet evaporation in air at 25°C. b.) Droplet evaporation in dodecane with and without the use of a coverslip during a complete 30-cycle PCR thermal cycling run.
The hydrophobic interaction between water, dodecane, and electrowetting surface is expected to show a temperature dependence. Using the same actuation test in which a 1.2-nL droplet is transported between two adjacent electrowetting pixels, the threshold voltage necessary for droplet actuation is plotted as a function of temperature in Fig. 12. Note that a significant increase in the electrowetting threshold voltage for our particular configuration does not occur until temperatures over 110°C. The available 90-V actuation voltage satisfies the necessary temperature range for this device.

Performance of Real Time PCR
As a proof-of-concept for our device, we amplify and detect a 364-base sequence characteristic for *S. aureus* (Norgen Biotek Corporation, Thorold, Canada). Fig. 11 shows qPCR data from individual droplets with serial dilutions of the target strand from 1000 down to 1 copy per droplet.

![Micrograph of 1.2-nL droplet within the array.](image)
Each test droplet comes from mixing PCR reagents, intercalator dye, DNA target, and buffer (Fig. 13). While we have successfully actuated 1.2-nL droplets, we get more consistent results from 3-nL droplets, which occupy approximately four pixels in the electrowetting array. Larger droplets impinge more of the higher fringing fields found at the edges of the square electrodes, allowing for easier movement. A video showing droplet transport for 3-nL droplets on the surface of the chip is given in Supplementary Video 2.

![Graph showing cycle threshold as a function of the log of the initial target copy number per 1.2 nanoliter droplet. Error bars represent four experimental trials.](image)

Fig. 14. Cycle threshold as a function of the log of the initial target copy number per 1.2 nanoliter droplet. Error bars represent four experimental trials.

The reagent consumption per qPCR test is reduced more than 40,000-fold compared to a conventional qPCR assay. For the full range of initial concentrations from 1 to 10000 copies per droplet, we obtain the expected linear relationship between cycle threshold and the logarithm of initial concentration (Fig. 14) with a standard deviation of less than 1 PCR cycle over four measurements at each concentration.

Conclusions

We have demonstrated the potential that CMOS ICs have for forming the basis for a nearly-fully-integrated droplet-based qPCR platform. All of the sensing and fluidics are embedded within the 4-mm-by-4-mm lab-on-chip device, running from a 3.3-V supply at 10mA, enabling battery-powered operation. Temperature control, droplet-based microfluidic transport, and SPAD-based fluorescence measurements are combined to perform proof-of-concept qPCR for the identification and quantification of *S. aureus*. Initial target copy numbers as low as one copy per 1.2 nanoliter droplet have been reliably amplified and detected. This work demonstrates how lab-on-chip devices can exploit the existing fabrication infrastructure and mature design capabilities of silicon CMOS foundries with modest additional post-processing steps and packaging techniques.

Affiliation

Columbia University, Department of Electrical Engineering, 500 W. 120th St., New York, New York 10027, United States
Notes and references