A 12.8K Current-Mode Velocity-Saturation ISFET Array for On-Chip Real-Time DNA Detection

Nicholas Miscourides, Student Member, IEEE, Ling-Shan Yu, Jesus Rodriguez-Manzano, and Pantelis Georgiou, Senior Member, IEEE

Abstract—This paper presents a large-scale CMOS chemical sensing array operating in current-mode for real-time ion imaging and detection of DNA amplification. We show that current-mode operation using ISFETs operating in velocity saturation devices can be exploited to achieve almost perfect linearity regarding their input-output characteristics (pH-current) which is aligned with the continuous scaling trend of transistors in CMOS. The array is implemented in a 0.35 µm process and includes 12.8K sensors configured in a 2T per pixel topology. We characterize the array taking into account non-ideal effects observed with floating gate devices such as increased pixel mismatch due to trapped charge and attenuation of the input signal due to the passivation capacitance and show that the selected biasing regime allows for a sufficiently large linear range which ensures linear pH to current despite the increased mismatch. The proposed system achieves a sensitivity of 1.03 µA/pH with a pH resolution of 0.101 pH and is demonstrated for the real-time detection of the NDM carbapenemase gene in E. Coli using loop-mediated isothermal amplification.

Index Terms—ISFET, array, current-mode, velocity saturation, DNA, LAMP, NDM, carbapenemase, E. coli

I. INTRODUCTION

The last decade has overseen a proliferation of CMOS-based chemical sensing systems being used in a variety of biomedical applications [1]. One enabling pH sensor which has contributed towards the successful development of such fully integrated sensing systems in CMOS is the Ion-Sensitive Field-Effect Transistor (ISFET) [2]. Owing to their compatibility with standard CMOS processes, ISFETs require no post-processing steps after fabrication which would typically require clean room facilities and expensive equipment. As a result, ISFETs are uniquely suited to address and bridge major limitations in the area of electrochemical instrumentation offering sensor miniaturization, scalability, mass fabrication and low cost. Numerous sensor architectures have appeared in the literature which leverage on the coexistence of sensors and instrumentation on the same IC to facilitate the transition from System to Lab-on-Chip (LoC) [3]–[6]. Recently, ISFET-based LoC platforms have been used to expand the field of rapid diagnostics with applications in infectious diseases [7], cancer [8], SNP identification [9] and multi-ion detection [10].

From a commercial perspective, ISFET sensors are most widely used as the sensing front-end in massively-parallel and large-scale arrays targeting DNA detection and sequencing [11], [12]. Interestingly, the introduction of ISFET technology was accompanied by a paradigm shift in sequencing methods since for the first time the economies of scale of semiconductors were leveraged to drive the cost of DNA detection down, eventually to the $1000 genome milestone [13]. Such sequencing platforms are designed in an imager-like configuration where each pixel aims to identify a fraction of the target sequence. As a result, the sensor density directly affects the throughput of these platforms and the pixel area dictates the overall silicon area and thus cost. Therefore, scalable sensor architectures have apparent benefits in this regard in order to facilitate chemical sensing arrays with similar order of magnitude resolution as optical imagers.

CMOS compatibility ensures that the sensors conform to a scaling trend according to Moore’s law [14] which has apparent benefits in terms of sensor density and thus increased in-parallel detection. However, analogue operation in sub-micron processes is significantly impacted by short-channel effects which have changed the operational characteristics of transistors. In sub-micron processes the typical FET regions of operation have shrunk to include the velocity saturation (VS) region at high electric fields. In general, this region is being avoided by designers due to its low-gain properties for single transistor amplifiers [15] since the transconductance of the device no longer follows a quadratic relationship but reduces to linear due to carrier saturation effects. In practice, this region is suitable for all applications which benefit from a linear gate voltage - drain current (VGS – ID) relationship, something that we have previously argued that is the case for the pH-induced gate voltage variations with ISFETs operating in current-mode [16]. Consequently, velocity saturation holds potential as an alternative region of operation since it provides a linear input-output relationship (pH – ID) for sensing and is expected to become more dominant in the increasingly accessible sub-micron processes.

Linearity is crucial for sensor instrumentation design, yet there are still inherent non-idealities for ISFET sensing that limit its wider adoption for fully-electronic pH detection. Major limitations include pixel mismatch due to trapped charge in the passivation layer typically built up during fabrication and drift of the sensing signal due to interfacial phenomena at the sensing surface. There are two main methods for addressing the aforementioned which call for additional complexity in the form of introducing a programmable gate [17] or gate reset switch [18]. The first method reduces the effective pH sensitivity whereas the latter introduces leakage at the gate which
manifests as electrical drift on top of the chemical. Additionally, the capacitive network created due to the formation of a passivation capacitance in conjunction with the gate parasitics attenuates the input signal from the electrolyte especially in the case of the passivation capacitance being of similar order of magnitude as the parasitics. Current approaches typically trade-off different attributes such as pixel size vs attenuation to tailor the configuration to its target application.

In this paper, we present a scalable architecture for ISFET arrays implemented in 0.35\(\mu m\) CMOS which operates in the velocity saturation (VS) regime. We show that adopting VS biasing can be exploited to make significant simplifications to the pixel design since the ISFET serves as both the sensing and transconductor device while ensuring that the output current is linearly dependent on the pH. This way, minimum size or very small devices are required which result in a small area footprint, large pixel density and thus array resolution. Furthermore, we demonstrate that contrary to typical calibration schemes which operate on a pixel-by-pixel basis, the large linear range of VS allows for the trapped charge spread to be accommodated. Therefore, global biasing of the sensors through the reference electrode can be optimized to maximize the number of active pixels irrespective (to a degree) of trapped charge mismatch while delivering linear input-output transduction. Finally, we demonstrate the system for the rapid and real-time on-chip DNA amplification and detection of the NDM-1 gene isolated from Escherichia coli using the Loop-Mediated Isothermal Amplification (LAMP) method. Bacterial strains that carry the NDM gene represent a major health challenge worldwide since they are resistant even to carbapenems antibiotics which are used as a last resort when common antibiotics have failed [19].

The paper is organized as follows: Section II describes the theoretical overview of ISFETs in CMOS and description of the velocity saturation origin. Section III presents the ISFET sensor design and pixel configuration. Section IV shows the system architecture and block description whereas Section V presents the chip characterisation and measured results. Finally, section VI demonstrates the system application for isothermal DNA amplification and detection using LAMP.

II. THEORETICAL OVERVIEW

A. ISFET Fundamentals

ISFETs designed in standard CMOS processes have been possible since 1999 [20] with a typical structure as shown in Fig. 1. Compared to earlier approaches, this method circumvented a major issue with fabrication, that it is not possible to obtain a polysilicon-insulator structure in a typical CMOS process which prevents the gate of the readout transistor from being in direct contact with the pH-sensitive insulating layer. The workaround was that an electrically-floating electrode was left between the polysilicon layer (needed to define the gate) and the top metal layer in the CMOS process. This ensures that the surface properties of the passivation layer are not affected while removing the need for further post processing in order to expose the gate.

What has established however the popularity of ISFETs as a semiconductor-based ion-sensitive sensor is that the typical material serving as the top insulating and passivation layer in the CMOS stack is silicon nitride (\(Si_3N_4\)) which is a pH sensitive material. Therefore, during metal layer deposition, ISFETs are created with metal layers and interconnecting vias providing a path from the gate of the underlying FET to the top metal layer which resides directly below the passivation. When in contact with an aqueous solution, silicon nitride is used as a hydrogen ion sensing layer due to the electrochemical phenomena taking place at the electrolyte/insulator interface, specifically the binding of hydrogen ions to the insulator surface and the formation of a double layer capacitance [2]. As a result, ion activity taking place at the passivation layer as well as the DC potential of the solution set by an external reference electrode \(V_{\text{ref}}\) (in Fig. 1) are capacitively coupled to the electrically-floating gate of the transistor and jointly define the operating point of the device.

CMOS compatibility ensures a number of desirable properties for fabricating chemically-sensitive yet electronic sensors. The system design can be significantly simplified by using a commercial CMOS process since it takes advantage of the foundry well-established design ecosystem. Compared to conventional potentiometric sensors such as ISEs, monolithic integration ensures miniaturization of sensors, scaling according to Moore’s, low-cost and mass-fabrication. Furthermore, it enables the utilization of the ISFET as a transistor which opens up the possibility of adding signal/data processing circuitry on the same chip as the sensor.

The chemical-electrical conversion at the electrolyte-insulator interface is explained by the site-binding theory which leads to the formation of a Gouy-Chapman-Stern double-layer capacitance [21]. Specifically, the double layer capacitance leads to the formation of an additional term, \(V_{\text{chem}}\), in the standard transistor model which is pH dependent. As modelled by [22]–[24], ISFET characteristics are described by:

\[
V_{\text{OV(ISFET)}} = V_{\text{chem}} + V_{\text{OV(MOSFET)}}
\]

where \(V_{\text{OV}}\) is the overdrive voltage \(V_{\text{OV}} = V_{GS} - V_{th}\) and

\[
V_{\text{chem}} = \gamma + \alpha S_{\text{Nernst}} pH
\]

![Fig. 1. ISFET cross-section in unmodified CMOS and equivalent macromodel.](image)
where \( \gamma \) groups all the non-pH dependent terms and \( \alpha \) represents the deviation from the maximum Nernstian sensitivity \( S_{\text{Nernst}} \).

\[
\alpha = \frac{S}{S_{\text{Nernst}}} \quad (3)
\]

where \( S \) indicates the observed sensitivity of the device, \( S < S_{\text{Nernst}} \) and \( S_{\text{Nernst}} = 2.3kT/q \approx 59 \text{ mV/pH} \) where \( kT/q \) is the thermal voltage.

**B. Velocity Saturation Fundamentals**

In sub-micron processes, second-order effects such as carrier velocity saturation, threshold voltage variations and hot carrier effects which were often ignored with long channel devices (\( L > 1 \mu m \)) are becoming more prominent and have already started to dictate the operation of devices. Velocity saturation of the charge carriers when travelling along the channel is the most impactful effect on the characteristics of the device and is described in more detail below since it is a mode of operation that we are exploiting to linearise the voltage-to-current operation of the device. On the other hand, hot carrier effects (which cause a slow modification of \( V_{th} \) from extensive use due to electrons tunnelling in the oxide) and additional variations to the threshold voltage (due to \( V_{th} \) becoming a function of \( L, W \) and \( V_{DS} \) on top of \( V_{RS} \) and process parameters) cause smaller deviations especially when the device is biased in velocity saturation. In this case, we focus on the most impactful effect and how the deviation from the expected characteristics is caused.

In brief, the behaviour of a MOSFET is typically described with the model shown below:

\[
I_D = \mu_n C_{ox} \frac{W}{L} \left( (V_{GS} - V_{th})V_{DS} - \frac{V_{DS}^2}{2} \right) \quad (4)
\]

where \( \mu_n \) is the mobility of carriers (electrons), \( C_{ox} \) the oxide capacitance and \( W \& L \) the device dimensions. The carrier mobility \( \mu_n \) determines the electron velocity \( v_n \) scaled by the electric field \( E \) along the channel (\( E \propto V_{DS} \)).

\[
v_n = -\mu_n E(x) \quad (5)
\]

where \( x \) denotes distance. Typically, for decreasing transistor size, the effective channel length decreases which causes the electrical field along the channel \( (E(x)) \) to increase. Consequently, the electron velocity should increase at a rate proportional to the rising electric field as described in Eq. 5. However when a critical value is reached, \( E_C = 1.5 \times 10^6 \text{ V/m} \) for p-type silicon, the carrier velocity tends to saturate at a value \( v_{sat} = 10^5 \text{ m/s} \) (shown in Fig. 2a) due to collision and scattering effects. The drain-source potential that corresponds to the \( E_C \) field value is denoted \( V_{DS,sat} \), above which the carrier velocity stays approximately constant. At higher \( V_{DS} \) potentials, carriers reach the saturated velocity earlier along the channel as they propagate from the source to drain diffusion regions. Under these conditions we can make a first order approximation for modelling the device by assuming that:

\[
v = \begin{cases} 
\mu_n E(x) & \text{for } E \leq E_C \\
\mu_n E_C = v_{sat} & \text{for } E \geq E_C 
\end{cases} \quad (6)
\]

Additionally, we can assume that the drain-source potential at which \( E_C \) is reached is constant and given by:

\[
V_{DS,sat} = L \times E_C = L \frac{v_{sat}}{\mu_n} \quad (7)
\]

Therefore, operation in the triode region remains the same as the case of a long-channel device whereas the saturated current \( I_{DS,sat} \) equation is obtained by substituting Eq. 7 to Eq. 4. As shown in Fig. 2b, we can see that \( V_{DS,sat} \) occurs earlier than \( V_{GS} - V_{th} \) and an extended saturation region is obtained. Furthermore, \( I_{DS,sat} \) and the maximum transconductance (gain) in this region are described by:

\[
I_{DS,sat} = v_{sat} C_{ox} W \left( (V_{GS} - V_{th}) - \frac{V_{DS,sat}}{2} \right) \quad (8)
\]

\[
g_m = \frac{\partial I_D}{\partial V_{GS}} = v_{sat} C_{ox} W \quad (9)
\]

According to the equations above, in velocity saturation the drain current is linearly dependent on the overdrive voltage \( V_{GS} - V_{th} \) and the width \( W \) of the device while being insensitive to \( V_{DS} \). Additionally, the saturated current is less dependent on second order effects such as mobility (\( \mu \)) degradation as well as the device’s length (\( L \)) which suggests that infinite output impedance is expected and that operation is insensitive to length mismatches which usually occur with minimum length devices. In practice, the device exhibits a high output impedance and channel length modulation similar to normal saturation due to the fact that different regions along
the channel might behave as in different regions of operation, specifically velocity saturation at the start of the channel and normal saturation throughout the rest. However, higher output impedance is expected in general which reduces the susceptibility to bias voltage variations and increases its usability as a programmable current source. Further information on device modelling and higher order effects can be found in [15], [25].

C. Expected Attenuation

The passivation layer \( (Si_{3}N_{4}) \) that is used for sensing, leads to the formation of an equivalent capacitance, \( C_{\text{pass}} \), and thus a series capacitive network in conjunction with the gate parasitic capacitances as shown in Fig. 3. This capacitance can be estimated using the process parameters (passivation thickness and dielectric constant) and the size of the top metal electrode. Additionally, the attenuation is a function of the gate parasitics which depend on the gate area and device biasing region as described in [14]. Different readout schemes will also affect the attenuation, for instance \( C_{gs} \) in a source follower configuration has a non-unity gain across it (typically around 0.85) and will scale according to Miller effect. Nevertheless, in current mode the source and drain potentials are kept constant therefore the overall effect is that any input voltage signal will be attenuated according to:

\[
\frac{V_{g}}{V_{in}} = \frac{C_{\text{pass}}}{C_{\text{pass}} + C_{gd} + C_{gb} + C_{gs}} \tag{11}
\]

where \( C_{gd} + C_{gs} + C_{gb} = C_{gg} \) denote the gate parasitic capacitances. From Eq. 11 we can see that the attenuation minimizes for a large passivation capacitance and/or small gate parasitics. In a given CMOS process (thus materials with fixed thickness and dielectric constants), \( C_{\text{pass}} \) is directly proportional to the size of the top metal electrode therefore a larger metal area is beneficial. Nevertheless, this comes at the expense of a larger pixel area with trade-offs associated with increased silicon area, array resolution and compactness. Alternatively, the attenuation can be reduced by using a small size device which ensures \( C_{gg} \) is small. In current-mode, having a small device is favourable to reach the in velocity saturation region. Therefore, the gate parasitics are reduced to a minimum thus attenuation in current-mode is a function of the size of \( C_{\text{pass}} \) which constitutes a trade-off with pixel size. In our design we optimised the pixel size vs attenuation trade-off so we could create a dense array, but also still achieve sufficient resolution for the target application.

III. ISFET SENSOR AND PIXEL DESIGN

A. ISFET in Velocity Saturation

In the case of an ISFET, an additional \( V_{\text{chem}} \) term is included in the device model which appears in series with the overdrive voltage. Therefore, in velocity saturation an ISFET will behave as:

\[
I_{D,\text{ISFET}} = v_{\text{sat}} C_{ox} W \left[ (V_{GS} - V_{\text{chem}} - V_{th}) - \frac{V_{DS,sat}}{2} \right] \tag{12}
\]

Thus, \( I_{D} \) is linearly proportional to \( pH \) and the \( pH \) sensitivity is given by:

\[
g_{m}(pH) = \frac{\partial I_{D}}{\partial pH} = -v_{\text{sat}} C_{ox} W (\alpha S_{\text{Nernst}}) \tag{13}
\]

We can see that the \( pH \) transconductance of the device evaluates to a constant which depends on the device’s width, process parameters and the quality of the sensing layer. As long as the transconductance is constant, given the operating
conditions, then linear \(pH\)-to-current conversion is ensured. From the above equations we can also gain insight on how to size an ISFET for operation in velocity saturation. Specifically, a short length is naturally selected to ensure that velocity saturation is reached whereas a short width might also be preferred in order to limit the biasing current occurring at high \(V_{GS}\) and \(V_{DS}\) potentials. Effectively, sizing the sensing device comes down to a trade-off between DC current bias in velocity saturation and device gain in this region (\(g_{m,PH}\)). In the 0.35\(\mu\)m process, the device has bee sized such that the velocity saturation current is below 200\(\mu\)A which is in accordance with the expected current range for the peripherals blocks while providing sufficient gain for the target application.

To verify the expected characteristics, Figs. 4a-4c show simulation results of the drain current and transconductance of a short length device in a 0.35\(\mu\)m commercial CMOS process. Both \(I_D\) and \(g_m\) are plotted on linear axes and shown as a function of the \(V_{GS}\) and \(V_{DS}\) potentials of the device with the substrate tied to ground (\(V_{BS} = 0\)V).

We can see from Fig. 4a that velocity saturation causes the device to display an extended saturation region since \(V_{DS,sat}\) occurs earlier than the typical “knee” of the \(I_D-V_{DS}\) response (when \(V_{DS} \geq V_{GS} - V_{th}\)) with a long channel device. As a result, we expect a linear \(I_D-V_{GS}\) response with the gradient (gain) given by the transconductance of the device. Looking at the 3D transconductance of the same transistor on Fig. 4b, we can see that for a given \(V_{DS}\), \(g_m\) increases with increasing \(V_{GS}\) until it reaches a relatively flat peak and subsequently decreases due to second order effects such as mobility degradation. This behaviour is similar to the expected response for a long channel transistor which exhibits only traditional saturation however in this case \(g_m\) will not reach as high values and will show a flatter peak. Additionally, for very high \(V_{DS}\) potentials, \(V_{DS} > \frac{V_{DD}}{2}\), \(g_m\) increases monotonically with increasing \(V_{GS}\) and reaches its highest point when \(V_{GS}\) approaches the power supply.

In the case of designing an ISFET for linear \(V_{GS}-I_D\) conversion, we are interested in a high \(g_m\) to maximize the intrinsic gain as well as obtaining a wide peak to ensure linearity. Furthermore, it is preferable to do so without resorting to very high \(V_{GS}\) and \(V_{DS}\) potentials (and thus drain current) which will increase the power consumption of the sensor. As a result, the best biasing point for a current-mode ISFET is annotated in Fig. 4b where the widest linear range is recorded as a trade-off with biasing current.

To determine the expected operating range for a single device, we assume a maximum \(pH\) sensitivity of 59 mV/pH

\[2\text{Linearity has been evaluated as the extent for which the output is linear for the available input range i.e. derivative is constant.}

Fig. 5. \(I_D-V_{GS}\) and \(g_m\) characteristics in a 65nm commercial process with \(V_{DS} = V_{DD}/2 = 0.6V\). The device in (a) is minimum size and in (b) is the same size as the device in Fig. 4.

<table>
<thead>
<tr>
<th>Configuration</th>
<th>Input Linear Range</th>
<th>Linearity (^2)</th>
<th>Gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Velocity Saturation</td>
<td>Large</td>
<td>Large</td>
<td>100 (\mu A/V)</td>
</tr>
<tr>
<td>Common Source</td>
<td>Small</td>
<td>Large</td>
<td>&gt; 10</td>
</tr>
<tr>
<td>Source Follower</td>
<td>Large</td>
<td>Large</td>
<td>~ 0.85</td>
</tr>
</tbody>
</table>

\(^2\text{Linearity has been evaluated as the extent for which the output is linear for the available input range i.e. derivative is constant.}

\(\alpha = 1\) therefore the entire \(pH\) scale would need 14 \times 59 = 826mV. Nevertheless, this model doesn’t take into account the attenuation that charge coupling at the passivation layer will experience to reach the floating gate as well as trapped charge at the passivation layer which typically occurs with floating gate devices. For a single device, both effects can be mitigated with precise biasing of the reference electrode, however for an array of sensors sharing a common reference electrode different biasing points across devices can be expected. As a result, it is preferable to operate in a region which offers the widest linear range possible.

In order to quantify the linearity of the device for a given \(V_{GS}\) range, the coefficient of determination \(R^2\) is used which indicates the extent to which the data points approximate a straight line. Table I shows indicative \(V_{GS}\) ranges and the corresponding linearity in \(I_D\) for the device plotted in Figs. 4a-4c. We can see that a short channel device biased in velocity saturation achieves almost perfect linearity, remarkably consistent throughout a large range of \(V_{GS}\). This is a consequence of carrier velocity saturation (\(V_{DS} = 1.5V\)) which corresponds to a peak \(g_m = 100.9\mu S\) varying only by \(\pm 3.88\%\) across a 826mV range. Therefore, it is possible to convert the \(pH\)-induced \(V_{GS}\) variations to drain current variations in a linear fashion with no linearisation method considered other than designing the underlying device to operate in velocity saturation.

For velocity saturation ISFET sensors both linearity and linear range are important since high linearity allows us to obtain linear input-output characteristics (\(pH\)-to-current) which is crucial in sensor instrumentation to facilitate subtraction/double sampling whereas a large linear range allows us to accommodate a large extent of trapped charge across multiple sensors. Typically, for ISFET sensing we are interested in the differences in \(pH\) of successive measurements rather than the absolute value which can be achieved this way by subtracting...
a reference frame from any subsequent while still producing linearly proportional differences to successive samples.

Comparing with other topologies, Table II shows a qualitative comparison of an ISFET in velocity saturation with regards to normal saturation configurations. Even though the comparison is not straightforward given the two different modes of operation (current vs voltage), operating in VS ensures linearity over a large range, similar to a source follower, with a transconductance gain as opposed to attenuation with the source follower. On the contrary, a common source amplifier has a very short input range which might saturate the output for the expected pH range. Overall, operating in velocity saturation allows for a wide pH range and is consistent with modern CMOS processes.

The 0.35\(\mu\m\) process ensures that velocity saturation is reached with a short channel device while allowing for a wide range of input \(V_{GS}\) voltage (1.4V) for which the output current is linear. As a result, this wide range can accommodate trapped charge commonly found in ISFET sensors to a large extent without the need for mismatch calibration since the difference in successive measurements will be linear even if they start at different DC points. Compared to more advanced processes, a standard device in a 65 nm commercial process operates with a power supply of 1.2 V which reduces the available range of linear \(V_{GS}\). To demonstrate this, Fig. 5 shows simulation results of the I-V characteristics in velocity saturation in a 65 nm process and Table III shows the corresponding \(V_{GS}\) range for which \(I_D\) if perfectly linear. This range has been obtained as the \(V_{GS}\) range at which \(g_m\) is maximally flat to within \(\pm 10\%\) of its maximum. Alternatively, advanced process nodes typically offer the option of higher \(V_{DD}\) devices for analogue applications. In the 65nm process, devices with 2.5 V supply can be used to increase the available linear \(V_{GS}\) range in velocity saturation. However, these devices have an increased minimum device size of \(W/L = 400/280 \text{nm/nm}\) which is comparable to minimum size in the 0.35\(\mu\m\) process and offer comparable performance with similar silicon area.

B. Pixel Design

Following the previous analysis, we can see that biasing the ISFET in velocity saturation allows for a single device to act as both the sensing and gain components of the input chemical signal. For optimum operation in VS, this device is of small size (shown in Fig. 6a) with the pixel gain given by the pH transconductance \(g_m(pH)\). Overall the pixel comprises two devices, the ISFET and a column selection switch which is designed to have a very small drop across it when the pixel is selected. The pixel size is defined by the size of the top metal area that capacitively senses the changes at the passivation layer. In this case, the pixel size is set as \(6.5 \times 7.8 \mu m^2\) with a \(2\mu m\) inter pixel distance as shown in Fig. 8 (left). Each pixel is accessed by establishing a non-zero \(V_{DD}\) across the pixel and the output is recorded as the drain current of the ISFET. With this configuration scaled in an array only one pixel is selected at any given time, to ensure that no more than one ISFET is sinking the velocity saturation current which is typically high in this region (\(> 80\mu A\)) in the 0.35\(\mu\m\) process. Similar imager configurations for photo sensing in current mode have been presented in [26], [27].

Fig. 7a shows the expected output current range for step pH changes with VS biasing assuming 30mV/pH net effect at the floating gate which is expected for unmodified CMOS [28]. The corresponding step changes in drain current effectively show the transconductance of the device and are perfectly linear as shown in Fig. 7b. This is a consequence of the \(g_m\) being approximately flat throughout VS and shows that linear pH-current conversion occurs with around \(3\mu A/pH\) sensitivity which requires no additional amplification for detection.
IV. SYSTEM IMPLEMENTATION

Following the analysis on the current-mode pixel and ISFET linearity in velocity saturation, this section describes the implementation of a complete system-on-chip which demonstrates the feasibility and scalability of this approach. Fig 6 shows the complete system architecture comprising the major block schematics whereas Fig. 8 shows the chip floorplan. Descriptions of major design blocks are shown below.

A. ISFET Array

The sensing part of the system consists of a 64 x 200 ISFET array with a global reference electrode used to bias the solution and determine the large signal response. As a result, all ISFETs will establish a $V_{GS}$ set by the combined effect of the reference electrode and any attenuation occurring due to the passivation capacitance. Since the reference electrode is externally controlled, it can be arbitrarily adjusted to maximize the number of sensors with a uniform output, correspondingly minimizing the effect of trapped charge and mismatch.

Addressing the array takes place by setting a fixed potential to the drain of each ISFET using a current conveyor on a row-by-row basis. This way a single switch can be used to address each row independently rather than in-pixel switch, thus its aspect ratio can be made very large to minimize any voltage drop across. Subsequently, an in-pixel column switch allows for individual pixel selection as shown in Fig. 6. Accessing each pixel takes place by establishing the current conveyor reference potential $V_D$ as the drain row potential $V_{DR}$ and thus as the $V_{DS}$ of each ISFET, assuming minimum voltage drop across the row and column switches. As a result, this configuration allows for programmable $V_{DS}$ and $V_{GS}$ ISFET voltages to target the velocity saturation region. Additionally, the high output impedance in this region reduces the variability of the output current to $V_{DR}$ variations observed at different places regions of the array and different bus lengths.

B. Current Conveyor

The current conveyor block is shown on Fig. 6c and is used to provide the velocity saturation current to each pixel while applying a fixed potential $V_{DR}$. Additionally, it mirrors the velocity saturation current to both an on-chip I-V & ADC block and to an analogue pad for external sampling. The current conveyor is based on a stacked configuration of cascode current mirrors operating such that the upper current mirror sets equal currents flowing through each branch whereas the bottom one will have equal currents in both branches only when the source voltage of the bottom devices is the same. Cascoding boosts the output impedance and increases the current range for which $V_{DR} \approx V_D$ whereas the aspect ratio of all devices is designed very large to allow a sufficient voltage headroom for $V_{DR}$ to be set in velocity saturation i.e. up to 1.5V for a current range $I_{ROW} \leq 200 \mu A$.

C. Digital Control Block

System timing and synchronization is carried-out by an on-chip control block comprising a synthesized FSM and an SPI communication block. The FSM includes multiple modes which allow for individual pixel addressing, continuous array raster scan and timing control for the on-chip or off-chip ADCs. All these actions can be carried-out at various
integer multiples of clock cycles with the maximum readout speed set at 1 clock cycle per pixel using external sampling and 11 clock cycles per pixel using internal sampling. Data communication takes place through a 16-bit SPI protocol which includes incoming commands specifying the FSM state and corresponding output depending on the command selected.

D. Readout Block

The VS current is converted to a voltage using a transimpedance amplifier (TIA) based on a wide-bandwidth op-amp provided by the 0.35μm process.1 The TIA is designed to convert a 0 – 200μA current range to a 1.6V range. The converted value is sampled by an on-chip 10 bit SAR ADC also provided as a standard cell. To maximize bit utilisation, the positive reference \( V_{RP} \) is the same as the TIA reference (2V) to sample a 0.4 – 2V linear signal. Additionally, the analogue signal is output for external sampling of either the current or the voltage signal. In this case, the voltage signal after the transimpedance amplifier is sampled using an external 16 b ADC for higher resolution.

Using the on-chip ADC and a clock frequency of 1.1 MHz, a raster scan requires 11 clock cycles per pixel which corresponds to 7.8 fps. The readout time can be adjusted by specifying a programmable digital instruction that controls the number of clock cycles per pixel in case of external sampling or ADC conversions per pixel in case of the on-chip ADC. Additionally, in both cases the readout speed can be varied by modifying the master clock frequency.

V. EXPERIMENTAL CHARACTERIZATION AND RESULTS

The following section describes the materials and methods used to experimentally verify the operation of the chip and the sensing platform. All following measurements describe wet measurements with the setup described below.

A. Fabricated Chip

The fabricated chip is shown on Fig. 8 with the sensing array highlighted and the block floorplanning on the right. This system forms part of a larger platform fabricated using the AMS 0.35μm process on a 2 × 4 mm² silicon chip. All peripheral circuits and bondpads are placed at the bottom to ensure that encapsulation will not affect the middle part of the chip which will be exposed to a solution. A custom made PCB cartridge has been designed to host the chip and microfluidic arrangement whereas a motherboard PCB is used to handle communication and power to the cartridge via a ribbon cable. The motherboard also includes a microcontroller, bias voltage generation and an external ADC with a higher resolution readout than the on-chip.

B. Encapsulation and Microfluidic Manifold

For wet experiments, the bare chip is glued on a ground plane (gold) on the cartridge PCB shown in Fig. 9a. Additionally, electrically insulating epoxy (EPO-TEK T7139) was used to cover the bond pads and serve as glob top encapsulant while only covering the bottom part of the chip to keep all arrays exposed.

A custom build microfluidic manifold forms a reaction chamber and brings the top surface of the chip in contact with a solution as shown on Fig. 9b. The chamber allows for ~300μm of space on each side of the chip which ensures that the larger part of the arrays are exposed. The manifold and chamber have been laser cut from an acrylic sheet and are screwed down on the cartridge PCB. Additionally, double sided biocompatible adhesive tape has been added around the microfluidic chamber that is in contact with the chip to ease the stress from screwing and prevent leakage during the amplification reaction. Large-signal biasing of the solution is provided by a reference Ag/AgCl electrode obtained by galvanostatic oxidation on silver tubes as proposed in [29]. Following the oxidation process, an aluminium wire was soldered on both tubes and connected to a voltage reference to be used for biasing. Two
silver tubes of outside diameter 1mm and wall thickness 0.175mm were used which also provided the inlet and outlet of the reaction chamber. Additionally, a heatshrink tube was added for protection of the tubes and prevention of any other conducting material getting in touch with the reference.

C. Trapped Charge and Attenuation

To verify the sensitivity to the reference voltage and bias of the solution, the reference electrode was swept in a solution of constant pH=7 and the array output recorded as shown in Fig. 10. The bottom part of the array is covered by the acrylic manifold and is not in contact with the solution causing 2816 out of 12800 pixels to show no sensitivity to the sweep. This part has been disregarded from measurements. The remaining part of the array shows the effect of trapped charge at the floating gate which causes the pixels to operate above saturation as shown in Fig. 10. To counteract this effect, the reference electrode voltage is varied to modify the $V_{GS}$ of each sensor and allow for the target biasing region (velocity saturation) to be reached.

Trapped charge is randomly left during fabrication at the floating gate of ISFETs which effectively creates significant mismatch between pixels. Current mode is beneficial in this case since it ensures linearity over a wide $V_{GS}$ range and reduces the susceptibility to mismatch. We use the standard deviation of active pixels to quantify the extent of trapped charge across the array. For the active pixels shown in Fig. 10 the trapped charge standard deviation is 11.3µA. Therefore, this will not affect the measurements since the output current is linear for a wider range and the difference in measurements will be accurate even if the DC starting point is different.

When the exposed part of the array is operating in velocity saturation and maximally linear range (i.e. 80 < $I_D$ < 200µA with $V_{ref}$ = −2V), 83.1% of the available pixels are used for sensing. To estimate the attenuation, the average current difference of active pixels for step changes in $V_{ref}$ was compared against $I_D-V_{GS}$ simulations such that the actual $V_{GS}$ values at the gate could be determined. This resulted into an average attenuation of $A_{actual} = 0.424$ that the input signal experiences to reach the floating gate. In contrast, the expected attenuation obtained through simulating an ISFET in Cadence Virtuoso was $A_{sim} = 0.57$ i.e. less attenuation. The expected attenuation was obtained by estimating the passivation capacitance value $C_{pass} = 1.14fF$ (see footnote 1) and obtaining a simulated value of the gate parasitic capacitances in velocity saturation as $C_{gg} = 0.865fF$, where $C_{gg} = C_{gs} + C_{gd} + C_{gb}$. The deviation of the actual and expected values can be explained if the parasitics are larger and/or the passivation capacitance is smaller than the estimated. Since, the $C_{pass}$ estimate ignores any fringing effects on the passivation layer, it is thus a conservative estimate and the actual value is probably larger than smaller. As a result, the above indicate that the parasitic capacitances at the floating gate are larger than $C_{gg}$ which is expected due to additional parasitics from metal buses, adjacent gates etc. which increase the attenuation of the input signal to the observed value $A_{actual}$.

D. pH Sensitivity

The pH sensitivity of the sensor array was verified by inserting in sequence pH buffers ranging 4-10 with a step change of 1. All buffers were prepared by increasing or decreasing the pH of a template solution with 1M HCl (SigmaAldrich 318949) or 1M NaOH (SigmaAldrich 71463). The template solution was prepared using 3M KCl (Sigma Aldrich 60137), 1M Tricine (Sigma Aldrich T0377) and de-ionized water. Subsequently, a peristaltic pump was used to create a continuous flow and sequentially pump each solution to the chip surface using the manifold of Fig. 9a. In between
measurements, a solution of pH 7 was inserted to accurately record the differences between pH 7 and the step change.

Pixels not in contact with the solution were disregard from any subsequent analysis such that the mean of the array can provide a suitable indicator of the pH change. Fig. 12 shows the array output as a histogram for representative solutions of pH 4, 7 and 10. The trapped charge spread causes an output current spread of around 50µA with an average variance σ ≈ 11.3 µA. Nevertheless, since the sensors are biased in velocity saturation which allows for a large linear range, pH differences will cause a linear change in current after subtracting the initial value. Finally, Fig. 11 shows the final pH sensitivity obtained throughout this characterization step. Using pH 7 as the reference point, the final sensitivity is $S_{pH} = -1.03\mu A/pH$ with $R^2 = 99.72\%$ linearity. Taking into account the attenuation experiment result, the intrinsic pH sensitivity of the passivation layer is determined at $S_{Si_3N_4} = 24.1mV/pH$, which is expected for ISFETs in CMOS [28].

Evidently the quality of the measurements is affected by drift at the sensor’s output, as shown in Fig. 11, caused by a slow modification of the $Si_3N_4$ insulator surface after exposure to an electrolyte. Silicon nitride has been shown to slowly convert to a hydrated $SiO_2$ or oxynitride layer which modifies the equivalent insulator capacitance over time [30]. For long measurements, an exponential decay model has been shown to describe drift [31] which can be reduced to linear for sufficiently short intervals. Assuming a relatively small instantaneous drift rate, derivative-based methods have been developed that deal with drift even for long term measurements [31]. In our case, the time course of drift is much slower compared to pH changes therefore it can be assumed to be locally linear. As such it can be subtracted for short time intervals around one pH step change which has been taken into account when obtaining the pH sensitivity of the sensor.

E. Current-Mode Noise

To characterize the total noise of the pixel including both electrical and chemical contributions, a current-mode pixel was connected to the on-chip transimpedance amplifier and sampled externally using a high precision data acquisition unit (PowerLab 8/35, ADInstruments). The output signal was continuously sampled for 10 minutes at $f_s = 20 kHz$ with 16-bit accuracy and a ±2V maximum input signal i.e. $31\mu V$ resolution. Subsequently, the input referred current noise at the input of the TIA was obtained using $S_i = S_v/R^2_T$ as in [32] where $S_v$ and $S_i$ represent the power spectral densities at the output and input of the TIA specified in $V^2/Hz$ and $A^2/Hz$.

Fig. 13 shows the power spectral density of the current-mode signal in $\frac{A}{\sqrt{Hz}}$, whereas the inset shows the sampled signal in the time domain which contains chemical drift with an average rate of 9.2 mV/min. Evidently, at frequencies of typical pH reactions ($<100Hz$) flicker noise is dominant which is expected since we have opted for a very small sensing pixel area. Additionally, chemical noise contributes to the overall noise response since it is typically low frequency and has been shown to increase ISFET flicker noise by at least an order of magnitude compared to its MOSFET counterpart [33]. Since our target application is low frequency and the DNA amplification reaction occurs in the time span of several minutes, the current noise has been obtained at $1Hz$, which corresponds to $104.28\frac{A}{\sqrt{Hz}}$. As such, the pH resolution is 0.101 pH with an SNR of 19.9dB.

VI. Real-time DNA Amplification Detection

To demonstrate the feasibility of the proposed system as a Lab-on-Chip platform, this section presents its application in real-time DNA amplification and the rapid detection of the NDM gene taking place in a microfluidic chamber on top of the chip. For this experiment, the acrylic manifold was replaced with one of open top such that cartridge PCB can be inserted in a flat block PCR machine (Veriti Thermal Cycler, Applied Biosystems) for temperature control. One drop of mineral oil was added to the reaction chamber to prevent evaporation during the amplification reaction. Additionally, a 0.1mm Ag/AgCl wire was inserted in the chamber for biasing. DNA amplification and a corresponding pH change is facilitated using the loop-mediated isothermal amplification technique [34] which was selected since it imposes no requirements for temperature cycling typically needed by PCR. On the contrary, DNA detection using LAMP takes place
at a constant temperature, typically 63°C, thus ensuring no temperature interference in CMOS at the time of amplification.

For this experiment, genomic DNA was isolated from pure cultures of carbapenem-resistant *Escherichia coli* carrying the NDM gene. The purified DNA was analysed immediately or stored at -80°C until further experiments. Specific LAMP primers published in [35] were used to confirm the presence of target NDM gene in the extracted samples. All oligonucleotides used in this experiment were synthesised by IDT (Integrated DNA Technologies, Germany). Negative and positive real-time LAMP experiments (10⁵ genomic copies/reaction) were performed in parallel using a conventional real-time instrument (LightCycler 96 System, Roche) and on-chip. The LAMP assay was optimised for pH-based on-chip detection. Twenty-µL solution was loaded into 0.2mL PCR tubes or into the microfluidic chamber on the chip and heated at 63°C for 20min. A fluorescent intercalating dye (SYBR green) was added for the detection of the dsDNA products using the LightCycler 96 System, which was used to compare with the on-chip positive and negative reactions.

Fig. 15a shows the fluorescent response over time obtained from the LightCycler instrument whereas Fig. 15b shows the array mean output of all active pixels for both negative and positive reactions. We can see that in both cases DNA amplification takes place and the positive is clearly identifiable. The reaction starts after the temperature has settled and lasts approximately 1000s after which DNA amplification plateaus as shown in Fig. 15a. At that point, the array produced a mean output change of 0.53µA between negative and positive runs which corresponds to 0.55 pH using the S_pH figure obtained earlier. After that point, the pH stays the same with differences in output attributed to drift. In comparison, the corresponding pre and post pH values of the solutions obtained using a commercial pH meter is shown in Table IV and indicates a pH change of 0.6 which is in agreement with the measured. We anticipate that the high temperature adds to the drift which is similar during both the negative and positive runs. Nevertheless, since the reaction is isothermal, temperature transients do not interfere with the output signal and differences in drift rate between the positive and negative reactions can be used to distinguish the pH difference. Additionally, Fig. 16 shows the distribution of active pixels at the beginning, post-amplification and end of the DNA amplification reaction. In this case, the distributions capture the differences in output current after subtracting the initial offsets due to trapped charge which is possible since all the active pixels are operating in the linear velocity saturation region.

![Fluorescence output from positive and negative reactions conducted in the LC96 System.](image)

(a) Fluorescence output from positive and negative reactions conducted in the LC96 System.

![Mean output of exposed pixels in VS after positive and negative pH LAMP reactions with the standard deviation indicated by the shaded bars.](image)

(b) Mean output of exposed pixels in VS after positive and negative pH LAMP reactions with the standard deviation indicated by the shaded bars.

Fig. 15. Results from the DNA amplification experiment for the detection of the NDM gene in *E. coli* in (a) the benchtop instrument and (b) on-chip.

**TABLE IV**

<table>
<thead>
<tr>
<th>Concentration (ng/µl)</th>
<th>pH</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre</td>
</tr>
<tr>
<td>Negative 1</td>
<td>17.9</td>
</tr>
<tr>
<td>Negative 2</td>
<td>592</td>
</tr>
<tr>
<td>Positive 1</td>
<td>9.13</td>
</tr>
<tr>
<td>Positive 2</td>
<td></td>
</tr>
<tr>
<td>Positive 3</td>
<td></td>
</tr>
</tbody>
</table>

**TABLE V**

<table>
<thead>
<tr>
<th>Process</th>
<th>2P4M 0.35µm CMOS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Power Supply</td>
<td>3.3V</td>
</tr>
<tr>
<td>Area</td>
<td>Die Size</td>
</tr>
<tr>
<td>Array size</td>
<td>64 × 200 pixels</td>
</tr>
<tr>
<td>Array Area</td>
<td>1.06 mm²</td>
</tr>
<tr>
<td>Chemical sensing area per pixel</td>
<td>6.5 × 7.8 µm²</td>
</tr>
<tr>
<td>All blocks combined</td>
<td>1.9 mm²</td>
</tr>
</tbody>
</table>

**Passivation effects**

- Passivation Capacitance: 1.15 fF
- Capacitive attenuation factor (A): 0.42
- S_{13N4} Sensitivity: 24.1 mV/pH

**System Description**

- pH-current conversion: 1.03 µA/pH
- ADC Resolution: 10b (int), 16b (ext)
- Frame rate: 7.8 fps \( \pm 0.1 \) MHz clock
- Instrumentation current range: 0-200 µA
- Measured \( I_D \) noise: 104.28nA @ 1Hz
- pH resolution: 0.101 pH

![Current (µA) vs Time (s) for the DNA amplification experiment for the detection of the NDM gene in *E. coli*.](image)

(a) \( t = 2000s \)  (b) \( t = 12000s \)  (c) \( t = 20000s \)

Fig. 16. Distribution of active pixels along the timecourse of DNA amplification.
This paper presents a large-scale ISFET array which demonstrates novel operation in current-mode regarding the sensing front-end. By exploiting short-channel characteristics in standard CMOS, we show that the velocity saturation region of operation is suitable and beneficial for ISFET-based pH sensing since it linearly converts the small-signal pH variations to the drain current of the device. Furthermore, by investigating the short channel I-V characteristics we show that the linear region covers a very wide range of $V_{GS}$ voltages where the transconductance stays approximately flat. As a result, an ISFET in standard CMOS with appropriate biasing can serve as a front-end for pH sensing while being relatively insensitive to the effects of trapped charge (as long as the device is on) and without the need to further amplify the ISFET transconductance. Moreover, current-mode operation allows for a scalable pixel architecture which has been demonstrated as a 12.8K pixel array spanning approximately 1mm$^2$ silicon area. Compared to previous works, shown in Table VI, this is the only approach demonstrating an array in current-mode of the smallest footprint size and linear pH to current conversion.

Furthermore, this paper shows that the real-time rapid detection (within 15min) of genes associated to antibiotic resistance can be achieved with minimal equipment while bypassing the need for fluorescent tagging. As a result, this approach shows significant potential as a rapid point-of-care diagnostic test with demonstrated application in detecting infectious diseases. Moving forward, we anticipate that such a platform can leverage on the large resolution of current-mode arrays to facilitate the simultaneous detection of multiple samples on the passivation surface, paving the way for new applications in fully-electronic diagnostic technology.

ACKNOWLEDGMENT

The authors would like to thank the EPSRC Centre for Doctoral Training in High-Performance Embedded and Distributed Systems (HiPEDS) under Grant EP/L016796/1.

REFERENCES


Nicholas Missourides (S’15) received the M.Eng. degree in Electrical and Electronic Engineering in 2014 and the M.Res. degree in Advanced Computing in 2015 both from Imperial College London, UK. He is currently working towards the Ph.D. degree at the High-Performance Embedded and Distributed Systems (HiPEDS) EPSRC Centre for Doctoral Training and the Centre for Bio-inspired Technology, Department of Electrical and Electronic Engineering, Imperial College London, UK.

His research interests include mixed-signal IC design for integrated chemical sensing in CMOS and large-scale ion imaging. He was the recipient of the the Sir Bruce White Prize 2014 for the best M.Eng. project with a focus on analogue integrated circuit design.

Ling-Shan Yu is a visiting researcher at Centre for Bio-inspired Technology at Imperial College London. She received her B.S (Hons) in 2010 and PhD in 2016. She was awarded a Taiwanese scholarship in 2011 and an outstanding performance PhD student scholarship in the Department of Life Sciences at Imperial College London from 2013-2015. In 2017, she received a postdoc fellowship sponsored by the Taiwan and UK governments. Her areas of expertise are bioinformatics, phylogenetic analysis and developing isothermal chemistries for viral detections. Her latest research is focusing on developing point-of-care devices for detecting infectious diseases.

Jesus Rodriguez-Manzano received the B.Sc. degree in biological sciences and the Ph.D. degree in biotechnology and environmental microbiology from the University of Barcelona, Spain, in 2005 and 2012, respectively.

He is currently a Research Fellow with the Department of Electrical and Electronic Engineering, Centre for Bio-Inspired Technology, Imperial College London. His background in microbiology and molecular biology was followed by a three-year Postdoctoral position with the Division of Chemistry and Chemical Engineering, California Institute of Technology, where he was developing cutting edge Point-of-Care diagnostic technologies for infectious diseases. He is the author or co-author of 25 scientific publications and 3 international patents with more than 1200 citations. His research interest focuses on the interface of biology and engineering, including: development of low-cost sample preparation methods, innovative molecular tools for detection, quantification and typing of microbial pathogens, microfluidics, isothermal nucleic acid amplification chemistries, and digital single-molecule analysis.

Pantelis Georgiou (AM05SM08SM13) received the M.Eng. degree in electrical and electronic engineering and the Ph.D. degree from Imperial College London (ICL), London, U.K., in 2004 and 2008, respectively.

He is currently a Reader with the Department of Electrical and Electronic Engineering, ICL, where he is also the Head of the Bio-Inspired Metabolic Technology Laboratory, Centre for Bio-Inspired Technology. His research includes bio-inspired circuits and systems, CMOS based Lab-on-Chip technologies, and application of microelectronic technology to create novel medical devices. He has made significant contributions to integrated chemical-sensing systems in CMOS, conducting pioneering work on the development of ISFET sensors, which has enabled applications, such as point-of-care diagnostics and semiconductor genetic sequencing and has also developed the first bio-inspired artificial pancreas for treatment of Type I diabetes using the silicon-beta cell. He received the IET Mike Sergeant Medal of Outstanding Contribution to Engineering in 2013. In 2017, he was also awarded the IEEE Sensors Council Technical Achievement award. He is a member of the IET and serves on the BioCAS and Sensory Systems technical committees of the IEEE CAS Society. He is also the CAS representative on the IEEE Sensors council and the IEEE Distinguished Lecturer in Circuits and Systems.