ARTICLE

Unraveling the Impact of Nano-Scaling on Silicon Field-Effect Transistors for the Detection of Single-Molecules

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

Sybren Santermans *[a,b], Geert Hellings [a], Marc Heyns [a,b], Willem Van Roy [a], Koen Martens [a]

Electrolyte-gated Silicon Field-Effect Transistors (FETs) capable of detecting single molecules could enable high-throughput molecular sensing chips to advance, for example, genomics or proteomics. For solid-gated silicon FETs it is well-known that nano-scaled devices become sensitive to single elementary charges near the silicon-oxide interface. However, in electrolyte-gated FETs, electrolyte screening strongly reduces sensitivity to charges near the gate oxide. The question arises whether nano-scaling electrolyte-gated FETs can entail a sufficiently large signal-to-noise ratio (SNR) for the detection of single molecules. We enhanced a technology computer-aided design tool with electrolyte screening models to calculate the impact of the FET geometry on the single-molecule signal and FET noise. Our continuum FET model shows that a sufficiently large single-molecule SNR is only obtained when nano-scaling *all* FET channel dimensions. Moreover, we show that the expected $1/\sqrt{\text{surface area}}$ scaling trend of the single-molecule SNR breaks down and no longer results in improvements for geometries approaching the decananometer size. This is the characteristic size of the FET channel region modulated by a typical molecule. For gate lengths below 50 nm, the overlap of the modulated region with the highly conductive junctions leads to a saturation of SNR. For cross-sections below 10-30 nm, SNR degrades due to the overlap of the modulated region with the convex FET corners where a larger local gate capacitance reduces charge sensitivity. In our study, assuming a commercial solid-state FET noise amplitude, we find that a suspended nanowire FET architecture with 35 nm length and 5×10 nm^2 cross-section results in the highest SNR of about 10 for a 15-base DNA oligo in a 15 mM electrolyte. In contrast with typical silicon nanowire FET sensors which possess micron-scale gate lengths, we find it to be key that all channel dimensions are scaled down to the decananometer range.

Introduction

The ion-sensitive field-effect transistor (ISFET) developed by Bergveld initiated FET-based sensing in liquid environments.[1] Later the functionalization of the FET sensor surface allowed for the specific detection of target molecules and further boosted the interest in these devices for applications in the life sciences.[2,3] When a charged biomolecule binds to the FET sensor surface, the channel resistance, or equivalently, the threshold voltage V_T, is changed. These bioFETs enable the real-time and label-free detection of various kinds of therapeutically relevant molecules. FET sensors have manifested themselves in many different configurations to improve their performance. Especially nanoscale sensors gained a lot of interest to boost sensor sensitivity significantly.[2,4–7] Several key advantages can be expected from downscaling FET molecular sensors. First, from the scaling of Metal-Oxide-Semiconductor FETs (MOSFET) with a surface area A, it is known that the threshold shift associated with single oxide defects scales as $\sim 1/A$,[8,9] while the FET noise scales as $\sim 1/\sqrt{A}$.[10,11] Also, for electrolyte-gated nanoscale FETs it is pointed out that single oxide charges can become detectable.[12] Since the discrete charge nature of single molecules is comparable to single oxide defects, the single-molecule signal-to-noise ratio (SNR) is expected to scale with $\sim 1/\sqrt{A}$ as well.[13] This makes nanoscale FET sensors potential candidates for enabling single-molecule detection. However, in contrast to sensing oxide charges, sensing molecules is impacted by electrolyte screening which diminishes the signal and complicates the realization of single-molecule detection.[14] Second, nano-scaled silicon FETs can potentially provide a platform to do massively parallel high-throughput molecular sensing and imaging at densities and rates exceeding the stat-of-the-art enabled by the proven scalability and electrical properties of silicon FET technology.[15] An electrical platform that can perform massively parallel single-molecule sensing could entail major advances in various domains: quantitative molecular analysis,[16] DNA sequencing,[15,17] proteomics,[18,19] read-out of DNA memories[20], etc. Moreover, proven mass-manufacturability is a key enabler for applications. Mass-manufacturability is established for the nanoscale silicon FET

[a] *imec, Kapeldreef 75, 3001 Leuven, Belgium*
[b] *Department of Materials Engineering, University of Leuven, Kasteelpark Arenberg 44, 3001 Leuven, Belgium*

Electronic Supplementary Information (ESI) available: 1. Source-drain dopant profile. 2. Discussion on sensitivity metric. 3. Effective channel length. 4. FET short-channel effects. 5. Charge carrier concentration along channel length. 6. Impact of cross-section scaling on drain current. 7. Impact of geometry scaling on FET noise. 8. Impact of ionic strength on FET noise 9. Impact of gate length on cross-section scaling. 10. Impact of channel carrier mobility and model on geometry scaling. 11. Impact of pH sensitivity on geometry scaling trends. 12. Impact of recognition molecule on FET signal and geometry scaling trends. 13. TCAD simulation of single-molecule experiments. 14. Overview of model parameters

platform. However, single-molecule detection with mass-manufacturable silicon FET sensors is not yet established.

To analyze the impact of device scaling on the molecular signal, it is essential to distinguish between continuum charge sensing and single-molecule or discrete charge sensing. For bulk concentration sensing many molecules bind to the surface resulting in an approximately homogeneous or continuum charge density at the FET surface. Shoorideh et al. showed that for continuum charge sensing, FET downscaling only minorly impacts the molecular signal.[21] The molecular signal only varies with about 10% for FET widths ranging from the micrometer to the decananometer range. This small variation is attributed to increased or reduced electrolyte screening at convex or concave corners. We have experimentally confirmed that the molecular signal for continuum charge sensing is largely independent of device geometry using silicon FETs,[14] and our supporting simulations matched the experimental results. This means that for continuum charge sensing, downscaling FETs is even expected to decrease the SNR due to the increase in FET noise.[13] Only for discrete charge sensing, when sensing a single or a fixed number of molecules, we expect a significant signal and SNR boost for nano-scaled FET devices.

The limited literature work shows that detecting single molecules with silicon FETs remains an outstanding challenge. Sensitivity to a single molecule has thus far been reported by only two academic groups using bottom-up carbon nanotube (CNT) based devices with a typical diameter < 3 nm and several microns in length.[22–30] The small diameter and high mobility charge carriers of the CNTs endow these tubes with 1D quantum transport. For CNTs, an exceptional charge sensitivity is observed, even for micron-scale lengths. This is attributed to their ballistic transport properties.[31,32] Despite continuing progress, the integration issues of carbon nanotube FETs still limit their manufacturability such that silicon remains the material of choice for chips.

Work on electrical single-molecule detection with silicon FETs has been reported by the Guo group only. The group reports on the observation of discrete random telegraph signals in micron length bottom-up silicon nanowire FETs with diameters ranging from 20-40 nm.[33–37] The work ascribes these signals to single DNA hairpins,[34] binding of single protein complexes [35] as well as the real-time conformation changes of single protein complexes.[36,37] Random telegraph noise (RTN) due to interface traps can result in similar discrete signals. A thorough analysis of RTN with sufficient device statistics has not been reported in this work to unequivocally exclude traps as a cause of the signal.

In this paper, we present a detailed modeling study of the sensor geometry's influence on a discrete single-molecule charge signal. Technology computer-aided design (TCAD) simulation software specialized in simulating semiconductor devices allows for studying the geometrical dependence. Various papers have demonstrated the use of TCAD software to simulate ISFETs and molecular FET sensors including a continuum electrolyte ion distribution model.[38–40] Here, we use a 3D TCAD model that includes the Bikerman ion distribution model[41–43] and accounts for steric effects.[42] This framework has previously been used to quantitatively describe the measured signal for a layer of DNA molecules binding to a silicon FET sensor.[44] With this study, we seek to understand and specify the geometrical requirements to achieve the single-molecule detection limit with silicon FET sensors. More specifically, we show the impact of the geometrical scaling of silicon FET sensors on the single-molecule SNR and identify the underlying physical mechanisms explaining the geometrical trends. We first discuss the discrete nature of the single-molecule charge modulation and the FET models that are used. Subsequently, we investigate the impact of FET gate length scaling on the single-molecule signal. Afterward, the impact of the FET width and height on the single-molecule signal is shown. The FET architecture's impact on scaling is also studied. Finally, we simulate FET sensor noise calibrated with experimental noise data to establish the potential single-molecule SNR and determine the feasibility of single-molecule detection with silicon FETs.

Results and discussion

Single-molecule FET sensor model

To study the impact of the FET geometry on single-molecule sensing both semiconductor transport in the FET channel and electrolyte charge screening must be accounted for. In the FET channel, the typical drift-diffusion equation is solved including density gradient quantum corrections which adjust the carrier density at the silicon-oxide interface.[45] Since the impact of single oxide traps on typical MOSFETs can be described without considering charge-induced scattering,[46,47] we do not include these effects in our model. Random discrete dopant effects are not considered as well.[46,48] Discrete dopant effects induce variability in the single-molecule signal but the average signal is expected to be approximately in line with a continuous model.[49] The electrolyte ion distribution giving rising to electrolyte screening is described by the Bikerman model. This model accounts for steric effects when high ion concentrations arise.[41–43] To identify the underlying physical mechanisms that explain the influence of the FET geometry on the single-molecule SNR, we initially omit the presence of additional phenomena like the typical pH sensitivity of a charged oxide surface, the dynamic behavior of molecules, a functional layer, etc. At a later stage, we specifically treat the impact of oxide surface charges on the geometry dependence of the single-molecule signal.[44,50,51] Since the pH sensitivity does not significantly influence the geometric scaling trends, it is only discussed in the supplementary information (section 10).

We use a positive singularly charged dielectric cube of 1 nm^3 as our biomolecule model system. This simple model system can be considered as a building block of more complex, larger and multiply charged molecules like for example DNA which is considered at the end of the paper. Also, as a small building block, it will be even more difficult to detect than a larger charged molecule such as DNA, and hence will be a more demanding detection criterium.

A single molecule is positioned in the middle of the FET sensor at a distance of 1 nm above the gate oxide (Figure 1a). For simplicity, we neglect the presence of a functional self-assembled monolayer (SAM) which hinders the presence of electrolyte ions close to the surface but is not expected to affect the geometric scaling trends. As such the 1 nm molecule distance corresponds to a realistic 1 nm linker length.[34,35,44] This setup corresponds to the direct binding of a single molecule to the FET sensor surface. In practice, recognition molecules are typically required to add selectivity to the FET sensor. For DNA FET sensors, complementary peptide nucleic acid (PNA) or DNA probes can be used to bind the target DNA molecule at a linker length distance from the sensor surface. Here, PNA probes are preferred over charged single-stranded DNA probes to limit additional nonlinear screening (see SI section 11). Other kinds of target molecules like proteins might require other recognition molecules of different sizes such as antibodies, aptamers or nanobodies resulting in different distances of the target molecule to the FET surface. Small recognition molecules which minimize the distance to the surface are preferred. However, the target molecule distance does not significantly affect the optimal FET design (see SI section 11).

The signal due to the singularly charged molecule is calculated in the subthreshold regime where the channel charge carrier density is low and semiconductor charge screening is small. Previous reports identified the subthreshold regime as the optimal regime for molecular charge sensing.[52,53]

We consider p-type inversion mode devices with a conducting channel of positive charge carriers, i.e. holes. We evaluate three different FET architectures: tri-gate (Figure 1ab), gate-all-around suspended (Figure 1c) and single-gate embedded (Figure 1d) devices. The architectures differ in the number of surfaces that are exposed to the electrolyte.

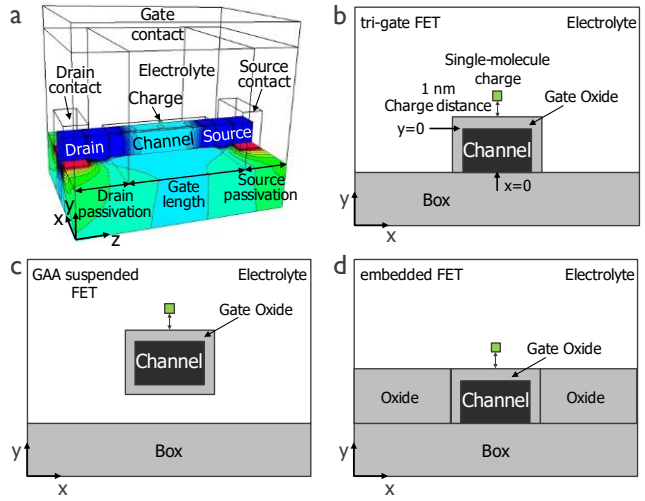

Figure 1: a) Structure of the simulated tri-gate FET sensor with the potential distribution across the FET channel in the presence of a single-molecule charge. Gate, source and drain contacts are indicated. Oxide source and drain passivation shield the electrolyte segment from the source-drain contacts. Channel source-drain junctions described by a Gaussian diffusion profile are situated at the interface between the source-drain passivation and electrolyte segment (see SI section 1). b, c, and d) Cross sections for the tri-gate, suspended and embedded FET. The origin of the coordinate system is situated in the middle of the channel (x = 0) at the channel top surface (y = 0).

Single-molecule modulation

The singularly charged molecule induces mirror charges in its environment. These mirror charges arise in the FET channel as well as in the electrolyte and consist of a change in the distribution of FET channel charge carriers and electrolyte ions respectively. In Figure 2a, we show that the presence of these mirror charges induces a local modulation of the potential (ΔΨ) that extends in both the FET channel and electrolyte. Unfortunately, a major part of the charge and potential modulation occurs in the electrolyte. This means that the molecule charge is dominantly mirrored by electrolyte ions present in the well-known electrolyte double layer.[54] This effect is commonly referred to as electrolyte screening and limits FET sensitivity. The ionic strength $\mathrm{I} = 1/2 \sum_{i=1}^{n} c_i z_i^2$ determines electrolyte screening with c_i and z_i the ion concentration and valency for the i[th] ion in an electrolyte with n ions. As expected, for a lower ionic strength I, the double layer characterized by the Debye length ($\lambda_d \sim \sqrt{1/\mathrm{I}}$) becomes thicker, and charge screening is reduced. Consequently, with a low ionic strength electrolyte, the magnitude of the potential modulation increases, but the modulation also extends over a larger volume (Figure 2).

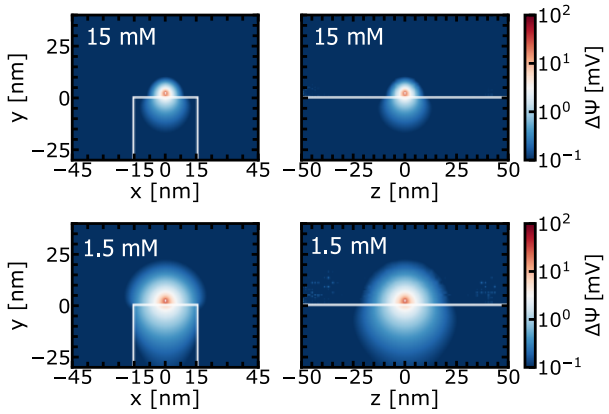

Figure 2: Local nature of the modulation by a singularly charged molecule in the electrolyte, with the molecule located 1 nm above the gate oxide of a tri-gate FET. Potential modulation (ΔΨ) in cross-sections perpendicular to the channel (left) and along the channel (right). Electrolyte ionic strength is 15 mM (top) and 1.5 mM (bottom). The tri-gate FET is operated in the subthreshold regime at 100 mV underdrive: V_{GS}-V_T = 100 mV. The drain voltage V_D = 0 V and the source voltage V_S = 50 mV.

In Figure 3a, we show the initial current density distributions in the FET channel which are identical for an electrolyte ionic strength of 15 and 1.5 mM. This denotes the negligible impact of the electrolyte ionic strength on FET gating. If a singularly charged molecule is present, the local mirror charge in the FET channel directly modulates the local channel resistivity and alters the current density along the channel (Figure 3b). Because of current continuity, the current density does not only change locally but also changes at the source and drain regions further away from the molecule charge. The increase in magnitude and volume of the local current density modulation for reduced electrolyte screening agrees with the larger magnitude and volume of the local channel mirror charge. The volume of the mirrored molecule charge or modulated potential depends on parameters such as molecule charge and

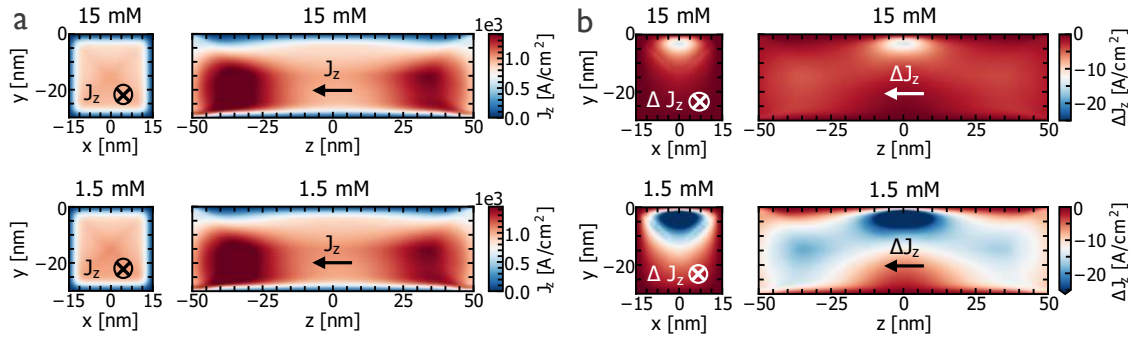

Figure 3: a) p-type tri-gate FET channel current density (J_z) in cross-sections perpendicular to the channel (left) and along the channel (right). The direction of the current density J_z is indicated. b) Modulation of the FET channel current density (ΔJ_z) in cross-sections perpendicular to the channel (left) and along the channel (right). The positive direction of the current density modulation ΔJ_z is indicated. The physical gate length of the tri-gate FET is 100 nm and the channel cross-section is 30×30 nm^2. Electrolyte ionic strengths I are 1.5 mM and 15 mM. The FET is operated in the subthreshold regime at 100 mV underdrive. V_D = 0 V and V_S = 50 mV.

position, electrolyte ionic strength, oxide surface charge, etc, and will play a crucial role in understanding the impact of the geometrical parameters on the single-molecule signal.

The channel drain current modulation ΔI_{DS} caused by the singularly charged molecule directly results in a threshold voltage shift ΔV of the FET I_{DS}-V_{GS} curve. We will use this threshold voltage shift as our single-molecule sensitivity metric S. For a singular charge at room temperature, the ΔV is small compared to the thermal voltage $kT/q \approx 25$ mV. Consequently, the I_{DS}-V_{GS} curve can be linearized and the ΔV becomes proportional to the relative drain current change $S = \Delta V \propto \Delta I_{DS}/I_{DS}$ (SI section 2). Due to the small ΔV, the single-molecule sensitivity also approximately scales with the ratio of the modulated channel resistance and the total resistance of the FET channel $S = \Delta V \propto \Delta R/R$. These alternative expressions will be used when we discuss the impact of the FET geometry on single-molecule sensitivity.

FET gate length scaling

Along the gate length direction, the FET channel can be considered as a series connection of local channel resistances (Figure 4). Therefore, to explain the impact of gate length scaling we use the following expression for the single-molecule sensitivity $S = \Delta V \propto \Delta R/R$. This expression directly shows that the single-molecule sensitivity can benefit from downscaling the physical gate length (L_{gate}) since $R \propto L_{gate}$. If the resistance modulation ΔR is not impacted by gate length downscaling, an inversely proportional trend between the single-molecule sensitivity and the gate length is expected.

In Figure 4, we show the simulated singularly charged molecule signals for tri-gate FETs (Figure 1b) with 10 nm and 30 nm square cross-sections as a function of the physical gate length. The signal increases by more than an order of magnitude when downscaling the gate length from 1 µm to tens of nanometers. For long gate lengths, the signal $S = \Delta V \propto \Delta R/R$ is inversely proportional to the gate length, which indicates that ΔR does not change with gate length. The reduction of the channel series resistance thus boosts the molecule signal.

For gate lengths below 200 nm, the signal behavior deviates from the $1/L_{gate}$ trend First, the increase in signal exceeds the $1/L_{gate}$ trend with a further reduction of the gate length. This behavior is related to a difference between the physical gate length and the effective channel length. The penetration of holes from the highly doped source-drain regions into the channel makes the physical gate length an improper measure for the total channel resistance R. Second, for even smaller gate lengths, the signal saturates. This saturation occurs when the region where the channel charge is modulated by the molecule starts to overlap with the region affected by the source-drain junctions.

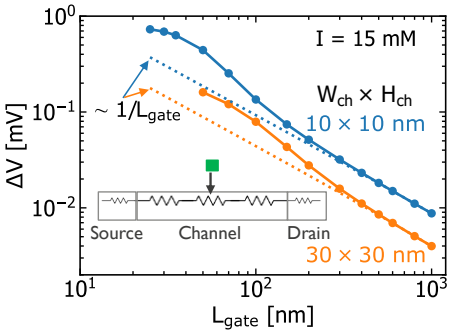

Figure 4: The impact of physical gate length (L_{gate}) scaling on the singularly charged molecule signal ΔV for the tri-gate FET sensor with cross-sections of 10x10 nm^2 and 30x30 nm^2. The dotted lines show the $1/L_{gate}$ trend expected from a series resistance effect. Inset: Representation of the FET channel as a series connection of resistances of which only the local resistance is affected by the single molecule.

To obtain a better metric for the total channel resistance compared to the physical gate length L_{gate}, we now consider the effective channel length L_{eff} (SI section 3). This effective channel length includes the effect that holes from the source-drain regions penetrate the channel and locally reduce the channel resistance. The depth of hole penetration is solely determined by FET electrostatics and does not depend on the electrolyte environment. Since the channel can be considered as a series connection of resistances, the change of the total channel resistance and thus the signal is low when a highly conductive location of the channel is modulated i.e. at the source-drain junctions. This effect is accounted for when using the effective channel length. In Figure 5a, we plot the signal for a singularly charged molecule in the center as a function of this effective channel length. This figure now shows the inversely

proportional trend as expected for the series resistance effect down to much shorter lengths.

However, the singularly charged molecule signal still starts to saturate for small effective channel lengths (< 20 nm) or gate lengths (< 50 nm). Moreover, the lower the ionic strength I of the electrolyte, the earlier and the more significant the saturation of the signal ΔV becomes (Figure 5b). To understand the origin of the saturation of the molecule signal for small effective channel lengths, we must consider electrolyte charge screening. Two electrolyte screening effects contribute to the earlier saturation of gate length downscaling for lower ionic strengths.

First, when the electrolyte ionic strength reduces, the double layer formed at the molecule charge penetrates farther into the electrolyte (Figure 2) and a larger region of the channel is modulated. As explained earlier a modulation that overlaps with the lowly-resistive source-drain regions minorly contributes to the signal $\Delta V \propto \Delta R/R$. Consequently, a molecule that modulates a larger region of the channel results in an earlier saturation of the signal for downscaled channel lengths. This effect is visualized by the longitudinal position dependence of the trap and molecule signals along the channel (Figure 5c). When moving the charge from the center to the source-drain junctions, the signal drops first for the charge that induces the largest modulated region, i.e. a molecule in low ionic strength followed by a molecule in high ionic strength and eventually a trap.

Second, a larger local ionic strength I_{local} at the source-drain junctions leads to locally enhanced screening and reduces the local charge sensitivity (Figure 5d). This effect thus also contributes to the saturation of the molecule signal for small channel lengths. In low ionic strength electrolytes, the relative enhancement of electrolyte charge screening at the source-drain regions is stronger. Moreover, at low ionic strength, the double layers formed at both the molecule charge and the source-drain junctions penetrate farther into the electrolyte. Therefore, in low ionic strength electrolytes, enhanced molecule charge screening already occurs further away from the source-drain junctions and the molecule signal saturates for larger gate lengths.

Because of the negative impact of the source-drain junctions, the singularly charged molecule signal in an electrolyte solution with ionic strength of 1.5 mM does not significantly improve by downscaling the physical gate length beyond 50 nm or equivalently the effective channel length beyond 15 nm (Figure 5b). Notice that for a 1.5 mM electrolyte with a Debye length of ~7 nm the molecule modulates a region of ~15 nm. For a 10x10 nm^2 FET, a maximum signal of about 1 mV is obtained.

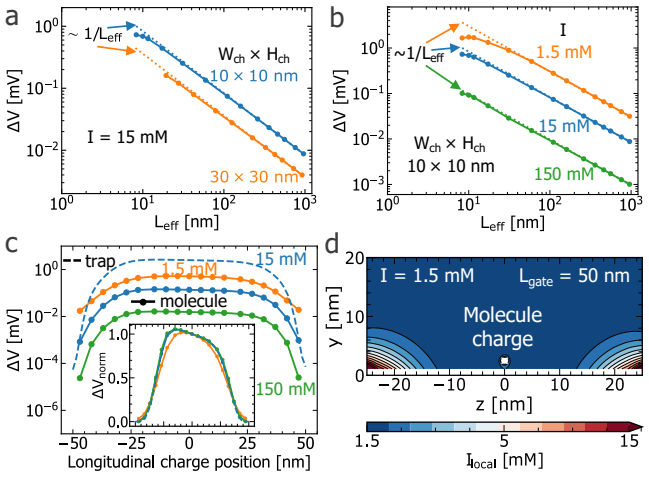

Figure 5: a) The singularly charged molecule signal (ΔV) as a function of L_{eff} for tri-gate FETs with cross-sections of 10x10 nm^2 and 30x30 nm^2. The singularly charged molecule is positioned 1 nm above the gate oxide in the center of the electrolyte. The dotted lines show the $1/L_{eff}$ trend. b) The singularly charged molecule signals as a function of L_{eff} for different electrolyte ionic strengths. c) The longitudinal position dependence of the trap (dashed) and molecule (dots) signal. Colors indicate different ionic strengths. The FET cross-section is 10x10 nm^2 and the gate length is 100 nm. The inset shows the signals normalized to the signal in the center. d) The distribution of the ionic strength (I_{local}) in an electrolyte with 1.5 mM bulk ionic strength. The gate length is 50 nm.

Cross-section scaling

The FET cross-section with a channel width W_{ch} and height H_{ch} can be considered as a parallel connection of local channel resistances. Therefore, we now consider the single-molecule sensitivity as a ratio of the drain current modulation and the initial drain current $S = \Delta V \propto \Delta I_{DS}/I_{DS}$. For large cross-sections, it is expected that cross-section downscaling does not significantly change the drain current modulation ΔI_{DS}. Consequently, we expect the sensitivity to scale with $\sim 1/I_{DS}$. However, when the size of the cross-section approaches the size of the region modulated by the molecule (Figure 2), an impact on the drain current modulation is likely to occur.

We elaborate further on the cross-section dependence of the single-molecule signal to determine the optimal FET width and height that maximize the single-molecule sensitivity. A long gate length of 600 nm is selected to limit short-channel effects. This allows for studying a wider range of cross-section dimensions. Also, we show in the supplementary information section 9 that the gate length does not significantly impact the cross-section scaling trends. We assume a singularly charged molecule located 1 nm above the gate oxide at the center of the FET channel. We first cover the width and height dependence for the tri-gate FET and subsequently discuss the other FET architectures.

The drain currents for 600 nm long tri-gate FETs as a function of FET width and height are shown in Figure 6ab. We observe that the drain current for a fixed height and width scales approximately with $W_{ch}+2H_{ch}$ as a function of the channel width and height respectively. For a tri-gate FET, $W_{ch}+2H_{ch}$ is the area exposed to the electrolyte and represents the gated surface of the FET channel. However, a general $W_{ch}+2H_{ch}$ relationship cannot describe the variation of the drain current across all

geometries (SI section 6). Based on the drain current scaling trend, downscaling the channel height is thus expected to give a larger boost in the single-molecule signal compared to width downscaling.

The drain current modulation ΔI_{DS} as a function of channel width and height is shown in Figure 6cd. For width downscaling (Figure 6c), the drain current modulation is constant for large channel widths. It becomes slightly larger for moderate widths (< 100 nm) but decreases strongly for small channel widths (< 50 nm). Moreover, for lower FET heights, the drain current modulation is larger (Figure 6d) and the drop in the drain current modulation for narrow channels is reduced (Figure 6c). For channel height downscaling, the drain current modulation is again constant for large heights but starts to increase for small heights (< 50 nm). A physical interpretation of these trends is given later.

The molecule signals as a function of the channel width and height directly result from the drain current I_{DS} and drain current modulation trends ΔI_{DS} (Figure 6ef). For large channel widths, the signal $\Delta V \propto \Delta I_{DS}/I_{DS}$ scales approximately with the inverse of the FET drain current because of a constant drain current modulation. However, if the channel width is further decreased, the increase in the single-molecule signal saturates (< 70 nm) to reach an optimum at moderate channel widths (W_{ch} ~= 10-30 nm). Eventually, for even smaller channel widths, the signal decreases due to the strong drop in the drain current modulation. The decrease in the signal for these small channel widths is less outspoken for smaller channel heights. Therefore, lower channel heights lead to smaller optimal channel widths and an increased optimal molecule signal. In contrast to width scaling, the signal (Figure 6f) exceeds the $1/I_{DS}$ trend for channel heights below ~50 nm. This directly follows from the increase in the drain current modulation. The saturation of the signal for channel heights below 10-20 nm is expected from the saturation of the $1/I_{DS}$ trend due to the fixed width. A boost in singularly charged molecule signal of about two orders of magnitude can be obtained when the channel width and height are scaled from 400 nm to the 10 nm range.

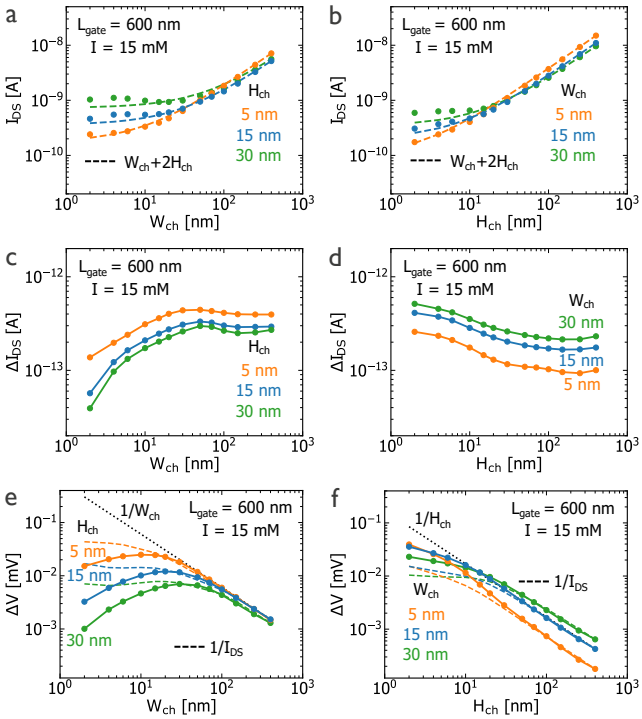

Figure 6: a and b) The drain current I_{DS} as a function of channel width (W_{ch}) and height (H_{ch}) for tri-gate FET sensors. For each series, dashed lines represent the W_{ch}+2H_{ch} trends normalized towards the largest width in panel a) and largest height in panel b). c and d) The drain current modulation ΔI_{DS} as a function of channel height and width. e and f) The impact of width and height downscaling on the singularly charged molecule signal $\Delta V \propto \Delta I_{DS}/I_{DS}$. The $1/I_{DS}$ trends are represented with a dashed line. The dotted lines indicate the inversely proportional trend with the FET channel width and height respectively. Colors indicate different channel heights in panels a,c,d and different channel widths in panels b,d,f. The gate length is 600 nm and the ionic strength is 15 mM. The molecule is positioned 1 nm above the gate oxide in the center of the electrolyte.

To understand the non-constant drain current modulation, we study the drain current modulation ΔI_{DS} and threshold voltage shift ΔV caused by a singularly charged trap at the silicon-oxide interface (Figure 7). This approach allows distinguishing between the well-known FET electrostatics which determines the trap signal and electrolyte screening effects.

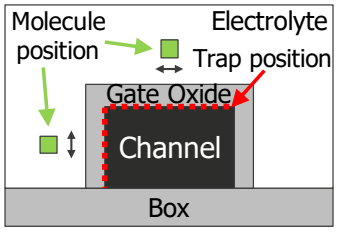

Figure 7: a) Position of the single trap at the top silicon-gate oxide interface and singularly charged molecule above the top gate oxide surface in the electrolyte.

Since we do not assume changes in mobility, the drain current modulation ΔI_{DS} is directly proportional to the amount of charge that is mirrored in the FET channel. From FET electrostatics, the fraction of the trap charge that is mirrored in the inversion layer of the channel can be written as:

$$f = C_{inv}/(C_{inv} + C_d + C_g)$$

with C_{inv}, C_d and C_g the capacitances of the inversion layer, depletion layer and gate respectively.[55] The gate capacitance consists of the oxide capacitance C_{ox} and electrolyte double layer capacitance C_{DL} in series: $1/C_g = 1/C_{ox} + 1/C_{DL}$. Here, the oxide capacitance is typically the lowest and therefore dominates the total gate capacitance (SI section 8). For a uniform planar FET, the channel and gate capacitances are constant across the device width. However, for a tri-gate FET, this is not valid. Therefore, we must consider the variation of these capacitances across the FET perimeter.

First, when downscaling the FET width and height from about 100 to 50 nm the current density or equivalently the carrier density increases at both the top and side surfaces of the FET (SI section 6). This leads to a larger inversion layer capacitance.[56] As a result, a larger fraction of the trap charge is mirrored in the channel and the drain current modulation increases (Figure 8ab).

When downscaling the FET width further, we must consider that the convex FET corners locally increase the gate capacitance and locally decrease the channel capacitances.[21] Therefore, at the FET corners the fraction of the trap charge mirrored in the channel f decreases and a larger fraction is screened in the electrolyte gate. Consequently, the drain current modulation (Figure 8ab) and the trap signal $\Delta V \propto \Delta I_{DS}/I_{DS}$ drop at the FET corners (Figure 8cd). For small FET widths (< 50 nm), this corner effect starts to influence the trap signal at the center (fig. 8ac). The drop in the trap signal for FET width downscaling (Figure 9a) is thus expected from the increased impact of the FET corners.

For height scaling, the impact of the FET corners does not increase for a molecule in the center of the top surface. On the other hand, the current density at the top surface becomes larger for FET heights below 50 nm (SI section 6). Consequently, at the top surface, the inversion layer capacitance and the fraction of mirrored charge in the FET channel increase. For small FET heights, this directly results in a larger drain current modulation and the increase of the trap signal beyond the $1/I_{DS}$ trend (Figure 9b).

Furthermore, notice that the presence of the oxide passivation at the bottom surface of the FET reduces the local gate capacitance and counteracts the corner effect (Figure 8bd). Therefore, the drain current modulation and trap signal at the concave bottom corners are larger compared to those at the convex top corners (see also SI figure S11).

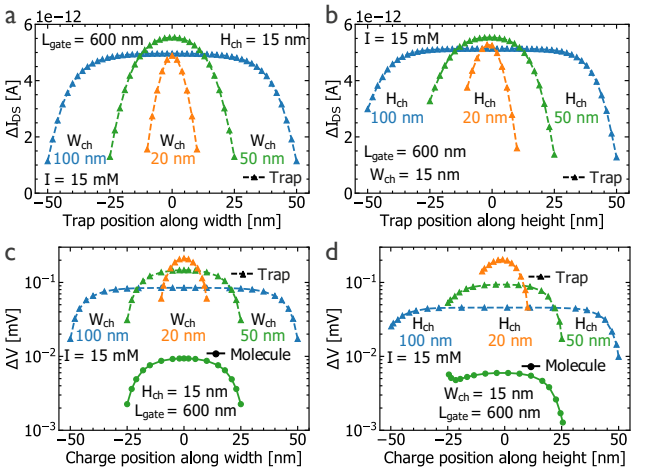

Figure 8: a,c) The positional dependence of the drain current modulation and signal for a single trap (triangle dashed) and molecule (circle) at the top of the channel. Colors indicate different channel widths. b,d) The positional dependence of the drain current modulation and signal for a single trap (triangle) and molecule (circle) at the side of the channel. Colors indicate different channel heights. In a and b the separation between the markers is 2 nm.

Next, to evaluate the impact of the electrolyte on the corner effect, we compare the position dependence and the scaling trends of the molecule and trap signal (Figure 8cd). The positional dependence of the trap and molecule signal only differs when the trap/molecule is situated at the concave FET bottom corner between the gate and bottom oxide. At this bottom corner, the molecule signal increases relative to the trap signal (Figure 8d). This is explained by the concave nature of the bottom corner that locally reduces electrolyte screening.[21] The largely similar position dependence of the trap and molecule signal shows that FET electrostatics dominates the observed trends in the FET signal variation across the FET surface. This implies that molecule charge screening by the electrolyte is not the origin of the corner effect that results in the reduction of the drain current modulation for small FET widths.

Furthermore, we study the impact of the electrolyte ionic strength on the scaling trends. We show the normalized molecule signals for different ionic strengths and the normalized trap signal as a function of FET width and height (Figure 9). Normalization is done w.r.t the largest FET width and height respectively. Differences between the scaling trends can be explained by differences in the size of the modulated region. The size of the modulated region increases in the following order: a trap, a molecule in high ionic strength and a molecule in low ionic strength electrolyte. The larger the size of the modulated region the earlier the molecule signal is impacted by the FET corners. Therefore, when scaling down the width, the molecule signal in a low ionic strength electrolyte already saturates at larger FET widths and a larger optimal FET width is obtained (Figure 9a). For FET height downscaling there is a minor effect of the FET corners. Consequently, a minor impact of the ionic strength on the saturation of the molecule signal occurs.

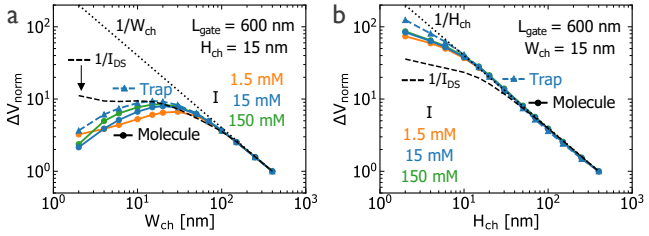

Figure 9: a) and b) The normalized signal for a single trap (triangle dashed) and singularly charged molecule (circle) as a function of FET channel width and height respectively. The trap charge is positioned at the center of the top silicon-gate oxide interface. The molecule charge is situated above the center of the channel in the electrolyte. Colors indicate different electrolyte ionic strengths I. The signals are normalized towards the largest device. The $1/I_{DS}$ trends are represented with a dashed line. The dotted lines indicate the inversely proportional trend with the FET channel width and height.

Not only an increase in ionic strength increases electrolyte screening. Also, charged oxide surfaces increase the local electrolyte screening strength due to nonlinear electrolyte screening.[57,58] Consequently, large oxide surface charges strongly reduce the FET molecule signal (Figure 10). Moreover, enhanced electrolyte screening at the charged oxide surface reduces the size of the modulated region. As such the FET corners have a smaller impact and the optimal FET width is somewhat reduced (Figure 10).

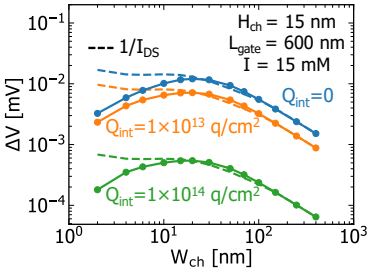

Figure 10: The impact of the oxide surface charge density (Q_{int}) on the width-dependent singularly charged molecule sensitivity ΔV of the tri-gate FET. Dashed lines represent the $1/I_{DS}$ trends normalized to the largest channel width.

Finally, we show the impact of the different FET architectures (tri-gate, suspended, embedded see Figure 1) on the cross-section dependence of the molecule signal (Figure 11a). In contrast to the tri-gate FET for which the bottom surface is passivated, all surfaces of the suspended FET are gated by the electrolyte. For wide suspended FETs, this leads to a larger drain current compared to the tri-gate FET (SI section 6). Since, the impact of the corner effect on the drain current modulation is similar for the suspended and the tri-gate FET (Figure 11b), the signal $\Delta V \propto \Delta I_{DS}/I_{DS}$ is smaller for the suspended FET (Figure 11a). For small FET widths, the suspended and tri-gate FET architectures become equivalent, and the molecule signals become equal (Figure 11a). Width downscaling thus gives a stronger signal boost for the suspended FET and results in a smaller optimal width for the suspended compared to the tri-gate FET. However, at the optimal FET width, the single-molecule signal reaches a higher value for the tri-gate compared to the suspended FET.

The difference between the embedded and the tri-gate FET is the passivation of the sidewalls (Figure 1). For large FET widths, the sidewall passivation minorly impacts the FET characteristics. This results in a similar molecule signal for both FET architectures (Figure 11a). When the FET width is scaled down, the signal of the embedded FET increases compared to the tri-gate FET. First, this is explained by the lower drain current for the embedded FET (SI section 6). Second, the sidewall passivation reduces the impact of the FET corners relative to the tri-gate FET. The signal reduces less at the FET corners and leads to a more uniform molecule signal across the FET surface (Figure 11b). The reduced corner effect makes the embedded architecture the most sensitive and results in a small optimal FET width of about 5 nm.

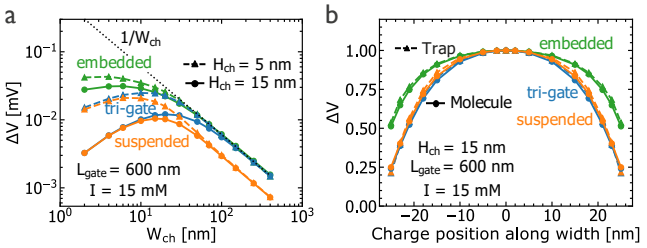

Figure 11: a) The singularly charged molecule signal as a function of FET width for different FET architectures with FET heights of 5 and 15 nm. b) The positional and architecture dependence of the FET normalized signal towards a single trap at the top silicon-gate oxide interface (triangle dashed) and a singularly charged molecule (circle) above the channel in the electrolyte. The signals are normalized towards the maximum FET signal. Colors indicate different FET architectures.

Single-molecule signal-to-noise ratio

The feasibility of sensing single molecules with silicon FET sensors is determined by the single-molecule signal-to-noise ratio (SNR). To calculate the SNR = $\Delta V/V_{noise}$, FET sensor noise must be considered. Several studies showed that noise in liquid-gated FET sensors is not significantly influenced by the liquid environment but is dominated by typical FET 1/f noise.[11,13,59,60] 1/f noise in FETs originates from the charging and discharging of many single oxide traps near the silicon-gate oxide interface.[55,61,62] For downscaled FETs the impact of individual traps can be observed and models are introduced that describe FET noise based on the ensemble impact of individual traps.

Banaszeski et al.[55] used such a framework to model the expected value of the noise power spectral density (PSD). The expected gate voltage-referred noise power spectral density reads as:

$$\mathrm{E}[S_{Vg}(f)] = \frac{kT}{\gamma f {g_m}^2}\mathrm{E}[N_{T,tot}]\mathrm{E}[\Delta {I_{tr}}^2 f_{E_T}(E_F)]$$

Here, $\gamma = \ln(\bar{\tau}_{max}/\bar{\tau}_{min})$ where $[\bar{\tau}_{min}, \bar{\tau}_{max}]$ indicates the observation window of the trap time constant $\bar{\tau}$. $\bar{\tau}$ is assumed to be a log uniformly distributed random variable. k, T and f represent the Boltzmann constant, the temperature, and the frequency respectively. Next, g_m is the FET transconductance which translates drain current-referred FET noise to a gate voltage-referred value. $\mathrm{E}[N_{T,tot}]$ represents the expected total number of traps in the gate oxide. Finally, ΔI_{tr} is the drain current shift induced by a single trap at the silicon-oxide interface and $f_{E_T}(E_F)$ is the probability density function of the trap energy evaluated at the channel Fermi level E_F (eV^{-1}).

Assuming a uniform spatial trap density along the silicon-oxide interfaces, $\mathrm{E}[S_{Vg}(f)]$ can be expressed as an integral over the entire silicon-oxide interface area A = 2×(W_{ch}+H_{ch})×L_{gate}:

$$\mathrm{E}[S_{Vg}(f)] = \frac{kT}{f {g_m}^2} \int \Delta {I_{tr}}^2 N_{tr}(E_F)\, dA$$

Here, $N_{tr}(E_F) = N'_{tr} f_{E_T}(E_F)/\gamma$ is the trap density per area and energy in the log domain of time constants ($cm^{-2}eV^{-1}$). Notice that here N'_{tr} is merely the spatially trap density (cm^{-2}). However, the trap density is determined by the fabrication process and can consequently be obtained from experimental calibration only. Therefore, we introduce a normalized trap density $N_{tr,norm}(E)$ and a constant calibration/scaling factor s such that $N_{tr}(E_F) = sN_{tr,norm}(E_F)$. The normalized trap density $N_{tr,norm}(E)$ includes the typical energy dependence observed for p-channel FETs with a SiO_2 gate oxide. More specifically, $N_{tr,norm}(E)$ equals unity at the silicon valence band edge energy and increases for decreasing trap energies (see methods).[55] The constant scaling factor s not only corrects for the real spatial trap density but also corrects for the variation of the trap's position along the thickness of the gate oxide. Remember that we assume that ΔI_{tr} is due to traps at silicon-oxide interface. As such this scaling factor rescales the normalized trap density to match the quality of the silicon-oxide interface of the used technology.

The scaling factor s is calibrated based on the measured noise power at 1 Hz of our 30×30×1000 nm^3 liquid-gated FETs.[63] We calculate the expected value of the noise amplitude V_{noise} by integrating $S_{Vg}(f)$ over a 1 to 1000 Hz frequency range:

$$V_{noise} = \sqrt{\frac{kT}{{g_m}^2} s \int \Delta {I_{tr}}^2 N_{tr,norm}(E_F)\, dA \times [\ln 1000 - \ln 1]}$$

The noise amplitudes for the different FET architectures as a function of the FET width and length are shown in Figure 12ab. The noise amplitudes are the smallest for the suspended FET and increase for the tri-gate FET and the embedded FET. This trend is due to the larger threshold voltage shift caused by traps that are situated at the oxide passivated surfaces where the local gate capacitance is lower $\Delta V_{tr} = \Delta I_{tr}/g_m \sim q/C_g$. The larger threshold voltage shift for traps at the passivated surfaces also leads to different noise trends as a function of the FET geometry for the different FET architectures (Figure 12a). For the suspended FET, the expected $1/\sqrt{A}$ trend occurs. For the embedded FET with larger FET heights, the noise increases beyond $1/\sqrt{A}$ due to the increased impact of the traps at the passivated sidewalls (SI section 7). For the tri-gate FET, the noise increases by less than $1/\sqrt{A}$ and even starts to decrease for small widths (< 50 nm). This is a direct result of the reduced contribution of the passivated bottom surface where $\Delta V_{tr} \sim q/C_g$ is large w.r.t. the total surface area. For all architectures, the noise increases for very small FET widths (< 5 nm). This follows from an increase in the trap density $N_{tr}(E_F)$ since band bending in the channel becomes stronger i.e. E_F decreases (SI section 7).

The noise amplitude as a function of gate length follows the expected $1/\sqrt{A}$ or $1/\sqrt{\mathrm{L_{gate}}}$ trend for gate lengths larger than 200 nm (Figure 12b). For smaller gate lengths, the noise amplitude exceeds the $1/\sqrt{A}$ trend. Again, this follows from an overall decrease in the channel Fermi level relative to silicon midgap increasing the trap density for p-type FETs. The overall decrease in Fermi level w.r.t. midgap for small FET gate lengths, originates from the increasing contribution of traps close to the source-drain regions where band bending is stronger (SI section 7).

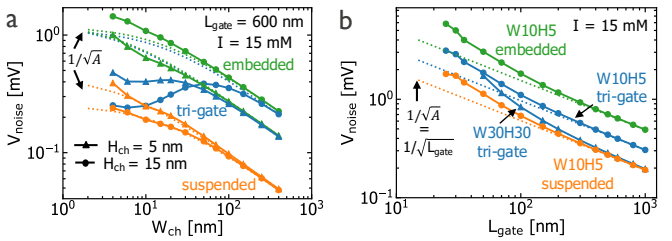

Figure 12: a) The FET noise amplitude as a function of width for different FET architectures with heights of 5 and 15 nm. b) The noise amplitude as a function of FET gate length for a tri-gate FET with 30×30 nm^2 cross-section and FETs with optimal cross-section W_{ch}×H_{ch} = 10×5 nm^2. The ionic strength is 15 mM. Colors indicate different FET architectures. The expected $1/\sqrt{A}$ trend is indicated with A = 2×(W_{ch}+H_{ch})×L_{gate} the FET interface area. The noise bandwidth is 1-1000 Hz.

Both the single-molecule signal and noise amplitude have now been discussed. To determine the single-molecule SNR, we assume that the surface in contact with the electrolyte is uncharged and pH insensitive. Such a well-passivated oxide surface limits additional nonlinear screening and the pH interference effect, which are typically present at realistic oxide surfaces and can strongly reduce the FET sensitivity.[44,50] Recent studies however showed that by tuning the electrolyte pH[44] and/or blocking the oxide surface groups,[51] these sensitivity-limiting effects can be curtailed.

The width and gate length dependence of the singularly charged molecule SNR for the different FET architectures is shown in Figure 13ab. In the supplementary information section 9, we show that the SNR cross-section scaling trends do not significantly change for short channel lengths. The suspended FET has the highest SNR followed by the tri-gate FET and eventually the embedded FET. This is the reverse order compared to the singularly charged molecule signal and follows from the increased noise amplitude for the tri-gate and the embedded FET compared to the suspended FET. For the three different FET architectures, the increase in SNR somewhat exceeds the $1/\sqrt{\mathrm{W_{ch}}}$ dependency when scaling FET widths down to 50 nm which is caused by the stronger increase in signal ~$1/W_{ch}$ compared to FET noise ~$1/A$. This also explains the stronger increase in SNR for the FETs with lower FET height. For FET widths below 50 nm, the molecule signal typically starts to saturate and eventually reduces. This results in an optimal FET width of about 10 nm for the different FET architectures with a 5 nm FET height. Also, notice that for large FET widths, the tri-gate and embedded FET are approximately equivalent which results in a similar SNR. On the other hand, for small FET widths, the tri-gate FET and suspended FET become equivalent and result in a similar SNR. On the other hand, for small FET widths, the tri-gate FET and suspended FET become equivalent and

result in a similar SNR. Downscaling the FET width from 400 nm to the optimal width results in a ~5× boost of the SNR. For height downscaling, an increase in the SNR is observed down to very small FET heights (Figure 13a).

Gate length scaling also strongly boosts the single-molecule SNR as depicted in Figure 13b. The impact of gate length scaling is shown for the 30×30 nm^2 tri-gate FET used for calibration of the noise model and the three different FET architectures with optimal cross-sections. When downscaling the gate length, the SNR boost initially exceeds the $1/\sqrt{\mathrm{L_{gate}}}$ trend due to a stronger increase in molecule signal compared to the FET noise. For FET gate lengths below 50 nm, the molecule signal starts to saturate. The SNR reaches an optimum for a gate length of ~35 nm. Downscaling the gate length from 1000 nm to the optimal gate length results in a close to 10-fold increase in the single-molecule SNR. Our study shows that due to its highest optimal SNR, the suspended FET with 10×5 nm^2 cross-section and 35 nm gate length is the best architecture to perform single-molecule detection.

The reported experimental single-molecule signals in refs. [34,35] exceed our simulated single-molecule signals with multiple orders of magnitude and hence cannot be explained by an electrostatic modulation of mobile charge carriers in the FET channel since this is captured by our model (see SI section 13 for more details). Therefore, in contrast to the reported single-molecule sensing experiments using micron-long silicon FETs, our model predicts single-molecule sensing not to be possible with micron-long FETs. Further work could resolve the encountered contradictions between the reported single-molecule sensing experiments and our model. Further theoretical modeling could elucidate the impact of quantum mechanical effects, random discrete dopants and ballistic transport on the bioFET signal. Also, there is very limited experimental work on single-molecule detection with bioFETs. Further experimental work could validate the currently claimed single-molecule signals.

The benefit of nano-scaling silicon FETs to boost the single-molecule detection limit could raise the question of whether the variability is a concern for such nano-scaled devices.[64] However, variability did not turn out to be a showstopper for silicon FET technology in computer chips. Currently, the state-of-the-art commercial 5 nm technology node features transistor densities exceeding 100 million per mm^2. For single-molecule bioFET sensors, the variability induced by the functionalization of single molecules including SAMs and linkers must be considered as well.[65] Research reports and even commercial technologies have shown the successful functionalization of single molecules on different kinds of devices.[17,22,35,66–68] Moreover, with a silicon FET sensor platform one targets a high-throughput and parallelized detection of whether a specific type or building block of a molecule is present or not. This type of digital sensing is less sensitive to variability issues compared to traditional analog sensing. Furthermore, the larger the signal-to-noise ratio the more robust the sensing platform becomes towards for example variability issues.

One such parameter that could further boost the single-molecule SNR, is the electrolyte ionic strength. We calculate the singularly charged molecule SNR for the optimized suspended FET architecture for three different electrolyte ionic strengths. When the electrolyte ionic strength is reduced from 15 to 1.5 mM (Figure 13c), the SNR increases approximately 3-fold. This is the result of the 3-fold increase in molecule signal as expected from the reduction of electrolyte screening ($\lambda_\mathrm{d} \sim \sqrt{1/\mathrm{I}}$). The noise amplitude only minorly increases (SI section 8). Decreasing the ionic strength further to 0.1 mM increases the SNR, but is less than expected from the decrease in electrolyte screening strength. This can be explained by the large modulation region for which the molecule signal is significantly affected by the FET corners, but also by a small increase in FET noise. An SNR of nearly 2 is obtained for a singularly charged molecule at an optimal gate length of about 35 nm. For even smaller gate lengths, the SNR decreases due to the negative impact of the source-drain junctions (Figure 4).

Also, we show the impact of a reduced noise power spectral density from 500 μV^{2}μm^2/Hz (our bioFETs) to 50 μV^{2}μm^2/Hz which is the value for a commercial solid-state technology (Technology Roadmap for Semiconductors ITRS). The expected 3-fold increase in SNR is obtained across the entire gate length range due to a 3× reduction of the noise amplitude ($V_{noise} \sim \sqrt{S_{V_g}}$). The reduced noise power density results in an SNR of about 6 for a singularly charged molecule in a 0.1 mM electrolyte solution. This indicates that the FET sensor platform has the potential to enable on-chip protonation/deprotonation proteomics.[19]

Finally, we show the SNR for a 15 base-pair DNA oligo (15bp DNA) with 30 negative charges in a 1.5 mM electrolyte solution (see triangles Figure 13d). Notice that the DNA SNR does not increase with a factor of 30 compared to the singularly charged molecule. This is caused by nonlinear screening which leads to the saturation of the molecule signal as a function of increasing molecule charge.[43] But also due to the more elongated shape of the DNA molecule such that the charges are on average further away from the surface. The DNA SNR for an optimal FET device with high and low noise power density is ~11 and ~36 respectively. Also, we calculate the SNR for the practically more relevant hybridization of a 15-base DNA oligo to a PNA probe in a 1.5 mM electrolyte solution (see also SI section 11). Here, we obtain an SNR of ~28 assuming the low noise power density of 50 μV^{2}μm^2/Hz. These are promising results that indicate that the realization of single-molecule sensing with nano-scaled silicon FETs is feasible.

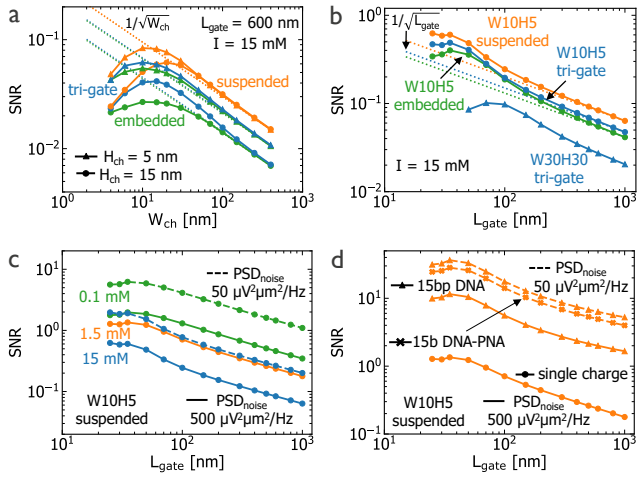

Figure 13: a) The impact of width and height scaling on the singularly charged molecule SNR for different FET architectures indicated by different colors. b) The impact of gate length scaling on the singularly charged molecule SNR for a FET with a larger cross-section (30×30 nm) and the different FET architectures with optimal cross-sections. The different FET architectures are indicated by different colors. c) The SNR for a singularly charged molecule as a function of gate length for the optimal suspended FET architecture (W_{ch}×H_{ch} = 10×5 nm²). Blue lines, orange lines and green lines represent simulated SNR values for 15 mM, 1.5 mM and 0.1 mM electrolyte ionic strengths respectively. Solid lines and dashed lines represent simulated SNR values for a device with low and high noise power spectral densities. d) The SNR for a 15 base-pair DNA molecule with 30 negative charges (triangle) and a 15-base DNA-PNA hybridization with 15 negative charges (cross) compared to the singularly charged molecule (circle). The electrolyte ionic strength is 1.5 mM. The noise bandwidth is 1-1000 Hz.

Conclusions

We quantified the impact of nano-scaling silicon FET sensors on the single-molecule SNR and also identified the physical mechanisms underlying these geometrical scaling trends. A channel series resistance effect explains the approximately inversely proportional trend between the gate length and the single-molecule signal for gate lengths larger than 50 nm. For smaller gate lengths, the boost in signal is limited due to an increased modulation of the source-drain junctions where the channel resistance is smaller and electrolyte screening is enhanced. Consequently, for effective channel lengths approaching the size of the channel region modulated by the single molecule, further downscaling does not significantly improve the single-molecule signal. The lower the electrolyte ionic strength, the larger the size of the modulated region and the stronger the saturation of the molecule signal becomes. Our results show that gate length downscaling from the micron range to 10s of nanometers can boost the single-molecule signal by almost two orders of magnitude.

Also, for large FET widths, the FET signal of a singularly charged molecule situated at the top surface is approximately inversely proportional to the width. However, for small widths, the molecule signal saturates and eventually decreases. This is explained by the proximity of the molecule to the convex top FET corners where the local gate capacitance increases and less charge is mirrored in the FET channel. For a larger size of the modulated region, the impact of the FET corners increases, and a larger optimal FET width arises. For a 15 mM ionic strength, the maximum molecule signal for a tri-gate FET is achieved for a width of about 20 nm. The detrimental impact of the FET corners can be reduced by using an embedded FET architecture resulting in a 2× boost in maximum signal compared to the tri-gate and suspended FET architecture. A second positive effect of the embedded FET architecture and the reduced corner effect is the reduced variation of the signal across the FET surface.

In contrast to width downscaling, the FET signal due to a single molecule at the top surface keeps increasing across the entire FET height range. For height downscaling the impact of the FET corners only minorly changes. An increase in molecule signal close to two orders of magnitude is obtained when the FET height is reduced from 400 nm to 5 nm.

Finally, we show that the single-molecule SNR, which determines the feasibility of single-molecule sensing, is also boosted by downscaling the FET sensor. For large FET dimensions, length, height and width downscaling results in a boost of the SNR somewhat exceeding the $1/\sqrt{\text{dimension}}$ trend. The optimal SNR arises for small FET widths ~10-20 nm and small FET gate lengths ~35 nm. Like the signal, the optimal FET dimensions for a maximum SNR increase when the size of the modulated region increases. In contrast to the signal, the largest SNR is obtained for the suspended FET due to a lower noise level compared to the tri-gate and embedded FET architectures. Moreover, the identification of the critical mechanisms underlying the single-molecule SNR geometric scaling trends aids the future design of better-performing single-molecule FET sensors. For a noise bandwidth of 1-1000 Hz and assuming an ideal pH insensitive and uncharged surface, the optimally-sized suspended FET results in an SNR of about 0.6 for a singularly charged molecule in a 15 mM solution. When we assume a commercial solid-state FET noise level, and a 1.5 mM and 15 mM electrolyte solution, we respectively predict an SNR of about 28 and 10 for a 15-base DNA oligo hybridizing to a PNA probe. This shows the potential of silicon FET devices as future single-molecule sensors.

Experimental

FET sensor structure

3D simulations are performed using Global TCAD Solutions GTS Framework, a TCAD software package specialized for simulating semiconductor devices. The FET architectures: tri-gate, suspended and embedded employed for the geometrical study are depicted in Figure 1. For all architectures, the source and drain regions are passivated with a 40 nm thick SiO_2 layer. The electrolyte sits in between the source and drain passivation and is contacted by the gate. The physical gate length is defined as the distance between the source and drain passivation or equivalently the length of the electrolyte segment. The source-drain junctions i.e. where the source-drain doping falls off are aligned with the passivation edges. The channel length is therefore equal to the physical gate length. The channel doping is 1×10^{17} cm^{-3}. The source and drain regions are highly doped (2×10^{20} cm^{-3}). The diffusion of the source-drain dopants into the

channel is characterized by a Gaussian function with σ = 3 nm (SI section 1). The gate oxide is a 1 nm thick SiO_2 layer. A 20 nm thick SiO_2 bottom oxide (Box) is present at the bottom of the FET architectures. Simulated physical gate or channel lengths range from 25 to 1000 nm. The FET channel widths and heights range from 2 nm to 400 nm.

FET sensor model

In the silicon FET channel and source-drain regions, the drift-diffusion equation is solved. The standard GTS MINIMOS 6 mobility model is selected for the silicon carriers.[69,70] Further, we apply the density-gradient model which includes a quantum correction on the charge carrier density at the silicon-oxide interfaces.[70,71] The electrolyte is modeled as an intrinsic semiconductor with the density of states set to 1×10^{-30} cm^{-3}. The low value for the density of states makes sure that the intrinsic semiconductor carriers do not play a role in charge screening. The ion distribution in the electrolyte segment is described by the Bikerman model which includes steric effects.[42] The Bikerman model for a binary symmetric electrolyte is implemented in the GTS framework with the bulk ion concentration c_0 and the ion diameter a as inputs:[42]

$$c_i = \frac{c_0 e^{-z_i q\psi/kT}}{1 + 4a^3 c_0 \sinh^2\left(\frac{z_i q\psi}{2kT}\right)}$$

The valency z_i is positive for the cation and negative for the anion species. We assume an ion valency of 1. ψ is the local potential referred to as the bulk electrolyte potential at the electrolyte gate contact. q is the elementary charge. k and T are the Boltzmann constant and the temperature. a is the solvated ion diameter which is set to 0.6 nm.[42] The hydronium ion concentration linked to the electrolyte's pH is not included in the ion distribution. Notice that a pH of 7 relates to a bulk hydronium ion concentration of only 10^{-7} M which is 3 orders of magnitude lower than the lowest considered background ion concentration of 0.1 mM. However, the hydronium ion concentration serves as an input for the pH-dependent oxide surface charging model. This model is only included in the supplementary information section 11, which elaborates on the impact of the oxide surface's pH sensitivity on the signal.

In the electrolyte segment, only electrostatics is considered. Feeding the Bikerman ion distribution model to the TCAD's Poisson equation then results in the following expression that is solved in the electrolyte segment:

$$\nabla^2\psi = -\frac{z_i q}{\varepsilon_r \varepsilon_0} \frac{c_0 e^{-z_i q\psi/kT}}{1 + 4a^3 c_0 \sinh^2\left(\frac{z_i q\psi}{2kT}\right)}$$

Here, ε_0 is the vacuum permittivity. The dielectric constant ε_r of the electrolyte segment is set to 78 which equals the dielectric constant of water at 25°C.[72] Ohmic contacts are assumed at which the potential is fixed. At all other simulation boundaries, the Neumann boundary condition guarantees that the simulation domain is self-contained.

The singularly charged molecule is modeled as a 1 nm^3 dielectric cube containing one positive elementary charge. The cube has a dielectric constant of 5. The 15 base-pair DNA molecule and 15-base DNA-PNA hybridization complex are modeled as dielectric boxes with a width and depth equal to 2 nm and a height of 5 nm. The dielectric boxes contain 30 and 15 negative elementary charges for the 15 base-pair DNA molecule and 15-base DNA-PNA complex respectively. The dielectric constant of the boxes is 5. For both the singularly charged and DNA molecule the distance of the bottom of the molecule to the gate oxide surface is set to 1 nm.

FET charge signal calculation

To determine the FET charge signal, we simulate I_{DS}-V_{GS} curves by sweeping the gate-source voltage V_{GS} in steps of 20 mV. The drain voltage V_D = 0 V and the source voltage V_S = 50 mV. The FET signal is expressed as a gate voltage shift using a constant current method. The simulated I_{DS}-V_{GS} curve without biomolecule charge is fitted with the EKV MOSFET model to calculate the threshold voltage V_T.[73] From this initial I_{DS}-V_{GS} curve, we determine the drain current I_{DS} at 100 mV underdrive V_{GS}-V_T = 100 mV (subthreshold). At this specific drain current, we extract the gate voltage for the I_{DS}-V_{GS} curve with the charged molecule added. This allows for determining the gate voltage shift or FET signal. The methodology makes sure that FETs with various geometries are compared in the same operating regime.

Signal-to-noise ratio calculation

The singularly charged molecule and 15 base-pair long DNA SNR are defined as the ratio of the molecule signal and the FET noise. The single-molecule signal is extracted using the constant current method as described in the previous paragraph. The gate voltage referred noise amplitude is calculated as follows:

$$V_{noise} = \sqrt{\frac{kT}{{g_m}^2} s \int {\Delta I_{tr}}^2 N_{tr,norm}(E_F)\, dA \times [\ln 1000 - \ln 1]}$$

k and T are the Boltzmann constant and temperature respectively. We assume a frequency range from 1 to 1000 Hz. The integral is done across the entire area of the silicon-oxide interface A = 2×(W_{ch}+H_{ch})×L_{gate}. The drain current modulations ΔI_{tr} induced by single traps across the entire device area are simulated using the linear perturbation method in GTS TCAD software.[45] The perturbation method is standardly applied at a 100 mV underdrive (V_{GS}-V_T = 100 mV). The transconductance g_m is extracted from the simulated I_{DS}-V_{GS} curve to translate the noise amplitude to a gate voltage-referred value.

We account for the typically observed energy dependence of the trap density for p-type FETs with a SiO_2 gate oxide by assuming the normalized factor $N_{tr,norm}(E)$ (Figure 14).[55] This factor is equal to 1 for a trap energy level E equal to the silicon valence band edge energy, $N_{tr,norm}(E_V)$ = 1. Traps near the Fermi level are approximately half occupied and therefore dominantly contribute to FET noise.[55] The variation of the Fermi level energy across the silicon-oxide interface is extracted from our TCAD simulations and serves as the input to calculate $N_{tr,norm}(E_F)$. We calibrate the scaling factor s using the mean measured noise power spectral density at 1 Hz $kTs/{g_m}^2 \int {\Delta I_{tr}}^2 N_{tr,norm}(E_F)\, dA$ for our 30 by 30 nm^2 tri-gate

liquid-gated FET with 1 µm gate length. The measured area normalized power spectral density (PSD) of our liquid-gated FETs equals 500 $\mu V^2\mu m^2/Hz$ which results in a scaling factor of 6.73×10^{10}. For commercial solid-state FETs with noise PSD of 50 $\mu V^2\mu m^2/Hz$, a scaling factor of 6.73×10^{9} is obtained. These scaling factors thus indicate trap densities at the silicon valence band edge of approx. 6.73×10^{10} and 6.73×10^{9} $cm^{-2}eV^{-1}$ for our liquid-gated and commercial solid-state FETs respectively. The trap densities are in line with the values reported in the literature.[55]

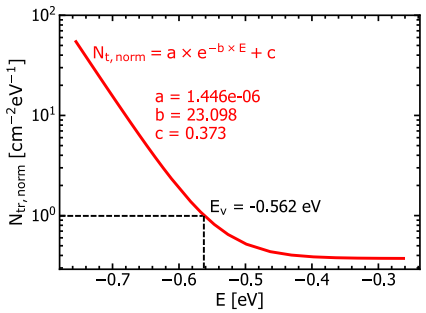

Figure 14: The normalized trap density as a function of the trap energy level relative to silicon midgap. Normalization of the density is done towards a trap with an energy level equal to the valence band energy.[55]

Measurement of FET sensor noise

The FETs to calibrate the noise characteristics are p-channel silicon-on-insulator (SOI) devices fabricated using CMOS technology as described in reference.[63] The Silicon channel thickness is approx. 30 nm. Channel widths range from 60 nm to 1000 nm and lengths from 450 nm to 3000 nm. The channel doping is set to 1×10^{17} cm^{-3} (n-type) and the source/drain regions received high dose p-type doping ($>1 \times 10^{20}$ cm^{-3}). This was followed by the growth of a 4 nm thick thermal SiO_2 layer as a gate oxide. Afterward, SiO_2 is used to passivate the FET source-drain contacts from the electrolyte environment. 15 mM phosphate-buffered saline (PBS) is used as the electrolyte solution for the noise measurement. We use a flow-through true Ag/AgCl electrode (by Microelectrodes Inc.) to ensure a stable and reliable gate voltage. The FET source and drain are contacted using a probe station. Source and substrate are grounded. Noise spectra are measured using a Keysight E4727A low-frequency noise analyzer. The measurements are performed in a dark and electrically shielded environment with drain-source bias (V_{DS}) of -50 mV.[40,63]

Author Contributions

S. Santermans performed research. S. Santermans, G. Hellings, W. Van Roy and K. Martens discussed and developed models and interpreted results. M. Heyns guided research. S. Santermans wrote the manuscript. G. Hellings, W. Van Roy and K. Martens reviewed and edited the paper.

Acknowledgments

The authors would like to thank Z. Stanojevic, M. Karner, and C. Kernstock (Global TCAD Solutions GmbH, Austria) for supporting the GTS TCAD simulations.

References

1 P. Bergveld, *IEEE Transactions on Biomedical Engineering*, 1970, **BME-17**, 70–71.
2 Y. Cui, Q. Wei, H. Park and C. M. Lieber, *Science (1979)*, 2001, **293**, 1289–1292.
3 F. Patolsky, G. Zheng and C. M. Lieber, *Nanomedicine*, 2006, **1**, 51–65.
4 J. I. Hahm and C. M. Lieber, *Nano Letters*, 2004, **4**, 51–54.
5 F. Patolsky, G. Zheng, O. Hayden, M. Lakadamyali, X. Zhuang and C. M. Lieber, *Proc Natl Acad Sci U S A*, 2004, **101**, 14017–14022.
6 F. Patolsky, G. Zheng and C. M. Lieber, *Nature Protocols*, 2006, **1**, 1711–1724.
7 E. Stern, J. F. Klemic, D. A. Routenberg, P. N. Wyrembak, D. B. Turner-Evans, A. D. Hamilton, D. A. LaVan, T. M. Fahmy and M. A. Reed, *Nature*, 2007, **445**, 519–522.
8 J. Franco, B. Kaczer, M. Toledano-Luque, P. J. Roussel, J. Mitard, L. Å. Ragnarsson, L. Witters, T. Chiarella, M. Togo, N. Horiguchi, G. Groeseneken, M. F. Bukhori, T. Grasser and A. Asenov, in *IEEE International Reliability Physics Symposium Proceedings*, 2012.
9 B. Kaczer, J. Franco, P. Weckx, P. J. Roussel, M. Simicic, V. Putcha, E. Bury, M. Cho, R. Degraeve, D. Linten, G. Groeseneken, P. Debacker, B. Parvais, P. Raghavan, F. Catthoor, G. Rzepa, M. Waltl, W. Goes and T. Grasser, *Solid-State Electronics*, 2016, **125**, 52–62.
10 M. von Haartman and M. Östling, in *Low-Frequency Noise In Advanced Mos Devices*, Springer Netherlands, 2007, pp. 1–26.
11 N. K. Rajan, D. A. Routenberg, J. Chen and M. A. Reed, *IEEE Electron Device Letters*, 2010, **31**, 615–617.
12 N. Clément, K. Nishiguchi, J. F. Dufreche, D. Guerin, A. Fujiwara and D. Vuillaume, *Applied Physics Letters*, 2011, **98**, 014104.
13 J. Männik, I. Heller, A. M. Janssens, S. G. Lemay and C. Dekker, *Nano Letters*, 2008, **8**, 685–688.
14 E. Stern, R. Wagner, F. J. Sigworth, R. Breaker, T. M. Fahmy and M. A. Reed, *Nano Letters*, 2007, **7**, 3405–3409.
15 J. K. Rosenstein, S. G. Lemay and K. L. Shepard, *Wiley Interdisciplinary Reviews: Nanomedicine and Nanobiotechnology*, 2015, **7**, 475–493.
16 J. J. Gooding and K. Gaus, *Angewandte Chemie International Edition*, , DOI:10.1002/anie.201600495.
17 S. Ardui, A. Ameur, J. R. Vermeesch and M. S. Hestand, *Nucleic Acids Research*, , DOI:10.1093/nar/gky066.
18 S. G. Rodriques, A. H. Marblestone and E. S. Boyden, *PLOS ONE*, 2019, **14**.
19 G. Miclotte, K. Martens and J. Fostier, *PLOS ONE*, 2020, **15**, e0238625.
20 V. Zhirnov, R. M. Zadegan, G. S. Sandhu, G. M. Church and W. L. Hughes, *Nature Materials*, , DOI:10.1038/nmat4594.
21 K. Shoorideh and C. O. Chui, *Proc Natl Acad Sci U S A*, 2014, **111**, 5111–5116.
22 Y. Choi, I. S. Moody, P. C. Sims, S. R. Hunt, B. L. Corso, I. Perez, G. A. Weiss and P. G. Collins, *Science (1979)*, 2012, **335**, 319–324.
23 Y. Choi, I. S. Moody, P. C. Sims, S. R. Hunt, B. L. Corso, D. E. Seitz, L. C. Blaszcazk, P. G. Collins and G. A. Weiss, *J Am Chem Soc*, 2012, **134**, 2032–2035.

24 Y. Choi, G. A. Weiss and P. G. Collins, *Physical Chemistry Chemical Physics*, 2013, **15**, 14879–14895.

25 Y. Choi, T. J. Olsen, P. C. Sims, I. S. Moody, B. L. Corso, M. N. Dang, G. A. Weiss and P. G. Collins, *Nano Letters*, 2013, **13**, 625–631.

26 T. J. Olsen, Y. Choi, P. C. Sims, O. T. Gul, B. L. Corso, C. Dong, W. A. Brown, P. G. Collins and G. A. Weiss, *J Am Chem Soc*, 2013, **135**, 7855–7860.

27 P. C. Sims, I. S. Moody, Y. Choi, C. Dong, M. Iftikhar, B. L. Corso, O. T. Gul, P. G. Collins and G. A. Weiss, *J Am Chem Soc*, 2013, **135**, 7861–7868.

28 S. Sorgenfrei, C. Y. Chiu, M. Johnston, C. Nuckolls and K. L. Shepard, *Nano Letters*, 2011, **11**, 3739–3743.

29 S. Sorgenfrei, C. Y. Chiu, R. L. Gonzalez, Y. J. Yu, P. Kim, C. Nuckolls and K. L. Shepard, *Nature Nanotechnology*, 2011, **6**, 126–132.

30 A. Setiadi, H. Fujii, S. Kasai, K. I. Yamashita, T. Ogawa, T. Ikuta, Y. Kanai, K. Matsumoto, Y. Kuwahara and M. Akai-Kasaya, *Nanoscale*, 2017, **9**, 10674–10683.

31 F. Liu, M. Bao, H. Kim, K. L. Wang, C. Li, X. Liu and C. Zhou, *Applied Physics Letters*, 2005, **86**, 1–3.

32 N.-P. Wang, S. Heinze and J. Tersoff, *Nano Letters*, 2007, **7**, 910–913.

33 J. Wang, F. Shen, Z. Wang, G. He, J. Qin, N. Cheng, M. Yao, L. Li and X. Guo, *Angewandte Chemie International Edition*, 2014, **53**, 5038–5043.

34 G. He, J. Li, H. Ci, C. Qi and X. Guo, *Angewandte Chemie International Edition*, 2016, **55**, 9036–9040.

35 J. Li, G. He, H. Ueno, C. Jia, H. Noji, C. Qi and X. Guo, *Nanoscale*, 2016, **8**, 16172–16176.

36 J. Li, G. He, U. Hiroshi, W. Liu, H. Noji, C. Qi and X. Guo, *ACS Nano*, 2017, **11**, 12789–12795.

37 W. Liu, J. Li, Y. Xu, D. Yin, X. Zhu, H. Fu, X. Su and X. Guo, *Advanced Science*, 2021, **n/a**, 2101383.

38 I. Y. Chung, H. Jang, J. Lee, H. Moon, S. M. Seo and D. H. Kim, *Nanotechnology*, , DOI:10.1088/0957-4484/23/6/065202.

39 F. Pittino, P. Palestri, P. Scarbolo, D. Esseni and L. Selmi, *Solid-State Electronics*, , DOI:10.1016/j.sse.2014.04.011.

40 M. Gupta, S. Santermans, B. Du Bois, R. Vos, S. Severi, G. Hellings, L. Lagae, K. Martens and W. Van Roy, *IEEE Sensors Journal*, 2020, **20**, 8956–8964.

41 J. J. Bikerman, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, , DOI:10.1080/14786444208520813.

42 M. Z. Bazant, M. S. Kilic, B. D. Storey and A. Ajdari, *Advances in Colloid and Interface Science*, 2009, **152**, 48–88.

43 S. Santermans, F. Schanovsky, M. Gupta, G. Hellings, M. Heyns, W. Van Roy and K. Martens, *ACS Sensors*, , DOI:10.1021/acssensors.0c02285.

44 S. Santermans, F. Schanovsky, M. Gupta, G. Hellings, M. Heyns, W. Van Roy and K. Martens, *ACS Sensors*, , DOI:10.1021/acssensors.0c02285.

45 GTS – Advanced TCAD and Physical Device Simulation - Global TCAD Solutions, http://www.globaltcad.com/en/home.html, (accessed February 6, 2020).

46 A. Asenov, R. Balasubramaniam, A. R. Brown and J. H. Davies, *IEEE Transactions on Electron Devices*, 2003, **50**, 839–845.

47 M. F. Bukhori, A. R. Brown, S. Roy and A. Asenov, in *2009 IEEE International Integrated Reliability Workshop Final Report*, IEEE, 2009, pp. 82–85.

48 M. F. Bukhori, T. Grasser, B. Kaczer, Hans. Reisinger and A. Asenov, in *2010 IEEE International Integrated Reliability Workshop Final Report*, IEEE, 2010, pp. 76–79.

49 L. Gerrer, J. Ding, S. M. Amoroso, F. Adamu-Lema, R. Hussin, D. Reid, C. Millar and A. Asenov, *Microelectronics Reliability*, 2014, **54**, 682–697.

50 D. Landheer, G. Aers, W. R. McKinnon, M. J. Deen and J. C. Ranuarez, *Journal of Applied Physics*, , DOI:10.1063/1.2008354.

51 M. Gupta, S. Santermans, G. Hellings, L. Lagae, K. Martens and W. Van Roy, *IEEE Sensors Journal*, 2021, **21**, 4143–4151.

52 X. P. A. Gao, G. Zheng and C. M. Lieber, *Nano Letters*, 2010, **10**, 547–552.

53 I. Heller, J. Mãnnik, S. G. Lemay and C. Dekker, *Nano Letters*, 2009, **9**, 377–382.

54 J. N. Israelachvili, *Intermolecular and Surface Forces*, Elsevier, 2011.

55 M. Banaszeski Da Silva, H. P. Tuinhout, A. Zegers-Van Duijnhoven, G. I. Wirth and A. J. Scholten, *IEEE Transactions on Electron Devices*, 2016, **63**, 3683–3692.

56 G. Reimbold, *IEEE Transactions on Electron Devices*, 1984, **31**, 1190–1198.

57 S. Santermans, F. Schanovsky, M. Gupta, G. Hellings, M. Heyns, W. Van Roy and K. Martens, *ACS Sensors*, 2021, acssensors.0c02285.

58 M. Gupta, S. Santermans, G. Hellings, L. Lagae, K. Martens and W. Van Roy, *IEEE Sensors Journal*, 2020, 1–1.

59 C. G. Jakobson, M. Feinsod and Y. Nemirovsky, *Low frequency noise and drift in Ion Sensitive Field Effect Transistors*, 2000, vol. 68.

60 K. Bedner, V. A. Guzenko, A. Tarasov, M. Wipf, R. L. Stoop, S. Rigante, J. Brunner, W. Fu, C. David, M. Calame, J. Gobrecht and C. Schönenberger, *Sensors and Actuators B: Chemical*, , DOI:10.1016/j.snb.2013.09.112.

61 K. K. Hung, P. K. Ko, C. Hu and Y. C. Cheng, *IEEE Transactions on Electron Devices*, 1990, **37**, 654–665.

62 A. J. Scholten, L. F. Tiemeijer, R. van Langevelde, R. J. Havens, A. T. Zegers-Van Duijnhoven and V. C. Venezia, *IEEE Transactions on Electron Devices*, 2003, **50**, 618–632.

63 K. Martens, S. Santermans, M. Gupta, G. Hellings, R. Wuytens, B. Du Bois, E. Dupuy, E. Altamirano-Sanchez, K. Jans, R. Vos, T. Stakenborg, L. Lagae, M. Heyns, S. Severi and W. Van Roy, *IEEE International Electron Devices Meeting*.

64 M. Tintelott, V. Pachauri, S. Ingebrandt and X. T. Vu, *Sensors*, 2021, **21**, 5153.

65 B. Dorvel, B. Reddy and R. Bashir, *Analytical Chemistry*, 2013, **85**, 9493–9500.

66 J. Korlach, P. J. Marks, R. L. Cicero, J. J. Gray, D. L. Murphy, D. B. Roitman, T. T. Pham, G. A. Otto, M. Foquet and S. W. Turner, *Proceedings of the National Academy of Sciences*, 2008, **105**, 1176–1181.

67 E. Pibiri, P. Holzmeister, B. Lalkens, G. P. Acuna and P. Tinnefeld, *Nano Letters*, 2014, **14**, 3499–3503.

68 T. Plénat, S. Yoshizawa and D. Fourmy, *ACS Applied Materials & Interfaces*, 2017, **9**, 30561–30566.

69 S. Selberherr, W. Hänsch, M. Seavey and J. Slotboom, *Solid-State Electronics*, 1990, **33**, 1425–1436.

70 *MinimosNT - User Manual*, Global TCAD Solutions GmbH, Wien, release 2022.03.

71 A. Wettstein, *Quantum Effects in MOS Devices*, Zürich, 2000.

72 C. G. Malmberg and A. A. Maryott, *Dielectric Constant of Water from 0° to 100° C*, 1956, vol. 56.

73 C. C. Enz, F. Krummenacher and E. A. Vittoz, *Analog Integrated Circuits and Signal Processing*, 1995, **8**, 83–114.