Simulating Capacitances to Silicon Quantum Dots: Breakdown of the Parallel Plate Capacitor Model

Ted Thorbeck, Student Member, IEEE, Akira Fujiwara, and Neil M. Zimmerman

Abstract—Many electrical applications of quantum dots rely on capacitively coupled gates; therefore, to make reliable devices we need those gate capacitances to be predictable and reproducible. We demonstrate in silicon nanowire quantum dots that gate capacitances are reproducible to within 10\% for nominally identical devices. We demonstrate the experimentally that gate capacitances scale with device dimensions. We also demonstrate that a capacitance simulator can be used to predict measured gate capacitances to within 20\%. A simple parallel plate capacitor model can be used to predict how the capacitances change with device dimensions; however, the parallel plate capacitor model fails for the smallest devices because the capacitances are dominated by fringing fields. We show how the capacitances due to fringing fields can be quickly estimated.

Index Terms—Capacitance, quantum dots (QDS), semiconductor device modeling, single electron transistors.

Quantum dots (QDs) with capacitively coupled gates have enabled the manipulation of individual electrons, which leads to many attractive applications. For example, the single-electron transistor (SET) [1] consists of a QD tunnel coupled to a source and drain, with a capacitively coupled gate that controls the average charge on the QD. The SET, in which the current can change by orders of magnitude when the charge on a nearby gate changes by less than the charge of an electron, is the world’s most sensitive charge electrometer [2]. By pulsing the gates on a multiple QD device, electrons can be pumped through the QDs to define either the ampere or the farad [3], [4]. SET-based single electron logic and memory have been proposed as the ultimate in classical electronic scaling [5], [6]. Going beyond classical electronics, both the charge and the spin of a single electron in a QD have been proposed as the basis for a quantum computer [7], [8]. The potential for single-electron QDs is enormous, but limitations in the reproducibility and predictability of QDs have limited these devices [6]. In particular, some of these applications are impossible without reproducible gate capacitances. For example, a useful QD quantum computer or multiple-QD pump would need many devices operating in parallel with the same gate voltages pulses, which would require reproducible and predictable capacitances.

In this paper, we show that the gate capacitances in our devices, QDs in a silicon nanowire, are reproducible for devices with the same lithographic parameters, and are predictable over a wide range of capacitances for devices with different dimensions. For nominally identical devices, the measured gate capacitances are reproducible to about within 10\%. In addition, using a capacitance simulator, we can predict the gate capacitances to within 20\%. In the capacitance simulation, we used only the parameters used during device fabrication, and no adjustable parameters, to determine the device dimensions.

For a single set of lithographically identical devices, we have previously shown that gate capacitances are reproducible [9]. Additionally, it has been shown that gate capacitances can scale with lithographic parameters [10]–[12]. In this paper, we will numerically describe the reproducibility of gate capacitances. Other groups have demonstrated the ability to simulate gate capacitance [13], [14]; however, most have done so for only one device, and some rely on several fitting parameters to match the experimental capacitances. We demonstrate capacitance simulations for a wide variety of device dimensions with no fitting parameters. Some previous work [11], [12] has relied upon a parallel plate capacitance model to predict capacitances. In this approximation, only field lines from the gate straight to the QD are included. In contrast, for our smallest devices the capacitances are instead dominated by fringing field lines from the gate to the QD.

Our devices [9] [see Fig. 1(a)] consist of a lightly p-doped or undoped nanowire surrounded by two gate layers. The nanowire is mesa-etched into a silicon on insulator wafer. Devices with a wide range of dimensions are presented in this paper. The minimum (maximum) dimensions of the nanowire: 17 (21) nm thick ($t_{ox,1}$), 10 (30) nm wide ($w_{si}$), 300 (510) nm long. Thermally grown SiO$_2$ isolates the nanowire from the gates ($t_{ox,1}$, $t_{ox,2}$ = 20 nm). The lower gate layer consists of three heavily doped poly-silicon gates that wrap around three sides of the nanowire like a tri-gate-FET (field effect transistor). The lower gates are 10 (40) nm long ($L_{UG}$) and spaced by 70 (130) nm ($S$). More SiO$_2$ separates the lower gate layer from the global upper gate ($t_{ox,2}$ = 30 nm). The length of the upper gate ($L_{UG} = S - 2t_{ox,2}$) ranges between 40 nm and 100 nm, and corresponds to the length the lower gates fills in between the two lower gates and is closest to the nanowire.

A positive voltage on the upper gate inverts the nanowire. Negative voltages on the lower gates deplete the nanowire,
geometry that we simulated is shown in Fig. 1(b) and (c). All structures in the simulator were terminated 50 nm away from the QDs because conductors and dielectrics further away had little impact on the simulated capacitances but increased the runtime of the simulation.

Table I contains the full set of measured gate capacitances (UG, LGS, LGC, LGD) to each of the three QDs (full, source half, and drain half) for two lithographically identical devices as well as the simulated capacitances. There are several qualitative dependences that make intuitive sense. 1) For the full QD, the barrier gates (LGS and LGD) have similar capacitances to the QD and LGC has a larger capacitance to the QD, as expected. 2) For the source half-QD, LGS, and LGC have similar capacitances to the QD whereas LGD has a much smaller capacitance because it is far away and screened by the other gates. 3) The capacitance from UG to the full QD is roughly the sum of the capacitances from UG to the source and drain half-QDs. To quantitatively describe the reproducibility in Table I, we define deviation as the difference between an individual data point and the average divided by the average. All the deviations in Table I are less than 15% and the average deviation is 6%. This demonstrates the ability to reproduce capacitances to lithographically identical devices.

Fig. 2 shows the measured and simulated values for the capacitance from LGC to the full QD. To compare devices of different lithographic parameters, we combine the thickness of the nanowire $t_{si}$, width of the nanowire $w_{si}$, and length of the lower gate $L_{LG}$ into a single parameter that describes the surface area of the nanowire directly below the gate $(2t_{si} + 2w_{si})L_{LG}$. Three points stand out when comparing the measured and simulated capacitances in Fig. 2. 1) For devices with the same surface area parameter, the maximum deviation is 9% and the average deviation is 7%. This demonstrates the reproducibility of the gate capacitance for nominally identical devices over a wide range of device dimensions. 2) The spread of capacitances for a single surface area parameter, the maximum deviation is 9% and the average deviation is 7%. This demonstrates the reproducibility of the gate capacitance for nominally identical devices over a wide range of device dimensions. 3) The simulated capacitances accurately predict the measured capacitances. The average deviation of the simulation (absolute value of the difference between the average measured and simulated capacitance divided by the simulated capacitance) is 17%. We note that, for larger devices, the capacitance simulator does tend to predict a larger $C_{LGC}$ than measured. One possible explanation is that the thermal oxidation rate depends on the stress in the silicon; in a larger device less stress might build up than expected, allowing for more silicon to be consumed by the oxide therefore making the capacitance smaller than we calculated (resulting in a larger $t_{ox},2$) and smaller $L_{LG}$) [10].

Fig. 2 also shows the breakdown of the parallel plate capacitance model. Although the parallel plate capacitance model captures the slope of the experimentally measured $C_{LGC}$, it predicts an intercept of zero whereas the measured $C_{LGC}$ has an intercept of 4.7 aF. For the smallest devices studied, the average measured $C_{LGC}$ is 5.1 aF; the capacitance simulator predicts 5.6 aF, a 10% error, whereas the parallel plate model predicts a capacitance of 1.0 aF. For smaller devices this

---

1Software is named for informational purposes only; it does not imply an endorsement or a recommendation by NIST.
discrepancy will increase. This breakdown is not surprising; we only expect the parallel plate capacitance model to work if the plates of the capacitor are much larger than their separation. In contrast, for our smallest devices, fringing electric field lines dominate the capacitance between gate and dot [see Fig. 1(d)]. A simple way to estimate the fringing fields is to note that the fringing fields should fall off on a length scale about equal to the separation of the gate and dot, in our case $t_{1,ox}/2$. Therefore, a corrected parallel plate capacitance estimate would look like: $C_{LGC} = \varepsilon(2t_{si} + w_{si})(L_{LG} + t_{1,ox})/t_{1,ox}$. For our smallest devices, this corrected parallel plate capacitor predicts $C_{LGC} = 3.9 \text{ aF}$, which favorable compares to our average measured $C_{LGC} = 5.1 \text{ aF}$.

Fig. 3 shows both the simulated and measured capacitance from UG to either the source or drain half-QD (which by symmetry should have identical capacitances to UG). Here, we parameterize the area as $(2t_{si} + 2w_{si})L_{UG}$, where $L_{UG}$ is the length over which the upper gate is closest to the QDs [see Fig. 1(b)]. This plot makes the same three points the previous plot did. 1) The spread of capacitances for a single surface area parameter is small. The average deviation is 5% and the maximum deviation is 6%. 2) The lithographic parameters control the gate capacitances. 3) The simulation predicts the measured gate capacitances; the average simulated deviation is 14%. Fig. 3 also show how for the smallest devices the capacitance simulator does a much better job of predicting gate capacitances than the uncorrected parallel plate capacitance model. The parallel plate capacitance model again underestimates the capacitance. For the smallest measured devices shown in Fig. 3, the average measured capacitance is $C_{UG} = 3.6 \text{ aF}$. The uncorrected parallel plate capacitor estimates $C_{UG} = 1.6 \text{ aF}$. Once we add the fringing fields to the corrected parallel plate capacitor, we estimate a capacitance of 4.0 aF. This again demonstrates that for our smallest devices the capacitance is dominated by fringing electric fields. To explain the difference between the simulated and measured $C_{UG}$ for the largest devices in Fig. 3, we suspect that differences in wire length affects thickness of the oxide $t_{ox,1}$.

For both Figs. 2 and 3 the parallel plate capacitance model did a better job of predicting measured capacitances when the full four-sided perimeter of the wire $(2t_{si} + 2w_{si})$ was used in calculating the surface area than when only the three-sides adjacent to the gates $(2t_{si} + w_{si})$ was used. This suggests that although we have a trigate-FET, conduction is occurring along all four surfaces of the wire.

In contrast to our success in simulating gate capacitances, the simulation was unable to match the measured barrier capacitances (the capacitance from the source or drain to the QD). We have previously shown that the barrier capacitance is a function of gate voltage [16], with a minimum of 15 aF and a maximum of 50 aF (for a device $w_{si} = 30 \text{ nm}, t_{si} = 10$), whereas our simulation predicted 4 aF. To explain the gate voltage dependence of the barrier capacitances, either the length of the barrier or the dielectric constant $\varepsilon$ must be a function of gate voltage. If the length of the barrier were a function of the gate voltage, then the length of the QD and hence the gate capacitances to the QD would be a function of the gate voltage, which we do not observe. A more likely explanation is related to $\varepsilon$ as $V_{LG}$ becomes

<table>
<thead>
<tr>
<th>Cap (aF)</th>
<th>UG</th>
<th>LG</th>
<th>LG</th>
<th>LG</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Dev. 1</td>
<td>Dev. 2</td>
<td>Sim.</td>
<td>Dev. 1</td>
</tr>
<tr>
<td>Full Island</td>
<td>22</td>
<td>22</td>
<td>25</td>
<td>2.7</td>
</tr>
<tr>
<td>Source Half</td>
<td>11</td>
<td>10</td>
<td>12</td>
<td>2.3</td>
</tr>
<tr>
<td>Drain Half</td>
<td>11</td>
<td>11</td>
<td>12</td>
<td>0.1</td>
</tr>
<tr>
<td>LGC</td>
<td>LGC</td>
<td>LGC</td>
<td>LGC</td>
<td>LGC</td>
</tr>
<tr>
<td>Dev. 1</td>
<td>Dev. 2</td>
<td>Sim.</td>
<td>Dev. 1</td>
<td>Dev. 2</td>
</tr>
<tr>
<td>2.5</td>
<td>2.8</td>
<td>2.6</td>
<td>5.1</td>
<td>5.1</td>
</tr>
</tbody>
</table>

Measured gate capacitances for device 1 and device 2 (nominally identical) and simulated capacitances for each of the four gates to each of the three QDs. For comparison to Figs. 2 & 3 the area of the full QD under the lower gate is 940 nm², and the area of the half QD under the upper gate is 3760 nm².

Fig. 2. Measured capacitances from LGC to the full island (blue x’s), simulated capacitances (green circles) and capacitances calculated from the parallel capacitance model (black line), which show reproducibility of measured capacitances and that the simulation predicts measured capacitances much better than the parallel plate model. The horizontal axis represents the area of the nanowire directly below the lower gate $(2t_{si} + 2w_{si})L_{LGC}$, based on the parameters used in lithography. Arrows with a number correspond to multiple data points on top of each other. The uncertainties are not shown, but are typically smaller than the plotting symbols.

Fig. 3. Experimental and simulated capacitances from UG to either source or drain half QD parameterized for the area of the nanowire closest to the upper gate $(2t_{si} + 2w_{si})L_{UG}$. Symbols are the same as in Fig. 2.
less negative the nanowire carrier density $\rho$ increases, and the silicon changes from insulating to semiconducting. Also, as $\rho$ increases, we expect that $\varepsilon$ increases from its textbook value, which corresponds to low $\rho$. Such a capacitance enhancement has been observed in doped [17] and undoped [16] silicon tunnel barriers, but no simulation has succeeded in predicting measured capacitances. Significant additional work is needed to include the gate voltage dependence on the simulation. To generalize from our experience, capacitors with a chemically defined insulator (like our gate capacitances) can be successfully simulated, whereas capacitors with an electrically defined insulator (like our barrier capacitances) cannot be successfully simulated.

In summary, the combination of Table I and Figs. 2 and 3 demonstrate experimental control over gate capacitances using the fabrication parameters. We showed that gate capacitances are typically reproducible to within 10% for lithographically identical devices. We showed how the often used parallel plate capacitance model breaks down for small devices because the capacitance is dominated by fringing electric fields. We showed how the parallel plate model can be improved by estimating the fringing capacitance. We predicted the gate capacitances in a capacitance simulator to within 20% using only lithographic parameters and no fitting parameters. Having a well calibrated capacitance simulator for these devices allows us to apply it to other ends, such as determining the location of QDs not defined by the gates which frequently occur in these and other silicon devices [18].

ACKNOWLEDGMENT

The authors would like to acknowledge helpful conversations with M. Stewart, P. Koppinen, R. Lake, J. Pomeroy, and H. Stalford.

REFERENCES
