Benchmarking high fidelity single-shot readout of semiconductor qubits

D Keith, S K Gorman, L Kranz, Y He, J G Keizer, M A Broome and M Y Simmons
Centre of Excellence for Quantum Computation and Communication Technology, School of Physics, University of New South Wales, Sydney, New South Wales 2052, Australia
1 Current address: Department of Physics, University of Warwick, Coventry CV4 7AL, United Kingdom

Abstract
Determination of qubit initialisation and measurement fidelity is important for the overall performance of a quantum computer. However, the method by which it is calculated in semiconductor qubits varies between experiments. In this paper we present a full theoretical analysis of electronic single-shot readout and describe critical parameters to achieve high fidelity readout. In particular, we derive a model for energy selective state readout based on a charge detector response and examine how to optimise the fidelity by choosing correct experimental parameters. Although we focus on single electron spin readout, the theory presented can be applied to other electronic readout techniques in semiconductors that use a reservoir.

1. Introduction
Quantum computing relies on the preparation, control and measurement of quantum states [1]. In order to achieve scalable universal quantum computation the error rate of all these processes needs to be less than ~1%—known as the fault-tolerant threshold for two-dimensional surface codes [2–5]. Recently, an emphasis has been placed on the quality of single and two-qubit gates through randomised benchmarking [6–9]. The usefulness of randomised benchmarking comes from the removal of state preparation and measurement errors. It also scales polynomially with the number of qubits making it an efficient verification and validation method [10]. However, state preparation and measurement errors will also lower the overall fidelity of the quantum computer’s operation and always need to be considered for fault-tolerant quantum computation [11]. Recent large-scale proposals for quantum computers in semiconductors utilise single electron or nuclear spins as the qubits [4, 5, 12–15]. The measurement of the electron spin in these semiconductor architectures can be performed using a weakly coupled reservoir to the quantum dot/donor. Motivated by these proposals, we examine the fidelity of correctly identifying the spin of the electron using a nearby reservoir that is monitored by a charge sensor.

Over a decade ago, single-shot spin readout of an electron was first achieved in a semiconductor by Elzerman et al [16]. Elzerman et al tuned the reservoir Fermi level between the Zeeman split electron spin states when in a global magnetic field, as in figure 1(a). This is done in such a way that only the excited spin state |1⟩ possesses enough energy to tunnel to the reservoir, to then quickly be replaced by a ground state electron |0⟩ tunnelling back from the reservoir during the readout time. These tunnelling events are what cause the ‘blip’ in the response of a nearby charge detector (see figure 1(b)). During the qubit readout process the detector response, x is monitored and the maximum value is recorded. In general, we assume that the detector is monitored for a readout time, t, and sampled at a rate of 1/t. The detection event is represented by a ‘blip’ in the charge sensor response as a function of the number of samples n made over time, t. The detection event corresponds to one of the two possible initial states (|0⟩ or |1⟩), which should ideally only occur when the electron is initially |1⟩ to perform accurate readout. The quality of the readout can be affected by such factors as thermal excitations, qubit state relaxation, the signal-to-noise ratio (SNR), and the ratio between filter rates, data sampling rates and electron tunnel rates. Subsequently, single-shot electron spin readout has been demonstrated using single
Figure 1. Single-shot readout using a reservoir. (a) Schematic of the Zeeman split energy levels ($0\rangle$ and $1\rangle$) separated by the Zeeman energy $E_Z$ in the system showing a charge detector that is used to determine if an electron has tunneled between the qubit and reservoir. The electron tunnel rates are dictated by the Fermi broadening of the reservoir about the Fermi level $E_F$, according to the reservoir temperature. (b) Illustrative response from the detector for a low level, $0\rangle$ and high level, $1\rangle$. The 'blip' in the detector response, $x$ indicates the measurement of $|1\rangle$. Right: corresponding detector histograms showing the mean, $\mu_i$, and noise (variance), $\sigma_i^2$, for both $i = 0, 1$. (c) Assignment of the qubit states based on the charge detector response. The qubit can be prepared in either the $|0\rangle$ or $|1\rangle$ spin state. Whether or not these states are preserved throughout spin-to-charge conversion depends on the conditional probability that the electron tunneled ($F_{BC}$), or not ($F_{BC}^\perp$). Similarly, the qubit states are preserved throughout electrical readout depending on the conditional probability that a 'blip' was successfully measured ($F_1$), or not ($F_2$). By following the arrows, the probability of correctly identifying the qubit state can be obtained.

donors [17] and donor quantum dots in semiconductors [18], Si/SiGe quantum dots [19], Si-MOS quantum dots [20] and nitrogen vacancy centres [21]. The ability to perform high fidelity single-shot readout has improved over the years reaching the point where single-shot spin readout can be performed above the fault-tolerant threshold [22]. However, the method used to determine the fidelity of single-shot readout has not been consistent between experiments making it difficult to directly compare one system with another.

Independent of the semiconductor system that is studied, the ability to distinguish between quantum states relies upon separating measurement distributions for each state using a particular threshold with respect to a signal from a detector. The measurement distributions are only dependent on a few experimentally accessible parameters including the noise spectrum, measurement bandwidth, temperature, magnetic field strength, tunnel rate to the reservoir, qubit energy separation, and the timing of the state conversion process. These factors have not always been consistently accounted for, and the fidelity analysis using energy selective spin readout has evolved since its first demonstration by Elzerman et al [16]. Elzerman et al [16] characterised the measurement fidelity of a single electron spin qubit by measuring the impact on detection errors from unwanted spin-flips, due to temperature or relaxation [23], as well as charge sensor dark counts during readout. Since then the semiconductor electronic qubit field has mostly performed measurements and fidelity calculations utilising a peak filter to distinguish spin states, where the measured detector signal must cross a particular threshold value within a given readout time [17, 20, 23–26].

Reference [17] employed the now commonly used Monte-Carlo method, where repeated random samples are produced to numerically fit simulated signal histograms to experimental histograms to optimise the readout threshold and take post-processing errors into account. However, in this analysis the effect of the finite electron temperature and spin relaxation on the spin state during readout was not included. In the papers to follow, the effect of the finite electron temperature was determined using a variety of methods, such as electron spin resonance measurements and the increase of thermal excitations with longer microwave bursts [20, 24], direct temperature measurements of the reservoir Fermi distribution [18], and tunnel rate measurements which change with quantum dot detuning based on temperature [25, 26]. The previously used Monte-Carlo method involves simulating the readout traces (generally with white Gaussian noise) based on the experimental parameters calculated from the experiment and comparing the resulting histograms to the experimental histograms. Other post-processing techniques have also been proposed such as wavelet-edge detection to filter out low frequency signal drift [27], maximum-likelihood estimation to recursively calculate the probability an event occurred [28] and nonlinear filtering to weight measurements differently over time [29] to better detect the qubit states, yet these do not remove the need to fit to a histogram of numerically simulated readout traces.
It is important that an agreed upon methodology is used to both benchmark results and help optimise readout for future experimental work. Therefore, we propose a comprehensive, analytical approach with the capability to determine optimal thresholds for performing single-shot readout with the highest possible fidelity without relying on arbitrary numerics. The model we present is extendable to multi-qubit systems with regards to sequential readout and can be generalised to any noise spectrum. Speciﬁcally, we demonstrate its performance in the case of white Gaussian noise as an example, which is commonly used to model detector noise [17, 18, 23, 25]. Using our model, we describe different limiting factors in readout fidelity, how to identify them and strategies to increase the fidelity once they have been identified. Finally, our method removes the need to rely on Monte-Carlo simulations [17] to calculate fidelities which we show in supplementary material I available online at stacks.iop.org/NJP/21/063011/mmedia, can introduce a large error on the readout fidelity by inappropriate bin and simulation numbers, which are often not quoted.

We outline our analytic approach in as general terms as possible, with sufﬁcient detail, and based in experimental parameters to encourage applicability, consistency, and practicality. For the proceeding sections, we make no assumptions on the energy levels of the qubit and use the terminology of high level (|1⟩), which is generally taken to be the excited state of the qubit and the low level (|0⟩), usually assigned to the ground state as shown in ﬁgure 1(a). We also assume that the four tunnel times of the individual qubit states to a reservoir (or similar structure), {t_{IN}^{0}, t_{OUT}^{0}, t_{IN}^{1}, t_{OUT}^{1}} are known.

To determine the readout fidelity we must identify the source of all relevant errors. The errors in single-shot readout can come from either the conversion of quantum states to a measurable voltage or current signal or from post-processing where the qubit state can be incorrectly assigned from the measurement. The individual state fidelities F_i are the conditional probabilities of correctly identifying the qubit states, |0⟩ and |1⟩, after the qubit is initially in the corresponding state. The state fidelity F_stc (F_E) can be broken down into fidelities related to different stages of the readout process as depicted inﬁgure 1(c): the state-to-charge conversion (STC), which is related to F_{STC}^0 (F_{STC}^1) and electrical detection, related to F_E^0 (F_E^1) and are given by

\[ F_0 = F_{STC}^0 F_E^0 \quad \text{and} \quad F_1 = F_{STC}^1 F_E^1, \]

\[ F_0 = 1 - F_{STC}^0 (1 - F_E^0), \]

\[ F_1 = 1 - F_{STC}^1 (1 - F_E^1), \]

such that \( F_{STC}^0 F_E^0 \) is the conditional probability that the |0⟩ qubit state did not tunnel to the reservoir and did not give a 'blip' in the detector response. Similarly, \( F_{STC}^1 F_E^1 \) is the conditional probability that the |1⟩ qubit state did tunnel to the reservoir and a 'blip' was detected by the sensor. The second terms in equations (1) and (2) result from consecutive errors that cancel each other out. The overall measurement fidelity, \( F_M \), which we define as the average detection fidelity of each qubit state is given by

\[ F_M = \frac{F_0 + F_1}{2}. \]

We now detail the calculations of the STC and electrical detection fidelities required to calculate the overall measurement fidelity.

2. State-to-charge conversion

The STC process is used to maximise the probability (and hence, fidelity/visibility) that the detected 'blip' is resultant from the |1⟩ state, and not the |0⟩ state, by optimising the readout time \( t_r \). The STC visibility, \( V_{STC} \) is dependent on \( t_{OUT}^0, t_{OUT}^1 \) and the |1⟩ relaxation time, \( T_1 \)

\[ V_{STC}(t_r; t_{OUT}^0, t_{OUT}^1, T_1) = F_{STC}^0 + F_{STC}^1 - 1, \]

where the |0⟩ level fidelity \( F_{STC}^0 \) is the probability that |0⟩ does not tunnel to the reservoir due to thermal ﬂuctuations and the |1⟩ level fidelity \( F_{STC}^1 \) is the probability that |1⟩ has tunneled to the reservoir in a time, \( t_r \). Assuming perfect electrical ﬁdelity we can categorise the individual fidelities based on STC as per ﬁgure 1(c) (left-hand side)

| Qubit | Measurement outcome |
|-------|---------------------|
|       | No Tunnel           | Tunnel              |
| |0⟩   | \( F_{STC}^0 \)     | 1 - \( F_{STC}^0 \) |
| |1⟩   | 1 - \( F_{STC}^1 \) | \( F_{STC}^1 \)     |

Errors creep in when electron tunnelling occurs when the qubit is initially |0⟩, or when there is no tunnelling resultant from the initial |1⟩ state, caused mainly by the temperature of the reservoir and the energy separation of the two qubit states. High temperatures can thermally excite ground-state electrons to incorrectly tunnel, while
high magnetic fields can increase the effects of relaxation. These give rise to two limiting situations for $V_{\text{STC}}$: readout time limited and $T_1$ limited.

Assuming we have detected a 'blip' using the electrical threshold, there is some probability that it could be due to a $\{0\}$ tunnelling out due to the finite temperature in the system. Determining this probability is the goal of the state-to-charge conversion analysis and was previously derived by Buch [18]; however, we repeat it here for completeness. We denote the state of the qubit by $\psi$ in the basis $\{\{0\}, \{1\}\}$. We define the initial time at the start of the readout phase as $t = 0$, and describe the system dynamics using the rate equation

$$\frac{d\psi}{dt} = \begin{pmatrix} \frac{1}{T_0^{\text{OUT}}} & \frac{1}{T_1} \\ 0 & \frac{1}{T_0^{\text{OUT}}} - \frac{1}{T_1} \end{pmatrix} \psi,$$

which has the solution

$$\psi_0(t) = \frac{\psi_0(0)T_0^{\text{OUT}} + \psi_1(0)t_0^{\text{OUT}}t_1^{\text{OUT}}}{T_0^{\text{OUT}}e^{\frac{t}{T_0^{\text{OUT}}}}} - \frac{\psi_0(0)t_0^{\text{OUT}}t_1^{\text{OUT}}}{T_0^{\text{OUT}}e^{\frac{t}{T_0^{\text{OUT}}}}}e^{\frac{-t}{T_0^{\text{OUT}}}},$$

$$\psi_1(t) = \psi_1(0)e^{\frac{-t}{T_1}},$$

where, $\psi_i(0)$ is the initial probability of being in state $i$ and $T_0^{\text{OUT}} = T_1(t_0^{\text{OUT}} - t_1^{\text{OUT}}) + t_0^{\text{OUT}}t_1^{\text{OUT}}$. The probability that either qubit state generates a 'blip' in the time, $t$, $N_{\text{off}}(t)$ is

$$N_{\text{off}}(t) = 1 - \psi_0(t) - \psi_1(t),$$

and similarly, we label the probability that either qubit state does not generate a 'blip' as

$$N_{\text{on}}(t) = \psi_0(t) + \psi_1(t) = (1 - F_{\text{STC}}^{\text{OUT}})\psi_0(0) + F_{\text{STC}}^{\text{OUT}}\psi_0(0).$$

From these two equations we can calculate the probability that the detected 'blip' was from $|1\rangle$ and not from $|0\rangle$. First, we find $N_{\text{on}}(t)$ when $\psi_0(0) = 1$ as this gives the probability that the $|0\rangle$ does not generate a 'blip',

$$F_{\text{STC}}^{\text{OUT}} = e^{-\frac{t}{T_0^{\text{OUT}}}}.$$
\[ \Gamma_{\text{OUT}}^0(\epsilon) = |1 - f(\epsilon \pm E_Z/2)| \Gamma_{\text{OUT}}, \]
\[ \Gamma_{\text{IN}}^i(\epsilon) = f(\epsilon \pm E_Z/2) \Gamma_{\text{IN}}, \]
\[ f(\epsilon \pm E_Z/2) = \frac{1}{1 + e^{-2\mu_B E_Z/k_B T}}, \]

where \( f(\epsilon \pm E_Z/2) \) is the Fermi–Dirac function with \(-\) for \(|0\rangle\) and \(+\) for \(|1\rangle\). \( \Gamma_{\text{OUT}}^0(\Gamma_{\text{IN}}^i) \) is the maximum tunnel rate out (in) of both qubit states, and \( k_B T \) is the thermal energy of the system. We can use two tunnel rates to find the detuning position about the reservoir, \( \epsilon_{\text{det}} \). Typically, \( \Gamma_{\text{OUT}}^1 \) and \( \Gamma_{\text{IN}}^0 \) are measured during the spin readout protocol. Assuming, \( \Gamma_{\text{OUT}} = \Gamma_{\text{IN}} \), then the ratio between the two times is given by
\[ \frac{t_{\text{OUT}}^0}{t_{\text{IN}}^i} = \frac{f(\epsilon - E_Z/2)}{1 - f(\epsilon + E_Z/2)} = R_t, \]
which gives
\[ R_t = \frac{1 + \frac{\epsilon - E_Z}{\Gamma_{\text{IN}}}}{1 + \frac{\epsilon + E_Z}{\Gamma_{\text{IN}}}}. \]

After rearranging and solving for \( \epsilon \) (neglecting the imaginary solution), we have
\[ \epsilon_{\text{det}} = k_B T \ln \left[ \frac{-E_Z}{2R_t} (1 - R_t + \sqrt{(1 - R_t)^2 + 4R_t e^{\Gamma_{\text{IN}}^0/2}}) \right]. \]

This value of detuning can then be used to obtain the other two tunnel times in the system. For example, \( t_{\text{OUT}}^0 \) can be found from using,
\[ t_{\text{OUT}}^0 = \frac{1 - f(\epsilon_{\text{det}} - E_Z/2)}{1 - f(\epsilon_{\text{det}} + E_Z/2)} t_{\text{OUT}}^1. \]

Alternatively, if only one tunnel time is known, we can find the relative magnitude between two tunnel rates at \( \epsilon = 0 \)
\[ \frac{\Gamma_{\text{OUT}}^1}{\Gamma_{\text{OUT}}^0} = \frac{t_{\text{OUT}}^1}{t_{\text{OUT}}^0} = e^{\Gamma_{\text{IN}}^0}, \]
\[ t_{\text{OUT}}^0 = e^{\Gamma_{\text{IN}}^0} t_{\text{OUT}}^1. \]

The ratio between the \( t_{\text{IN}}^0 \) and \( t_{\text{OUT}}^1 \) states can also be found by making similar assumptions. In this case
\[ \frac{\Gamma_{\text{IN}}^0}{\Gamma_{\text{OUT}}^0} = \frac{t_{\text{OUT}}^0}{t_{\text{IN}}^0} = \frac{f(-E_Z/2)}{1 - f(-E_Z/2)} = e^{\Gamma_{\text{IN}}^i}, \]
\[ t_{\text{OUT}}^0 = e^{\Gamma_{\text{IN}}^i} t_{\text{IN}}^0, \]
and similarly
\[ t_{\text{OUT}}^1 = e^{-\Gamma_{\text{IN}}^i} t_{\text{IN}}^i. \]

Equations (14) and (24) allow for a relatively accurate estimate of the readout visibility for a particular temperature and the qubit energy separation. For spin qubits in a magnetic field \( B \), the qubit energy separation is given by \( E_Z = g \mu_B B \), where \( g \) is the gyromagnetic ratio and \( \mu_B \) is the Bohr magneton. Hence, the tunnel rates which help determine \( V_{\text{STC}} \) have an exponential dependence on both the reservoir temperature and magnetic field strength. The magnetic field also reduces the excited state relaxation time \( T_1 \) at large fields with a \( T_1 \propto B^{-\alpha} \) dependence [17]. Having shown that \( V_{\text{STC}} \) has a strong dependence on the magnetic field and reservoir temperature we now move on to discuss the readout visibility in each of the limiting cases caused by these key factors and how \( V_{\text{STC}} \) can be optimised.

2.1. Optimal state-to-charge visibility

In figure 2(b) we plot the fidelities, \( F_{\text{STC}}^0 \) and \( F_{\text{STC}}^1 \) as well as \( V_{\text{STC}} \) as a function of the readout time. Since \( F_{\text{STC}}^0 \) corresponds to the probability of \(|0\rangle\) not tunnelling out to the reservoir, at \( t = 0 \), \( F_{\text{STC}}^0 = 1 \). The fidelity \( F_{\text{STC}}^1 \) then follows an exponential decay as \(|0\rangle\) becomes more likely to tunnel off to the reservoir. The fidelity \( F_{\text{STC}}^1 \)
represents the probability that the $|1\rangle$ state has tunneled off to the reservoir. At $t = 0$, $F_{\text{STC}}^0 = 0$ as there has not been a chance for the qubit state to tunnel to the reservoir. As $t \to \infty$, $F_{\text{STC}}^0 \to 0$ and $F_{\text{STC}}^1 \to 1$ as the probability that the qubit tunnels approaches unity. Therefore, there is an optimum readout time, $t_{\text{opt}}$ that maximises $V_{\text{STC}}(t)$ and offers the best compromise between $F_{\text{STC}}^0$ and $F_{\text{STC}}^1$. The state-to-charge conversion visibility follows the $F_{\text{STC}}^0$ curve for short $t$ and then follows $F_{\text{STC}}^1$ after the optimal readout time, $t_{\text{opt}}$. The best $V_{\text{STC}}$ is obtained for low reservoir temperatures and for a large qubit energy splitting since this maximises the ratio of $t_{\text{OUT}}^0$ to $t_{\text{OUT}}^1$.

2.2. Readout time limited
The readout time limit occurs when the optimum readout time of the system calculated from the state-to-charge analysis is close to the individual tunnel out times, $t_{\text{OUT}}^0$ and $t_{\text{OUT}}^1$ (figure 2(c)). This will occur when the difference between $t_{\text{OUT}}^0$ and $t_{\text{OUT}}^1$ becomes very small, that is $t_{\text{OUT}}^0/t_{\text{OUT}}^1 \to 1$ making it difficult to find an optimal readout time, hence decreasing both $F_{\text{STC}}^0$ and $F_{\text{STC}}^1$. This is generally an indication that the state levels are not sufficiently separated in energy compared to the temperature broadening of the reservoir. If the relative tunnel rates cannot be changed then this scenario is extremely difficult to overcome since it means the temperature in the system needs to be reduced. We determine that for $F_M > 99\%$ then $t_{\text{OUT}}^0/t_{\text{OUT}}^1 \gtrsim 800$, which for a qubit energy splitting, $E_Z$ corresponds to $E_Z/k_B T \approx 13$. For electron spins in silicon, this corresponds to magnetic field to temperature ratio of $B/T \approx 10$ T K$^{-1}$.

2.3. $T_1$ limited
Finally, the last situation is when the $T_1$ of the $|1\rangle$ level is close to the optimal readout time calculated from the state-to-charge analysis. A distinct plateau in the state-to-charge visibility can be observed that limits $F_{\text{STC}}^1$, shown in figure 2(d). This is due to the large fraction of $|1\rangle$ states relaxing to $|0\rangle$ and not causing a ‘blip’ in the charge sensor response. We find that provided $T_1 \gtrsim 100t_{\text{OUT}}^0$ then $F_M$ can be above 99%. Again, this situation is difficult to overcome without the ability to change the relative tunnel rates.

3. Electrical readout
The electrical visibility $V_E$ is the ability of the charge detector to resolve the ‘blip’ in the detector response, $x$ and is related to the sample rate of the charge sensor, $\Gamma_{\text{IN}} t_{\text{OUT}}$, $t_{\text{OUT}}^0$, and the sensitivity index, $D'$ defined as
\[
D' = \frac{\mu_i - \mu_0}{\sqrt{\frac{1}{2}(\sigma^2_i + \sigma^2_0)},}
\]

where, \(\mu_i(\sigma_i)\) is the mean (standard deviation) of the \(i = 0, 1\) levels of the charge sensor response. In the case when \(\sigma_0 = \sigma_1, D'\) is equivalent to the SNR. Lastly, \(V_s\) also depends on the filter cut-off frequency, \(f_c\), used to filter the charge sensor response

\[
V_E(x; D', \Gamma, t_{\text{OUT}}, t_{\text{IN}}, f_c) = F_F^0 + F_F^1 - 1,
\]

where \(F_F^0\) is the probability of \(|0\rangle\) not causing a 'blip' above a threshold value \(x\) and \(F_F^1\) is the probability of \(|1\rangle\) generating a 'blip' (assuming that the \(V_{\text{STC}} = 1\)) in the charge detector response as per figure 1(c) (right-hand side)

| Qubit | Measurement outcome |
|-------|---------------------|
| \(|0\rangle\) | \(F_F^0\) | \(1 - F_F^0\) |
| \(|1\rangle\) | \(1 - F_F^1\) | \(F_F^1\) |

Errors arise when a 'blip' occurs within the readout time when the qubit is initially \(|0\rangle\), or when there is no 'blip' within the readout time when the initial state is \(|1\rangle\). The key factors that cause these errors, and hence reduce \(V_E\), include a sample rate too slow to detect fast 'blips', high noise that disguises a potential 'blip', or filtering such that fast 'blips' are removed from the measured signal.

Optimisation of the electrical fidelity requires finding the threshold that gives the maximum ability to distinguish between \(|0\rangle\) and \(|1\rangle\) in the detector response. Therefore, we must calculate the value of the detector response, \(x\) over the duration of \(t_r\) which maximises \(V_E\), as it will be used as the optimal threshold value, \(x_{\text{opt}}\), to distinguish between the two states.

First we want to find the probability \(P_{\text{miss}}\) of missing a 'blip' due to the finite sample time of the detector. The tunnel out event of the \(|1\rangle\) state can occur anywhere within the interval, \(\delta t = t_r - \tau\) where \(\tau\) is some point in time less than the sample time, \(t_r = 1/\Gamma\). Therefore, the probability of detecting a high level 'blip' is a sum between the exponential distribution of \(|1\rangle\) normalised over the interval 0 to \(t_r\)

\[
p_1(t) = \frac{e^{-t/\tau_{\text{det}}}}{t_{\text{OUT}}^1(1 - e^{-t/\tau_{\text{det}}})}
\]

and the distribution of \(|0\rangle\)

\[
p_0(t) = \frac{e^{-t/\tau_{\text{det}}}}{t_{\text{IN}}^0
\]

such that,

\[
P_{\text{det}} = \int_0^t p_1(t - \tau) \cdot p_0(\tau) \, d\tau
\]

\[
= \int_0^t \frac{e^{-t/\tau_{\text{det}}}}{t_{\text{OUT}}^1(1 - e^{-t/\tau_{\text{det}}})} \frac{e^{-\tau/\tau_{\text{det}}}}{t_{\text{IN}}^0} \, d\tau.
\]

This convolution gives the probability of detecting a 'blip' of length, \(t\)

\[
P_{\text{det}} = \left(1 - e^{t/\tau_{\text{det}}}ight) t_{\text{IN}}^0 t_{\text{OUT}}^1
\]

where \(R_{\text{det}} = t_r/\tau_{\text{OUT}}\). Next, we find the total probability of missing all 'blips', \(P_{\text{miss}}\). We are interested in the time \(t_{\text{OUT}}^1 + t_{\text{IN}}^0 \leq t_r\) for \(t_{\text{IN}}^0 > 0\) which, for simplicity, is equivalent to \(t_{\text{OUT}}^1 \leq t_r/2\) as depicted by figure 3(a), hence

\[
P_{\text{miss}} = 1 - \int_0^{t_r/2} P_{\text{det}} \, dt
\]

which results in

\[
P_{\text{miss}} = 1 - \frac{(1 - e^{t_{\text{OUT}}^1/\tau_{\text{det}}}) R_{\text{det}}^1}{(1 - e^{t_{\text{OUT}}^1}) (R_{\text{IN}}^1 - R_{\text{IN}}^0)}
\]

where \(R_{\text{IN}}^0 = t_r/\tau_{\text{IN}}\). All the 'blips' with values of \(t_{\text{OUT}}^1\) and \(t_{\text{IN}}^0\) that fall within the blue shaded region of figure 3(a) contribute to \(P_{\text{miss}}\) and have the general shape of the example shown in figure 3(b).

Assuming we know the type and magnitude of the noise and average levels of the two states \(\mu_i\), we can write the individual readout state fidelities as a function of the detector response.
The missed ‘blips’ due to the finite readout time are taken into account in the STC. The electrical visibility that distinguishes between the two qubit levels, from combining equations which corresponds to the value of $t_{\text{fin}}$. If $t_{\text{fin}} + t_{\text{OUT}} \leq t_{r}$ as well, then the ‘blip’ will be undetectable in the detector response, and hence contribute to $P_{\text{miss}}$. The points that fulfil the criteria of an undetectable ‘blip’ in terms of $t_{\text{OUT}}$ and $t_{\text{fin}} + t_{\text{OUT}}$ are depicted by the blue shaded region. The shaded area is equivalent to an area with height $t_{\text{fin}} + t_{\text{OUT}} = t_{r}$ and width $t_{\text{fin}}$ which is used in the integration for calculating $P_{\text{miss}}$.

Figure 3. Electrical readout probabilities (a) depiction of an example ‘blip’ that would be undetectable and contribute to $P_{\text{miss}}$. ‘Blips’ in the detector response can be described using $t_{\text{OUT}}$ and $t_{\text{fin}} + t_{\text{OUT}}$. By definition, for a ‘blip’ to physically exist, $t_{\text{fin}} > 0$. The time until the initial edge of the ‘blip’ is given by $t_{r}$. If $t_{\text{fin}} + t_{\text{OUT}} \leq t_{r}$, then the ‘blip’ will be undetectable in the detector response, and hence contribute to $P_{\text{miss}}$. (b) The points that fulfil the criteria of an undetectable ‘blip’ in terms of $t_{\text{OUT}}$ and $t_{\text{fin}} + t_{\text{OUT}}$ are depicted by the blue shaded region. The shaded area is equivalent to an area with height $t_{\text{fin}} + t_{\text{OUT}} = t_{r}$ and width $t_{\text{fin}}$ which is used in the integration for calculating $P_{\text{miss}}$. (c) Distribution histograms for the different limiting cases for $V_{E}(x)$.

The missed ‘blips’ due to the finite readout time are taken into account in the STC. The electrical visibility that distinguishes between the two qubit levels, from combining equations (30), (38) and (39), is then given by

$$V_{E}(x) = (1 - P_{\text{miss}}) \left[ C_{0}(x) - C_{1}(x) \right].$$

We can immediately see the importance of $P_{\text{miss}}$ on the readout fidelity as it limits the maximum $V_{E}(x)$. The optimum threshold, $x_{\text{opt}}$ is the value of $x$ that maximises $V_{E}(x)$. This can be found by differentiating equation (41) and setting the condition

$$\frac{dV_{E}(x)}{dx} = (1 - P_{\text{miss}}) \left[ \frac{dC_{0}(x)}{dx} - \frac{dC_{1}(x)}{dx} \right] = 0,$$

which corresponds to the value of $x$ where the two partial readout distributions are equal, $\frac{dC_{0}(x)}{dx} = \frac{dC_{1}(x)}{dx}$. We can now use these equations to investigate various limiting cases of the system. Similarly to STC, we now describe the optimal scenario (not limited by any particular experimental parameter). We also show the schematic phase diagram for the different limiting cases for $V_{E}(x)$ in figure 4(a).

3.1. Optimal electrical visibility

In figure 4(b)(i) we plot the fidelities, $F_{E}^{0}$ and $F_{E}^{1}$ as well as the electrical visibility, $V_{E}(x) = F_{E}^{0} + F_{E}^{1} - 1$ as a function of the detector response threshold, $x$. The $F_{E}^{0}$ level corresponds to the probability that the maximum value of the readout trace of $|0\rangle$ will be less than $x$. Therefore, in figure 4(b)(i) the probability, $F_{E}^{0}$ begins at zero for
small x and increases with a skewed Gaussian distribution as x increases until it reaches 1, indicating that the distribution will always be less than those values of x where $F_E^0 = 1$. For $F_E^0$, the condition is reversed; that is, $F_E^0$ corresponds to the probability that the maximum value of the readout trace of |1⟩ will be greater than x. For small x, $F_E^0 = 1$ indicating that the readout trace will always have a maximum value above x. As x increases the probability that the maximum value lies above x decreases and eventually there will never be a maximum value of the readout trace that is above x which corresponds to $F_E^0 = 0$. The optimal threshold value, $x_{opt}$ can vary dramatically and its exact position will depend on what factor is limiting $V_G(x)$.

The lower panel in figure 4(b)(ii) shows the distributions $P_F^0$ (green) and $P_F^1$ (blue) that can be used to visualise the difference between the two measured levels in the readout trace. For high fidelity readout, these two distributions do not overlap and are well separated in x.

### 3.2. Sample rate limited

The first limiting case we consider is when the sample rate of the detector is too low to be able to detect fast tunnelling events. This is characterised by a flat plateau region in the electrical fidelity in figure 4(c)(ii) at value below $V_G(x) = 1$. The sample rate does not affect the state-to-charge conversion fidelity and hence it can be arbitrarily high. The individual readout distributions are clearly distinguishable; however, there is a large number of events in the $P_F^0$ state that lie under $P_F^0$ shown in figure 4(c)(ii). These events are faster than $\Gamma_1$ and are not measured by the detector, hence reduce $F_E^0$. Note that since the tunnelling events are stochastic there will always be a finite number of events faster than $\Gamma_1$. This limiting case can be easily remedied by increasing the sample rate of the sensor. Using equation (37) for $P_{\text{miss}}$, we find that the required sample rate for $F_M = 99\%$ is $\Gamma_1 \gtrsim 12/\Gamma_0$.

### 3.3. Noise limited

If the detector is not sufficiently filtered or has poor noise characteristics then the ability to distinguish between the two levels becomes difficult. When this is the case, the electrical visibility is limited by the noise of the charge detector. NL electrical fidelities are characterised by an almost symmetrical peak in $V_G(x)$ (figure 4(d)(i)) where the two readout distributions clearly overlap with each other, see figure 4(d)(ii). This makes it difficult to optimise the detector response threshold, and reduces both $F_E^0$ and $F_E^1$. The NL scenario is more difficult to overcome compared to the SRL situation. We can optimise the charge sensor using low-noise amplifiers [30, 31] or adjust the filter frequency to limit some of the noise in the device. However, reducing the filter frequency can
also have a detrimental effect on the readout fidelity as high frequency ‘blips’ can also be attenuated. Assuming white Gaussian noise, we find that it is possible that $F_M > 99\%$ can be achieved with a sensitivity index $D'$ as low as 3 provided $\Gamma$, is sufficiently fast to resolve the ‘blips’ in the readout trace. Note that this $D'$ is based on the integration time and sample rate rather than the readout time which is required to be longer due to the stochastic tunnelling processes.

3.4. Filter limited
The detector is normally low pass filtered in readout experiments to remove high frequency noise from the readout trace. However, this also filters the high frequency ‘blips’ which reduces $F_M$, as well as the overall electrical visibility. In the case of a FL charge detector the peak in the electrical visibility will be asymmetrical with the 0 distribution being quite sharp compared to the 1 level as shown in figure 4(e)(i). This is due to $P^0_F$ exhibiting an extremely long tail extending towards $P^1_F$ (figure 4(e)(ii)). This scenario can be readily fixed by increasing the filter cut-off frequency. However, as the filter frequency is increased more noise couples into the charge sensor. Therefore, there is a trade off between the FL and NL scenario. The filter limit is much easier to improve and should essentially be increased until the noise in the system begins to dominate the electrical fidelity, that is, when the peak becomes symmetrical as in figure 4(d)(i).

4. Applications and discussion
In this section we describe a number of applications and extensions of the model presented in the paper.

4.1. Initialisation fidelity
The initialisation fidelity can be found using a similar method to the STC visibility calculation. The rate equation model in the basis $(|z\rangle, |0\rangle, |1\rangle)$ where $|z\rangle$ is the state when the qubit is emptied is given by

$$\frac{d\psi}{dr} = \begin{pmatrix} -\frac{1}{\tau_{\text{IN}}} & \frac{1}{\tau_{\text{OUT}}} & \frac{1}{\tau_{\text{OUT}}} \\ \frac{1}{\tau_{\text{IN}}} & -\frac{1}{\tau_{\text{OUT}}} & \frac{1}{\tau_{\text{OUT}}} \\ \frac{1}{\tau_{\text{IN}}} & 0 & -\frac{1}{\tau_{\text{OUT}}} \end{pmatrix} \psi,$$

and assume the system starts in this state, $\psi_z(0) = 1$. The solution to this system of equations can be found analytically; however, the solutions are rather unwieldy. Instead, to get an approximation for the ideal initialisation time, we assume the electron cannot tunnel back to the reservoir, that is $t_{\text{OUT}} = \infty$. It is worth noting that having a short $T_1$ aids in initialisation as the $|1\rangle$ is expected to decay quickly to the $|0\rangle$ state even if it were accidentally loaded. Therefore, we are interested in the regime where $\tau_{\text{IN}} \ll T_1$.

The solution to the system of equations when $\psi_z(0) = 1$ is given by

$$\psi_0(t) = e^{-\frac{t(T_1 + T_N)}{T_N^2}} \left[ t_{\text{IN}} T_1 \left( e^{-\frac{(T_1 + T_N)}{T_N}} - e^{-t/T_1} \right) + t_{\text{IN}} T_1 \left( e^{-\frac{(T_1 + T_N)}{T_N}} - e^{t/T_1} \right) + t_{\text{IN}} T_1 \left( e^{-\frac{(T_1 + T_N)}{T_N}} - e^{t/T_1} \right) \right],$$

$$\psi_1(t) = t_{\text{IN}} e^{-\frac{t+\frac{t^2}{T_N}}{T_N^2}(e^{t/T_1} - e^{-t/T_1})},$$

where, $T_N = T_1 (t_{\text{IN}} + t_{\text{OUT}}^0) - t_{\text{IN}} t_{\text{IN}}$ and $T_1 = t_{\text{IN}}^0 + \frac{t_{\text{IN}}}{t_{\text{IN}}^0}$. Finally, we define the initialisation fidelity as the probability of loading a spin down electron

$$F_I = \psi_0(t).$$

The optimum $F_I$ is more difficult to define compared to $F_M$ because of the influence of $T_1$. Therefore, we define the optimal $F_I$ that maximises $F_I$ while also minimising the initialisation time, $t_I$. This time corresponds to the maximum chance of loading a $|1\rangle$ state. Therefore, $t_I$ can be calculated by differentiation of $\psi_I(t)$.
Figure 5. Calculation of the initialisation fidelity and optimisation of the initialisation time. Probability of the three qubit states during initialisation calculated using equation (43) with the parameters from Broome(L). The system begins in the empty state, \(|z\rangle\) and then \(|0\rangle\) state quickly becomes populated due to tunnelling from the reservoir to the qubit. The \(|1\rangle\) state remains almost completely unpopulated due to the long tunnel in time, \(t_{\text{IN}} \approx 2 \text{ s}\). Two estimates for the initialisation time are shown by the dotted lines, \(t_I\) (red) and \(\sqrt{2\pi} t_I\) (black). The latter of these two times is more conservative and has an initialisation fidelity, \(F_I = 98.9\%\). The inset shows the initialisation fidelity as a function of initialisation time.

This time represents the time where the \(T_1\) process starts dominating the initialisation process. That is, where the majority of the \(|0\rangle\) are due to relaxation of the \(|1\rangle\) state. Therefore, it represents a minimal time that maximises the initialisation fidelity.

To demonstrate the initialisation fidelity calculation, in figure 5 we show the solution to equation (43) for the data of Broome(L) [32]. We initialise the system in the empty state and then watch how the qubit states are populated as a function of time. The \(|0\rangle\) state quickly becomes populated due to direct tunnelling from the reservoir to the qubit. The \(|1\rangle\) state becomes slightly populated during the initial tunnelling period \(t < 20 \text{ ms}\); however, due to the slow tunnel rate into the \(|1\rangle\) state from the reservoir there is never any significant population. We then show two different initialisation times, \(t_I\) calculated from equation (49) and \(\sqrt{2\pi} t_I\). The first time, \(t_I = 26 \text{ ms}\) gives an initialisation fidelity, \(F_I = 97.2\%\). The second initialisation time estimate, \(\sqrt{2\pi} t_I = 65 \text{ ms}\) is more conservative and gives an initialisation fidelity, \(F_I = 98.9\%\). The factor \(\sqrt{2\pi}\) was chosen based on examining a number of different initialisation fidelity calculations and shows a good compromise between initialisation time and fidelity. To obtain the actual initialisation fidelity (and time) the full system should be used to calculate \(\psi(t)\). However, the above analysis offers a simple estimate to set the initialisation time.

4.2. Calculation of \(V_E\) assuming white Gaussian noise
White Gaussian noise assumes that the levels, \(|0\rangle\) and \(|1\rangle\) have a Gaussian noise distribution with a constant spectral density at all frequencies. In this section, we find \(C_0\) and \(C_1\) for this type of noise.

We denote the mean levels of \(|0\rangle\) and \(|1\rangle\) in the x domain as \(\mu_0\) and \(\mu_1\). Both \(|0\rangle\) and \(|1\rangle\) have a Gaussian noise distribution centred about their mean, \(\mu_i\)

\[ \mathcal{N}(x; \mu_i, \sigma_i^2) = \frac{1}{\sqrt{2\pi} \sigma_i} e^{-\frac{(x - \mu_i)^2}{2\sigma_i^2}}. \]  

The levels have an associated noise \(\sigma_i^2\) and we define the sensitivity index,

\[ D' = \frac{\mu_1 - \mu_0}{\sqrt{\sigma_1^2 + \sigma_0^2}}, \]  

which reduces to the SNR if \(\sigma_1 = \sigma_0\).

During the readout process, we are interested in the maximum value of the detector response during the readout time. Therefore, we want to build our state distributions by taking the maximum of \(\mathcal{N}(x; \mu_i, \sigma_i^2)\) over a single readout trace.

First, we will consider the lower level, \(|0\rangle\). For a fixed sample number of the readout trace \(n_r = t_r/t_s\), the CDF for the maximum of a sampled Gaussian is simply the product of \(n_r\) individual Gaussian distributions.
Therefore, $C_0(x)$ is given by

$$C_0(x) = P_0(x)^{n_r},$$

where $P_0(x)$ is the CDF of a single Gaussian, given by

$$P_1(x) = \frac{1}{2} \left[ 1 + \text{erf} \left( \frac{x - \mu_1}{\sqrt{2} \sigma_1} \right) \right].$$

$$E(n, n_i, n_{\text{max}}) = \frac{1}{n_r (1 - e^{-(n_{\text{max}} - 1)/n_r})},$$

where $n = t/t_s$ is the sample number in the readout trace, $n_{\text{max}}$ is the maximum number of samples in the distribution, and $n_i = t_{\text{OUT/IN}}^i/t_s$ is the characteristic length of $|n|/|0\rangle$ in the readout trace. The $C_0(x)$ probability distributions by $E(n)$

$$C_i(x) = \int_{|s| = 1}^{n_{\text{max}} - 1} E(s, n_0, s_r - 1) \times \left[ \int_{n_{\text{min}} - i}^{n_{\text{min}} - s} \eta(n)S_n dn + \int_{n_{\text{min}} - i}^{\infty} \eta(n)S_n dn \right] ds,$$

where,

$$S_n = \left( \frac{n}{n_r} P(x) + \frac{n_i - n}{n_r} P_0(x) \right)^{n_r},$$

and

$$\eta(n) = e^{-(n - n_i)/n_1}. $$

Here, the integration is carried out over two different scenarios. The first integral in the brackets describes the 'blips' of a length, $n_r - s$, where $s$ is the length of the readout trace before the 'blip'. These 'blips' are fully resolved during the readout. The second integral describes 'blips' that are actually longer than the $n_r - s$ and therefore become artificially shortened to exactly $n_r - s$. $S_n$ is the relative probability of $|0\rangle$ and $|1\rangle$ CDFs over the readout trace weighted by $\eta(n)$ over the entire readout trace. Equation (55) can be numerically integrated to obtain $C_i(x)$.

At this time, we introduce the effect of filters on the readout trace. The filter in readout experiments is usually low-pass and is characterised by a cut-off frequency, $f_c$ which we define as the -3 dB attenuation in the gain amplitude factor, $G(f_1, f_2)$. The noise, $\sigma^2_i(f_c)$ is then given by

$$\sigma^2_i(f_c) = 2A^2_n f_c,$$

where $A_n$ is the noise power spectral density in units of $x^2/\sqrt{\text{Hz}}$. This value can be found experimentally by simply measuring $\sigma^2_i(f_c)$ at a known $f_c$ and inverting equation (58).

Finally, the filter also attenuates the amplitude of the 'blip' in the readout trace since it has some frequency components above $f_c$. To account for this, we convert the cut-off frequency of the filter into the inverse number domain, $m = 1/n$ of the readout trace

$$m_r = \frac{f_c}{\Gamma_1},$$

and apply it to $\mu_1$

$$\mu_1(n) = \max[h_p(n, m_r)](\mu_1 - \mu_0) + \mu_0,$$

where $h_p(n, m_r)$ is the pulse response of the filter with cut-off, $m_r$. Therefore, any $|1\rangle$ levels with a frequency number much less than $m_r$ will be limited to $\mu_0$. These new noise, $\sigma^2_i(f_c)$ and mean, $\mu_1(n)$ parameters then need to be included in $P(x)$ in equations (52) and (55). The values of $\sigma^2_i(f_c)$ and $\mu_1(n)$ will be different depending on the filter used in the experimental setup. However, in supplementary materials II we give the calculation for an 8th order Bessel filter commonly used in experiments.

4.3. An example of the optimisation of experimental parameters

We now apply the model to a real experiment and investigate the parameter space to maximise the readout fidelity. To demonstrate how to further improve readout fidelity we use the results from Broome(L)[32] who already achieved high fidelity measurements of electron spin states in a 2P donor dot system. In figure 6(a) we
use our model, and the experimental values from Broome(L) to plot the phase diagram of the optimal measurement fidelity by sweeping the sample rate of the charge sensor, \( \Gamma_s \), and the filter cut-off frequency, \( f_c \). By directly comparing the phase diagram to that in figure 4(a) and examining the state distributions in figure 6(b) we can immediately determine that the slow sample rate of the charge sensor in this experiment is the main factor that limits the readout fidelity, similar to figure 4(c). The Shannon–Nyquist sampling theorem is usually assumed to set the sample rate of the charge sensor, \( \Gamma_s = 2f_c \) [33] (red line in figure 6); however, the Shannon–Nyquist theorem only applies to signals that contain no frequency components above the filter cut-off frequency. When performing readout this is never true as the tunnel events follow an exponential distribution and there will always be some events that are above the filter frequency. Using the theory presented here, we obtain a fidelity of 97.0% using \( \Gamma_s = 5 \text{ kHz} \) and \( f_c = 1 \text{ kHz} \) as in the experiment (black diamond in figures 6(a) and (b)). The readout fidelity was limited by a combination of the low filter frequency and a slightly slow sample rate. However, our analysis shows this could be improved to \( F_M = 97.9\% \) by simply using \( \Gamma_s = 5.5 \text{ kHz} \) and \( f_c = 2 \text{ kHz} \), shown in figure 6(c). This is a small but significant increase in fidelity that can be easily identified, demonstrating the value of the model presented here.

By using our model with parameters obtained from previous results (see supplementary materials III), we can compare calculated fidelities with those quoted in past work, which we show in table 1. Whilst the fidelities of the model agree well with previous quoted fidelities, since the methods used to calculate the fidelities in each paper have differed, it has not been possible to make direct comparisons between them. The analysis presented here makes it possible to compare different samples for the first time and we show that by including the state-to-charge conversion, some of the previously quoted fidelities will be reduced. Also, the error in \( F_M \) is lower for calculated values than for reported values (e.g. Broome(L) and Broome(R) [32]) as expected since our analytic model used for calculations removes numerical errors obtained using a Monte-Carlo simulation. Note that the parameters required to calculate the readout fidelity are often not quoted. This makes it difficult to calculate the correct readout fidelity and approximations must be used.

The limiting factor in the majority of the previous results is the electrical visibility. This is mainly due to the sensitivity index \( D' \) of the charge detector used in the experiment. The state-to-charge conversion visibility is typically lower in GaAs gate defined quantum dots [16, 34] due to the comparatively lower Zeeman splitting at the same magnetic field values. Therefore, larger magnetic fields must be applied to achieve the same quality of state-to-charge conversion as compared with silicon based devices. In addition, this has the adverse effect of decreasing the electron \( T_1 \) relaxation times and hence, further decreases the state-to-charge conversion visibility.

Figure 6. Optimising the sample rate and filter frequency for single-shot spin readout. (a) The measurement fidelity, \( F_M \), as a function of sample rate, \( \Gamma_s \), and filter frequency, \( f_c \), for an electron spin qubit on a 2P donor dot (Broome(L)) [32]. The black diamond is \( F_M \) for \( \Gamma_s = 5 \text{ kHz} \) and \( f_c = 1 \text{ kHz} \) used in the experiment. The blue star is the optimum filter cut-off frequency, \( \Gamma_s = 2f_c \), the sample rate usually assumed to be correct based on the Shannon–Nyquist sampling theorem [33], which falls below the optimal fidelity point. (b) The electrical visibility and state distributions as a function of the detector response for \( \Gamma_s = 5 \text{ kHz} \) and \( f_c = 1 \text{ kHz} \) used in the experiment (black diamond). There is a significant number of 1 states that lie underneath the 0 state distribution (orange arrow). There is also a large tail on the 1 state distribution indicating that the measurement fidelity is limited by a combination of sample rate and filter frequency. (c) The electrical visibility and state distributions as a function of the detector response for \( \Gamma_s = 5.5 \text{ kHz} \) and \( f_c = 2 \text{ kHz} \) (blue star). Using these parameters there is a clear change (1.9%) in the optimal \( V_E \) caused by fewer missed 1 states underneath the 0 state readout distribution (orange arrow). This amounts to an increase in \( F_M \) by 0.9%.
Table 1. Comparison of reported fidelities and those calculated using the model in this paper with the same reported readout time $t_{\text{rep}}$ and detector response threshold $x_{\text{rep}}$. We observe that the state-to-charge visibilities $V_{\text{STC}}$ have a large impact on the overall measurement fidelity, thus should always be taken into account.

| References | $t_{\text{rep}}$ (ms) | $x_{\text{rep}}$ (nA) | $V_{\text{STC}}$ (%) | $V_E$ (%) | $F_M$ (%) | $V_{\text{STC}}$ (%) | $V_E$ (%) | $F_M$ (%) |
|------------|----------------------|------------------------|----------------------|----------|----------|----------------------|----------|----------|
| [16] Elzerman | 0.5 | 0.73 | N/A | 65.0 | 82.5 | 79.9 ± 1.8 | 52.7 ± 0.2 | 71.0 ± 0.6 |
| [17] Morello | 100 | 1.1 | N/A | 92.0 | 96.0 | 100 | 74.7 | 87.3 |
| [19] Simmons | 200 | 8.8 | N/A | 94.0 | 97.8 ± 0.3 | N/A | N/A |
| [34] Nowack(R) | 2 | 220 | N/A | 86 ± 1 | 77.1 ± 1.8 | 94.8 | 86.5 ± 0.9 |
| [24] Pla | 1 | 72 ± 1 | 82 ± 2 | 77 ± 2 | 40.1 | 89.4 | 67.9 |
| [18] Buch | 40 | 5.4 | N/A | 96.5 | 96.1 | 96.1 | 92.9 | 94.6 |
| [20] Veldhorst | 1 ms$^a$ | 97.0 | 96.2 | 96.5 | 95.7 | N/A | N/A |
| [22] Watson(D$^0$) | 55 ± 0.05 | 120 | 99.6 | 99.6 | 99.6 ± 0.2 | 99.4 | 99.5 ± 0.1 |
| [22] Watson(D$^-$) | 1 ± 0.005 | 1.2 | 99.5 | 97.4 | 98.4 | 99.2 ± 0.1 | 96.6 ± 0.5 | 97.9 ± 0.3 |
| [25] Watson(D1) | 58 | 170 | 99.9 | 99.8 | 99.9 | 98.4 ± 0.1 | 99.2 | 92.1 |
| [25] Watson(D2) | 62 | 170 | 99.9 | 99.7 | 99.8 | 99.9 | 98.3 ± 0.1 | 99.1 |
| [32] Broome(L$^i$) | 10.5 ± 0.1 | 22 ± 1 mV | 97.9 ± 0.1 | 94.6 ± 1.0 | 96.2 ± 1.1 | 97.9 ± 0.5 | 96.2 ± 0.1 | 97.1 ± 0.3 |
| [32] Broome(R) | 109 ± 30 | 26 ± 2 mV | 98.7 ± 0.2 | 96.5 ± 2.0 | 97.6 ± 2.1 | 98.7 ± 0.6 | 96.5 ± 0.1 | 97.6 ± 0.3 |

$^a$ Names within parenthesis are taken from initial reference to distinguish between readout performed on different quantum dots/transitions.

$^b$ Value estimated from the figures appearing in a given reference.
since there is an increased chance that the spin state relaxes before being measured. Nevertheless, straight forward improvements to readout fidelities are possible for a range of previous results by optimising the values used for the readout time \( t_{\text{opt}} \) and detector response threshold \( x_{\text{opt}} \). We perform this optimisation for each previously reported experiment and present the results in table 2. Most of the optimisations result in small, yet significant improvements to the measurement fidelity \( F_M \) shown by the gain (optimised \( F_M \) minus calculated \( F_M \)) with the most notable increasing \( F_M \) by over 8%.

### 4.4. Minimisation of the readout time though optimisation of the qubit tunnel rates

Ideally the qubit should be readout as fast as possible while still maintaining high measurement fidelity. This is particularly important when the qubits are measured sequentially and will be vital for making scalable quantum computers as fast as possible. The main limiting factor to the speed of the readout is the noise as it scales with increasing filter cut-off frequency. Therefore, we need to find the highest filter frequency where we can still perform high fidelity readout. To investigate this we need to find the dependency of \( t_{\text{OUT}}^0 \) as a function of the sensitivity index \( D' \) (we assume \( \sigma_0 = \sigma_1 \) for simplicity). We first calculate \( f_c \) as a function of \( D' \)

\[
f_c = \frac{(\mu_1 - \mu_0)^2}{2D'^2A_n^2}. \tag{61}
\]

The optimisation of the qubit tunnel rates that determine the 0 and 1 levels is difficult due to the many factors involved in the fidelity calculations. Firstly, we assume that the readout is performed at zero detuning between the spin states, such that \( t_{\text{OUT}}^0 = t_0^0 \). Secondly, \( T_1 \) and \( t_{\text{OUT}}^0 \) are both much longer than \( t_{\text{OUT}}^0 \) so that the STC visibility is not limiting the overall readout fidelity (\( > 10^3 t_{\text{OUT}}^0 \), corresponding to \( \Delta E \approx 18k_BT \)). Finally, for simplicity we use the Shannon–Nyquist theorem to set the sample rate of the charge sensor, \( \Gamma_s = 2f_c \) despite this sample rate not being optimal.

Using the assumptions outlined in the above, we show in figure 7(a) the normalised fastest tunnel rate where \( F_M > 99\% \), \( \Gamma_{\text{OUT}} = 1/t_{\text{OUT}}^0 \) of the qubit \( |1\rangle \) state for \( \Delta E \approx 18k_BT \) (red) and \( \Delta E \approx 13k_BT \) (blue) as a function of SNR \( D' \) obtained by changing the filter cut-off frequency. The latter case (\( \Delta E \approx 13k_BT \) is plotted as this is approximately the lowest qubit energy splitting where \( F_M > 99\% \) can still be achieved. We can see a clear peak in the tunnel rate where the fastest readout time can be achieved. For \( \Delta E \approx 18k_BT \) the fastest tunnel rate occurs near \( D' \approx 5.75 \), while for \( \Delta E \approx 13k_BT \) the peak is slightly shifted to \( D' \approx 6 \). The fact that there is an optimal \( D' \) may be somewhat surprising. For low \( D' \) the filter frequency is high and as a result to achieve high fidelity readout the tunnel times must be quite long compared to the filter frequency to ensure there are enough high level points in the charge sensor trace. This is to account for the lower noise, which essentially means that the 1 state must be sampled more to obtain a high probability of a high maximum charge sensor response. At high \( D' \) the filter frequency is low and as a result the tunnel time must be slow to ensure that none of the tunnel events are attenuated and occur below the charge sensor threshold. This means that although the ratio \( f_c/\Gamma_{\text{OUT}} \) may be the smallest for high \( D' \) the tunnel time is still slower compared to the lower noise case. The optimal \( D' \) occurs where these two competing effects are minimised.

### Table 2. Readout fidelities calculated using the analytic model presented here while using the original reported experimental parameters (see supplementary material III) with optimised values for the readout time \( t_{\text{opt}} \) and detector response threshold \( x_{\text{opt}} \). This optimisation improved fidelities up to 8% compared to fidelities calculated with reported thresholds (gain is equal to optimised \( F_M \) minus the calculated \( F_M \) from table 1).

| References | \( t_{\text{opt}} \) | \( x_{\text{opt}} \) | \( V_{\text{STC}}(\%) \) | \( V_E(\%) \) | \( F_M(\%) \) | Gain (%) |
|------------|-----------------|-----------------|-----------------|----------|----------|-------|
| [16] Elserman | 0.46 ± 0.01 ms | 0.74 ± 0.01 | 79.9 ± 1.8 | 67.6 ± 0.2 | 75.8 ± 0.6 | +4.8 |
| [17] Morello | 175 ps | 1.52 nA | 100.0 | 92.4 | 96.2 | +8.9 |
| [19] Simmons | 139 ± 7 ms | 2118 ps | 98.1 ± 0.3 | 92.5 ± 0.1 | 95.4 ± 0.2 | N/A |
| [34] Nowack(R) \(^a\) | 1.65 ± 0.04 ms | 260 ps | 77.6 ± 1.8 | 97.2 | 87.7 ± 0.9 | +1.2 |
| [24] Pla | 0.55 ms | 646 ps | 47.9 | 92.7 | 72.2 | +4.3 |
| [20] Buch | 22 ms | 7.2 nA | 97.4 | 94.2 | 95.9 | +1.3 |
| [20] Veldhorst | 0.15 ms | 344 ps | 99.2 | 91.6 | 95.4 | N/A |
| [22] Watson(D) \(^a\) | 53.4 ± 5 ms | 241 ps | 99.6 ± 0.2 | 99.4 | 99.5 ± 0.1 | 0.0 |
| [22] Watson(D) \(^b\) | 98.0 ± 0.06 ms | 1.32 nA | 99.2 ± 0.1 | 97.1 ± 0.5 | 98.2 ± 0.3 | +0.3 |
| [23] Watson(D1) \(^b\) | 58.5 ± 2.6 ms | 188 ± 1 nA | 99.9 | 99.5 ± 0.1 | 99.7 | +0.5 |
| [25] Watson(D2) \(^b\) | 57.4 ± 3 ms | 187 ± 1 nA | 99.9 | 99.3 ± 0.1 | 99.6 | +0.5 |
| [32] Broome(L) \(^b\) | 10.6 ± 0.2 ms | 22.8 ± 0.2 mV | 97.9 | 96.2 ± 0.1 | 97.1 ± 0.3 | 0.0 |
| [32] Broome(R) \(^b\) | 211 ± 7 ms | 27.2 ± 0.1 mV | 98.7 | 96.0 ± 0.1 | 97.7 ± 0.3 | +1.0 |

\(^a\) Names within parenthesis are taken from initial reference to distinguish between readout performed on different quantum dots/ transitions.
The minimum $t_{\text{OUT}}^i$ for $F_M = 99\%$ is $t_{\text{OUT}}^i = \Delta E = 13k_BT$ for $\Delta E = 18k_BT$ (red) and $t_{\text{OUT}}^i = 50/f_i$ for $\Delta E = 13k_BT$ (blue). Note that these plots represent the minimum tunnel rate where $F_M = 99\%$. Therefore, as an additional investigation we also plot $t_{\text{OUT}}^i$ as a function of $D'$ to find the border between $F_M > 99\%$ and $F_M < 99\%$ in figure 7(b). The region above the lines show where $F_M > 99\%$ can be obtained for both $\Delta E = 18k_BT$ (red) and $\Delta E = 13k_BT$ (blue). Below the line the total measurement fidelity is always less than 99\%. The results here show there is a large variation of the fastest tunnel rates that can be obtained depending on the $D'$ of the charge sensor. Importantly, it appears that increasing the $D'$ above $\sim 6$ appears to have a minimal effect on the overall measurement fidelity (demonstrated by the plateau region in figure 7(b)). Finally, we note that these plots were generated assuming white Gaussian noise and therefore may not be entirely applicable to charge sensors with, for example, $1/\Omega$ dominated noise.

4.5. Extension to sequential multi-qubit readout

Single electron spin measurement has already been demonstrated over ten years ago [16] and the semiconductor quantum computing field is moving towards sequential multi-spin readout [25, 34]. As such, we note the only extension to the model presented here to incorporate multi-spin readout is to take into account the extra wait time while reading out the other qubit(s). Neglecting any cross-talk between the qubits during the individual qubit readout time we only need to take into account the relaxation of the $|1\rangle$ state. For the following analysis we assume that when the other qubit(s) is/are not being measured they have no probability of tunnelling out to the reservoir, that is, $t_{\text{OUT}}^i = t_{\text{OUT}}^i \to \infty$. Therefore, the only relevant time scale is the relaxation time of $|1\rangle$, $T_1$.

We assume that qubit $i$ is measured for a time, $t_{m,i}$, which can be optimised independently using equation (15) and that the total readout time for all qubits is $T_m = \sum_i t_{m,i}$. In addition, we define the total time before qubit $i$ is measured as the wait time for qubit $i$

$$t_{w,i} = \sum_{j=1}^{i-1} t_{m,j}, \quad \forall i > 1, \quad \text{and} \quad t_{w,1} = 0.$$

The only modification to account for the extended wait time is a multiplicative factor in equation (13) that accounts for the probability that $|1\rangle$ relaxes during the measurement of the other qubits

$$F_{\text{SCT},i} = \Lambda_i F_{\text{SCT},i}(t_{\text{OUT},1}, t_{\text{OUT},i}, T_{1,i}),$$

where $\Lambda_i = \exp(-t_{w,i}/T_{1,i})$, $T_{1,i}$ is the relaxation time of qubit $i$ and $F_{\text{SCT},i}$ is defined in equation (13). Equation (63) can then be used instead of equation (13) for calculation of $F_M$. Note that $t_{\text{OUT}}$ obtained using equation (15) will still give the optimal values for equation (63). We now want to find the optimal ordering of the qubits to achieve the highest fidelity readout across all those being measured. We will demonstrate this with an example. We want to measure three different qubits, $\{Q_1, Q_2, Q_3\}$ sequentially with the following values for $t_m$ and $T_1$:

|       | $Q_1$ | $Q_2$ | $Q_3$ |
|-------|-------|-------|-------|
| $t_m$ | 3     | 1     | 2     |
| $T_1$ | 5     | 2     | 10    |

Since we want to maximise the fidelity across all qubits, we are only interested in the multiplicative factor in equation (63). Therefore, we need to calculate $\Lambda_i$ for every order of measurement, $M = \{M_1, M_2, M_3\}$. This means there are $N!$ measurement combinations we must consider where $N$ is the total number of qubits to be
read out. Whichever qubit is measured first, by definition, has $\Lambda_1 = 1$ since $t_{m,1} = 0$. To determine the best measurement order we calculate $\sum \Lambda_i / N$ which corresponds to the average reduction in $F_{STC}$ across all the qubits. As $\sum \Lambda_i / N \rightarrow 1$ the higher the overall fidelity will be obtained using the given measurement order. From the calculations in the table below we can see that the measurement $\{M_1, M_2, M_3\} = \{Q_2, Q_1, Q_3\}$ will be optimal for mitigating the effect of sequential readout on the individual qubits since this combination has the largest $\sum \Lambda_i / N$.

| $M_1$ | $M_2$ | $M_3$ | $\Lambda_1$ | $\Lambda_2$ | $\Lambda_3$ | $\sum \Lambda_i / N$ |
|-------|-------|-------|-------------|-------------|-------------|------------------|
| $Q_1$ | $Q_1$ | $Q_1$ | 1.0000      | 0.2231      | 0.6703      | 0.6311           |
| $Q_1$ | $Q_1$ | $Q_2$ | 1.0000      | 0.7408      | 0.0821      | 0.6076           |
| $Q_2$ | $Q_2$ | $Q_1$ | 1.0000      | 0.8187      | 0.6703      | 0.8297           |
| $Q_2$ | $Q_2$ | $Q_2$ | 1.0000      | 0.9048      | 0.5488      | 0.8179           |
| $Q_3$ | $Q_3$ | $Q_2$ | 1.0000      | 0.6703      | 0.0821      | 0.5841           |
| $Q_3$ | $Q_3$ | $Q_2$ | 1.0000      | 0.3679      | 0.5488      | 0.6389           |

We can immediately see here that the average $F_{STC}^3$ across all qubits is reduced by a factor of $\sum \Lambda_i / N = 0.8297$ compared to measuring each qubit simultaneously, that is, without waiting between measurements. The method outlined above can also be used to optimise the measurement order for the readout of specific qubits.

5. Summary

Single-shot electron spin readout fidelity calculations will become increasingly important as experiments push towards the fault-tolerant threshold for two-dimensional surface codes [35–37]. The improvement of gate fidelities will place increasingly more emphasis on state preparation and measurement errors as these become the limiting source of infidelity. The current state of analysis for measurement fidelity varies considerably [16–18] and in this current paper we propose a standard approach, which we have used to make a comparison between previous experimental results.

We have presented a method to calculate the single-shot readout fidelity of a detector based on an a comprehensive statistical analysis of the system. We first provided a simple formula to calculate the sample rate and readout time required to achieve high fidelity readout, where we emphasise the importance of choosing a sufficiently fast sample rate. Using our model, we describe different fidelity limiting factors, how to identify them from the model and strategies to increase the fidelity once they have been identified. To illustrate this we use the results from a 2P donor dot system that had already achieved high fidelity electron spin readout (Broome(L)) [32] to show that the fidelity can be further increased by 0.9% since the previous measurements had been limited by the sample rate and filter frequency. Assuming white Gaussian noise and provided the key experimental parameters are met: charge sensor sample rate $\Gamma_s \gtrsim 12/t_{m,S}$, $D' \gtrsim 3$, qubit energy splitting $\Delta E \gtrsim 13k_B T$ and a long relaxation time, $T_1 \gtrsim 100\mu$s then fidelities greater than 99% can be achieved.

Acknowledgments

The research was conducted by the Australian Research Council Centre of Excellence for Quantum Computation and Communication Technology (project number CE170100012) and Silicon Quantum Computing Pty Ltd. MYS acknowledges an Australian Research Council Laureate Fellowship.

ORCID iDs

M Y Simmons © https://orcid.org/0000-0002-6422-5888

References

[1] DiVincenzo D P 2000 Fortschr. Phys. 48 771
[2] Wang D S, Fowler A G and Hollenberg L C J 2011 Phys. Rev. A 83 020302
[3] Fowler A G, Mariantoni M, Martinis J M and Cleland A N 2012 Phys. Rev. A 86 032324
[4] Hill C D, Peretz E, Hile S J, House M G, Fuechsle M, Rogge S, Simmons M Y and Hollenberg L C J 2015 Sci. Adv. 1 9
[5] O’Gorman J, Nickerson N H, Ross P, Morton J J L and Benjamin S C 2016 NPJ Quantum Inf. 2 15019
[6] Knill E, Leibfried D, Reichle R, Britton J, Blakestad R B, Jost J D, Langer C, Ozeri R, Seidelin S and Wineland D J 2008 Phys. Rev. A 77 012307
[7] Magean E, Gambetta J M and Emerson J 2011 Phys. Rev. Lett. 106 180504
[8] Barends R et al 2014 Nature 508 500
[9] Wallman J and Flammia S T 2014 New J. Phys. 16 103032
[10] Cross A W, Magesan E, Bishop L S, Smolin J A and Gambetta J M 2016 NPJ Quantum Inf. 2 16012
[11] Harty T P, Allcock D T C, Ballance C J, Guidoni L, Janacek H A, Linke N M, Stacey D N and Lucas D M 2014 Phys. Rev. Lett. 113 220501
[12] Loss D and DiVincenzo D P 1998 Phys. Rev. A 57 120
[13] Kane B E 1998 Nature 393 133
[14] Tosi G, Mohiyaddin F A, Schmitt V, Tenberg S, Rahman R, Klimeck G and Morello A 2017 Nat. Commun. 8 450
[15] Veldhorst M, Eerink H G J, Yang C H and Drzaek A S 2017 Nat. Commun. 8 1766
[16] Elzerman J M, Hanson R, Willems van Beveren L H, Witkamp B, Vandersypen L M K and Kouwenhoven L P 2004 Nature 430 431–5
[17] Morello A et al 2010 Nature 467 687
[18] Büch H, Mahapatra S, Rahman R, Morello A and Simmons M Y 2013 Nat. Commun. 4 2011 Phys. Rev. Lett. 106 156804
[19] Veldhorst M et al 2014 Nat. Nanotechnol. 9 981
[20] Robledo L, Childress L, Bernien H, Hensen B, Alkemade P F A and Hansen R 2011 Nature 477 574
[21] Watson T F, Weber B, House M G, Bäch H and Simmons M Y 2015 Phys. Rev. Lett. 115 166806
[22] D’Anjou B and Coish W A 2014 Phys. Rev. A 89 023131
[23] Pla J, Tan K Y, Dehollian I P, Lim W H, Jamieson D N, Drzaek A S and Morello A 2012 Nature 489 541
[24] Watson T F, Weber B, Hsueh Y-L, Hollenberg L C L, Rahman R and Simmons M Y 2017 Sci. Adv. 3 3
[25] Prance J, Bael B J V, Simmons C B, Savage D E, Lagally M G, Friesen M, Coppersmith S N and Eriksson M A 2015 Nanotechnology 26 215201
[26] Gorman S K, He Y, House M G, Keizer J G, Keith D, Fricke L, Hile S J, Broome M A and Simmons M Y 2017 Phys. Rev. Lett. 115 146806
[27] Prance J, Bael B J V, Simmons C B, Savage D E, Lagally M G, Friesen M, Coppersmith S N and Eriksson M A 2015 Nanotechnology 26 215201
[28] House M G, Xiao M, Guo G, Li H, Cao G, Rosenthal M M and Jiang H 2013 Phys. Rev. Lett. 111 126803
[29] Gambetta J, Braff W A, Wallraff A, Girvin S M and Schoelkopf R J 2007 Phys. Rev. A 76 012325
[30] Ho Eom B, Day P K, Le Duc H G and Zmuidzinas J 2012 Nat. Phys. 8 623
[31] Tracy L A, Luhman D R, Carr S M, Bishop L C, Eyck G A T, Playm T, Wendl J R, Lilly M P and Carroll M S 2016 Appl. Phys. Lett. 108 063101
[32] Broome M A et al 2018 Nat. Commun. 9 980
[33] Shannon C E 1998 Proc. IEEE 86 447
[34] Nowack K C, Shafei M, Laforest M, Prawiroatmodjo G E D K, Schreiber I R, Reichl C, Wegscheider W and Vandersypen L M K 2011 Science 333 1269
[35] Brunner R, Shin Y-S, Obata T, Pioro-Ladrière M, Kubo T, Yoshida K, Taniyama T, Tokura Y and Tarucha S 2011 Phys. Rev. Lett. 107 146801
[36] Shulman M D, Dial O E, Harvey S P, Bluhm H, Umansky V and Yacoby A 2012 Science 336 202
[37] Veldhorst M et al 2015 Nature 526 410