Supplemental Material for
"Optimal control for quantum detectors"

In this supplemental material we provide additional details for the main text. In Supplementary Note I, we review the results from quantum state discrimination, and the average error rates for single-shot and multiple shot state discrimination. In Supplementary Note II we plot $F_t(\omega)$ defined in Eq. (3) in the main text for the controls considered in Fig. 4 in the main text. In Supplementary Note III, we provide figures showing the dependence of the optimal control with the correlation time and experimental limitations.

Supplementary Note I. QUANTUM STATE DISCRIMINATION

In this section, we connect well-known results in optimal quantum state discrimination to the objective function necessary for the optimal control for detection in the main text.

Consider the task of discriminating two density matrices $\hat{\rho}_\eta(t_\mathrm{s})$ and $\hat{\rho}_{\eta+\mathrm{s}}(t_\mathrm{s})$ obtained in the presence and absence of the signal respectively that where the signal and noise both cause qubit dephasing. These density matrices are given by

$$\rho_\eta = \frac{1}{2}\begin{pmatrix} 1 & e^{-\chi_\eta(t)} \\ e^{-\chi_\eta(t)} & 1 \end{pmatrix}, \quad \text{and} \quad \rho_{\eta+\mathrm{s}} = \frac{1}{2}\begin{pmatrix} 1 & e^{-\chi_{\eta+\mathrm{s}}(t)} \\ e^{-\chi_{\eta+\mathrm{s}}(t)} & 1 \end{pmatrix} \tag{1}$$

Results from the early works of Helstrom state that two quantum states denoted as $\hat{\rho}_1$ and $\hat{\rho}_2$ with prior ($p = 1/2$) is optimally discriminated by measuring in the basis $\hat{\rho}_1 - \hat{\rho}_2$, with the positive eigenspace corresponding to $\hat{\rho}_1$ [1, 2]. Clearly, these two quantum states in Eq. 1 are discriminated optimally by the $\hat{E}_\pm = |\pm\rangle\langle\pm|$ operator. This is precisely what is measured in detection protocol [in step (iii)] defined in the main text (by applying a Hadamard rotation and measurement). The measurement outcome is the result of a Bernoulli trial with probability $P_{|0\rangle,\eta}(t_\mathrm{s}) = \frac{1}{2}(1 + e^{-\chi_\eta(t_\mathrm{s})})$ or $P_{|0\rangle,\eta+\mathrm{s}}(t_\mathrm{s}) = \frac{1}{2}(1 + e^{-\chi_{\eta+\mathrm{s}}(t_\mathrm{s})})$ depending on whether the signal is absent or present respectively. Notice that $P_{|0\rangle,\eta+\mathrm{s}}(t_\mathrm{s}) < P_{|0\rangle,\eta}(t_\mathrm{s})$. In the following we discuss the optimal detection protocol that minimizes the error rate for single shot and multiple shot detection schemes.

Average Error rate in discrimination with a single shot

Let us now calculate the average error rate in discriminating the two quantum states $\hat{\rho}_\eta$ and $\hat{\rho}_{\eta+\mathrm{s}}$ from the measurement outcome of one shot of the detection protocol shown in Fig. 1. The threshold for single shot detection is defined as follows: if the outcome of the measurement is $|0\rangle$, then the signal is not detected, and conversely when the outcome is $|1\rangle$, then the signal is detected. Assuming no prior information (i.e equal prior probability of $\hat{\rho}_\eta$ and $\hat{\rho}_{\eta+\mathrm{s}}$) the average probability of error (False Positive and False Negative) in the detection is,

$$P_\mathrm{err} = \frac{1}{2}\left[\underbrace{P(|0\rangle\,|\rho_{\eta+\mathrm{s}})}_{\mathrm{FN}} + \underbrace{P(|1\rangle\,|\rho_\eta)}_{\mathrm{FP}}\right] \tag{2}$$

$$= \frac{1}{2}\left(P_{|0\rangle,\eta+\mathrm{s}} + 1 - P_{|0\rangle,\eta}\right) \tag{3}$$

$$\Rightarrow P_\mathrm{err} = \frac{1}{2} - \frac{1}{2}\left(P_{|0\rangle,\eta} - P_{|0\rangle,\eta+\mathrm{s}}\right) = \frac{1}{2} - \frac{1}{2}\mathcal{O} \tag{4}$$

where, in the first step, $P(|0\rangle\,|\rho_{\eta+\mathrm{s}})$ denotes the probability of measuring a '0' given that the signal is present and similarly for $P(|1\rangle\,|\rho_\eta)$ and $\mathcal{O}$ is defined in Eq. (5) in the main text. Therefore, the probability of mis-detection is minimized by maximising the difference $\mathcal{O} = \frac{1}{2}(e^{-\chi_\eta(t)} - e^{-\chi_{\eta+\mathrm{s}}(t)})$, which is what is used as the criteria for obtaining the optimal control protocols in the main text.

Error in discrimination with $N_\mathrm{s} > 1$ shots

For experiments where multiple shots (N_s) are available, the aim is to identify protocols that minimize the average error rate (average probability of FP and FN) for the discrimination problem. According to the Neymann-Pearson lemma, the statistic that minimizes the error rate is the likelihood ratio [3]. Given the the measurement outcomes correspond to N_s independent

Bernoulli trials, the likelihood ratio is a function of N_0: the number of measurement outcomes that are $|0\rangle$. Therefore, N_0 is the optimal decision statistic that minimizes the error rate. The result of the detection protocol is determined by the threshold, $N_{\rm th}$, with $N_0 < N_{\rm th}$ and $N_0 \geq N_{\rm th}$ corresponding to signal present or absent respectively.

Given that N_0 is drawn from a binomial distribution given by $B(N_0, N_{\rm s}, P_{|0\rangle,\eta+{\rm s}})$ or $B(N_0, N_{\rm s}, P_{|0\rangle,\eta})$ depending on whether the signal is present or absent respectively , the average error rate is defined as,

$$P_{\rm err} = \frac{1}{2}\left[\Pr(N_0 \geq N_{\rm th}|\hat{\rho}_{\eta+{\rm s}}) + \Pr(N_0 < N_{\rm th}|\hat{\rho}_\eta)\right] \tag{5}$$

$$\Pr(N_0 \geq N_{\rm th}|\hat{\rho}_{\eta+{\rm s}}) = \sum_{N_0, N_0 \geq N_{\rm th}} B(N_0, N_{\rm s}, P_{|0\rangle,\eta+{\rm s}}) \tag{6}$$

$$\Pr(N_0 < N_{\rm th}|\hat{\rho}_\eta) = \sum_{N_0, N_0 < N_{\rm th}} B(N_0, N_{\rm s}, P_{|0\rangle,\eta}) \tag{7}$$

where $B(N_0, N_{\rm s}, P) = \binom{N_{\rm s}}{N_0} P^{N_0}(1-P)^{(N_{\rm s}-N_0)}$ is the Binomial distribution and $N_{\rm th}$ is the threshold. The threshold is chosen to minimize the average error, which corresponds to the value of N_0 where the two binomial distributions cross,

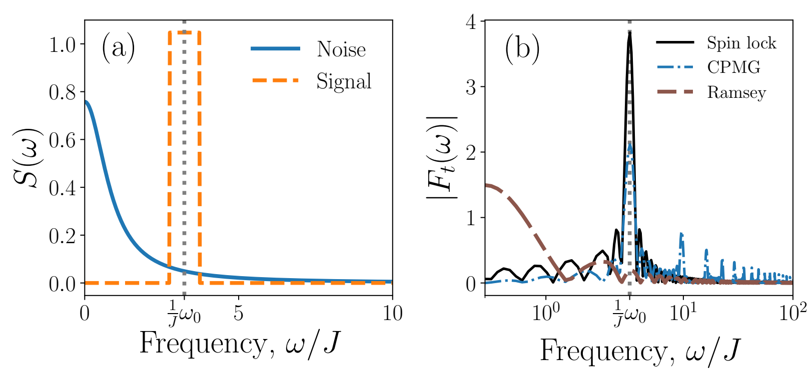
$$N_{\rm th} = \underset{\tilde{N}}{\rm argmin}\left|\sum_{N_0 \geq \tilde{N}} B(N_0, N_{\rm s}, P_{|0\rangle,\eta+{\rm s}}) + \sum_{N_0 < \tilde{N}} B(N_0, N_{\rm s}, P_{|0\rangle,\eta})\right|. \tag{8}$$

Note that the expression for single shot average error rate in Eq. (5) reduces to the single shot case in Eq. (4) in the limit, $N_0 = 1$ and $N_{\rm s} = 1$ and $N_{\rm th} = 1$.

Therefore, for the case of multiple shots, the average error rate is dependent on the total number of shots $N_{\rm s}$ as well as the outcome probabilities, $P_{|0\rangle,\eta}$ and $P_{|0\rangle,\eta+{\rm s}}$. Unlike the single-shot discrimination case, there is not a simple expression for the objective function that is differentiable, and therefore, we choose the single-shot average error as the objective function for optimizing the control protocols. Note that, in the limit of large number of shots (in which case the error rate is small), a differentiable objective function may be obtained using a Gaussian approximation for the binomial distribution; but we leave this for future investigations. The accuracy of the normal approximation also depends on the probabilities $P_{|0\rangle,\eta}$, $P_{|0\rangle,\eta+{\rm s}}$.

Supplementary Note II. FILTER FUNCTION IN THE SECOND CUMULANT APPROXIMATION

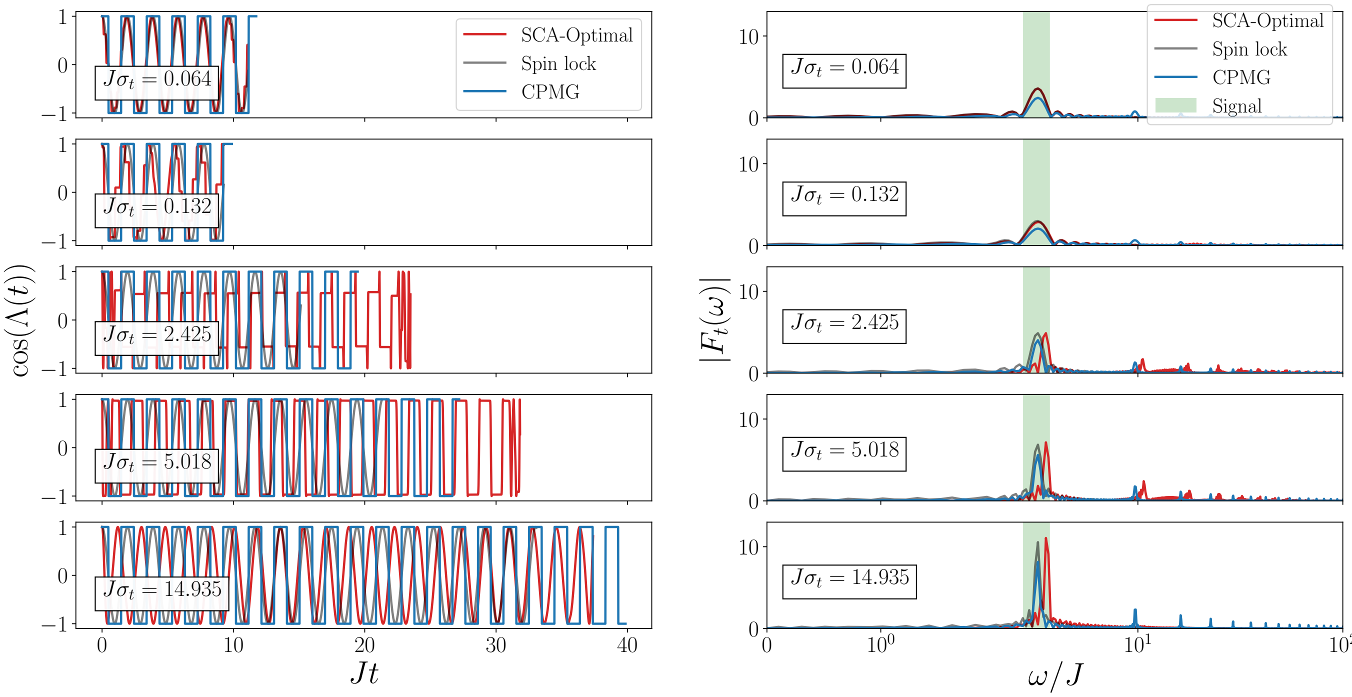
We plot the filter functions of the controls considered in Fig. 4 (a).



Supplementary Figure 1. Additional details of the signal for the detection protocols implemented using parameters in Fig. 4. (a) Noise and Signal spectrum corresponding to the parameters in Fig. 4 in the main text. (b) Filter Functions corresponding to different control schemes (Spin-Lock, CPMG and Ramsey) ; the time-domain profile of these controls are shown in Fig. 4 (a).

Supplementary Note III. DETAILS FOR NUMERICAL OPTIMIZATION

The optimal protocol compared with Spin-lock and CPMG for different values of σ_t, together with their corresponding filter functions, are shown in Supplementary Figures 2 and 3 for different signals. In Supplementary Figure 2, we show the optimal control protocols for single shot detection for the signal spectrum considered in the main text. The optimal control crosses over from spin-locking to CPMG before crossing back over to spin-locking for higher correlation times. Interestingly, the optimal control protocol has shapes that are neither CPMG or Spin-locking. We also test the optimization method for narrower signal bandwidth (with same SNR); the optimal controls are shown in Supplementary Figure 3. In this case, the we see a similar crossover from spin-locking to CPMG, before crossing over to control shapes that are neither of the two at higher correlation times.



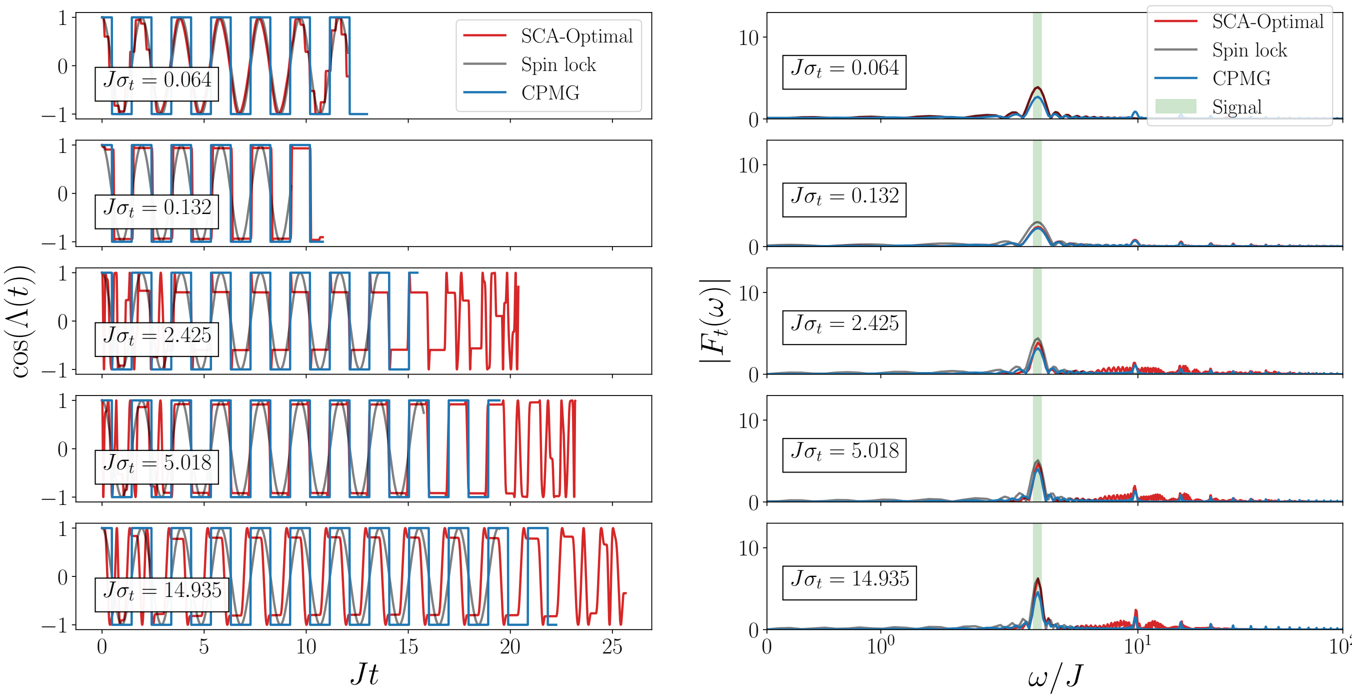
Supplementary Figure 2. Comparison of the numerically optimized protocol (SCA-optimal) with spin lock and CPMG protocols for different values of noise correlation σ_t for a white-cutoff signal centered at $\omega_0/J = 3.24$ with a width of $2\delta_0/J = 0.97$. The left panel shows the time integrated control protocol $\cos(\Lambda(t))$, and the right panel shows the associated filter functions $|F_t(\omega)|$. Note that for small values of $J\sigma_t$, noise spectrum is nearly white and the optimal control coincides with spin-lock protocol. As $J\sigma_t$ increases the optimal control gets closer to CMPG protocol, and later assumes solutions that differ from both spin-lock and CPMG protocols.

To take experimental limitations into account, we also consider a scenario where the maximum power is limited. This is manifested in a constraint on the maximum value of Ω at any time. We can implement this constraint by either adding an $L2$ regularization term to the objective function, that is $\mathcal{O}_{L2} = \beta||\mathbf{\Omega}||_2^2$, where β is a hyper parameter that controls the magnitude, or imposing a hard cutoff Ω_c on the value of the optimization variable. Note that close to an optimal point, where the cost function is approximately quadratic, the $L2$ regularization is equivalent to imposing an upper bound on the root mean square (rms) of the control defined as

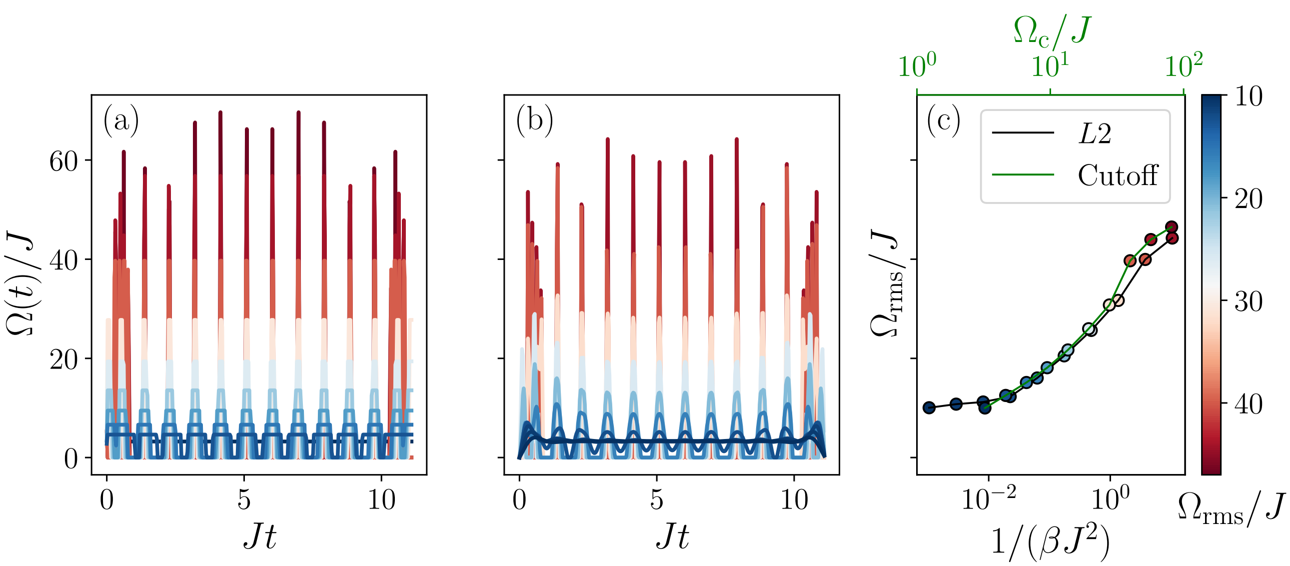
$$\Omega_{\mathrm{rms}} = \sqrt{\frac{1}{t_{\mathrm{s}}}\int_0^{t_{\mathrm{s}}} |\Omega(t)|^2 dt}. \tag{9}$$

We observe that in a regime where the optimizer finds a CPMG-like control scheme to be optimal, adding these constraint transforms instantaneous pulses to finite-width version of the pulse that ultimately reaches the spin-lock solution, see Supplementary Figure 4.

[1] C. W. Helstrom, Information and Control **10**, 254 (1967).
[2] C. W. Helstrom, Journal of Statistical Physics **1**, 231 (1969).
[3] J. Neyman and E. S. Pearson, Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character **231**, 289 (1933).



Supplementary Figure 3. Comparison of the numerically optimized protocol (SCA-optimal) with spin lock and CPMG protocols for different values of noise correlation σ_t for a white-cutoff signal centered at $\omega_0/J = 3.24$ with a width of $2\delta_0/J = 0.49$. The left panel shows the time integrated control protocol $\cos(\Lambda(t))$, and the right panel shows the associated filter functions $|F_t(\omega)|$. Note that for small values of $J\sigma_t$, noise spectrum is nearly white and the optimal control coincides with spin-lock protocol. As $J\sigma_t$ increases the optimal control gets closer to CMPG protocol, and later assumes solutions that differ from both spin-lock and CPMG protocols.



Supplementary Figure 4. The effect of constraints on the optimal protocol for $J\sigma_t = 0.56$, $\omega_0/J = 2.59$, and $2\delta_0/J = 0.97$. The plots are color coded by their root mean square (rms) value $\Omega_{\rm rms}$. (a) Controls $\Omega(t)$ obtained where a hard cutoff Ω_c is imposed on the amplitude of the drive. As the cutoff value is decreased the instantaneous pulses are transformed to finite width pulses with lower rms.(b) With $L2$ regularization, we observe that as β is increased, the pulse solution $\Omega(t)$ gets smoother and pulses are widened with a lower rms. (c) The correspondence between $\Omega_{\rm rms}$ and the cutoff frequency Ω_c (green) and the hyperparameter β (black). We see that increasing (decreasing) the value of $\beta(\Omega_c)$ results in a solution with a lower $\Omega_{\rm rms}$.