PAPER • OPEN ACCESS

Bayesian quantum noise spectroscopy

To cite this article: Chris Ferrie *et al* 2018 *New J. Phys.* **20** 123005

Recent citations

- Magnetic-Field Learning Using a Single Electronic Spin in Diamond with One-Photon Readout at Room Temperature
R. Santagati *et al*

View the article online for updates and enhancements.

New Journal of Physics

The open access journal at the forefront of physics

CrossMark

PAPER

Bayesian quantum noise spectroscopy

Chris Ferrie[1,6]**, Chris Granade**[2,3]**, Gerardo Paz-Silva**[4,5] **and Howard M. Wiseman**[4,5]

[1] University of Technology Sydney, Centre for Quantum Software and Information, Ultimo NSW 2007, Australia
[2] Centre for Engineered Quantum Systems, University of Sydney, Sydney, NSW, Australia
[3] School of Physics, University of Sydney, Sydney, NSW, Australia
[4] Centre for Quantum Dynamics, Griffith University, Brisbane, Queensland 4111, Australia
[5] Centre for Quantum Computation and Communication Technology, Griffith University, Brisbane, Queensland 4111, Australia
[6] Author to whom any correspondence should be addressed.

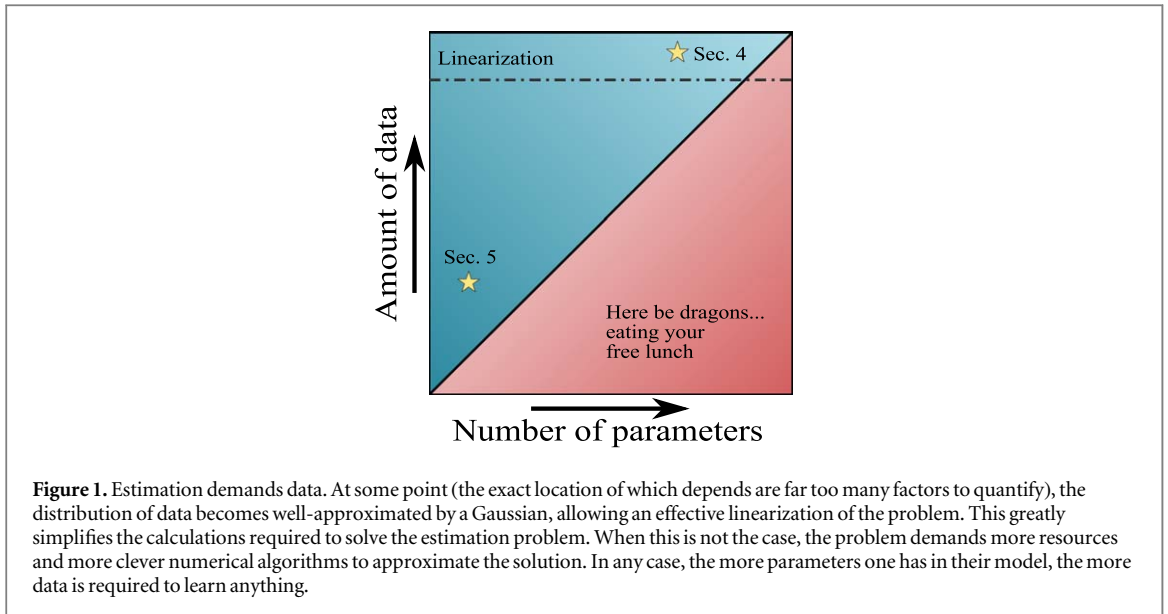**E-mail:** christopher.ferrie@uts.edu.au

Abstract

As commonly understood, the noise spectroscopy problem—characterizing the statistical properties of a noise process affecting a quantum system by measuring its response—is mathematically ill-posed, in the sense that no unique noise process leads to a set of responses unless extra assumptions are taken into account. Ad-hoc solutions assume an implicit structure, which is often never determined. Thus, it is unclear when the method will succeed or whether one should trust the solution obtained. Here, we propose to treat the problem from the point of view of statistical estimation theory. We develop a Bayesian solution to the problem which allows one to easily incorporate assumptions which render the problem solvable. We compare several numerical techniques for noise spectroscopy and find the Bayesian approach to be superior in many respects.

1. Introduction

Quantum technologies will very likely ultimately rely on active error correction. However, at every stage—crucially in current experiments—open-loop control techniques to *suppress* errors need to be employed [1]. They can be thought of as a layer-0 level of protection designed to make the errors in any operation as small as possible, before the machinery of quantum error correction and fault-tolerant quantum computing takes over.

These techniques can be roughly classified in terms of their robustness to uncertainty in the knowledge of the noise they attempt to suppress. Exemplifying this, dynamical decoupling (DD) [2] sits on one end of the spectrum as a highly robust technique—DD sequences only require the noise to be 'slow' (in some appropriate metric we will discuss in more detail later), but beyond this the details of the noise are not important. On the other hand, optimal control techniques can be used to design pulse sequences capable of efficiently suppressing a wide range of noises, both 'fast' and 'slow', but only if detailed knowledge of the noise is available [3, 4]. The ideal strategy is thus a function of the knowledge available about the noise. Unfortunately this detailed information is often absent, since in an open quantum system scenario the bath or environment generating such noise cannot be directly measured or controlled. Phenomenological models and intimate control over the fabrication process of a given quantum system can alleviate this, but can at most give partial information about the noise sources affecting a qubit.

In order to bridge this gap in knowledge, *quantum noise spectroscopy* protocols of varying generality have been developed and implemented in recent years [5–13]. Their objective is to characterize the actual noise affecting a quantum system of interest, regardless of its source, in terms of its correlations, or more specifically the set of power polyspectra [14]. The key point is that the information these protocols output, together with optimal control techniques, should enable one to design control routines tailored to suppress the actual noise affecting the quantum system of interest [15]. Very recently, for example, a 10 min record-breaking coherence time was achieved in trapped-ions [16] using this principle. Operationally, spectroscopy protocols measure the response of a quantum system to the noise affecting it. This is achieved by exploiting the ability to prepare a set of

Figure 1. Estimation demands data. At some point (the exact location of which depends are far too many factors to quantify), the distribution of data becomes well-approximated by a Gaussian, allowing an effective linearization of the problem. This greatly simplifies the calculations required to solve the estimation problem. When this is not the case, the problem demands more resources and more clever numerical algorithms to approximate the solution. In any case, the more parameters one has in their model, the more data is required to learn anything.

initial states, to apply suitable control sequences, and to measure a set of convenient observables. The main difficulty is that noise correlations influence the dynamics of the quantum system in a highly nonlinear way. Thus, inferring these correlations in detail from the response of the quantum system is generally an ill-posed problem, unless *a priori* information on the noise is assumed. Even when standard assumptions, such as Gaussian noise or a dephasing coupling are satisfied, the problem remains nonlinear and inverting it carries along a set of non-trivial complications that in turn constrain the type of noise that can be characterized. For example, in [8–12] a control-induced frequency comb approach is used in order to overcome the nonlinear character of the problem but it comes at the cost of being only effective when the noise correlations are *smooth* functions in frequency space.

We propose that many of these problems can be ameliorated, or at least properly quantified, using a statistically principled approach. Within the statistical phrasing of the problem we provide a Bayesian solution [17], complete with a numerical implementation. We show the problem can be solved analytically in the limit of large of amounts of experimental data. At the other extreme—the small-data limit—a numerically stable Monte Carlo algorithm [18] approximates the full Bayesian solution. Our two approaches provide a robust solution to the software side of the noise spectroscopy problem, and can be used to improve state-of-the-art spectroscopy protocols [8–12]. These two regimes are schematically depicted in figure 1. Though the physical model we consider is simplified to illustrate the novel aspects of this work, we note that our approach is very versatile. We discuss this later.

We summarize the performance of our approach for the large- and small-data limits in section 6, with complete details including all source code in the supplementary material. In the large data limit, our approach can give almost an order of magnitude improvement in performance over state-of-the-art estimation strategies. By contrast, in the small-data regime, we can even achieve two orders of magnitude improvement with just 2500 bits of data. Thus, our approaches yield immediate and dramatic benefits in terms of experimental costs.

An expert reader may have realized that our work seems related to recent work of Zwick, Alvarez and Kurizki (ZAK) [19]. It is useful to highlight the main differences between that work and ours. There the authors discuss the problem of designing experiments to maximize the expected information of select spectral properties ('bath parameters'). The main difference is that ZAK treat the single parameter case, and are thus able to obtain clever analytic solution to the experiment design problem. In contrast, our approach is inherently multidimensional and thus more general, but this means we only provide numerical algorithms. These works can thus be seen as complementary[7]. Recent work has utilized Bayesian parameter estimation in trap ion noise spectroscopy experiments [20, 21] as well as optomechanical systems [22].

Our paper is organized as follows. In section 2 we motivate the problem from a physical perspective. In section 3 we extract the core mathematics, simplifying as much as possible the physical equations in order to phrase the problem as one of statistical estimation theory. We provide two solutions (as discussed in figure 1) at varying degrees of complexity and economy. In section 4 we provide the Gaussian process model which is valid in the large data limit. In section 5 the full Bayesian solution is outlined and applied to the ubiquitous $1/f$ noise

---

[7] Moreover, though ZAK suggest a Bayesian approach to the problem, they do not detail what numerical Bayesian updating algorithm is used. Our software is capable of realizing the protocol of ZAK for their specific problem and its multidimensional generalization.

model. A summary of the findings from our numerical experiments is given in section 6. We wrap up in section 7 with a brief discussion. Full implementation details and code to reproduce the results are listed in the ancillary files for this paper. In particular, this paper can be seen as a *living document* as follows: when the source is downloaded and run, new random variables will be drawn and the data will change. Therefore, the figures will also change, while—hopefully!—the conclusions drawn from them will not.

2. Physical motivation

In order to present our results, it is useful to introduce a concrete physical model that is simple enough that the open quantum systems and control language does not distract from the main features of our statistical approach, but that is sufficiently non-trivial to be a relevant model from the physical point of view. This is in line with our core intention in this paper, to introduce specialized tools from statistical analysis into the very relevant problem of noise characterization.

Our toy physical model is the dynamics of a two-level system (TLS) in the presence of dephasing noise generated by a zero-mean, Gaussian, stationary process. The TLS and its environment are assumed to be initialized in a factorizable state of the form $\rho_{SB} = \rho_S \otimes \rho_B$, with ρ_B being the initial state of the bath. Furthermore, their dynamics is ruled by a Hamiltonian that, in the interaction frame with respect to the natural dynamics of the bath, takes the form

$$H(t) = \sigma_z \otimes B(t) + H_{\mathrm{ctrl}}(t), \tag{1}$$

where $B(t)$ represents the bath noise, and $H_{\mathrm{ctrl}}(t)$ is a control Hamiltonian acting solely on the TLS. We will describe the bath as if it is quantum but a classical bath is of course just a special case where the bath operator that appears in our equations commutes at different times, i.e. $[B(t_1), B(t_2)] = 0$. For simplicity, we shall assume $H_{\mathrm{ctrl}}(t)$ to enact instantaneous σ_x pulses at times $\{t_i\}$, i.e. $H_{\mathrm{ctrl}}(t) = \frac{\pi}{2}\sum\delta(t - t_i)\sigma_x$, such that, in the so-called toggling frame with respect to the control, the Hamiltonian takes the even simpler form

$$H(t) = y(t)\sigma_z \otimes B(t),$$

with $y(t)$ a binary function taking values in $\{-1, 1\}$ and switching at times $\{t_i\}$. We will be interested in the expectation value of a Pauli operator σ_α at a time T, given by $\langle\sigma_\alpha(T)\rangle = \mathrm{Tr}[U(T)\rho_S \otimes \rho_B U(T)^\dagger(\sigma_\alpha \otimes \mathbf{1})]$, with the unitary evolution given by the appropriate time ordered exponential $U(T) = \mathcal{T}_+(\mathrm{e}^{-\mathrm{i}\int_0^T H(s)\mathrm{d}s})$. It is important to highlight that when the bath $B(t)$ has a component $\beta(t)$ that is a classical stochastic process, one can only access the average expectation value of σ_α over many realizations of $\beta(t)$. That is one can measure $\langle\!\langle\sigma_\alpha(T)\rangle\!\rangle_c$ with $\langle\cdot\rangle_c$ denoting average over realizations of $\beta(t)$. To ease the notation we will denote simply by $\langle\!\langle\cdot\rangle\!\rangle$ when both averages are taken. Under these conditions, the expectation value of an operator σ_α is given by [12]

$$\langle\!\langle\sigma_\alpha(T)\rangle\!\rangle = \mathrm{Tr}_S\left[\mathrm{Tr}_B\left[\left\langle\mathcal{T}_+\left(\mathrm{e}^{-\mathrm{i}\int_{-T}^T \mathrm{d}t\tilde{H}(t)}\right)\right\rangle_c \rho_B\right]\rho_S\sigma_\alpha\right], \tag{2}$$

with

$$\tilde{H}(t) = \begin{cases} -\sigma_\alpha H(T - t)\sigma_\alpha & \text{for} \quad 0 \leqslant t \leqslant T \\ H(T + t) & \text{for} -T \leqslant t < 0 \end{cases}.$$

The assumption that $B(t)$ is a Gaussian, zero-mean, stationary process, implies that only the second cumulant of the process is non-vanishing, i.e. $C^{(2)}(B(t_1)B(t_2)) = \mathrm{Tr}[\langle B(t_1)B(t_2)\rangle_c \rho_B] = \mathrm{Tr}[\langle B(t + t_1 - t_2)B(t)\rangle_c \rho_B]\ \forall\ t$. In turn this leads to the exact solution, in both time and frequency domains, being given by [12, 14]

$$\langle\!\langle\sigma_\alpha(T)\rangle\!\rangle = \mathrm{e}^{-\frac{1}{2}\left(\frac{1-f_{z,\alpha}}{2}\right)^2 \int_0^T \mathrm{d}t_1 \int_0^{t_1} \mathrm{d}t_2\, y(t_1)y(t_2)C^{(2)}(\{B(t_1),B(t_2)\})}\langle\sigma_\alpha(0)\rangle,$$

$$= \mathrm{e}^{-\frac{1}{2}\left(\frac{1-f_{z,\alpha}}{2}\right)^2 \int_0^\infty \frac{\mathrm{d}\omega}{2\pi}F(\omega,T)S(\omega)}\langle\sigma_\alpha(0)\rangle. \tag{3}$$

Here $f_{\alpha,z} = \mathrm{tr}[\sigma_\alpha\sigma_z\sigma_\alpha\sigma_z]/2$ takes values in $\{-1, 1\}$, $F(\omega, T) = |F^{(1)}(\omega, T)|^2$ is the filter function, and $S(\omega) = \int_{-\infty}^\infty \mathrm{d}t C^{(2)}(\frac{\{B(t),B(0)\}}{2})\mathrm{e}^{-\mathrm{i}\omega t}$, where $\{\cdot, \cdot\}$ denotes the anticommutator, is the symmetric or classical power spectrum of the noise process. Additionally, $F^{(1)}(\omega, T) = \int_0^T \mathrm{d}t\, y(t)\,\mathrm{e}^{\mathrm{i}\omega t}$ is the so-called order-one fundamental filter function which depends purely on the control [23, 24]. These equations capture the dephasing behaviour of a qubit in the presence of a classical—as in the various semiclassical approximations often made in NV centers [25] or NMR [10]—noise, or quantum noise—as in the case of a bosonic bath in a thermal state [14, 26].

In this language, suppressing the decoherence is akin to minimizing the value of the exponent on the right hand side of equation (3). This can be achieved via the use of control: one can choose a routine whose filter has a small overlap with the power spectrum. DD sequences, for example, achieve this by generating filters which

vanish at $\omega = 0$ and are flat around it. They are thus very effective against decoherence generated by 'slow' noise, where $S(\omega)$ is mostly supported around $\omega = 0$. For noise with considerable high frequency contributions, the DD sequence can be ineffective or can, in analogy to the anti-Zeno effect [27], even generate the opposite effect, i.e. decoherence enhancement. On the other hand, if information about the power spectrum is available then optimal control techniques can be used to find a control routine that minimizes the overlap and thus the decoherence [23, 24]. This is clearly the ideal situation, but it raises the question if the necessary information can be obtained in an open quantum system scenario.

Complete knowledge of the Hamiltonian describing the decoherence process and of the initial state of the environment would grant us perfect knowledge of the power spectrum and, in such situations, optimal control methods could be used to minimize the decoherence of the TLS. However, even when Gaussianity is imposed *a priori* [14], such knowledge is rarely available. For example, the temperature of the thermal state of a bosonic environment or the dispersion relation for the bosonic modes is usually unknown. Fortunately, as seen from the equations above, the decoherence process only depends on $S(\omega)$, i.e. not on the actual form of $B(t)$ or even ρ_B but on the bath correlations they induce. This quantity, while not directly measurable can be inferred from the measurable (average) response of the TLS to different control Hamiltonians.

This is the working principle behind recently proposed *noise spectroscopy* protocols. Schematically, one such protocol would work as follows. Imagine preparing $a + 1$ eigenstate of σ_x at time $t = 0$, in such way that the expectation value of the observable σ_x at the final time $t = T$, given $H_{\text{ctrl}}(t)$, is as in equation (3). Different choices of $H_{\text{ctrl}}(t)$ result in different filters $F(\omega, T)$, and different experimentally accessible values of $\langle\!\langle \sigma_x(T) \rangle\!\rangle|_{H_{\text{ctrl}}(t)}$. In principle, it should be possible to choose a sufficiently large set of different control sequences in such way that the integral in the exponent can be deconvolved, and information about $S(\omega)$ can be inferred. Different approaches to this problem, under different simplifying assumptions, have been proposed and even experimentally implemented [10, 11, 13, 16, 28]. More exotic protocols have been proposed to characterize more general noise processes, such as non-Gaussian noise [11] or noise affecting multiple qubits [12, 29]. While we will not consider them here in detail, we note that the statistically motivated methods can in principle be also used there via an appropriate generalization. It should be highlighted that by measuring the response of the qubit to the presence of the bath and qubit control the best one can expect to estimate are the bath correlation functions alluded to earlier. From them, information about physical parameters of the bath can be extracted but only under concrete assumptions. For example, if one is given the prior that the bath consists of collection of harmonic oscillators in a thermal state and system and bath couple through linearly as in [12] then one can extract the temperature of the bath from $S(\omega)$.

In the remainder, we will abstract as much physical detail as possible for brevity and generality. This allows us to easily apply techniques from statistical decision and estimation theory.

3. Bayesian spectral estimation

In the physical description above, we make reference to observations as being the average values of observables. By contrast, in real experiments, observations are made by acquiring single bits of data at a time through projective measurements of single quantum systems. These two views of experimental observations agree only in the limit that very large numbers of projective measurements are made on identical copies of the system. Reasoning about noise spectroscopy in the presence of experimental constraints is thus, at its core, a statistical problem not suited to the 'data-fitting' paradigm we are more used to. That is, when statisical fluctuations are dominated by those force on us by quantum measurement uncertainties, fitting points to lines without regard for the statisical models they were generated from is not a robust approach.

To make this precise we first extract the core mathematical elements of the problem. Mathematically, we are interested in the exponent appearing in (3),

$$\chi(S; F_j) = \frac{1}{2\pi} \int_0^\Omega S(\omega) F_j(\omega) \, d\omega, \tag{4}$$

where Ω is a high frequency cut-off which is often imposed experimentally and required for the numerical integration we use. So as to not introduce too many complications, we assume that Ω is known.

Recall that we do not have direct access to χ as it is only exposed experimentally through the statistical model in (3). Moreover, expectation values of observables also cannot be measured directly and will always come with fluctuations due to finite sample sizes. Thus, we prefer to work from the bottom up, considering the precise distribution of each bit of data. To this end, let r be a the binary random variable with distribution

$$\Pr(r = 1 | S; F_j) = \frac{1}{2}(1 + e^{-\chi(S; F_j)}), \tag{5}$$

such that the expectation value in (3) obeys

$$\langle\langle \sigma_x \rangle\rangle = \Pr(r = 1|S; F_j) - \Pr(r = -1|S; F_j).$$

This is the most fundamental statistical model and we should process data at this level whenever possible. What exactly we mean by *process data* will be made explicit next.

The notation $\Pr(A|B)$ is read 'the probability of *A* being true given *B* is known to be true'. So, $\Pr(r = 1|S; F_j)$ is the probability of observing $r = 1$ *given* the filter F_j is used and the spectrum is *S*. However, the spectrum is the thing we are not *given*. To rectify this, we *invert* the probability using Bayes' rule:

$$\Pr(S|r; F_j) = \frac{\Pr(r|S; F_j)\Pr(S|F_j)}{\Pr(r|F_j)}. \tag{6}$$

Some terminology [30]: $\Pr(r|S; F_j)$ is called the *likelihood function* and in physics it is always given by the physical model; $\Pr(r|F_j)$ is called the *evidence* and it is usually ignored as it can be determined by normalization; $\Pr(S|F_j)$ is called the *prior* and encodes the information we have about the spectrum before the data is take; and finally, $\Pr(S|r; F_j)$ is called the *posterior*, which is the information we have about the spectrum *after* the experiment— exactly what we want to know!

In general, performing this inversion is both analytically and computationally intractable. There are two general approaches to solving this problem. Either we make analytical approximations or we employ clever numerical integration techniques. Here we demonstrate both. But, the problem and solutions are also not decoupled from how much can be assumed known about the spectrum—the *dimension* of model—and the amount of data available, such that the domain of applicability of each solution is restricted in subtle ways. This is shown pictorially in figure 1.

3.1. Big data: analytical approximations with weak assumptions

In the large data[8] limit, we can appeal to the central limit theorem. In the Gaussian limit of the likelihood function we effectively linearize the model. For brevity, we will denote $\chi(S; F_j) =: \chi_j$. Suppose the number of binary samples taken per filter function used is *N*. Denote each binary sample r_{ji} and

$$\hat{y}_j = \frac{1}{N}\sum_{i=1}^{N} r_{ji}. \tag{7}$$

Then $\hat{y}_j$ is a binomial random variable with mean and variance given by

$$\mathbb{E}[\hat{y}_j] = \frac{1}{2}(1 + e^{-\chi_j}) \text{ and} \tag{8a}$$

$$\mathbb{V}[\hat{y}_j] = \frac{1}{4N}(1 - e^{-2\chi_j}). \tag{8b}$$

Consider the random variable $\hat{\chi}_j = -\log(2\hat{y}_j - 1)$. Taylor expanding about the mean—that is, about the variable $\hat{y}_j - \mathbb{E}[\hat{y}_j]$—we have

$$\mathbb{E}[\hat{\chi}_j] \approx \chi_j, \tag{9a}$$

$$\mathbb{V}[\hat{\chi}_j] \approx \frac{e^{2\chi_j} - 1}{N}. \tag{9b}$$

Another way to specify data when this approximation is valid is to treat $\hat{\chi}_j \sim \mathcal{N}(\chi_j, \sigma_j^2)$, for each filter F_j, where

$$\sigma_j^2 = \frac{e^{2\chi_j} - 1}{N}. \tag{11}$$

Though this approximation will be valid when strong assumptions are made to reduce the model dimension on *S*, it's real utility is in allowing a tractable solution for weak assumptions on *S*. In section 4, we will specify precisely how we model *S* when this approximation holds.

3.2. Small data: sequential Monte Carlo (SMC)

For smaller data sets, the normality assumptions made in the previous section are difficult to justify, and may fail altogether. Thus, we must take an alternative approach to calculating the posterior distribution given by (6). To do so, we note that the main advantage of the Gaussian approach of the previous section was that it will allow us to represent the prior and the posterior distributions as being different members of the same family of distributions. A promising alternative approach, then, is to consider more general families of distributions, perhaps at cost of greater computational effort. In particular, we will use the SMC approximation, which represents the distributions $\Pr(S|F_j)$ and $\Pr(S|r; F_j)$ appearing in (6) by weighted sums of δ-distributions. This

---

[8] The definition of 'large' is intentionally left ambiguous as it depends on far too many things to give a precise number to.

Figure 2. The filter functions of the 25 sequences we consider in this work.
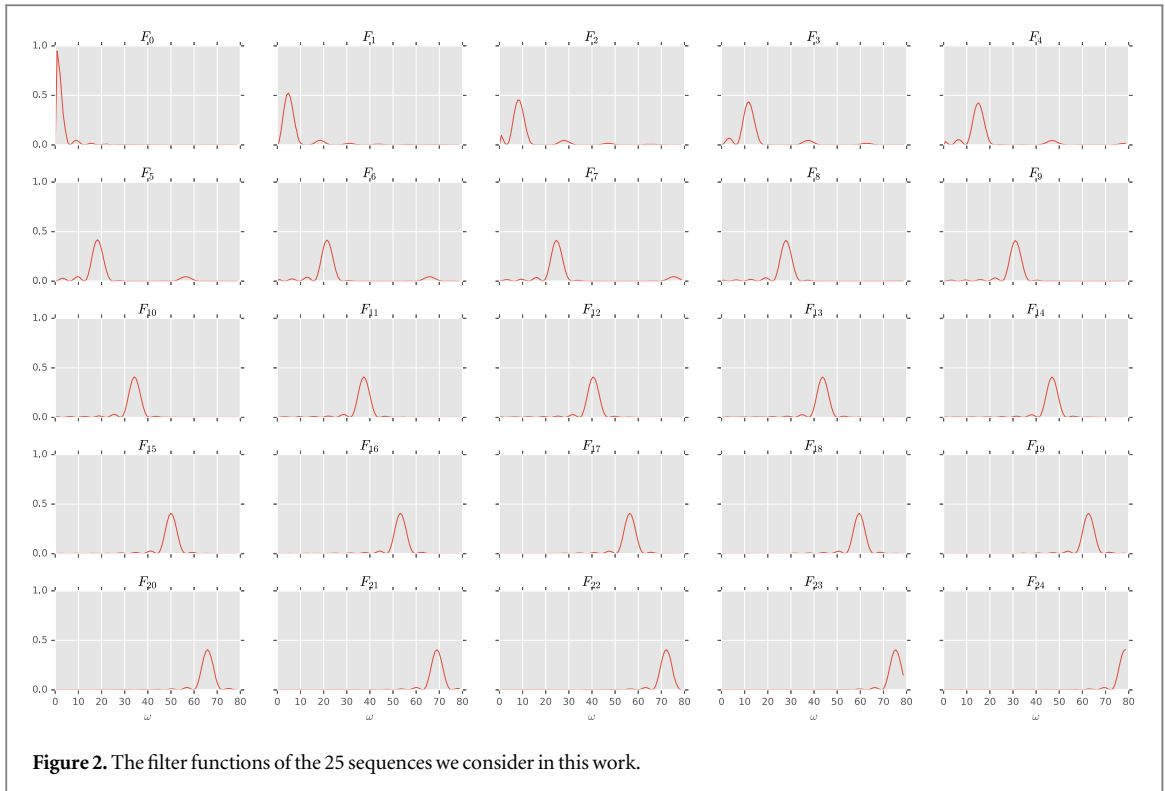
approximation is very general, and will allow us to be much more general in our treatment of $\Pr(r|S; F_j)$. In particular, using SMC will allow us to easily express models for spectral density functions that can be described using a small number of parameters, such as $1/\omega^\alpha$ for an unknown power α.

On the other hand, using SMC forgoes the benefits of the analytic approximations described in the previous section, such that there is a natural tradeoff between the two approaches with the amount of data being taken, and with the form of the models under consideration. We detail the SMC-based approach and compare it to the Gaussian process approach in section 5.

3.3. Filter functions and 'naive' estimator

For our numerical experiments we consider filters that arise from compressing so-called CPMG sequences [31] of increasing number of pulses in a fixed total time T, as was done in [32] for example. For a CPMG sequence of p pulses, $y(t)$ is a function which switches between -1 and 1 at every pulse time $t_i = (2i + 1) T/(2p)$. More involved sequence choices can be made, if one is interested in exploring higher frequency regimes, for example [12, 33], but this simple choice is enough for our purposes. A useful feature of this choice is that it provides an intuitive way of producing filters whose main support is in a given frequency range. More specifically, the larger p is the higher in frequency the main peak of $F(\omega)$ is. This is important because if no filter had support in a given frequency range it would be impossible to accurately estimate the power spectra in that regime. We plot the filter functions used in this paper in figure 2.

For the large data limit, we can use a simple data fitting estimator for a point of reference. In forming our naive spectral density estimate, we will require the normalizations $f_j := \int_0^\Omega \mathrm{d}\omega F_j(\omega)$ and the maxima $\overline{\omega}_j := \mathrm{argmax}_\omega F_j(\omega)$. For the filters we consider (see figure 2), the normalizations and maxima are plotted in figure 3. Next, we approximate each filter function as

$$F_j(\omega) \approx \delta(\omega - \overline{\omega}_j) f_j. \tag{12}$$

The above leads to

$$\chi_j \approx \frac{S(\overline{\omega}_j) f_j}{2\pi}. \tag{13}$$

Suppose we identify the experimentally observed random variables $\hat{\chi}_j$ with the theoretical values χ_j. Then, inverting (13), we define the naive estimate
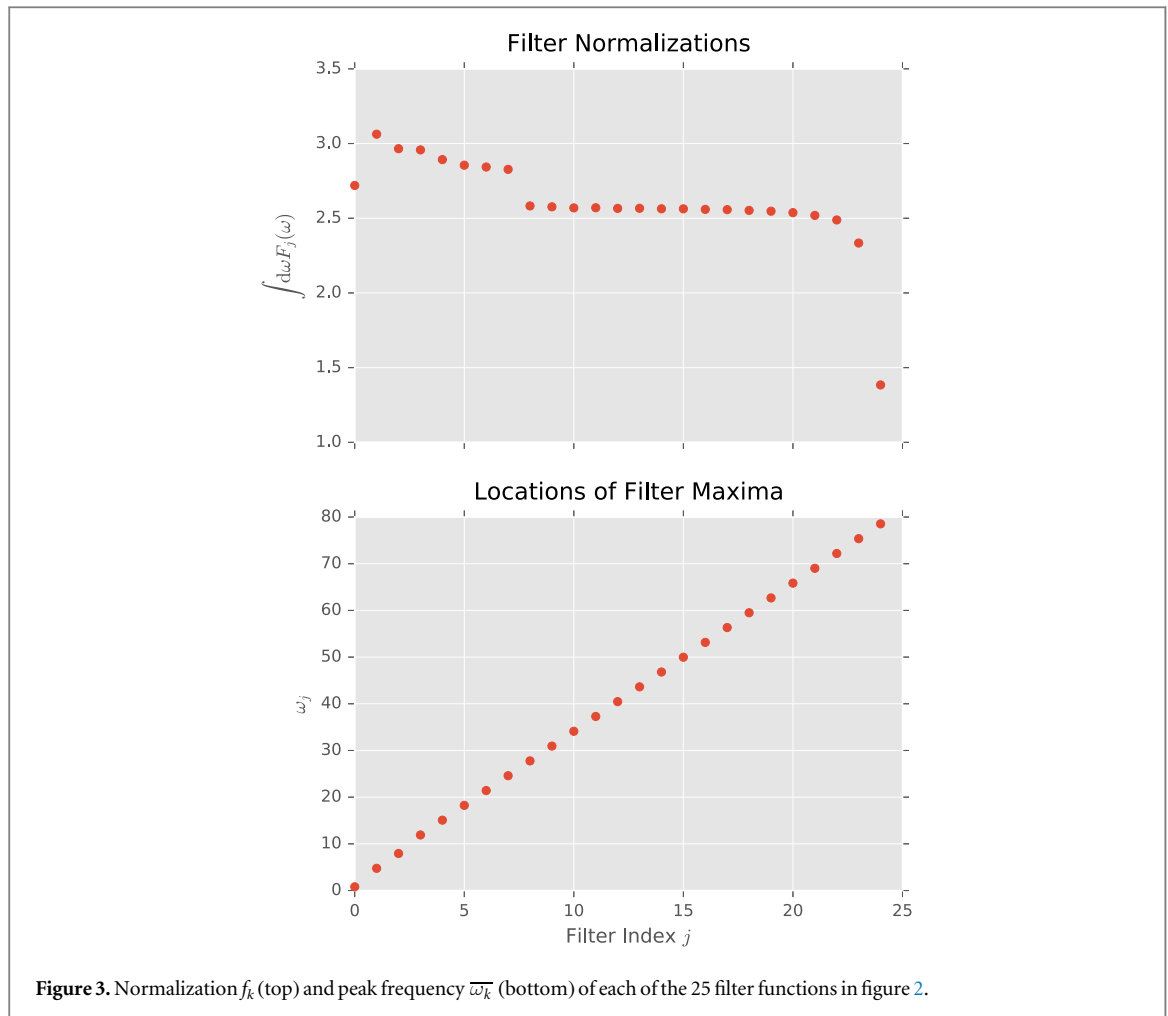
Figure 3. Normalization f_k (top) and peak frequency $\overline{\omega_k}$ (bottom) of each of the 25 filter functions in figure 2.

$$\hat{S}_{\text{naive}}(\overline{\omega}_j) = \frac{2\pi\hat{\chi}_j}{f_j}. \tag{14}$$

We highlight that this is the same approximation made in earlier spectroscopy experimental studies, see for exampled [28], and thus our naive estimator is essentially equivalent to those approaches. This will be compared to some more sophisticated, but more computationally expensive estimators. An example output of the naive estimator is shown in figure 4. We stress that more advanced spectroscopy protocols do exist, based on control generated frequency combs for example (see also [34] for a recent comprehensive review on the topic), that are capable of providing more detailed estimates of the spectrum, but we chose here the most basic protocol available to avoid distracting an interested reader.

4. Big data: Gaussian process regression

Suppose we are in the large data limit. That is, N is big enough that all distributions are roughly Gaussian. To deal with the notion of a prior, or measure, on functions we treat the unknown spectrum as a random function $S(\omega)$. Denote the distribution of S as $\Pr(S)$. To specify this concretely we discretize the support of the distribution to the set $W = \{\omega_1, \ldots, \omega_M\}$. This means that, whenever a numerical calculation is performed, we really only consider random variables $S_k := S(\omega_k)$, which can be represented collectively as the vector $\boldsymbol{S}$.

The simplest non-trivial distribution is Gaussian (or normal): $\boldsymbol{S} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{k})$, where $\boldsymbol{k}_{jk} := k(\omega_j, \omega_k)$ and $\boldsymbol{\mu}_j := \mu(\omega_j)$ are the covariance and mean. In the function space picture we write this $S(\omega) \sim \mathcal{GP}(\mu(\omega), k(\omega, \omega'))$, where $\mathcal{GP}$ stands for *Gaussian process* [35], μ is the *mean function* and k is the covariance function, or *kernel*. In standard notation,

$$\mu(\omega) = \mathbb{E}_S[S(\omega)] \text{ and} \tag{14}$$

$$k(\omega, \omega') = \mathbb{E}_S[(S(\omega) - \mu(\omega))(S(\omega') - \mu(\omega'))]. \tag{15}$$

In principle we can choose any functions μ and k as our mean and kernel functions. However, there are natural choices and ones that have been found to perform well in a broad range of problems. The most common kernel
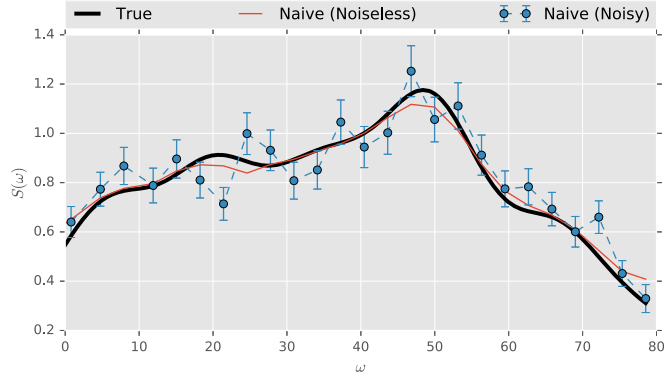
Figure 4. A randomly chosen true spectrum, compared to naive estimates from noiseless and noisy data. The data for each estimate is simulated using the 25 control sequences discussed in the main text, with the noisy data being simulated for $N = 1000$ repetitions per control sequence, and with the noiseless data being simulated for the limit of infinite repetitions per sequence. The discrepancy between the noiseless data curve and the true curve is due to the approximation of the filters as delta functions in the theory.
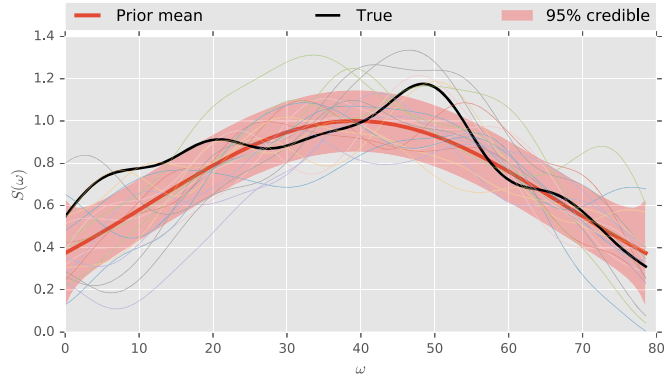

Figure 5. A visualization of a Gaussian Process. Here the mean function μ is taken to be a Gaussian function and we use the squared exponential kernel in (16) with parameters $\kappa = 0.02$ and $\delta = 100$. In red, the mean and 95% credible band is plotted. The other curves are samples from this GP. One of them, in solid black, we take to be the true spectrum.

is the so-called squared exponential[9],

$$k(\omega, \omega') = \kappa e^{-\frac{(\omega - \omega')^2}{\delta}},$$
(16)

where δ is a hyperparameter which controls the correlation in S for nearby ω and κ controls the overall prior uncertainty. In a purely Bayesian context, we should have *a priori* values for μ, κ, and δ. In other words, we believe the 'true' spectrum is drawn according to a GP with these parameters. If this is the case, then no more needs to be done. If not, we would need to perform *model selection* [36]. The topic of model selection is beyond the scope of this work and so we will chose specific values for μ, κ, and δ. To get some intuition for how this relates to qubit noise spectra, we have plotted a visualization of the GP prior we will use in figure 5.

If we begin with a GP prior and the distribution of data is also Gaussian, then the posterior is Gaussian and we can derive an analytic expression for its mean function and kernel. As discussed above, in the large data limit we can do just this. Moreover, since each experiment is uncorrelated, we can process the data at once treating $\hat{\chi} \sim \mathcal{N}(\chi, \Sigma)$, where Σ is a diagonal covariance matrix with entries given by σ_j^2.

Since the prior is Gaussian and the likelihood function is Gaussian, the posterior is also Gaussian. Determining its mean and covariance is a simple exercise in multivariate completing the square. Denote G as the matrix with entries $G_{kj} = F_j(\omega_k)(\omega_k - \omega_{k-1})/4\pi$, such that the trapezoidal rule applied to χ_j is written

$$\chi_j \approx \frac{1}{4\pi} \sum_k F_j(\omega_k) S(\omega_k)(\omega_k - \omega_{k-1}),$$
(17)

$$= \sum_k G_{jk} S_k.$$
(18)

---

[9] Note that a Gaussian function describes five different things in this paper!
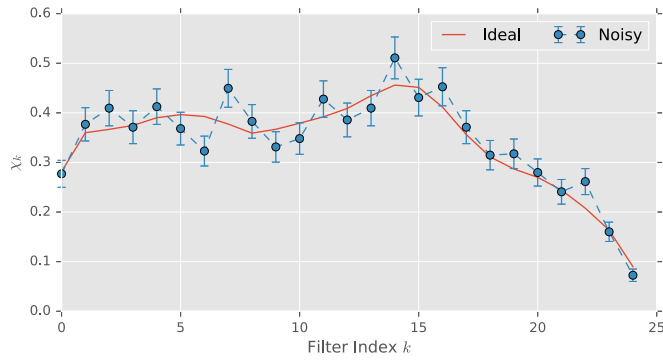
Figure 6. For $N = 1000$ repetitions, and using the true spectrum in figure 5, we plot the observed data superposed over the theoretical (infinite precision) χ's.
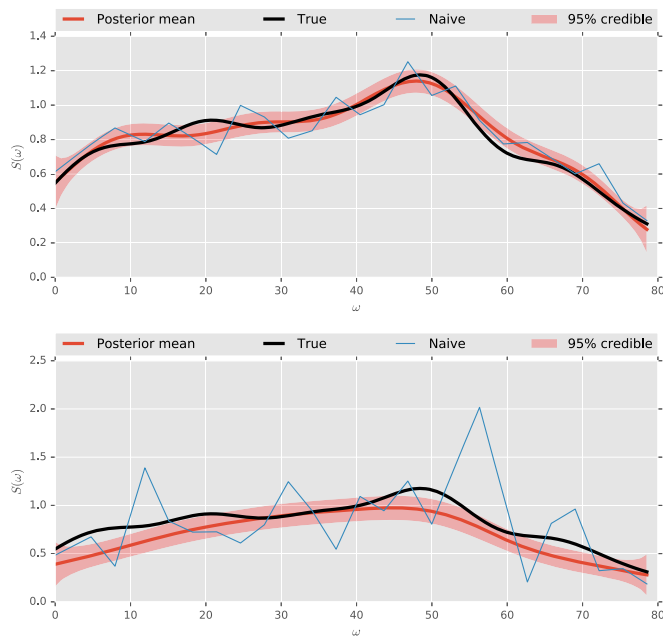

Figure 7. The posterior Gaussian Process, naive estimator and true spectrum using the data plotted in figure 6 (top), and with similar data taken using 100 shots per filter function (bottom).

Then, Bayesian updating amounts to updating the covariance and mean as follows [35]:

$$k \mapsto k' = G^{\mathrm{T}} \Sigma^{-1} G + k^{-1}, \tag{19}$$

$$\mu \mapsto k'^{-1}(\chi^{\mathrm{T}} \Sigma^{-1} G + \mu^{\mathrm{T}} k^{-1}). \tag{20}$$

Note that these equations do not take account of postivity of S. To a large extent, this is guaranteed by the prior. However, outlier data can force the credible regions to extend to $S < 0$. In such cases, we treated those simulations as heralded outliers. We also briefly investigated forced truncations as well as truncated Guassian approximations. Both worked as we expected, but offered only a small increase in accuracy at the cost of simplicity and computation. Thus, we provide the results of the unconstrained algorithm.

Let us take a look at an example simulation and use this GP estimator to find the spectrum. In all simulations, the resolution on the vector of frequencies is taken to be $M = 100$. First, in figure 6 we plot the simulated data for $N = 1000$ repetitions per experiment.

Using the data plotted in figure 6, we apply equations (19) and (20) to get the posterior Gaussian Process. We plot the mean function and 95% credible band in figure 7. We see that the naive and GP estimator agree reasonably well when plenty of data is available. We also plot the same procedures for much less data ($N = 100$), where it is evident that naive estimator fails completely. Whereas, the GP estimator correctly hedges its bets by not suggesting any extreme features deviating the prior GP—it 'knows' it does not have enough data to do so. A more extensive analysis of this difference is presented in section 6.

Figure 8. The posterior Gaussian Process and true spectrum for a $1/f^\alpha$ model (21), demonstrating that the Gaussian Process formalism does not consider the additional structure provided by the spectral model. For this simulation, each experiment was repeated $N = 50$ times and α was chosen uniformly at random in the interval $[1/2, 1]$. ($A = 10, c = 3$.)
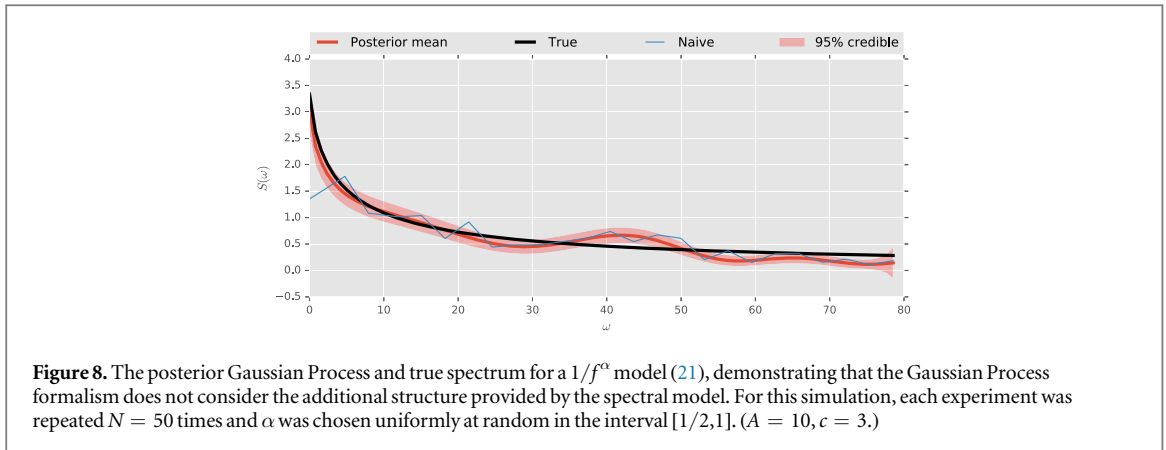
The GP estimator is not a silver bullet, however. First, as already noted, it is only valid when the data are drawn according to a Gaussian distribution—for example, when the Central Limit theorem applies. Second, unless augmented with more sophisticated machine learning algorithms, the GP estimator can only reliably estimate features with sizes the order of δ. To illustrate this, consider now the one-on-f model (or, more generally the $1/f^\alpha$ model, with f and ω being interchangeable here). This corresponds to a spectrum,

$$S(\omega; A, \alpha, c) = \frac{A}{\omega^\alpha + c}, \tag{21}$$

where c gives an effective low frequency cut-off. In this case, the spectrum has a high amount of structure. Indeed, the entire functional form is dictated by only a few parameters. We do not expect, then, that with the freedom allowed by the GP, it will be able to find this structure without a large amount of data. In figure 8, we see that the GP is blind to the structure in the $1/f^\alpha$ model and estimates additional non-existent features. While the naive estimator also suffers from this problem, in the next section we will show how to incorporate information about this structure by the method of hyperparameters. This provides an alternative approach for cases in which a global property of the spectrum is of more experimental interest than the spectrum itself (as in, for example, [37]).

After we describe the final estimator, we will provide an extensive numerical comparison of all estimators in section 6.

5. Small data: hyperparameterized nonlinear regression

In the $1/f^\alpha$ example discussed in the previous section, the prior uncertainty was concentrated on a small number of parameters, such that the distribution of the spectra at each point is very highly correlated. That is, if we perfectly knew the parameters A, α, and c as they appear in (21), we would be able to precisely predict the value of $S(\omega)$ at arbitrary ω.

Though we can describe these correlations using the methods of the previous section, such that the covariance kernel is effectively a low-rank linear operator, it can also be very useful to describe our learning problem more directly. For example, in (21), we can interpret (A, α, c) as a vector of parameters in its own right, as this vector describes the distribution over the alternative parameterization implied by our discretization of $S(\omega)$.

In making this interpretation, we will use the method of hyperparameters to incorporate our knowledge of an appropriate functional form for spectra into our estimation model directly. This is the standard approach in Hamiltonian parameter estimation, for example, where prior knowledge of the physics allows for a drastic reduction in model dimension [38]. Bayesian inference can be then applied directly to the such hyperparameters instead of at the level of the bare physical model.

Generally, if the likelihood function $\Pr(r|\boldsymbol{\eta})$ depends on a model vector $\boldsymbol{\eta}$ that itself is distributed as $\Pr(\boldsymbol{\eta}|\boldsymbol{\theta})$ for some other vector $\boldsymbol{\theta}$, then we can consider the marginalized distribution

$$\Pr(r|\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{\eta}|\boldsymbol{\theta}}[\Pr(r|\boldsymbol{\eta}, \boldsymbol{\theta})] \tag{22a}$$

$$= \int_{\text{supp}\boldsymbol{\eta}} \mathrm{d}\boldsymbol{\eta} \Pr(r|\boldsymbol{\eta}) \Pr(\boldsymbol{\eta}|\boldsymbol{\theta}) \tag{22b}$$

as a likelihood function in its own right.

Returning to the problem of spectral density estimation, we note that we can readily define the model vector $\boldsymbol{\eta}$ by the inner product $\langle S, F_k \rangle = \chi(S; F_k) = \frac{1}{2\pi} \int S(\omega) F_k(\omega) \mathrm{d}\omega$, as predicting each inner product χ_k is

Figure 9. An illustrative example. The difference in $1/f^\alpha$ functions when evaluated at the estimated $\hat{\alpha} = \mathbb{E}[\alpha]$ and the Bayes estimate of the spectrum $\hat{S} = \mathbb{E}[S(\alpha)]$. This illustrates that it matters quite a bit where the mean in a 'mean estimate' is taken. In this example, the hyperparameters are chosen according to (26).
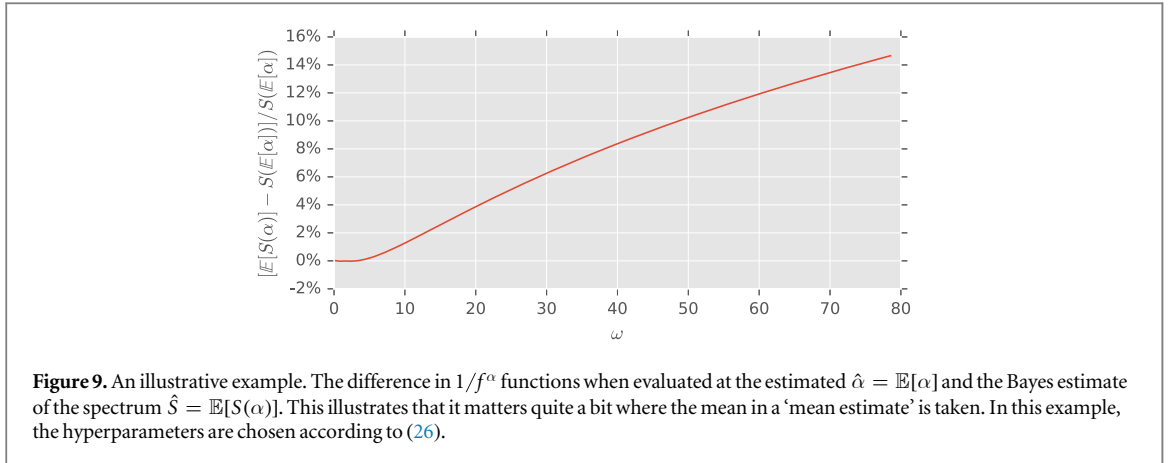
sufficient to reproduce the entire likelihood function. From this perspective, if we can reproduce each χ_k from a lower-dimensional model $\boldsymbol{\theta}$ (that is, a model with less parameters than the number of filter functions used to gather data), then $\boldsymbol{\theta}$ represents a more efficient hyperparameterization than taking $\boldsymbol{\eta}$ directly. For example, consider hyperparameterizing (4) following a $1/f^\alpha$ model,

$$\chi_k(\boldsymbol{\theta}) = \int_0^\Omega \frac{AF_k(\omega)}{\omega^\alpha + c} d\omega, \tag{23}$$

for a given ultraviolet cut-off Ω and with $\boldsymbol{\theta} = (A, \alpha, c)$. This model has been studied experimentally, in particular as a diagnostic for superconducting and spin qubits [6, 7], such that improvements even in this simple example immediately yield experimental benefits.

By expressing the estimation problem in terms of the hyperparameters $\boldsymbol{\theta}$, we introduce a subtle distinction in how we report our final estimates $\hat{S}$ once we have obtained a datum r. We can report the spectrum evaluated at the estimated hyperparameters $\hat{\boldsymbol{\theta}} = \mathbb{E}[\boldsymbol{\theta}|r]$ for our recorded data. This estimate achives the best possible mean-squared error for reporting the hyperparameters themselves, but does not necessarily provide the best estimate of S. As an alternative, we can instead report the Bayesian mean estimate of the spectrum directly,

$$\hat{S}(\omega) = \mathbb{E}_{\boldsymbol{\theta}}[S(\omega; \boldsymbol{\theta})|r]. \tag{24}$$

That is, by taking the spectra and then the mean, we obtain the Bayesian estimate of the spectrum, using our knowledge of the hyperparameters. Though these two methods coincide for spectrum models that are linear functions of their hyperparameters, for models such as $1/f^\alpha$, the difference can be quite significant, as demonstrated in figure 9.
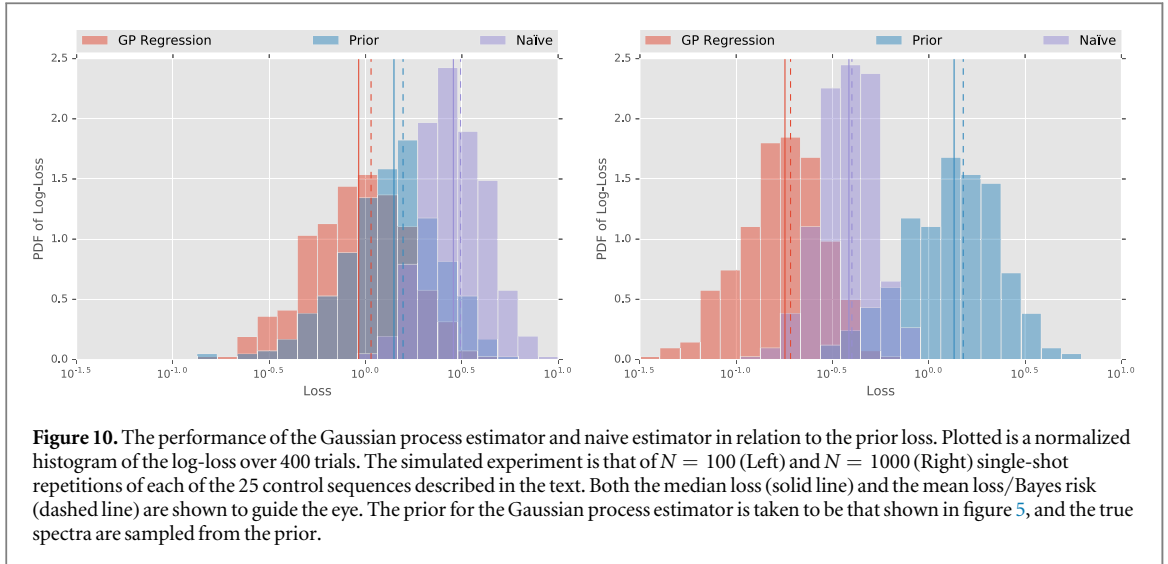
Critically, both methods for estimating S from a posterior over hyperparameters can be generated easily from the same data without additional analysis. Thus, we are free to report the optimal estimate for questions of experimental interest, rather than assuming *a priori* that only one question will be asked of our data.

In any case, we treat the spectrum more generally as being drawn from a parameterized distribution of spectra, such that $\Pr(S(\omega)|\boldsymbol{\theta}) = \delta(S(\omega) - S_{\boldsymbol{\theta}}(\omega))$ and $\boldsymbol{\theta}$ is a real-valued vector for a functional form $S_{\boldsymbol{\theta}}(\omega)$ such as the $1/f^\alpha$ model discussed above. Therefore, once we have specified $\boldsymbol{\theta}$, we have specified the unknown spectrum. Following Bayes' rule (6) as usual gives us a posterior distribution over the parameters $\boldsymbol{\theta}$, conditioned on a data record r,

$$\Pr(\boldsymbol{\theta}|r) = \frac{\Pr(r|\boldsymbol{\theta})\Pr(\boldsymbol{\theta})}{\Pr(r)}. \tag{25}$$

Although the denominator looks like an innocuous normalization constant, producing accurate estimates of $\boldsymbol{\theta}$ requires its calculation. Since we do not assume that $S_{\boldsymbol{\theta}}$ is a linear function of $\boldsymbol{\theta}$, and since we are concerned with efficiently utilizing small amounts of data, the analytic solution in terms of Gaussian process regression used for the process-model case cannot be directly applied here. In lieu of that, our preferred method is SMC [18], also known as particle filtering. This algorithm computes posterior distributions of the form given as (25) by evaluating the likelihood at each of many different *particles*, each of which represents a particular hypothesis about the true model vector $\boldsymbol{\theta}$ and an associated weight. Expectation values over the posterior can then be replaced by finite sums over the particles in the SMC approximation.

We will use the implementation provided by the QInfer package for Python [39]. In the following section, we detail and present results obtained from a QInfer model for hyperparameterized spectral density estimation, and compare these results to those obtained from the Gaussian process regression method of section 4. Our QInfer model will depend on the specification of a set of test frequencies $W = \{\omega_1, ..., \omega_M\}$ and a spectral model

Figure 10. The performance of the Gaussian process estimator and naive estimator in relation to the prior loss. Plotted is a normalized histogram of the log-loss over 400 trials. The simulated experiment is that of $N = 100$ (Left) and $N = 1000$ (Right) single-shot repetitions of each of the 25 control sequences described in the text. Both the median loss (solid line) and the mean loss/Bayes risk (dashed line) are shown to guide the eye. The prior for the Gaussian process estimator is taken to be that shown in figure 5, and the true spectra are sampled from the prior.

function $S(\omega; \boldsymbol{\theta})$, where M specifies the resolution of test frequencies. The inner products $\langle S(\omega; \theta), F_k(\omega) \rangle$ will then be approximated by numerically evaluating the integral (4) in terms of the trapezoidal rule (18) applied to the integrand $\{S(\omega_i; \theta) F_k(\omega_i) : \omega_i \in W\}$. This design allows for our model to be very general with respect to the particular choice of $S(\omega; \boldsymbol{\theta})$.

In this paper, we will work with one such description by writing the $1/f^\alpha$ model as a *hierarchal model* in which r is a random variable that is defined by its distribution conditioned on our new hyperparameters $\boldsymbol{\theta}$. In particular, the conditional distribution of r is given by[10]

$$r|A, \alpha, c \sim \text{Bernoulli}\left(\frac{1 + e^{-\chi_k(A,\alpha,c)}}{2}\right), \tag{26}$$

with our prior on the hyperparameters $\boldsymbol{\theta} = (A, \alpha, c)$ given by

$$A \sim \text{Normal}(10, 0.025), \tag{27a}$$

$$\alpha \sim \text{Uniform}([0.5, 1]), \tag{27b}$$

$$\text{and } c \sim \text{Exponential}(0.1, 3). \tag{27c}$$

For more details, please see the complete source code provided in the supplemental material.

6. Numerical experiments

We have already demonstrated some comparisons between the different approaches in the previous sections. The purpose of this section is to consolidate and expand on those illustrative comparisons. Namely, we will compare the performance of the naive, GP and hyperparameter estimator over many randomly chosen spectra. The comparison is facilitated by the mean-squared error metric. Let $\hat{S}$ be an estimate of the true spectrum S. Then, the error—or *loss*—is defined as
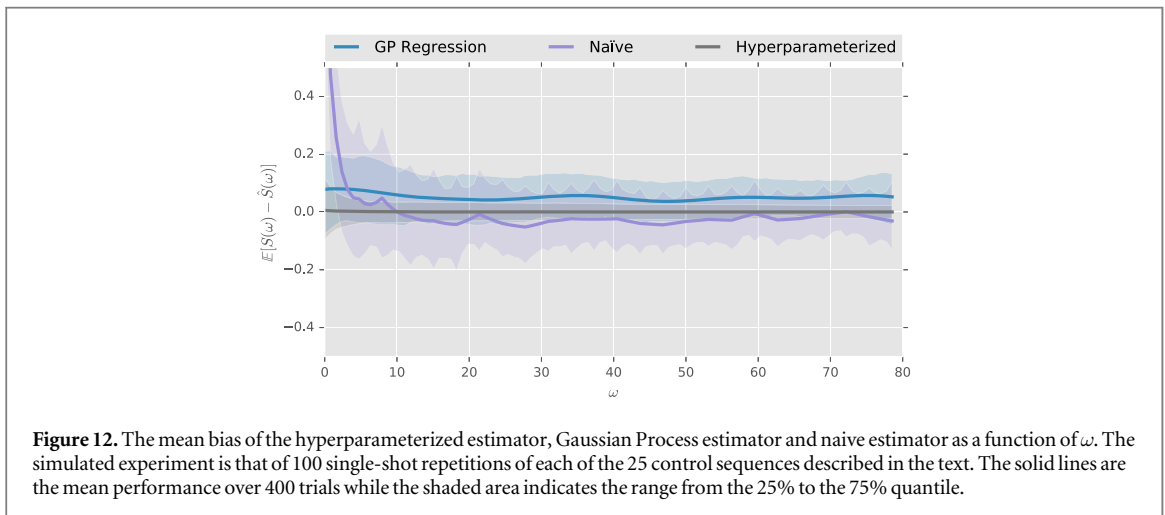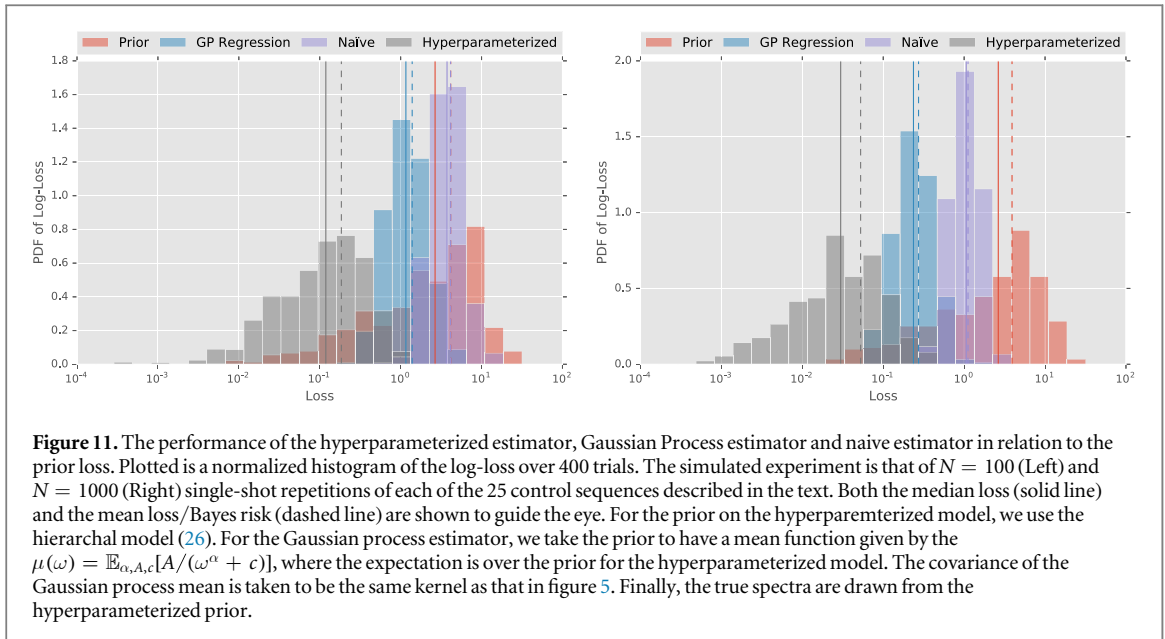
$$L(S, \hat{S}) = \int_0^\Omega |S(\omega) - \hat{S}(\omega)|^2 d\omega. \tag{28}$$

To compare different estimators, we select a true spectrum randomly from the prior, simulate experiments, calculate the estimators, record the loss of each, and repeat. Then we plot a histogram of the achieved accuracy.

We start with the true spectrum randomly selected by sampling a Gaussian Process, as exemplified in figure 5. In this case, the hyperparameterized estimator is not useful[11], so we only compare the GP estimator with the naive estimator. As a point of reference, we can also treat the prior mean function as an estimator and calculate its loss. This is equivalent to the earlier comparison in figure 7, but we average the results of many trials. The result of 400 trials is shown in figure 10. As expected, the posterior loss is lower than the prior loss, indicating that the algorithm is learning. The naive loss does a respectable job as well, but is convincingly beaten by the GP

---

[10] An exponentially distributed random variable is denoted $x \sim \text{Exponential}(x_l, \lambda)$ and has a probability density function $f(x) = \lambda \exp(-\lambda(x - x_l))$ for $x > x_l$.

[11] The hyperparameters are either not defined or have equivalent dimension to the GP estimator, in which case they would serve only to approximate something that can be analytically calculated.

Figure 11. The performance of the hyperparameterized estimator, Gaussian Process estimator and naive estimator in relation to the prior loss. Plotted is a normalized histogram of the log-loss over 400 trials. The simulated experiment is that of $N = 100$ (Left) and $N = 1000$ (Right) single-shot repetitions of each of the 25 control sequences described in the text. Both the median loss (solid line) and the mean loss/Bayes risk (dashed line) are shown to guide the eye. For the prior on the hyperparemterized model, we use the hierarchal model (26). For the Gaussian process estimator, we take the prior to have a mean function given by the $\mu(\omega) = \mathbb{E}_{\alpha,A,c}[A/(\omega^{\alpha} + c)]$, where the expectation is over the prior for the hyperparameterized model. The covariance of the Gaussian process mean is taken to be the same kernel as that in figure 5. Finally, the true spectra are drawn from the hyperparameterized prior.

Figure 12. The mean bias of the hyperparameterized estimator, Gaussian Process estimator and naive estimator as a function of ω. The simulated experiment is that of 100 single-shot repetitions of each of the 25 control sequences described in the text. The solid lines are the mean performance over 400 trials while the shaded area indicates the range from the 25% to the 75% quantile.
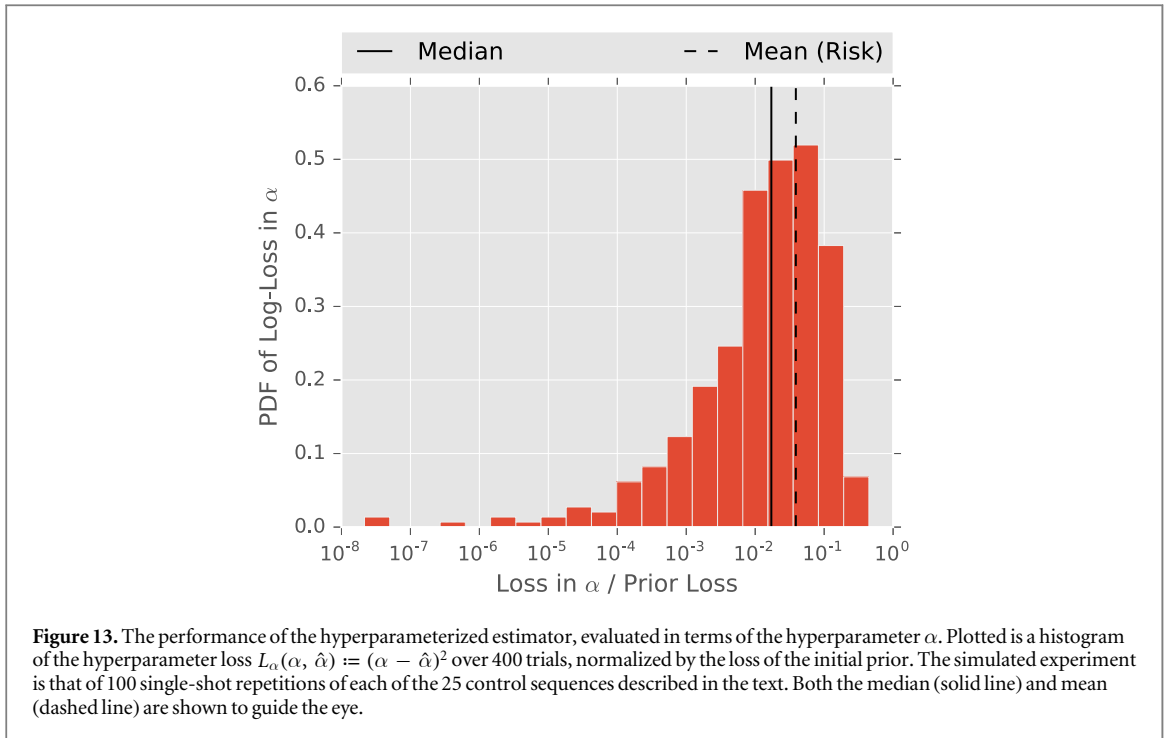
estimator—especially given the fact that GP estimator comes with all the added benefits of the Bayesian methodology discussed above.

In the case of the $1/f^{\alpha}$ model we compare all estimators. The results are presented in figure 11 using another 400 trials. Again, we see that each estimator demonstrate genuine learning by reducing the loss over the prior. The GP still outperforms the naive estimator, but here the hyperparameterized estimator shines, demonstrating that knowledge of the additional structure is immensely beneficial for learning. We also plot the bias of each estimator in figure 12. The hyperparameterized estimator is extremely robust, reducing both the variance and bias over its competitors. By contrast, the GP estimator and naive estimator are reliably biased, and in opposite directions. We do not yet have a strong theoretical explanation for this behaviour.

When performing parametric estimation, as is implied by our hyperparameterized estimator, it is more common to define loss functions on the parameters themselves. The standard loss function is the squared error:

$$L(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}) = \|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\|^2. \tag{29}$$

As this loss function is not defined for the naive or GP estimator, we only report the results for the hyperparameterized estimator. This appears in figure 13. Since we are only considering a single parametric estimator, in figure 13 we plot the ratio of the posterior loss to the prior loss of the hyperparameterized estimator. This demonstrates that the parameter of interest, α, can be learned to two orders of magnitude better accuracy than the prior with no more than 2500 single-shot measurements.

Figure 13. The performance of the hyperparameterized estimator, evaluated in terms of the hyperparameter α. Plotted is a histogram of the hyperparameter loss $L_\alpha(\alpha, \hat{\alpha}) := (\alpha - \hat{\alpha})^2$ over 400 trials, normalized by the loss of the initial prior. The simulated experiment is that of 100 single-shot repetitions of each of the 25 control sequences described in the text. Both the median (solid line) and mean (dashed line) are shown to guide the eye.

7. Discussion

In this work, we formulated the noise spectroscopy problem in the language of statistical estimation theory. This allows us to provide a robust and principled solution to the problem using Bayesian analysis. Considering figure 1 again, we have demonstrated two separate numerical solutions suited to two different regimes in the continuum of possibilities: the big data limit and the low-data/low-dimension limit.

In the large data limit, we can linearize the data into a 'signal plus noise' model and use a non-parametric approach to capture a broad class of spectra. The resultant spectra are not by eye different from a naive data fitting procedure in many cases. However, for a minimal amount of added computation, our approach gives a statistically rigorous accounting of the error bars in the reported spectra.

In the low-data limit we use the exact statistical model. In this case, estimation—or, *learning*—is aided by use of prior knowledge on the model which reduces its dimension. The SMC method is then applied to the resultant parameter estimation problem, which allows accurate inference even for highly nonlinear models in the low-data setting. While here we estimated parameters describing the functional form of $S(\omega)$, it is also possible to use the hyperparameters method to estimate physical quantities of interest when further assumptions about the physics of bath are in place. For example, if the bath is compromised of a collection of harmonic oscillators in a thermal bath and the system couples to the bath in a linear fashion, then there is fixed relation between $S(\omega)$ and the temperature β of the thermal state. As discussed in [12], this relation can be leveraged to estimate β. With the methods described here we can go further, by setting β as the hyperparameter to be determined we can use Bayesian assisted thermometry on the quantum bath requiring a reduced number of measurements.

In sum, we have treated what we consider the core inference problem in noise spectroscopy in order to provide the cleanest demonstration of our algorithms. Within our approach, however, there is no limit to the model complexity that can be treated—requiring only more computational resources. We comment briefly on some generalization that straightforwardly build on these methods. For instance, if one expects the power spectrum to have delta-like peaks—as would be the case if the probe was coupled to a finite number of harmonic oscillators, for example—it is possible to use the position and height of the peaks as hyperparameters in our routines. In an analogous way, our methods can be used for *model selection*—that is, to discriminate between various proposed models for our environment. For instance, making contact with our previous example, one can determine how many oscillators are coupled to our probe. More importantly, the statistical methods developed here are not constrained to the spectroscopy scenario we considered: Gaussian, zero mean noise. More general spectroscopy protocols are based on inverting multidimensional integrals of the form

$$\int d\omega_1 \cdots d\omega_m F^{(m)}(\omega_1, \cdots, \omega_m, T) S^{(m)}(\omega_1, \cdots, \omega_m),$$

where $S^{(m)}(\omega_1, \cdots, \omega_m)$ is the mth order polyspectra and $F^{(m)}(\omega_1, \cdots, \omega_m, T)$ an mth generalized filter function [12, 33], essentially an mth dimensional time ordered Fourier transform of the product $y(t_1) \cdots y(t_m)$. Estimating

$S^{(m)}(\omega_1, \cdots, \omega_m)$ given our ability to manipulate $F^{(m)}(\omega_1, \cdots, \omega_m, T)$ is then a generalization of our current methods to higher dimensional integrals.

Finally, we note that our inferential algorithm can easily be embedded into control software for online (real-time) estimation and, more interestingly, closed-loop adaptive control. These are exciting possibilities not currently offered without significant modification by traditional approaches.

Acknowledgments

References

[1] Lidar D A and Brun T A (ed) 2013 *Quantum Error Correction* (Cambridge: Cambridge University Press)

[2] Viola L and Lloyd S 1998 Dynamical suppression of decoherence in two-state quantum systems *Phys. Rev.* A **58** 2733

[3] Brif C, Chakrabarti R and Rabitz H 2010 Control of quantum phenomena: past, present and future *New J. Phys.* **12** 075008

[4] Glaser S J *et al* 2015 Training Schrödinger's cat: quantum optimal control *Eur. Phys. J.* D **69** 279

[5] Yan F, Bylander J, Gustavsson S, Yoshihara F, Harrabi K, Cory D G, Orlando T P, Nakamura Y, Tsai J-S and Oliver W D 2012 Spectroscopy of low-frequency noise and its temperature dependence in a superconducting qubit *Phys. Rev.* B **85** 174521

[6] Yan F, Gustavsson S, Bylander J, Jin X, Yoshihara F, Cory D G, Nakamura Y, Orlando T P and Oliver W D 2013 Rotating-frame relaxation as a noise spectrum analyser of a superconducting qubit undergoing driven evolution *Nat. Commun.* **4** 2337

[7] Dial O E, Shulman M D, Harvey S P, Bluhm H, Umansky V and Yacoby A 2013 Charge noise spectroscopy using coherent exchange oscillations in a singlet-triplet qubit *Phys. Rev. Lett.* **110** 146804

[8] Yuge T, Sasaki S and Hirayama Y 2011 Measurement of the noise spectrum using a multiple-pulse sequence *Phys. Rev. Lett.* **107** 170504

[9] Young K C and Whaley K B 2012 Qubits as spectrometers of dephasing noise *Phys. Rev.* A **86** 012314

[10] Álvarez G A and Suter D 2011 Measuring the spectrum of colored noise by dynamical decoupling *Phys. Rev. Lett.* **107** 230501

[11] Norris L M, Paz-Silva G A and Viola L 2016 Qubit noise spectroscopy for non-Gaussian dephasing environments *Phys. Rev. Lett.* **116** 150503

[12] Paz-Silva G A, Norris L M and Viola L 2017 Multiqubit spectroscopy of Gaussian quantum noise *Phys. Rev.* A **95** 022121

[13] Muhonen J T *et al* 2014 Storing quantum information for 30 seconds in a nanoelectronic device *Nat. Nano* **9** 986

[14] Breuer H-P and Petruccione F 2002 *The Theory of Open Quantum Systems* (Oxford: Oxford University Press)

[15] Shulman M D, Harvey S P, Nichol J M, Bartlett S D, Doherty A C, Umansky V and Yacoby A 2014 Suppressing qubit dephasing using real-time Hamiltonian estimation *Nat. Commun.* **5** 5156

[16] Wang Y, Um M, Zhang J, An S and Lyu M Single-qubit quantum memory exceeding 10 min coherence time *Nat. Photonics* **11** 646–50

[17] Bretthorst G L 1988 *Bayesian Spectrum Analysis and Parameter Estimation* (New York: Springer)

[18] Doucet A and Johansen A M 2011 A tutorial on particle filtering and smoothing: fifteen years later *The Oxford Handbook of Nonlinear Filtering* ed D Crisan and B Rozovskii (Oxford: Oxford University Press)

[19] Zwick A, Álvarez G A and Kurizki G 2016 Maximizing information on the environment by dynamically controlled qubit probes *Phys. Rev. Appl.* **5** 014007

[20] Frey V M, Mavadia S, Norris L M, de Ferranti W, Lucarelli D, Viola L and Biercuk M J 2017 Application of optimal band-limited control protocols to quantum noise sensing *Nat. Commun.* **8** 2189

[21] Norris L M, Lucarelli D, Frey V M, Mavadia S, Biercuk M J and Viola L 2018 Optimally band-limited spectroscopy of control noise using a qubit sensor *Phys. Rev.* A **98** 032315

[22] Ng S, Ang S Z, Wheatley T A, Yonezawa H, Furusawa A, Huntington E H and Tsang M 2016 Spectrum analysis with quantum dynamical systems *Phys. Rev.* A **93** 042121

[23] Paz-Silva G A and Viola L 2014 General transfer-function approach to noise filtering in open-loop quantum control *Phys. Rev. Lett.* **113** 250501

[24] Kofman A G and Kurizki G 2004 Unified theory of dynamically suppressed qubit decoherence in thermal baths *Phys. Rev. Lett.* **93** 130406

[25] Bar-Gill N, Pham L M, Belthangady C, Sage D L, Cappellaro P, Maze J R, Lukin M D, Yacoby A and Walsworth R 2012 Suppression of spin-bath dynamics for improved coherence of multi-spin-qubit systems *Nat. Commun.* **3** 858

[26] Caldeira A O and Leggett A J 1981 Influence of dissipation on quantum tunneling in macroscopic systems *Phys. Rev. Lett.* **46** 211

[27] Kaulakys B and Gontis V 1997 Quantum anti-Zeno effect *Phys. Rev.* A **56** 1131

[28] Bylander J, Gustavsson S, Yan F, Yoshihara F, Harrabi K, Fitch G, Cory D G, Nakamura Y, Tsai J-S and Oliver W D 2011 Noise spectroscopy through dynamical decoupling with a superconducting flux qubit *Nat. Phys.* **7** 565

[29] Szańkowski P, Trippenbach M and Cywiński L 2016 Spectroscopy of cross correlations of environmental noises with two qubits *Phys. Rev.* A **94** 012109

[30] Sivia D and Skilling J 2006 *Data Analysis: A Bayesian Tutorial* (Oxford: Oxford University Press) Oxford Science Publications

[31] Meiboom S and Gill D 1958 Modified spin-echo method for measuring nuclear relaxation times *Rev. Sci. Instrum.* **29** 688

[32] Álvarez G A and Suter D 2011 Measuring the spectrum of colored noise by dynamical decoupling *Phys. Rev. Lett.* **107** 230501

[33] Norris L M, Paz-Silva G A and Viola L 2016 Qubit noise spectroscopy for non-Gaussian dephasing environments *Phys. Rev. Lett.* **116** 150503

[34] Szańkowski P, Ramon G, Krzywda J, Kwiatkowski D and Cywiński Ł 2017 Environmental noise spectroscopy with qubits subjected to dynamical decoupling *J. Phys.: Condens. Matter* **29** 333001

[35] Rasmussen C E and Williams C K I 2005 *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)* (Cambridge, MA: MIT Press)

[36] Edwards W, Lindman H and Savage L J 1963 Bayesian statistical inference for psychological research *Psychol. Rev.* **70** 193

[37] Medford J, Cywiński L, Barthel C, Marcus C M, Hanson M P and Gossard A C 2012 Scaling of dynamical decoupling for spin qubits *Phys. Rev. Lett.* **108** 086802

[38] Granade C E, Ferrie C, Wiebe N and Cory D G 2012 Robust online Hamiltonian learning *New J. Phys.* **14** 103013

[39] Granade C, Ferrie C, Hincks I, Casagrande S, Alexander T, Gross J, Kononenko M and Sanders Y 2017 QInfer: statistical inference software for quantum applications *Quantum* **1** 5