

# The ubiquitous problem of learning system parameters for dissipative two-level quantum systems: Fourier analysis versus Bayesian estimation

Sophie G. Schirmer<sup>1</sup> and Frank C. Langbein<sup>2</sup>

<sup>1</sup>*College of Science (Physics), Swansea University,  
Singleton Park, Swansea, SA2 8PP, United Kingdom*

<sup>2</sup>*College of Physical Sciences & Engineering (Computer Science & Informatics),  
Cardiff University, 5 The Parade, Cardiff, CF24 3AA, United Kingdom*

(Dated: December 16, 2014)

We compare the accuracy, precision and reliability of different methods for estimating key system parameters for two-level systems subject to Hamiltonian evolution and decoherence. It is demonstrated that the use of Bayesian modelling and maximum likelihood estimation is superior to common techniques based on Fourier analysis. Even for simple two-parameter estimation problems, the Bayesian approach yields higher accuracy and precision for the parameter estimates obtained. It requires less data, is more flexible in dealing with different model systems, can deal better with uncertainty in initial conditions and measurements, and enables adaptive refinement of the estimates. The comparison results shows that this holds for measurements of large ensembles of spins and atoms limited by Gaussian noise as well as projection noise limited data from repeated single-shot measurements of a single quantum device.

PACS numbers: 03.67.Lx, 03.65.Wj

## I. INTRODUCTION

Quantum systems play an important role in atomic and molecular physics, chemistry, material science and many important current technologies such as nuclear magnetic resonance imaging [1] and spectroscopy [2], promising nascent quantum technologies such as spintronic devices [3], and potential future technologies such as quantum information processing [4]. Novel applications require increasingly sophisticated control, and accurate and precise models to facilitate controlled manipulation of their dynamics.

Although theoretical device modelling remains important, system identification and data-driven models are becoming increasingly important in many areas of science and technology to accurately describe individual systems [5]. System identification comprises a range of problems including model identification, model discrimination and model verification. Once a model has been selected, the task often reduces to identifying parameters in the model from experimental data. In the quantum domain this is often data from one of the many types of spectroscopy, from magnetic resonance to laser to electron transmission spectroscopy, depending on the physical system. More recently single shot measurements of quantum systems have also become important for quantum devices relying on individual quantum states.

Fourier analysis of the spectra is frequently used to identify model parameters such as chemical shifts and relaxation rates by examination of the positions and shape of peaks in a free-induction-decay (FID) spectrum [6]. Fourier analysis of Rabi oscillation spectra has also been used to identify Hamiltonians [7, 8], as well as decoherence and relaxation parameters for two-level systems [9], and concurrence spectroscopy [10] has been applied to determine information about coupling between qubits. For

more complex systems, Bayesian techniques and maximum likelihood estimation [11] have proved to be extremely valuable to construct data-driven models to identify Hamiltonian parameters [12] and decoherence parameters for multi-level systems [13]. Bayesian techniques have also been applied for adaptive Hamiltonian learning using sequential Monte-Carlo techniques [14].

In this work we revisit simpler systems: two-level systems subject to decoherence, one of the simplest but arguably most important models in quantum physics. The model is ubiquitous in magnetic resonance imaging, where the magnetization signal from protons (spin- $\frac{1}{2}$  particles) precessing and dephasing in a magnetic field is the basis for non-invasive, in-vivo imaging. In quantum information it describes qubits as the fundamental building blocks subject to decoherence. Therefore, characterization of two-level systems is extremely important. We compare two frequently used estimation strategies based on Fourier analysis and a Bayesian approach combined with maximum likelihood estimation, for the ubiquitous parameter estimation problem of a two-level system subject to decoherence. We consider accuracy, precision and efficiency for different systems and noise models, including Gaussian noise, typically encountered for large ensemble measurements, and projection noise, typically present in data from repeated single-system measurements.

## II. SYSTEM AND EXPERIMENTAL ASSUMPTIONS

In this section we introduce our dynamic model of the physical system and our assumptions about initialisation and measurement of the system. We focus in particular on the different options for the measurements depend-

ing on the nature of the physical system and hence the measurements from which we wish to estimate the parameters.

### A. Dynamic system model

The state of a quantum system is generally described by a density operator  $\rho$ , which, for a system subject to a Markovian environment, evolves according to a Lindblad-type master equation

$$\begin{aligned} \dot{\rho}(t) &= [H_0, \rho(t)] + \mathcal{D}[V]\rho, \\ \mathcal{D}[V] &= V\rho V^\dagger - \frac{1}{2}(V^\dagger V\rho + \rho V^\dagger V), \end{aligned} \quad (1)$$

where  $H$  represents the Hamiltonian and  $V$  the dephasing operator. If the dephasing occurs in the same basis as the Hamiltonian evolution then we can choose a basis in which both  $H_0$  and  $V$  are diagonal. For a two-level system we can thus write  $H = \omega\sigma_z$  and  $V = \tilde{\gamma}\sigma_z$ , where  $\tilde{\gamma} \geq 0$ , leaving us essentially with two core system parameters to identify,  $\omega$  and  $\tilde{\gamma}$ , or often  $\gamma = 2\tilde{\gamma}^2$ .

### B. Initialization and Readout

A basic experiment involves initializing the system in some state  $|\psi_I\rangle$  and measuring the decay signal, a so-called free-induction decay experiment. The measured signal depends on the system parameters as well as the initial state and the measurement. Taking the measurement operator to be of the form

$$M = \begin{pmatrix} \cos\theta_M & \sin\theta_M \\ \sin\theta_M & -\cos\theta_M \end{pmatrix}, \quad (2)$$

and taking the initial state to be

$$|\psi_I\rangle = \cos(\theta_I)|0\rangle + \sin(\theta_I)|1\rangle, \quad (3)$$

the measurement signal is of the form

$$p(t) = e^{-\gamma t} \cos(\omega t) \sin(\theta_I) \sin(\theta_M) + \cos(\theta_I) \cos(\theta_M). \quad (4)$$

Assuming the system is initially in the ground state  $|0\rangle$ , e.g., corresponding to spins being aligned with an external magnetic field, the initialization procedure corresponds to applying a short pulse to put the system into a superposition of the ground and excitation state. Notice if the system is not well characterized then it is likely to be infeasible to prepare the system in a well-defined superposition state with a known angle  $\theta_I$ . Rather,  $\theta_I$  becomes an additional parameter to be estimated.

The operator  $M$  corresponds to measuring the system with regard to an axis tilted by an angle  $\theta_M$  from the system axis in the  $(x, z)$  plane, which can describe many different experimental situations. In an FID experiment in NMR, for example, an  $x$ -magnetization measurement corresponds to setting  $\theta_M = \frac{\pi}{2}$ . In a Rabi spectroscopy

experiment of a quantum dot, where the population of the ground and/or excited state is measured, e.g., via a fluorescence measurement, we would typically set  $\theta_M = 0$ . In some situations, such as the examples mentioned, the Hamiltonian and measurement bases may be well-known. In other situations, however, such as in a double quantum dot system with charge state read-out via a single electron transistor perhaps,  $\theta_M$  may a priori at most be approximately known. In this case  $\theta_I$  becomes an additional parameter to be estimated. In this work we employ a formalism that does not require either the initial state or measurement to be known a priori.

### C. Continuous vs discrete-time and adaptive measurements

In an FID experiment we could in principle measure the decay signal continuously. However, modern receivers typically return a digitized signal, i.e., a vector of time samples, usually the signal values integrated over short time intervals  $\Delta t$ . For this type of readout, the number  $N$  of time samples and their spacing  $\Delta t$  are usually fixed, or at least selected prior to the start of the experiment. In this set-up there is usually little opportunity for adaptive refinement short of simply repeating the entire experiment with shorter  $\Delta t$  or larger  $N$ .

In other situations, such as Rabi spectroscopy [15], each measurement corresponds to a separate experiment. For example, we prepare the system in a certain initial state, let it evolve under some Hamiltonian (with parameters to be estimated) for some time  $t$  before performing a measurement to determine the state of the system. In this case we are more flexible and can in principle choose the measurement times adaptively, trying to optimize the times to maximize the amount of information obtained in each measurement.

Here we mainly consider the case of a regularly sampled measurement signal but we also briefly consider how the estimation can be improved in the latter case by adaptive sampling with particular focus on the comparison between the different estimation strategies.

### D. Ensemble vs single-system measurements

In many settings from NMR and MRI to electron spin resonance (ESR) to atomic ensembles in atom traps, large ensembles of spins or atoms are studied resulting in ensemble average measurements. In this setting, the backaction from the measurement is negligible and the system can be measured continuously to obtain a measurement signal  $s(t)$ . The noise in the signal is well approximated by Gaussian noise, which can be simulated by adding a zero-mean Gaussian noise signal  $g(t)$  to the ideal signal  $p(t)$ , i.e., the measured signal  $d(t) = p(t) + g(t)$ . By the Law of Large Numbers and Iterated Logarithm Law [16] this gives a Gaussian distribution for  $d(t)$  with

mean  $p(t)$  and variance  $\sigma^2 \sim \frac{\log \log N_e}{2N_e}$  for  $N_e \rightarrow \infty$ . This is a good error model for simulating physical systems and estimating the noise in actual measurement data when the ensemble size  $N_e$  is large.

More recently single quantum systems, such as trapped ions [17], trapped atoms [18], single electron spins [19], and charge states in Josephson junctions [20], have become an important topic for research because of their potential relevance to quantum technologies. Given a single copy of a two-level system, measurement of any observable yields only a single bit of information indicating a 0 or 1 result. To determine the expectation value of an observable the experiment has to be repeated many times and the results averaged. Furthermore, due to the backaction of the measurement on the system, we can generally only perform a single projective measurement. To obtain data about the observable at different times the system has to be re-initialized and the experiment repeated for each measurement. In this context the ensemble size  $N_e$  is the number of times each experiment on a single copy of the system is repeated. As repetitions are time- and resource-intensive, it is desirable to keep  $N_e$  small. However, this means the precision of the expectation values of observables becomes limited by projection noise, following a Poisson distribution. To simulate experiments of this type we compute the probability  $\hat{p}_1$  of measurement outcome 1 for the simulated system, generate  $N_e$  random numbers  $r_n$  between 0 and 1, drawn from a uniform distribution, and set  $p_1 = N_1/N_e$ , where  $N_1$  is the number of  $r_n \leq \hat{p}_1$ .

### III. PARAMETER ESTIMATION STRATEGIES

This section introduces the three parameter estimation strategies based on Fourier and Bayesian analysis we wish to compare.

#### A. Fourier-spectrum based estimation

A common technique to find frequency components in a noisy time-domain signal is spectral analysis. Consider a measurement signal of the form

$$p(t) = a + be^{-\gamma t} \cos(\omega_0 t), \quad t \geq 0, \quad (5)$$

which corresponds directly to measurement (4) if we set  $a = \cos \theta_I \cos \theta_M$  and  $b = \sin \theta_I \sin \theta_M$ . Subtracting the mean of the signal  $\langle p(t) \rangle = a$  and rescaling gives  $f(t) = (p(t) - a)/b$ . To account for the fact that  $f(t)$  is defined only for  $t \geq 0$  we multiply  $f(t)$  by the Heaviside function

$$u(t) = \begin{cases} 0 & \text{if } t < 0 \\ 1 & \text{if } t \geq 0. \end{cases}$$

The Fourier transform of  $u(t)f(t) = u(t)e^{-\gamma t} \cos(\omega_0 t)$  is

$$F(\omega) = \frac{\gamma + i\omega}{(\gamma + i\omega)^2 + \omega_0^2}$$

and the power spectrum is  $P(\omega) = |F(\omega)|^2$ . Differentiating with respect to  $\omega$  and setting the numerator to 0 shows that  $|F(\omega)|^2$  has extrema for  $\omega = 0$  and  $(\gamma^2 + \omega^2)^2 - \omega_0^2(4\gamma^2 + \omega_0^2) = 0$ . The real roots  $\omega_*$  of this equation satisfy

$$\mathcal{E}_1(\omega_0, \gamma) = \omega_*^2 + \gamma^2 - \omega_0 \sqrt{4\gamma^2 + \omega_0^2} = 0 \quad (6)$$

and the corresponding maximum of the power spectrum

$$P_* = P(\omega_*) = \frac{\omega_0^2 + \omega_0 \sqrt{4\gamma^2 + \omega_0^2}}{8\gamma^2 \omega_0^2} = \frac{\omega_0^2 + \omega_*^2 + \gamma^2}{8\gamma^2 \omega_0^2}.$$

Defining the error term

$$\mathcal{E}_2(\omega_0, \gamma) = 8\gamma^2 \omega_0^2 P_* - \omega_0^2 + \gamma^2 + \omega_*^2, \quad (7)$$

we can estimate the frequency  $\omega_0$  and dephasing rate  $\gamma$  from the peak height  $P_*$  and position  $\omega_*$  via

#### Strategy 1:

$$\{\omega_0, \gamma\} = \arg \min_{\omega'_0, \gamma'} \{|\mathcal{E}_1(\omega'_0, \gamma')| + |\mathcal{E}_2(\omega'_0, \gamma')|\}. \quad (8)$$

Determining the maximum  $P_*$  and its location  $\omega_*$  from  $|F(\omega)|^2$ , we may choose  $\omega'_0 = \omega_*$  and  $\gamma' = \sqrt{2\omega_*/(8\omega_*^2 P_* - 1)}$  as starting point for a local minimization routine provided  $\gamma \ll \omega_0$  as is usually the case.

Instead of estimating the height of the peak, estimates for  $\omega_0$  and  $\gamma$  can also be obtained using the width of the peak. Let  $\omega_{1,2}$  be the (positive) frequencies for which  $|F(\omega)|$  assumes half its maximum. One way to estimate  $\omega_{1,2}$  is to take the minimum and maximum of  $\{\omega : |F(\omega)| \geq \max(F)\}$ , assuming that sufficient measurements have been made such that  $F$  is symmetric and peaked, i.e., it has low skewness and high kurtosis.

The full-width-half-maximum  $2d$  of  $|F(\omega)|$  is  $|\omega_2 - \omega_1|$  and we can derive the following expression:

$$\begin{aligned} d &= \left[ \sqrt{\omega_0^2 - \gamma^2 + 2\sqrt{3}\omega_0\gamma} - \sqrt{\omega_0^2 - \gamma^2} \right] \\ &= \left[ \sqrt{\omega_*^2 + 2\sqrt{3}\gamma\sqrt{\omega_*^2 + \gamma^2}} - \omega_* \right]. \end{aligned}$$

Hence, given the location  $\omega_*$  and half-width  $d$  of the peak solving for  $\gamma$  gives the alternative

#### Strategy 2:

$$\gamma = \frac{1}{6} \sqrt{6g(\omega_*, d) - 18\omega_*^2}, \quad \omega_0 = \sqrt{\omega_*^2 + \gamma^2}, \quad (9)$$

where  $g(\omega_*, d) = \sqrt{9\omega_*^4 + 12d^2\omega_*^2 + 12d^3\omega_* + 3d^4}$ .

Strategy 2 based on peak-positions and linewidths is probably the most common approach for estimating frequencies and  $R_2$ -relaxation rates from FID signals in NMR and in many other contexts. The expressions for  $|P(\omega)|^2$ , the peak heights and linewidth are more complicated than those for quadrature measurements as we only have a real cosine signal but the approach is fundamentally the same.

## B. Bayesian and Maximum Likelihood Approach

Given discrete time-sampled data represented by a row vector  $\mathbf{d}$  of length  $N_t$  containing the measurement results obtained at times  $t_n$  for  $n = 1, \dots, N_t$ , let  $\mathbf{p}$  be the vector of the corresponding measurement outcomes predicted by the model.  $\mathbf{p}$  depends on the model parameters, here  $\omega_0$  and  $\gamma$ . Assuming Gaussian noise with variance  $\sigma^2$  we define the joint likelihood [11]

$$P(\mathbf{p}, \mathbf{d}, \sigma) = \frac{1}{(\sqrt{2\pi}\sigma)^{N_t}} \exp \left[ -\frac{\|\mathbf{p} - \mathbf{d}\|_2^2}{2\sigma^2} \right]. \quad (10)$$

If the noise level  $\sigma$  of the data is not known a priori, we can eliminate this parameter following the standard Bayesian approach by integrating over  $\sigma$  from 0 to  $\infty$ , using the Jeffrey's prior  $\sigma^{-1}$ . This gives

$$P(\mathbf{p}, \mathbf{d}) = \frac{\Gamma(\frac{N_t}{2} - 1)}{4\pi^{N_t/2}} \|\mathbf{p} - \mathbf{d}\|_2^{2-N_t} \quad (11)$$

where  $\Gamma$  is the Gamma function. It is usually more convenient and numerically robust to work with the (negative) logarithm of the likelihood function, the so-called log-likelihood. When the noise level  $\sigma$  is known the log-likelihood reduces to

$$L(\mathbf{p}, \mathbf{d}, \sigma) = -\log P(\mathbf{p}, \mathbf{d}, \sigma) = \frac{1}{2\sigma^2} \|\mathbf{p} - \mathbf{d}\|_2^2 + \text{const}, \quad (12)$$

where the constant is usually omitted; when  $\sigma$  is not known a priori we obtain instead

$$L(\mathbf{p}, \mathbf{d}) = -\log P(\mathbf{p}, \mathbf{d}) = \frac{1 - N_t}{2} \log \|\mathbf{p} - \mathbf{d}\|_2^2 + \text{const}. \quad (13)$$

The idea of maximum likelihood estimation is to find the model parameters that maximize this (log-)likelihood function. To simplify this task, we follow a similar approach as in previous work [11–13] and express the signals as linear combinations of a small number  $m_b$  of basis functions  $g_m(t)$  determined by the functional form of the signals. In our case the measurement signal  $p(t)$  can be written as a linear combination of  $m_b = 2$  basis functions

$$p(t) = \alpha_1 g_1(t) + \alpha_2 g_2(t). \quad (14)$$

with  $g_1(t) = 1$  and  $g_2(t) = e^{-\gamma t} \cos(\omega_0 t)$ . As the basis functions are not orthogonal, we define an orthogonal projection of the data onto the basis functions sampled at times  $t_n$  as follows. Let  $G$  be a matrix whose rows are the basis functions  $g_m(t)$  evaluated at times  $t_n$ ,  $G_{mn} = g_m(t_n)$ , and  $E \text{diag}(\alpha_m) E^\dagger$  be the eigendecomposition of the positive-definite matrix  $GG^\dagger$ . Then  $H = \text{diag}(\alpha_m^{-1/2}) E^\dagger G$  is a matrix satisfying  $H^\dagger H = GG^\dagger$ , whose rows form an orthonormal set,  $HH^\dagger = I$ , and we define the orthogonal projection of the data vectors onto the basis function by  $\mathbf{h} = H\mathbf{d}^\dagger$ .

Projecting the data onto a linear combination of basis functions introduced  $m_b$  nuisance parameters  $\alpha_m$ . Using a standard Bayesian approach we can eliminate them

by integration using a uniform prior, and following further simplifications [11], it can be shown that the log-likelihood (11) becomes

$$L(\omega_0, \gamma | \mathbf{d}) = \frac{m_b - N_t}{2} \log \left[ 1 - \frac{m_b \langle \mathbf{h}^2 \rangle}{N_t \langle \mathbf{d}^2 \rangle} \right] \quad (15)$$

where  $\langle \mathbf{d}^2 \rangle = \frac{1}{N_t} \sum_{n=0}^{N_t-1} d_n^2$  and  $\langle \mathbf{h}^2 \rangle = \frac{1}{m_b} \sum_{m=0}^{m_b-1} h_m^2$  and we have dropped the constant offset. This log-likelihood function can be evaluated efficiently, and we can use standard optimization algorithms to find its maximum, motivating

$$\text{Strategy 3: } \{\omega_0, \gamma\} = \arg \max_{\omega'_0, \gamma'} L(\omega'_0, \gamma' | \mathbf{d}). \quad (16)$$

Note that in general, finding the global maximum of the log-likelihood function is non-trivial as it is non-convex, tends to become sharply peaked, especially for large data sets, and may have many local extrema, necessitating global search techniques. However, for our two-parameter case, finding the global optimum over reasonable ranges for  $\omega$  and  $\gamma$  proved straightforward using either standard quasi-Newton or even Nelder-Mead Simplex optimization. For more complex functions a density estimator such as particle filters (sequential Monte Carlo methods) or kernel density estimators may be used, which also enable effective determination of the maximum.

## IV. EVALUATION AND COMPARISON OF ESTIMATION STRATEGIES

We now compare the three strategies introduced in the previous section for ensemble and single-shot measurements and also discuss the uncertainty in the estimated parameters and show how Strategy 3 enables the estimation of additional initialisation and measurement parameters. For this we use 10 systems with different values for  $\omega$  and  $\gamma$ , given in Table I, and collect measurement data from simulations with the relevant noise models. For each system the signal was sampled uniformly at  $N_t = 100$  time points  $t_k \in [0, 30]$ . We assume that we have some order of magnitude estimate of the system frequency  $\bar{\omega}$  based on the physical properties of the system, giving us a range for the values of  $\omega$ . Without loss of generality we can express both  $\omega$  and  $\gamma$  in units of  $\bar{\omega}$ . Accordingly all times quoted in the following will be in units of  $\bar{\omega}^{-1}$ . In our simulations we choose  $\omega \in [0.2, 2]$  and  $\gamma \in [0.05, 0.4]$  in units of  $\bar{\omega}$ .

To calculate an average relative error for the parameter estimates,  $N_s = 1000$  runs were performed for each system and noise level and the error computed as

$$e(\omega) = \frac{1}{N_s} \sum_{n=1}^{N_s} \omega^{-1} |\omega_{\text{est}}^{(n)} - \omega| \quad (17a)$$

$$e(\gamma) = \frac{1}{N_s} \sum_{n=1}^{N_s} \gamma^{-1} |\gamma_{\text{est}}^{(n)} - \gamma| \quad (17b)$$

$\omega$	1.0000	0.9000	0.5003	0.7304	1.2161	1.6211	0.2218	1.5195	0.7551	0.8029
$\gamma$	0.1000	0.1000	0.1243	0.1875	0.2031	0.0993	0.1234	0.0751	0.0533	0.1921

TABLE I: Model parameters for 10 models compared below (in units of  $\bar{\omega}$ ).

where  $\omega$  and  $\gamma$  are the actual parameters of the simulated system and  $\omega_{\text{est}}^{(n)}$  and  $\gamma_{\text{est}}^{(n)}$  are the estimated values for the  $n$ th run.

### A. Ensemble measurements with Gaussian noise

To compare the different estimation strategies for discretely sampled signals with Gaussian noise we simulate the measurement result  $d_k$  at time  $t_k$ . The expected signal  $p(t_k)$  was calculated based on the selected model and Gaussian noise of mean 0 and standard deviation  $\sigma$  added to each value. Fig. 1 (left) shows an example of an ideal measurement signal and simulated data with uniform sampling at times  $t_n = n\Delta t$  with  $\Delta t = 0.3$ .

Fig. 2 compares the errors according to (17) for the three strategies. Strategy 2, probably the most common technique for estimating the frequency and dephasing parameter using the position and width of the peak in the Fourier spectrum, actually gives the least accurate and least precise estimates — the median error of the estimated values is large, as is the spread of the errors for different systems as indicated by the large error bars. Strategy 1 produces slightly improved estimates, but parameter estimates based on Strategy 3 are significantly better. The results are similar for  $\omega$  and  $\gamma$ . Fig. 3 furthermore suggests that Strategies 1 and 2 are *not* unbiased estimators. The mean of the distribution over the estimation runs does not appear to converge to the true value of the parameter even for very lowest noise level and 1000 runs. Strategy 3, however, appears to be an unbiased Gaussian estimator.

One interesting feature of Strategies 1 and 2 is that the median estimation errors appear to be almost constant over the range of noise levels considered, while for Strategy 3 the error increases with increasing noise level, as one would expect. A probable reason for this is that the uncertainties in the position, and indirectly the width, of the peaks in the Fourier spectrum primarily depend on the length of the signal  $T$ . Specifically, for a fixed number of samples, [9] found that the uncertainty in the parameter estimates was mainly proportional to  $1/\sqrt{T}$ . This would explain why the accuracy of the estimates obtained from the Fourier-based strategies appears roughly constant as the signal length and number of samples were both fixed in our simulated experiments ( $T = 30$ ,  $N_t = 100$ ). So it might be argued that the Fourier-based strategies are less sensitive to noise. However, it is important to notice that even for noise with  $\sigma = 0.1$ , Strategy 3 still outperforms the other strategies in all cases.

Furthermore, accurately and precisely estimating loca-

tion and width of a peak in the Fourier spectrum for a relatively short, noisy signal can be challenging, as illustrated by the power spectrum examples in Fig. 4. The blue bars show the  $|F(k)|^2$ , where  $F(k)$  is the discrete Fourier transform of the measured discrete signal

$$F(k) = \sum_{n=1}^{N_t} d'_n e^{-2\pi i(k-1)(n-1)/N_t}, \quad 1 \leq k \leq N_t, \quad (18)$$

computed using the Fast Fourier Transform (FFT), after centering and rescaling,  $\mathbf{d}' = (\mathbf{d} - \bar{d})/d_{\text{max}}$  with  $\bar{d} = \frac{1}{N_t} \sum_{n=1}^{N_t} d_n$  and  $d_{\text{max}} = \max |d_n - \bar{d}|$ . The red curve is an approximation to the continuous Fourier transform

$$F(\omega) = \int_{-\infty}^{\infty} f(t) e^{-i\omega t} dt \approx \sum_{n=1}^{N_t} d'_n e^{i\omega t_n} \frac{1}{2} (\Delta t_n + \Delta t_{n-1}) \quad (19)$$

where the integral has been approximated using the trapezoidal rule with  $\Delta t_n = t_{n+1} - t_n = T/N_t$  for  $n = 1, \dots, N_t - 1$  and  $\Delta t_0 = \Delta t_{N_t} = 0$ . The left figure shows a “good” power spectrum for a low-noise input signal. Even in this case the frequency resolution is limited but the peak has a more or less Lorentzian shape and the width is well defined. However, for increasing noise the peak can become increasingly distorted (center) and for very noisy signals it may even become split (right) making width estimation difficult and assumptions about kurtosis and skewness are no longer valid.

A further advantage of Strategy 3 is that it also provides direct estimates for the noise variance [11]

$$\sigma = \frac{1}{N_t - m_b - 2} (N_t \langle \mathbf{d}^2 \rangle - m_b \langle \mathbf{h}^2 \rangle) \quad (20)$$

and Fig. 5 shows that the estimates are very accurate across the board.

### B. Single-system measurements

To assess if there are significant differences in the performance of different estimation strategies in the presence of projection noise, we repeat the analysis in the previous subsection for the same 10 model systems, sampled over the same time interval  $[0, 30]$ , but with various levels of projection noise added instead of Gaussian noise. Fig. 1 (right) shows an example of an ideal measurement signal and simulated data. Fig. 6 shows the relative errors for the different estimation strategies for the same model systems but subject to (simulated) projection noise. Strategy 3 again performs significantly better than the other strategies. Fig. 7 shows that the likelihood of the estimates increases with increasing number of repetitions

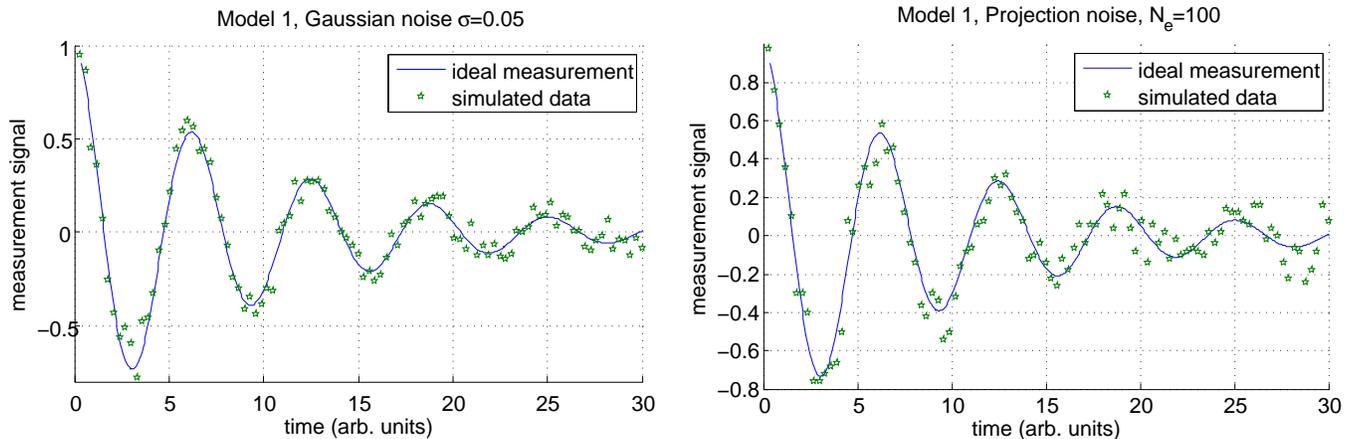


FIG. 1: Example of ideal measurement signal and data from simulated experiments with Gaussian noise ( $\sigma = 0.05$ , left) and projection noise (each data point is the average of  $N_e = 100$  binary-outcome single shot measurements, right).

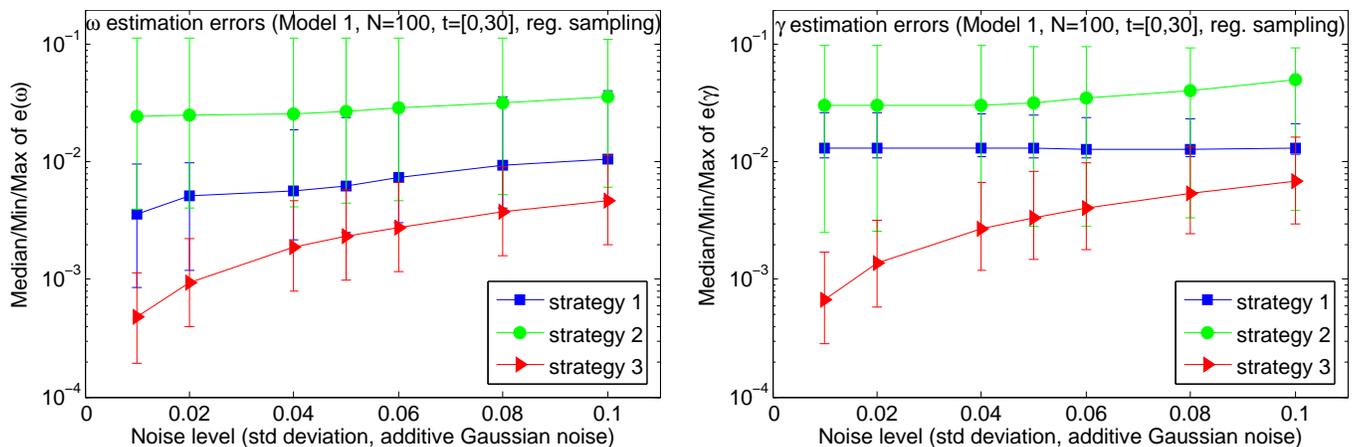


FIG. 2: Minimum, maximum and median of relative error (averaged over 1000 runs for each system and noise level) of  $\omega$  (left) and  $\gamma$  estimates (right) as a function of the magnitude of the Gaussian noise for 10 model systems (Table I).

$N_e$ , as expected. It also shows again that the maximum likelihood for some model systems is consistently higher than for others, as was observed for Gaussian noise.

Fig. 8 shows that even the estimates for the noise variance  $\sigma^2$  obtained automatically with Strategy 3 are very accurate in that the results obtained closely track the theoretical values  $\sigma^2 = 1/N_e$  expected for projection noise.

Overall this shows that although the noise strictly follows a Poisson distribution in this case, we still obtain very good estimates of the noise level for typical values of  $N_e$  using a Gaussian error model in the derivation of the maximum likelihood estimation strategy. So overall Strategy 3 appears to be consistently better than Strategies 1 and 2, independent of the types of measurements and their associated noise for the two-level frequency and dephasing estimation problem.

### C. Uncertainty in parameter estimates

The error statistics are useful for comparing different strategies in terms of both the accuracy (mean or median of error) and precision (spread of errors) of the estimated parameters, and the graphs above show that Strategy 3 outperforms the other strategies on both counts. However, obtaining such statistics requires data from many simulated experiments as well as knowledge of the actual system parameters. In practice, the actual values of the system parameters to be estimated are usually unknown, as otherwise there would be no need to estimate the parameters in the first place, so we cannot use error statistics directly to determine the accuracy and precision of our estimates. However, we can estimate the uncertainty of the parameter estimates, as discussed next.

For the Fourier-based strategies we have already mentioned that the uncertainty in the parameter estimates is mainly determined by the frequency resolution, limited by the sampling rate based on the Nyquist-Shannon

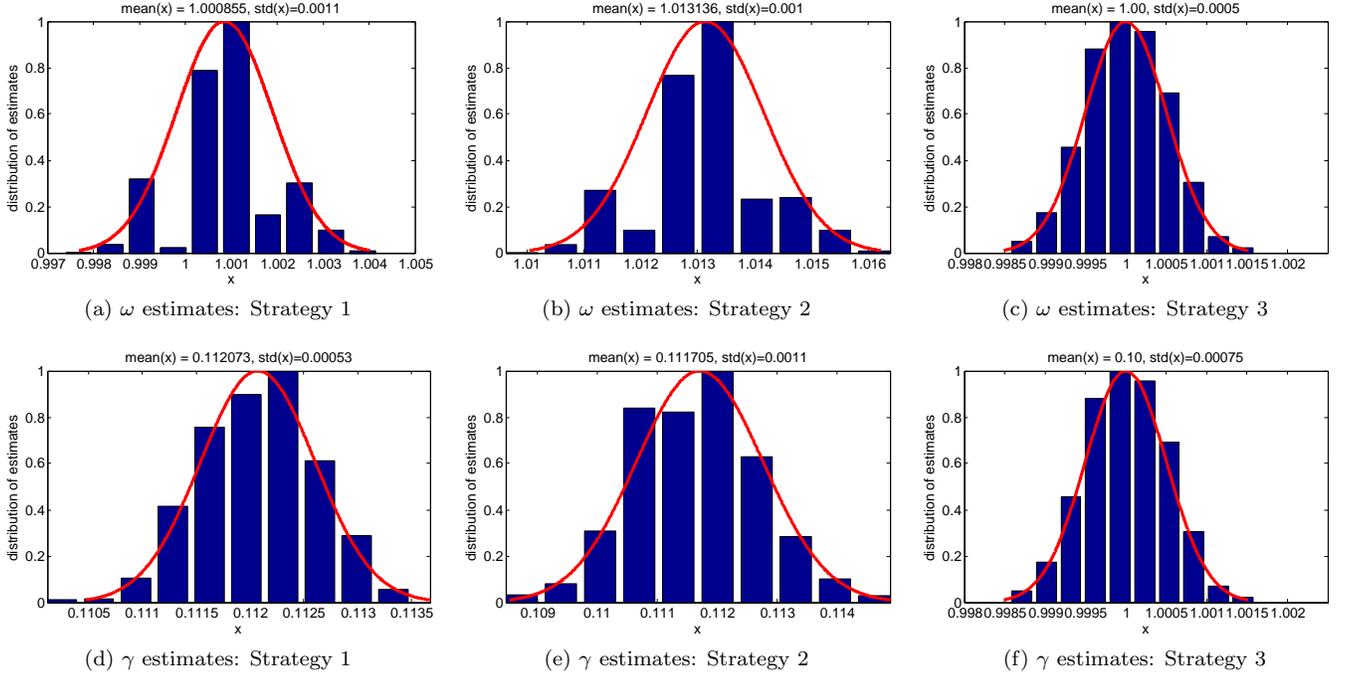


FIG. 3: Distribution of  $\omega$  and  $\gamma$  estimates for 1000 runs for model 1 with 1% Gaussian noise for strategies 1, 2 and 3.

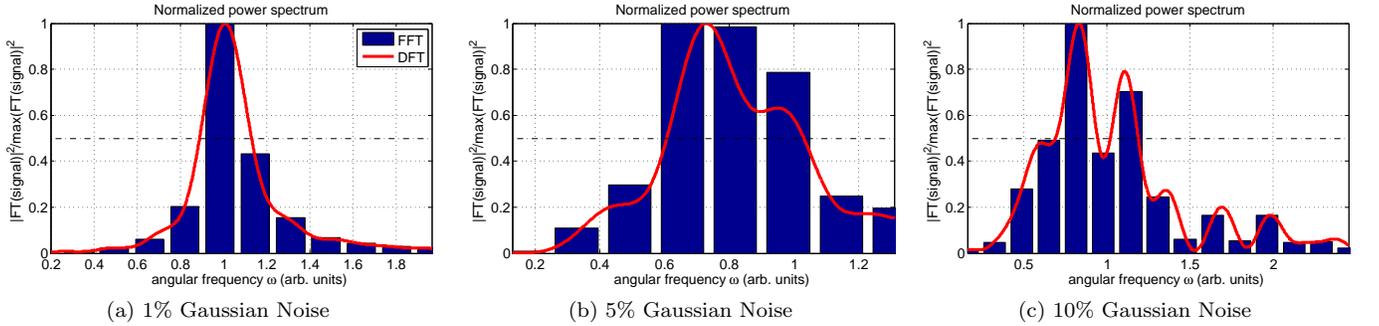


FIG. 4: Limits of Fourier resolution and difficulty in estimating peak width for short, noisy signals.

sampling theorem, which is fixed  $N_t/T$  in our case, and the length of the sampled input signal as the Gabor limit implies as trade-off between time- and band-limits.

For the maximum likelihood estimation we can obtain uncertainty estimates for the parameters from the width of the peak of the likelihood function around the maximum. We use the following simple strategy. Let  $(\omega, \gamma)$  be the parameters for which the log-likelihood assumes its (global) maximum  $L_{\max}$ . To estimate the uncertainty in  $\omega$  we compute the log-likelihood  $L(\omega + \delta\omega, \gamma | \mathbf{d})$  for values  $\delta\omega$  where  $L$  is significantly larger than 0 (implemented by sampling under the assumption that  $L$  is not too far off a peaked distribution). Then we find the range of  $\delta\omega$  for which the actual likelihood

$$\exp(L(\omega + \delta\omega, \gamma | \mathbf{d})) \geq \frac{1}{2} \exp(L_{\max}) \quad (21)$$

to determine the full width at half maximum (FWHM)

$\delta\omega^{\text{FWHM}}$  of the likelihood peak in the  $\omega$  direction. Assuming a roughly Gaussian peak the uncertainty in  $\omega$  is then given by

$$\Delta\omega = 2\sqrt{2\ln(2)} \delta\omega^{\text{FWHM}}, \quad (22)$$

and similarly for  $\gamma$ . Fig. 9 shows the resulting peaks in the likelihood function for a typical experiment together with the FWHM estimates, showing greater uncertainty in the  $\gamma$  estimates.

Fig. 10 show the resulting uncertainties for parameter estimates obtained by Strategy 3 for the ensemble measurements. The uncertainty in the  $\omega$  and  $\gamma$  estimates increases with the noise level, as one would expect, but for some systems the increase is steeper than for others. In particular, the uncertainties are greater for models 4, 5 and 10, for which  $\gamma$  is large, and lowest for model system 9, which has the lowest  $\gamma$  of the 10 models. The

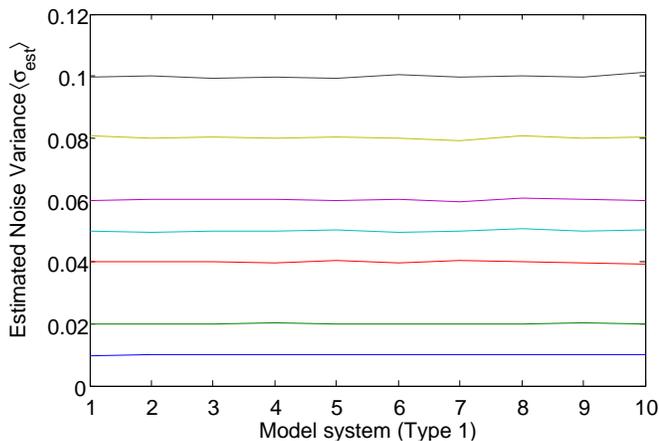


FIG. 5: The estimated noise level  $\sigma$  of the measurement data for 10 model systems of type 1 obtained from Strategy 3 closely track the actual noise levels of the simulated data (0.01, 0.02, 0.04, 0.05, 0.6, 0.8, 1.0).

higher uncertainties coincide with dips in the maximum of the log-likelihood in Fig. 12. Although there is some variation in the value of the maximum log-likelihood between different runs for the same model and error level, the differences between the average of the maximum log-likelihood over many runs for model systems 1 and 5 are several standard deviations, e.g.  $\max \log L \approx 47.9 \pm 3.2$  (for model 1,  $\sigma = 0.1$ ) vs  $34.3 \pm 3.3$  (model 5,  $\sigma = 0.1$ ). This is consistent with the peak of the (log-)likelihood being lower and broader for model 5, resulting in higher uncertainty, and narrower and higher for model 1, resulting in less uncertainty. Fig. 11 shows that the uncertainties for parameter estimates behave the same ways for single shot measurements as a function of the projection noise level  $1/\sqrt{N_e}$ .

This suggests that given the same amount of data the uncertainty of our estimates increases slightly with larger dephasing rate. A probable explanation for this is that the signal decays faster for higher dephasing and thus the signal-to-noise ratio of the later time samples is reduced. For higher dephasing rates the results could likely be improved by adding more samples for shorter times or introducing weights and reducing the latter for measurements obtained for longer times.

#### D. Estimating initialisation and measurement parameters

According to (14) Strategy 3 also provides information about the initialization and measurement procedure via estimates for the parameters  $\alpha_1$  and  $\alpha_2$ . For this model we obtain

$$\alpha_1 \pm \alpha_2 = \cos \theta_I \cos \theta_M \pm \sin \theta_I \sin \theta_M = \cos(\theta_I \mp \theta_M)$$

and thus

$$\theta_I = \frac{1}{2}[\arccos(\alpha_1 - \alpha_2) + \arccos(\alpha_1 + \alpha_2)], \quad (23a)$$

$$\theta_M = \frac{1}{2}[\arccos(\alpha_1 - \alpha_2) - \arccos(\alpha_1 + \alpha_2)]. \quad (23b)$$

Fig. 13 shows the estimates for the parameters  $\alpha_1$  and  $\alpha_2$  with error bars indicating uncertainty for the ensemble measurements. From the plot it is evident that  $\alpha_1 \rightarrow 0$  and  $\alpha_2 \rightarrow 1$  for  $\sigma \rightarrow 0$ , which suggests  $\theta_I = \theta_M = \frac{\pi}{2}$ , which agrees with the values of the initialization and measurement angles used in the simulated experiments. Fig. 14 shows that the same is true in the case of projection noise for single shot measurements. The associated estimates for the parameters  $\alpha_1$  and  $\alpha_2$  in converge to  $\alpha_1 \rightarrow 0$  and  $\alpha_2 \rightarrow 1$  for  $N_e \rightarrow \infty$ , which suggests  $\theta_I = \theta_M = \frac{\pi}{2}$ , which also agrees with the values of the initialization and measurement angles used in the simulated experiments. Similar behaviour is observed for other choice of the initialization and measurement angles.

#### E. Fisher Information and Cramer Rao Bound

The Fisher information matrix  $I = (I_{ij})$  is defined by

$$I_{ij} = E \left[ \frac{\partial L}{\partial \theta_i} \frac{\partial L}{\partial \theta_j} \right] = \int \frac{\partial L}{\partial \theta_i} \frac{\partial L}{\partial \theta_j} f(x|\theta) dx = -E \left[ \frac{\partial^2 L}{\partial \theta_i \partial \theta_j} \right] \quad (24)$$

where  $L(x, \theta)$  is the log-likelihood of the measurement outcome  $x$  given  $\theta$  and  $E$  the expectation w.r.t.  $x$ . If the estimator  $T$  for the parameters  $\theta$  is unbiased, i.e. the mean square error of  $T$  is

$$\text{MSE}(T) = \text{Bias}(T)^2 + \text{Var}(T) = \text{Var}(T) \quad (25)$$

where  $\text{Var}(T)$  is the covariance matrix of the estimator, then the matrix  $C = \text{Var}(T) - I^{-1}$  must be positive semi-definite and  $\|C\|$  gives an estimate of how close we are to the Cramer-Rao limit.

Applied to our case,  $\theta = (\omega, \gamma)$  and

$$L(\mathbf{x}|\theta) = -N \log(\sqrt{2\pi}\sigma) - \frac{1}{2\sigma^2} \sum_{n=1}^N [p(\theta, t_n) - x_n]^2$$

with  $p(\theta, t) = e^{-\theta_2 t} \cos(\theta_1 t)$ , we get

$$\frac{\partial L}{\partial \theta_1} = -\frac{1}{\sigma^2} \sum_{n=1}^N [p(\theta, t_n) - x_n] \frac{\partial p(\theta, t_n)}{\partial \theta_1} \quad (26a)$$

$$\frac{\partial L}{\partial \theta_2} = -\frac{1}{\sigma^2} \sum_{n=1}^N [p(\theta, t_n) - x_n] \frac{\partial p(\theta, t_n)}{\partial \theta_2} \quad (26b)$$

$$(26c)$$

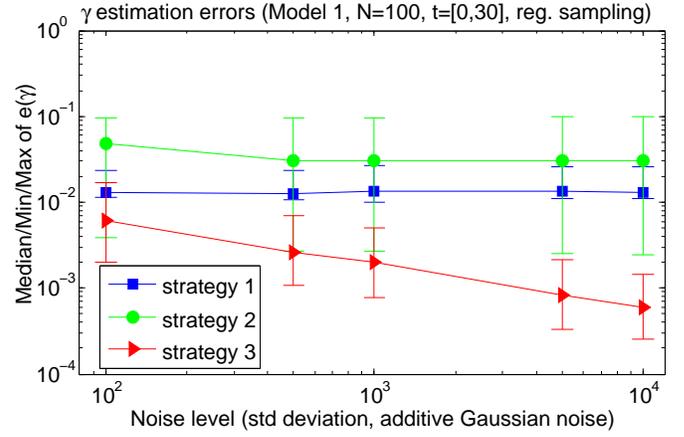
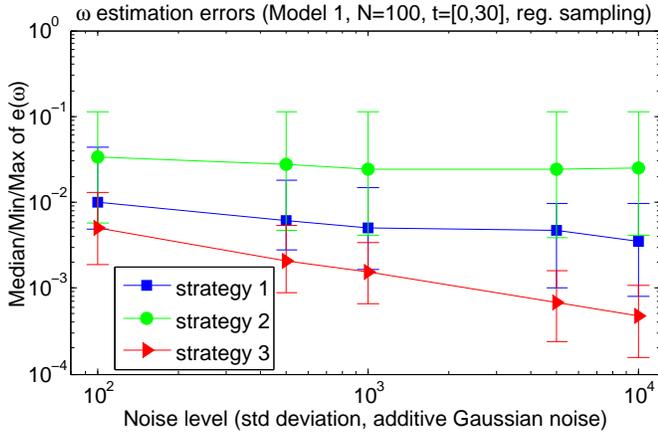


FIG. 6: Minimum, maximum and median of relative error of  $\omega$  (left) and  $\gamma$  estimates (right) as a function of the number of single-shot measurement repetitions per data point,  $N_e$ , for 10 model systems (Table I).

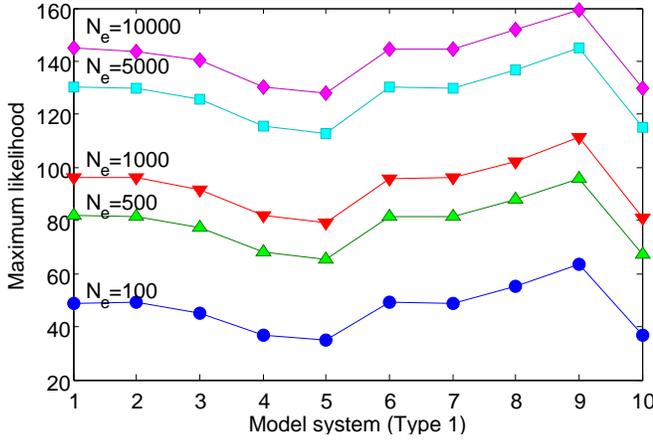


FIG. 7: Maximum likelihood for 10 model systems, averaged over 100 runs each, obtained from Strategy 3.

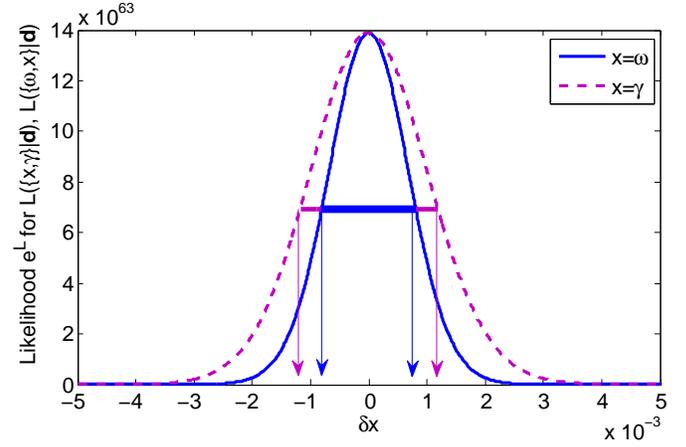


FIG. 9: Estimation of width of likelihood peak with regard to  $\omega$  and  $\gamma$ .

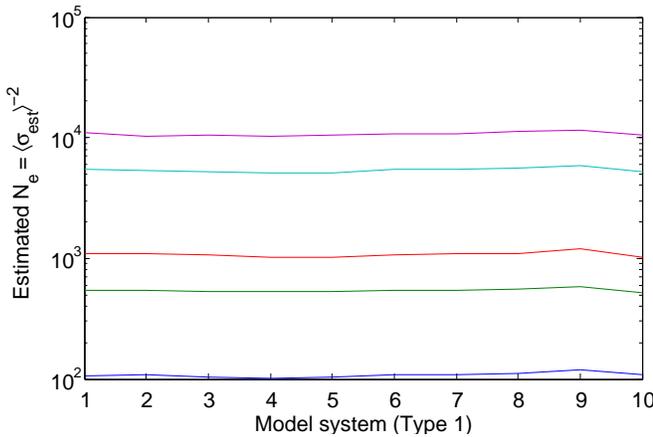


FIG. 8: Estimated  $N_e = \langle \sigma_{est} \rangle^{-2}$  for single shot measurements for 10 model systems, averaged over 100 runs each, obtained from Strategy 3. The  $N_e$  estimates closely track the actual number of repetitions of the single shot measurements for the simulated data (100, 500, 1000, 5000, 10000).

and

$$\frac{\partial p(\theta, t_n)}{\partial \theta_1} = -t_n e^{-\theta_2 t_n} \sin(\theta_1 t_n) =: \alpha_n \quad (27a)$$

$$\frac{\partial p(\theta, t_n)}{\partial \theta_2} = -t_n e^{-\theta_2 t_n} \cos(\theta_1 t_n) =: \beta_n. \quad (27b)$$

$$(27c)$$

Setting  $p_n = p(\theta, t_n)$  we have

$$\begin{aligned} \frac{\partial L}{\partial \theta_1} \frac{\partial L}{\partial \theta_2} &= \frac{1}{\sigma^4} \left( \sum_{n=1}^N \alpha_n p_n - \alpha_n x_n \right) \left( \sum_{n=1}^N \beta_n p_n - \beta_n x_n \right) \\ &= \sigma^{-4} \left( AB - \sum_{n=1}^N c_n x_n + \sum_{m,n=1}^N \alpha_m \beta_n x_m x_n \right) \end{aligned}$$

with  $A = \sum_n \alpha_n p_n$  and  $B = \sum_n \beta_n p_n$ ,  $c_n = \alpha_n B + \beta_n A$ . Similarly for the other partial derivatives. Noting

$$\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} x_n \exp \left[ -\frac{|p_n - x_n|^2}{2\sigma^2} \right] dx_n = p_n \quad (28)$$

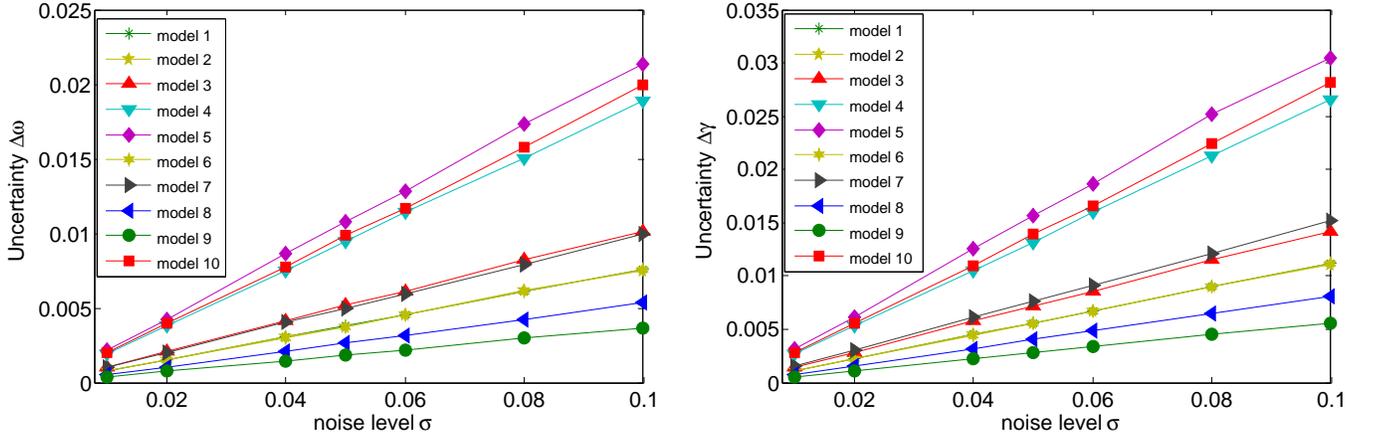


FIG. 10: Uncertainties of  $\omega$  (left) and  $\gamma$  estimates (right) for 10 model systems (Table I) as a function of Gaussian noise level.

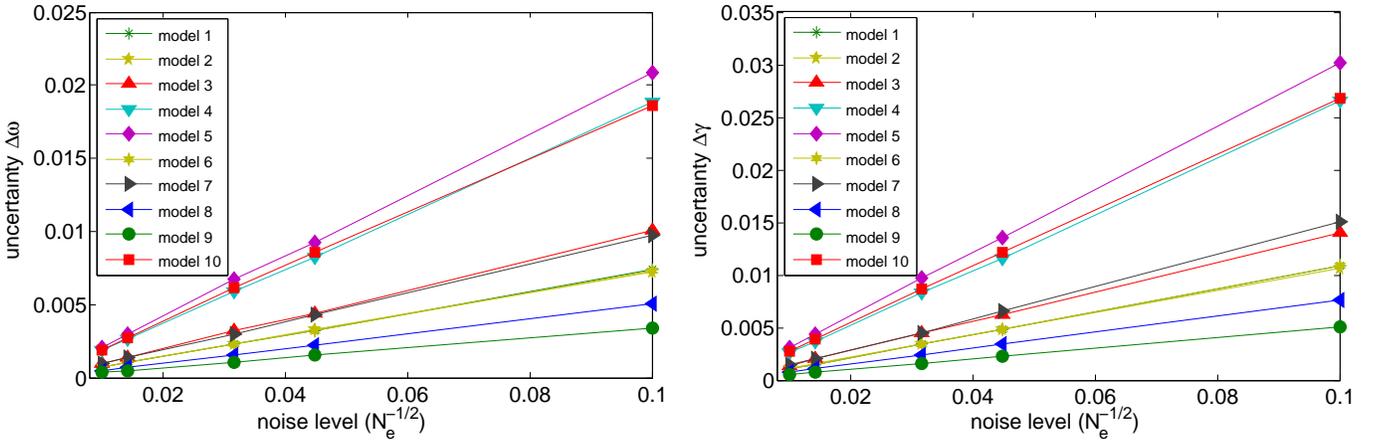


FIG. 11: Uncertainties of  $\omega$  (left) and  $\gamma$  estimates (right) for 10 model systems as a function of projection noise level.

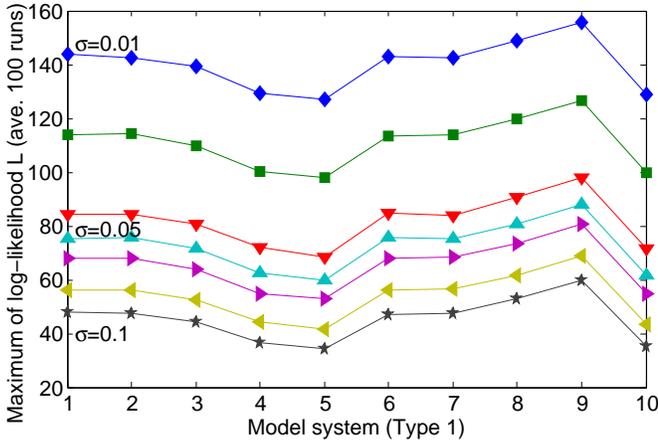


FIG. 12: Maximum of log-likelihood (Strategy 3) for 10 model systems (Table I) for different noise levels.

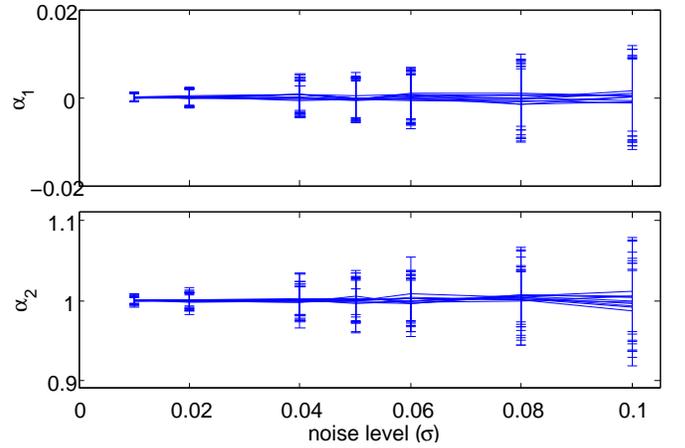


FIG. 13: Estimates for parameters  $\alpha_1$  and  $\alpha_2$  including uncertainty as a function of the noise level  $\sigma$  for 10 model systems.

and assuming the estimator is unbiased, we finally obtain

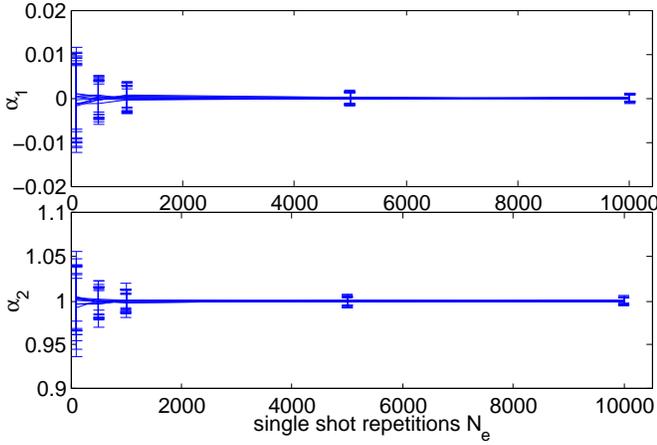


FIG. 14: Estimates for parameters  $\alpha_1$  and  $\alpha_2$  as a function of the number of single shot repetitions  $N_e$  for 10 model systems (Type 1, averaged over 100 runs each).

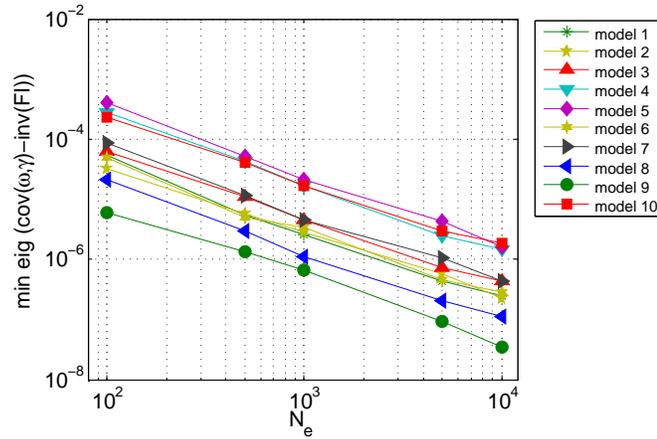


FIG. 15: Plot of the minimum eigenvalue of the covariance matrix of the estimator minus the inverse Fisher information for various models as a function of  $N_e$ .

the entries of the Fisher information matrix

$$\begin{aligned}
 I_{11} &= \sigma^{-4} \left( A^2 - 2A \sum_n \alpha_n p_n + \sum_{m,n} \alpha_m \alpha_n p_m p_n \right) \\
 I_{12} &= \sigma^{-4} \left( AB - \sum_n c_n p_n + \sum_{m,n} \alpha_m \beta_n p_m p_n \right) \\
 I_{22} &= \sigma^{-4} \left( B^2 - 2B \sum_n \beta_n p_n + \sum_{m,n} \beta_m \beta_n p_m p_n \right).
 \end{aligned} \tag{29}$$

While our simulations suggest that the estimators based on Strategies 1 and 2 are not unbiased, Strategy 3 appears to be unbiased. Fig. 15, showing the smallest eigenvalue of the matrix  $C$  for our various test systems subject to projection noise, suggests that we indeed approach the Cramer-Rao bound for  $N_e \rightarrow \infty$  and  $\sigma = N_e^{-1/2}$ .

## V. ADAPTIVE ESTIMATION STRATEGIES

We may find that the accuracy or precision of the parameters obtained from an initial data set is not sufficient and we would like to improve it by acquiring additional data. Adaptive refinement strategies depend on the experimental set-up and system and a detailed analysis of specific strategies is beyond the scope of this paper. However, we shall briefly discuss general approaches for iterative refinement for the Fourier and Bayesian estimation approaches and compare these for a few examples.

In some settings an entire measurement trace is obtained in a single experimental run and we are only able to sample the signal at regular time intervals restricted by the experimental equipment available. In this case the only options available to us are extending the signal length (keeping sampling density or number of sample points constant) or repeating the experiment. If Fourier-based estimation strategies are used, the only way to really improve the resolution of the Fourier spectrum, and thus the accuracy and precision of our estimates, is by increasing the signal length. However, for a decaying signal the signal-to-noise ratio progressively deteriorates until the signal vanishes, limiting the accuracy and precision that are attainable. This is illustrated in Fig. 16(left), which shows the (normalized) power spectrum for 1 to 1000 repetitions of the experiment for model parameters 4, assuming each individual measurement trace is subject to Gaussian noise at  $\sigma = 0.1$  and the signals are averaged. For a single run of the experiment with this level of noise, the peak is distorted but the power spectrum quickly converges. The corresponding estimates for  $\omega$  and  $\gamma$  (Fig. 16, center and right) also converge but not to the true value. For Strategy 2 the  $\omega$  and  $\gamma$  estimates are inaccurate. The optimization step in Strategy 1 appears to improve the accuracy of the  $\omega$  estimates but the  $\gamma$  estimates are still inaccurate. Strategy 3 does not suffer from these limitations and averaging multiple short traces should increase the accuracy of our estimates. Indeed the figure shows that this appears to be the case: both the  $\omega$  and  $\gamma$  estimates converge to the true values.

This shows that Strategy 3 allows adaptive refinement even if all we are able to do is to repeat the experiment multiple times and average the measurement traces. However, in some situations we have more freedom. For Rabi spectroscopy, for example, each data point, corresponding to a measurement at a particular time  $t_n$ , may be obtained in a separate experiment, and we may be free to choose the measurement times  $t_n$  flexibly. In this case, having obtained  $N_t$  measurements we can try to choose the next measurement time  $t_{N_t+1}$  such that it optimizes the amount of information we gain from the experiment. We could ask, for example, considering all possible outcomes of a measurement at time  $t$  and their probability based on our current knowledge, at what time should we measure next to achieve the largest reduction in the uncertainty of our estimates. However, this would require calculating the uncertainty of the parameters (e.g., by es-

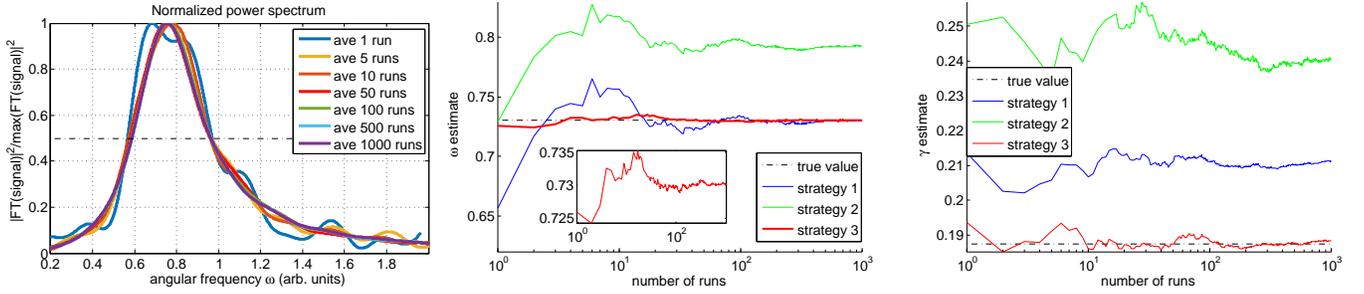


FIG. 16: Iterative refinement by averaging of signal traces: power spectra (left),  $\omega$  estimates (center) and  $\gamma$  estimates (right).

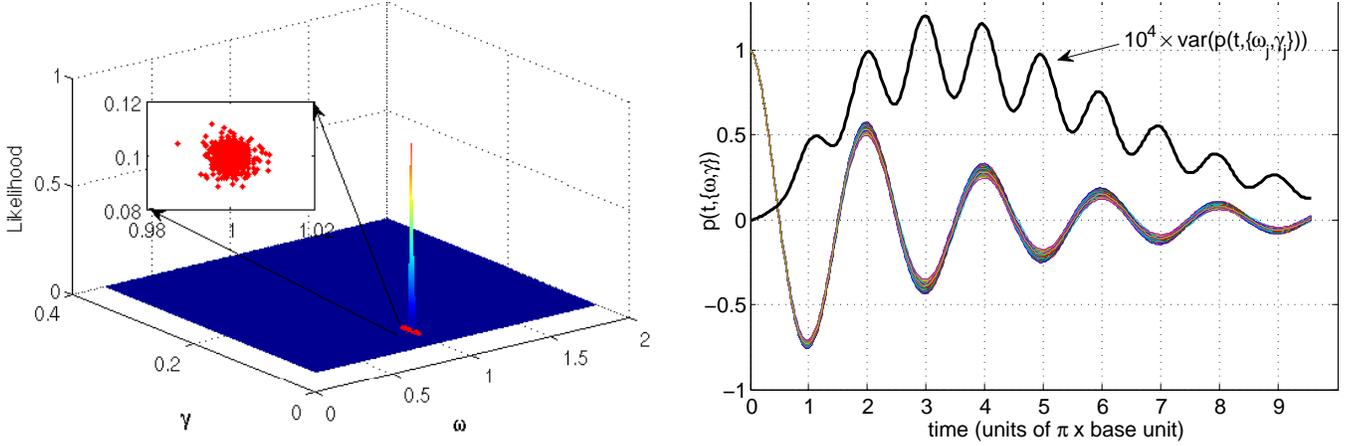


FIG. 17: Prior likelihood after 25 time samples ( $t_n = 1.2n$ ) for model 1 (left) with  $\{\omega_j, \gamma_j\}$  samples (red dots) and corresponding predicted measurement traces  $p_j(t) = p(t, \{\omega_j, \gamma_j\})$  and variance of  $p_j(t)$  as function of  $t$  (right).

timating the width of the likelihood peaks) for all possible measurement times and outcomes. Given the continuum of measurement outcomes and measurement times, this is generally too expensive to calculate.

We therefore consider a simpler heuristic. We generate a number of guesses  $\{(\omega_j, \gamma_j), j = 1, \dots, J\}$  for the parameters based on the current likelihood distribution for the parameters. We then calculate the measurement signal  $p(t, \{\omega_j, \gamma_j\})$  for a set of discrete times and select the next measurement time where the variance of the predicted measurement results is greatest. The idea behind this strategy is that a larger spread in the predicted results indicates greater uncertainty, and a measurement at such a time should result in a greater reduction of the uncertainty. We illustrate this strategy in Fig. 17. The variance of the predicted traces  $p_j(t) = p(t, \{\omega_j, \gamma_j\})$  exhibits oscillations at about twice the frequency of the signal, being largest around the minima and maxima of the oscillatory signal but due to the damping of the signal there is an overall envelope and a global maximum around 3 in units of  $\pi\bar{\omega}^{-1}$ . To avoid repeated sampling at the same time it is desirable to introduce a degree of randomness, e.g., by selecting the next measurement time based on the maximum of the variance of  $p_j(t_s)$  sampled over a discrete set of times  $t_s$ , such as a non-uniform low-discrepancy sampling of the time interval

$[0, T]$ . Furthermore, in practice it may be rather inefficient to recalculate the variance of the traces after a single measurement. Instead, it we shall acquire an initial set of  $N_0$  data points and then select the next  $N_1$  measurement times to coincide with peaks in the variance of the traces where we allow  $N_1$  to vary depending on the number of peaks. In Fig. 17, for example, there are eight local peaks and we would choose the next eight measurement times to coincide with these maxima and iterate the process.

An even simpler way of iterative refinement is via low-discrepancy (ld) time sampling, a generalization of uniform sampling that lends itself to easy iterative refinement. The basic idea of ld sequences is to ensure the largest gap between samples is asymptotically optimal, while there is little uniformity in the sampling points to avoid aliasing effects (see blue noise criterion). In this case the initial measurement times are chosen to be the first  $N_0$  elements in a low-discrepancy quasi-random sequence such as the Hammersley sequence [22], and in each subsequent iteration the next  $N_i$  elements of the sequence are used. The number of initial measurements  $N_0$  and subsequent measurements per iteration  $N_i$  are completely flexible, the elements of the sequence can be scaled to uniformly cover any desired time interval, and we can perform as many iterations as desired. Fig. 18 shows

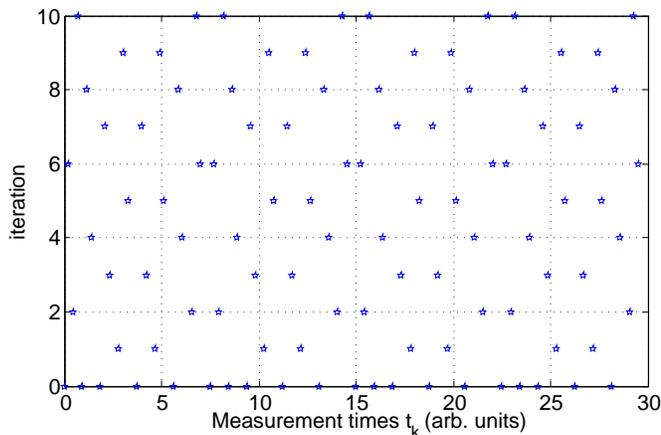


FIG. 18: Selection of measurement times for iterative low-discrepancy sampling. The new measurement times in each iteration as chosen such as to fill in the largest existing gaps.

the measurement times as a function of the iteration as determined by the Hammersley sequence with  $N_0 = 20$  and  $N_i = 8$  for 10 iterations and total sampling times  $T = 30$ , showing that uniform coverage of the sampling interval is maintained. For a fixed number of measurements  $N_t = 100$  we verified that there was no significant difference in the errors and uncertainties of the parameter estimates between low-discrepancy and uniform sampling for the cases considered above. Furthermore, iterative refinement based on ld-sampling performed very well. Fig. 19 for model system 4 with measurements subject to 5% Gaussian noise shows that simple iterative ld-sampling actually outperforms the adaptive refinement strategy based on the trace-variance described above. While this may not be universally the case, and may be due to the variations in the trace variance being relatively small in our example, it shows that simple strategies such as iterative ld-sampling can be highly effectively.

## VI. GENERALIZATION TO OTHER MODELS

So far we have considered a particular model of a dephasing two-level system with dephasing acting in the Hamiltonian basis. However, if control fields are applied, as in a Rabi oscillation experiment for example, then the effective Hamiltonian and the dephasing basis may not coincide. For example, for two-level atoms in a cavity driven resonantly by a laser, the effective Hamiltonian with regard to a suitable rotating frame is  $H = \Omega\sigma_x$ , where  $\Omega$  is the Rabi frequency of the driving field. Assuming the driving field does not alter the dephasing processes, so that we still have  $V = \sqrt{\frac{\gamma}{2}}\sigma_z$ , the resulting measurement trace is given by [21]:

$$p(t) = e^{-\gamma t} \sin \theta_I \sin \theta_M + \Phi_3^x(t) \cos \theta_I \cos \theta_M \quad (30)$$

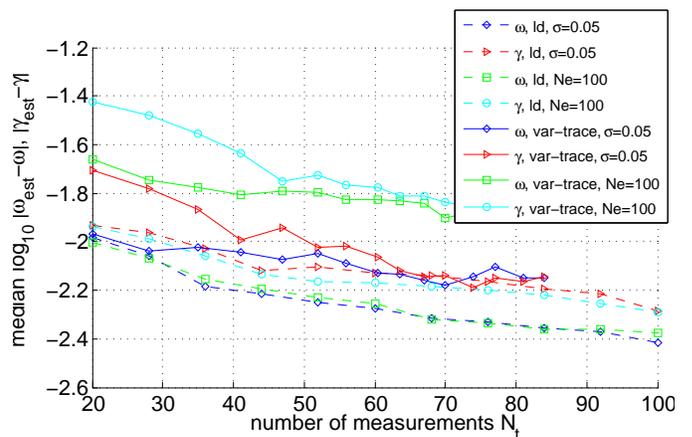


FIG. 19: Median error of  $\omega$  and  $\gamma$  parameter estimates for iterative ld-sampling and adaptive sampling based on trace variance for model system 4 with measurements subject to 5% Gaussian noise and projection noise  $\sigma = N_e^{-1/2}$ , respectively.

where

$$\Phi_3^x(t) = e^{-\frac{\gamma}{2}t} \left[ \cos(\omega t) + \frac{\gamma}{2\omega} \sin(\omega t) \right], \quad (31)$$

$$\omega = \sqrt{\Omega^2 - \frac{\gamma^2}{4}}. \quad (32)$$

If  $\Omega^2 < \gamma^2/4$  then  $\omega$  is purely imaginary and the sine and cosine terms above turn into their respective hyperbolic sine and cosine equivalents. If  $\Omega^2 = \gamma^2/4$ , the expression  $\omega^{-1} \sin(\omega t)$  must be analytically continued.

Due to the more complex nature of the signal, the Fourier estimation strategies are not directly applicable. However, we can very easily adapt Strategy 3. All that is required is a change in the basis functions, setting  $g_1(t) = e^{-\gamma t}$  and  $g_2(t) = \Phi_3^x(t)$ .

Fig. 20 shows the log-likelihood functions for a very sparsely sampled signal with significant projection noise for a system of type (30) for a simulated experiment performed with  $\theta_M = \frac{\pi}{4}$  and  $\theta_I = \frac{\pi}{3}$ . The signal is a damped oscillation, though not a simple damped sinusoid. Strategy 3 easily succeeds in identifying the model parameters and the log-likelihood function has a clearly defined peak. In fact, we are showing the log-likelihood here as the actual likelihood function is so sharply peaked that its internal structure, especially the squeezed nature, is not easy to see.

Finally, Fig. 21 (left) shows the error statistics for the  $\omega$  and  $\gamma$  estimates obtained using Strategy 3 for 10 models of type (30) with the same values for  $\Omega$  and  $\gamma$  as in Table I. We compare two experimental conditions:  $\theta_I = \theta_M = 0$ , which corresponds to maximum visibility of the oscillations and  $\theta_I = \frac{\pi}{3}$ ,  $\theta_M = \frac{\pi}{4}$ , for which the signal is more complex and the visibility of the oscillations is reduced as shown in Fig. 20. The estimation errors are very similar to those for models of type 1. For  $\gamma$  they are effectively identical for both experimental conditions; for  $\Omega$  they are slightly larger in case 2b, as might

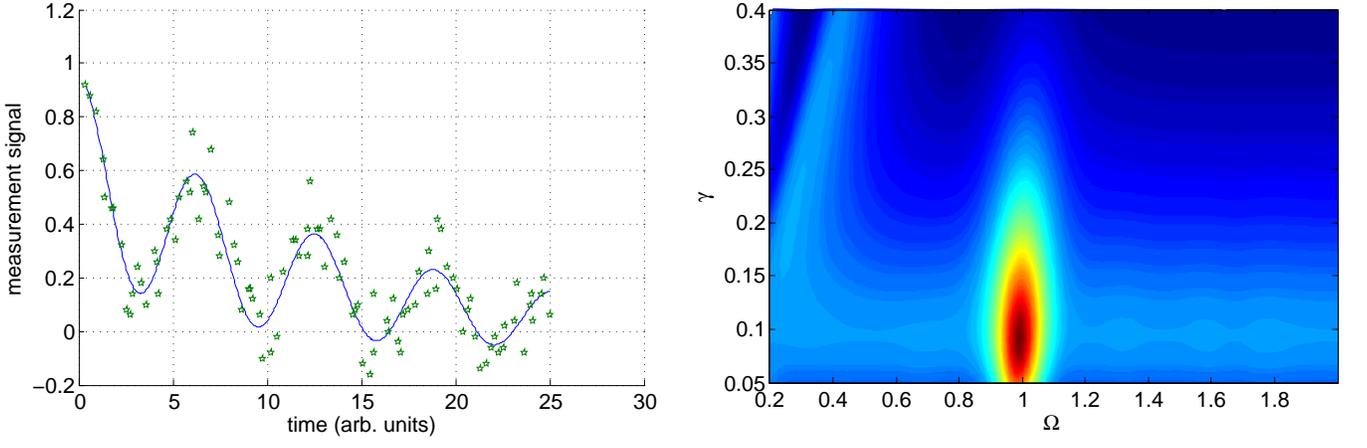


FIG. 20: Ideal signal (blue) and sparsely sampled noisy data (green \*,  $N_t = 100$ ,  $t \in [0, 25]$ ,  $N_e = 100$  single shot experiments per data point) for a system described by Eq. (30) with  $\omega = 1$ ,  $\gamma = 0.1$  (left) and corresponding log-likelihood (right).

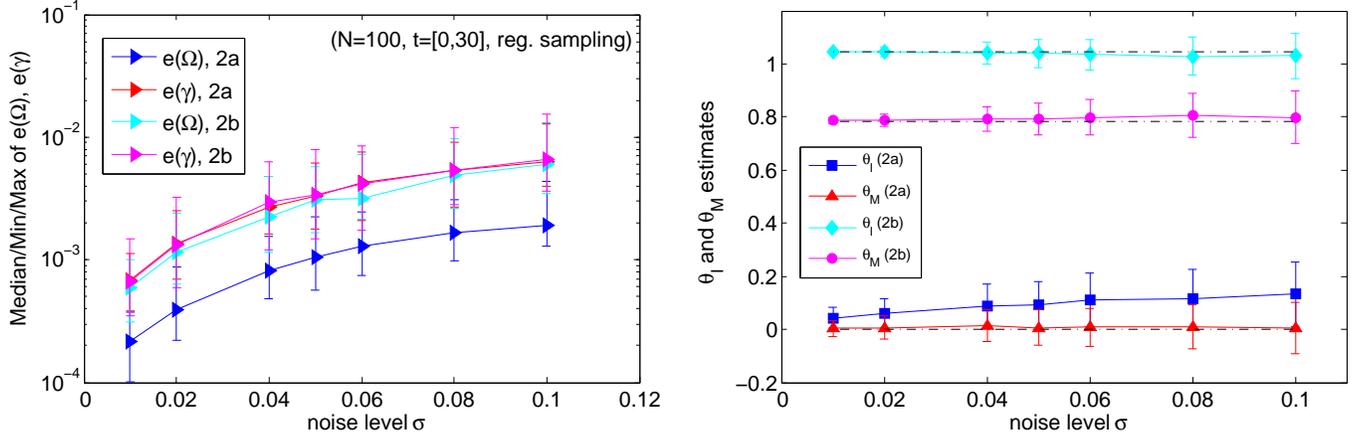


FIG. 21: Minimum, maximum and median of relative error (averaged over 100 runs for each system and noise level) of  $\omega$  and  $\gamma$  estimates as a function of noise level  $\sigma$  (left) and estimates for the initial state and measurement angles  $\theta_I$  and  $\theta_M$  (right) for 10 model systems of type (30) with model parameters given in (Table I) for two experimental conditions:  $\theta_I = \theta_M = 0$  (2a, maximum visibility) and  $\theta_I = \frac{\pi}{3}$ ,  $\theta_M = \frac{\pi}{4}$  (2b).

be expected as the visibility of the oscillations is reduced in this case.

In both cases we also obtain excellent estimates of the noise level  $\sigma$  of the data as well as estimates for the parameters  $\alpha_1$  and  $\alpha_2$ . As before, if the initial state prepared or the precise measurement performed are unknown a priori, as may well be the case for a system that is not yet well characterized, we can use these parameters to derive estimates for  $\theta_I$  and  $\theta_M$ :

$$\theta_I = \frac{1}{2}[\arccos(\alpha_2 - \alpha_1) + \arccos(\alpha_2 + \alpha_1)] \quad (33a)$$

$$\theta_M = \frac{1}{2}[\arccos(\alpha_2 - \alpha_1) - \arccos(\alpha_2 + \alpha_1)] \quad (33b)$$

Fig. 21 (right) shows the estimates derived for the angles  $\theta_I$  and  $\theta_M$  for both experimental conditions. The markers indicate the average of the estimate for all runs and all model systems, the errorbars indicate the standard

deviation of the estimates. The estimates are not as accurate as those for the system parameters, as one would expect as we have marginalized the amplitudes  $\alpha_1$  and  $\alpha_2$  and thus  $\theta_I$  and  $\theta_M$ . However, they are still quite close to the actual values (black dash-dot lines) with the exception of the  $\theta_I$  estimate for case (2a), which is slightly more biased and less accurate – it should be 0, coinciding with the measurement angle  $\theta_M$ .

## VII. CONCLUSIONS

We have investigated the ubiquitous problem of identifying crucial parameters from experimental data for two-level systems subject to decoherence. Comparing different strategies based on the analysis of Fourier spectra as well as Bayesian modelling and maximum likelihood estimation, the latter approach was found to be vastly supe-

rior to commonly used Fourier based strategies in terms of accuracy and precision of the estimates obtained.

Strategies based on simple Fourier analysis is limited by the accuracy with which the positions, heights and widths of the Fourier peaks can be determined. As the spectral resolution is limited by signal length and sampling rate, the accuracy of Fourier-based estimation schemes for short, decaying signals or sparse noisy data is limited. The Bayesian approach is not constrained in this way and yields uncertainties for the system parameters as well as information about the noise in the data.

An additional advantage of the Bayesian estimation is that it does not require a priori knowledge of the initialization or measurement angles  $\theta_I$  and  $\theta_M$ . Rather, the estimation procedure provides values for the coefficients of the basis functions, which are related to the parameters  $\theta_I$  and  $\theta_M$ .

The results are widely applicable to many experimental settings from the analysis for free-induction decay signals for spin systems, e.g., in NMR, MRI and ESR to Rabi spectroscopy for atomic ensembles, trapped ions, quantum dots or Josephson junction devices.

### Acknowledgments

We acknowledge funding from the Ser Cymru National Research Network in Advanced Engineering and Materials. SGS also thanks the Royal Society for funding through a Leverhulme Senior Fellowship grant and the UK Engineering and Physical Sciences Research Council for recent funding. FCL acknowledges funding from the Cardiff University Research Leave Fellowship Scheme.

- 
- [1] D. W. Mc Robbie, E. A. Moore, M. J. Graves, M. R. Prince *MRI from Picture to Proton* (Cambridge University Press, 2007)
  - [2] J. B. Lambert, E. P. Mazzola, *Nuclear Magnetic Resonance Spectroscopy: An introduction to Principles, Applications and Experimental Methods* (Pearson Education, 2004)
  - [3] T. Shinjo, *Nanomagnetism and Spintronics* (Elsevier, London, 2014)
  - [4] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, 2000)
  - [5] J. Nathan Kutz, *Data-Driven Modeling and Scientific Computation* (Oxford University Press, 2013)
  - [6] M. Balci, *Basic 1H and 13C NMR Spectroscopy* (Elsevier, 2005)
  - [7] S. G. Schirmer, A. Kolli and D. K. L. Oi, *Phys. Rev. A* **69** 050306(R) (2004)
  - [8] J. H. Cole *et al.* *Phys. Rev. A* **71** 062312 (2005)
  - [9] J. H. Cole *et al.* *Phys. Rev. A* **73** 062333 (2006)
  - [10] S. J. Devitt, J. H. Cole, L. C. L. Hollenberg, *Phys. Rev. A* **73** 052317 (2006)
  - [11] G. L. Bretthorst, *Bayesian Spectrum Analysis and Parameter Estimation* (Springer, Berlin, 1998)
  - [12] S. G. Schirmer and D. K. L. Oi. *Phys. Rev. A* **80** 022333 (2009)
  - [13] D. K. L. Oi and S. G. Schirmer. *Laser Physics* **20**(5) 1203-1209 (2010)
  - [14] C. E. Granade *et al.*, *New J. Phys.* **14** 103013 (2012)
  - [15] S. Blatt *et al.* *Phys. Rev. A* **80** 052703 (2009)
  - [16] W. Feller, *An Introduction to Probability Theory and Its Applications* (John Wiley & Sons, 1968)
  - [17] R. Blatt and C. F. Roos *Nature Physics* **8** 277-284 (2012)
  - [18] A. M. Kaufman *et al.*, *Science* **345**, 306-309 (2014)
  - [19] V. S. Pribiag *et al.* *Nature Nanotechnology* **8** 170-174 (2013)
  - [20] Y. Nakamura, Yu. A. Pashkin and J. S. Tsai, *Phys. Rev. Lett.* **87** 246601 (2001)
  - [21] Erling Gong, Weiwei Zhou and S. G. Schirmer, *Model discrimination for dephasing two-level systems*. *J. Phys. A*, in press.
  - [22] J. M. Hammersley and D. C. Handscomb *Monte Carlo Methods* (Springer 1964)