University of Strathclyde Science

Quantum System Identification with Bayesian and Maximum Likelihood Estimation

Daniel K. L. Oi
Computational Nonlinear & Quantum Optics
University of Strathclyde

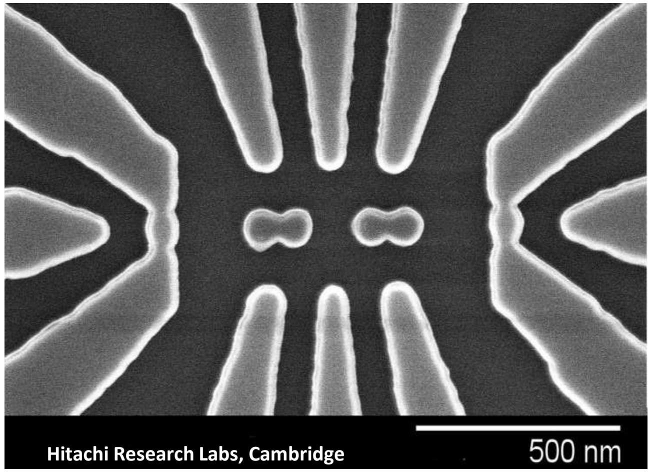
QUISCO Quantum Information Scotland

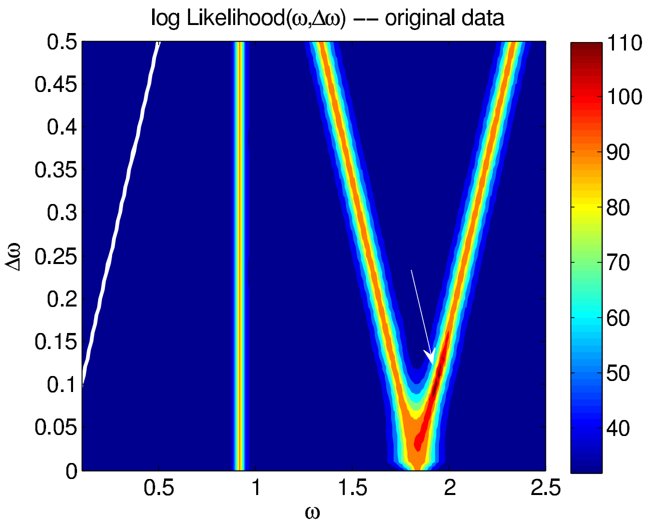
Collaborators:
S. G. Schirmer (Swansea), J. H. Cole (RMIT),
S. J. Devitt (NII), A. D. Greentree (Melbourne),
L. C. L. Hollenberg (Melbourne), A. Kolli (UCL)

SUPA

Talk Outline

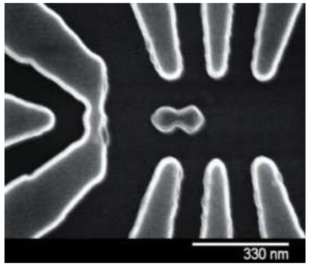
- The system characterisation problem
- Characterisation scenarios
  - » Single Qubit
  - » Qubit Confinement
  - » Two Qubits
  - » Decoherence
- Scalability and Efficiency
  - » Adaptive Estimation
  - » Multiparameter Estimation
- Future directions





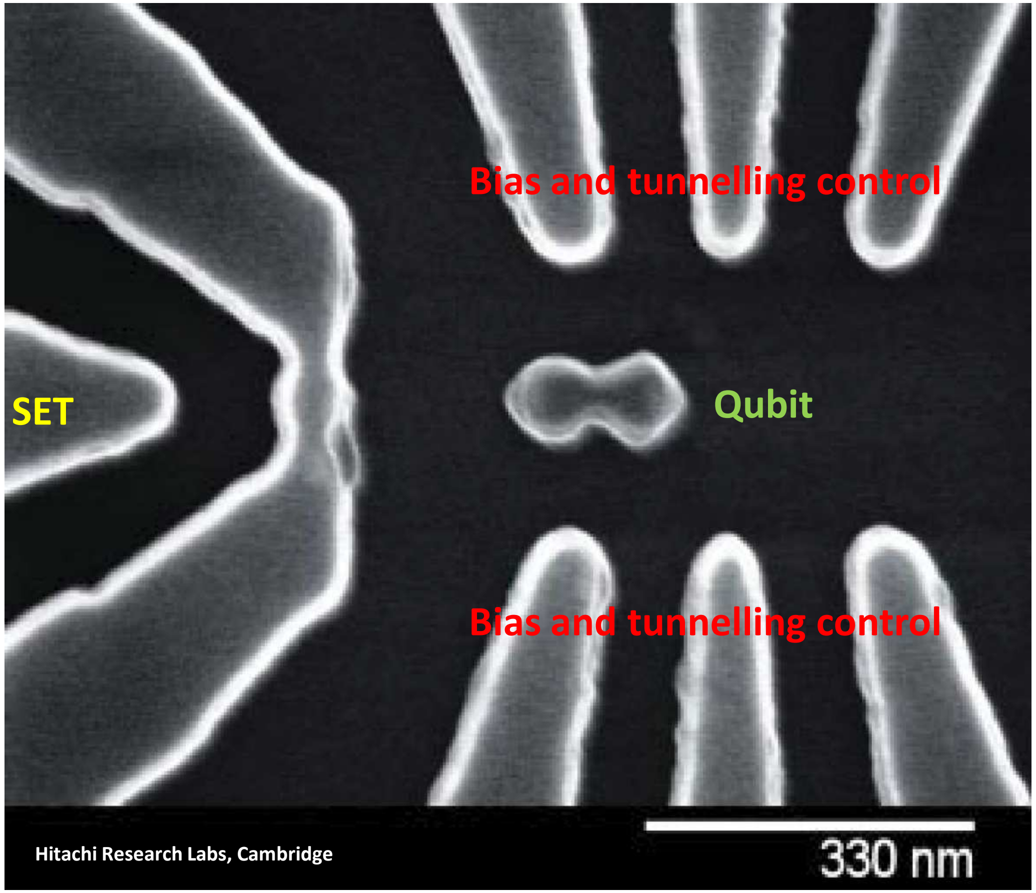
Physics-based mathematical models for quantum devices via experimental system
S. G. Schirmer, D. K. L. Oi and S. J. Devitt, Journal of Physics: Conference Series 107 (2008).

Why characterise a system?

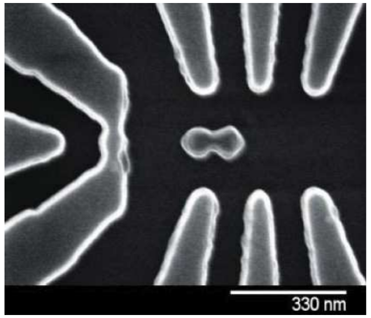


- Quantum Control
- Usually need to know how a system behaves in order to control it. Need to have an accurate dynamical model
- Want to know the Hamiltonian of the system, and how it responds to the application of external control inputs (e.g. fields, voltages, lasers)
- Need to estimate noise and decoherence in the system, do we have a sufficiently coherent system? Is the noise of the expected type? Can we improve the operating regime of our device?
- Is our system actually what we think it is: not all "qubits" are actually qubits
- "System" means not only the physical register, but also the preparation, measurement, and control aspects of the device

Example: Solid State Qubit



What are the problems?



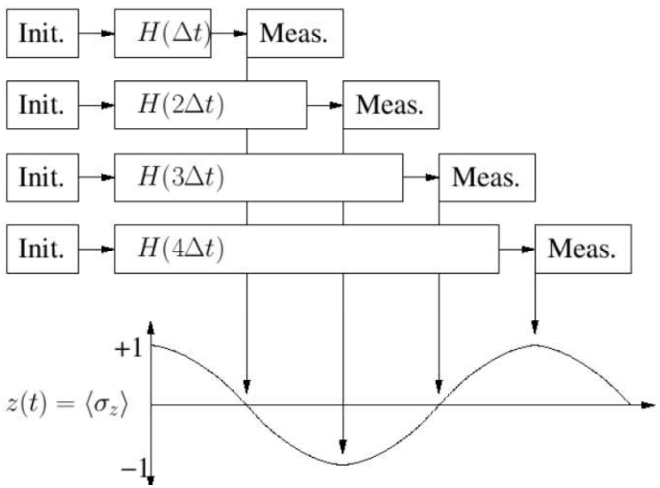
- Some quantum systems are identical, e.g. atoms, ions
- Some are **not**, e.g. quantum dots, SQUIDS
- Even if the qubits are nominally identical, their **environment or control structures may vary** (e.g. diamond NV centres or donors in Si)
- Manufacturing tolerances/variation means that *ab initio* modelling may not be enough to predict with any great accuracy the response
- Diagnostic tools (such as spectroscopy or SEM, AFM of geometry) is intrusive and it would be good to be able to avoid resources other than that ***in situ*** necessary for normal operation
- Often, **readout is also limited** to a fixed (computational) basis or even just a single population, else direct access to the registers may be limited

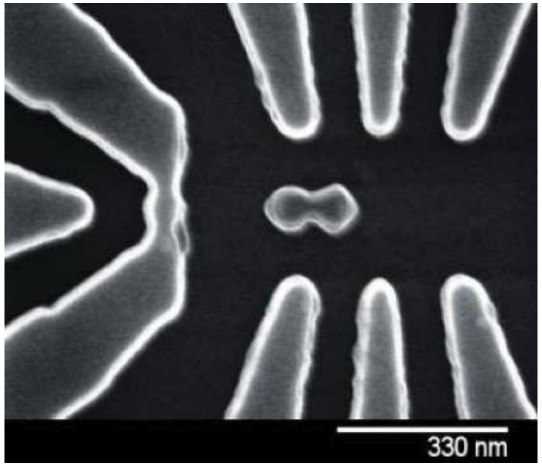
What about quantum process tomography?

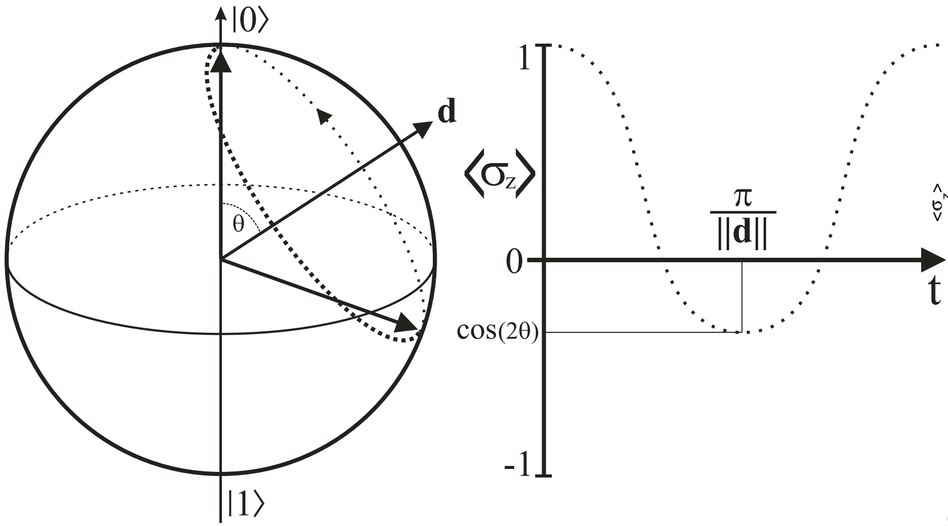
- A. QPT requires **initialisation** in a spanning set of (known) states
- B. Requires **tomographic measurement**, e.g. Different bases or IC-POVMs
- It gives the **discrete** process (CP map), dynamics (e.g. Hamiltonian or Lindblad operators) still needs to be reconstructed from stroboscopic estimation of how the map evolves in time
- Some systems have restricted initial resources, i.e. Initialisation and measurement in a fixed basis (by the physical architecture). Achieving A and B requires (coherent) control over your system (to unitarily rotate initial state and measurement basis), **Catch-22** situation
- Can we **bootstrap** knowledge of system dynamics to enable effective control?

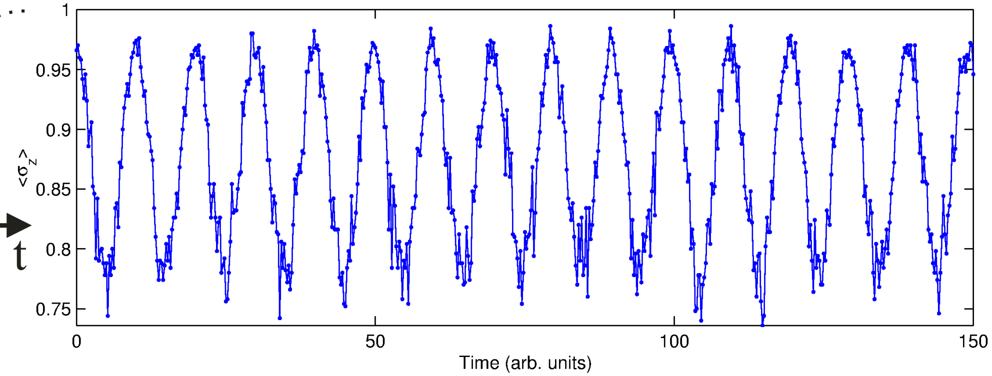
Qubit Example

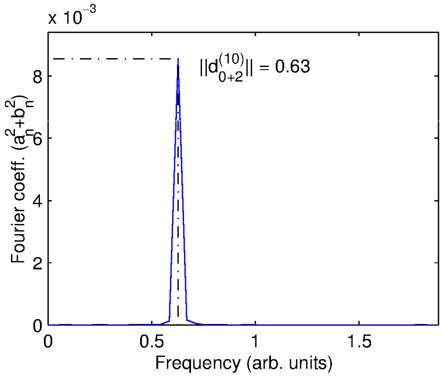
- Simple case: assume a two-level system with completely unknown (constant) Hamiltonian
- Projective measurement, defines computational basis, want to find Hamiltonian in this basis.
- Initialise system in one of the computational basis elements
- Timing control between initialisation and measurement
- From the statistics of the measurement results at different times, can we reconstruct the Hamiltonian?

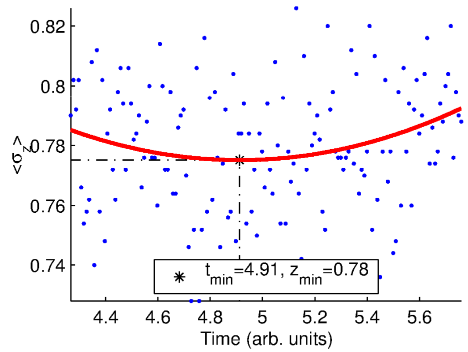


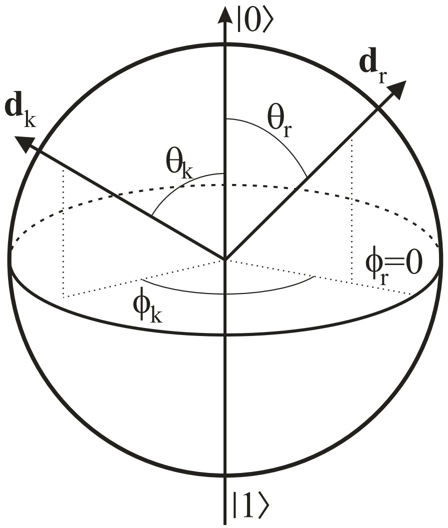










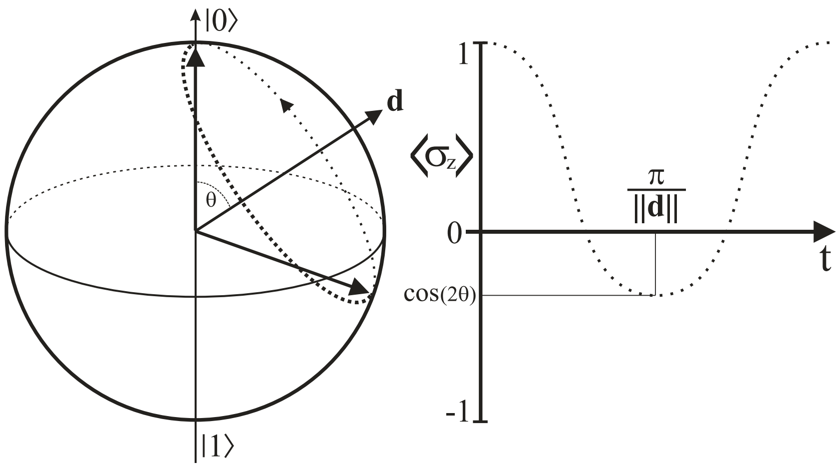


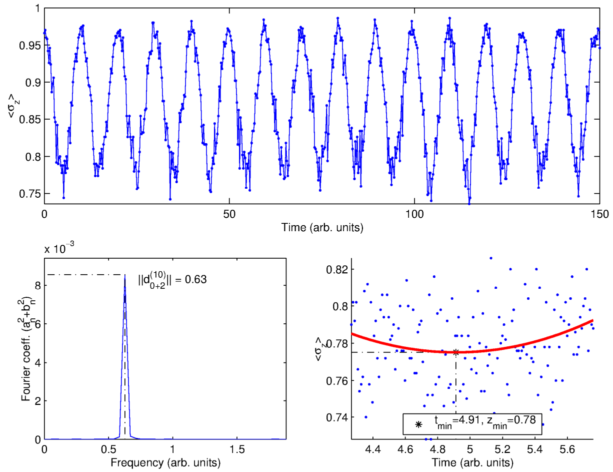
Experimental Hamiltonian identification for controlled two-level systems
S. G. Schirmer, A. Kolli, D. K. L. Oi, Phys. Rev. A 69, 050306(R) (2004)

Related work:
Optimal Experiment Design for Quantum State and Process Tomography and Hamiltonian Parameter Estimation
Robert Kosut, Ian A. Walmsley, Herschel Rabitz, arXiv: quant-ph/0411093

Signal extraction and inversion

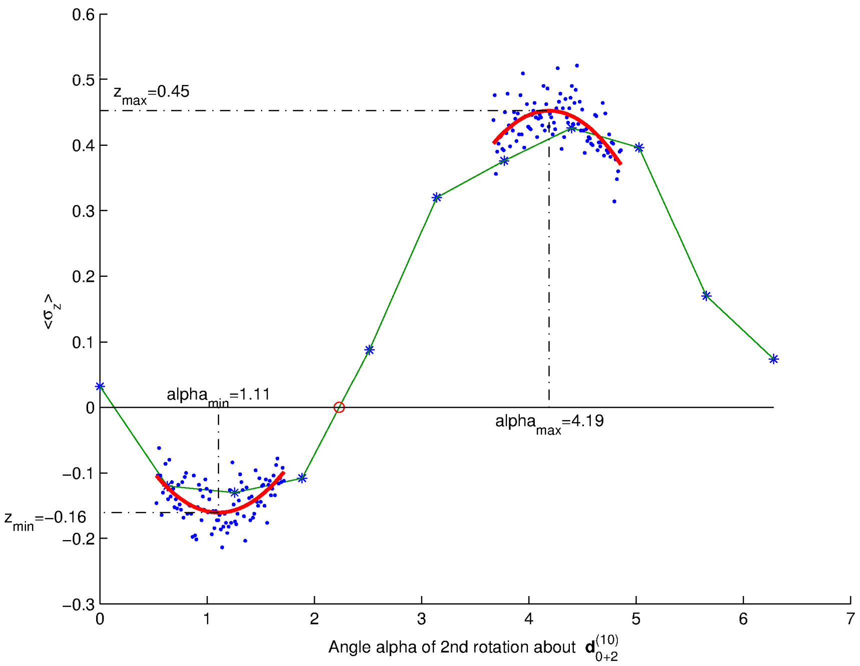
- Signal consists of a single trace, noisy (finite sampling, **shot noise**)
- Cosine with frequency and amplitude (visibility)
- Mapping between these with two of the Hamiltonian parameters
- Simple inversion process (for generic case)
- Cannot determine "phase" of Hamiltonian, **not observable** in limited preparation/measurement setting

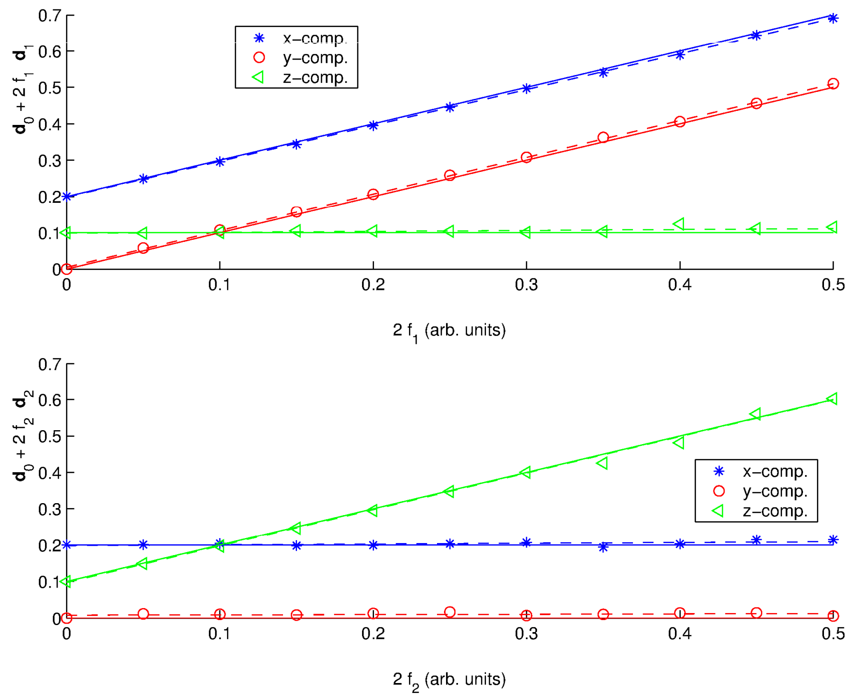




Identifying an experimental two-state Hamiltonian to arbitrary accuracy
Jared H. Cole, Sonia G. Schirmer, Andrew D. Greentree, Cameron J. Wellard, Daniel K. L. Oi, Lloyd C. L. Hollenberg, Phys. Rev. A 71, 062312 (2005)

- We can lift some ambiguity if we can introduce a **second** Hamiltonian, e.g. a control parameter
- Apply composite pulses, use first Hamiltonian to prepare state outside of computation basis (in generic case), then reconstruct additional Hamiltonian **relative** to the original one
- Can then map the control space



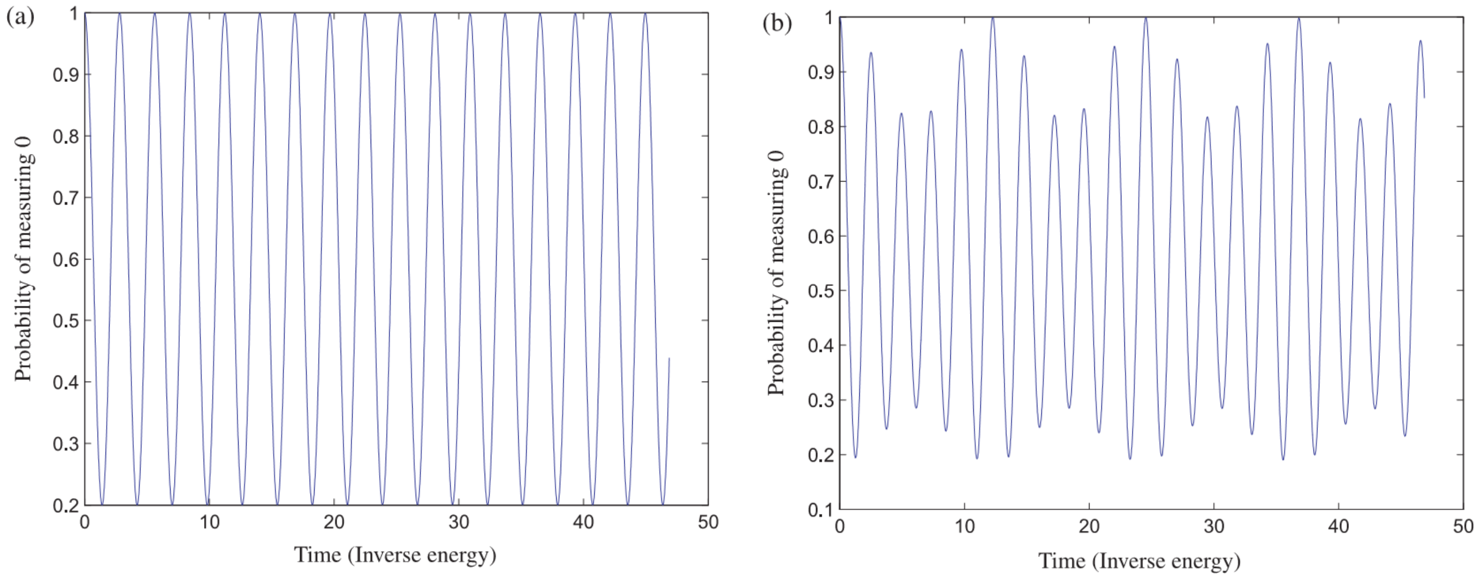


Experimental Hamiltonian identification for Qubits subject to multiple independent control mechanisms
S. G. Schirmer, A. Kolli, D. K. L. Oi, J. H. Cole, Proc. 7th Int. Conf. QCMC, Glasgow 25-29 July 2004 (AIP 2004)

Related work: , *Hamiltonian identification through enhanced observability utilizing quantum control*
Zaki Leghtas, Gabriel Turinici, Herschel Rabitz, and Pierre Rouchon, arXiv:arXiv:1102.2717v1

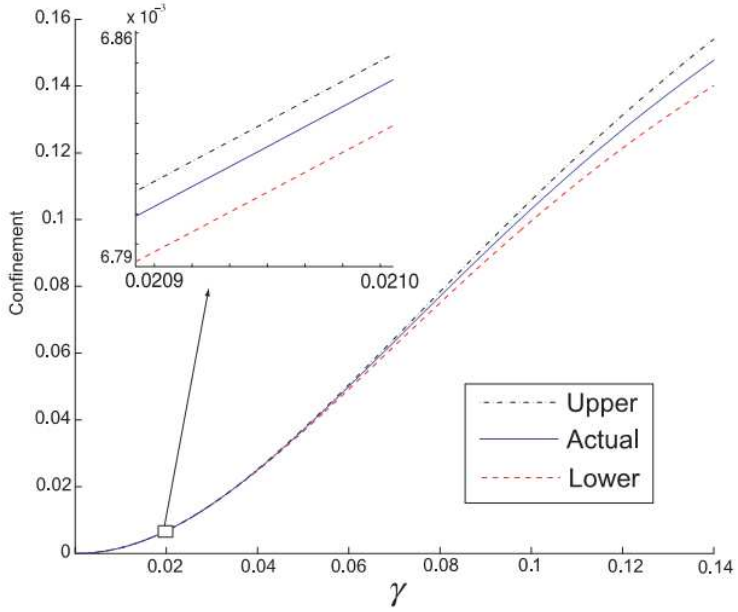
Subspace confinement

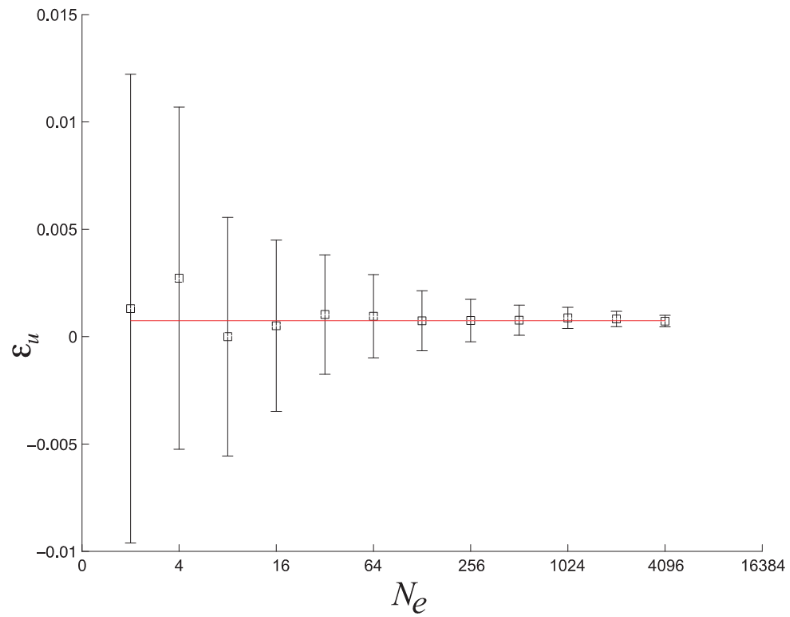
- When is a "qubit" not a qubit?
- Other levels can be involved, e.g. Atoms, quantum dot levels, SQUIDs
- Many measurement schemes only measure **population of a single level**, infer the complementary result, e.g. Fluorescence shelving



Subspace Confinement: How good is your qubit?
Simon J. Devitt, Sonia G. Schirmer, Daniel K.L. Oi, Jared H. Cole, Lloyd C.L. Hollenberg, New Journal of Physics 9 (2007) 384

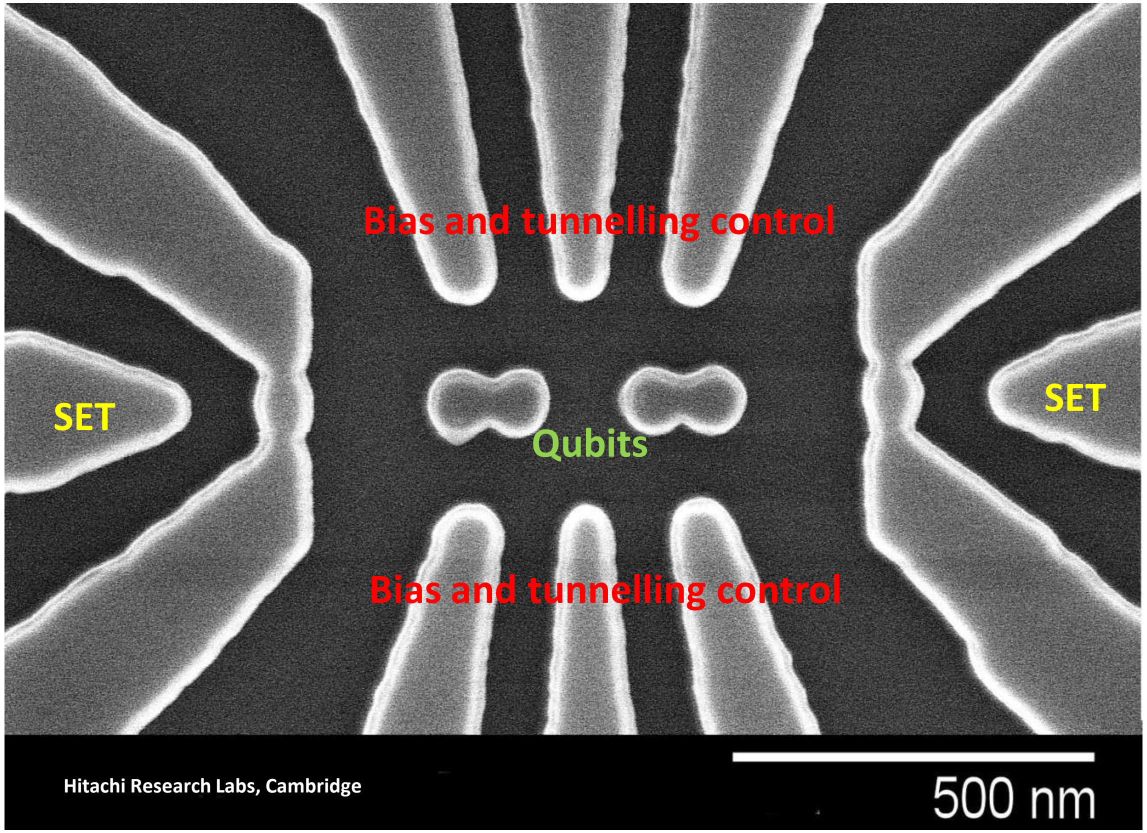
- Out of Hilbert space error much more deleterious than normal errors, loss-resistant codes have much greater overhead
- Important to be able to bound the magnitude of any involvement a third or more levels (projection onto qubit complement)
- Deviations from a pure two-level Hamiltonian can be detected from the power spectrum of the signal data trace and bounds calculated





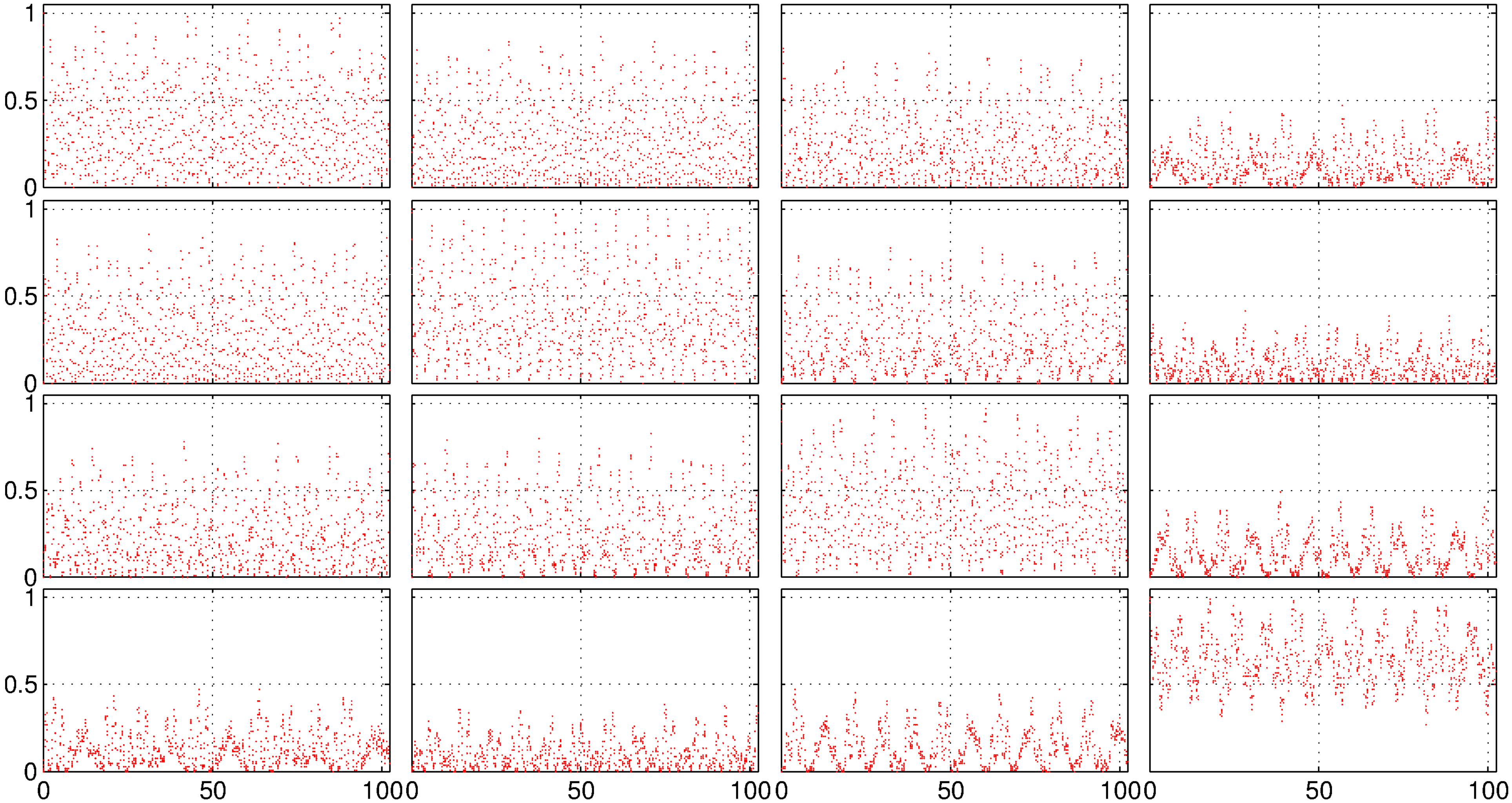
More complicated example: 2 qubits

- Consider a 4 level Hamiltonian system e.g. 2 qubits
- Measure and initialise in a single fixed basis

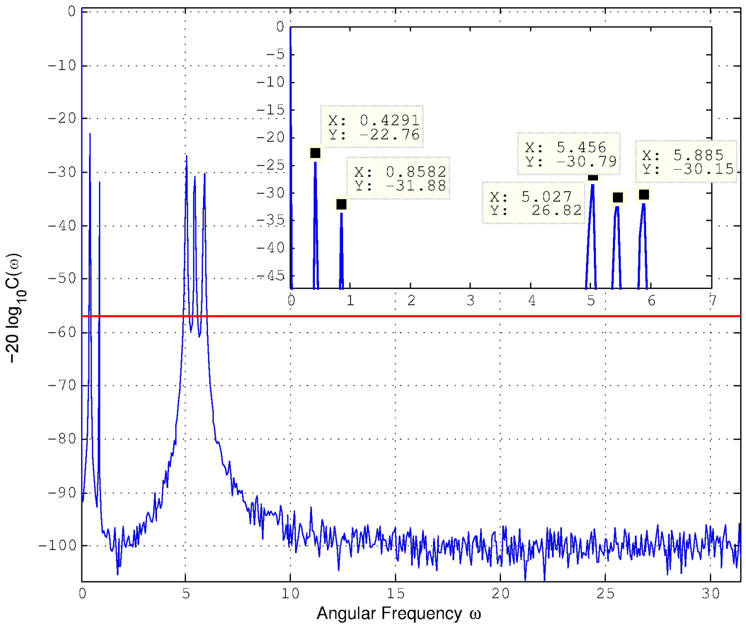


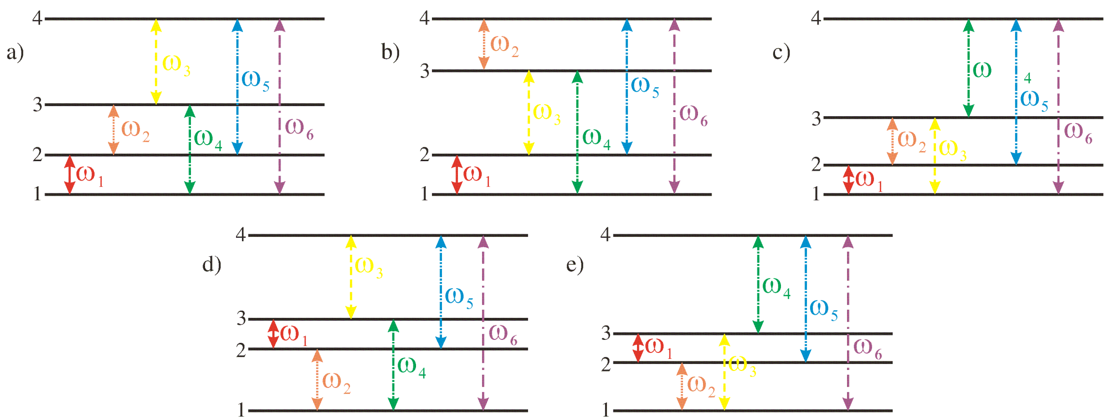
Two-Qubit Hamiltonian Tomography by Bayesian Analysis of Noisy Data
S. G. Schirmer, D. K. L. Oi, Phys. Rev. A 80, 022333 (2009)
Related work:
Scheme for direct measurement of a general two-qubit Hamiltonian
Simon J. Devitt, Jared H. Cole, and Lloyd C. L. Hollenberg
Phys. Rev. A 73, 052317 (2006)

- Signal consists of 16 data traces
- There are 214 signal parameters

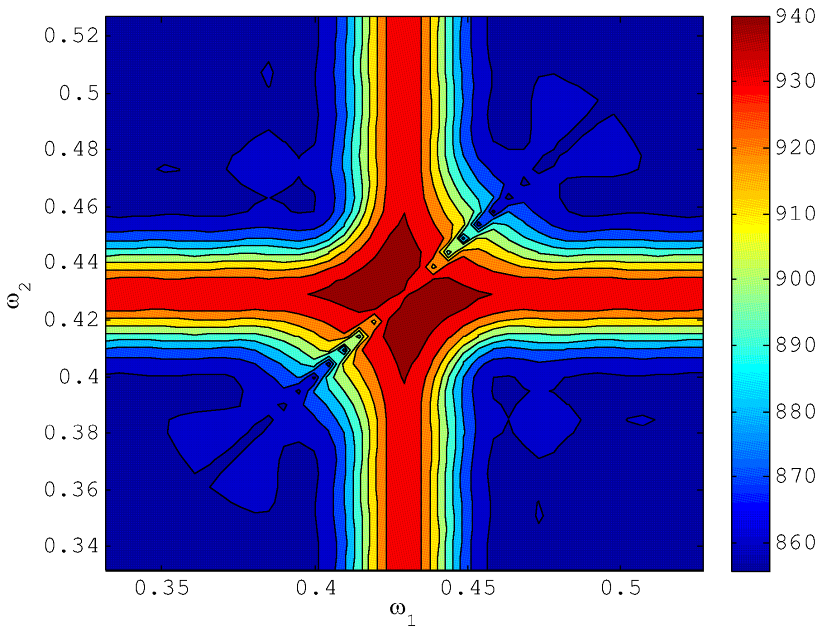


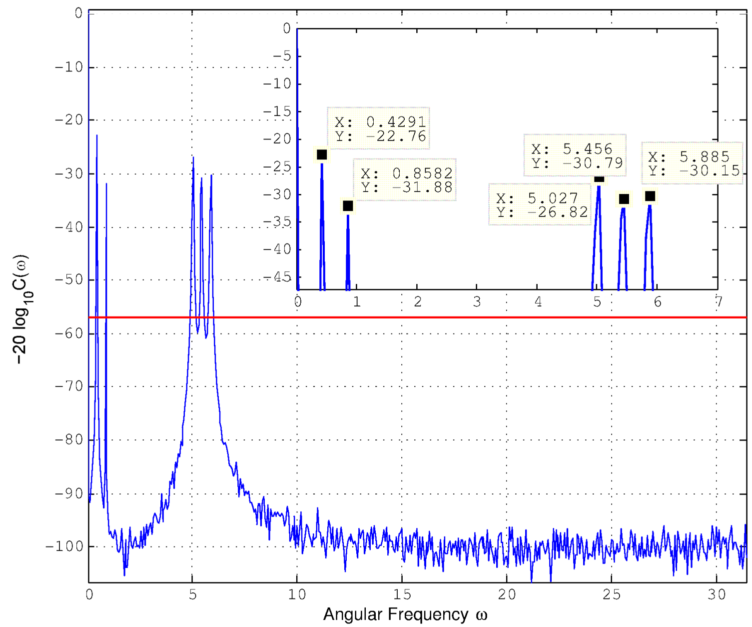
- **Interdependence** of the parameters, e.g. The six frequencies are made up of the sums of three underlying transitions
- Direct estimation of signal parameters intractable, search space too large to find any optimal solutions
- Fourier analysis not sufficient (multiple peaks, non-optimal estimator, frequency resolution, amplitude)





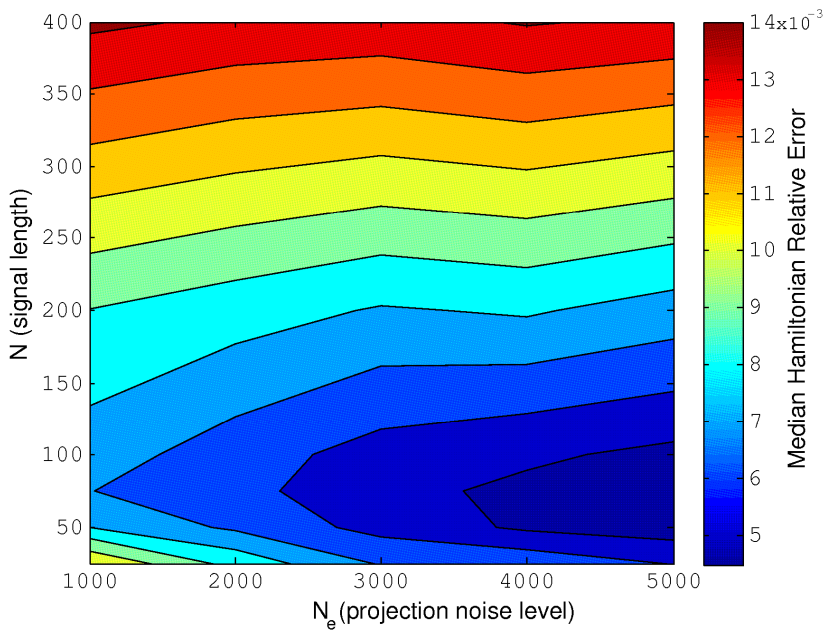
- **Bayesian signal analysis** to perform parameter estimation and reduce complexity of estimation task
- Frequency resolution much better than Nyquist (but assumes underlying model).
- Can estimate frequency independently of amplitudes
- Still search intensive but use power spectra as initial starting point

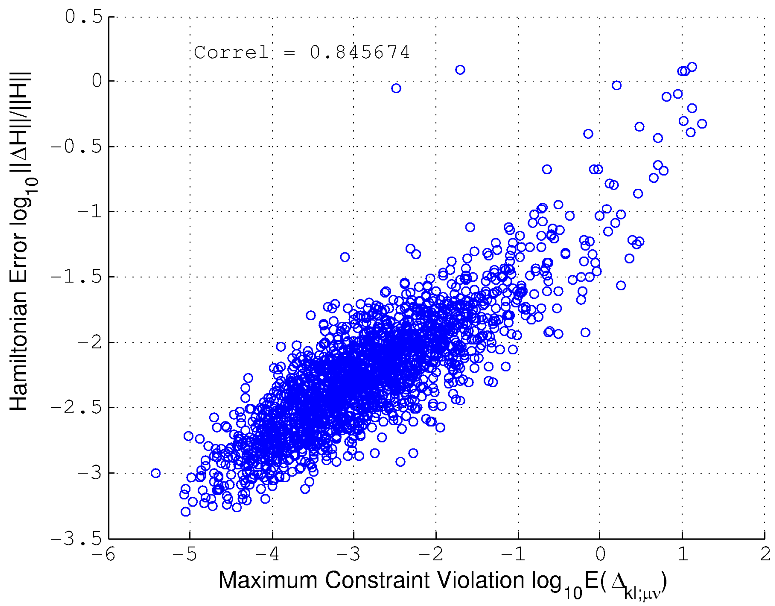




Bretthorst, G. Larry, 1988, "Bayesian Spectrum Analysis and Parameter Estimation," in Lecture Notes in Statistics, 48, Springer-Verlag, New Yorkc

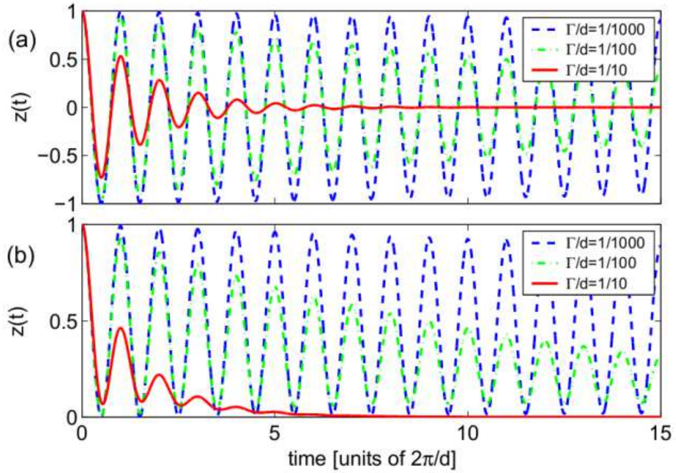
- Inversion of signal parameters to obtain Hamiltonian **sensitive** to violations of consistency. Cf. state reconstruction positivity
- Suggests imposing model at the parameter estimation step or else Bayesian estimation or Maximum Likelihood **directly** on Hamiltonian
- Can also estimate second Hamiltonian wrt reference Hamiltonian, but accuracy of first estimation lead to optimum sampling time due to **accumulated errors**





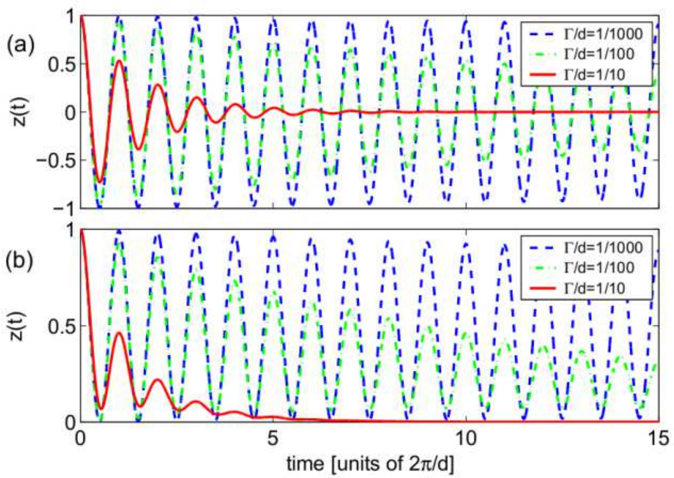
Open Systems

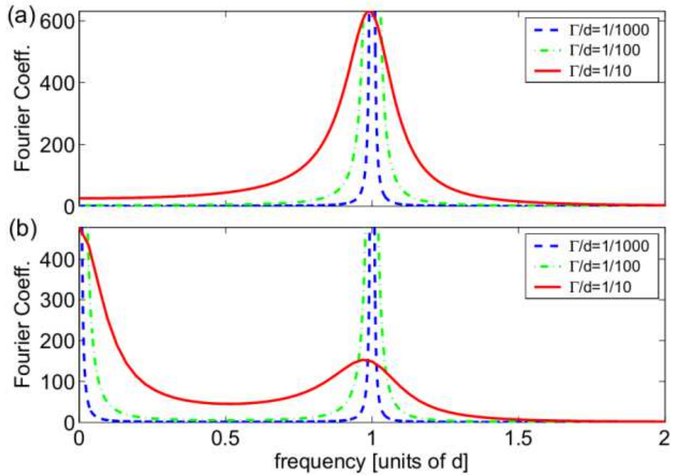
- Previous examples included “projection noise”, sampling statisics
- Real systems also suffer from decoherence, dominant form of which is dephasing
- For coherent systems suitable for QIP, these effects should be small (otherwise build a better qubit!)
- Assume weak damping limit, dephasing as a perturbation of predominantly Hamiltonian dynamics
- Can we still accurately estimate system dynamics in the presence of the noise?



Qubit with noise

- Lorentzian shape of Fourier spectrum to estimate dephasing and amplitude damping terms
- Damping limits ultimate length of the signal that can be usefully acquired
- Need to assume a particular noise model
- Hamiltonian reconstruction accurate despite non-trivial phase and amplitude damping



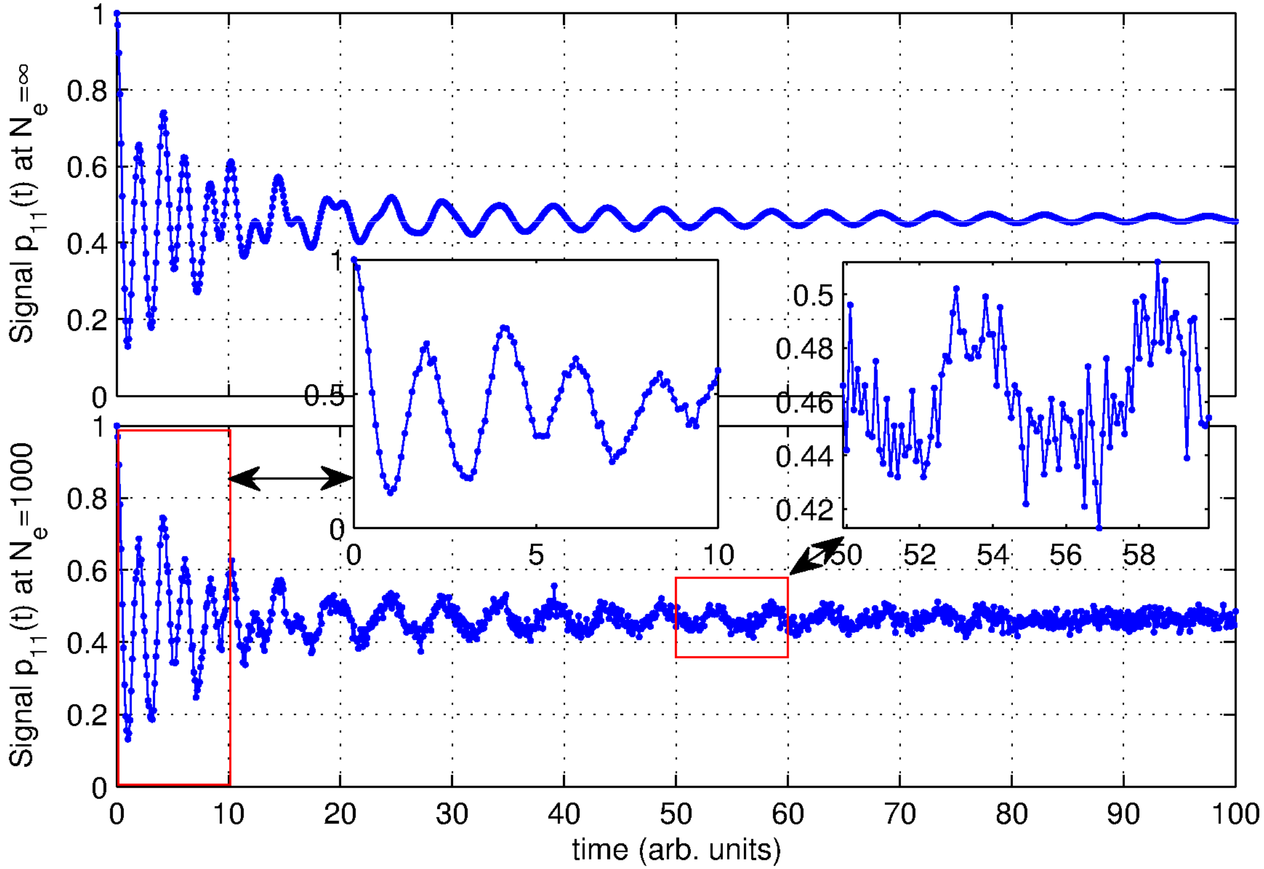


Qubit with noise

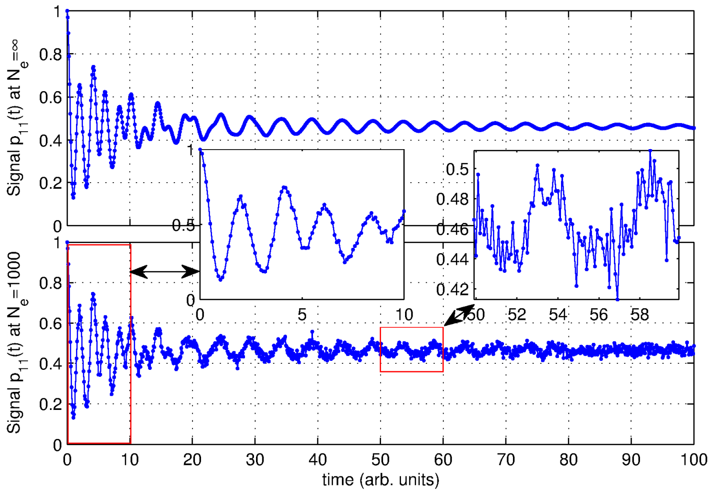
- Can also do Bayesian signal estimation to extract signal parameters from a single trace, single initial state
- Instead of Fourier Transform, use the Laplace Transform
- Reconstruct Hamiltonian from a set of polynomial equations
- Can examine the question of identifiability, what types of open quantum system dynamics can be identified
- Some issues with instabilities of inversion process, trade-off between robustness and initial state and number of identified parameters

Qutrit with dephasing

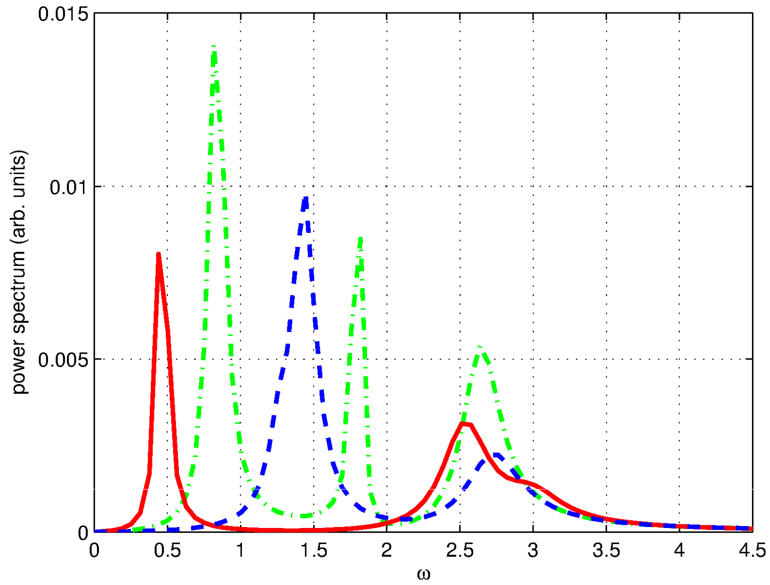
- More complicated system with multiple damping terms
- Arbitrary (generic) Hamiltonian but dephasing restricted to the eigen-energy basis

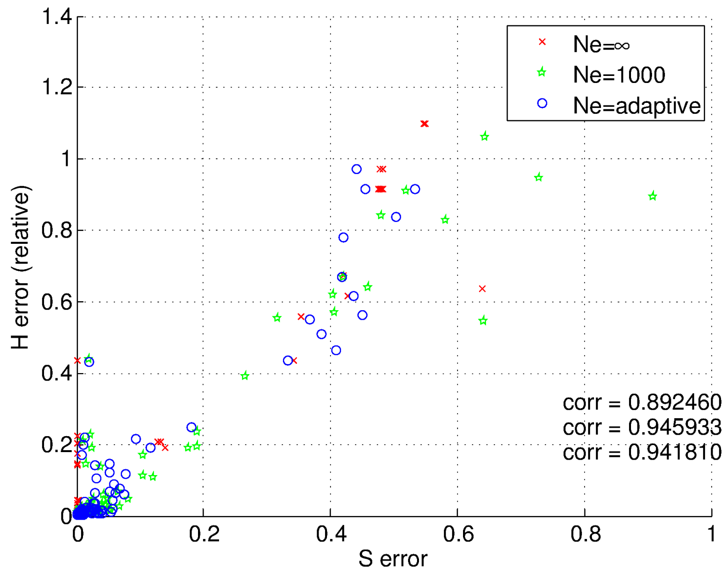


Quantum System Identification by Bayesian Analysis of Noisy Data: Beyond Hamiltonian Tomography
S. G. Schirmer, D. K. L. Oi, Laser Physics 20, 1203 (2010)



- Bayesian signal analysis still effective with weak to moderate damping, even when peaks disappear or merge
- Hamiltonian reconstruction accurate but damping rate estimation a lot worse, though precise knowledge not critical?





Challenges

- Improve **efficiency** of estimation of Hamiltonian parameters
- Increase the **size/complexity** of a system that can be characterized
- Incorporating **prior knowledge** system structure greatly increases efficiency and tractability of many problems
- Access to the entire system may not be possible, e.g. **Restricted access** to only the ends of a spin chain. Under certain assumptions about the connectivity of form of the Hamiltonian [1], initialisation and measurement of a single spin may be sufficient for characterisation [2]

[1] *Estimation of Coupling Constants of a Three-Spin Chain: Case Study of Hamiltonian Tomography with NMR*
E. H. Lapasar et al., arXiv:1111.1381
[2] *Bypassing state initialization in Hamiltonian tomography on spin-chains*
C. Di Franco, M. Paternostro, M. S. Kim, arXiv:1105.3667

Adaptive Bayesian Hamiltonian Estimation

- For single parameter estimation, e.g. Pure sigma-X Hamiltonian frequency for a qubit, it has been suggested that adaptive Bayesian estimation can give approximately exponential scaling in the accuracy with the number of samples, compared with a power law using off-line methods [1,2]
- Selecting the optimum adaptive measurement is simple in principle (maximise the mean conditional information) but extremely difficult to perform in practice for systems with more than a few parameters in the general case (even just 2 parameters difficult)
- Are there approximate adaptive schemes which can be more easily implemented in practice and will scale?

[1] *Characterization of a qubit Hamiltonian using adaptive measurements in a fixed basis*
Alexandr Sergeevich et al., Phys. Rev. A 84, 052315 (2011)
[2] *Adaptive Hamiltonian Estimation Using Bayesian Experimental Design*
C. Ferrie, C. E. Granade, D. G. Cory, arXiv:1111.0935

Bayesian Experimental Design

- Assume we have already performed an experiment E and obtained some data D about a system
- We estimate the value of the system parameters Θ

$$\Pr\left(\Theta \mid D, E\right) = \frac{\Pr\left(D \mid \Theta, E\right)\Pr\left(\Theta\right)}{\Pr\left(D \mid E\right)}$$

- We now want to choose a new experiment E_1, giving possible outcomes $D_{1,}$ which on average will give us the most information about the system
- We can calculate the probability of obtaining new data D_1 given our current estimate of Θ

$$\Pr\left(D_1 \mid E_1, D, E\right) = \int \Pr\left(D_1 \mid \Theta, E_1\right)\Pr\left(\Theta \mid D, E\right) d\Theta$$

On the measure of the information provided by an experiment, D. V. Lindley, Ann. Stat. 27, 986 (1956)
Bayesian Adaptive Exploration, T. J. Loredo, arXiv:astro-ph/0409386
Adaptive Hamiltonian Estimation Using Bayesian Experimental Design
C. Ferrie, C. E. Granade, D. G. Cory, arXiv:1111.0935

- For each of these possible outcomes, we calculate the average effect on our estimate of Θ

$$U(E_1) = \sum_{D_1} \Pr(D_1 | E_1, D, E) U(D_1, E_1)$$

- The utility function of an outcome is typically taken as the information gain

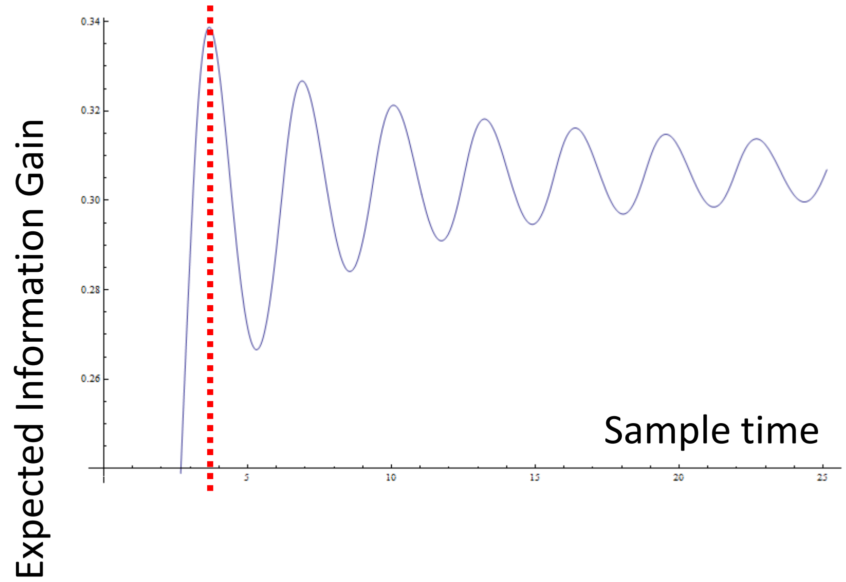
$$U(D_1, E_1) = \int \Pr(\Theta | D_1, E_1, D, E) \log\left(\Pr(\Theta | D_1, E_1, D, E)\right) d\Theta$$

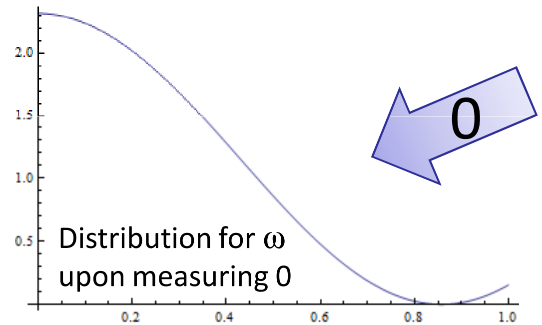
- The optimum experiment $\hat{E}$ maximises the expected information gain

$$U(\hat{E}) = \max_{E_1}\left[\sum_{D_1} \Pr(D_1 | E_1, D, E) \int \Pr(\Theta | D_1, E_1, D, E) \log\left(\Pr(\Theta | D_1, E_1, D, E)\right) d\Theta\right]$$

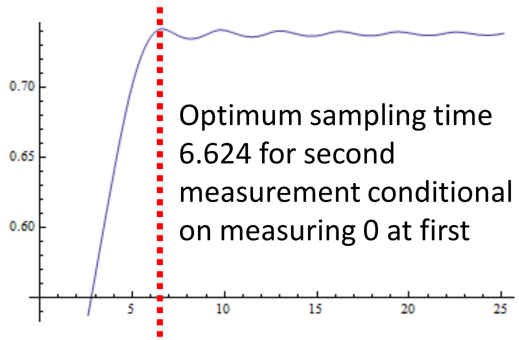
Single Qubit Example $H = \frac{\omega}{2}\sigma_x$

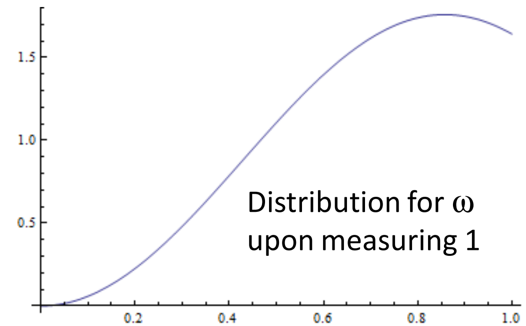
- Initialise and measure $|0\rangle$
- Flat prior [0,1] for ω
- Utility is information gain
- First measurement optimal at t=3.67
- Result is either 0 or 1



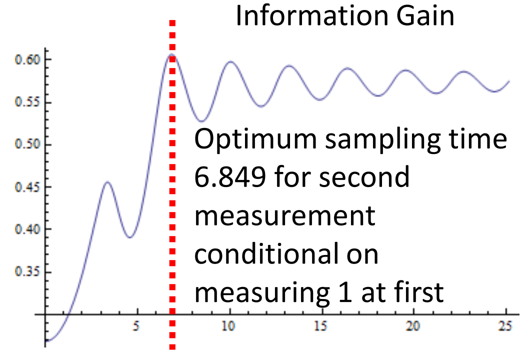


Information Gain

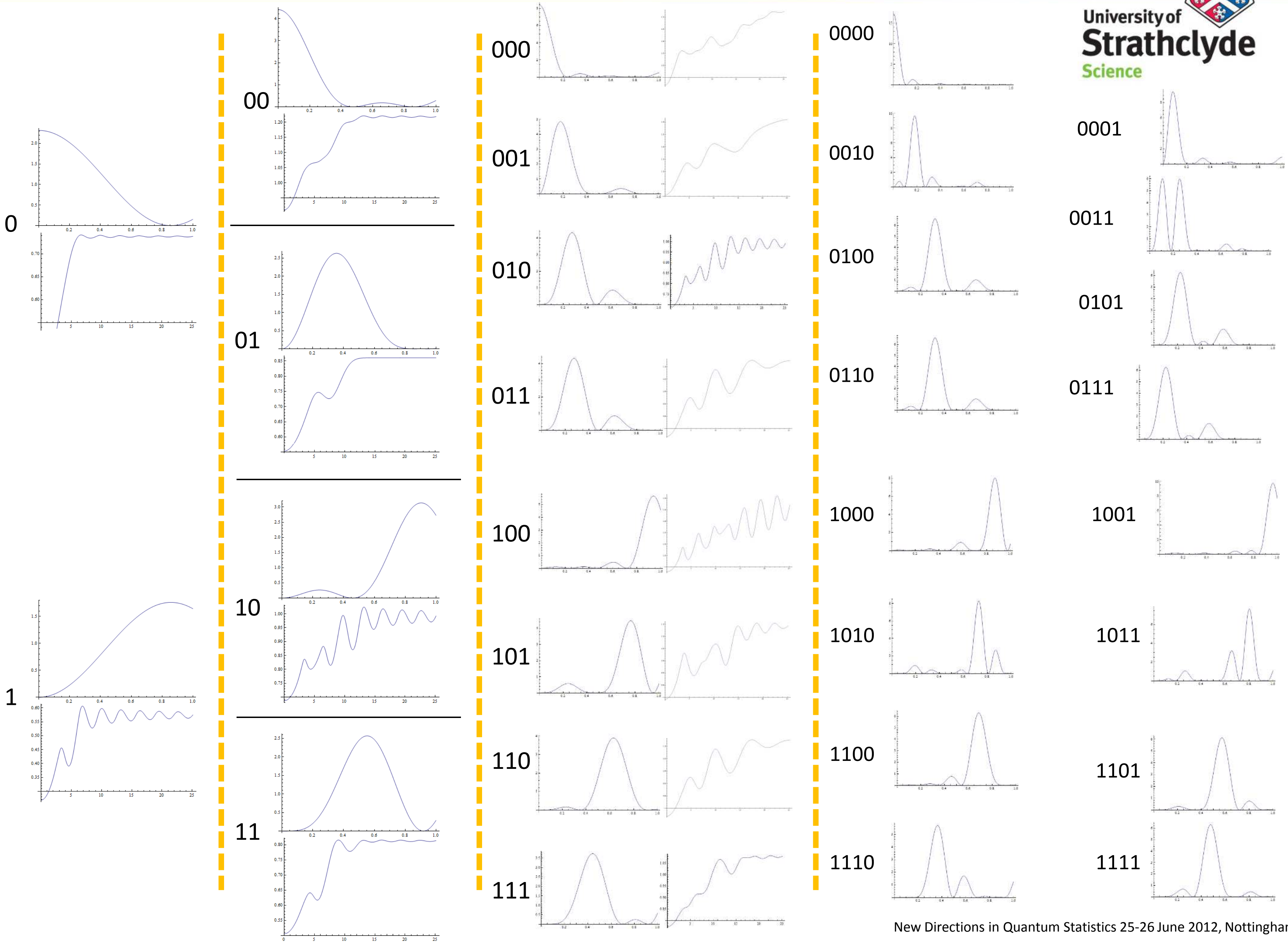




Information Gain



[1] *Characterization of a qubit Hamiltonian using adaptive measurements in a fixed basis*
Alexandr Sergeevich et al., Phys. Rev. A 84, 052315 (2011)
[2] *Adaptive Hamiltonian Estimation Using Bayesian Experimental Design*
C. Ferrie, C. E. Granade, D. G. Cory, arXiv:1111.0935


0
1
00
01
10
11
000
001
010
011
100
101
110
111
0000
0010
0100
0110
1000
1010
1100
1110
0001
0011
0101
0111
1001
1011
1101
1111

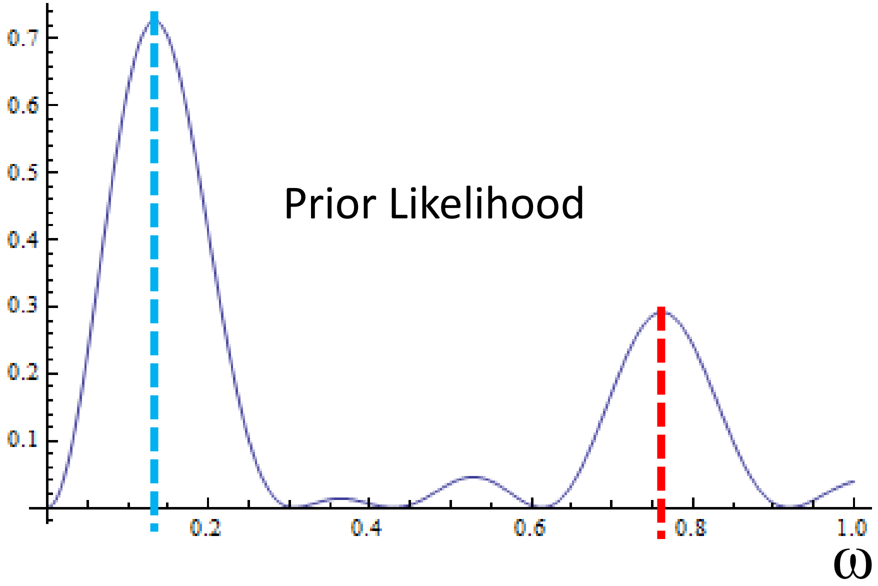
Remarks

- Easy to calculate the (un-normalised) likelihood of a model
- Not easy to integrate (normalise, calculate marginals, or expectations) peaked distributions
- Extremely computationally intensive to maximise expected information gain for arbitrary Hamiltonian, even for only a few parameters
- Alternatively, minimise the expected variance of the distribution at each step, “Greedy algorithm” for minimising expected mean square error of the Bayesian mean estimator [1]
- For simple qubit Hamiltonian, simple mapping between “signal” and the parameter of interest
- What happens in less straightforward situations?

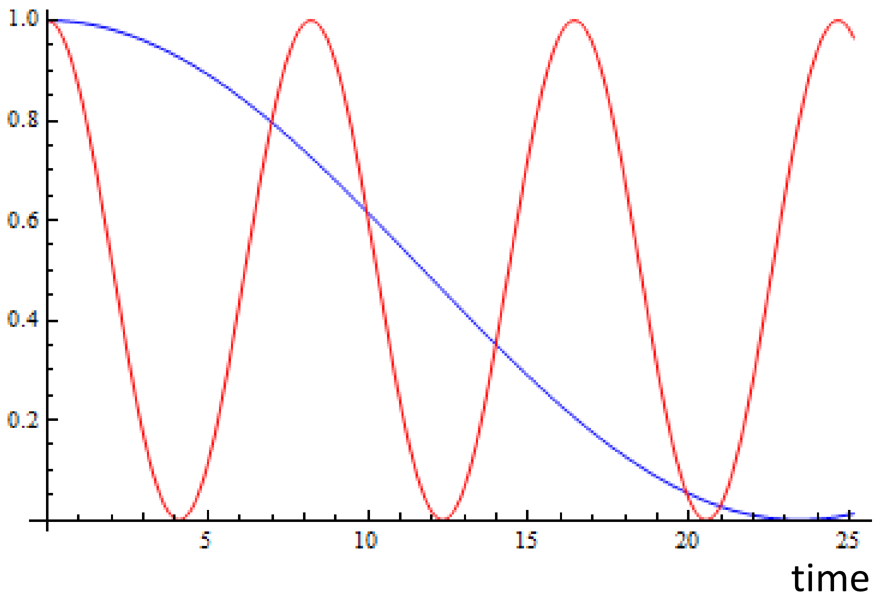
[1] *Characterization of a qubit Hamiltonian using adaptive measurements in a fixed basis*
Alexandr Sergeevich et al., Phys. Rev. A 84, 052315 (2011)

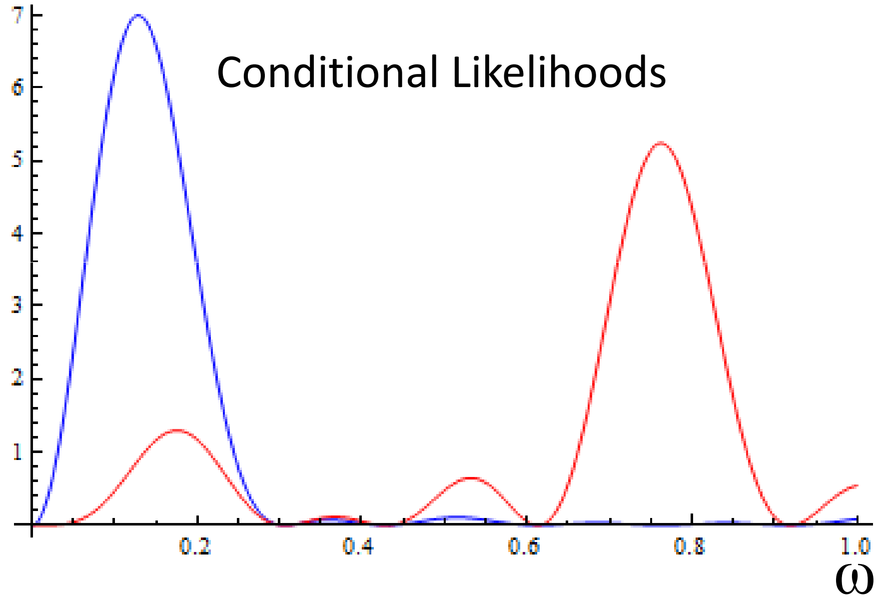
“Pretty good” adaptive sampling?

- Instead of maximising over the entire distribution, we can instead concentrate on differentiating between a few alternative hypotheses
- For a qubit, we simply select the measurement to maximise the distinguishability between two likely models
- Example: single parameter estimation (frequency of pure sigma-X Hamiltonian), select two peaks on likelihood plot and select the measurement time where the probabilities differ the greatest

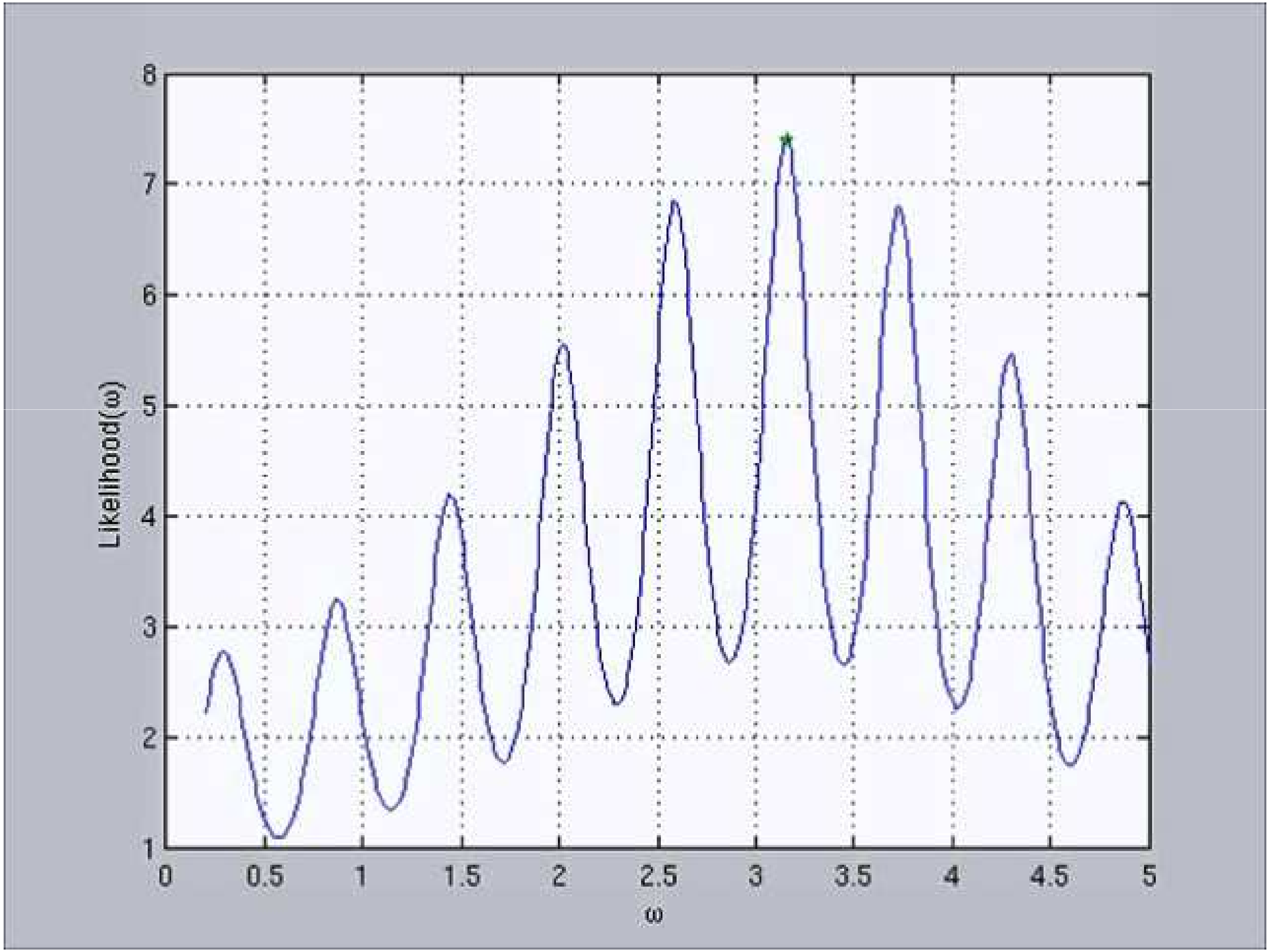


Probabilities for different models



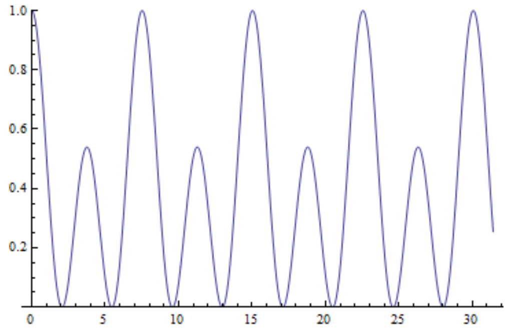


Single parameter estimation (20 measurements)



Two-Parameter Bayesian Estimation

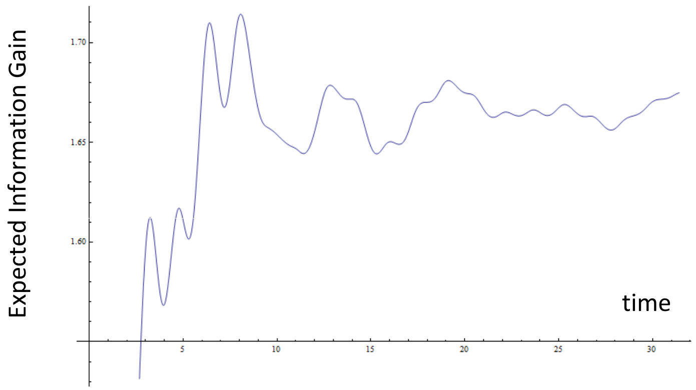
$$H = \begin{pmatrix} 0 & d_1 & 0 \\ d_1 & 0 & d_2 \\ 0 & d_2 & 0 \end{pmatrix} \quad \Pr\left(0 \mid t, \{d_1, d_2\}\right) = \left(\frac{d_2^2 + d_1^2 \cos\left(\sqrt{d_2^2 + d_1^2}\, t\right)}{d_2^2 + d_1^2} \right)^2$$

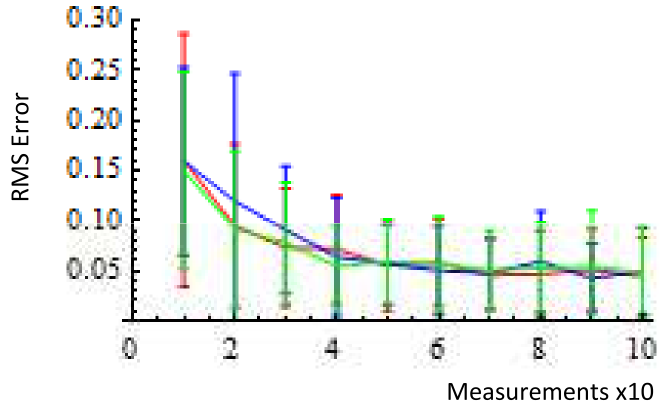


- Model of a qubit embedded in a larger state manifold
- Prepare and measure population of ground state (restricted measurement)
- Coupling to nuisance level leads to leakage from qubit subspace
- Want to determine couplings
- Assume $0.5<d_1<1$ and $0<d_2<0.5$, flat prior over region (could use Gaussian/Lorentzian?)
- More complex relationship between system parameters and signal probabilities, non-trivial correlation between the “visibility” of each parameter

Sampling strategies

- Ideally, choose the time at which to make a measurement to maximise the information gain (Difficult!)

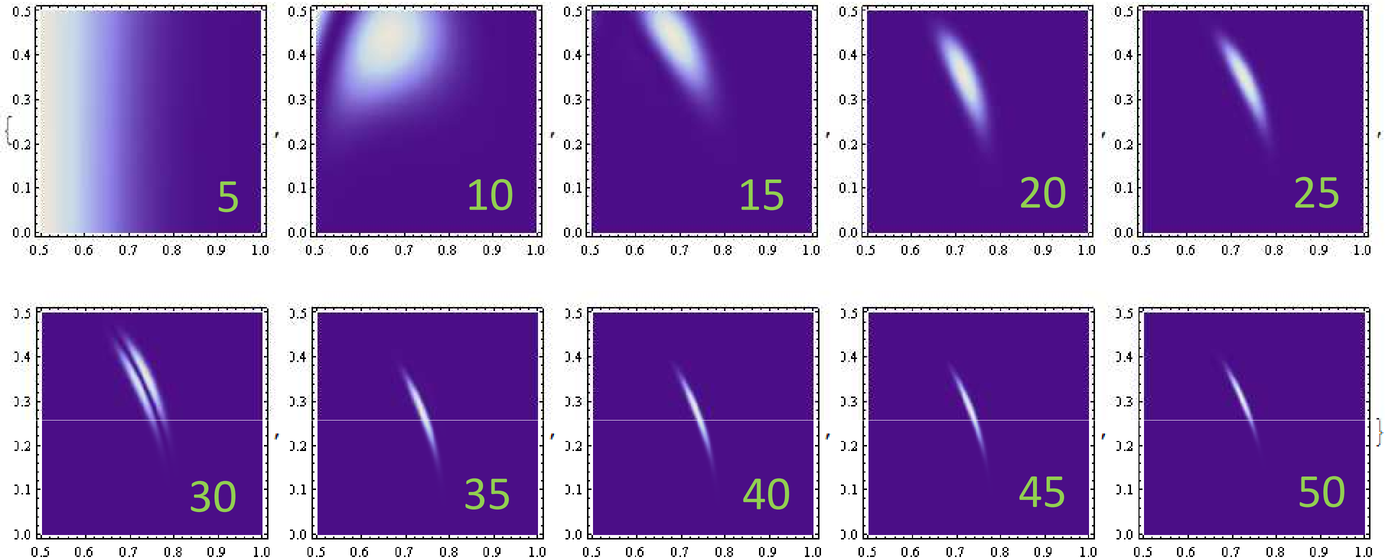


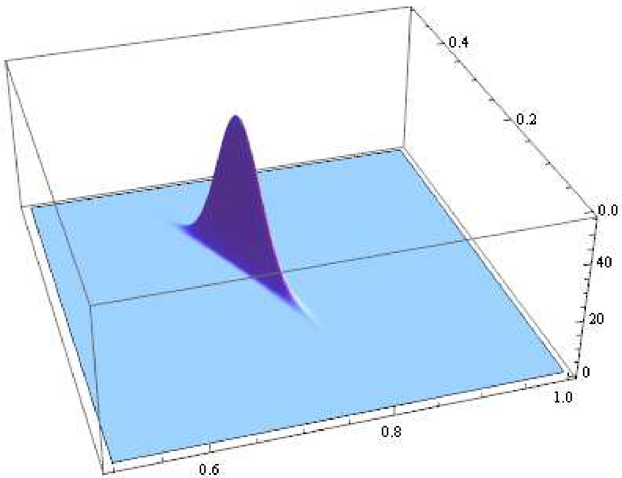


- Alternatively, non-adaptively sample
  - Regular intervals ("Fourier sampling" red)
  - Choose random times (blue)
  - Low discrepancy sampling (green)
- Investigating which non-adaptive sampling strategy works best, no clear trends so far.

University of Strathclyde
Science

2-Parameter Distribution Functions





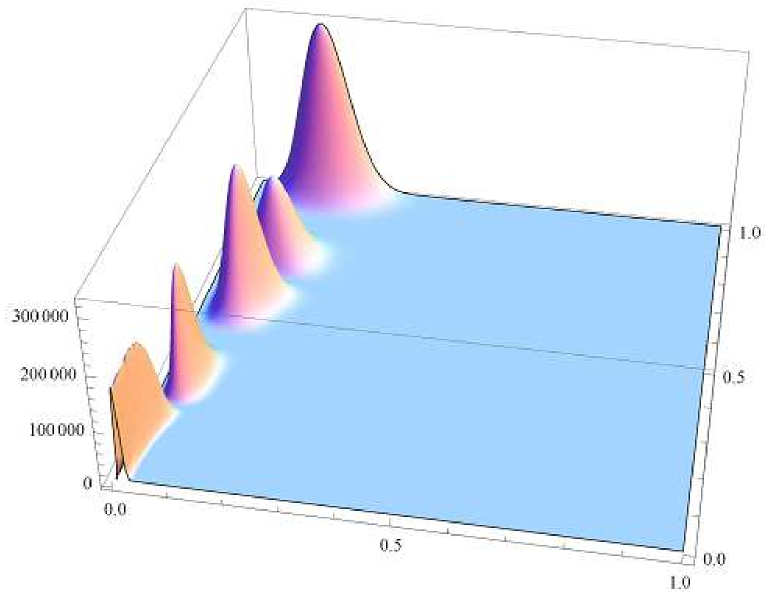
- Ridge indicative of greater uncertainty in d_2
- Perhaps analysis of the Fisher Information?
- Re-parameterising, signal is:

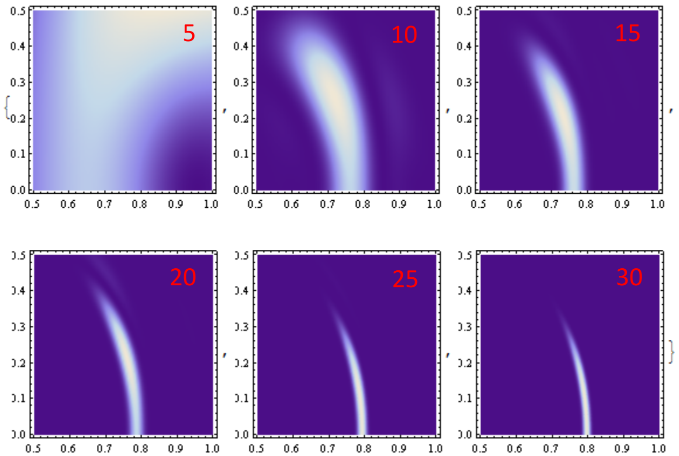
$$\Pr\left(0\,|\,t,\Omega,\alpha\right)=\left(\cos\left(\alpha\right)^2\cos\left(\Omega t\right)+\sin\left(\alpha\right)^2\right)^2$$

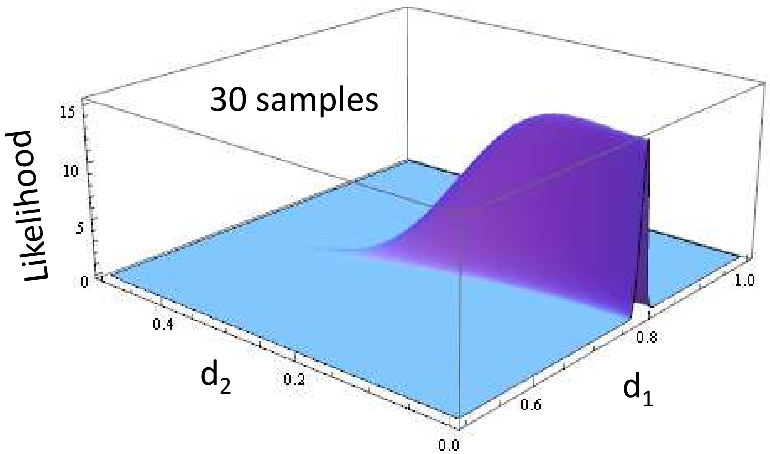
$$d_1=\Omega\cos\left(\alpha\right),\ d_2=\Omega\sin\left(\alpha\right)$$

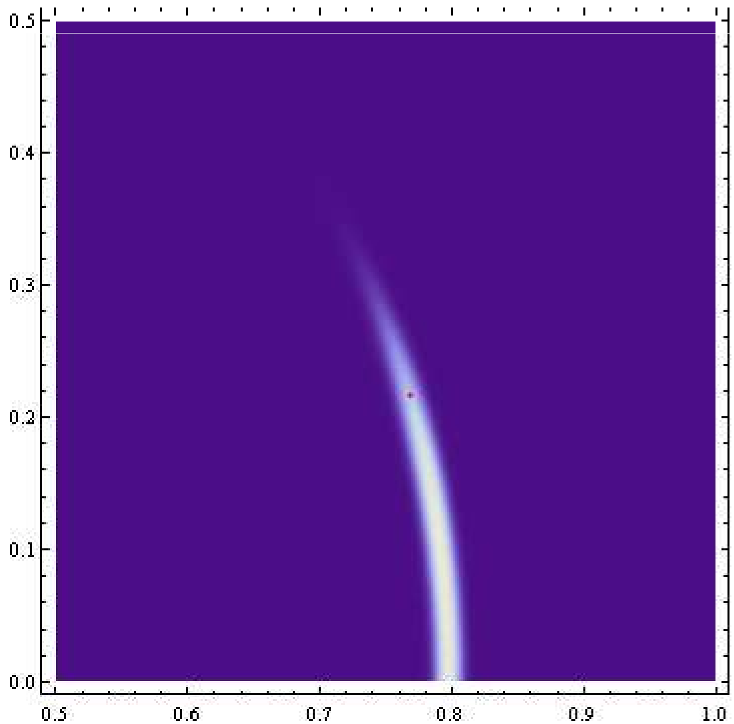
Adaptive Sampling

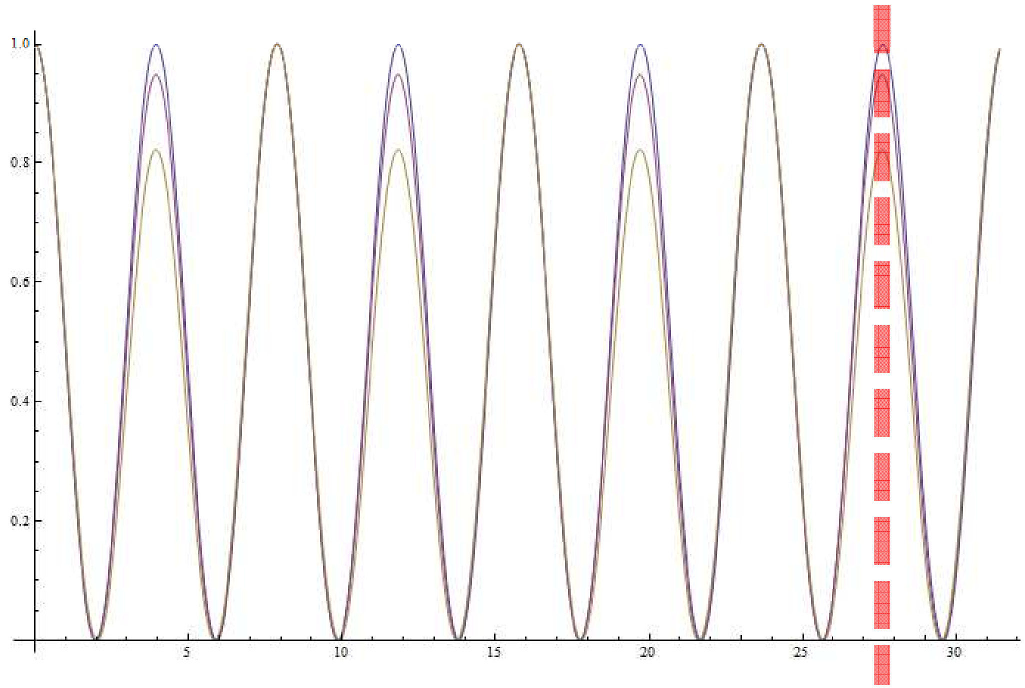
- If several regions show peaks, we can distinguish between them by sampling at times where signals differ the most
- We can reduce the width of a single ridge along one direction by a "Pretty Good Sampling", concentrating measurement times at which models distributed along the ridge have greatest variation
- This can equalise the spread of the distribution, lower aspect ratio of likelihood distribution



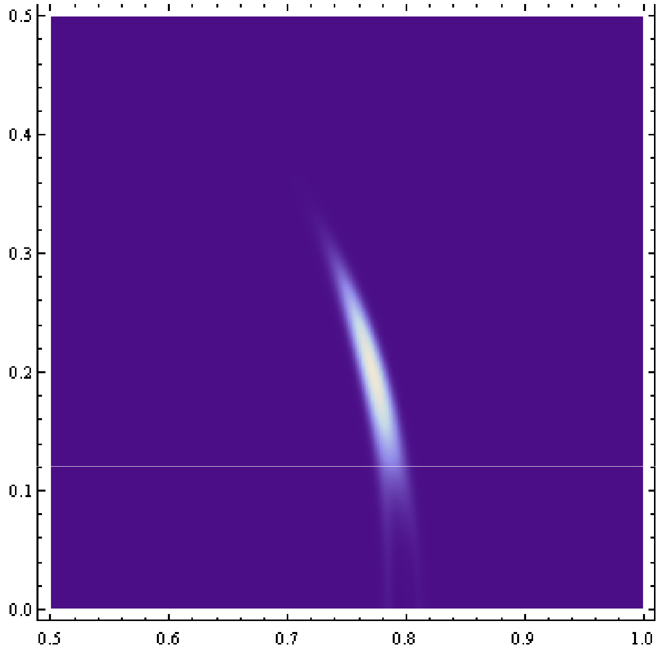

5
10
15
20
25
30
0.5 0.6 0.7 0.8 0.9 1.0
0.0 0.1 0.2 0.3 0.4 0.5

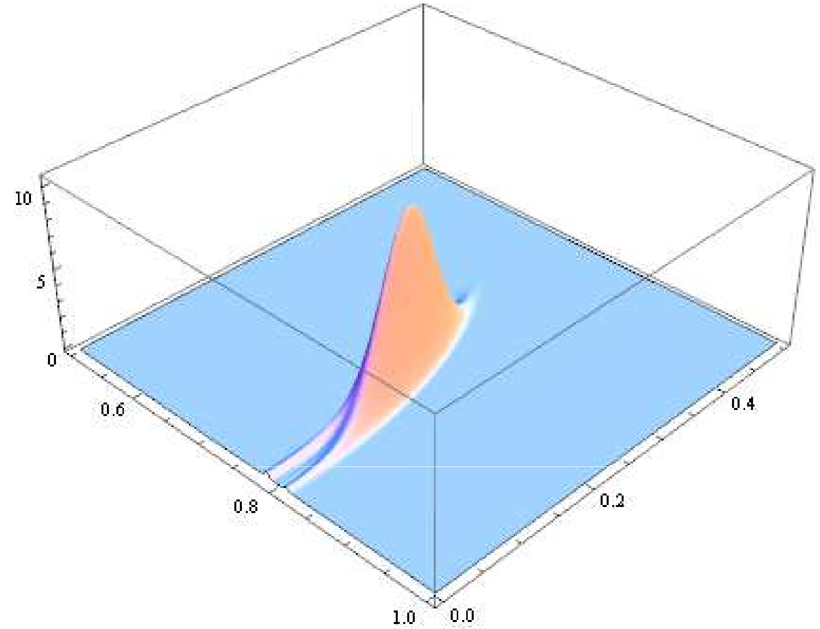

30 samples
Likelihood
15
10
5
0
d2
0.4
0.2
0.0
d1
0.6
0.8
1.0

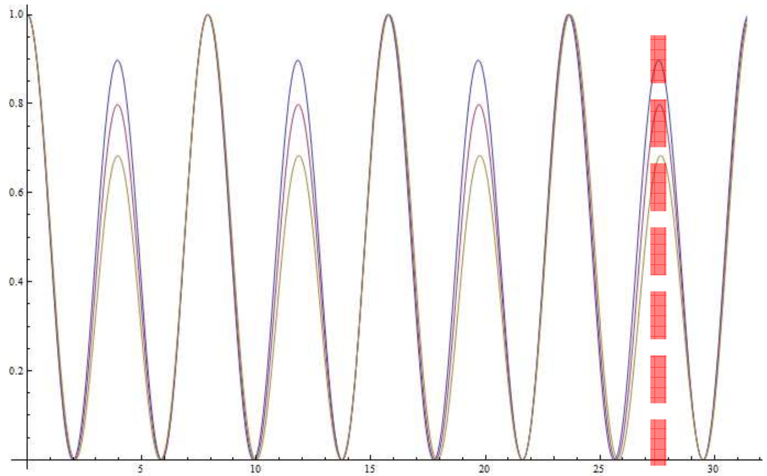

0.5
0.4
0.3
0.2
0.1
0.0
0.5 0.6 0.7 0.8 0.9 1.0


1.0
0.8
0.6
0.4
0.2
5 10 15 20 25 30

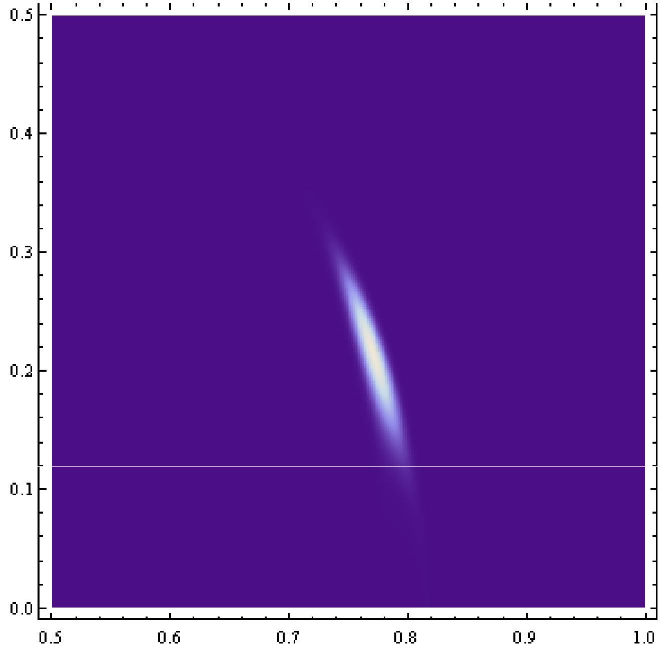
Plus 10 samples (40 Total)

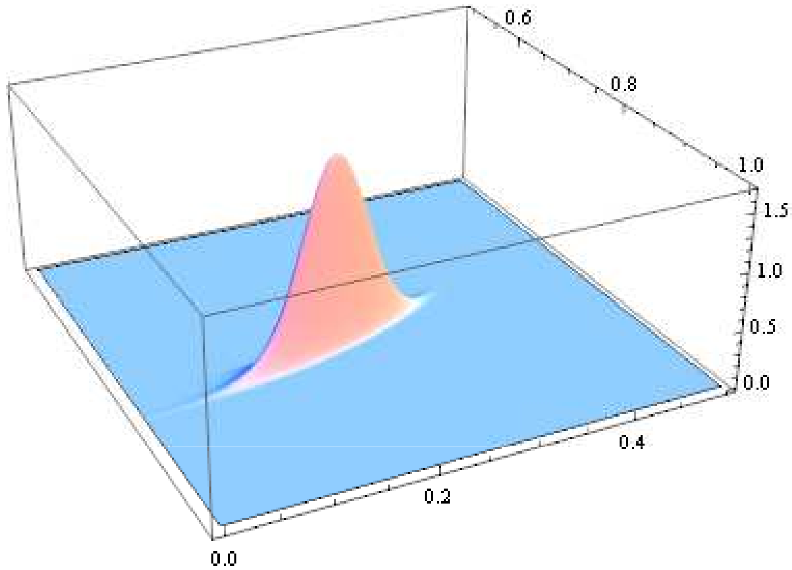


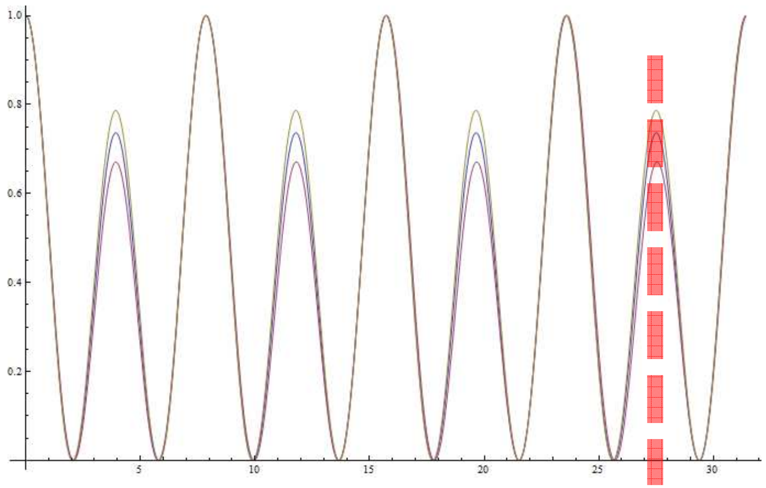




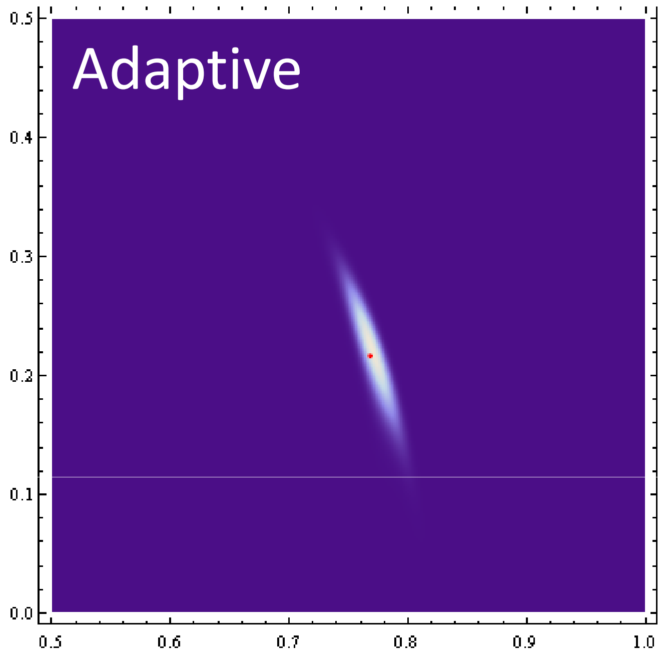
Plus 10 samples (50 Total)

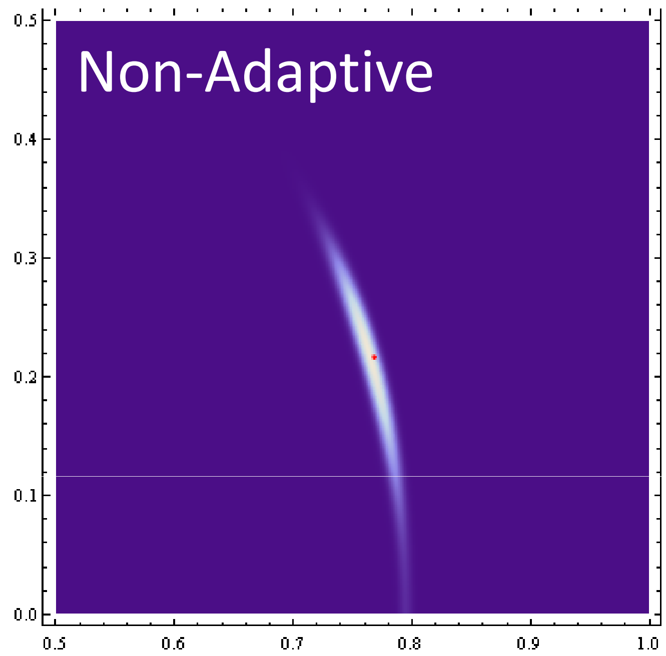


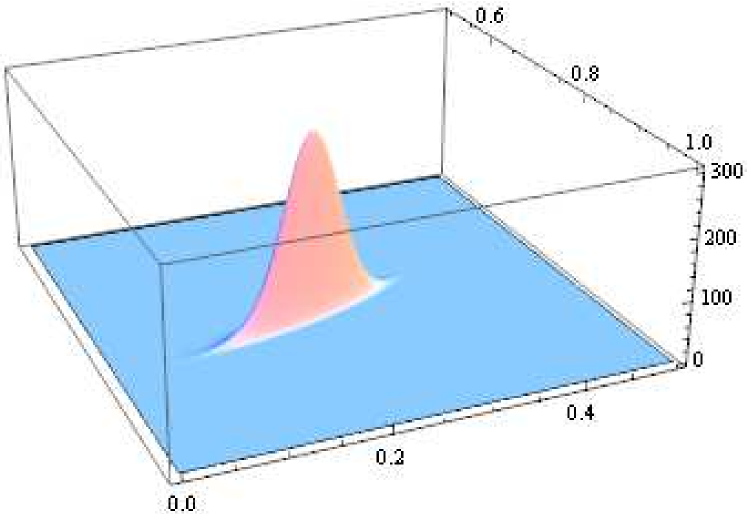


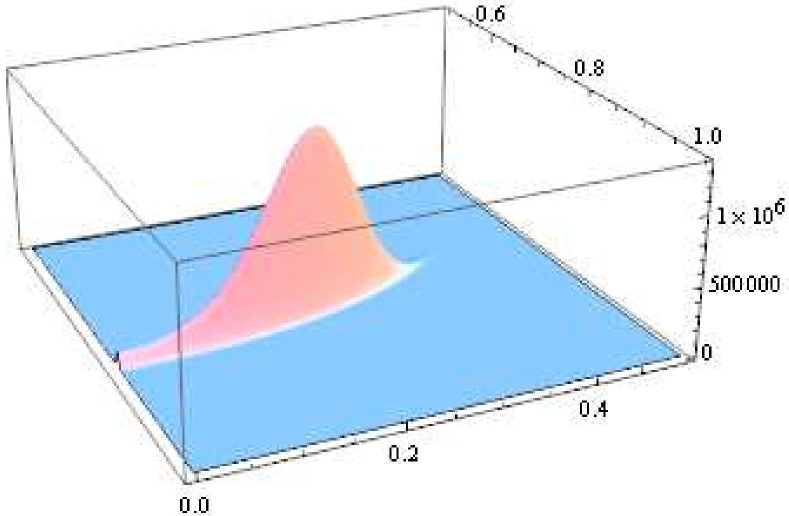


After 60 Measurements

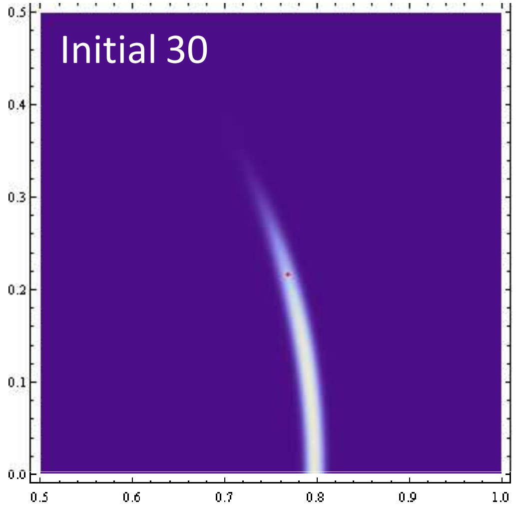


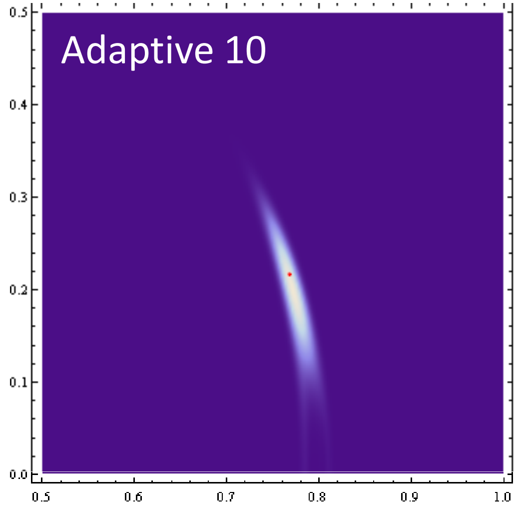


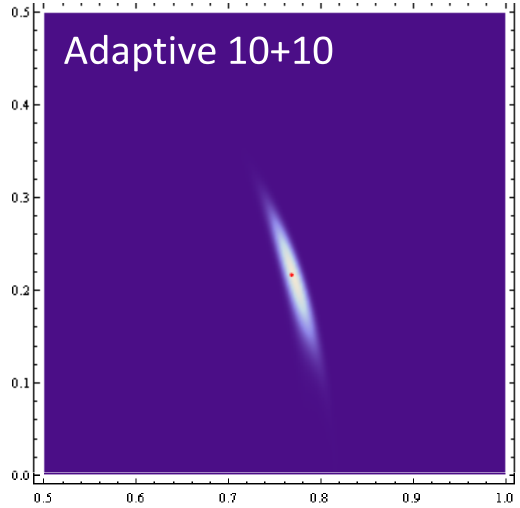


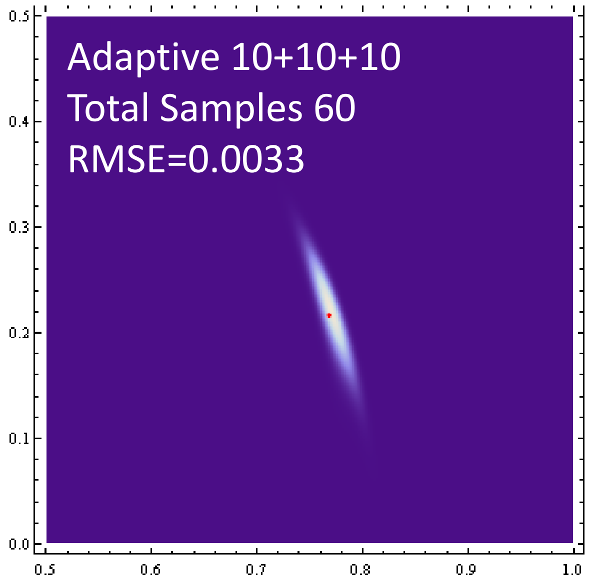


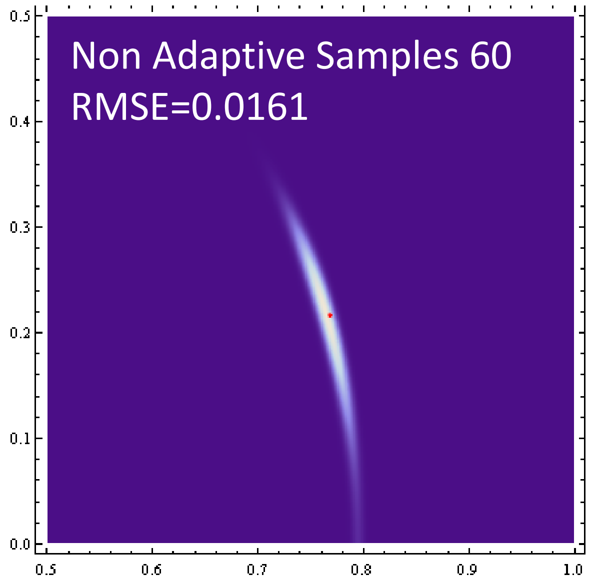
Comparison


Initial 30

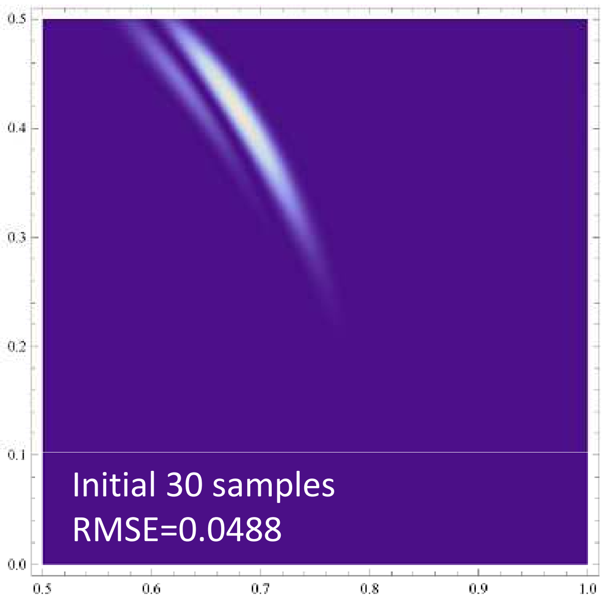

Adaptive 10

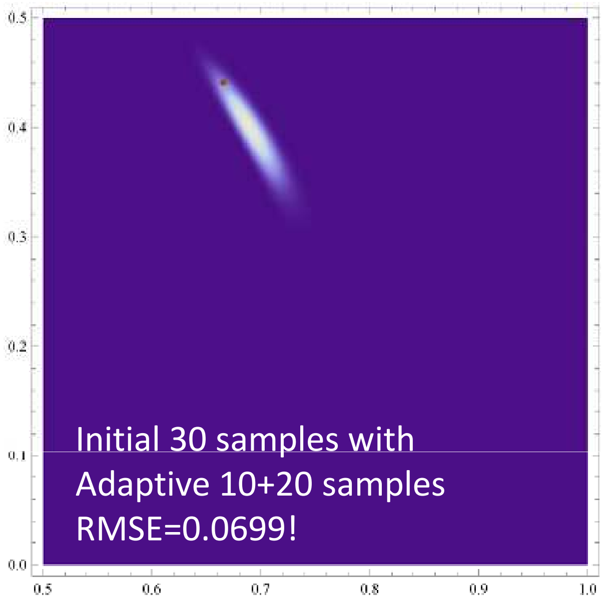

Adaptive 10+10

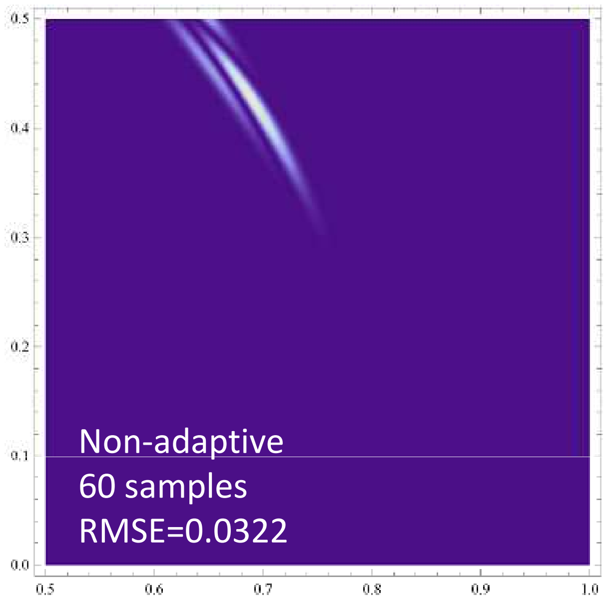

Adaptive 10+10+10
Total Samples 60
RMSE=0.0033

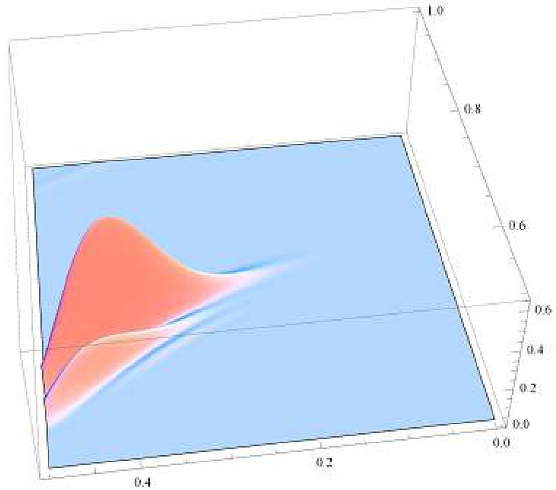

Non Adaptive Samples 60
RMSE=0.0161

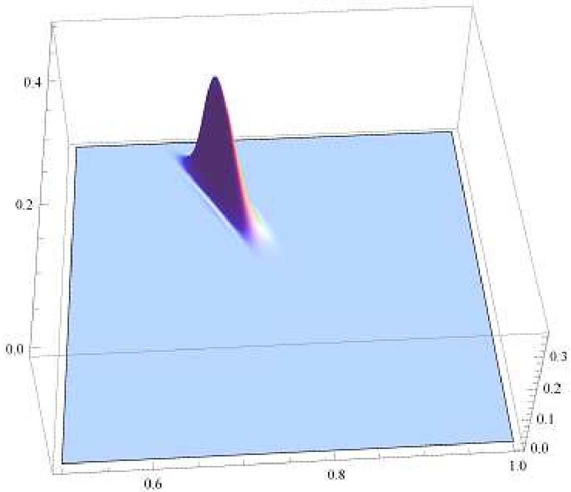
Negative Example

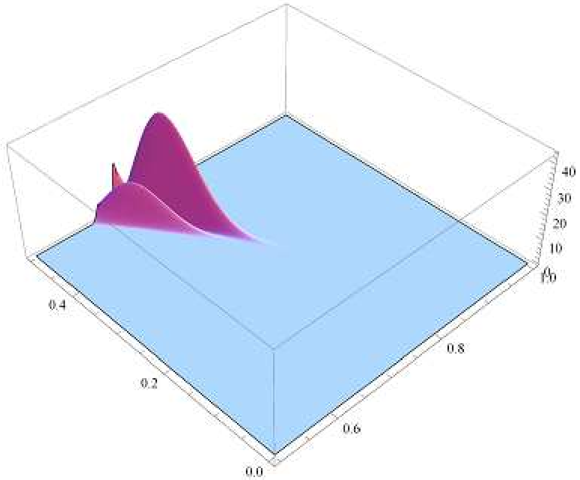










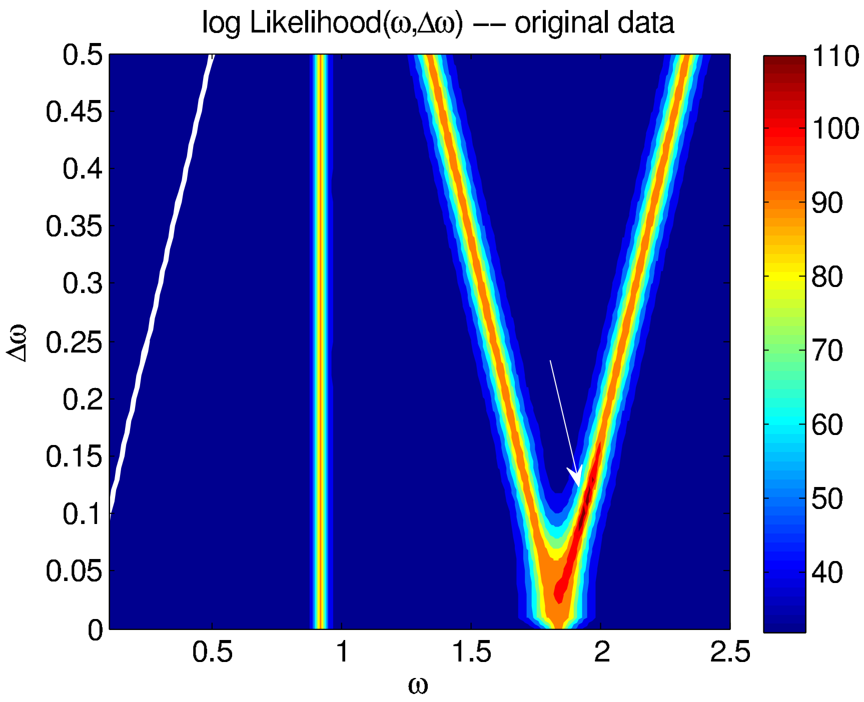


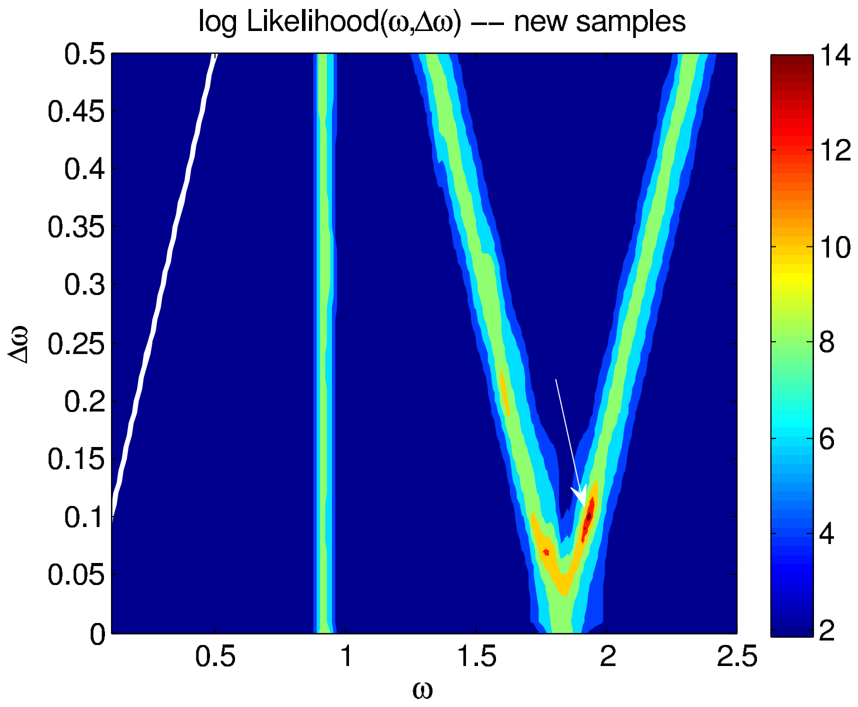
Observations

- Not necessarily RMS error better with adaptive vs non-adaptive
- "Coherence" of many samples surrounding a time point, high correlation hence reduced entropy of data leads to less efficient overall information gain
- Mean of non-adaptive distribution typically better estimator for the same number of measurements, despite broader width
- More fine-grained adaptive steps (after every measurement instead of every 10) may give advantage to adaptive measurements

Cf. Adaptive Bayesian Signal Estimation

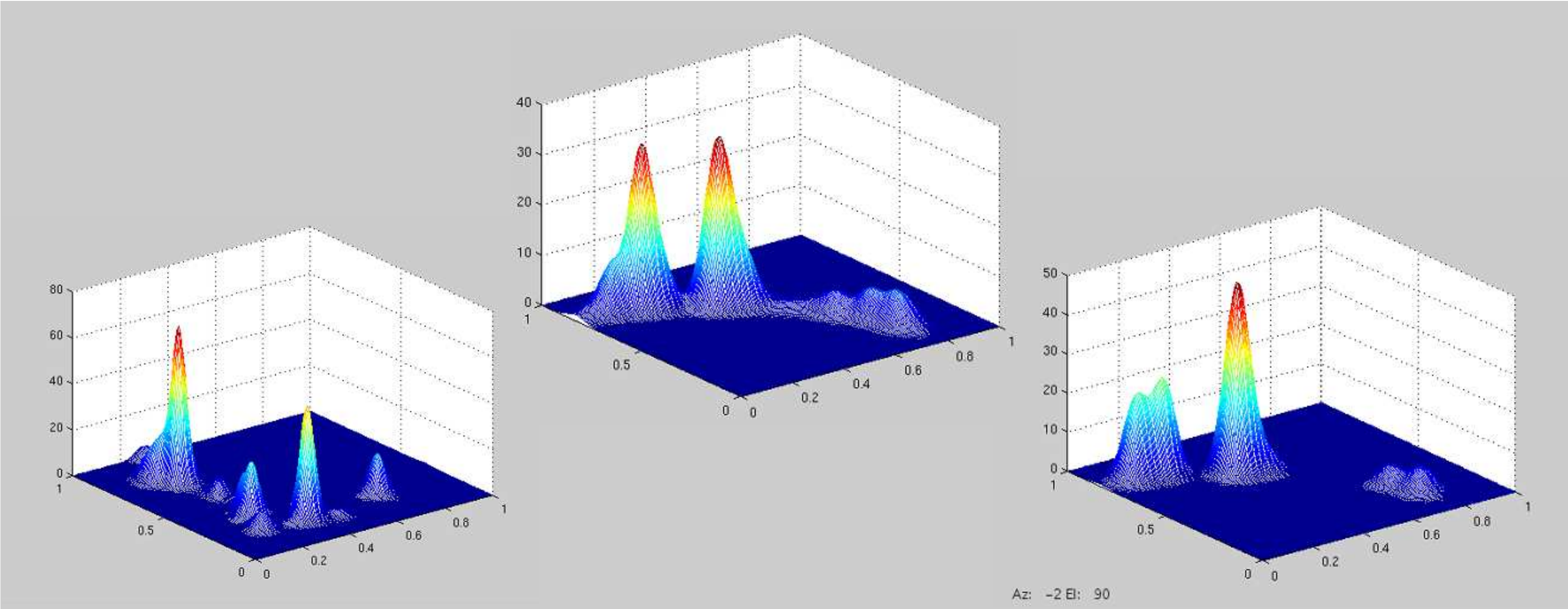
- Partially adaptive signal estimation gives much better accuracy than simply taking more random/LDS samples





PDF Approximation by KDE

- Approximate the PDF by a kernel distribution estimation (KDE)*
- Choose gaussian basis functions to approximate the actual PDF
- Use KDE to estimate statistics and peaks of the PDF
- Combine with (more) exact techniques once ROI known
- Could lead to faster optimal adaptive schemes



* Work by F. Langbein and S. G. Schirmer

Conclusions

- From limited initial resources, we can identify completely unknown Hamiltonians, at least in low dimensional cases
- We can handle decoherence, as long as the model is appropriate. Use QPT with identified Hamiltonian control for better estimates
- Blow-up in signal complexity as dimension increases, constrains what can be done in practise for the general case
- Prior knowledge of structure of Hamiltonian helps a lot, especially when access is limited
- Optimal adaptive estimation difficult for multi-parameter problems in general (highly peaked distribution), suggest Pretty Good Adaptive Sampling strategy to distinguish between likely models
- Analysis required of various approximate adaptive sampling schemes, trade-offs between complexity and efficiency

Open Questions?

- Fast methods for adaptive sampling
- Decoupling of parameters, better sets of variables to estimate
- Many parameter estimation
- Bias in adaptive sampling?
- Fisher information of signals for error bounds
- Model selection
- ???