Above Bandgap Hyperpolarization Mechanism in Isotopically Purified Silicon and Optimal Bayesian Experiment Design for T_1 Estimation

by

Thomas Alexander

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Science
in
Physics (Quantum Information)

Waterloo, Ontario, Canada, 2018

Abstract

This thesis is concerned with the mechanism underlying the above bandgap illumination Dynamic Nuclear Polarization (DNP) of phosphorus donors in isotopically purified silicon-28. Two proposed DNP models are introduced and compared. A series of NMR saturation experiments are performed in which modified buildup dynamics are observed when the saturation tone is applied at the bare phosphorus resonance. This effect is attributed to the phosphorus donor being ionized via the Auger process resulting in dynamics which are modelled as a set of coupled Bloch equations. The donor bound exciton capture and neutralization rates are extracted, and a paramagnetic shift of the bare phosphorus resonance is observed. These observed dynamics strongly imply the DNP mechanism is due to phononic modulation of the donor electron spatial wavefunction inducing cross-relaxation between the hyperfine coupled electron and nuclear spins.

The framework of Bayesian parameter estimation and its Sequential Monte Carlo(SMC) numerical implementation for continuous outcome probability distributions are introduced. Next, an introduction to Bayesian experiment design and its incorporation within the SMC framework is provided. A discussion of the computational challenges for continuous outcome distributions is given. To resolve these difficulties Monte Carlo Maximum Importance Sampling(MIS) numerical methods are developed which allow the evaluation of Bayesian experimental design heuristics such as the Bayes risk. These design strategies are applied to the problem of T_1 relaxation rate estimation with inversion recovery experiments. Experiments are optimized both respect to per-experiment performance and total experiment time. These techniques are shown to have substantial improvements over baseline methods. Furthermore, they compare favourably with previous frequentist experimental designs for IR experiments and demonstrate significant improvements.

Acknowledgements

Over the last three years, I have had the opportunity to work on numerous subjects and projects including silicon phosphorus, Bayesian parameter estimation, superconducting qubits, neutron interferometry, NV centres and optimal pulse control. This would not be possible in any group other than the Cory group. I would like to thank Professor Cory for his support and guidance in developing the skills and intuition to connect these subjects. To Professor Pushin I would like to thank you for your patience, and support. I know I was not the student you hoped for, yet you always did your best to make sure I was on the right track.

I have received an enormous amount of guidance and friendship from the Cory group, and I wish to thank every one of you. If I have forgotten you, it was not intentional, and I appreciate everything you have done for me. Feel free to bug me about it over a drink. To Holger and Rahul, I would like to thank you for your help and education in performing our silicon phosphorus NMR work. Long nights in the lab were made tolerable and even fun with you two. Additionally, I would like to give special thanks to Holger for helping edit the first half of this thesis. To Ian, I would like to thank you for the enormous help, and education you have given me when it comes to statistics and general quantum information theory. Your ability to carefully and eloquently explain a subject at a moments notice still surprises me on top of that I've had fun sharing an office the last few years.

George, you have made the lab a cheerful place, while teaching me low-temperature physics. You have the most charming way of grumpily brightening everyone's day. To Chris Granade and Chris Wood, while you have been gone a couple of years, I got my start in the Cory lab as an undergrad learning from you guys and you both greatly influenced my research trajectory for which I am grateful. To Mads and Dusan, it seems so long since we were doing assignments together in the first month of our degree, it is exciting to see how far you two have come and I am excited to see what you will do next.

To Carly, Chris and Sara thank you for trying to bring the chaos of the lab into order. I know it can be a thankless task, but we all benefit greatly from your support and would be helpless without you.

To my parents, I would like to thank you for your constant support and love. It has not been a smooth ride to completing this degree, but you have supported me the entire way. The same goes for all of my friends both new and old, your support and willingness to listen has meant a great deal to me.

And finally last, but certainly not least. To Kealin, You made a huge sacrifice to move to Waterloo to be with me while I worked on my education. I have watched you work so hard the last two years and establish yourself in your field, and I am so proud of you. You have supported me in a way you cannot even imagine from listening to me ramble about my work and even taking an interest, to picking me up late at night at the lab. You have been my constant companion and best friend. I do mean it when I say I could not have done this without you. I look forward to seeing what the coming years will bring for us.

Dedication

To Kealin for your unending patience, support and love.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

D^+ ionized donor. xi, xii, 17, 18, 35–38
D^0 neutralized donor. xi, 17, 18
D^0X donor bound exciton. xi, 17, 18, 35
T_1 T_1 relaxation time. 3, 8
LN$_2$ liquid nitrogen. 32
^{28}**Si** silicon-28. 8
^{29}**Si** silicon-29. 8, 9, 25
^{31}**P** phosphorus-31. x, 9–11, 26, 28, 32, 37

AEDMR auger electron detected magnetic resonance. 10, 42

BCRB Bayesian Cramér-Rao lower bound. xvi, xvii, 54, 57, 79, 88, 89
BFGS Broyden–Fletcher–Goldfarb–Shanno algorithm. 109
BIM Bayesian information matrix. 54

CLT central limit theorem. 65
CRB Cramér-Rao lower bound. viii, 54, 55, 62, 64, 91, 92
CW continuous wave. 10

DNP dynamic nuclear polarization. xii, 3, 7, 8, 11, 32–34

EDMR electrically detected magnetic resonance. 10
EMF electromotive force. 10
ESM effective strong measurements. xv, 67, 74–76

ESR electron spin resonance. 3, 10

FI fisher information. ix, xvii, 54, 55, 63–66, 77, 89–91, 105

FID free induction decay. 35, 46

FIM fisher information matrix. 55, 63, 77, 89, 91

FIR faster inversion recovery. x, xiii, xvi–xviii, 44, 45, 82, 84–89, 92, 97, 100, 104, 110

IG information gain. 57, 80, 81

IR inversion recovery. x, xiii, xv, xvi, 44, 45, 48, 49, 55, 72, 80, 84–88, 91, 92, 97, 101

KL divergence Kullback–Leibler divergence. 57

L-BFGS limited-memory BFGS optimization algorithm. xviii, 109, 110

LHe liquid helium. xii, 27, 30–33, 42

MCMC Markov chain Monte Carlo. 47, 66

MIS maximum importance sampling. ix, xiv–xvii, 60, 61, 66–70, 74–77, 79, 80, 85, 90, 94, 96, 98, 104, 105, 108, 109

MLE maximum likelihood estimate. 64

MRI magnetic resonance imaging. 44

MSE mean squared error. 52, 53

MW microwave. 3

NMR nuclear magnetic resonance. xi, 4, 7, 10, 22, 24, 27, 32, 44, 67, 104, 106

OD optical density. 33

RF radio frequency. 9

RWA rotating wave approximation. 9

RWMCMC random walk Markov chain Monte Carlo. 47, 48, 104

SGU signal generation unit. 32

Si:P silicon-phosphorus. vii, 3–15, 42

SMC sequential Monte Carlo. xiii, 47–50, 52, 58, 60, 97

SNR signal to noise ratio. viii–x, xiii, xv–xvii, 35, 37, 46, 49, 55, 62–68, 71, 72, 74–80, 82–86, 89, 92–94, 96–98, 100, 102, 105

SR saturation recovery. 44, 45

VRM valley recombination model. 11

Chapter 1

Introduction

Over the last six decades silicon has been the primary enabler of the unprecedented technological revolution that has transformed the world in a short period of time. Silicon based devices have changed the way, directly or indirectly, that nearly every activity takes place. Examples include communication, electricity generation, and computation [10, 32, 7].

Much of the utility in silicon comes from its position in group 14 of the periodic table. Pure crystal silicon is an intrinsic semiconductor, the presence of impurity donors from neighbouring periodic groups will modify the conductive properties of the crystal [95]. Doping has enabled the manufacturing of sophisticated and compact nonlinear electrical elements such as the diode, rectifier, and transistor [7]. Continual process refinement has reduced the CMOS element densities at an astonishing rate, approximately following Moore's law [60]. The current state of the art node size is approaching 10nm, with further improvements expected in the near future [76]. However, it is expected that Moore's law cannot continue indefinitely, as device dimensions start approaching the size of the silicon lattice constant. On these length scales, they will no longer behave as semi-classical semiconductors, and quantum mechanical effects will hinder device operation [69]. It has been predicted that the rules of quantum mechanics may be harnessed to solve specific problems such as prime-factorization exponentially faster than within the classical computing paradigm [73], which has spurred a dramatic increase in research towards the implementation and applications of quantum computing. It was quickly recognized that if quantum computing could be integrated into a CMOS architecture, the silicon industry's fabrication techniques could be used to accelerate the development of large-scale quantum computers [84]. Silicon has previously proven itself to be the element that keeps giving, and naturally, it has become a significant avenue in which quantum devices may be pursued. This thesis focuses on the substitutional phosphorus defect in silicon, which is a promising qubit system for sensing and computing.

1.1 Donors in Silicon

Donors in silicon are dopant atoms with five or more valence electrons and are substitutional defects in the silicon lattice [95]. As silicon has four valence electrons, donors with five or more valence electrons will donate the extra electrons to the conduction band. It is also possible to dope silicon with atoms with fewer than four valence electrons; such defects will leave a hole in the valence band. If there are an excess of conduction(valence) band electrons(holes) the silicon will behave as an N(P)-type semiconductor [95]. While the properties above are seen at room temperature, where thermal energies are large compared to the binding energy of the donor potential when the temperature is lowered the donor electrons(holes) will no longer have sufficient thermal energy to escape the attractive potentials of defect nuclei. The electrons will become spatially localized around the defect nuclei inside the silicon lattice and form neutral donor systems [49].

If the defect nuclei have nonzero spin the bound electrons(holes) will couple to the nuclei via the hyperfine interaction [75]. A splitting of the resonance will be seen when electrons localize around the spin. Such systems are of interest in electron spin resonance (ESR) as they provide a localized donor electron, which also interacts with nearby nuclei within the lattice. Examples of neutral donors in silicon that have been resolved include phosphorus, arsenic, antimony and bismuth [22]. These donor nuclei have been observed to have relatively long dephasing times, on the order of several milliseconds. It has been demonstrated that these coherence times may be increased dramatically with a combination of isotopic purification, refocusing techniques and ionization of the nucleus [68].

Due to the localization of the electron-nuclear system, its long coherence times, and controllability the Si:P system was one of the first candidates proposed for the construction of a scalable quantum computer [45]. Much work has been devoted towards making this a reality with precise single donor implantation [28] and control [85] having been demonstrated. A vital component of any quantum computer is the ability to reliably and rapidly initialize the system in a pure state. Which requires that the system be thermalized at mK temperatures in a moderately high magnetic field such that it approaches a pure nuclear state as its equilibrium distribution[75]. The characteristic time over which this process occurs is the spin-lattice relaxation time, T_1 and can be on the order of hours at low temperatures [33].

A possible solution to circumvent the long relaxation times for low-temperature solid-state nuclei is to use DNP to pump the system into its ground state [75]. A variety of DNP methods have been demonstrated for the Si:P system including resonant bound exciton transitions and a microwave (MW) induced Overhauser effect [94, 43]. Perhaps the least understood hyperpolarization mechanism is due to the illumination of the silicon crystal by above band-gap light [58]. The polarization technique is easy to use and consists of shining above-bandgap light directly onto a doped silicon crystal at low temperatures, and high magnetic field. The mechanism behind the above band-gap polarization is not well understood. There have been multiple conflicting proposals in the literature such as phonon induced oscillations of the electron-nuclear hyperfine interaction[67, 58] and bound exciton angular momentum conservation via spin-orbit coupling [72]. A reliable understanding of this mechanism might open the door to its use as an initialization technique in quantum devices such as qubits, magnetometers, and gyroscopes.

1.2 Outline and Results

This thesis will focus on a series of NMR experiments performed on the Si:P system in isotopically purified silicon. These experiments were performed in collaboration with graduate students Holger Haas and Rahul Deshpande.

Here the Si:P system is introduced and a brief introduction to the NMR theory required for interpreting the results of this thesis. In chapter 2 the phononic cross-relaxation mechanism of Pines, Bardeen, and Slichter which is the critical component of McCamey's above bandgap phosphorus hyperpolarization process [58] is introduced. This mechanism and an alternative spin exchange mechanism proposed by Sekiguchi *et al.*, which is a consequence of angular momentum conservation during the free exciton capture event [72] are compared and contrasted. The bound exciton process and resultant Auger ionization of the Si:P donor electron is introduced, and the detectable consequences of the hyperpolarization being a result of angular momentum conservation are noted. In chapter 3 the experimental setup and its capabilities are given, and a series of experiments and their accompanying results are described in chapter 4. These provide the first direct observation of the auger ionization process with NMR, the rates of ionization and recombination processes are extracted and provide evidence supporting the phononic hyper-polarization model. Finally, in chapter 5 conclusions are given, and a discussion of future work is presented.

1.3 Description of the Si:P System

Due to the isolated nature of the Si:P system for the isotopically purified silicon sampled studied which has a phosphorus concentration of $1.5 \times 10^{15}\text{cm}^{-3}$ and 99.9954% silicon-28 composition, the description of the system is relatively simple consisting of two coupled spin 1/2 particles. At high fields, the Hamiltonian is a sum of three terms: two Zeeman terms $\hat{H}_Z^e$ and $\hat{H}_Z^n$ describing the donor electron and phosphorus nuclei, respectively, and an isotropic hyperfine component $\hat{H}_{HF}$ coupling the two spins

$$\hat{H} = \hat{H}_Z^e + \hat{H}_Z^n + \hat{H}_{HF}. \tag{1.1}$$

The Zeeman Hamiltonian term describes the coupling of spins' magnetic moment to an external magnetic field

$$\hat{H}_Z^i = -\hat{\boldsymbol{\mu}}_i \cdot \boldsymbol{B_0}, \tag{1.2}$$

where $\hat{\boldsymbol{\mu}}_i$ is the magnetic dipole moment of the respective spin and $\boldsymbol{B_0}$ is the externally applied static magnetic field. The effect of the Zeeman component is to induce an energy splitting between the spin-up $|\uparrow\rangle$ and spin-down $|\downarrow\rangle$ eigenstates. The electron spin is represented with single-barred arrows $|\uparrow\rangle(|\downarrow\rangle)$ and the nuclear spin with double-barred arrows $|\Uparrow\rangle(|\Downarrow\rangle)$. Product states may be conjugated into a single ket $|\uparrow\rangle \otimes |\Uparrow\rangle = |\uparrow\Uparrow\rangle$. In practice it is easier to work with the spin angular momentum operators which obey the relation

$$\boldsymbol{\mu}_i = \gamma_i \hat{\boldsymbol{S}}, \tag{1.3}$$

where γ_i is the gyromagnetic ratio and describes the relationship between the spins magnetic moment and angular momentum. For convenience all Hamiltonians will be written with $\hbar = 1$. The applied quantizing static field $\boldsymbol{B}$ is assumed to be aligned along the z-axis giving the Hamiltonian

$$\hat{H}_Z^i = -\gamma_i B_0 \hat{S}_z. \tag{1.4}$$

The Larmor frequency is defined as $\omega_i = \gamma_i B_z$, this corresponds to the angular frequency at which a spin vector placed in the transverse plane will circumnavigate the Bloch sphere. This thesis will differentiate between electron and nuclear spins by using the $\hat{\boldsymbol{S}}$ and $\hat{\boldsymbol{I}}$ angular momentum operators, respectively. The gyromagnetic strengths for the electron and phosphorus nuclei are $\gamma_e/2\pi$=-28.024 GHz/T and $\gamma_n/2\pi$=17.23 MHz/T, respectively [33].

The Si:P system approximately corresponds to that of a hydrogenic atom [95] with a screened central potential. Consequently the donor electron spatial wavefunction is well described by effective mass theory and is approximately an s-orbital state with a modulating envelope composed of the six silicon Bloch function minima imposed by the crystalline C_{3v} symmetry [95]. The electron wavefunction's ground state is nonzero at the phosphorus nucleus due its s-character. The coupling between the electron and phosphorus spins consists entirely of a Fermi contact interaction [75, 62]

$$\hat{H}_{hf} = \frac{8\pi}{3} \hbar \gamma_e \gamma_n \hat{\boldsymbol{S}} \cdot \hat{\boldsymbol{I}} |\Psi(0)|^2, \tag{1.5}$$

where $|\Psi(0)|^2$ is the magnitude squared of the electron wavefunction at the location of the phosphorus nucleus. Similar to the Zeeman terms, a frequency associated with the hyperfine interaction strength may be defined

$$\hat{H}_{hf} = \frac{2\pi}{\hbar} A \hat{\boldsymbol{S}} \cdot \hat{\boldsymbol{I}}. \tag{1.6}$$

where $A = \frac{8\pi}{3}\hbar\gamma_e\gamma_n|\Psi(0)|^2$. In the Si:P system A =117.54 MHz [33].

The full system Hamiltonian may be written as

$$\hat{H} = -\omega_e \hat{S}_z + \omega_n \hat{I}_z - 2\pi A \hat{\boldsymbol{S}} \cdot \hat{\boldsymbol{I}}. \tag{1.7}$$

The experiments described in this thesis were performed at a field of B_0 =6.69 T. At this field the Larmor frequencies of the donor electron and phosphorus nuclei are $\omega_e/2\pi$=187.5 GHz and $\omega_n/2\pi$=115.30 MHz, respectively. Due to the electron's large Larmor frequency the components of the hyperfine interaction $\hat{S}_x\hat{I}_x$ and $\hat{S}_y\hat{I}_y$, which do not commute with $\hat{H}_Z^e$,

Temperature (K)	e^- Polarization	^{31}P Polarization
4	79%	0.09%
1.25	99.8%	0.33%

Table 1.1: Electron and nuclear spin polarizations at experimental temperatures where the hyperfine interaction has been accounted for.

will effectively be averaged to zero and have negligible effect on the system dynamics [75]. The system Hamiltonian may be written as

$$\hat{H} = -\gamma_e B_z \hat{S}_z + \gamma_n B_z \hat{I}_z - 2\pi A \hat{S}_z \hat{I}_z. \tag{1.8}$$

In high fields the eigenstates of the system are approximately described by the product state basis of $\{|1\rangle \approx |{\downarrow}{\Uparrow}\rangle, |2\rangle = |{\downarrow}{\Downarrow}\rangle, |3\rangle \approx |{\uparrow}{\Uparrow}\rangle, |4\rangle = |{\uparrow}{\Downarrow}\rangle\}$. The effect of the hyperfine interaction is to modify the nuclear Zeeman splitting between the lower and upper electron manifolds, which may be seen by rewriting the Hamiltonian in the form

$$\hat{H} = -\omega_e \hat{S}_z + (\omega_n - \pi A) |{\uparrow}\rangle \langle{\uparrow}| \otimes \hat{I}_z + (\omega_n + \pi A) |{\downarrow}\rangle \langle{\downarrow}| \otimes \hat{I}_z. \tag{1.9}$$

From (1.9) one can see that in the spin down manifold where at low temperatures the vast majority of electron polarization resides, the nuclear spin energy levels splitting acquires an additional amount of $A/2$. Define $\omega_{n+} = \omega_n + \pi A/2$, and $\omega_{n-} = \omega_n - \pi A/2$ to denote the spin down, and spin up electron manifold hyperfine shifted nuclear Larmor frequencies. Rewriting the system Hamiltonian in this parametrization

$$\hat{H} = -\omega_e \hat{S}_z - \omega_{n-} |{\Uparrow}\rangle \langle{\Uparrow}| \otimes \hat{I}_z - \omega_{n+} |{\Downarrow}\rangle \langle{\Downarrow}| \otimes \hat{I}_z. \tag{1.10}$$

The thermal equilibrium system state is given by the Boltzmann distribution [75]

$$\hat{\rho}_{eq} = e^{-\frac{\hat{H}}{k_b T}}/Z, \tag{1.11}$$

where k_b is Boltzmann's constant, T is the temperature of the system, and $Z = Tr[e^{-\frac{\hat{H}}{k_b T}}]$ is the partition function. The polarizations at a field of B_0 =6.71 T have been calculated in table 1.1. As the electron is nearly fully polarized at 1.25 K, experiments are performed in the $|{\Downarrow}\rangle$ manifold in order to maximize the observed signal. It is important to note that even at 1.25 K the nuclear polarization is practically negligible. As experiments are performed with a low ^{31}P concentration silicon sample nuclear thermal populations would not be observable. As a result DNP mechanisms are absolutely critical to performing inductively detected NMR with ^{31}P defect nuclei in silicon. The T_1 of the ^{31}P nuclei has previously been observed to be on the order of minutes at 4.2 K extending to many hours at 1.7 K such that it has not been possible to observe magnetization decay with the experimental setup [33].

1.4 Dynamics

The unitary dynamics of the Si:P system evolution may be determined from the Liouville-von Neumann equation [75]

$$\frac{d\hat{\rho}(t)}{dt} = -i[H, \hat{\rho}(t)]. \tag{1.12}$$

Note that any identity component in $\hat{\rho}$ will commute with $\hat{H}$ at all times and they need not be considered. Since the ^{31}P nuclei exist within a silicon crystal lattice, a variety of environmentally induced non-unitary dynamics are introduced, which may be included into the Lindblad master equation provided they are Markovian [53].

$$\frac{d\hat{\rho}(t)}{dt} = \hat{\mathscr{L}}\hat{\rho}(t), \tag{1.13}$$

here $\mathscr{L}$ is the Lindbladian and may be separated into its unitary and dissipative dynamics

$$\mathscr{L}\hat{\rho} = -i\left[\hat{H}, \hat{\rho}\right] + \hat{\mathscr{D}}(\hat{\rho}). \tag{1.14}$$

The dissipator $\hat{\mathscr{D}}(\hat{\rho})$ has the form

$$\hat{\mathscr{D}}(\hat{\rho}) = \frac{1}{2}\sum_{l=0}^{d^2-1} \alpha_l \left(\hat{A}_l \hat{\rho} \hat{A}_l^\dagger + \left\{\hat{A}_l^\dagger \hat{A}_l, \hat{\rho}\right\}\right). \tag{1.15}$$

The Si:P system experiences two kinds of dissipative dynamics. The first being thermal relaxation towards the equilibrium state, which is represented as an T_1 process with a rate $\Gamma_1 = 1/T_1$ and a dissipator

$$\hat{A}_0 = (1 - p_\uparrow)\sqrt{\Gamma_1}\hat{\sigma}_- + p_\uparrow\sqrt{\Gamma_1}\hat{\sigma}_+. \tag{1.16}$$

Where $p_\uparrow$ is the thermal occupation probability of the excited state. The T_1 process physically corresponds to the system exchanging energy with a connected environment in a non-reversible manner. In the Si:P system T_1 relaxation time (T_1) relaxation primarily affects the donor electron. The primary relaxation mechanism for the ${}^{31}P$ nuclei is due to electron-nuclear cross relaxation and at low temperatures is on the order of hours [74, 33]. As a consequence of the ${}^{31}P$ relaxation occurring over durations much greater than experiment lengths it may be neglected.

The second dissipative mechanism that plays a role in the dynamics is pure dephasing or phase damping. Phase damping may be associated solely with a decay of the off-diagonal density matrix terms, as opposed to amplitude damping, which may also affect diagonal terms. The phase damping dissipator has the form

$$\hat{A}_1 = \sqrt{\frac{\Gamma_2}{2}}\hat{\sigma}_z. \tag{1.17}$$

It should be noted that amplitude damping will introduce an effective dephasing process with a rate $\Gamma_2^{eq} = \Gamma_1/2$. In the case where multiple phase damping processes are present, they may be combined into an effective dephasing time constant

$$T_2^{(\mathrm{eff})} = \left(\sum_{i=0}^{n} \Gamma_2^{(i)} \right)^{-1} . \tag{1.18}$$

The dephasing rate T_2 is a measure for the lifetime of superposition states of energy eigenstates, and is the primary parameter of interest for both error rates in quantum computing [63] and sensitivities of quantum sensors [14].

For Si:P a major source of donor electron dephasing has been shown to be from nearby donor electrons and nearby silicon-29 nuclei. The impact of donor electron dephasing may be easily controlled by reducing the doping concentrations. The downside is that this will also reduce the number of nuclei available for detection, resulting in the need for DNP techniques to enhance the net nuclear magnetization [5]. An additional source of line broadening for the phosphorus nuclei is due to lattice strain resulting from local silicon isotopic mismatch. Reducing concentrations of silicon-29 (^{29}Si) from its 4.7% natural abundance to ultra-high purity levels of (99.99% > silicon-28 (^{28}Si))[13] provides an immense reduction in dephasing and has opened a pathway towards ultra-long lifetime Si:P devices. Coherence times may be further increased by utilizing decoupling sequences to remove the effect of the remaining ^{29}Si [68].

In the regime of low doping concentrations and isotopically purified samples, the remaining principal dephasing mechanism for the ^{31}P nucleus is the induced T_2 caused by its donor-bound electron. The effect will occur due to an accumulation of random phase by the ^{31}P nucleus as the electron rapidly fluctuates between its spin manifolds as a consequence of its short T_1e [33]. The electron induced T_2 may be removed entirely by ionizing the ^{31}P and ejecting the donor electron into the conduction band. Extraordinary coherence times of several hours have been measured for ionized ^{31}P in conjunction with decoupling sequences at low temperatures, and up to 39 minutes at room temperature [68]. The ionized state of the Si:P system in isotopically purified lowly doped silicon crystals is noise free to such a degree that it is often colloquially described as existing in a semiconductor vacuum.

1.5 Resonant Control

Nuclei are controlled with the application of resonant radio frequency (RF) pulses applied perpendicular to the static field direction [75] with Hamiltonian of the form

$$\hat{H}_{rf} = -\gamma_n B_x(t) \cos(\omega_{rf} t + \phi) \hat{I}_x = -\Omega(t) \cos(\omega_{rf} t + \phi) \hat{I}_x, \tag{1.19}$$

where $B_x(t)$ is the amplitude modulation function ($dB_x(t)/dt \ll \omega_n$) of the RF field and $\Omega(t) = \gamma_n B_x(t)$. Taking the static Hamiltonian (1.1) and entering a doubly rotating frame

$$H_{rot} = -\omega_e \hat{S}_z - \omega_{rf} \hat{I}_z, \tag{1.20}$$

and applying the rotating wave approximation (RWA) to discard the double frequency terms [75] arriving at

$$\hat{H}^{(rf)} = -\delta\omega_{n-} \left|\Uparrow\right\rangle \left\langle\Uparrow\right| \otimes \hat{S}_z - \delta\omega_{n+} \left|\Downarrow\right\rangle \left\langle\Downarrow\right| \otimes \hat{S}_z + \Omega(t)(\cos(\phi)\hat{I}_x + \sin(\phi)\hat{I}_y), \tag{1.21}$$

where $\delta\omega_{n\pm} = \omega_{n\pm} - \omega_{rf}$. Note that the static field dynamics of the electron are no longer present in this frame. There is no coupling matrix element between the electron spin up(down) manifolds. The action of the applied RF field on the ^{31}P nuclei in the rotating frame is to cause the nuclear spin to precess around a new effective field with Larmor frequency $\omega_{eff}^{(\pm)} = \sqrt{\delta\omega_{n\pm}^2 + \Omega^2}$, and tipping angle $\theta = \text{atan}\left(\frac{\Omega}{\delta\omega_{n\pm}}\right)$. If the pulse is applied directly on resonance with the electron spin down nuclear transition ($\delta\omega_{n+} = 0$) the ^{31}P spin will undergo Rabi oscillations along the axis defined by the control field phase. In the opposing electron spin manifold the effect will be negligible provided $\frac{\Omega}{\delta\omega_{n-}} \ll 0$. The addition of the control Hamiltonian (1.19) introduces the ability to coherently drive the ^{31}P nuclear spin to any position on the Bloch sphere.

1.6 Signal Detection

^{31}P spins in silicon have been previously detected in a variety of manners, with the vast majority using the bound donor electron to indirectly detect the ^{31}P polarization at the surface, and in the bulk. Initial observations in the bulk were performed using continuous wave (CW) ESR measurements [22]. More recently electrically detected magnetic resonance (EDMR)[61] and auger electron detected magnetic resonance (AEDMR) [94] techniques have been used to readout the spin state near the crystal surface.

Direct inductive detection of the ^{31}P nuclei is performed which allows the spin state in bulk to be observed. Inductive detection has the additional benefit that the ^{31}P spin may still be observed when the nucleus is ionized; ESR, EDMR, and AEDMR require the donor electron be bound to the ^{31}P nucleus in order to read out the nuclear spin state.

Precessing spins in the transverse plane induces an electromotive force (EMF) across the NMR probe coil [40]. The induced EMF is observed by first mixing down the signal with a reference frequency near the Larmor frequency of the sample, filtering the doubly rotating terms, and then detecting the resulting signal in quadrature [29]. The observed spin measurement operator is $\hat{I}_+ = \hat{I}_x + i\hat{I}_y$. It should be noted that all measurements performed in this work are ensemble, and not projective measurements, consequently all measurements give expectation values.

1.7 Bloch Equation Description

Provided there is no coherent exchange between the lower and upper electron manifolds the ^{31}P system dynamics in the lower manifold may be traced over, allowing a simple description of the Si:P system in terms of a single-spin Hamiltonian with an effective T_2^* due to the electron. The dynamics in the rotating frame is described by athe Bloch equation representation [75].

$$\frac{dm_x(t)}{dt} = +\delta\omega_{n+}\ m_y(t) + \Omega m_z(t) - \frac{m_x(t)}{T_2}, \tag{1.22a}$$

$$\frac{dm_y(t)}{dt} = -\delta\omega_{n+}\ m_x(t) - \frac{m_y(t)}{T_2}, \tag{1.22b}$$

$$\frac{dm_z(t)}{dt} = -\Omega\ m_x(t) + \frac{(m(\infty) - m_z(t))}{T_1}. \tag{1.22c}$$

Where m_x, m_y and m_z correspond to the expectation values of measurement operators $\hat{I}_x$, $\hat{I}_y$ and $\hat{I}_z$ respectively. The steady state polarization is determined by the parameter $m(\infty)$. In ensemble systems these expectation values correspond to the total nuclear magnetizations of the sample under study M_x, M_y and M_z. In practice, the T_1 of the ^{31}P nucleus at low temperatures is effectively infinite with respect to experiment times and is neglected. The DNP process may be modeled as an effective T_1 process with time constant T_1^{DNP}.

1.8 Si:P Hyperfine Interaction

The hyperfine interaction between the phosphorus nucleus and the donor electron is dominated by the Fermi contact interaction, [67] and its strength is proportional to electron and nuclear wavefunction overlap

$$A \propto |\psi_e(0)|^2. \tag{1.23}$$

When modelling the Si:P hyperfine interaction care must be taken to account for the silicon lattice in which the system resides. The donor electron wavefunction may be modelled within the framework of effective mass theory [48], in which the electron wavefunction is assumed to be a linear combination of the silicon conduction band Bloch functions with an effective mass due to the silicon band structure. Symmetry analysis of the system allows the existence of A_1 singlet, E doublet and T_2 triplet wavefunction states. Only the A_1 singlet state, which is an equal superposition of the six band minima, has a nonzero wavefunction amplitude at the position of the phosphorus nucleus, and consequently a nonzero Fermi contact hyperfine interaction. Many experiments have confirmed this singlet state to be the ground state of the orbital electron wavefunction [95].

Wilson and Ferher demonstrated that crystal strain will result in a quadratic reduction in the hyperfine interaction strength when a uniaxial stress along the [100] crystal direction is applied [90]. Hasegawa used perturbation theory to calculate that under the valley recombination model (VRM) model the effect of crystal strain on the hyperfine interaction strength is to shift it by

$$\delta A = -A\left(\frac{\Xi_u}{3E_{12}}\right)^2\left(\frac{3}{2}(\epsilon_{xx}-\epsilon_{yy})^2+\frac{1}{2}(\epsilon_{xx}+\epsilon_{yy}-2\epsilon_{zz})^2\right), \tag{1.24}$$

where $\Xi_u = 8.6\ eV$ is the deformation of pure shears, $E_{12} = 12.95\ eV$ the energy separation of the electron ground and excited states, $\{\epsilon_{ii}\}$ the axial crystal strains and $\{\epsilon_{ij}\}$ the shear strains in the crystallographic coordinate system [37]. The reduced hyperfine interaction is attributed to the mixing of the singlet ground state with the excited doublet state under crystal strain. As the doublet component has no overlap with the nucleus it will have no Fermi contact interaction. Furthermore, they demonstrated how a time dependent modulation of strain (phonons) may lead to a flip-flop interaction between the electron and the phosphorous spin degrees of freedom. Pines, Bardeen and Slichter showed how the flip-flop interaction in conjunction with a rapid electron relaxation rate, non-thermal equilibrium nuclear populations may be achieved [67].

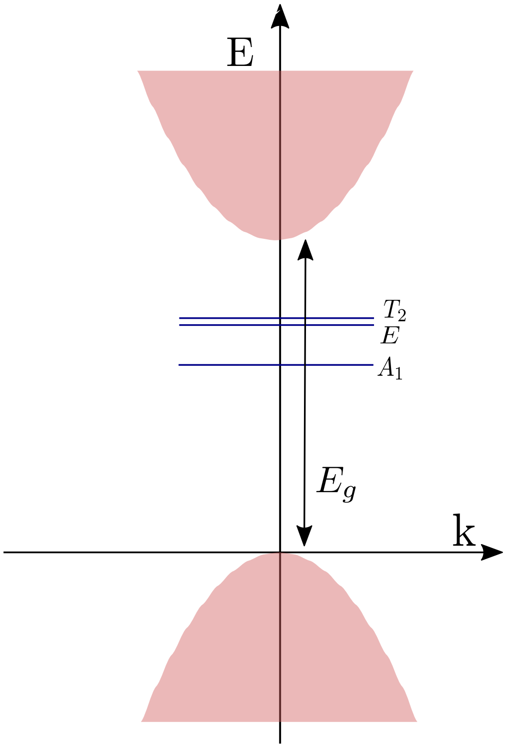


Figure 1.1: Spatial donor electron wavefunction energy levels.

1.9 Phononic Hyperpolarization Process

Phonon induced electron-nuclear cross relaxation of the Si:P system has long been considered as a potential mechanism for achieving non-thermal nuclear polarizations outside of the traditional Overhauser effect [64]. The original proposal dates back to Pines, Bardeen and Slichter [67] with further exploration by Nakayama [62] which suggested that an electron-nuclear cross-relaxation mechanism $T_x^{(en)}$ in conjunction with an electron relaxation mechanism $T_1^{(e)}$ will drive the nuclear polarization towards the steady state electron thermal polarization as visualized in figure 1.2. Feher experimentally confirmed the ability to achieve such spin-spin cross relaxation with the injection of hot conduction electrons [23]. More recently McCamey *et. al* attribute their observations of non-thermal nuclear polarizations, achieved with the illumination of above bandgap light, to this mechanism [58]. An approximate semiclassical treatment of this process is given in [39] which is followed closely, with a focus on providing a clear exploration of the mechanism and therefore do not consider the full symmetries of the crystal strain on the hyperfine interaction as described in section 1.8.

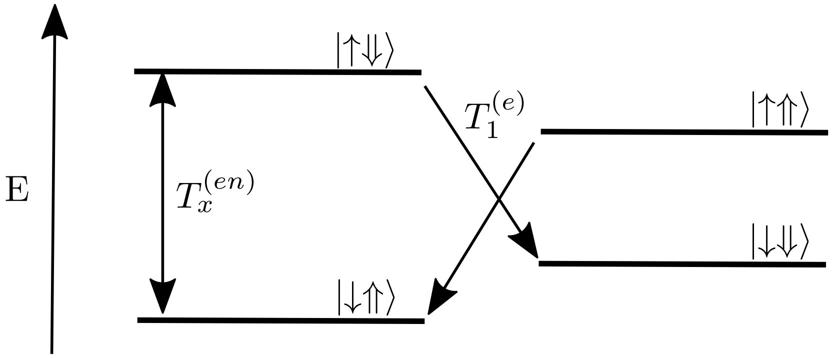


Figure 1.2: Nuclear non-thermal polarization is induced by the combination of an electron-nuclear cross-relaxation process in combination with simultaneous electron relaxation.

The process considered is the result of above bandgap photons being absorbed by the silicon crystal, which generates free electron-hole pairs in the conduction and valence bands. As silicon is an indirect bandgap material the absorption process will involve the emission or absorption of a phonon(s) such that energy and crystal momentum is conserved [95]. The electrons and holes will eventually recombine and give up their kinetic energy through a series of cascading one-phonon scattering events with attractive lattice impurities generating predominantly acoustic phonons at low temperatures [49]. The phonon bath corresponds to a spectrum of periodic crystal strains which will modify the Si:P hyperfine coupling.

Considering second order strain perturbations of the Si:P hyperfine interaction of the form

$$A \to A(1-\nu\epsilon^2)\hat{\boldsymbol{S}}\cdot\hat{\boldsymbol{I}}, \tag{1.25}$$

where $\epsilon = \sum_{i=1}^{n} \sqrt{I}\epsilon_i \cos(\omega_i + \phi_i)$ is the resultant strain field at the site of the phosphorus donor, I is the illumination intensity, ϵ_i the photon to phonon conversion ratio for a phonon mode ω_i, and ϕ_i the uniformly distributed random phase. The root dependence on intensity is a result of the net strain being the sum of many individual phononic modes with uniform random phase. Finally, ν is the strain-hyperfine dependent modulation strength. By writing the interaction in this way, several assumptions are made including the adiabatic approximation, that the motion of the electron wavefunction under the strain field is instantaneous [1]. Only the secular portion of the hyperfine interaction equation (1.25) is considered which can be written as

$$A\hat{S}_z\hat{I}_z + \frac{IA\nu}{4}\left(\sum_{\substack{\omega_i+\omega_j=\\ \omega_e+\omega_n}} + \sum_{\substack{\omega_i-\omega_j=\\ \omega_e+\omega_n}}\right)\epsilon_i\epsilon_j\left(\hat{S}_x\hat{I}_x + \hat{S}_y\hat{I}_y\right). \tag{1.26}$$

Importantly, the second term in equation (1.26) is capable of transferring polarization between the electron and ^{31}P nucleus. Note that although presented as a coherent process, in reality this process will be averaged over the random photon absorption and electron-hole recombination processes removing any phase coherence at given donor sites. The surviving terms correspond to two phonon absorption and Raman scattering with the energy conservation conditions being $\omega_i + \omega_j = \omega_e + \omega_n$ and $\omega_i - \omega_j = \omega_e + \omega_n$, respectively [62]. It is known that modulation of the spin-orbit coupling at frequency $\omega_i \approx \omega_e$ is responsible for the electron relaxation rate $T_1^{(e)}$ due to the electron g-factor's linear dependence on crystal strain fields [91]. Under these conditions the interaction with rapidly thermalized free carrier conduction band electrons is likely to be a much stronger contributor to the electron thermalization rate [24]. The net result is an electron-nuclear cross-relaxation rate $T_x^{(en)}$ in conjunction with the electron relaxation rate $T_1^{(e)}$. For the Si:P system $T_1^{(e)} \ll T_1^{(n)}$ and consequently the nuclear relaxation rate is disregarded. Assuming effective bath temperatures Θ_1 and Θ_x for $T_1^{(e)}$ and $T_x^{(en)}$ respectively, the steady state electron and nuclear polarizations are $\rho_e = (2q-1)\hat{S}_z$ and $\rho_n = \frac{p-q}{p(2q-1)-q}\hat{I}_z$, where $p = \left(1 + e^{\hbar(\omega_e+\omega_n)/(k_B\Theta_x)}\right)^{-1}$ and $q = \left(1 + e^{\hbar(\omega_e/(k_B\Theta_x)}\right)^{-1}$ are thermal occupation probabilities for the $T_1^{(e)}$ and $T_x^{(en)}$ process Hamiltonians. If $\Theta = \Theta_x = \Theta_1$ the expected thermal polarization of the separable electron-nuclear system of $\tanh[\hbar\omega_e/(2k_B\Theta)]$ and $\tanh[\hbar\omega_n/(2k_B\Theta)]$ is obtained. However, when $\Theta_x \neq \Theta_1$ it is possible to achieve nuclear polarizations approaching that of the thermal electron polarization. In the limit $T_1^{(e)} \ll T_x^{(en)}$ the steady state polarization is approached with a time constant

$$T_{\text{DNP}} = \frac{T_x^{(en)}}{p+q-2pq}. \tag{1.27}$$

For large illumination intensity I it is expected that $p > q$, which drives the nuclear polarization towards the $|\downarrow\rangle$ state which is anti-aligned with its $T_1^{(n)}$ driven thermal polarization.

Chapter 2

Auger Ionization and Phononic Hyperpolarization

As outlined in chapter 1 a key component of interest for the Si:P system is the hyperpolarization mechanism under above bandgap illumination. A study of the mechanism is performed making use of the observation of modified buildup dynamics which are a consequence of the periodic ionization of the phosphorus donor. During polarization buildup, a low power (Hz-KHz) saturation Rabi drive is applied at the bare Phosphorus Larmor frequency as determined from liquid state measurements. Modified build up dynamics are observed, which depend strongly on the applied saturation power and frequency. The dynamics are attributed to the phosphorus being ionized and neutralized during the hyperpolarization process. If the saturation drive is sufficiently strong and close to resonance, it will tilt the phosphorus spin into the x-y plane, at which point the defect will be subsequently re-neutralized after some random period effectively destroying its coherence. The ionization process is attributed to bound exciton induced Auger ionization [70], which has been observed previously in Si:P under above bandgap illumination [94].

A more detailed explanation of the modelled mechanism follows, a semi-classical approach primarily utilizing modified Bloch Equations is given. The initial state is assumed to be a thermally polarized (to a good approximation 0%) phosphorus nuclear ensemble with a hyperpolarization rate $T_1^{D_h^0}$ due to the applied above bandgap illumination. It is important to note that the thermal and hyperpolarization T_1 processes steady state magnetizations have opposite signs, as the hyperpolarization mechanism effectively exchanges polarization with the thermally polarized electrons which have an inverted alignment with respect to the phosphorus nuclei. The system also experiences Larmor oscillations with rate $\omega_{^{31}P}$, and there is a strong effective T_2 process as discussed in 1.4 system. The reduced D^0 Hamiltonian in the frame rotating at the saturation frequency ω_Ω is given by

$$\frac{\mathscr{H}_{D^0}}{2\pi} = \Omega \hat{I}_x + \left(\frac{A_e}{2} - \Delta v\right)\hat{I}_z, \tag{2.1}$$

where Ω is the applied Rabi drive, A_e is the electron hyperfine shift, and $2\pi\Delta v = \omega_{^{31}P} - \omega_{\Omega}$ is the difference between the drive and the bare phosphorus resonance frequencies. Note that as $\Omega \ll \left(\frac{A_e}{2} - \Delta v\right)$ the Hamiltonian is well approximated by

$$\frac{\mathscr{H}_{D^0}}{2\pi} \approx \left(\frac{A_e}{2} - \Delta v\right)\hat{I_z}. \tag{2.2}$$

Define the free exciton capture rate $\kappa_{D^0 \to D^0X}$. Due to the Pauli exclusion principle, it is required that the donor-bound electron and free exciton electron form a singlet state. Such a state will display no hyperfine splitting. However, it is possible that the hyperfine shift due to the hole could be observed. There has yet to be a direct measurement of such a shift, although the expected strength has been estimated to be in the range 2-20 MHz [93]. A refined technique could be used to measure the hole hyperfine splitting by sweeping the saturation tone over the expected hole Fermi contact interaction strength range and looking for a signature of reduced observed magnetizations. Assuming that the capture process occurs on a much quicker time scale than the spin evolution timescale $\omega_{^{31}P}$, that there is negligible relaxation $T_1^{D_0X}$ in the bound exciton state, and that the bound exciton dephasing rate $T_2^{D^0X}$ is negligible, the Hamiltonian dynamics in this manifold are modelled as

$$\frac{\mathscr{H}_{D^0X}}{2\pi} = \Omega\hat{I}_x + \left(\frac{A_h}{2} - \Delta v\right)\hat{I}_z, \tag{2.3}$$

where A_h is the hole hyperfine strength. The bound exciton will decay, with a recombination of the hole and electron at a rate $\kappa_{D^0X \to D^+}$. There are two primary decay channels: the energy may be released as a photon, or approximately 10000 times more frequently it will ionize the phosphorus donor electron through the Auger ionization process [70]. As the Auger process is four orders of magnitude more likely, only this decay channel is considered. After ionization, the phosphorus spin is left in an effective semiconductor vacuum and therefore all T_2 processes are disregarded in its ionized state. In the ionized state, the ^{31}P nuclei are no longer hyperfine shifted and remain at their bare Larmor frequency $\omega_{^{31}P}$, perhaps shifted by some small paramagnetic shift δ due to shielding from nearby silicon atoms. A saturation Rabi drive of strength Ω at frequency ω_Σ will induce Rabi oscillations on the D^+ nuclei. The Hamiltonian in this frame is

$$\frac{\mathscr{H}_{D^+}}{2\pi} = \Omega\hat{I}_x + (\delta - \Delta v)\,\hat{I}_z. \tag{2.4}$$

At a rate $\kappa_{D^+ \to D^0}$, the D^+ state will be neutralized via the acquisition of a free electron from the conduction band [80], thus reverting to the initial Hamiltonian, equation (2.1) and dissipative processes. As the recapture event is a telegraph process, provided $\kappa_{D^+ \to D^0} \ll \frac{A_e}{2}$ an individual ^{31}P will acquire a uniform random phase upon returning to the D^0 state. Averaging over this uniform phase for all donor nuclei will effectively destroy all transverse magnetization. This results in modified build up dynamics under applied saturation near the ionized phosphorus resonance frequency, with an observed reduction in the steady-state magnetization.

As noted in Chapter 1, one can equivalently model the system utilizing the Lindblad master equation, or the Bloch equations. For simplicity, the dynamics are described as a chemical exchange process using a series of coupled Bloch equations, otherwise known as the Bloch-McConnel equations [6]. These equations are given below

$$\frac{dm_x^{D^0}}{dt} = -\left(\frac{A_e}{2} - \Delta v\right) m_y^{D^0} - \left(\kappa_{D^0 \to D^0 X} + \frac{1}{T_2^{D^0}}\right) m_x^{D^0}, \tag{2.4a}$$

$$\frac{dm_y^{D^0}}{dt} = +\left(\frac{A_e}{2} - \Delta v\right) m_x^{D^0} - \left(\kappa_{D^0 \to D^0 X} + \frac{1}{T_2^{D^0}}\right) m_y^{D^0}, \tag{2.4b}$$

$$\frac{dm_z^{D^0}}{dt} = -\frac{(m(\infty) - m_z^{D^0})}{T_1} - \kappa_{D^0 \to D^0 X} m_z^{D^0} + \kappa_{D^+ \to D^0} m_z^{D^+}, \tag{2.4c}$$

$$\frac{dm_x^{D^0 X}}{dt} = -\left(\frac{A_h}{2} - \Delta v\right) m_y^{D^0 X} - \left(\kappa_{D^0 X \to D^+} + \frac{1}{T_2^{D^0 X}}\right) m_x^{D^X} + \kappa_{D^0 \to D^0 X} m_x^{D^0}, \tag{2.5a}$$

$$\begin{aligned} \frac{dm_y^{D^0 X}}{dt} &= \left(\frac{A_h}{2} - \Delta v\right) m_x^{D^0 X} + \Omega m_z^{D^0 X}, \\ &- \left(\kappa_{D^0 X \to D^+} + \frac{1}{T_2^{D^0 X}}\right) m_y^{D^0 X} + \kappa_{D^0 \to D^0 X} m_y^{D^0}, \end{aligned} \tag{2.5b}$$

$$\frac{dm_z^{D^0 X}}{dt} = -\Omega m_y^{D^0 X} - \kappa_{D^0 X \to D^+} m_z^{D^0 X} + \kappa_{D^0 \to D^0 X} m_z^{D^0}, \tag{2.5c}$$

$$\frac{dm_x^{D^+}}{dt} = -\left(\delta - \Delta v\right) m_y^{D^+} - \kappa_{D^+ \to D^0} m_x^{D^+} + \kappa_{D^0 X \to D^+} m_x^{D^0 X}, \tag{2.6a}$$

$$\frac{dm_y^{D^+}}{dt} = +\left(\delta - \Delta v\right) m_x^{D^+} + \Omega m_z^{D^+} - \kappa_{D^+ \to D^0} m_y^{D^+} + \kappa_{D^0 X \to D^+} m_y^{D^0 X}, \tag{2.6b}$$

$$\frac{dm_z^{D^+}}{dt} = -\Omega m_y^{D^+} - \kappa_{D^+ \to D^0} m_z^{D^+} + \kappa_{D^0 X \to D^+} m_z^{D^0 X}, \tag{2.6c}$$

where equations (2.4a-2.6c) describe the D^0, D^0X, and D^+ spin dynamics for the magnetizations of each state m_x, m_y, and m_z. Note that in the D^0X equations equation (2.4a) and in equation (2.4b) it is not possible to create magnetization in the transverse plane, as a result if the system is initialized without transverse components their dynamics may be dropped from the model. The entire process is visualized in figure 2.1.

As the modeled equations are a series of coupled linear first order differential, equations rewritten as a matrix equation they have the form

$$\frac{\partial}{\partial t}\vec{m} = \mathbf{A}\vec{m}, \tag{2.7}$$

where $\vec{m}$ is the total magnetization vector combining the D^0,D^0X and D^+ magnetization components

$$\vec{m} = \left(m_I^{D^0}, m_z^{D^0}, m_x^{D^0 X}, m_y^{D^0 X}, m_z^{D^0 X}, m_x^{D^+}, m_y^{D^+}, m_z^{D^+}\right)^T. \tag{2.8}$$

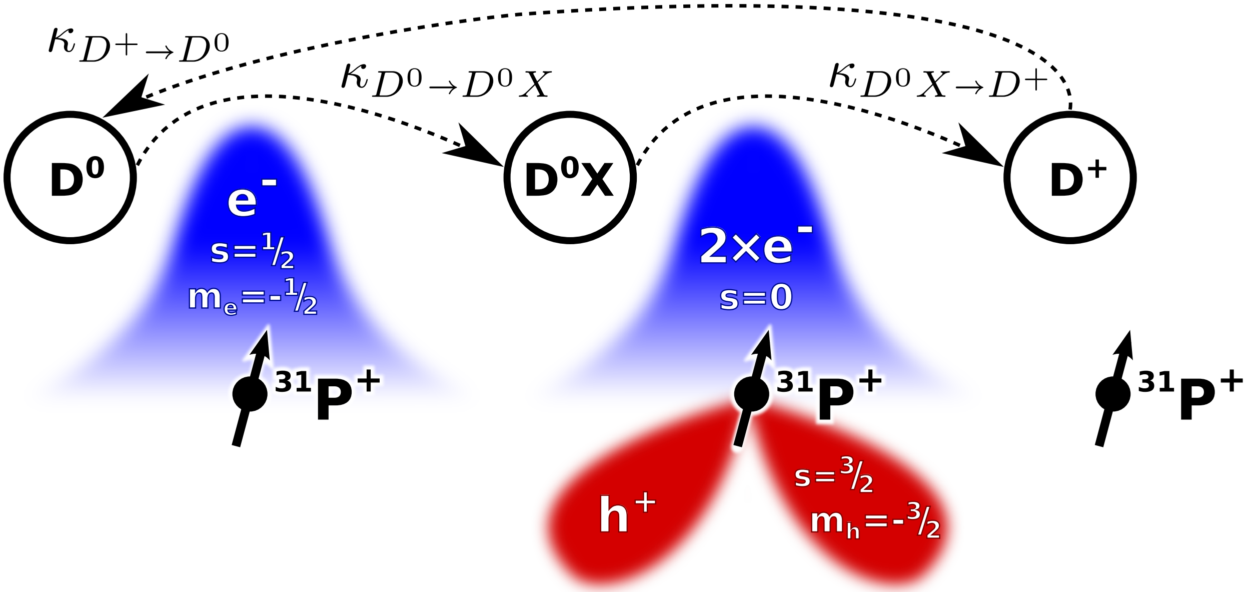


Figure 2.1: Visualization of the D^0 ionization process. D^0 captures a free exciton becoming D^0X. After some time the free exciton will decay via Auger ionization into the ionized D^+ state, where a resonant saturation drive will rotate the ^{31}P on the Bloch sphere. The ^{31}P nucleus will then capture a free electron from the conduction band returning to the neutralized D^0 state. Any created coherences will now rapidly be destroyed due to the electron mediated T_2 process. Image used with permission [39].

A is the transition matrix with elements

$$\mathbf{A} =
\begin{pmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\frac{m(\infty)}{T_1^{D^0}} & 0 & 0 & -\frac{1}{T_1^{D^0}}-\kappa_{D^0\to D^0X} & 0 & 0 & 0 & 0 & 0 & \kappa_{D^+\to D^0} \\
0 & 0 & 0 & 0 & -\kappa_{D^0X\to D^+} & -\frac{A_h}{2}+\Delta v & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \frac{A_h}{2}+\Delta v & -\kappa_{D^0X\to D^+} & \Omega & 0 & 0 & 0 \\
0 & 0 & 0 & \kappa_{D^0\to D^0X} & 0 & -\Omega & -\kappa_{D^+\to D^0} & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \kappa_{D^0X\to D^+} & 0 & 0 & -\kappa_{D^0X\to D^+} & -\delta+\Delta v & 0 \\
0 & 0 & 0 & 0 & 0 & \kappa_{D^0X\to D^+} & 0 & +\delta-\Delta v & -\kappa_{D^+\to D^0} & \Omega \\
0 & 0 & 0 & 0 & 0 & 0 & \kappa_{D^0X\to D^+} & 0 & -\Omega & -\kappa_{D^+\to D^0}
\end{pmatrix}. \tag{2.9}$$

Since the rate equation description found in equation (2.7) can easily be solved numerically, all data analysis is performed with the numerical solutions for the sake of accuracy. However, additional insight may be gained by solving the equations under further simplifying assumptions. In the limit where the bound exciton lifetime is short compared to the free exciton capture and donor neutralization rate $\kappa_{D^0X\to D^+} \gg \kappa_{D^0\to D^0X}, \kappa_{D^+\to D^0}$, the steady state magnetization has a Lorentzian dependence on both the saturation strength Ω and detuning from the paramagnetic shifted bare resonance frequency $\zeta = \delta - \Delta v$ given by

$$\lim_{t\to\infty} m_z^{D^0}(t) = \mu_z^{D^0} = \frac{\kappa_{D^+\to D^0}^2 + \zeta^2 + \Omega^2}{\kappa_{D^+\to D^0}^2 + \zeta^2 + \left(1 + \kappa_{D^0\to D^0X} T_1^{D^0}\right)\Omega^2}. \tag{2.10}$$

In equation (2.10) the observed signal has been normalized by $m(\infty)$ removing the explicit dependence on the absolute nuclear polarization. As expected, in the limit of a very weak saturation drive $\Omega \to 0$, free exciton capture $\kappa_{D^0\to D^0X} \to 0$, and large detuning $\zeta \to \infty$ a magnetization of unity is recovered. Under these conditions the time dependent buildup process is exponential

$$m_z^{D^0}(t) \approx \mu_z^{D^0}\left(1 - e^{-\frac{t}{\mu_z^{D^0} T_1^{D^0}}}\right). \tag{2.11}$$

It is observed that the approximate effect of the ionization process is to increase the effective hyperpolarization rate by a factor $\mu_z^{D^0}$. Evaluating the time derivative at the start of polarization ($t = 0$)

$$\left.\frac{\partial M_z^{D^0}}{\partial t}\right|_{t=0} = \frac{1}{T_1^{D^0}}, \tag{2.12}$$

it is apparent that their is no dependence on parameters relating to saturation, Ω, Δ, and $\kappa_{D^+\to D^0}$.

2.1 Alternative Hyperpolarization Process

In section 1.9, the phononic hyperpolarization process is introduced. Sekiguchi *et al.* have proposed an alternative process that predicts that the observed hyperpolarization is due to the conservation of net angular momentum during bound exciton formation[72]. The proposed polarization process is sketched below.

For a bound exciton ^{31}P system the electron spins in their ground state will be in an antiparallel singlet configuration due to the Pauli exclusion principle [79]. Conduction, free exciton, and phosphorus donor electrons will have thermal polarizations near unity at the magnetic field and temperature of study, while the nuclear thermal polarization is near zero. Consequently, system eigenstates $|1\rangle$ and $|2\rangle$ will have near equal thermal populations. While $|2\rangle = |\Downarrow\downarrow\rangle$ is a pure Zeeman state, $|1\rangle$ will be predominantly $|\Downarrow\uparrow\rangle$ with a small admixture of $|\Uparrow\downarrow\rangle$ due to the hyperfine mixing. To form a bound exciton D^0X state, an electron must flip its spin. While instances of $|0\rangle$ may only have their electron spin flip via spin-orbit coupling, Sekiguchi *et. al* argue that systems in state $|1\rangle$ may use the admixture component $|\Uparrow\downarrow\rangle$ to flip the bound exciton electron and form the D^0X state, which would also result in the flip of the nuclear spin for a net zero total spin change. After Auger ionization, the nuclei will recapture a new thermally polarized electron. Consequently, this process will drive the phosphorus polarization towards the thermal electron polarization and the system as a whole towards $|1\rangle$.

Crucially, this model has testable consequences. For bound exciton formation to proceed, the nuclear state must be anti-aligned with the electron. As the hyperpolarization process proceeds, the nuclear state will become increasingly polarized reducing the rate of exciton capture. While direct calculation of the free exciton-neutral impurity cross section by variational methods is possible [57, 71, 56], they are both complicated and agree poorly with experimental results. Of primary interest is discerning between the two polarizations mechanisms. As the formation of D^0X during the polarization process is directly observed, the polarization dependence of D^0X formation may be incorporated by letting the rate of D^0X formation linearly depend on the nuclear magnetization as in spin-exchange scattering [56]

$$\kappa_{D^0\to D^0X} \to \kappa_{D^0\to D^0X}\left(1 - m_z^{D^0X}\right). \tag{2.13}$$

The hyperpolarization process is shifted to the D^0X state, but the dependence on $m_z^{D^0}$ is kept, as both the capture and the polarization processes occur simultaneously. The full set of coupled nonlinear differential equations are found in equations (2.13a-2.15c).

$$\frac{dm_x^{D^0}}{dt} = -\left(\frac{A_e}{2} - \Delta v\right) m_y^{D^0} - \left(\kappa_{D^0\to D^0X}\left(1 - m_z^{D^0X}\right) + \frac{1}{T_2^{D^0}}\right) m_x^{D^0}, \tag{2.13a}$$

$$\frac{dm_y^{D^0}}{dt} = +\left(\frac{A_e}{2} - \Delta v\right) m_x^{D^0} - \left(\kappa_{D^0\to D^0X}\left(1 - m_z^{D^0X}\right) + \frac{1}{T_2^{D^0}}\right) m_y^{D^0}, \tag{2.13b}$$

$$\frac{dm_z^{D^0}}{dt} = -\kappa_{D^0\to D^0X}\left(1 - m_z^{D^0X}\right) m_z^{D^0} + \kappa_{D^+\to D^0} m_z^{D^+}, \tag{2.13c}$$

$$\begin{aligned}\frac{dm_x^{D^0X}}{dt} = &-\left(\frac{A_h}{2} - \Delta v\right) m_y^{D^0X} - \left(\kappa_{D^0X\to D^+} + \frac{1}{T_2^{D^0X}}\right) m_x^{D^X} \\ &+ \kappa_{D^0\to D^0X}\left(1 - m_z^{D^0X}\right) m_x^{D^0},\end{aligned} \tag{2.14a}$$

$$\begin{aligned}\frac{dm_y^{D^0X}}{dt} = &\left(\frac{A_h}{2} - \Delta v\right) m_x^{D^0X} + \Omega m_z^{D^0X} - \left(\kappa_{D^0X\to D^+} + \frac{1}{T_2^{D^0X}}\right) m_y^{D^0X} \\ &+ \kappa_{D^0\to D^0X}\left(1 - m_z^{D^0X}\right) m_y^{D^0},\end{aligned} \tag{2.14b}$$

$$\begin{aligned}\frac{dm_z^{D^0X}}{dt} = &-\frac{(m(\infty) - m_z^{D^0})}{T_1} - \Omega m_y^{D^0X} - \kappa_{D^0X\to D^+} m_z^{D^0X} \\ &+ \kappa_{D^0\to D^0X}\left(1 - m_z^{D^0X}\right) m_z^{D^0},\end{aligned} \tag{2.14c}$$

$$\frac{dm_x^{D^+}}{dt} = -\left(\delta - \Delta v\right) m_y^{D^+} - \kappa_{D^+\to D^0} m_x^{D^+} + \kappa_{D^0X\to D^+} m_x^{D^0X}, \tag{2.15a}$$

$$\frac{dm_y^{D^+}}{dt} = +\left(\delta - \Delta v\right) m_x^{D^+} + \Omega m_z^{D^+} - \kappa_{D^+\to D^0} m_y^{D^+} + \kappa_{D^0X\to D^+} m_y^{D^0X}, \tag{2.15b}$$

$$\frac{dm_z^{D^+}}{dt} = -\Omega m_y^{D^+} - \kappa_{D^+\to D^0} m_z^{D^+} + \kappa_{D^0X\to D^+} m_z^{D^0X}. \tag{2.15c}$$

Importantly, equation (2.10) holds for the bound exciton polarization model, which means that saturation swept resonance experiments will yield identical rate predictions for both models. To discern between the models, the buildup dynamics under saturation must be studied. This must be done numerically as the equations are nonlinear and do not present an analytic solution.

2.2 Light Distribution Dependence

The dynamics of the proposed model are strongly dependent on local light distribution variations over the silicon crystal. The free electron and exciton have been measured to have mean free paths of 10 um and 24 um respectively for similar temperatures and doping concentrations as used within the experiments described within this thesis [24, 19]. An additional factor to consider is that their respective production rates have different dependencies on applied light intensity. In silicon, free electrons are formed via photon absorption and then primarily recombine with holes, which is a bimolecular decay process. As a result, the saturated free electron density displays a $\sqrt{I}$ laser power dependence. Free excitons, on the other hand, are produced, via a bimolecular process and decay spontaneously with a constant rate, and therefore display a linear I laser power dependence [66, 86]. Consequently, the locally applied laser intensity must be accounted for in given model rates, which is accomplished by redefining the corresponding rate parameters to include explicit power dependence

$$\kappa_{D^0\to D^0X}(I) = \sqrt{I}\alpha_{D^0\to D^0X}, \tag{2.16a}$$

$$\kappa_{D^+\to D^0}(I) = I\alpha_{D^+\to D^0}. \tag{2.16b}$$

Note that $\alpha_{D^0\to D^0X}$ and $\alpha_{D^+\to D^0}$ have units of Hz/$\sqrt{W}$ and Hz/W, respectively. Reparameterizing equation (2.10) the intensity-dependent build up is given by

$$\mu_z^{D^0}(I) = \frac{(I\alpha_{D^+\to D^0})^2 + \Delta^2 + \Omega^2}{(I\alpha_{D^+\to D^0})^2 + \zeta^2 + \left(1 + \sqrt{I}\alpha_{D^0\to D^0X}T_1^{D^0}\right)\Omega^2}. \tag{2.17}$$

Assuming there is some light distribution over the sample $\rho(I)$, the saturated magnetization is now given by

$$\mu_z^{D^0} = \int_0^\infty \rho(I)\mu_z^{D^0}(I)dI \tag{2.18}$$

and the approximate buildup

$$\mu_z^{D^0}(t) = \int_0^\infty \rho(I)\mu_z^{D^0}(I)\left(1 - e^{-\frac{t}{\mu_z^{D^0}(I)T_1^{D^0}}}\right)dI \tag{2.19}$$

It is clear that the observed magnetization buildup will be the sum of many exponentials with the form of equation (2.10), with build up rates and final intensities determined by $\mu_z^{D^0}$.

The incident Gaussian beam from the laser is reflected from the NMR coil and forms a grating pattern, it is then focused by the insulating sapphire tube and finally enters the sample. The effect of reflections must also be taken into account. To determine the light distribution over the sample, the optical system was simulated with COMSOL ray tracing using 200000 rays. The laser power deposited in each 100 μm × 100 μm × 100 μm mesh element is extracted, and then bin and normalize the simulated intensities to generate an intensity distribution $\rho(I)$. Meshing dimensions were selected to be such that the element size was much larger than the exciton mean free path [19], while still being significantly smaller than the subterahertz phonon mean free path [35]. The simulated intensity pattern with accompanying intensity distribution is shown in figure 2.2. From simulations, the laser power absorbed per mesh element is calculated assuming a penetration depth of $\lambda = 26.9$ cm at 4.2 K as measured by Macfarlane *et. al* for 1047 nm light in silicon [55]. The estimated aggregate laser power deposited over the same at full laser power has been simulated to be 0.232 mW, this is equivalent to an input uniform intensity of 166 mW cm^{-2}. A diagram of the sample with surface incident laser intensity outside the sapphire isolation tube is depicted in figure 2.2, the grating pattern induced by the coil is observable.

2.2.1 Stark Ionization of the Phosphorus Defect

The D^+ ionized phosphorus defect in isotopically purified silicon is an extraordinarily noise free system. It has often been described as existing within a semiconductor "vacuum", with coherence times of up to hours being reached at low temperatures [68]. In the past ionization has been achieved with the use of narrow-band lasers tuned to the bound exciton transition. Such lasers are expensive and bulky, therefore, the technique is not well suited for ionization in microscopic devices such as gyroscopes or magnetometers.

An alternative mechanism to ionize the phosphorus defect is the application of a strong electric field across the silicon crystal. An electric field applied over a defect will shift the electron wavefunction density away from the defect nuclei through the Stark effect. As the field magnitude becomes large the ground $1s$-like electron state will anticross with the excited $2p$ state effectively allowing the electron to tunnel into the bulk and ionizing the donor past this threshold field. Although not explicitly measured for bulk phosphorus defects in silicon the ionization threshold has been estimated to be on the order of 1.55V/μm [65]. Achieving ionization in this way could yield devices that are simpler to fabricate and operate, as traditional lithography processes can be used to place electrodes rather than requiring a complex optical setup such as is required for Auger ionization.

Occurring in parallel with the bulk of the work within this thesis an effort has been ongoing to ionize the phosphorus donors within the experiment setup. Two capacitor plates have been added on either side of the silicon sample within the detection coil. It has been estimated with Comsol simulations that a voltage in excess of $4kV$ must be established across the two capacitor plates within the helium bath. This has proven to be a large experimental challenge as hinted at in section 3.0.2. The major difficulty is due to ionization of the helium bath. This is believed to be due to impurities within the helium and an effort is underway to engineer a superleak so as to purify the helium. It is important to note that if a method other than direct inductive detection of the phosphorus defect were to be used, a much smaller sample and consequently ionization field would be required.

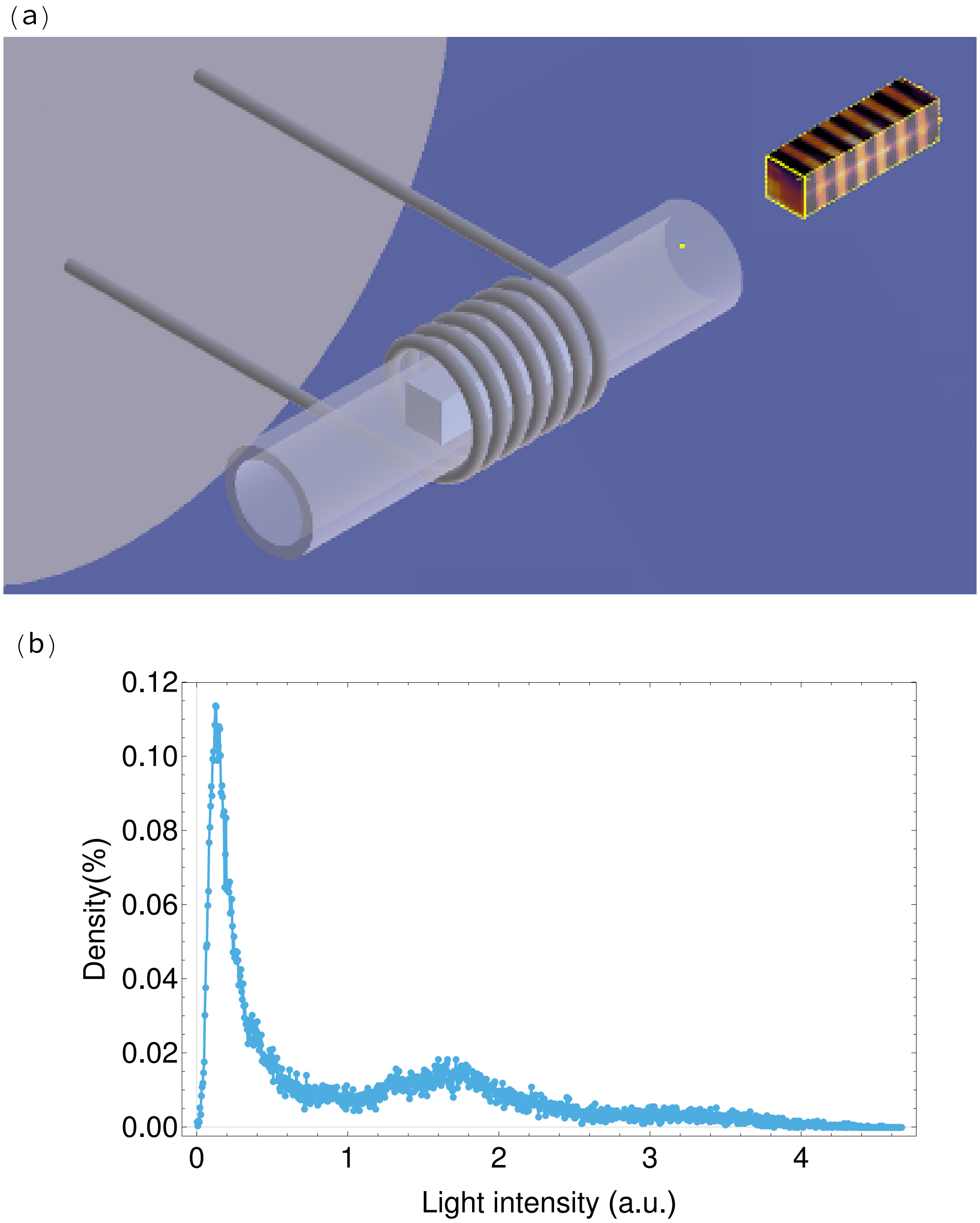


Figure 2.2: (a) Diagram of the NMR probe-head. COMSOL ray tracing simulations were used to map the hyperpolarization laser light distribution over the sample volume. (b) Simulated absorbed light intensities per mesh element binned to demonstrate their effective distribution.

Chapter 3

Experimental Sample, Apparatus and Methods

3.0.1 Sample

The sample used in this study is a 2.1×2.2×8 mm ^{29}Si crystal initially grown from charge 10 material as part of the Avogadro project [8]. The sample was cut from crystal 28Si-10Pr10.6.1PeFZ3, which was grown along the [100] crystalline plane [33]. The sample has been isotopically enriched to a concentration of 99.9954% silicon-28. The residual 0.0046% is assumed to consist primarily of silicon-29 [33]. The sample has a phosphorus concentration of 1.5×10^{15} cm^{-3}, which was introduced in the form of phosphine gas during the final float-zone crystal growth run. The sample also has a concentration of 1×10^{14} cm^{-3} of boron.

The sample was extracted with a diamond saw cutting which results in a rough surface finish and may induce strain in the crystal, which would lead to signal line broadening and increased surface recombination rates which results in reduced hyperpolarization rates for donors near the crystal surface [90]. The sample was chemically etched in 10:1 solution of HNO_3:HF to remove surface damage.

Characterized coherence times may be found for the sample in table 3.1.

Temperature(K)	T_2 (ms)	Sequence
4.2	52	Hahn echo
1.7	421	Hahn echo
1.7	1200	CPMG

Table 3.1: Sample's ^{31}P coherence times [33].

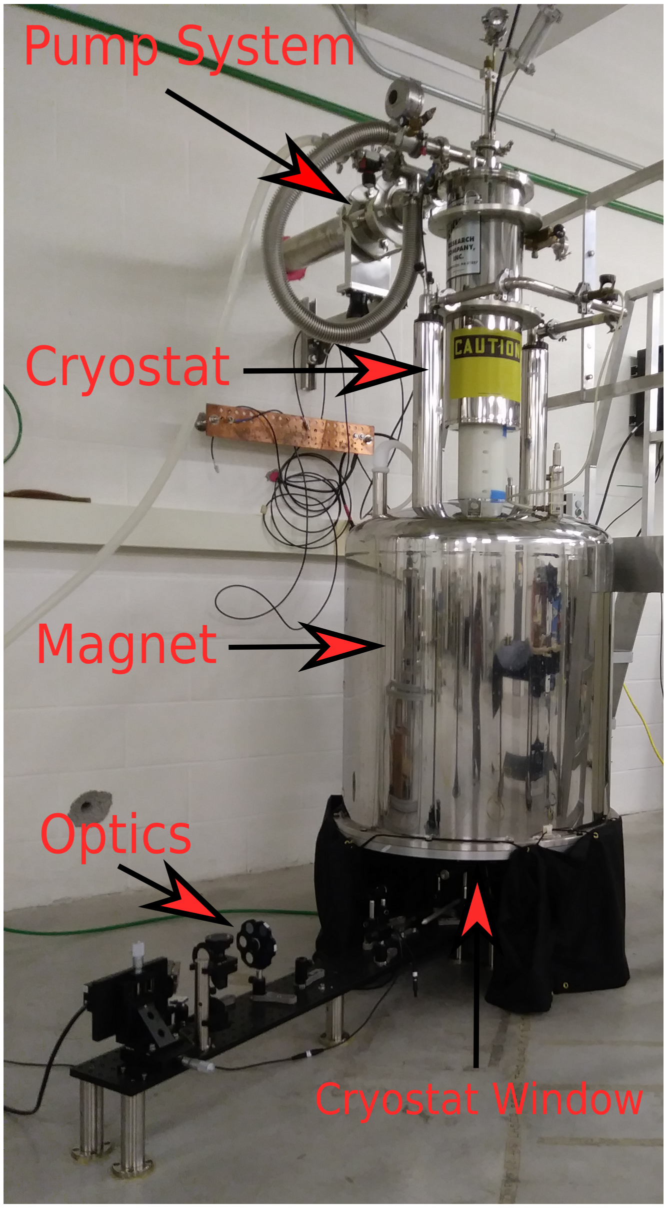


Figure 3.1: Diagram of the experimental setup with the cryostat inserted in the magnet and the probe in the cryostat. The optical table is aligned such that the beam may be guided through the cryostat window and incident onto the sample.

3.0.2 Probe

The probe is constructed around a central stainless steel tube. This tube provides a central supporting structure and is also used for filling the LHe cryostat. Copper baffles are spaced at uniform intervals along the probe and provide mounting locations. The last 15 cm of the probe is threaded to allow the Teflon sample holder to be positioned such that the sample is in the magnet sweet spot. The probe has a number of vacuum feedthroughs, four of which are used for fibreglass tuning rods, one for the circuit electronics. There are three additional ports at the top of the probe: one SHV connector for high voltage electronics, one thermometer connection, and a port connected to the central rod for LHe fills.

Probe Circuit

The magnetic resonance signal is observed with a home built cryogenic NMR probe. The probe is designed to be inserted at the top of the cryostat as seen in figure 3.1 and to have the sample sit in the sweet spot of the magnet. The probe circuit is a tunable split resonance circuit, which allows its two electrical resonances to be excited with a single input line [34]. The circuit diagram is shown in figure 3.2. The circuit coil is wound of 0.7 mm diameter rhodium flashed silver plated copper wire and has seven turns with a pitch of 1.2 mm. The lower resonance may be tuned over a range of 110-122 MHz, and the upper resonance of 165-190MHz using tunable capacitors, that are connected to fibreglass rods that pass through vacuum feedthroughs at the top of the probe and may be rotated to vary their capacitance. The lower resonance is used for the application of low power saturation pulses, and the upper resonance is used for the application of rotation and readout pulses, as well as signal detection.

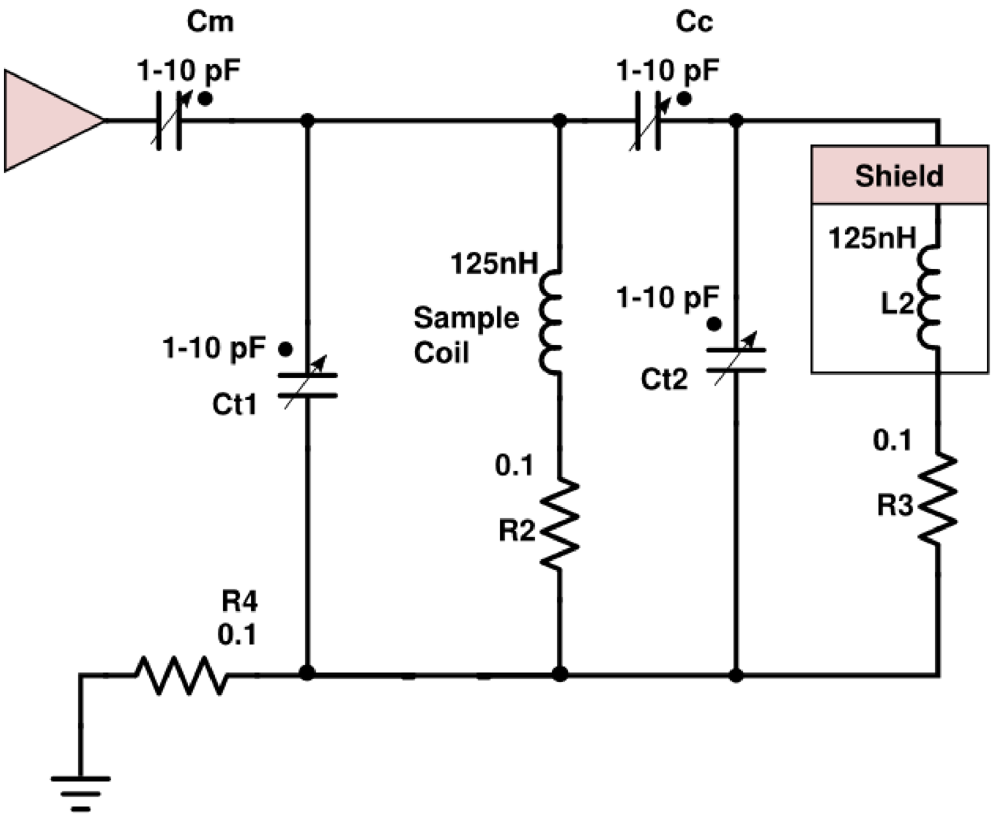


Figure 3.2: Circuit diagram for tunable split resonance probe circuit. The circuit has two resonances that are both tuned and matched via the tuning capacitors C_m, C_c, C_{t1} and C_{t2}. Both resonances are excited with a single circuit input. The sample sits within the central sample coil in the circuit.

As there is no directly observable signal at the lower resonance for use in calibrating the Rabi strength, calibration is done with the aid of a pickup coil. The coil is positioned near the sample coil so that it is weakly coupled. For each probe tuning the lower resonance Rabi strength is calibrated by first characterizing the S_{14} parameter of the pickup coil at the upper electrical resonance and then the lower resonance. By observing the hyperfine shifted ^{31}P signal the probe input power is calibrated for a given Rabi strength Ω_{174} with a Rabi sequence. The peak to peak voltage induced on the pickup coil by the Rabi drive is then measured with an oscilloscope. Using the measured pickup coil S_{14} parameter the peak to peak voltage necessary for the lower electrical resonance to have an equivalent drive strength is calculated. The spectrometer output power is then adjusted until the required oscilloscope peak to peak voltage is achieved. When calibrating for a specific Rabi strength, it is important to account for the ENDOR enhancement which is a result of the electron spin remaining adiabatically in the instantaneous ground state determined by the applied

RF field, [16]

$$\Omega_{174}^{(en)} = \gamma_{en}\Omega_{174} = \left(1 + \frac{A}{2\gamma_n B_0}\right)\Omega_{174} \tag{3.1}$$

$$= \left(1 + \frac{117.52\text{MHz}}{2 \times 115.3\text{ MHz}}\right)\Omega = 1.5096 \times \Omega_{174}. \tag{3.2}$$

The bare phosphorus resonance will not have such a Rabi enhancement and must have its magnitude rescaled by the factor $1/\gamma_{en}$.

High Voltage Wiring

One of the main design goals of designing the probe was to deliver both AC and DC electric field signals of up to 10 kV to the sample, which was a significant challenge as there are no off the shelf high voltage components that are rated to be reliable at cryogenic temperatures, high voltages, and frequencies up to 200 MHz. Due to the dangers inherent in working with such high voltages at currents up to 20 mA all efforts were made to conform to the IEEE 400-2012 safety standards [4]; including protective equipment, low resistance grounding, and an exclusion zone. In the design of the high voltage control chain, the three crucial components were cabling, connectors, and the vacuum feedthrough. While no compatible microwave cabling is rated to these voltages, dielectric strength calculations show that Teflon insulated RG-58 and SHV-5 connectors are capable of tolerating voltages of 20 kV and are routinely used for this purpose at Fermilab [2]. Teflon also has the benefit of being relatively well behaved at cryogenic temperatures. Extensive low-temperature high voltage insulation testing of the cabling and connectors was performed to verify that they were capable of handling voltages of up to 10 kV without excessive leakage currents or arcing. SHV-5 connectors supplied by Pasternack and SHV-5 vacuum feedthrough supplied by Allectra GmbH were used for cable mating. The vacuum feedthrough grounds the probe with the cryostat which is itself directly connected to the building ground. Care was taken to introduce no additional ground loops. Room temperature cabling utilized standard Teflon insulated RG-58 coaxial cable and low-temperature cabling utilized semi-rigid stainless steel Microcoax UT-141-SS-SS which transitions to UT-141 copper semi-rigid close to the sample as stainless steel is slightly ferromagnetic. It is crucial that coaxial cable with low-density PTFE insulation not be used as it will crack at low temperatures, which will lead to arcing at high voltages. There are no SHV connectors available for purchase designed to be mounted with semi-rigid coax. A modified connector mounting procedure for 0.141” semi-rigid coaxial cable which has the same inner conductor diameter as RG-58 was developed. The procedure is as follows:

A. Strip the semi-rigid cable outer-shield insulation as indicated in connector mounting instructions.

B. From an RG-58 cable strip a 10-15cm long portion of the braided outer conductor.

C. Align the stripped outer conductor on the semi-rigid cable such that the stripped RG-58 conductor extends past the semi-rigid stripped outer conductor by the amount indicated in the connector mounting instructions for the RG-58 outer insulation.

D. Heat shrink the stripped outer conductor in place on the semi-rigid cable such that the diameter of the heat shrunk semi-rigid component is comparable to the OD of an RG-58 cable (5 mm). Multiple layers of heat shrink may be necessary.

E. The component of the semi-rigid cable with the heat shrunk outer conductor now emulates a traditional RG-58 cable, mount the connector as instructed.

An image of a resulting set of connectors attached with this technique is shown in figure 3.3. The above connection procedure has been tested and have found it capable of making robust and repeatable matings that function both at room and cryogenic temperatures.

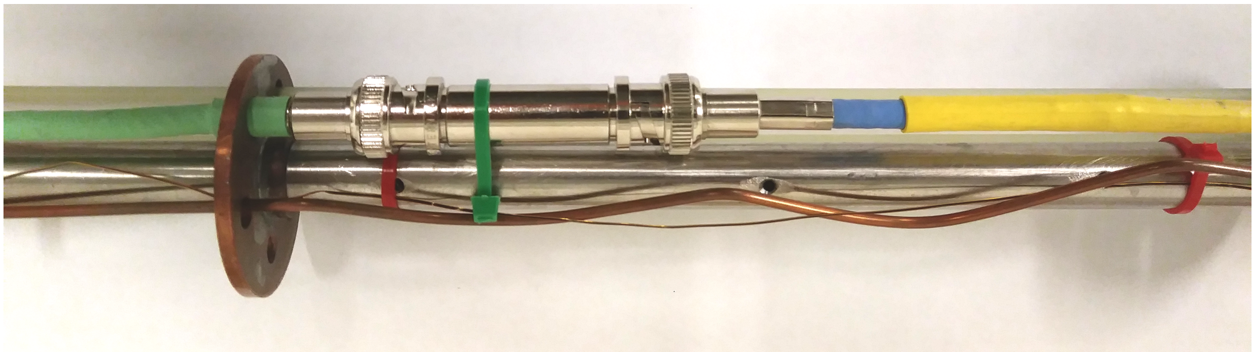


Figure 3.3: Demonstration of two SHV connectors that have been attached to the semi-rigid coaxial cable using the method outlined in section 3.0.2 and then mounted to the probe. The connector mating has survived many cryogenic and high voltage cycles.

DC high voltages are supplied by a Glassman FJ10R12 high voltage reversible polarity supply which can be triggered by the spectrometer.

Purified LHe Cell

Commercial LHe have many impurities that can reduce its dielectric strength well below the 30 kV/mm of pure LHe [12]. The reduced dielectric strength has caused unwanted arcing in high-voltage experiments. To increase the purity of the LHe a purified LHe cell has been engineered to surround the sample. The sample is inserted in a custom manufactured sapphire tube which is then inserted in the probe detection coil, which can be seen in figure 3.4. The tube provides high voltage isolation of the detection circuitry and spectrometer while allowing 90% of the applied infrared laser intensity to reach the sample. The ends of the tube are closed with leak-tight Teflon caps with vacuum grease applied. At cryogenic temperatures the caps contract around the tube forming a tight seal. One cap is fitted with a capillary which serves as both a purified LHe delivery capillary and a high voltage cathode while a copper wire provides the anode in the other cap. The LHe delivery capillary leads to a brass LHe cell.

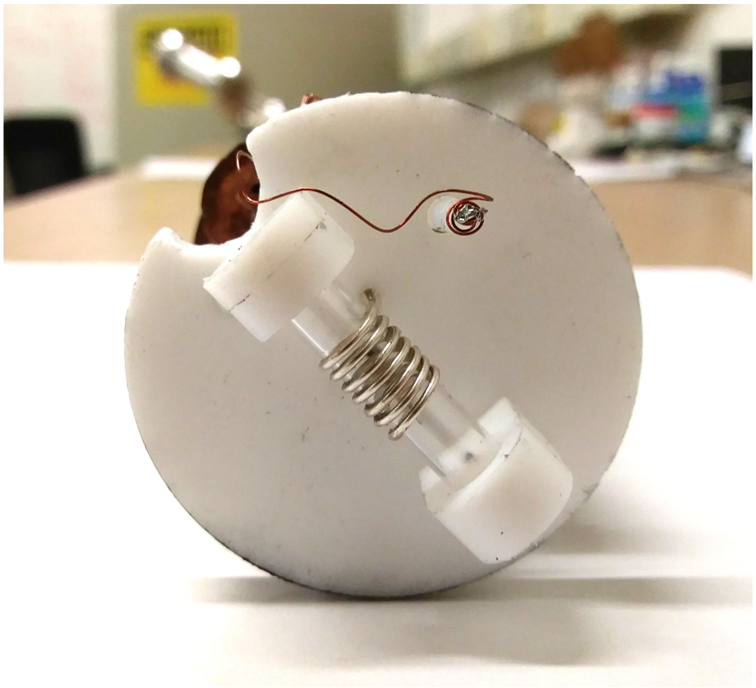


Figure 3.4: Sapphire tube inserted inside the detection coil. Tube is capped on both ends with Teflon and capillaries are inserted into the caps to supply both high voltage and purified LHe.

Two methods for delivering purified LHe to the cell are currently being tested. The cell has a window made of Vycor 7930 which has micropore channels that allow only purified superfluid helium to pass through [92]. Additionally, the cell is attached to a capillary that is connected to a gas handling system at the top of the probe. This capillary is used to condense helium into the cell. It has yet to be established which method is more reliable for obtaining pure LHe.

A ferroelectric disk composed of cesium nitrate mixed with Stycast epoxy is inserted in the interior of the cell [11, 15], which is used to attract and contain any charged impurities that may enter the cell.

3.0.3 Cryostat

Experiments were performed in a customized Janis CNDT8 LHe cryostat which is inserted in a 286 MHz (B_0 =6.7 T) wide-bore Bruker superconducting magnet. The cryostat has an exterior liquid nitrogen (LN_2) jacket and an interior LHe bucket dewar. The two cryogen reservoirs are separated from each other and the exterior walls by a shared vacuum space. The cryostat has a removable 2.5" OD tail that is designed to slide in the bore of the magnet. At the base of the tail, there is a series of three sapphire windows that are transparent to the DNP laser. The NMR probe described in 3.0.2 is inserted directly into the Helium bucket, and the inductive coil rests near the third sapphire window in the sweet spot of the magnet. Experiments may be performed helium temperature of 4.2 K at atmospheric pressure. However, all experiments in this thesis were performed at 1.3 ± 0.1 K which is reached by pumping on the helium bath through a series arrangement of a TRIVAC B D 65 B rotary vane vacuum pump in conjunction with a RUVAC WS1001 roots pump. The pressure of the helium dewar is measured while pumping with a Granville-Phillips 275 Mini-Convectron pressure sensor.

3.0.4 Electronics

A Bruker Avance 300 spectrometer provides probe tuning, pulse signal generation, and quadrature detection. Control signals applied to the upper ^{31}P hyperfine shifted resonance signal of 174 MHz are routed through a 300 W Bruker BLAX300RS X-channel amplifier. The signal is then routed through a Werlatone C5964 30dB dual directional coupler. The 174 MHz tone is applied to the input (J1) path of the coupler. Saturation pulses applied to the ionized ^{31}P (non-hyperfine shifted) resonance at 115 MHz are optionally routed through one of two possible channels

A. In the case where amplifier noise will not provide significant levels of experimental error the signal from the spectrometer is routed through the same amplifier, preamplifier, and directional coupler channels as the 174 MHz signal.

B. When amplifier noise is important, the signal is routed directly from the spectrometer signal generation unit (SGU) through a Mini Circuits ZHL-32A low noise amplifier into a K&L Bandpass filter tuned to the desired signal frequency. The signal is then coupled to the probe through the reverse (J4) port of the dual directional coupler. The coupled signal will be reduced by 30 dB.

The transmitted signal was monitored via the forward (J3) port with a Tektronix DPO 4104 oscilloscope. The electrical signal in the coil was directly observed via the probe pickup coil and may be compared to the transmitted signal with the oscilloscope.

3.0.5 Optics

A schematic of the optical setup is found in figure 3.5. Illumination of the sample for the DNP is provided by LaserGlow LRS-1047-CFM-0030-05 laser. The laser has a continuous wave output of 500 mW at 1047±1 nm. This wavelength was chosen as it is slightly greater than the silicon bandgap [95], such that it may excite free carriers, yet still be able to penetrate the silicon crystal to uniformly illuminate the sample given its long penetration constant of 26.9 cm [55]. During experiments, the laser is triggered via TTL by the spectrometer. The $1/e^2$ width of the laser beam is 2mm. The beam is optionally routed through an attenuator wheel with its values ranging from 0-1.0 optical density (OD). To uniformly illuminate the sample the beam is passed through a 4× beam expander which provides a spot size of approximately 8 mm at the sample. The beam is reflected into the Janis cryostat through the sapphire windows and onto the sample via a silvered mirror. The input power transmitted to the internal LHe bath space was measured to be 218 mW. The beam must be aligned onto the sample to illuminate the crystal uniformly. As the 1047 nm beam is part of the infrared spectrum and is invisible, a 532 nm Laserglow LBS-532 alignment laser was installed and may be coupled to the beam path with a flippable alignment mirror. The alignment procedure is as follows

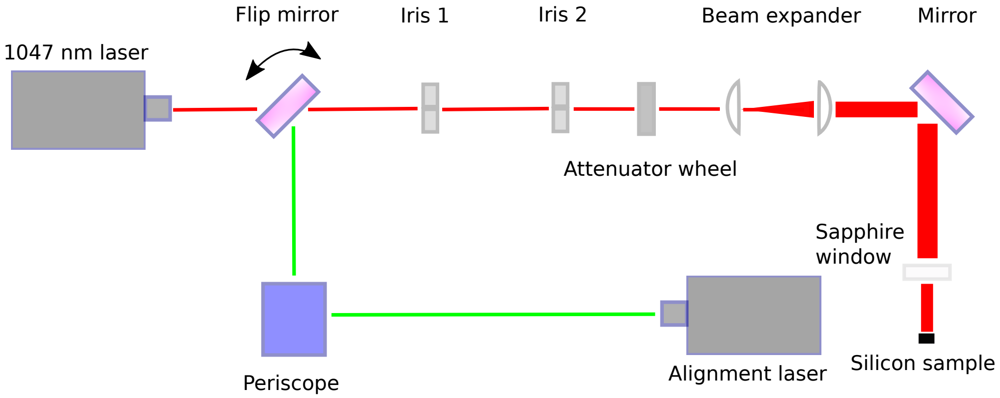


Figure 3.5: Schematic of DNP illumination optics. An alignment laser may be optionally coupled into the beam path to make alignment easier.

A. Configure the two irises such that they are effectively two pinholes.

B. Place a 1047 nm detector card behind the second iris.

C. Align the 1047 nm beam so that it passes through both irises and is visible on the detector card.

D. Disable the 1047nm laser and power on the alignment laser.

E. Flip alignment mirror into its upright orientation such that it redirects the alignment beam into the 1047 nm beam path.

F. Adjust the alignment laser's optical periscope and the flip mirror such that the alignment laser is visible on the viewing card behind both irises. At this point, the 1047 nm and the alignment laser are aligned to be co-linear.

G. Open both irises completely and remove the viewing card.

H. Adjust angled mirror under magnet bore such that the alignment laser is incident onto the sample.

I. Flip alignment laser mirror into down position and re-enable 1047nm laser.

J. The last two steps may be cycled in combination with a DNP experiment tuning step with the aim of maximizing the hyperpolarization rate, and consequently the laser intensity incident on the sample.

The entire optical setup is mounted to an optical breadboard which is, in turn, attached to the laboratory floor underneath the magnet bore-hole.

Chapter 4

Experiments and Results

4.0.1 Experimental Procedures

A typical experimental sequence is composed of an initial saturation train being applied at the D^0, ^{31}P resonance frequency to reset the phosphorus spins to an unpolarized state. The 1047 nm laser is then turned on, and the hyperpolarization commences. Simultaneously with the laser illumination, a saturation pulse is applied near the D^+ resonance frequency with detuning Δ and amplitude Ω. Hyperpolarization is allowed to proceed for some time τ, after which a readout $\pi/2$ pulse is applied at the D^0X resonance and free induction decay (FID) signal is acquired. This pulse sequence is visualized in figure 4.1

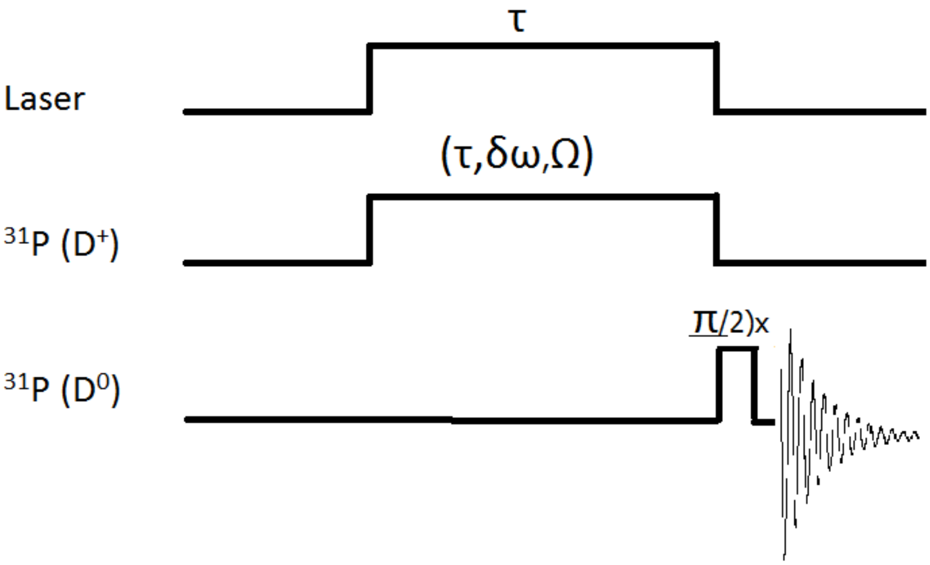


Figure 4.1: A typical hyperpolarization experiment with D^+ saturation pulse sequence.

Final net magnetizations are obtained by Fourier transforming the FID and then integrating the obtained spectrum over the signal region. The number of scans per point are chosen so as to have approximately the same SNR for each experiment point. For an exponential growth curve $M(t) = \left(1 - e^{-\frac{t}{T_1^{((\mathrm{DNP})}}}\right)$ the SNR is defined as $s(t,n) = M(t)\sqrt{n}/\sigma$, where σ^2 is the variance of the experimental noise assuming additive white noise that is constant over the entire experimental run and n is the number of experiment repetitions performed. For a desired SNR s and experiment buildup time t the number of experiments that must be performed to achieve the SNR is

$$n = \left(\frac{s\sigma}{M(t)}\right)^2. \tag{4.1}$$

Initially, the polarization buildup as a function of applied laser power under no saturation was obtained. For the model to be true, the hyperpolarization rate $1/T_1^{D^0}$ should be inversely proportional to the applied laser power. A series of hyperpolarization experiments at varying laser powers are performed with results shown in figure 4.2. A simple exponential function very well approximates the buildup, this is a strong indicator that the buildup is not dependent on local light intensity variations, which would create localized free exciton concentration imbalances, but rather long range phenomena within the crystal-like phonons as hypothesized in 1.9. The inset figure demonstrates that the buildup rate is linearly dependent on the applied lasing power, correctly predicting that the hyperpolarization should stop as the applied laser power becomes nil.

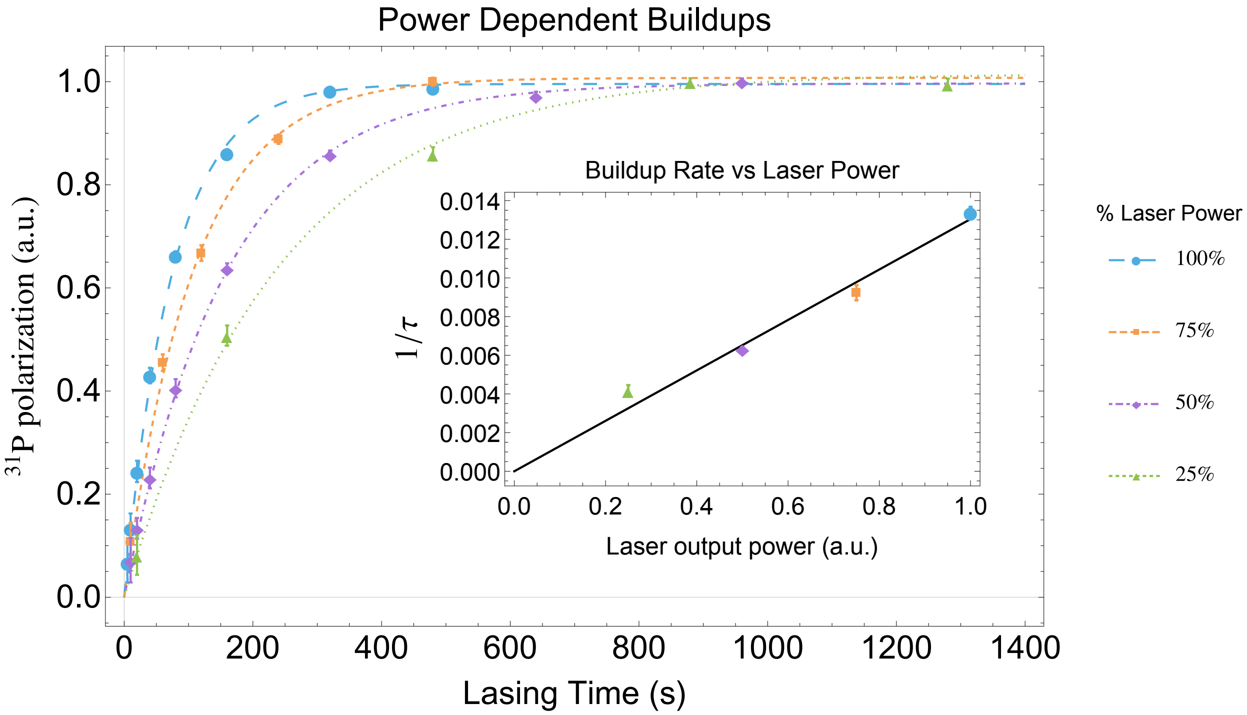


Figure 4.2: Polarization buildup time constant dependence on the laser power with no applied saturation ($\Omega = 0$). Importantly, the final buildup polarization is the same to within noise levels for all lasing power levels. (inset) The buildup curves are fit to an exponential buildup function and their polarization rates $T_1^{\text{DNP}}(I)$ are extracted. The extracted buildup rate as a function of laser power are fit to a linear model. The error bars are extracted from the T_1^{DNP} exponential growth model fit and are derived from a linearization of the model function. This demonstrates that the polarization buildup rate is a linear function of power.

Equation (2.10) predicts that fixing the magnetization buildup time and applying a saturating drive the magnetization should sweep out a Lorentzian as a function of the detuning from the D^+ ^{31}P resonance. Saturated (Ω=206 Hz) buildup experiments are performed with the maximum laser power and hyperpolarize for τ=160 s, which provides a good balance between experiment time, and single scan SNR. The datasets are normalized against an unsaturated (Ω=0 Hz) buildup and plotted with error bars in figure 4.3. Weighted least squared fits are made to the numerical model, both with and without the light distribution. The light distribution fit is marginally better. For this set of experiments the hyperpolarization rate is $T_1^{\text{DNP}} = 53.7 \pm 2.0$ s. From these fits the extracted rates are $\kappa_{D^0\to D^0X} = 731 \pm 63$ Hz and $\kappa_{D^+\to D^0} = 400 \pm 20$ kHZ. Normalizing by the average input laser intensity $\mathbb{E}[I]$ the intensity dependent rates $\alpha_{D^0\to D^0X} = 4.40 \pm 0.38$ kHz$/\left(\text{mW cm}^{-2}\right)$ and $\alpha_{D^+\to D^0} = 981 \pm 49$ kHz$/\sqrt{\text{mW cm}^{-2}}$ are calculated. A small paramagnetic shift of the ionized phosphorus resonance frequency of $\delta/(2\pi) = -10 \pm 1$ kHz with respect to a trisodium phosphate reference sample is observed. In all future experiments the central saturation frequency is shifted to account for this paramagnetic shift. It must be stressed that these rates are temperature, field, doping concentration, and wavelength dependent, and it may be of interest to map out as a function of these parameters in the future.

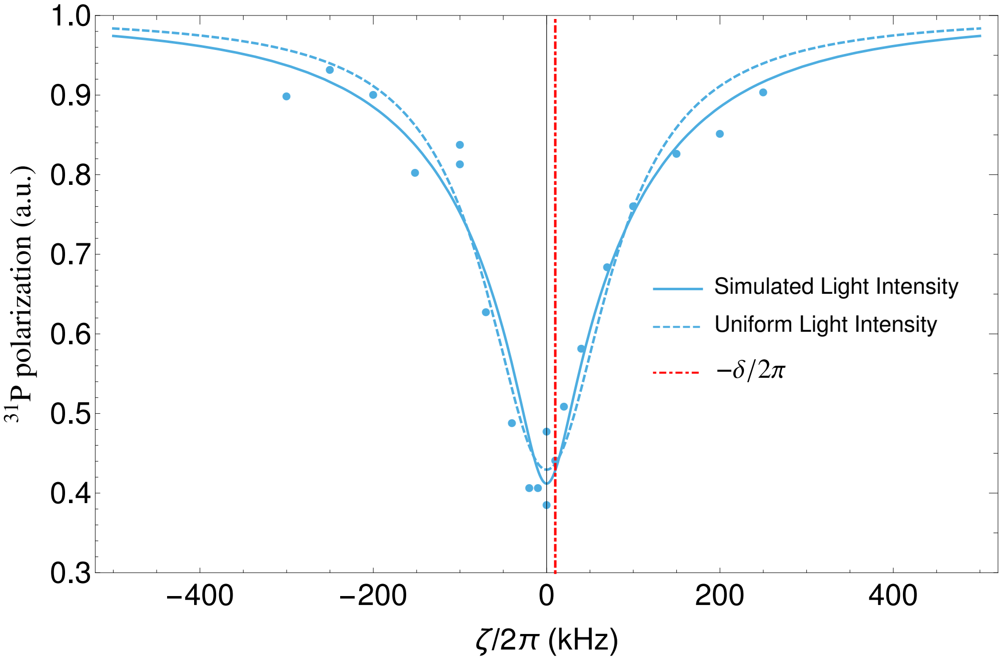


Figure 4.3: A saturation drive of Ω=207 Hz is applied while varying the frequency (Δv) and hyperpolarize for sufficiently long (160 s) to observe the steady state polarization. The observed magnetizations are normalized to Ω=0 reference experiment and plotted centered at $\zeta = 0$. The data is fit to the model found in equations (2.4a-2.6c), both accounting for the simulated light distribution (blue, solid) or assuming a uniform light distribution (blue, dashed). A D^+ paramagnetic shift of $\delta/(2\pi) = -10$kHz is observed which is included in the plot (red, dot dashed).

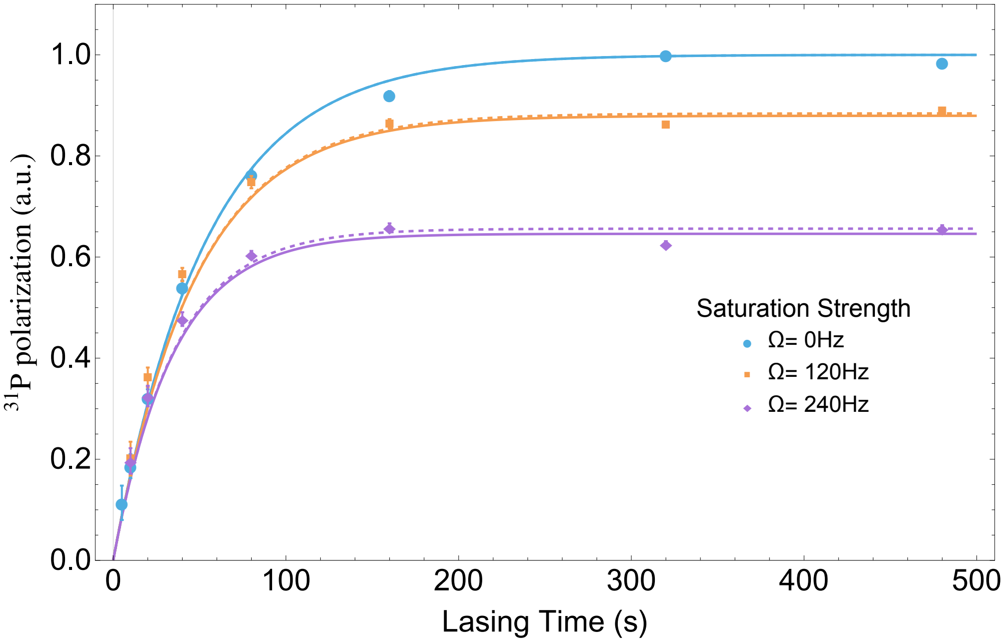
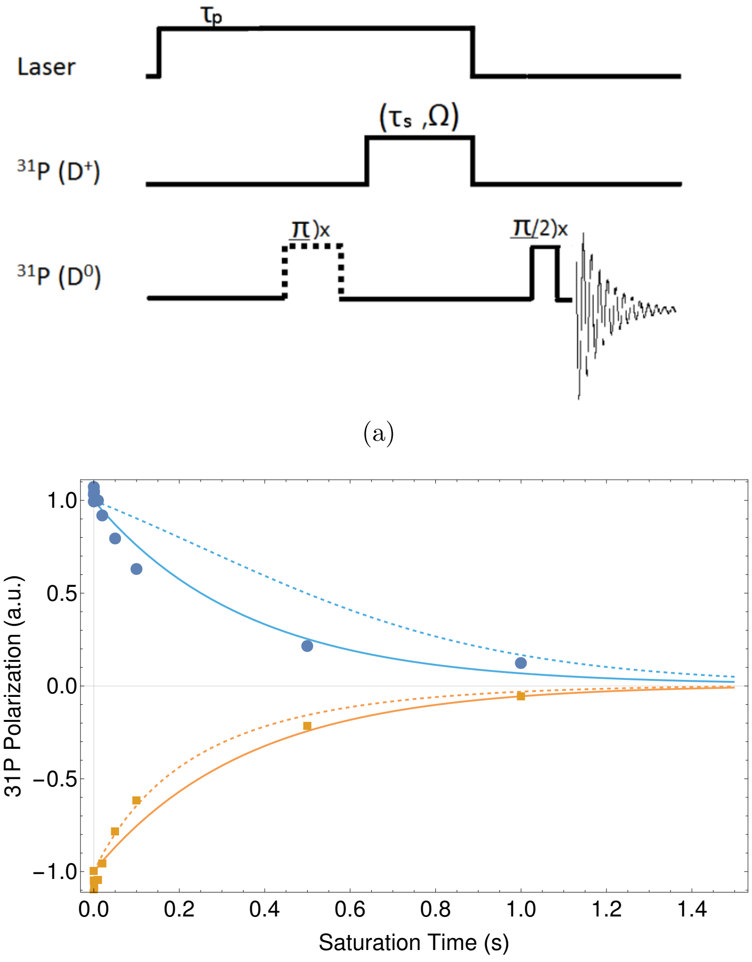


Figure 4.4: Buildup curves taken under different applied saturation strengths Ω= 0 (circles), Ω= 120 Hz (diamonds) and Ω= 240 Hz (squares).

4.1 Nuclear Polarization Dependent D^0X Formation

In section 2.1 an overview of the competing hyperpolarization model proposed by Sekiguchi *et al.* was given. In this section a series of experiments that provide strong evidence against this alternative model is reported. The alternative model predicts the rate of D^0X formation to be polarization dependent. The phosphorus ensemble is hyperpolarized and then optionally a π pulse is applied, these are then followed by a strong saturation drive of $\Omega = 3975$Hz as depicted in figure 4.5a. The excitonic polarization model predicts an asymmetry in the saturation dynamics dependent on whether the magnetization is aligned or anti-aligned with the electron magnetization provided that the nuclear polarization is sufficiently high. Previously, the nuclear polarization was estimated to be higher than 64%, which is used for comparison of the two models [33]. Note that larger polarizations will only serve to deviate the predictions of the two models further.

Experiment results are shown in figure 4.5b. It is clear that the saturation dynamics do not display a strong polarization dependent asymmetry. The phononic hyperpolarization fits the data relatively well, while the excitonic model fits the electron aligned polarization poorly, which is strong evidence against the excitonic model. This behaviour is more easily seen in the summed signal $m(t) = m_{\downarrow}(t) + m_{\uparrow}(t)$ as in figure 4.6. It is clear that for the hyperpolarization and bound exciton processes to depend linearly on the nuclear state, the absolute polarization must be much less ($< 10\%$) than the previously lower bounded absolute polarization of 64%.



(a)

(b)

Figure 4.5: (a) Pulse sequence for phosphorus polarization dependent Auger ionization experiment. The experiment begins with a polarization laser pulse of length τ_p, followed by an optional π pulse (dotted), which will invert the phosphorus polarization. A saturation tone is then applied with Rabi strength Ω for time τ_s. Finally, the signal is finally observed with a $\pi/2$ pulse. By varying the saturation time τ_s polarization dependent saturation rate may be extracted. (b) Experiment results for $\Omega = 3975$ Hz. The plotted lines are predictions from the process models, where the solid line corresponds to the saturation rate independent of nuclear polarization and the dashed line is a numerical simulation of the coupled Bloch equations when the rate of exciton capture $\kappa_{D^0 \to D^0X}$ is dependent on the nuclear polarization, as described in section 2.1, for an initial polarization of 64%. Higher initial polarizations will further enhance the asymmetry between the two curves. No saturation dependence on the spin state of the phosphorus nuclei is observed. Furthermore, the data agrees with the prediction of the independent polarization model relatively well.

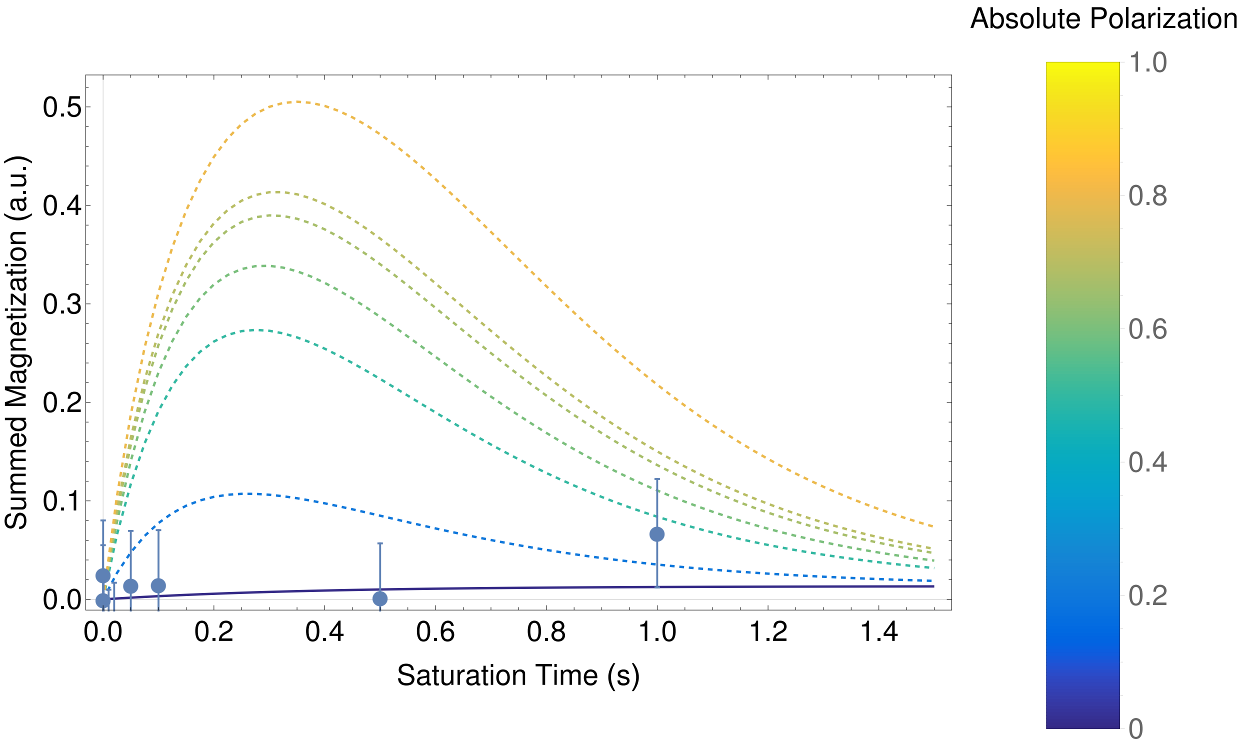


Figure 4.6: The same data as shown in figure 4.5b, but displayed as the sum of the two initial magnetization data points so as to enhance the visualization of the signal saturation asymmetry. (solid) Nuclear state independent hyperpolarization simulation. (dashed) A range of nuclear spin-state dependent simulations for varying levels of absolute nuclear polarization.

For further evidence in favour of the phononic model recall the initial observation that the phosphorus hyperpolarization process is well fit by a single exponential. Recalling that the mean free path of excitons is on the order of 10 μm the hyperpolarization rate under the excitonic would be strongly dependent on the local light intensity. The observed buildup would be integrated over the sample light distribution

$$\int dI p(I)(1 - e^{-\chi I t}) \tag{4.2}$$

Where $p(I)$ is the laser intensity distribution over the sample, and χ is the proportionality constant between the input laser intensity and the hyperpolarization rate. In figure 4.7 the observed hyperpolarization dynamics are compared with both a fit two a single exponential and a sum of exponentials proportional to the laser intensity as described by equation (4.2). It is clear that the hyperpolarization rate is not intensity dependent, which is consistent with a phononic hyperpolarization process which at sub-terahertz, phonons have mean free paths of several mm[35], which is on the order of the sample size.

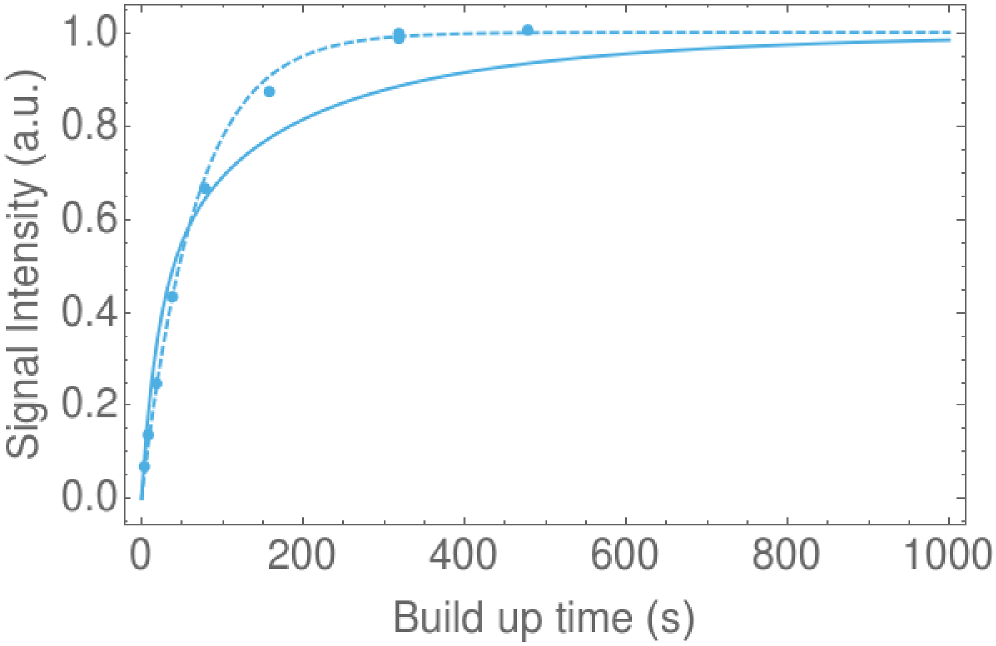


Figure 4.7: Comparison of a single exponential buildup fit (dashed) to a simulated polarization buildup assuming that the polarization rate is proportional to the local incident laser intensity (solid). It is clear that the hyperpolarization rate fits the dashed curved better and is therefore not a dependent on the local laser intensity.

Chapter 5

Future Work and Conclusion

5.0.1 Stark Ionization of the Phosphorus Donor

The examination of the bound exciton Auger ionization process in the previous section has proceeded in parallel with an effort to demonstrate Stark ionization of the phosphorus donor electrons. In section 3.0.2 an overview of the probe's ability to deliver high voltages directly to the sample in the cryostat and current efforts to design a purified LHe cell for the sample to reside within are given. By placing a set of conductive plates around the sample, the DC Stark effect may be used to manipulate the donor electron's wavefunction and at sufficiently high voltages (4̃.5 kV for the sample geometry) ionize the system. Saeedi *et. al* demonstrated that ionized phosphorus defects in isotopically purified silicon have coherence times up to hours [68]. Their work utilized AEDMR to observe the nuclear spin state and thus requires the donor electron to be present to measure the nuclear spin state. Additionally, they relied on narrowband lasers to excite bound exciton transitions to hyperpolarize and ionize the system. Our proposed protocol requires only the capability to illuminate the sample with above bandgap light and apply an electric field across the sample. The nuclear spin state may be observed inductively, and the dynamics of the ionized phosphorus nuclei may be studied directly. The ultra long coherence times of the system would be ideal for the construction of high sensitivity nuclear spin magnetometers [41] and gyroscopes [50, 9] as spin phase detection sensitivity is proportional to $1/T_2$.

5.0.2 Conclusion

The two competing models for the Si:P above bandgap hyperpolarization mechanism have been compared and contrasted. The first which predicts that the hyperpolarization is due to time-dependent strain fields modulating the electron spatial wavefunction, and the second model predicts hyperpolarization as a consequence of angular momentum conservation during the process of bound exciton capture. Crucially, these models predict different buildup dynamics as the former model's buildup rate should be independent of the nuclear state while the rate of the latter will be dependent.

A novel experimental procedure for studying the Auger ionization process of phosphorus donors in isotopically purified silicon with saturation experiment procedure given in section 4.0.1 has been demonstrated. The dynamics were modelled with three coupled sets of Bloch equations, and it was shown that the observed dynamics fit the model well. This technique could potentially be used for measuring the hyperfine shift of bound exciton hole as noted in section 2. The light distribution over the silicon sample is nonuniform and such effects were accounted for within the model with the aid of ray tracing simulations, which allows for the Auger ionization and photo-neutralization rates to be extracted for the specific experimental conditions. A small paramagnetic shift of the ionized phosphorus Larmor frequency of -10 kHz is observed.

Noting that the two hyperpolarization models predict different rates for bound exciton capture conditional on the polarization of the phosphorus nuclei, an experiment contrasting the saturation dynamics of phosphorus initially polarized aligned/anti-aligned with the electron thermal population is performed. The experiment results demonstrate that the saturation rate does not depend on the polarization of the phosphorus nuclei. This strongly suggests that the phononic hyperpolarization model underlies the hyperpolarization mechanism. This is also strongly implied by the observation that a single exponential well models the observed hyperpolarization buildup. If the hyperpolarization process is excitonic, it would depend on the local exciton, and free carrier densities, which would be strongly dependent on the local illumination intensity. Consequently, the observed buildup would be a sum over many exponentials due to varying light distribution over the sample. As the phonon mean free path is on the order of mm, the observation of a single exponential buildup is consistent with the first hyperpolarization model. This still leaves open the question of what exchanges angular momentum with the bound exciton electron to allow the capture process to proceed and is an avenue for future work.

Chapter 6

Bayesian Parameter Estimation of T_1 Relaxation Time

The relaxation time or T_1 of a spin ensemble contains information about spins' and their environment [75]. The T_1 relaxation time is a key parameter of interest in applications such as magnetic resonance imaging (MRI), and NMR logging [96, 46]. Depending on the system and experimental conditions the value of T_1 can vary from microseconds to hours or even days [33]. Therefore, parameter estimation of the relaxation time can be a very costly experiment. Numerous techniques for the optimization of T_1 relaxation measurements have been explored in the literature [27, 36, 89, 78, 96]. These works have focused on frequentist classical designs due to the numerical intractability of Bayesian parameter estimation. The development of numerical computational methods and computing power have made possible the use of Bayesian methods, which have traditionally been intractable. The objective of the following chapters is to explore the application of Bayesian parameter estimation and experimental design to optimally select experiments to minimize the variance of the posterior distribution over T_1. The implementation of such techniques could drastically reduce experiment times.

The T_1 is typically measured with an inversion recovery (IR), saturation recovery (SR) or faster inversion recovery (FIR) experiment [89]. The system is assumed to be initialized with thermal polarization $M(\tau < 0) = M(\infty)$, along the principle axis of the applied Zeeman field $B_z = B_0$. Applied pulses are assumed to be perfect with duration $\tau_p \ll T_1$, although imperfections may be accounted for [88]. As noted by Weiss et. al. the resultant magnetization may be ideally parameterized as

$$M(\tau) = M(\infty)\left(1 - ae^{-\frac{\tau}{T_1}}\right), \tag{6.1}$$

where τ is the experiment relaxation delay and T_1 is the relaxation rate to be estimated. The inversion recovery (IR), saturation recovery (SR) and faster inversion recovery (FIR) experiments are defined by the choice of a as shown in table 6.1. When experimental imperfections are included, the model may be extended to five parameters for a more robust fit [51].

Table 6.1: Parameterizations of a for various relaxation experiments and their associated pulse sequence.

Experiment	a	Pulse Sequence
SR	1	$\pi/2 - \tau - \pi/2$
IR	2	$\pi - \tau - \pi/2$ assuming$(\tau_{th} \gg \hat{T}_1)$
FIR	$2 - e^{-\tau_{th}}$	$\pi - \tau - \pi/2 - \tau_{th}$

For clarity the magnetization through an IR experiment in figure 6.1, where τ_{th} is the system thermalization time. A π pulse is applied which inverts the magnetization $M(\tau = 0) = -M_0$. The magnetization is then allowed to recover for some time τ under the dissipative action of T_1, and finally a readout $\pi/2$ pulse is applied and the signal observed. The magnetization is allowed to return to its thermal polarization by waiting for $\tau_{th} \gg T_1$ before repeating the cycle and recording another data point.

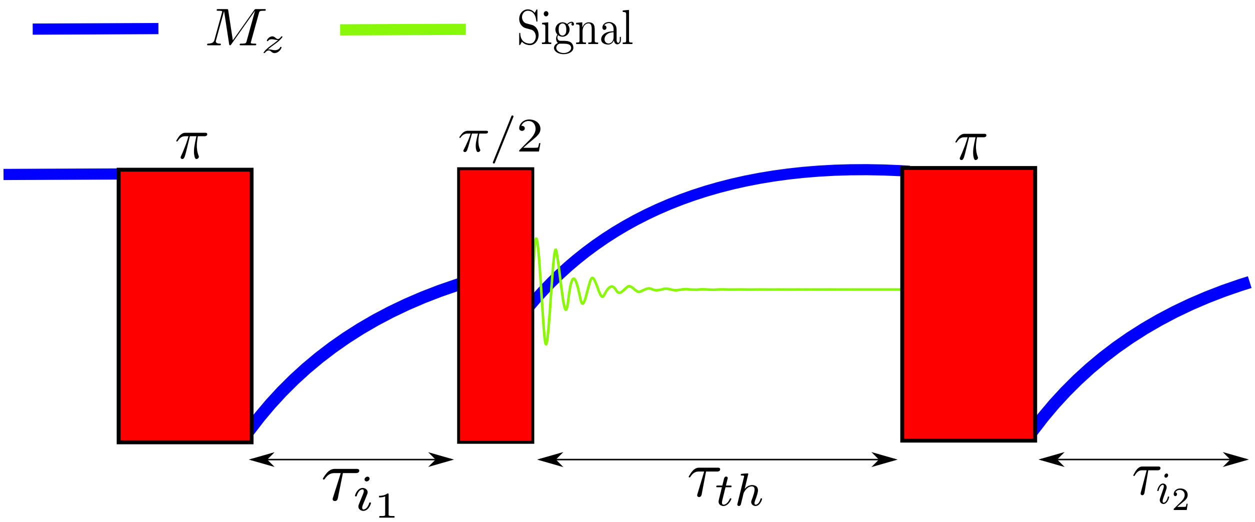


Figure 6.1: Inversion Recovery pulse sequence and magnetization. For the IR experiment the recovery time is set such that $\tau_{th} \gg \hat{T}_1$, whereas, in the FIR experiment the recovery time is set in an online manner.

Note that the thermalization step is troublesome for two reasons, first T_1 is not known, which is compensated for by waiting much longer than what the maximum possible T_1 value is "believed" to be thus increasing the interval between experiments which is natural within the Bayesian framework, and second this recovery step contains information about the value of T_1 which may be included by incorporating the recovery step into the model (see section 8.2.1). The detected FID is typically Fourier transformed and the peak of interest integrated to obtain the magnetization $M(\tau)$.

In practice the observed signal will have noise, which is assumed to be additive, normally distributed and time independent. The measurement thus consists of a signal $M(M_0, T_1; \tau_i)$

and noise ϵ_i.

$$M_i = M(M(\infty), T_1; \tau_i) + \epsilon_i \qquad \epsilon_i \sim \mathcal{N}(0, \sigma^2). \tag{6.2}$$

The variance of the noise σ^2 is assumed known and stationary. Note that it is simple to account for these noise variations with techniques described in the following chapters. The SNR is the ratio of the magnetization $M(\infty)$ to the noise standard deviation σ,

$$\text{SNR} = \frac{M(\infty)}{\sigma}. \tag{6.3}$$

Figure 6.2 demonstrates the effect of noise on the observed data, for unity SNR it is difficult to observe the signal with the eye. It is within this region that classical experimental design techniques based on model linearizations fail [88].

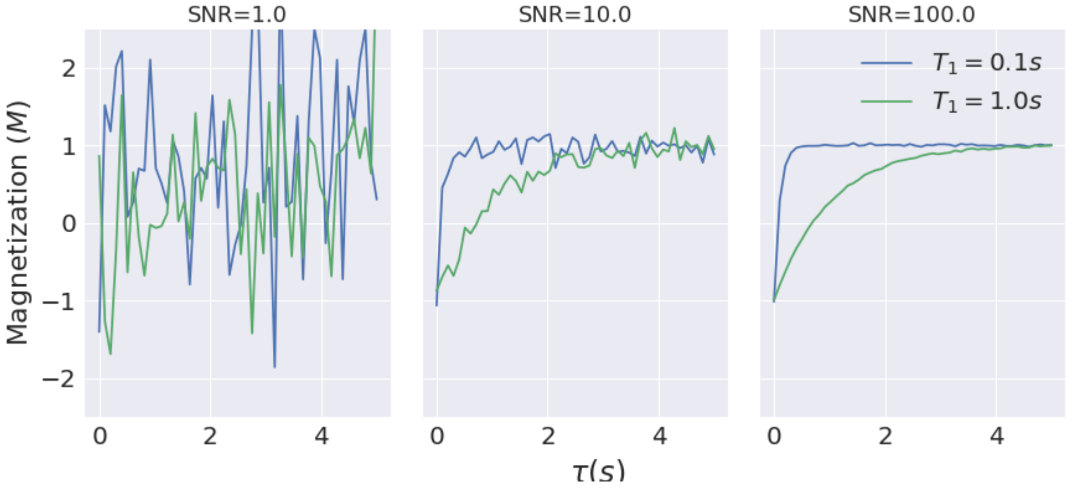


Figure 6.2: Simulated inversion recovery experiment data for varying SNR levels. Data points have been connected to guide the eye.

Under the above assumptions the model parameters $M(\infty)$ and T_1 are related to the observed datum M_i through the likelihood function

$$\text{Ł}(M_i; T_1, M(\infty)) = \Pr(M_i | T_1, M(\infty)) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(M_i - M(M(\infty), T_1; \tau_i))}{2\sigma^2}}. \tag{6.4}$$

Given prior distribution $\pi_{\mathbf{X}}(\mathbf{x})$ for the parameters $\mathbf{x} = \{M(\infty), T_1\}$ and applying Bayes' theorem to obtain the posterior probability distribution

$$\pi_{\mathbf{X}} | M_i(\mathbf{x}) = \frac{\text{Ł}(M_i; M(\infty), T_1)\pi_{\mathbf{X}}(\mathbf{x})}{\Pr(M_i)}. \tag{6.5}$$

Where $\Pr(M_i) = \int_{\mathbf{X}} d\mathbf{x} \text{Ł}(M_i | M(\infty), T_1)\pi_{\mathbf{X}}(\mathbf{x})$ is a normalization constant for the posterior distribution, and in general not analytically solvable. It is not possible to calculate analytical expectation values with respect to the above likelihood function and, much of this chapter concerns itself with solving this problem for application to parameter estimation and optimal experiment design.

6.1 Bayesian Sequential Monte Carlo Parameter Estimation

The two numerical algorithms typically employed for posterior evaluations are variants of the Markov chain Monte Carlo (MCMC) methods. The first, random walk Markov chain Monte Carlo (RWMCMC), attempts to walk through the parameter space, with a sampling density proportional to the underlying probability distribution. This guarantees convergence in the limit of many sampling steps. Popular variants include the Metropolis-Hastings algorithm, and the more recent No-U-Turn Sampler [59, 38]. The second approach, which we will focus on in this thesis is the sequential Monte Carlo (SMC) algorithm, often known as the particle filter [54]. The SMC algorithm approximates distributions as a discrete sum of delta functions with particle locations $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_{n_{par}}\}$, and weights $W = \{w_1, \ldots, w_{n_{par}}\}$,

$$\pi_{\mathbf{X}}(x) \equiv \sum_{i=1}^{n_{par}} w_i \delta(\mathbf{x} - \mathbf{x}_i). \tag{6.6}$$

To ensure that equation (6.6) represents a probability distribution we require that the weights be non-negative and sum to unity,

$$w_i \geq 0, \tag{6.7a}$$

$$\sum_{i=1}^{n_{par}} w_i = 1. \tag{6.7b}$$

In the limit of large particle numbers the true underlying distribution will be well approximated. Employing the particle approximation it is straightforward to apply the Bayes update rule to update particle weights $w_{i|O}$

$$w_{i|O} = N_O^{-1} \text{Ł}_{Oi} w_i, \tag{6.8}$$

and calculate the normalization factor

$$N_O = \sum_{i=1}^{n_{par}} L_{Oi} w_i = \Pr(O). \tag{6.9}$$

Where the $L_{Oi} = \Pr(O|\mathbf{x}_i)$ is the likelihood of observing datum O for a given particle with location $\mathbf{x}_i$. Note that the normalization factor N_O is equivalent to the likelihood of the observed data, marginalized with respect to the prior distribution. An important feature of the SMC update is that it is a sequential update procedure, allowing the algorithm to proceed iteratively, unlike RWMCMC methods that require a full walk sampling process at each step. SMC is particularly well suited to experimental parameter estimation in which experiments observe one datum o_{n+1} at a time allowing the prior distribution π_n to be updated to find the posterior π_{n+1}. Figure 6.3 demonstrates the ability of the SMC algorithm to iteratively learn the model parameters $M(\infty)$ and T_1 for an IR experiment.

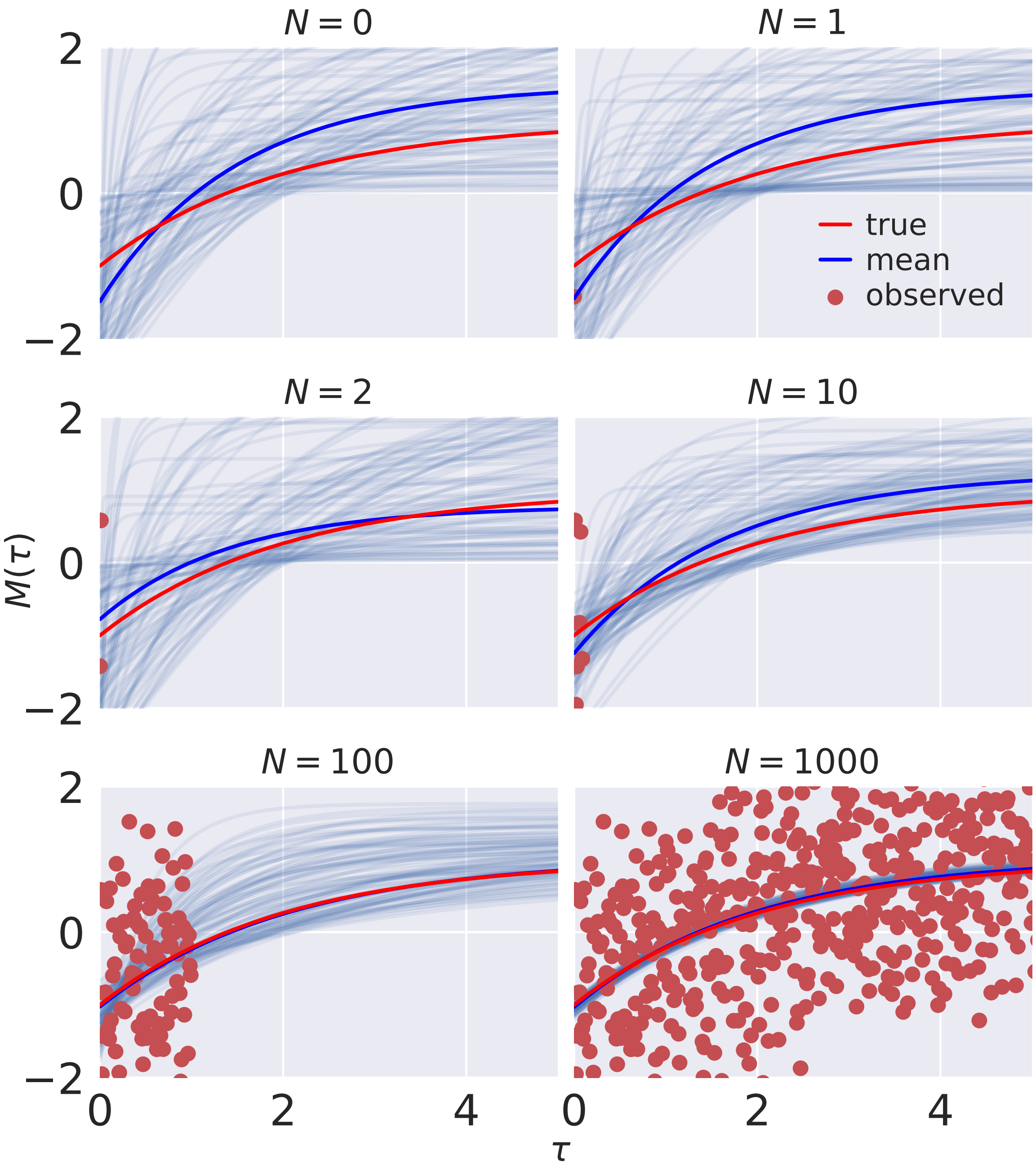


Figure 6.3: Demonstration of the ability of the SMC particle filter to estimate the parameters of an IR experiment for increasing number of particles for an SNR of 1. The translucent curves are IR curves with model parameters sampled from the current posterior distribution. Note that the updater is able to provide an estimate for $N = 0$ and 1 samples which is possible due to the incorporation of prior information for $M(\infty)$ and T_1 within the Bayesian framework.

In practice after many updates, all weights except one will tend to zero, resulting in numerical instabilities [3]. To rectify this resampling techniques such as Liu-West resampling are applied [54], which redistribute the particles within the parameter space and reset their weights to unity. Resampling allows the SMC particles to move toward the true parameter values and explore the high dimensional parameter space efficiently as demonstrated in figure 6.4.

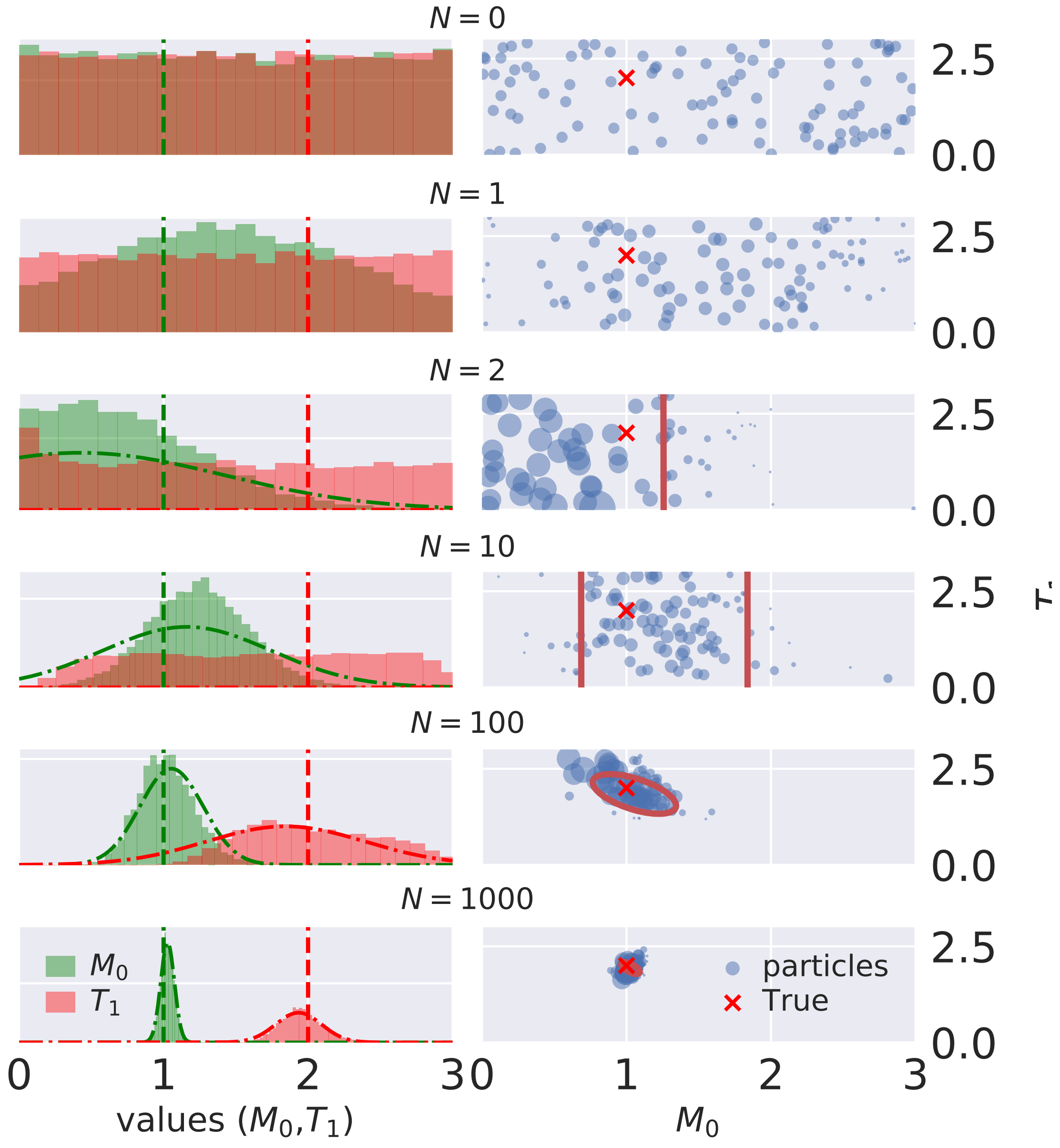


Figure 6.4: Simulation of a single linear spaced design (see equation (8.2)) with $\tau_0 = 0$ and τ_{max}=10, demonstrating the particle filter behaviour as additional data is acquired with an SNR of 10. (left) Marginalized posterior distributions for $M(\infty) = 1$ and $T_1 = 2$ with the true values given by the vertical dashed line. The dash-dotted curve is a traditional least squares fit for the dataset. (right) Posterior distribution particles with their radius being proportional to the respective particle weight. The red ellipse is a 95% confidence interval for the particle filter distribution. Note that as data acquisition proceeds the particles redistribute themselves through resampling.

The expectation value of an operator $F(\mathbf{x})$ is calculated under the SMC approximation

$$\mathbb{E}_{\mathbf{x}|\pi_{\mathbf{X}}}[F(\mathbf{x})] = \sum_{i=1}^{n_{par}} w_i F(\mathbf{x}_i). \tag{6.10}$$

The Python SMC package QInfer is used for all simulations in this thesis [31].

6.2 Sequential Bayesian Experiment Design

For an experiment there is a set of knobs (experimental control parameters) that may be varied. How to optimally set these to maximize how much may be learned about the experimental system is often not considered. An interesting question is whether the "best experiment" which extracts the most information about a set of parameters can be chosen. In a Bayesian framework using information theory this question may be posed as "Find the experiment that will maximize the information gain in our posterior probability distribution relative to the prior distribution?"

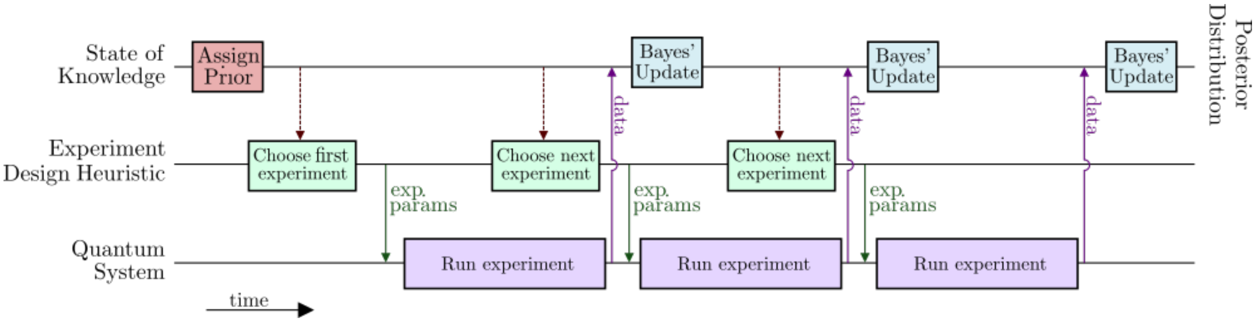


Figure 6.5: Diagram of the sequential online experiment design procedure we advocate. At every step the optimal experiment given the current prior distribution over model parameters is chosen. Figure used and modified with permission [42].

6.2.1 Loss Functions

This question is approached by attempting to minimize some appropriately chosen loss function. The naive (read "often best") way of doing this is to minimize the distance between the "true" value of the unknown parameters $\mathbf{x}$ to be estimated, and the current estimate which is a function of the data already observed. Put more formally the mean squared error (MSE) loss function will be minimized

$$L_Q(\mathbf{x}, \hat{\mathbf{x}}(O;C)) = (\mathbf{x} - \hat{\mathbf{x}}(O;C))^T \mathbf{Q} (\mathbf{x} - \hat{\mathbf{x}}(O;C)). \tag{6.11}$$

Where $\mathbf{x}$ is a vector of true parameter values to be estimated, and $\hat{\mathbf{x}}(O; C)$ is an estimate of these parameter values given the set of all observed datum $O = \{o_1, \cdots, o_n\}$, and experimental parameters for each observation $\mathbf{C} = \{\mathbf{c_1}, \cdots, \mathbf{c}_n\}$. The matrix $\mathbf{Q}$ is a positive semidefinite scaling matrix with inverse units that weights the relative importance of individual model parameters and their covariances to the loss. Typically $\mathbf{Q}$ is chosen to be diagonal, which has the effect of disregarding parameter covariances in the loss function. However, there is no such requirement in general. Note that equivalently the utility function defined as the negative loss

$$U(\mathbf{x}, \hat{\mathbf{x}}) = -L(\mathbf{x}, \hat{\mathbf{x}}), \tag{6.12}$$

may be maximized, rather than minimizing the loss function. In section A.3 it is shown how often it is useful to rescale the weighting matrix by the initial prior, to properly weight model parameter with different magnitudes.

In optimal experiment design it is desired to select the parameters $\mathbf{c}_{n+1}$ that minimize the loss considering all possible observations. With each experiment trial, the observed data is independent of prior experiments and sampled from a noise distribution which itself is a function of the unknown model parameters. The best guess for the model parameters given the observed data is given by the current prior distribution π_n. For every experiment outcome o_i there is an associated likelihood for this outcome under the settings $\mathbf{c}_i$, $l(o_i|\mathbf{x}, \mathbf{c}_i)$, which depends on the true values of the parameters to be estimated. Define the risk of the experiment to be the expected loss over all possible experiment outcomes given the "true" knowledge (which from a pure Bayesian perspective is not obtainable) of the unknown parameters

$$R(\mathbf{x}, \hat{\mathbf{x}}; \mathbf{c}_i) = \mathbb{E}_{o_i|x;\mathbf{c}_i}[L_Q(\mathbf{x}, \hat{\mathbf{x}}(o_i; \mathbf{c}_i))]. \tag{6.13}$$

Of course, the parameters $\mathbf{x}$ are not known precisely (except within simulations), and therefore the next best thing, the prior information, must be used. The prior is a probability distribution $\Pr(\mathbf{x}|I) = \pi(\mathbf{x})$, where I denotes the prior information. In the case of experimental design, previous experiments will be contained within the set of prior information I, and the prior information $\pi_n(\mathbf{x}) = \Pr(\mathbf{x}|O_n, I; \mathbf{C}_n)$. The Bayes risk is, therefore, the average risk value over the prior distribution $\pi(\mathbf{x})$

$$r(\pi; C) = \mathbb{E}_{\mathbf{x}}[R(\mathbf{x}, \hat{\mathbf{x}}; C)] = \int \pi(\mathbf{x}) R(\mathbf{x}, \hat{\mathbf{x}}; C) d\mathbf{x}. \tag{6.14}$$

The Bayes risk is also referred as Bayesian A-optimality [18]. When the MSE is the loss function of interest the unique unbiased estimator that minimizes the loss function is the mean of the posterior distribution $\hat{\mathbf{x}}(O; C) = \mathbb{E}_{\mathbf{x}|O;C}[\mathbf{x}]$.

6.3 Sequential Experimental Design

The goal of experimental design is to choose experimental parameters $\mathbf{c}_{opt}$ to gain the most information adaptively (minimize the loss function) about the unknown parameters $\mathbf{x}$ with each experiment.

$$\mathbf{c}_{opt} = \arg\min_{\mathbf{c}} L(\pi_{\mathbf{X}}, \mathbf{c}). \tag{6.15}$$

By greedily optimizing sequential experiments each experiment is chosen with the best current knowledge.

6.3.1 Fisher Information and the Cramér-Rao lower bound

The inverse fisher information (FI) of a probability distribution lower bounds the variance of any unbiased estimator, which is known as the Cramér-Rao lower bound (CRB) [21]. The Fisher information is the second moment of the score

$$\mathbf{I}(\mathbf{x}) = \mathbb{E}_{O|\mathbf{x};\mathbf{C}}\left[\nabla_{\mathbf{x}} \log(\Pr(O|\mathbf{x};\mathbf{C})) \cdot \nabla_{\mathbf{x}}^{T} \log(\Pr(O|\mathbf{x};C))\right]. \tag{6.16}$$

The score is a measure of how sensitive a given distribution is to parameter variations. The FI can be seen as a measure of the curvature around the maximum of the probability distribution and is additive for independent experiments,

$$\mathbf{I}_{\mathbf{C}_{n+1}}(\mathbf{x}) = \mathbf{I}_{\mathbf{c}_{n+1},\mathbf{C}}(\mathbf{x}) = \mathbf{I}_{\mathbf{c}_{n+1}}(\mathbf{x}) + \mathbf{I}_{\mathbf{C}}(\mathbf{x}).$$

It is, therefore, possible to compute the FI for experiments performed both in parallel and sequentially.

The Cramér-Rao lower bound (CRB), which states that given an unbiased estimator $\hat{\theta}$ the lower bound of the variance for the estimator is given by the inverse of the FI matrix.

$$\mathbf{cov}_{\hat{\mathbf{x}}} \geq \mathbf{I}(\theta)^{-1}. \tag{6.17}$$

This bound will be used extensivly and it is important to explore some aspects of it in detail. The CRB is defined for fixed known model parameters $\mathbf{x}_0$, *i.e.* a Dirac delta prior distribution. Importantly it is agnostic towards the choice of estimator provided the estimator is unbiased. The CRB allows the question “Let $\mathbf{x_0}$ be the true value of $\mathbf{x}$, how small can the variance of the estimator $\hat{\mathbf{x}}$ be made?” to be asked. The Bayesian generalization of the CRB is the Van-Tree’s inequality and has the form of the expectation of the FI matrix over the distribution $\pi(\mathbf{x})$ [83]. This weights the FI evaluation at a specific value of $\mathbf{x}$ proportionally to the belief it is the true value $\mathbf{x_0}$ according to the prior information. The Van-Trees inequality is otherwise known as the Bayesian Cramér-Rao lower bound (BCRB). The weighted FI matrix is known as the Bayesian information matrix (BIM)

$$\mathbf{J_D}(\pi; C) = \mathbb{E}_x\left[\mathbf{I}(\mathbf{x}; C)\right]. \tag{6.18}$$

The FI of the prior $\pi(\mathbf{x})$ is given by

$$\mathbf{J_P}(\pi) = \mathbb{E}_O\left[\nabla_{\mathbf{x}} \log(\pi(\mathbf{x})) \cdot \nabla_{\mathbf{x}}^{T} \log(\log(\pi(\mathbf{x})))\right] \tag{6.19}$$

The Van Tree’s inequality bounds the variance

$$\mathbf{cov}_{\hat{\mathbf{x}}|\pi} \geq \left(\mathbf{J_D} + \mathbf{J_P}\right)^{-1}. \tag{6.20}$$

Fisher Information of the Inversion Recovery Model

Under the assumption of normal additive noise, and a continuously differentiable model function $f(\mathbf{x})$, the fisher information matrix (FIM) has the simple form

$$I_{ij}(\mathbf{x}) = \frac{f_i(x) f_j(x)}{\sigma^2}, \tag{6.21}$$

where

$$f_i(\mathbf{x}) = \frac{\partial f(\mathbf{x})}{\partial x_i}, \tag{6.22}$$

is the derivative of the model function with respect to the j^{th} model parameter. The FIM for the IR model given in equation (6.1) with $a = 2$ is

$$\mathbf{I}_{ir}(M(\infty), T_1 | \tau; \sigma) = \frac{1}{\sigma^2} \begin{pmatrix} \left(1 - 2e^{-\frac{\tau}{T_1}}\right)^2 & -\left(1 - 2e^{-\frac{\tau}{T_1}}\right)\left(\frac{2M(\infty)\tau}{T1^2} e^{-\frac{\tau}{T_1}}\right) \\ -\left(1 - 2e^{-\frac{\tau}{T_1}}\right)\left(\frac{2M(\infty)\tau}{T1^2} e^{-\frac{\tau}{T_1}}\right) & \left(\frac{2M(\infty)\tau}{T1^2} e^{-\frac{\tau}{T_1}}\right)^2 \end{pmatrix} \tag{6.23}$$

Assuming perfect knowledge of the equilibrium magnetization only the inverse of the second diagonal index is examined. The CRB for estimation of the T_1 is

$$\operatorname{Var}[T_1] \geq \left(\frac{\sigma T1^2}{2M(\infty)\tau} e^{\frac{\tau}{T_1}}\right)^2 \tag{6.24}$$

Note that the CRB of T_1 is non-linear and blows up at both $\tau \to 0$ and $\tau \to \infty$ which demonstrates that measuring at these limits will give no information about the T_1. The variance grows quadratically with respect to the ratio of $\sigma / M(\infty)$, which agrees with the definition of the effective SNR in section 6.

Chapter 7

Numerical Techniques for Online Experimental Design

Here our goal for experiment design is to find optimal experiments given prior knowledge and an unbiased estimator. Consider the observation of a dataset with n datum O_n, and corresponding experimental settings $\mathbf{C}_n$. Take the initial prior to be $\pi(\mathbf{x})_n = \Pr(x|I)$ and apply Bayes' update rule to obtain the posterior $\pi(\mathbf{x})_n = \Pr(\mathbf{x}|O_n; \mathbf{C}_n)\pi(\mathbf{x})/\Pr(O_n)$. The objective is to choose a set of experimental parameters for the next experiment c_{n+1} that minimizes the expected loss. This is a greedy or sequential experiment design protocol. Alternatively the question "what are the M next best experiments to be performed?" could be posed. This is a simple generalization and is explored in section 7.3.

Let the set of all experimental parameters including the hypothetical next experiment be $\mathbf{C}_{n+1} = \mathbf{c}_{n+1} \cup \mathbf{C}_n$. Recalling that the goal of experiment design is to find the experiment c_{opt} that minimizes the Bayes' risk

$$\mathbf{c}_{opt} = \arg\min_{\mathbf{c_{n+1}}} r(\pi; \mathbf{C}_{n+1}). \tag{7.1}$$

The Bayes' risk has the form

$$\begin{aligned} r(\pi; \mathbf{c}_{n+1} \cup \mathbf{C}_n) &= \int \pi(\mathbf{x}|O_n, I; \mathbf{C}_n) R(\mathbf{x}, \hat{\mathbf{x}}; \mathbf{C}_{n+1}) d\mathbf{x} \\ &= \int \pi(\mathbf{x}|O_n, I; \mathbf{C}_n)\, \mathbb{E}_{o_{n+1}|x;\mathbf{c}_{n+1}}[L_Q(\mathbf{x}, \hat{\mathbf{x}}(O_{n+1}; \mathbf{C}_{n+1}))] d\mathbf{x} \\ &= \int \pi(\mathbf{x}|O_n, I; \mathbf{C}_n)\, \mathbb{E}_{o_{n+1}|x;\mathbf{c}_{n+1}}[L_Q(\mathbf{x}, \mathbb{E}_{\mathbf{x}|O_{n+1};\mathbf{C}_{n+1}}[\mathbf{x}])] d\mathbf{x}. \end{aligned} \tag{7.2}$$

Where $\mathbb{E}_{\mathbf{x}|O_{n+1};\mathbf{C}_{n+1}}[\mathbf{x}]$ is the posterior mean given the observation o_{n+1} and is given by

$$\mathbb{E}_{\mathbf{x}|O_{n+1};\mathbf{C}_{n+1}}[\mathbf{x}] = \int \mathbf{x} \frac{\Pr(o_{n+1}|\mathbf{c}_{n+1}, \mathbf{x})\pi(\mathbf{x}|O_n, I; \mathbf{C}_n)}{\Pr(\mathbf{x})} d\mathbf{x}. \tag{7.3}$$

The Bayes' risk may be bounded using the BCRB from equation (6.20) as discussed in section 6.3.1

$$r(\pi; C) \geq \mathrm{Tr}\left[\mathbf{Q}(\mathbf{J_D}(\pi; C) + \mathbf{J_P}(\pi; C))^{-1}\right]. \tag{7.4}$$

7.0.1 Information Gain

A downside to the use of loss functions is their dependence on the choice of estimator and the relative magnitudes of model parameters. It may be desirable to consider only the structure of the posterior distributions. Information theory may be used to compare the relative information content of two distributions using the Kullback–Leibler divergence (KL divergence)

$$O_{KL}[P(\mathbf{x})||Q(\mathbf{x})] = \mathbb{E}_{\mathbf{x}|P}\left[\log\left(\frac{P(\mathbf{x})}{Q(\mathbf{x})}\right)\right]. \tag{7.5}$$

Where $P(\mathbf{x})$ and $Q(\mathbf{x})$ are the distributions of interest. Use of the KL divergence allows optimal experiments to be selected without having to define an estimator explicitly. The objective is to choose the experiment ($\mathbf{c}_{n+1}$) that maximizes the negative KL divergence between the expected posterior distribution over all possible outcomes (weighted by the outcomes respective likelihood of occurring) and the posterior distribution from the first n experiments [44]. The risk predicts the mean information that will be gained by performing the measurement $\mathbf{c}_{n+1}$ given the current state of knowledge and is known as the expected information gain (IG). When applied to experimental design this is often referred to as Bayesian D-optimality [18]. The IG is given by

$$\begin{aligned}\overline{\mathrm{IG}}(\mathbf{c}_{n+1}) &= \mathbb{E}_{o_{n+1}}\left[\mathrm{IG}_{O_n;\mathbf{C_n}}(\mathbf{c}_{n+1})\right] & (7.6)\\ &= -\int_{\mathscr{O}}\int_{\mathbf{x}} O_{KL}[\Pr(\mathbf{x}|o_{n+1}, O_n; \mathbf{c}_{n+1}, \mathbf{C}_n)||\Pr(\mathbf{x}|O_n; \mathbf{C}_n)]\mathrm{d}\mathbf{x}\mathrm{d}o_{n+1}. & (7.7)\end{aligned}$$

The maximal c_{n+1} is optimal in the sense that it is the experiment that has the"tightest" expected posterior distribution. However, as discussed in the following sections, maximizing the information gain has several implementation issues. Choosing the experiment parameter that optimizes the information gain of the posterior distribution does not necessarily select the experiment that will minimize the covariance of the estimator. The Bayes' risk is upper bounded by the BCRB, whereas it is difficult to derive useful upper bounds for the expected information gain.

7.1 Evaluating the Bayes' Risk with Continuous Outcome Probability Distributions

The Bayes' risk, equation (7.2), provides a quantity for adaptively choosing experimental parameters. For most likelihood model functions the integrals in the equation are impossible to solve analytically. It is therefore necessary to use numerical algorithms to approximate the integrals over posterior distributions. The Bayes' risk (7.2) may be evaluated under the SMC approximation, equation (6.6). Assuming the previous execution of experiments $\mathbf{c}_{n+1}$ and corresponding observed data o_{n+1} the mean estimator is defined as

$$\begin{aligned}\hat{\mathbf{x}}(\pi_{\mathbf{X}}, o_{n+1}; \mathbf{c}_{n+1}) &\coloneqq \mathbb{E}_{\mathbf{x}|\pi_{\mathbf{X}}, o_{n+1}; \mathbf{c}_{n+1}}[\mathbf{x}] \\ &= \sum_{i=1}^{n_{par}} w_{i|o_{n+1}} \mathbf{x}_i. \end{aligned} \tag{7.8}$$

Applying the particle filtering approximation to the rest of (7.2)

$$\begin{aligned} r(\pi; \mathbf{c}_{n+1}) &= \sum_{i=1}^{n_{par}} w_i \int_{\mathscr{O}} \mathrm{d}o_{n+1} \Pr(o_{n+1}|\mathbf{x}_i; \mathbf{c}_{n+1})[L_Q(\mathbf{x}_i, \hat{\mathbf{x}}(\pi_{\mathbf{X}}, o_{n+1}; \mathbf{c}_{n+1}))] \\ &= \sum_{i=1}^{n_{par}} w_i \int_{\mathscr{O}} \mathrm{d}o_{n+1} \Pr(o_{n+1}|\mathbf{x}_i; \mathbf{c}_{n+1})(\mathbf{x}_i - \hat{\mathbf{x}}(\pi_{\mathbf{X}}, o_{n+1}; \mathbf{c}_{n+1}))^T \mathbf{Q}(\mathbf{x}_i - \hat{\mathbf{x}}(\pi_{\mathbf{X}}, o_{n+1}; \mathbf{c}_{n+1})). \end{aligned} \tag{7.9}$$

The integral in equation (7.9) is generally not analytically solvable. In the case where the possible outcomes are discrete and drawn from a known set of possible outcomes O, the Bayes' risk is calculable under the SMC approximation as

$$r(\pi; \mathbf{c}_{n+1}) = \sum_{i=1}^{n_{par}} w_i \sum_{j \in O} \Pr(o_j|x_i; \mathbf{c}_{n+1})(\mathbf{x}_i - \hat{\mathbf{x}}(\pi_{\mathbf{X}}, o_j; \mathbf{c}_{n+1}))^T \mathbf{Q}(\mathbf{x}_i - \hat{\mathbf{x}}(\pi_{\mathbf{X}}, o; \mathbf{c}_{n+1})). \tag{7.10}$$

Even the normal distribution, which is a well behaved continuous outcome distribution, is not analytically solvable apart from the case of a linear model function. This difficulty is bypassed by approximating the integral in equation (7.2) using Monte Carlo integration with importance sampling.

The first step is to draw a set of n_{sam} particles (model parameter vectors) $\mathbf{X}_{sam} = \{\mathbf{x}_1, \ldots, \mathbf{x}_{n_{sam}}\}$ from the prior distribution

$$\mathbf{x}_j \sim \pi_{\mathbf{X}}(\mathbf{x}) = \sum_{i}^{n_{par}} \omega_i \delta(\mathbf{x} - \mathbf{x}_i). \tag{7.11}$$

Given the outcome distribution $\Pr(o|\mathbf{x}_i)$, for each particle $\mathbf{x}_j$, sample a single outcome from the likelihood function

$$o_j \sim \Pr(o_{n+1}|\mathbf{x}_j; \mathbf{c}_{n+1}) \quad \forall \mathbf{x}_j \in \mathbf{X}_{sam}, \tag{7.12}$$

to generate the set of outcomes $O_{sam} = \{o_1, \ldots, o_{n_{sam}}\}$. The set of sampled outcomes values may be used to approximate $\Pr(o_{n+1})$ as a sum of delta functions with uniform weighting

$$\Pr(o_{n+1}) \approx \frac{1}{n_{sam}} \sum_{o_j \in O_{sam}} \delta(o_{n+1} - o_j). \tag{7.13}$$

Recalling the second line of equation (7.9) the integral-sum may be rewritten as

$$\int_{\mathscr{O}} \mathrm{d}o_{n+1} \left(\sum_{i=1}^{n_{par}} \pi_{\mathbf{X}}(\mathbf{x}_i) \Pr(o_{n+1}|x_i; \mathbf{c}_{n+1}) \right) = \int_{\mathscr{O}} \mathrm{d}o_{n+1} \Pr(o_{n+1}), \tag{7.14}$$

therefore the Monte Carlo integrated risk is

$$r(\pi; \mathbf{c}_{n+1}) = \frac{1}{n_{sam}} \sum_{i=1}^{n_{sam}} (\mathbf{x}_i - \hat{\mathbf{x}}(\pi_{\mathbf{X}}, o_i; \mathbf{c}_{n+1}))^T \mathbf{Q} (\mathbf{x}_i - \hat{\mathbf{x}}(\pi_{\mathbf{X}}, o_i; \mathbf{c}_{n+1})). \tag{7.15}$$

Selecting a larger n_{sam} will increase the accuracy of the Bayes' risk evaluation. In the limit of large samples, convergence is guaranteed [81], although a variety of issues arise in practice such as floating point rounding errors and excessive memory costs.

7.1.1 Bayes' Risk Evaluation: Maximum Importance Sampling

It is important to note that in the above description each sampled particle (model parameter) is associated with a single outcome, which is computationally inefficient. If an outcome is drawn under model parameter $\mathbf{x}_i$ it is likely a valid outcome for model parameter $\mathbf{x}_j$, just with a different likelihood. To calculate the risk with a single set of sampled outcomes integrated for all model parameters in the prior distribution, each outcome o_i is weighted by its likelihood under model parameter $\mathbf{x}_j$, which is known as maximum importance sampling (MIS) and allows exploration of a greater portion of the outcome domain for each particle [81]. Consider the Monte Carlo Integration example below,

$$\mathbb{E}_{\pi(\mathbf{x})}[f(\mathbf{x})] = \int_{\mathbf{X}} f(\mathbf{x}) \pi(\mathbf{x}) \mathrm{d}x \approx \frac{1}{n} \sum_{i=1}^{n} f(\mathbf{x}_i). \tag{7.16}$$

A set of points sampled from another distribution $g(x)$ may be used by multiplying the above integral with the unity factor $\frac{g(\mathbf{x})}{g(\mathbf{x})}$, where the numerator is approximated by a set of samples and the denominator is an analytic weighting for each sample

$$\mathbb{E}_{\pi(\mathbf{x})}[f(\mathbf{x})] = \mathbb{E}_{g(\mathbf{x})}[\frac{f(\mathbf{x})\pi(\mathbf{x})}{g(\mathbf{x})}] = \int_{\mathbf{X}} f(\mathbf{x}) \pi(\mathbf{x}) \frac{g(\mathbf{x})}{g(\mathbf{x})} \mathrm{d}x \approx \frac{1}{n} \sum_{i=1}^{n} \frac{f(\mathbf{x}_i)\pi(\mathbf{x}_i)}{g(\mathbf{x}_i)}. \tag{7.17}$$

Applying this technique to the Bayes' risk equation (7.15) for the sampling distribution $\Pr(o_{n+1})$, and the prior particle distribution $\pi_{\mathbf{X}}(\mathbf{x})$, the MIS approximation

$$\begin{aligned} r(\pi;\mathbf{c}_{n+1}) &= \int\limits_{\mathscr{X}} \mathrm{d}\mathbf{x}\pi(\mathbf{x}|O_n, I; \mathbf{C}_n) \left(\int\limits_{\mathscr{O}} \mathrm{d}o_{n+1} \Pr(o_{n+1}|\mathbf{x};\mathbf{c}_{n+1}) L_Q(\mathbf{x}, \hat{\mathbf{x}}(\pi_{\mathbf{X}}, o_{n+1};\mathbf{c}_{n+1})) \right) \\ &\approx \sum_{i=1}^{n_{par}} w_i \left(\int\limits_{\mathscr{O}} \mathrm{d}o_{n+1} \Pr(o_{n+1}|\mathbf{x}_i;\mathbf{c}_{n+1}) L_Q(\mathbf{x}_i, \hat{\mathbf{x}}(\pi_{\mathbf{X}}, o_{n+1};\mathbf{c}_{n+1})) \right), \end{aligned} \tag{7.18}$$

is obtained. Recall that the sampled set of outcomes O_{sam} approximates the distribution $\Pr(o_{n+1})$. Using MIS Monte Carlo integration with equation (7.18) and applying the marginalized probability SMC identity equation (6.9) the MIS approximation for the Bayes risk

$$\begin{aligned} r_{\text{mis}}(\pi;\mathbf{c}_{n+1}) &= \sum_{i=1}^{n_{par}} w_i \left(\frac{1}{n_{sam}} \sum_{j=1}^{n_{sam}} \frac{\Pr(o_j|\mathbf{x}_i;\mathbf{c}_{n+1})}{\Pr(o_j;\mathbf{c}_{n+1})} L_Q(\mathbf{x}_i, \hat{\mathbf{x}}_j) \right) \\ &= \frac{1}{n_{sam}} \sum_{i=1}^{n_{par}} \sum_{j=1}^{n_{sam}} \frac{w_i \Pr(o_j|\mathbf{x}_i;\mathbf{c}_{n+1})}{N_o} L_Q(\mathbf{x}_i, \hat{\mathbf{x}}_j) \\ &= \frac{1}{n_{sam}} \sum_{i=1}^{n_{par}} \sum_{j=1}^{n_{sam}} \frac{w_i \Pr(o_j|\mathbf{x}_i;\mathbf{c}_{n+1})}{N_o} L_Q(\mathbf{x}_i, \hat{\mathbf{x}}_j) \\ &= \frac{1}{n_{sam}} \sum_{j=1}^{n_{sam}} \sum_{k=1}^{n_m} Q_{kk} \sum_{i=1}^{n_{par}} w_{i|o_j} (\mathbf{x}_{ik} - \hat{\mathbf{x}}_{jk})^2 && (7.19) \\ &= \frac{1}{n_{sam}} \sum_{j=1}^{n_{sam}} \sum_{k=1}^{n_m} Q_{kk} \left(\widehat{\mathbf{x}^2}_{jk} - \widehat{\mathbf{x}_{jk}}^2 \right), && (7.20) \end{aligned}$$

is obtained. Where $\widehat{\mathbf{x}_{jk}} = \sum_{i=1}^{n_{par}} w_{i|o_j}\mathbf{x}_{ik}$ and $\widehat{\mathbf{x}_{jk}} = \sum_{i=1}^{n_p} w_{i|o_j}\mathbf{x}_{ik}^2$ are the first and second posterior moments of model parameter x_k after observation o_j. The MIS Bayes risk (7.20) is the mean posterior variance with respect to the sampled data values o_j. It should be noted that when the mean is much larger then the variance, evaluating the sample variance in the form of equation (7.20) is known to be numerically unstable. In this case equation (7.19) may be used which comes at the cost of numerical efficiency. A better single pass algorithm is Welford's method [47]. The computational complexity of the MIS Bayes risk is $\mathrm{O}(n_{par}n_{sam})$ and is linear in the number of outcome samples. The number of samples may be increased until sufficient convergence is obtained. The MIS Bayes risk is a random variable dependent on draws from the model likelihood posterior distribution, consequently it can be difficult to find the numerical minimum with traditional optimization methods. A method to resolve this for additive normal noise is provided in section A.1.

7.1.2 Bayes Risk Evaluation: Effective Strong Measurements

At its core, Monte Carlo integration functions by discretizing the outcome distribution for the model under study, such that the Bayes risk may be efficiently calculated. An alternative proposed heuristic has its roots in the microscopic nature of the macroscopic magnetization. The measurement of normal additive noise is discretized to N effective strong measurements of a binomial distribution such that the information content of a measurement from this distribution will on average be equal to that of the additive noise measurement. The outcome of the binomial distribution will have N discrete possibilities and allow a simple numerical evaluation of the Bayes risk with the discrete risk from equation (7.10).

Recalling the observed magnetization

$$M|M(\infty), \sigma, m \sim \mathscr{N}(M(\infty)m(t), \sigma^2). \tag{7.21}$$

Where $m(t)$ is the expectation value of σ_z

$$\mathrm{Tr}[m(t)\sigma_z]. \tag{7.22}$$

This may be rewritten in the form of a projection to the system ground state for a single spin

$$\mathrm{Tr}[m(t)\sigma_z] = \mathrm{Tr}[\rho(t)(2\left|0\right\rangle\left\langle 0\right| - \mathbb{1})] = 2\,\mathrm{Tr}[\rho(t)\left|0\right\rangle\left\langle 0\right|] - 1 = 2p(t) - 1, \tag{7.23}$$

where $p(t)$ is the probability of obtaining a ground state measurement. The explicit time dependence is now dropped for notational simplicity. Equation (7.21) may be rewritten as

$$M|M(\infty), \sigma, m \sim \mathscr{N}(M(\infty)(2p(t) - 1), \sigma^2). \tag{7.24}$$

The likelihood of having k ground state outcomes assuming N strong measurements, where $k, N \in \mathbb{Z}_{\geq 0}$ is given by the binomial distribution

$$\Pr(k; N, p) = \binom{N}{k} p^k (1-p)^{N-k}, \tag{7.25}$$

which has variance equal to the CRB

$$\mathrm{Var}[\hat{p}] = \mathbf{I}^{-1}(p) = p(1-p)/N. \tag{7.26}$$

As the binomial variance is dependent on the value of p which is not known, the average variance assuming a uniform prior for p

$$\overline{\mathrm{Var}[\hat{p}]} = \frac{1}{6N}, \tag{7.27}$$

is used. The CRB provides a lower bound for the variance of equation (7.23)

$$\mathrm{Var}[\hat{p}] \geq \mathbf{I}^{-1}(p) = \frac{\sigma^2}{4M(\infty)^2}. \tag{7.28}$$

Note that the above bound is not dependent on p. The two Fisher Informations are set equal and solved for N to obtain an estimate for the information content equivalent number of effective strong measurements

$$N = \left\lfloor \frac{2M(\infty)^2}{3\sigma^2} \right\rfloor = \left\lfloor \frac{2}{3}\mathrm{SNR}^2 \right\rfloor . \tag{7.29}$$

Note that N has been rounded down to be conservative. In the regime where $\mathrm{SNR} \leq \sqrt{3/2}$ the number of effective measurements will be $N = 0$ and evaluation of the Bayes risk will provide no information. Therefore $N \geq 1$ is enforced.

Note that the number of effective measurements scales quadratically in the SNR. In the regime where $\mathrm{SNR} \gg 1$ the effective strong measurement heuristic will become inefficient to calculate, and a new experiment optimization technique is needed.

7.1.3 Bayes Risk Evaluation: High-SNR Limit

The fisher information matrix (FIM) may be used to estimate instrument sensitivity. In the Bayesian setting for additive normal noise in the high SNR limit, the inverse FIM is equivalent to the covariance of the inferred true parameters $\mathbf{x}^{(0)}$ [82]. From this a high SNR closed form expression for the Bayes risk may be derived.

Ferrie *et. al* have observed that particle filtering distributions tend toward a normal posterior rapidly [25]. Assume the prior distribution after obtaining sufficient data O with experiments $\mathbf{C}$ over n experiments to be normal

$$\pi_{\mathbf{X}}(\mathbf{x}) = \Pr(\mathbf{x}|O;\mathbf{C}) = \frac{\mathbf{P}}{\sqrt{2\pi}} e^{-\mathbf{P}(\mathbf{x}-\hat{\mathbf{x}})(\mathbf{x}-\hat{\mathbf{x}})^T} . \tag{7.30}$$

The matrix $\mathbf{P} = \mathbf{\Sigma}^{-1} = \mathrm{cov}[\mathbf{x}]^{-1}$ is the inverse covariance matrix and $\hat{\mathbf{x}}$ the parameter means of the prior distribution, which is obtained from the particle distribution.

The likelihood of an observation o_{n+1} assuming the generic model $f(\mathbf{x};\mathbf{c}_{n+1})$ is

$$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(o - f(\mathbf{x};\mathbf{c}_{n+1}))^2}{2\sigma^2} \right) . \tag{7.31}$$

Under the assumption of Gaussian prior and noise, in the high SNR limit ($|f|^2/\sigma^2 \to \infty$) Vallisneri [82] has shown that the posterior mean is given by

$$\hat{x}_i(\pi_{\mathbf{X}}, o; \mathbf{c}_{n+1}) \approx \mathbf{\Sigma}^{(p)}_{ki} \left((o - f(\mathbf{x}^{(0)}) f'_i + \mathbf{\Sigma}^{-1}_{ik} x^{(0)}_i \right) . \tag{7.32}$$

Where $\mathbf{\Sigma}^{(p)} = (\mathbf{I} + \mathbf{P})^{-1}$ is the posterior's approximate covariance matrix with all FI components evaluated at the parameter mean $\mathbf{x}^{(0)}$ ie. $I_{ki} = I_{ki}(\mathbf{x}^{(0)}; \mathbf{c}_{n+1})$. Recall that

$$f'_k = \left. \frac{\partial f(\mathbf{x})}{\partial x_k} \right|_{\mathbf{x}=\mathbf{x}^{(0)}} , \tag{7.33}$$

and that for the case of a normal distribution the FI matrix elements are

$$I_{ij} = f_i' f_j'. \tag{7.34}$$

The intuition underlying this approximation is that if the SNR is sufficiently high, only a small region around the mean of the posterior distribution will have non negligible probability mass. The linearization of the model function may be used as it often will allow the posterior to be solved for analytically to give a linearized posterior mean. This is used to give an approximate expression for the Bayes risk in the high SNR limit gives

$$\begin{aligned} R(\mathbf{x}, \hat{\mathbf{x}}, \mathbf{c}_{n+1}) &\approx \sum_{i=1}^{N} Q_{ii}\, \mathbb{E}_{o|\mathbf{x}^{(0)};\mathbf{c}_{n+1}} \left[\left(x_i - \mathbf{\Sigma}_{ki}^{(p)} \left((o - f(\mathbf{x}^{(0)}) f_i' + \mathbf{\Sigma}_{ik}^{-1} x_i^{(0)} \right) \right)^2 \right] \\ &= \sum_{i=1}^{N} Q_{ii} {\mathbf{\Sigma}_{ki}^{(p)}}^2 \left(\mathbf{\Sigma}_{ik}^{-2} \left((x_i - \hat{x}_i) \right)^2 + {f_k'}^2 \right). \end{aligned} \tag{7.35}$$

Under the assumption of a normal prior distribution equation (7.30), equation (7.35) reduces to

$$r(\pi; \mathbf{c}_{n+1}) = \sum_{i=1}^{N} Q_{ii} \mathbf{\Sigma}_{ij}^{(p)} {\mathbf{\Sigma}_{jk}^{(p)}}^{-1} \mathbf{\Sigma}_{ki}^{(p)} = \sum_{i=1}^{N} Q_{ii} \mathbf{\Sigma}_{ii}^{(p)}. \tag{7.36}$$

As an important sanity check observe that in the limit of vanishing signal strength the risk simplifies to

$$r(\pi; \mathbf{c}_{n+1}) \xrightarrow{\mathbf{P}=0} \operatorname{Tr}\left[\mathbf{QI}\right], \tag{7.37}$$

which shows that the CRB is saturated in the limit of high SNR and an uninformative prior. This result has been derived and utilized many times within the context of Bayesian A-optimality for nonlinear model functions [18]. Equation (7.36) provides an analytical function that may be optimized to determine most informative experiments in the high signal limit. What constitutes high SNR is model dependent and must be determined numerically. Note that if the expectation of the inverse high SNR approximation over the entire prior distribution inversion is taken

$$\int (\mathbf{I}(\mathbf{x}) + \mathbf{P}) \Pr(\mathbf{x}) d\mathbf{x} = \int \mathbf{I}(\mathbf{x}) \Pr(\mathbf{x}) d\mathbf{x} + \mathbf{P} = \mathbf{J_D} + \mathbf{J_P}, \tag{7.38}$$

is obtained. This is exactly the Van Tree's inequality for the normal distribution. This may be used as an alternative risk metric, although performance is typically not much improved over the simple maximum likelihood estimate (MLE) approximation for the increased computational cost.

Note that when designing batched experiments with the high SNR Bayes risk the high SNR posterior variance is simply

$$\mathbf{\Sigma}^{(p)} = (\sum_{i=1}^{n} \mathbf{I}(\mathbf{c}_i) + \mathbf{P})^{-1}, \tag{7.39}$$

due to the additivity of the FI. Which may then be used to evaluate the Bayes risk with equation (7.36).

7.2 Comparing Bayes Risk Methods Numerical Performance

Three well motivated methods for numerically evaluating the Bayes risk have been provided in sections 7.1.1-7.1.3. These methods are complementary as they each operate in a specific regime.

The strong measurement risk is useful for evaluating the risk in the low SNR regime as it requires the sampling of very few outcomes; however, the number of the outcomes grows quadratically in the SNR. It also provides a valuable conceptual link to the discrete nature of the underlying projective measurements that are made.

As the strong measurement risk becomes inefficient to evaluate, the Bayes risk enters a middling regime where it becomes sensible to use the maximum importance sampling risk as it scales linearly with the number of samples. Provided the SNR is relatively high and the prior is not too wide, it is efficient to evaluate the risk in this manner. Note that this method makes no assumptions about the form of the likelihood or prior distributions. It may also be used to approximate the Bayes risk for finite or multi-outcome distributions such as the Poisson or multinomial distributions.

Finally, in the high SNR regime the Bayes risk becomes a functional of the FI and may be evaluated analytically after approximating the prior distribution as normal, and is, therefore, a constant time operation to evaluate. Although the high SNR risk was derived for a uniform or Gaussian prior, and a Gaussian likelihood function, in practice it can be a useful experimental design tool for more complex priors and likelihood distributions. The former due to the tendency for particle filters to approach normal posteriors due to resampling, and the latter as a result of the central limit theorem (CLT).

Studying the validity of the above regimes is necessary. While in the majority of cases the Bayes risk is not analytically solvable, it is for a linear model. Consider a scalar linear model $f(\lambda;t) = \lambda t$, with additive normally distributed noise of variance σ ie. an observation has the form

$$f_i = \lambda t_i + \epsilon_i \qquad \epsilon_i \sim \mathscr{N}(0, \sigma^2).$$

The experiment setting t is effectively a knob that may be turned to increase the SNR of the measurement and gives an effective variance $\sigma_{\text{eff}}(t) = \sigma/t$. Assuming a normally distributed prior with mean λ_0 and variance σ_0 the Bayes risk is

$$r(t) = \frac{\sigma_{\text{eff}}^2 \sigma_0^2}{\sigma_{\text{eff}}^2 + t^2 \sigma_0^2}. \tag{7.40}$$

Which as expected is the variance of the product normal distribution resultant from the likelihood and prior distribution's convolution. This simple form is a result of the linear model, and is much more complicated in general.

Evaluating the high-SNR approximate risk, the FI is simply the inverse effective variance $FI = 1/\sigma^2_{\text{eff}}$. The high-SNR risk is equal to the exact Bayes risk $r_{\text{snr}} = r$ which is not surprising as the high-SNR risk relies on the linearization of the model function to approximate the posterior variance, which in this case is linear.

The number of effective strong measurements is

$$N = \left\lfloor \frac{2}{3}\frac{1}{\sigma^2_{\text{eff}}} \right\rfloor$$

The strong measurement and MIS risks must be evaluated using particle filtering. The numerical performance of the MCMC methods for risk and information gain evaluation with the linear model are displayed in figures 7.1 and 7.2.

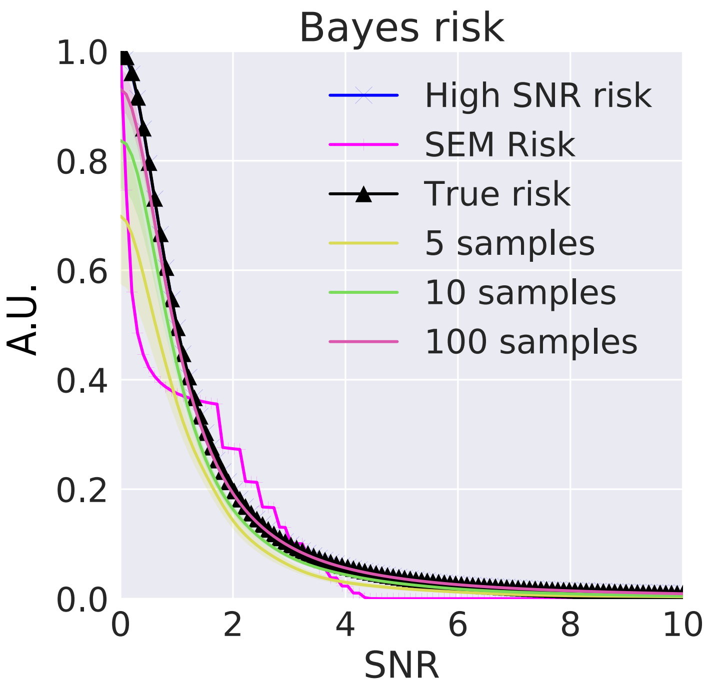


Figure 7.1: Bayes risk evaluated for the linear model with varying number of outcome samples. Same number of outcome samples and particles were used. Sampling based strategies quickly converge to the true risk.

The strong effective measurement model is discontinuous with the variation of the SNR, which is an effect of discretizing the number of effective measurements with the application of the floor function. Observe that while the ESM model follows the true Bayes risk, it is much less accurate than the sampling based risk. The poor performance of the ESM risk for the linear model is attributed to the assumption that the magnetization model function is constrained between $-M(\infty) \geq M \leq M(\infty)$ which was made in the derivation. This constraint is not satisfied for the linear model but is for the magnetization model.

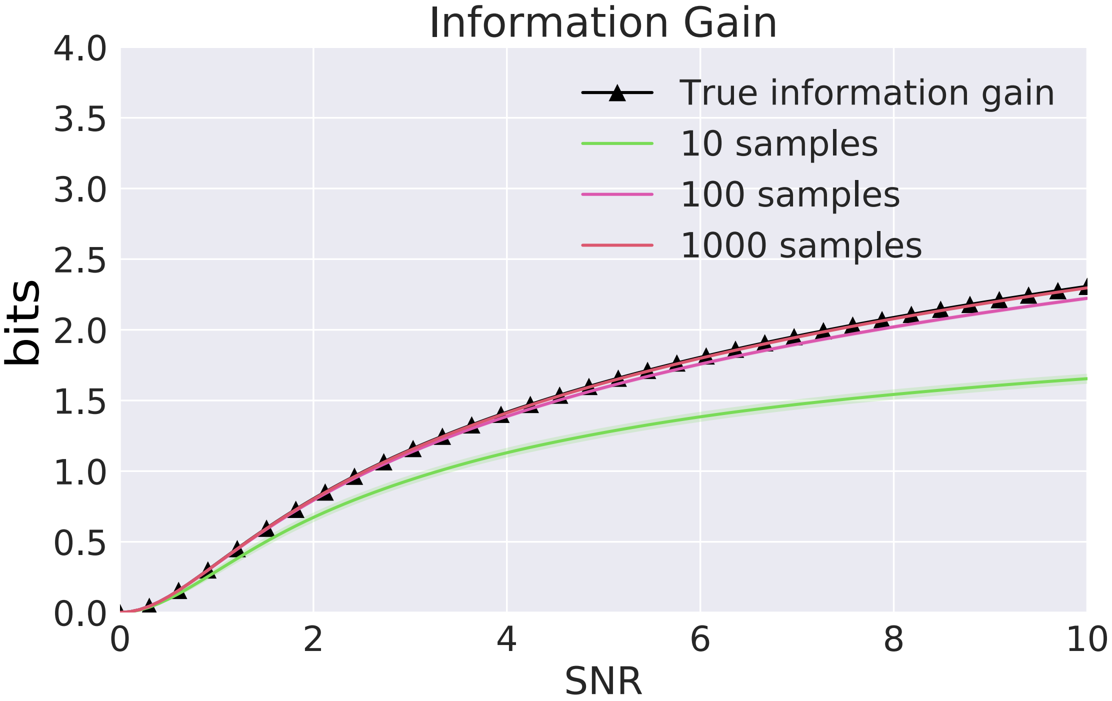


Figure 7.2: Information gain evaluated for the linear model with varying number of outcome samples. Same number of outcome samples and particles were used. Sampling based strategies quickly converge to the true risk.

The analytical infomation gain of the linear model is

$$ig(t) = \frac{1}{2}\log\left(\sigma_{eff}^2 + t^2\sigma_0^2\right) - \log\left(\sigma_{eff}^2\right). \tag{7.41}$$

In figure 7.1 observe that the MIS algorithm does a good job approximating the information gain.

7.2.1 Restricting the number of particles for Risk Evaluation

Typically when performing parameter estimation with a particle filter, the number of particles used will be on the order of thousands. As the computational complexity of the MIS sampling algorithm is $O(n_{sam}n_{par})$, if n_{par} is on the order of thousands in practice n_{sam} must be kept small to keep evaluation sufficiently fast for online experimental design. A small n_{sam}

will lead to inaccuracy in the integration procedure as it will explore only a small outcome space. The effect of downsampling the particle filter is examined by heuristically keeping only the n_{sam} most likely particles in the prior distribution $\pi_{\mathbf{X}}(\mathbf{x})$. The performance of the

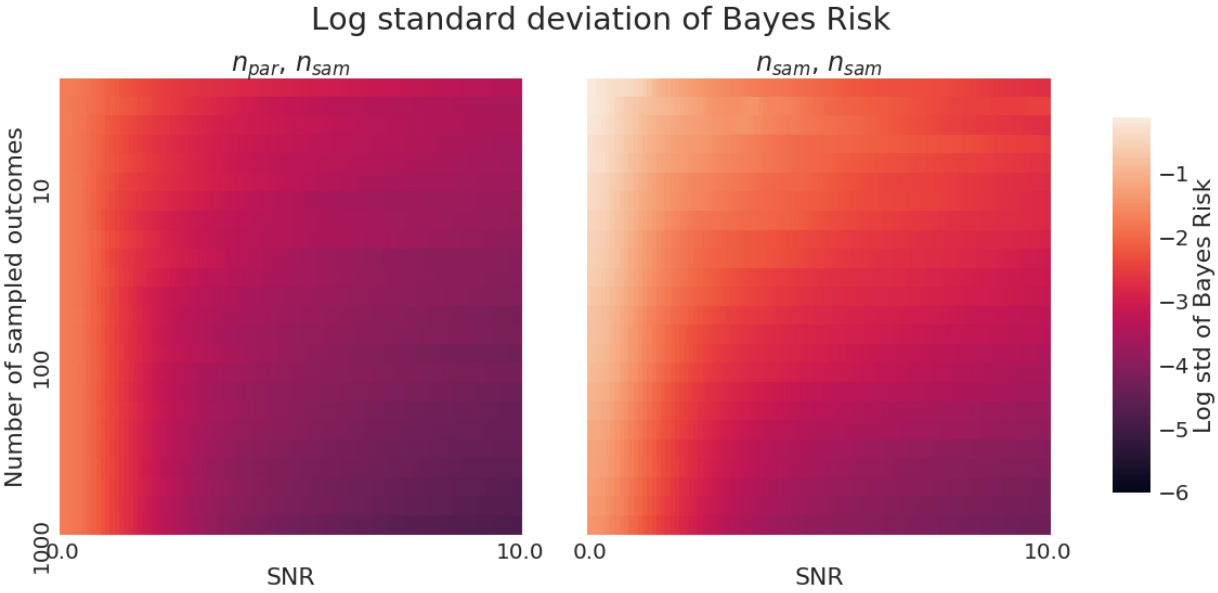


Figure 7.3: (left) Heat map of the standard deviation of the MIS Bayes risk evaluated with the linear model for a varying number of outcome samples and 5000 particles in the particle distribution. (right) Identical to (left) except with the same number of particles as outcomes samples used in the integration procedure.

full particle distribution in comparison to the above heuristic is demonstrated in figure 7.3, it is clear the performance of the full distribution is superior to that of the downsampled distribution. However, as both a greater number of samples are used and the SNR is increased the standard deviation of the heuristic becomes comparable to that of the full distribution. In practice, it is necessary to keep the evaluation time of the integration procedure small to perform online experiment design. In figure 7.4, the computational time-weighted Bayes risk standard deviation is shown. When the function evaluation time is of concern, it is useful to reduce the number of particles in the distribution and increase the number of sampled outcomes. In figure 7.5, the same behaviour holds for the evaluation of the information gain. In practice, $n_{sam} = n'_{par} = 300 - 400$ is sufficient to reduce the variation of the Bayes risk to a sufficiently low level.

7.3 Batched Experiment Design

An extension to the task of designing the best experiment given the current state of knowledge, is designing the next n experiments. This may be required depending on the capabil-

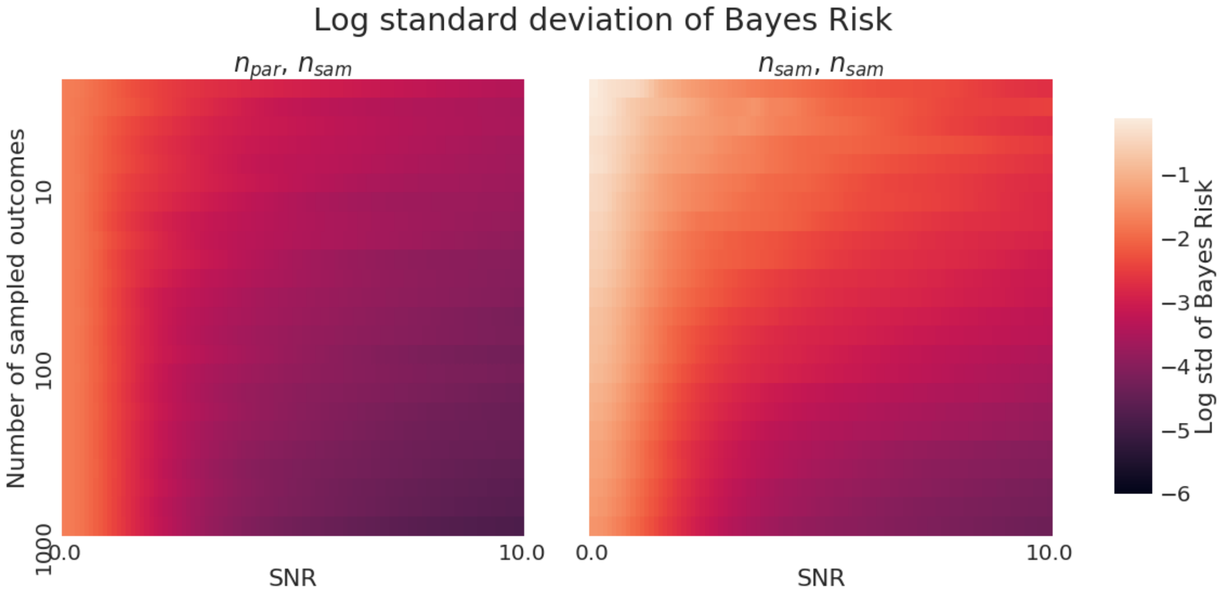


Figure 7.4: (left) Evaluation time-weighted heat map of the standard deviation of the MIS Bayes risk evaluated with the linear model with varying number of outcome samples and 5000 particles in the particle distribution. (right) Identical to (left) except with the same number of particles as outcomes samples used in the integration procedure. Performance of the downsampling heuristic is superior to that of the full distribution when evaluation time is taken into account.

ities of the experimental setup such as in the case where experiments are to be evaluated in batches due to limited device bandwidth. This question is simple to formulate within the Bayesian framework. Consider the evaluation of the Bayes risk for n experiments with experimental settings $\mathbf{C} = \{\mathbf{c}_1, \cdots, \mathbf{c}_n\}$ and associated outcomes $O = \{o_1, \cdots, o_n\}$. To evaluate the MIS Bayes risk we require the ability to calculate the likelihood distribution for a given set of experiments and the ability to draw outcomes $\Pr(O)$. The likelihood of the batched experiment is simply the product of the likelihoods of the individual experiments

$$L_{Oj} = \Pr(O|\mathbf{x}_j; \mathbf{C}) = \prod_{i=1}^{n} \Pr(o_i|\mathbf{x}_j; \mathbf{c}_i), \tag{7.42}$$

and modifying equation (7.43) to sample outcomes from $\Pr(O)$

$$o_{ji} \sim \Pr(o|\mathbf{x}_j; \mathbf{c}_i) \;\; \forall \mathbf{x}_j \in \mathbf{X}_{sam}, \; \forall \mathbf{c}_i \in \mathbf{C}. \tag{7.43}$$

With these two substitutions the MIS Bayes risk may be evaluated using equation (7.20). The computational complexity of evaluating the MIS Bayes risk for a batch of n_{exps} is $O(n_{exps} n_p n_s)$ and is therefore tractable to evaluate.

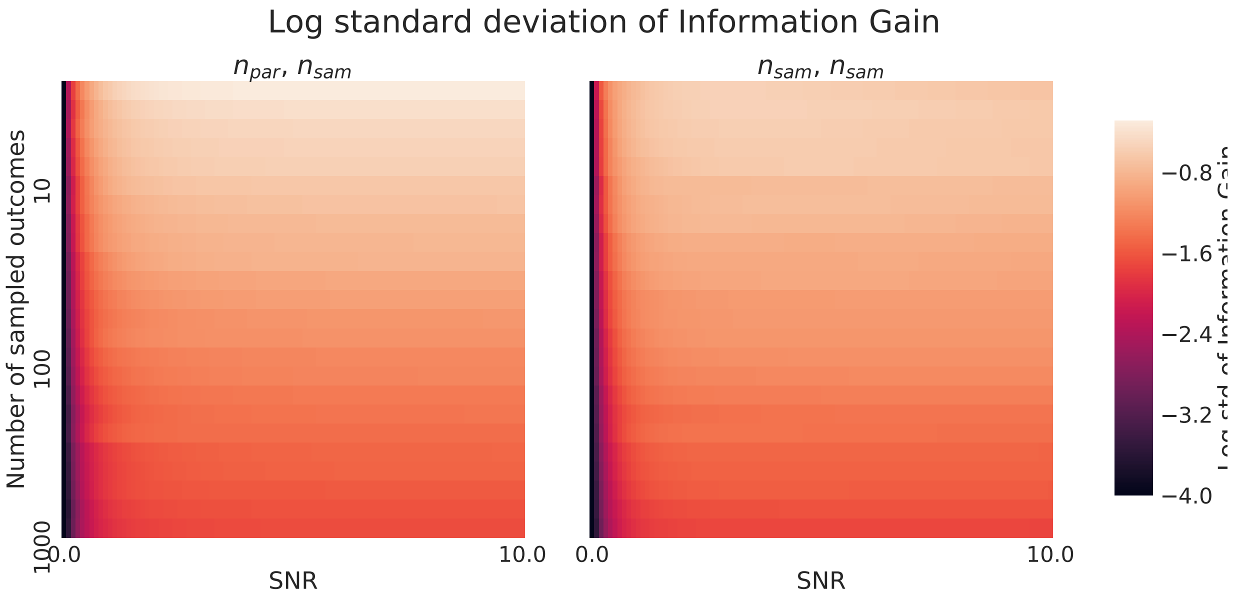


Figure 7.5: (left) Evaluation time-weighted heat map of the standard deviation of the MIS information gain evaluated with the linear model for varying number of outcome samples and 5000 particles in the particle distribution. (right) Identical to (left) except with the same number of particles as outcomes samples used in the integration procedure. Just as demonstrated in figure 7.4 for the Bayes risk the time-weighted performance is superior for the information gain when using the downsampling heuristic.

Chapter 8

Online Experimental Design applied to Inversion Recovery Experiments

In this chapter the improved estimation efficiency that may be obtained in a typical inversion recovery experiment by utilizing the experiment design heuristics described in chapter 7 is explored. From figure 8.1 the complexity of the Bayes' risk and its dependence on both the prior distribution and current SNR is clear. Observe that the full numerical Bayes' risk converges given a sufficient number of samples particles and outcomes. It is also clear that for various priors and SNR levels it diverges from the high SNR approximate Bayes risk. The reason for this is an open question; possible factors for this behavior are explored in chapter 9.

Prior to performing simulations, what constitutes a reasonable number of particle filter and outcome samples for the Bayes' risk must be determined. This is a necessary balancing act between accuracy and evaluation time. In figure 8.2 a grid search over these parameters is performed. Within this chapter 300 outcome and particle samples are used for all simulations, as the Bayes' risk varies less than 10^{-3} with respect to a 3000 outcome and particle reference Bayes' risk, while keeping the evaluation time around half a second for 300 different experiment settings on a desktop CPU.

8.1 Per Experiment Performance Enhancement of Optimal Inversion Recovery Experiments

Experiments are simulated with SNRs of 1 and 10. In any form of Bayesian analysis often the point of contention is what prior to use. When exploring the behavior of T_1 parameter estimation in inversion recovery experiments, simulations are initialized with a tight uniform prior over $M(\infty)$ (essentially perfect knowledge of the magnetization), and a broad uniform prior in the range of $0.1 - 3.0s$ for the T_1 value

$$\pi(M(\infty), T_1) = U(M(\infty), 0.999, 1.001) \cup U(\tau, 0.1, 3.0), \tag{8.1}$$

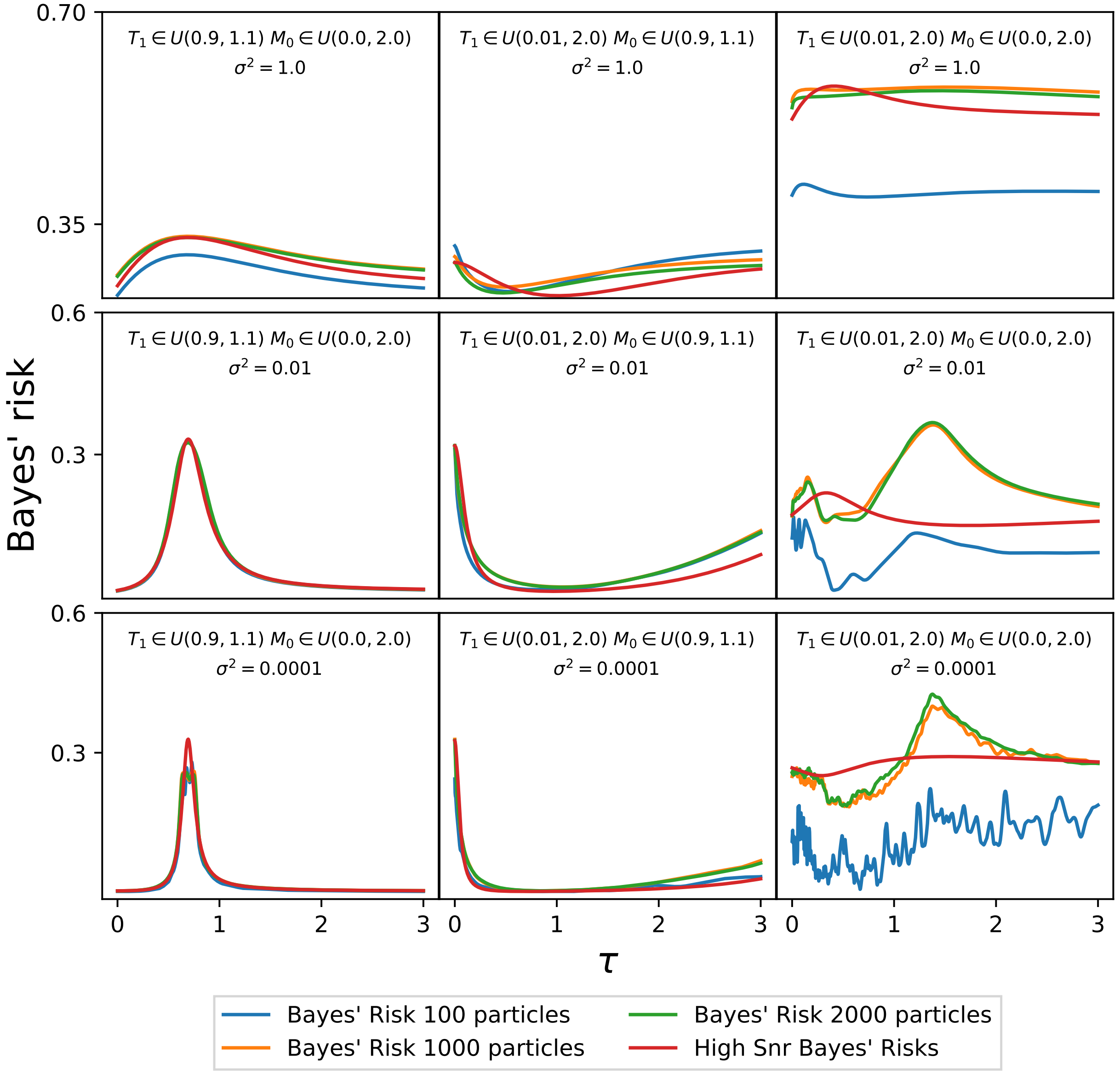


Figure 8.1: IR Bayes' risk evaluated for various prior distributions at different levels of SNR.

with 5000 particles. For a given simulation run, a set of true model parameters $\mathbf{x}_i = \{M(\infty)_i, T_{1i}\} \sim \pi(M(\infty), T_1)$ is sampled from the prior. This set of parameters is used for all experiment parameter selection techniques in a given iteration so that experiment designs may be consistently compared. Experiment times are selected from within the domain $\tau_i \in \mathrm{T} = (0, 5 \times 3.0)s$. Experiment optimization is then performed with the expected Bayes' risk over 300 linearly spaced values of τ in the range $(0, 5 \times 3.0)$ and selecting the experiment setting $\mathbf{c}_{n+1} = \tau_{n+1}$ that minimizes the respective loss function. An experiment outcome o_{n+1} is simulated for each experiment design heuristic and input into the respective particle

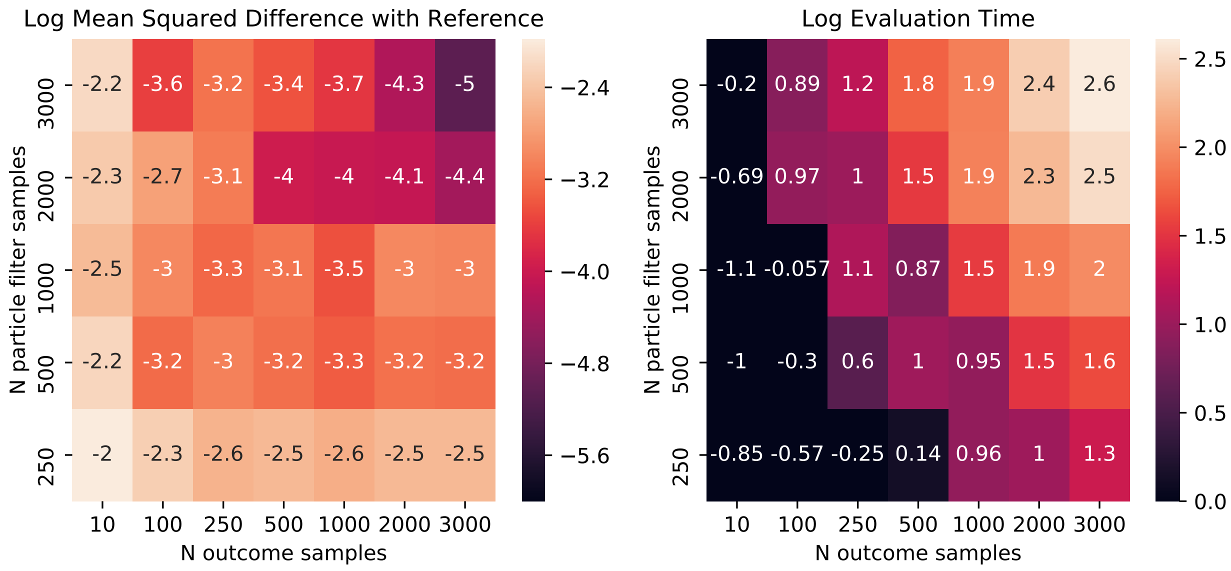


Figure 8.2: (left)Heatmap depicting the Bayes' risk variation as a function of the number of particle and outcome samples. (right)Heatmap depicting the Bayes' risk evaluation time as a function of the number of particle and outcome samples.

filter implementation to perform a Bayes' update. As outcomes are stochastically sampled it is unlikely that any two simulations will have the same results. Simulation results are then averaged over many values of x_i to obtain the average mean squared error $\overline{\mathrm{MSE}(T_1, \hat{T_1})}$ between the estimated $\hat{T_1}$, and the true value T_1, and the average variance of the posterior distributions.

For a baseline experiment design heuristic two methods are used. The first is a traditional sweep of τ where for n uniformly spaced increasing values of τ from the domain $\mathrm{T} = [\tau_1, \tau_n]$,

$$\tau_i = \frac{i-1}{n-1}(\tau_n - \tau_1) + \tau_1. \tag{8.2}$$

The second method is a uniformly random selection of N time values from $\tau_i \sim U(0, \mathrm{T})$. Note that in the limit of a large number of simulation runs and experiments per simulation the two methods should produce equivalent results as Bayes' rule is commutative for independent experiments. In practice this does not hold as resampling is used within the particle filtering algorithm which can cause experiment ordering to matter [54].

25 simulation runs, each consisting of 200 experiments are performed. The random, uniform linear sweep, MIS Bayes risk, ESM Bayes risk, high SNR Baye's risk and MIS Information gain experiment design heuristics are simulated. The ESM strategy is limited to $N_{meas} \leq 100$ to prevent excessive computational time in high SNR simulations. Note that this will introduce further error in this already approximate Bayes risk.

In figure 8.3 and figure 8.4 the average log MSE and posterior variance of T_1 are plotted respectively for varying SNRs with their 95% confidence intervals. It is immediately clear that certain strategies perform better than others.

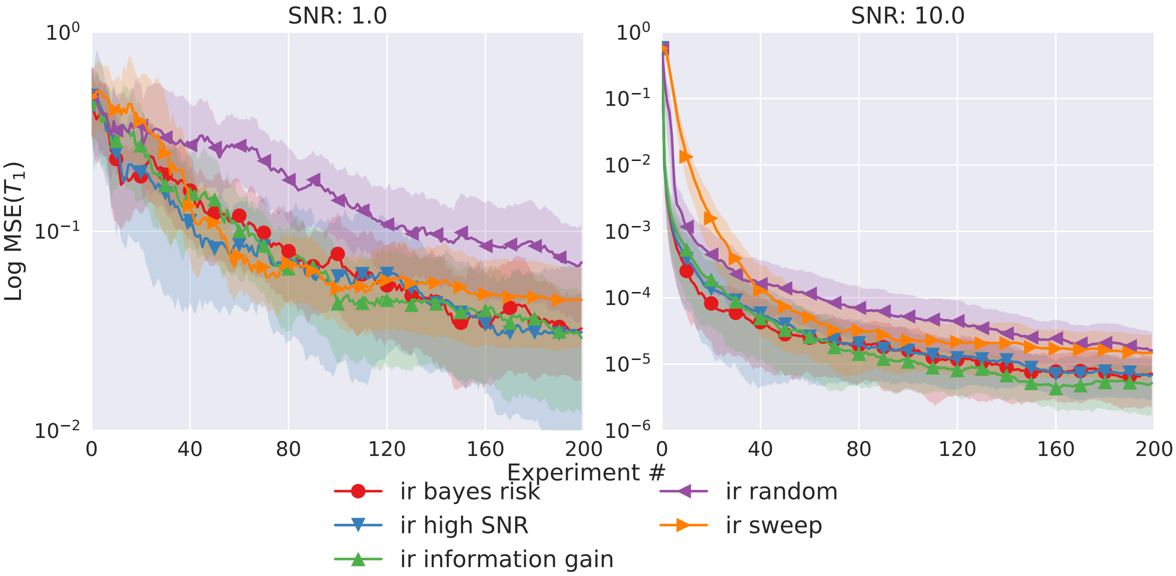


Figure 8.3: Simulated mean squared error plots for the estimated relaxation time compared to the true value for varying SNRs and fixed magnetization. Shaded regions give 95% confidence intervals. It is clear that online experimental design strategies perform better than both the swept and random baseline strategies.

The random experiment selection is clearly the worst performing design technique for learning the true T_1 value. Recalling the magnetization model function equation (6.1) this should be expected. Poor performance is attributed to the small derivative of the model function with respect to T_1 for large values of τ. For additive normal noise the result is an exponentially suppressed sensitivity of the model likelihood function to variations in T_1. Consequently the particle likelihood function at an incorrect $T_{1_\epsilon} = T_1 + \epsilon$ satisfies the relationship $\lim_{\tau\to\infty} L(T_{1_\epsilon}) = L(T_1)$ and consequently the updater learns nothing about the T_1 value in the limit of large $\tau \gg T_1$. Since both the uniform random sampling and linear sweep heuristics performs many experiments at values greater than T_1, and in this particular simulation a tight prior is placed on the magnetization, these experiments are wasted. For this reason the random experiment design also performs poorly compared to sequential Bayesian heuristics.

A quick survey of the online experimental design algorithms shows that, with exception of the ESM, they provide approximately a half order of magnitude improvement in mean squared error and variance of the T_1 posterior for a given number of experiments. In figure 8.5 it is observed that the ESM Bayes risk performs worse for all SNR values, and drastically worse performance at low SNR (although still better than baseline strategies), which is where it was hoped for the heuristic to be most accurate. As this metric is also the most

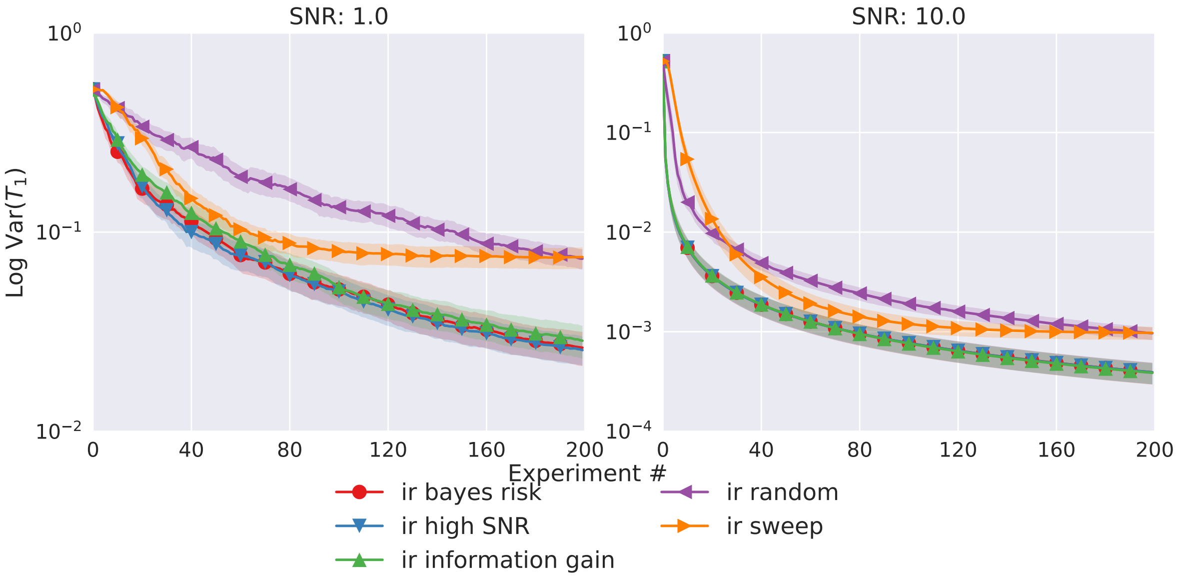


Figure 8.4: Simulated variance plots of the posterior relaxation time estimate for varying SNRs. Shaded regions give 95% confidence intervals. As in figure 8.3 it is clear that online experimental design strategies perform significantly better.

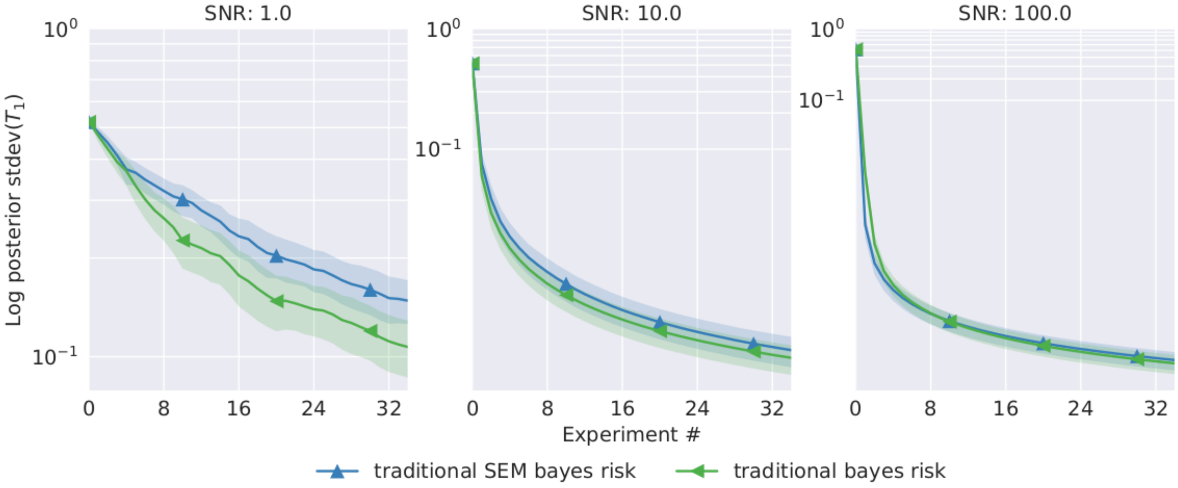


Figure 8.5: Comparison of the MIS Bayes risk with the ESM Bayes risk. The ESM approach is worse for all SNR values, with increasingly worse performance at low SNR.

expensive to compute at large SNR where it is most accurate, and the high SNR risk exists for this regime it is concluded that the ESM Bayes risk that is not useful for T_1 estimation with additive normal noise and omitted from future simulations. Examining only the online design strategies in figure 8.6, the MIS Bayes risk, information gain and high SNR Bayes risk all have comparable performance within their confidence intervals.

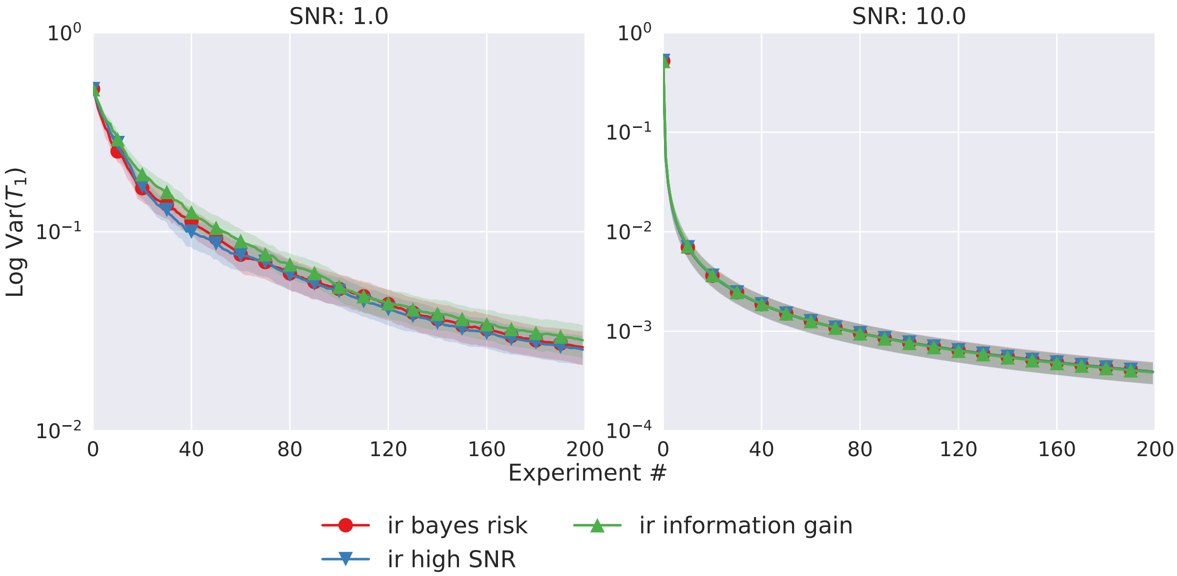


Figure 8.6: Simulated online experimental design variance plots. We note that all strategies have comparable performance regardless of SNR.

In figure 8.7 the selected experiment is plotted against the prior T_1 value at the current step. It is clear that the Bayes risk finds its minimum approximately at $\tau = T_1$, which as expected is the maximum value of the T_1 index of the FI. Similarly the same simulation as above is performed swapping the uniform prior bounds for T_1 and $M(\infty)$, ie. strong belief in the value of T_1 but not magnetization with results shown in figure 8.8 Observe that $\tau \approx 0$ is consistently selected, which is the maximal value for the magnetization index of the FIM.

8.1.1 Unknown Magnetization and Relaxation Time Simulations

The previous section was primarily concerned with learning T_1 when the magnetization is known. In practice typically both $M(\infty)$, and T_1 will have some initial non point-like prior. In figure 8.9 the average covariance of both the magnetization and relaxation time as a function of experiment number is depicted. An interesting feature is the early priority the Bayesian heuristics place on learning the magnetization. These in a sense are the 'easiest' experiments which the algorithm can choose to reduce the weighted risk quickly. After which it will focus on reducing the variance of T_1.

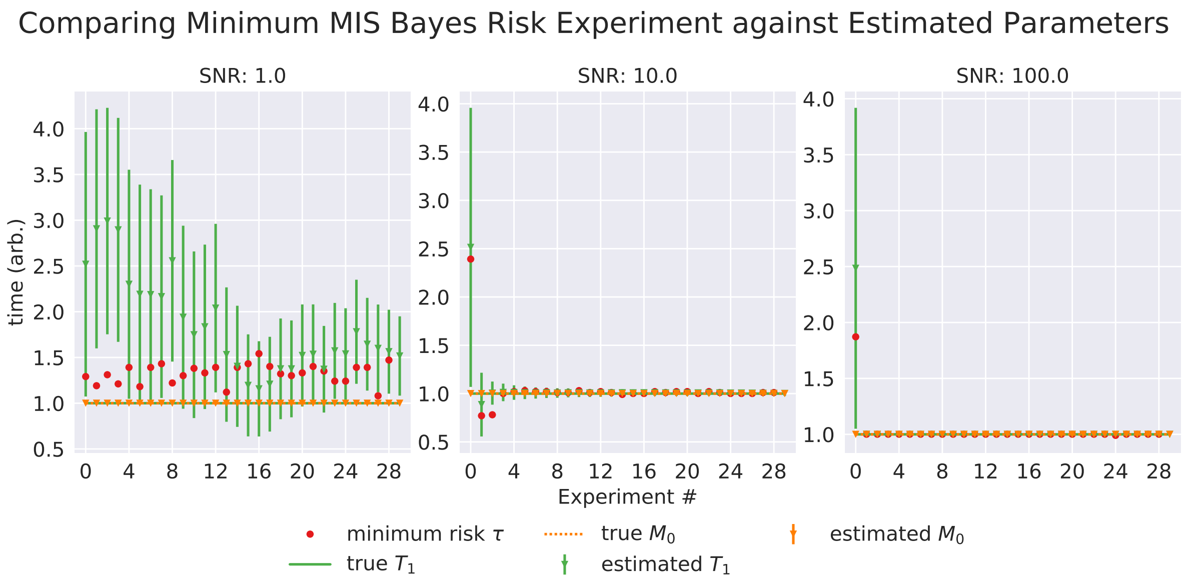


Figure 8.7: Comparison of selected experiment τ values chosen at the minimum Bayes risk compared to the mean $\hat{T}_1$ at each step for a simulation with $M(\infty) = 1$, $T_1 = 1$ true model parameters and a very tight prior over the magnetization. Observe that for low SNRs the MIS minimum risk roughly follows the current posterior belief of the T_1 model parameter, however in the higher SNR regime of 10, and 100 the minimum MIS Bayes risk is exactly the estimated T_1 value as is predicted by the high SNR Bayes risk of section 7.1.3.

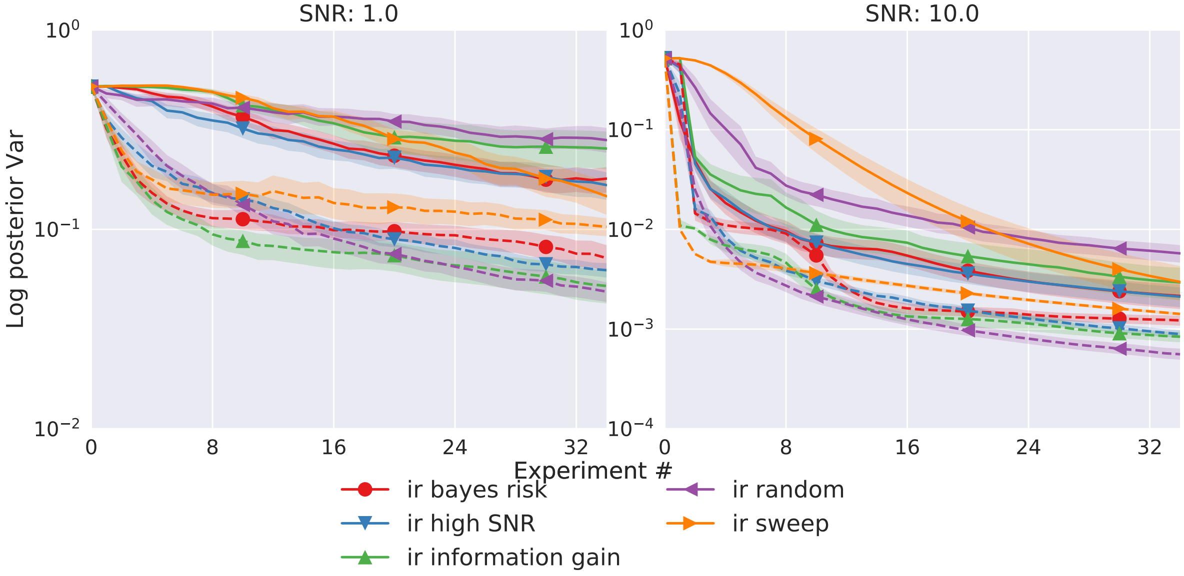


Figure 8.9: Variance plots when learning both $M(\infty)$(dashed) and T_1(solid).

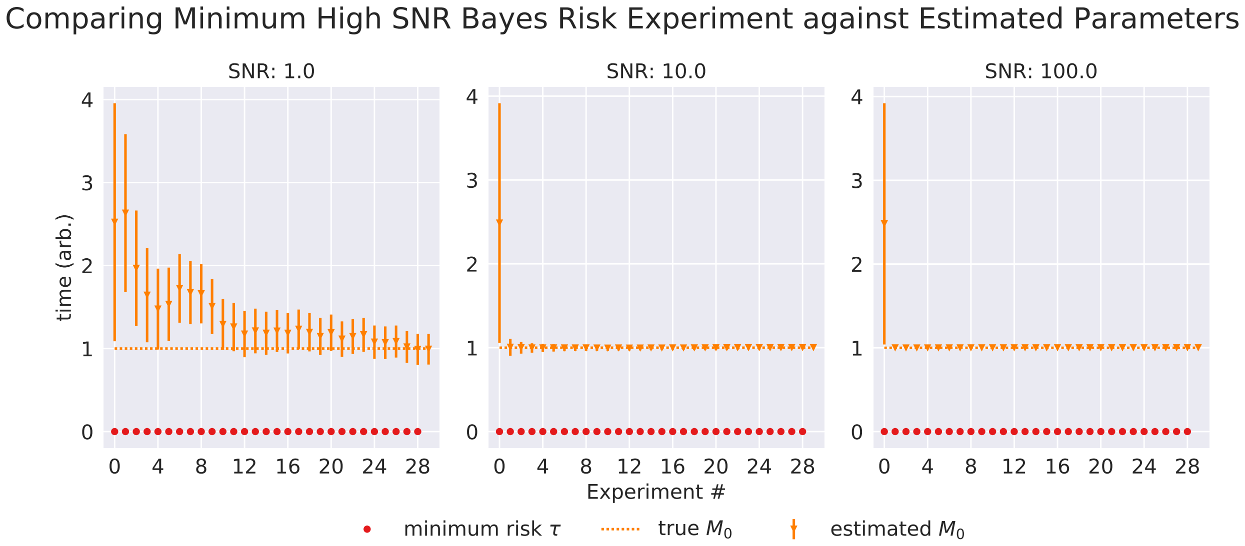


Figure 8.8: Comparison of selected experiment τ values chosen at the minimum Bayes risk compared to the mean $\hat{M(\infty)}$ at each step for a simulation with $M(\infty) = 1$, $T_1 = 1$ true model parameters and a very tight prior over the magnetization. Observe that in all SNR regimes the chosen experiment is $\tau = 0$ and is the optimal measurement point for learning of the magnetization irrespective of SNR. This the experiment that maximizes the signal regardless of the true value of T_1.

Plotting the mean squared variances compared with the BCRB value in figure 8.10 observe that the algorithm still approaches the optimal estimation strategy after a sufficient number of experiments.

8.2 Optimizing Experiments with Respect to Wall Time

Thus far minimizing the per experiment utility has been explored. This is fine for situations where the “cost” of an experiment is the same regardless of what experiment is run. For example this may correspond to cases where each experiment costs a fixed amount of currency or time, in which case there is a simple mapping between the cost and corresponding number of experiments. However in general this mapping need not, and will not be trivial. Redefining the cost weighted utility function as

$$U(\mathbf{c})_w = \frac{U(\mathbf{c})}{\mathrm{cost}(\mathbf{c})}. \tag{8.3}$$

For T_1 measurements it is generally desired to generate the best estimate for T_1 in the shortest period of time. Take the cost of an experiment to be the sum of the initialization time τ_i,

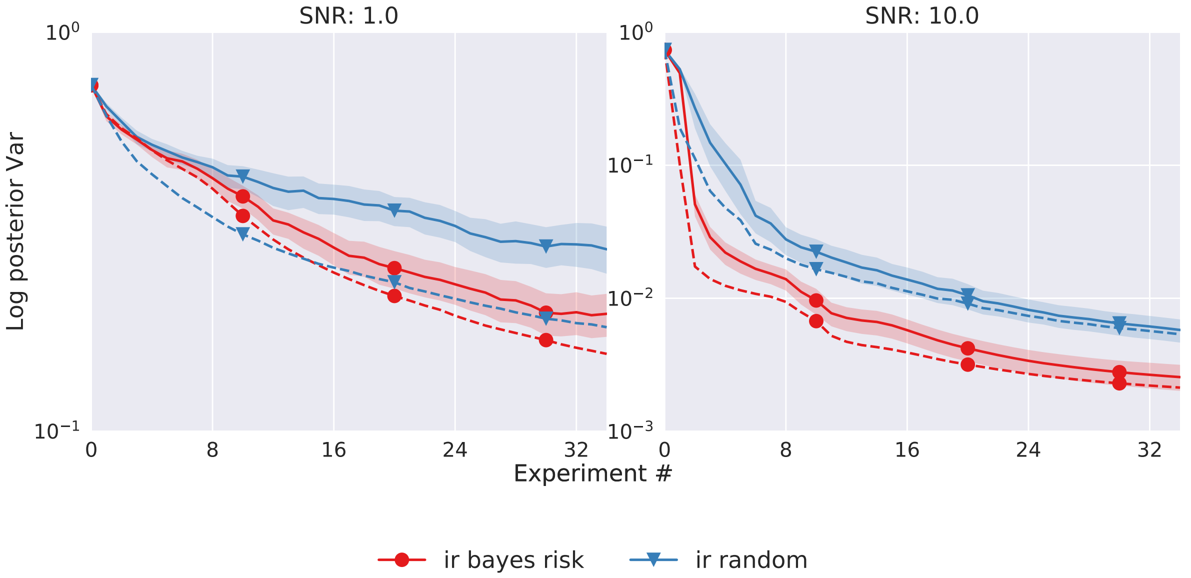


Figure 8.10: Mean squared variance plots compared to the BCRB values for both the MIS Bayes risk and random experiment design strategies when learning both $M(\infty)$(dashed) and T_1(solid). The particle filtering strategy rapidly approaches the BCRB in the high-SNR regime.

the recovery time τ and the thermalization time τ_{th} to be

$$\text{cost}(\tau) = \tau_{tot} = \tau_i + \tau + \tau_{th}. \tag{8.4}$$

The new cost weighted utility function becomes

$$U_{\tau_{tot}}(\tau) = \frac{U(\tau)}{\tau_i + \tau + \tau_{th}} \tag{8.5}$$

This is almost the utility function that must be optimized to perform a cost-weighted online experimental design of the IR experiment. However, if this is design heuristic is implemented as described above, it will be found that while the IG performs exactly as expected the Bayes risk does not.

To understand why, consider that the question the information gain utility asks is “How much more information will be contained in the average posterior distribution than the prior?”. When the experiment cost is accounted for this is now “What is the rate of information gain with respect to the experiment cost?”. This is a very intuitive question to ask and is equivalent to the rate of information gain. On the other hand, if the Bayes risk in the single model parameter case is considered the question becomes “What is the expected variance of the posterior distribution weighted by the cost of the experiment?”, which is not in the form of a rate. It is necessary to redefine the cost weighted Bayes risk in the form of a rate, in a similar manner to the IG. Define the risk improvement

$$\mathrm{ri}(\pi_{\mathbf{X}}, \mathbf{c}) = \mathrm{Tr}[\mathbf{Q}\mathrm{Cov}_{\pi_{\mathbf{X}}}(\mathbf{x})] - r(\pi_{\mathbf{X}}, \mathbf{c}). \tag{8.6}$$

The risk improvement asks "How much less will the $\mathbf{Q}$ weighted variance of the expected posterior distribution be than the weighted variance of the prior distribution?". Adding the experiment cost to the risk improvement this becomes "What is the expected rate of variance reduction per experiment cost?". This is in the same rate form as the IG and will perform comparably. Note that in the case where the experiment cost does not depend on the experimental setting $\mathbf{c}$ then

$$\arg\min_{\mathbf{c}} \frac{r(\pi_{\mathbf{X}}, \mathbf{c})}{\text{cost}} = \arg\max_{\mathbf{c}} \frac{ri(\pi_{\mathbf{X}}, \mathbf{c})}{\text{cost}},$$

is satisfied. The Bayes risk is often a satisfactory utility because the experiment cost has been implicitly taken to satisfy the above condition.

Cost weighted optimal design strategies with a $\tau_{th} = 5 \times max(T_1) = 15s$ thermalization time are explored. In practice this corresponds to each experiment starting with a minimum of 99% of the total magnetization. The interval time is set to $\tau_i = 0.5s$. Simulation results are shown in figure 8.11. The MIS and high SNR risks have near identical performance. Note that the Bayes risk strategies beat the information gain strategies for estimating relaxation time, but are edged out by the information gain strategy when learning the magnetization. Recall that the risk strategies optimize the weighted posterior variance and observe that in figure 8.13 the risk strategies as expected outperform the information gain.

When optimizing the amount learned per unit time, not experiment, it is expected that these strategies will perform poorly per experiment compared to the baseline strategies, as demonstrated in figure 8.12.

8.2.1 Faster Inversion Recovery Experiment

In the previous section the concept of minimizing the utility of an experiment not with respect to the experiment number, but instead the cost of a given experiment was introduced. For the case of a typical T_1 learning experiment the cost is the total length of an experiment, including the fixed minimum experiment interval time τ_i, and the tunable thermalization time τ_{th}. For samples with long T_1 times $\tau_{th} + \tau_i \gg \tau$ is usually satisfied. In order to increase the rate of information gain or risk improvement per unit time it is easiest to decrease τ_{th}.

Assuming a perfect inversion (π) and readout ($\pi/2$) pulses with the condition $T_1 \gg T_2$ satisfied, after an inversion recovery pulse sequence, the net magnetization is zero. The magnetization will then proceed to recover to its thermal equilibrium following the thermalization model function

$$M_{th}(M(\infty), T_1) = M(\infty)(1 - e^{-\frac{\tau_{th}+\tau_i}{T_1}}). \tag{8.7}$$

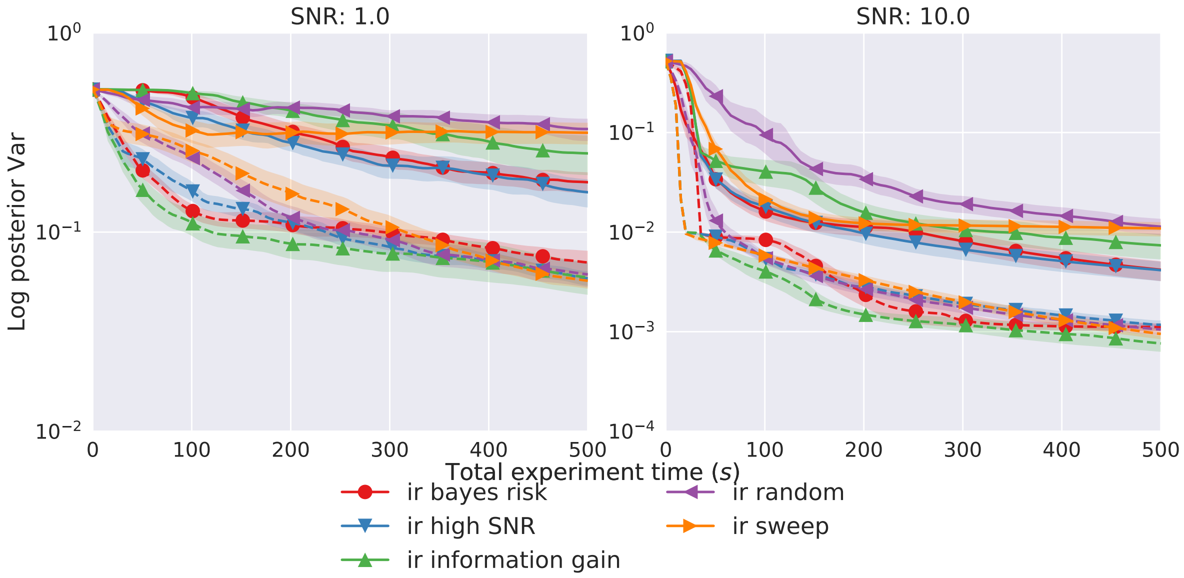


Figure 8.11: Variance plots when learning both $M(\infty)$(dashed) and T_1(solid) with cost weighted experimental design strategies.

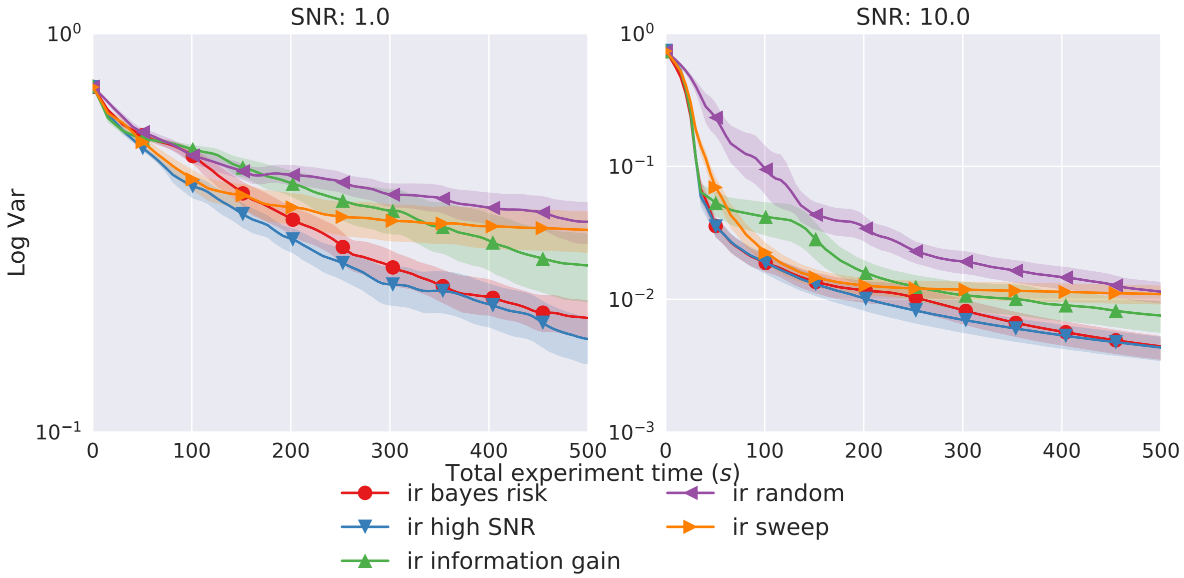


Figure 8.12: Identity weighted sum total variance plot using cost weighted experimental design strategies. Observe that the Bayes risk strategies outperform the information gain as expected.

At this point the experimentalist might realize that rather than allow the full magnetization to recover by waiting $\tau_{th} \gg \hat{T}_1$, one could let the magnetization recover only to some

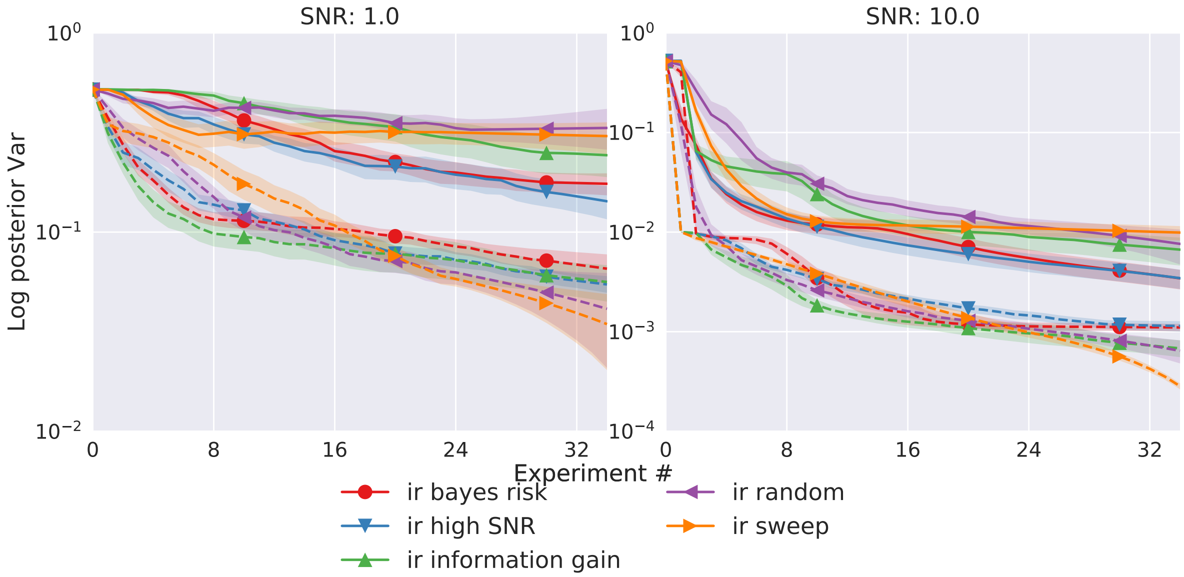


Figure 8.13: Variance plots when learning both $M(\infty)$(dashed) and T_1(solid) with cost weighted experimental design strategies as a function of experiment number. Note that as expected baseline strategies outperform the online strategies which are focused on maximizing how much can be learned in a period of time, and not per experiment.

given fraction of the total $M_{th} \leq M(\infty)$ and compensate for the decreased experiment SNR by averaging more measurements in a given period of time. This is known as the faster inversion recovery experiment. To gain an intuition for the benefit of this approach the effective SNR of the two thermalization strategies is compared.

If both the experiment noise and total experiment time is fixed the SNR is proportional to the product of the thermal magnetization and the root number of experiments performed. The ratio of the two approaches to thermalization demonstrates the improvement in the SNR one can achieve by selecting the optimal thermalization time

$$\begin{aligned} \alpha &= \frac{\sqrt{\frac{\tau_{exp}}{\tau_{tot}(\tau_{opt})}} M_{th}(\tau_{opt})}{\sqrt{\frac{\tau_{exp}}{\tau_{tot}(\tau_{tr})}} M_{th}(\tau_{tr})} \\ &= \sqrt{\frac{\tau_i + \tau + \tau_{opt}}{\tau_i + \tau + \tau_{th}} \frac{1 - e^{-\tau_{opt}}}{1 - e^{-\tau_{th}}}}, \end{aligned} \tag{8.8}$$

for a fixed measurement τ. Maximizing equation (8.8) for the thermalization time $\hat{\tau_{opt}}$ that will result in the maximum SNR gain gives

$$\hat{\tau_{opt}} = \arg\min_{\tau_{opt}} \alpha = -\tau_i - \tau - \frac{1}{2}T_1 - T_1 W_{-1}\left(-\frac{1}{2}e^{-\frac{1}{2}-\frac{\tau}{T_1}}\right). \tag{8.9}$$

Where W is the Lambert W Function [20].

In practice an experimentalist might wait anywhere from 3-5 times maximum expected T_1 to reset to the thermal state. Assuming the τ_i and τ are zero, the improvement in SNR the experimentalist can expect in a given period of time if the optimal thermalization time τ_{opt} is used compared to the previous thermalization time τ_{th} is shown in figure 8.14. For the simulations in question the the improvement is generally a factor of $1.2 - 1.4$ in the effective SNR.

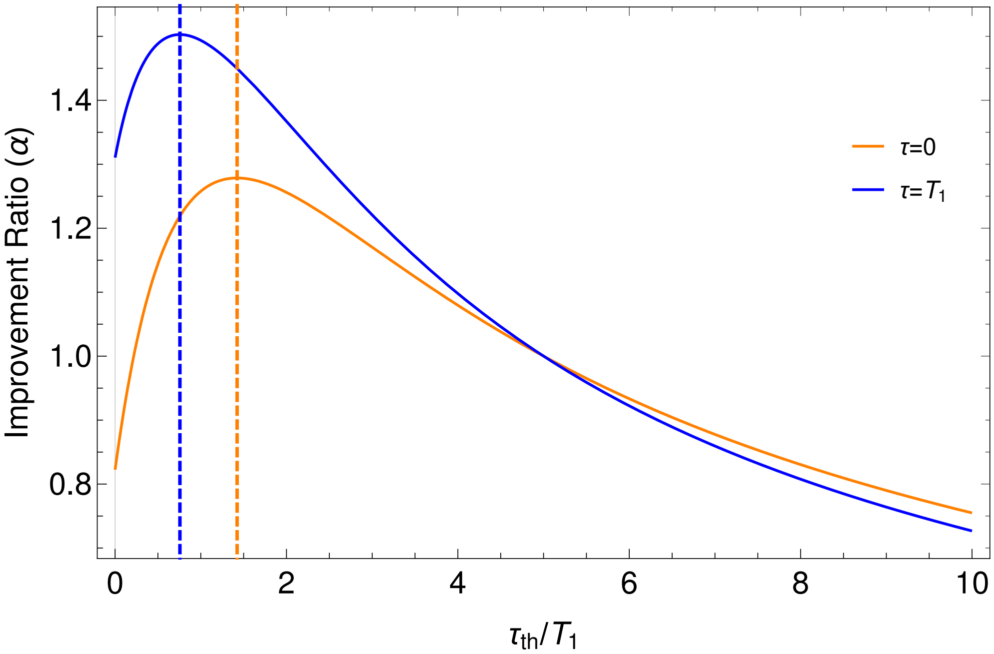


Figure 8.14: Improvement in SNR with variation of thermalization time. The optimal thermalization time $\hat{\tau_{opt}}$ plotted (dashed) to guide the eye. Figures were evaluated with $T_1 = 1.0\ s$, $\tau_{tr} = 5 \times T_1$ and $\tau_i = 0\ s$.

Reducing the thermalization time is a very useful technique and can be further improved by incorporating the thermalization time as an experimental setting in the experiment model function, consequently gaining information about the T_1 value in the thermalization step of the experiment. Let $a = 2 - e^{-\tau_{th}}$ in equation (6.1) to obtain the FIR model function

$$M_{fir}(M(\infty), T_1; \tau, \tau_{th}) = M(\infty)\left(1 - e^{-\frac{\tau_{th}}{T_1}}\right)\left(1 - 2e^{-\frac{\tau}{T_1}}\right). \tag{8.10}$$

This allows the online experimental design algorithm to effectively choose the SNR of each observation and to incorporate estimation of the T_1 parameter into the thermalization reset step, in which traditionally no information is gained. Of course the cost (time) of an experiment is not of interest, the answer is to always let $\tau_{th} \gg T_1$, ie. to perform the maximum SNR experiment. However, when time is of concern significant improvements may be obtained using the FIR experiment, with variable thermalization time, over the IR experiment and its fixed thermalization time.

In figure 8.15 the performance of the cost weighted rapid reset T_1 learning under a tight magnetization prior is compared with the Bayes risk with the same τ_{th} and τ_i settings as in the previous section. It is observed that a large improvement in the variance with respect to total experiment time for the FIR Bayes risk design strategy over the IR experiment is even obtained for the baseline random strategy. Note that although not plotted, a large improvement was shown for the information gain. However, the Bayes risk was superior in all cases. Once again it is found that the high SNR Bayes risk has superior performance to the MIS Bayes risk, which does not align with expectations.

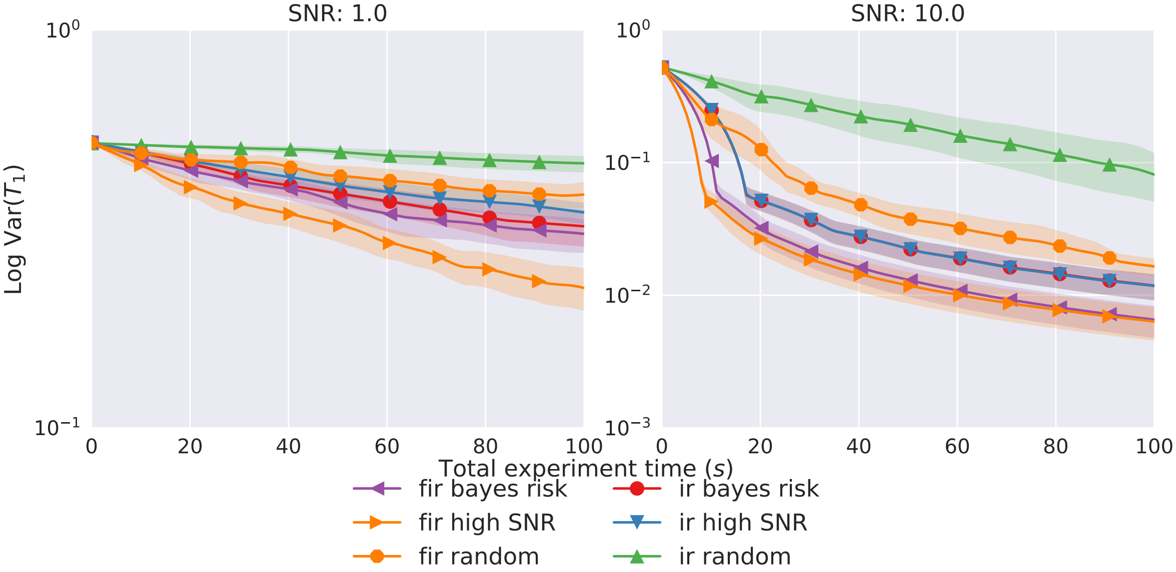


Figure 8.15: Variance plots for learning T_1 under a tight $M(\infty)$ prior comparing FIR and IR experiments against total experiment time, with selected time weighted online design heuristics.

Examining the performance of the FIR and IR experiments when optimizing per experiment and not experiment time in figure 8.16 it is seen that the rapid reset and traditional experiments converge in performance as the rapid reset optimal design will choose to let the thermalization time τ_{th} be as large as possible to maximize the SNR of the single measurement. This is expected as the model function equations (6.1) and (8.10) become identical in the limit $\tau_{th} \to \infty$.

As expected, looking at the per experiment performance of the time weighted online design strategies in figure 8.17, the FIR experiments are much less effective per experiment as they choose to sacrifice SNR in order to improve time efficiency.

Now the behavior when estimating both $M(\infty)$ and T_1 under identical priors is studied, as in the previous section. In figure 8.20 var($M(\infty)$) and var(T_1) as a function of total experiment time are shown. Interestingly it is only for the low SNR plot that the FIR

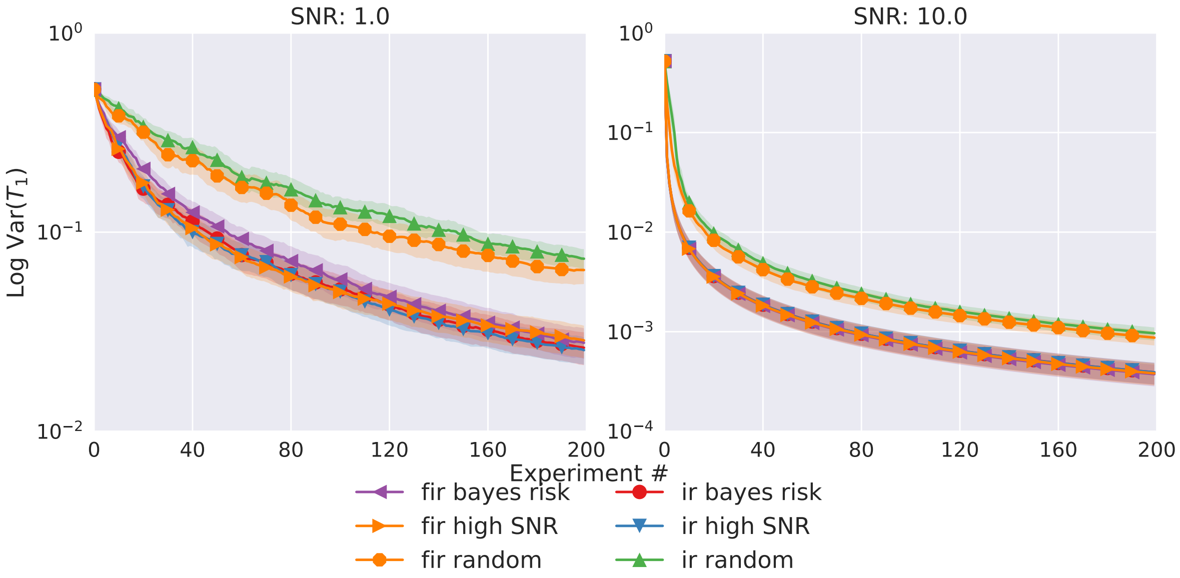


Figure 8.16: Variance plots for learning T_1 under a tight $M(\infty)$ prior comparing FIR and IR experiments against total experiment time, with selected online design strategies without cost-weighted experiment design.

experiment significantly outperforms the traditional IR experiment when only the variance of T_1 is considered. However, the FIR experiment's online strategies perform significantly better when considering the variance of $M(\infty)$. In figure 8.18 the combined weighted variance for $M(\infty)$ and T_1 is plotted. The performance of the FIR experiment while still superior is not as drastic compared with the IR experiment when learning the magnetization as well as the relaxation time in the high-SNR regime.

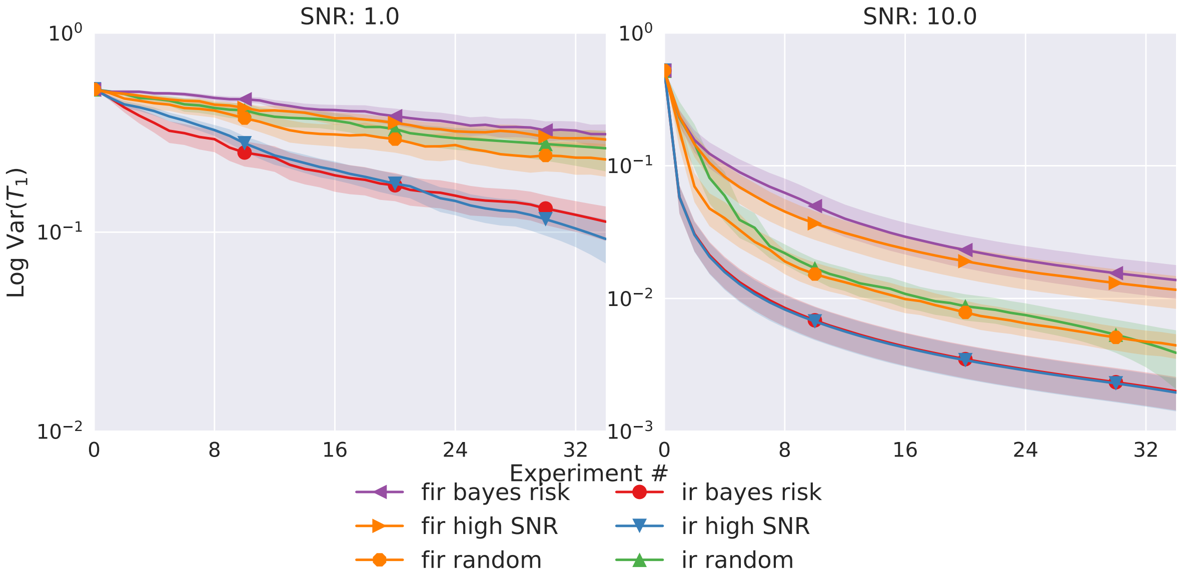


Figure 8.17: Variance plots for learning T_1 under a tight $M(\infty)$ prior comparing FIR and IR experiments per experiment, with selected time weighted online design strategies. The FIR experiments perform worse with respect to IR experiments when evaluated on a per experiment basis as they have had their recovery times selected to optimized for total experiment time.

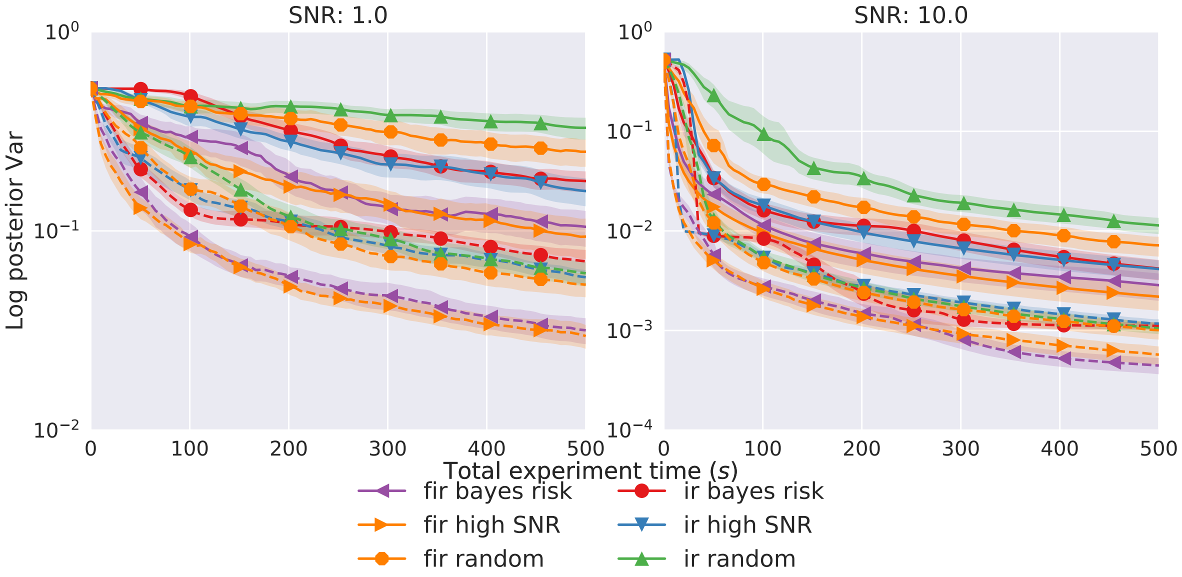


Figure 8.18: Variance plots for learning both $M(\infty)$ and T_1, comparing FIR and IR experiments against total time, with selected time weighted online design strategies.

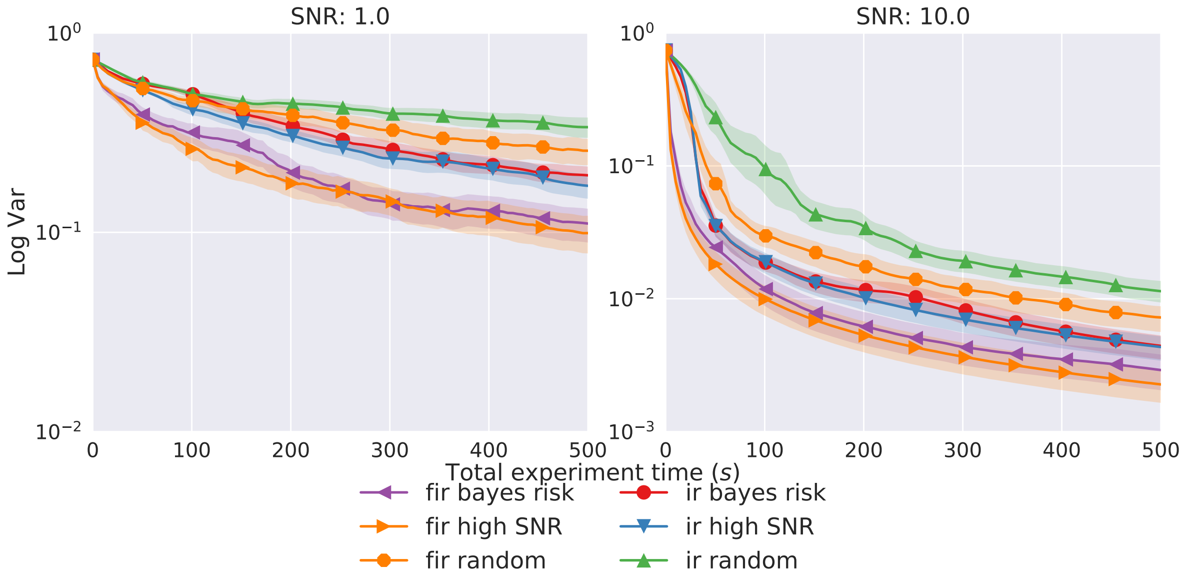


Figure 8.19: Combined weighted variance for $M(\infty)$ and T_1, comparing FIR and IR experiments against total time, with selected time weighted online design strategies.

The FIR experiment also approaches the BCRB in figure 8.19 for the combined weighted variance of $M(\infty)$ and T_1, verifying that the estimation strategy is still near optimal for the FIR experiment.

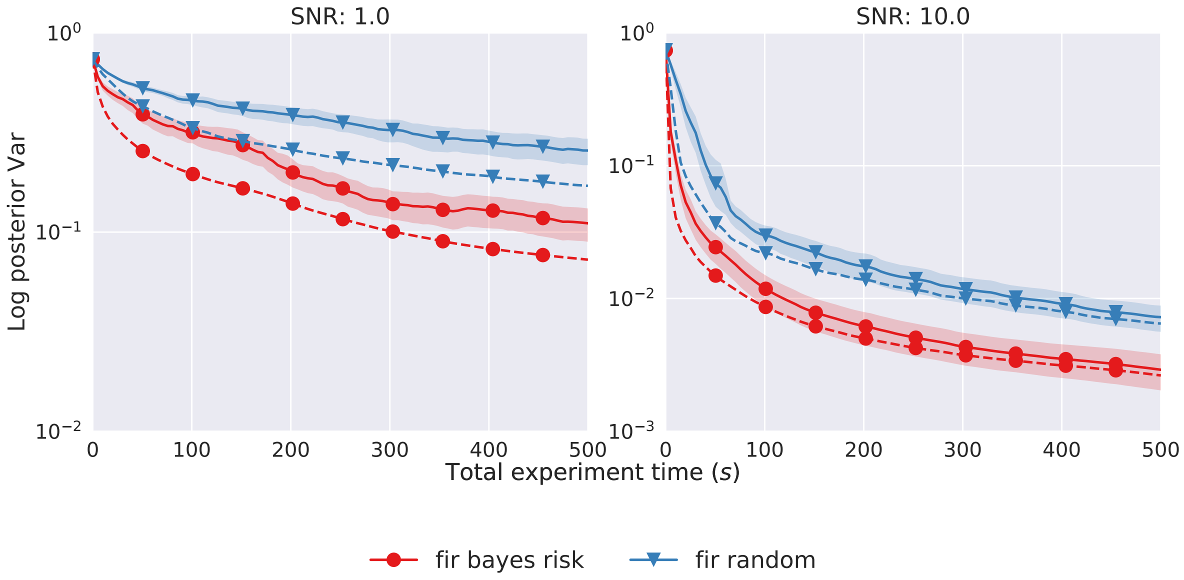


Figure 8.20: Combined weighted variance for $M(\infty)$ and T_1 for the FIR experiment with the BCRB plotted. Verifying that the weighted variance approaches the lower bound set by the BCRB for the FIR experiment.

8.2.2 FI Experiment Selection Heuristic

Previously it was noted that the mean maximal values of the FIM are the most likely experiment to be selected in most cases. The performance of the particle filtering techniques have been shown to approach the BCRB, and the high SNR Bayes risk is a functional of the FIM. It seems reasonable that the maximal diagonal values of the FIM will be the optimal measurement locations to reduce the variance of the respective model parameter. As the FIM for additive normal noise is a function of the model function's first derivatives, these optimal measurement locations may be solved for the inversion recovery experiments analytically.

Consequently the following heuristic online experiment design protocol is proposed. For a given prior $\pi(M(\infty), T_1)$ with means $M(\hat{\infty})$, $\hat{T_1}$ and covariance $Cov(M(\infty), T_1|\pi)$ define the maximal FIM τ values as τ_{T_1} and $\tau_{M(\infty)}$. The maximal FI values may be used to compute the cost weighted high-SNR risk improvement for τ_{T_1} and $\tau_{M(\infty)}$ yielding risk improvements $\mathrm{ri}(\tau_{M(\infty)})$ and $\mathrm{ri}(\tau_{T_1})$. The selected heuristic experiment τ is

$$\underset{\tau \in \left\{\tau_{M(\infty)}, \tau_{T_1}\right\}}{\arg\max} \quad \mathrm{ri}(\tau). \tag{8.11}$$

This online experimental design heuristic will alternate between the optimal measurements for learning the magnetization, and relaxation time proportionally to the weighted variance of the respective parameter. Note that it will still typically perform experiments at many different values of τ as the optimal τ will vary as data is collected and the prior is updated. This heuristic requires no numerical optimization over experimental parameters and greatly reduces the computational resource requirements for online experiment design, while producing comparable performance enhancements over baseline design strategies.

It is only when both the magnetization and relaxation time have same order of magnitude uncertain priors (or a weighting matrix $\mathbf{Q}$ that induces a near equal contribution to the Bayes risk for both parameters) that a strong deviation of the Bayes risk selected values compared to the heuristic values will be observed, as seen in the case of wide prior for both $M(\infty)$ and T_1 in figure 8.21 compared to figure 8.22 where the FI heuristic is used.

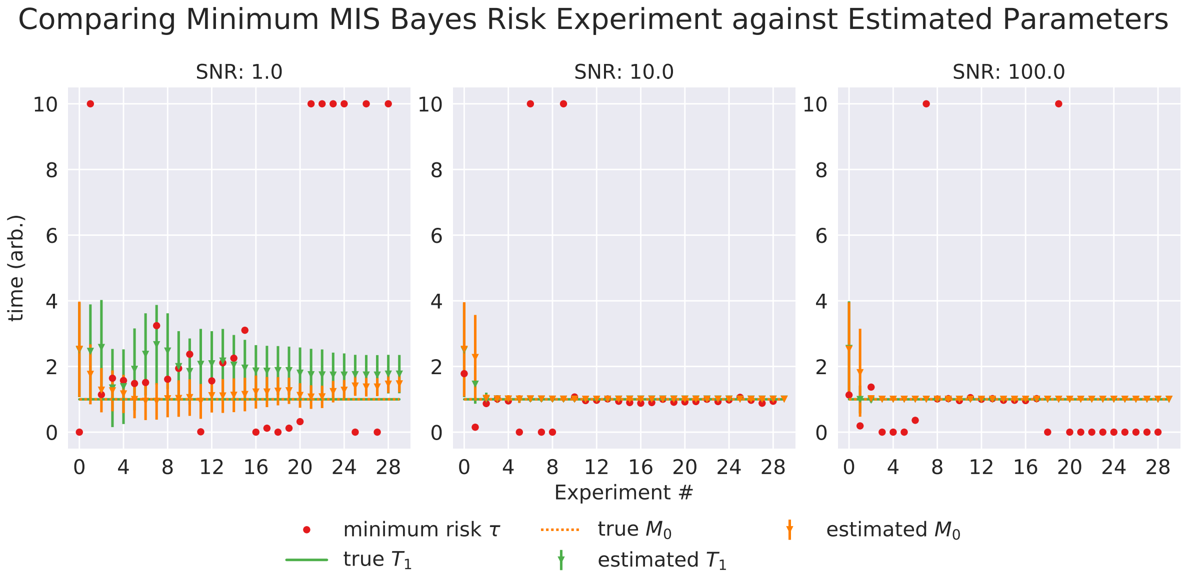


Figure 8.21: Comparison of selected experiment τ values chosen at the minimum MIS Bayes risk compared to the means $\hat{M(\infty)}$ and $\hat{T}_1$ at each step for a simulation with true model parameters $M(\infty) = 1$, $T_1 = 1$, and a wide prior over both the magnetization and relaxation time. The selected experiment values are not the optimal values of the FI heuristic discussed above, because the online design attempts to minimize the weighted posterior variance of both model parameters, and therefore samples experiments that are less easy to interpret.

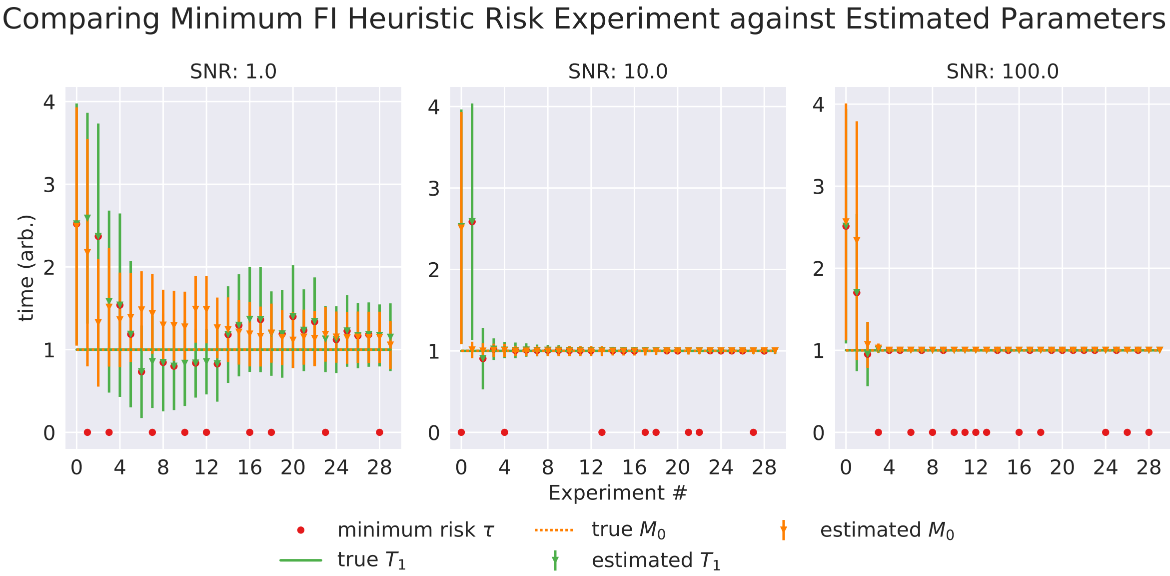


Figure 8.22: Comparison of selected experiment τ values chosen at each step using the FI heuristic for a simulation with true model parameters $M(\infty) = 1$, $T_1 = 1$, and a wide prior over both the magnetization and relaxation time. The heuristic alternates between sampling to improve the magnetization and relaxation time estimates, and samples at the optimal heuristic times. Its performance is expected to be worse on average than the Bayes risk, but to require fewer computational resources.

The simplest heuristic case is the IR experiment with a constant experiment cost. The Fisher information of the IR model function has maximal values $\tau_{M0} = 0$ and $\tau_{T_1} = T_1$. Results of heuristic simulation performance in figure 8.23 demonstrate that the heuristic is near optimal when $M(\infty)$ is initially known. In figure 8.24 simulations where $M(\infty)$ has an identical unknown prior to the relaxation time are plotted. Observed performance is slightly worse than the Bayes risk techniques, as the heuristic focuses on reducing the variance of a single model parameter at a time, whereas the Bayes risk optimization is capable of selecting experiments that reduce the total weighted variance.

The heuristic becomes slightly more complicated when experiment time(cost) is accounted for. The maximum FIM T_1 element is

$$\tau_{T_1} = \frac{1}{4}\left(\hat{T}_1 - 2(\tau_i + \tau_{th}) + \sqrt{\hat{T}_1^{\,2} + 12\hat{T}_1\tau_i + 4(\tau_i + \tau_{th})^2}\right). \tag{8.12}$$

From equation (8.12) observe that the maximal CRB value when accounting for cost varies from $\hat{T}_1$ in the limit of a very long experiment interval/thermalization time, and $\hat{T}_1/2$ in the limit of no interval/thermalization time. The CRB optimal value for learning the magnetization is still $\tau_{M(\infty)} = 0$. The heuristic is extended to the FIR experiment by defining

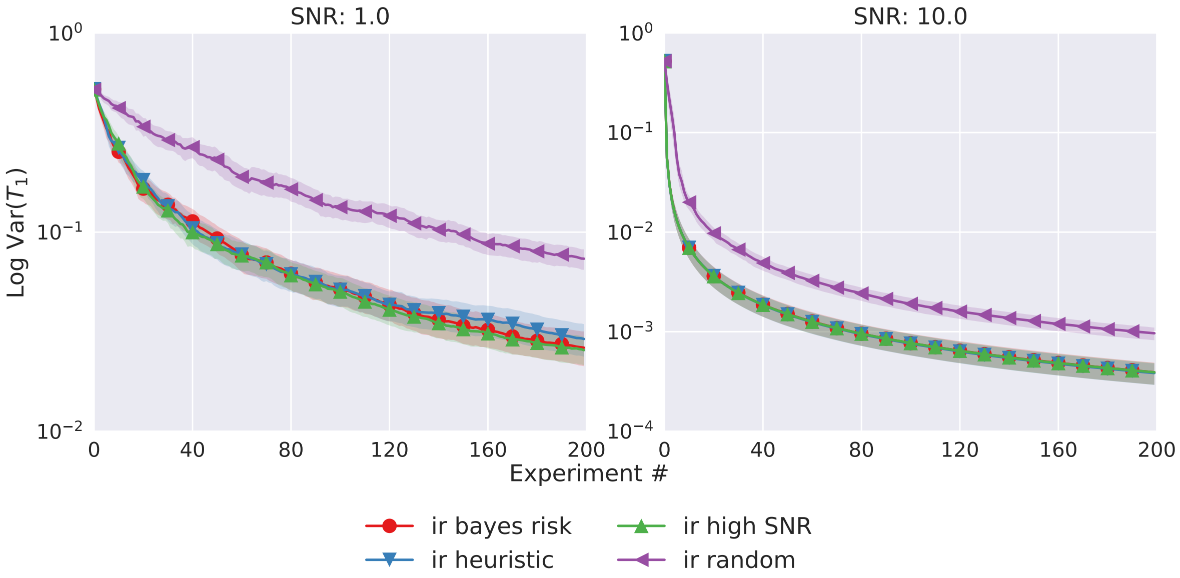


Figure 8.23: Covariance plot of T_1 estimate for heuristic online experimental design when $M(\infty)$ is assumed known and cost is constant. As expected the heuristic performance is equivalent to the Bayes risk.

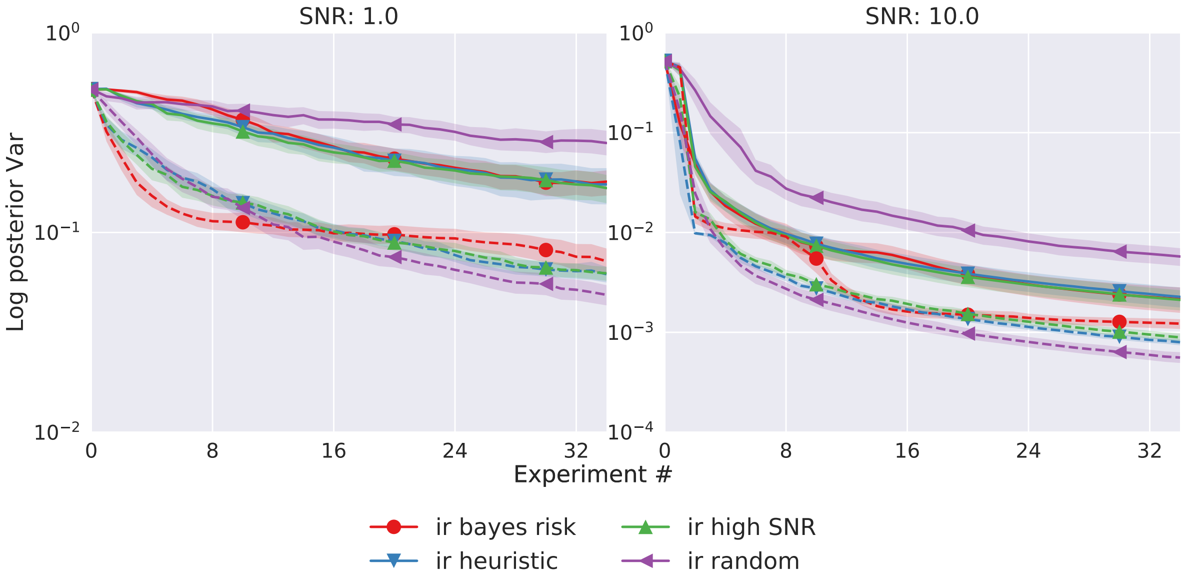


Figure 8.24: Covariance plot of $M(\infty)$(dashed) and T_1(solid) estimates for heuristic online experimental design under uninformative identical priors for $M(\infty)$ and T_1 when experiment cost is constant.

the optimal reset time to be the thermalization time that maximizes the SNR found in equation (8.9).

The performance of the experiment time weighted heuristic with unknown magnetization and relaxation time for the IR and FIR experiments are simulated in figure 8.25. Observed performance is equivalent to the high-SNR Bayes risk for both magnetization and relaxation time estimates.

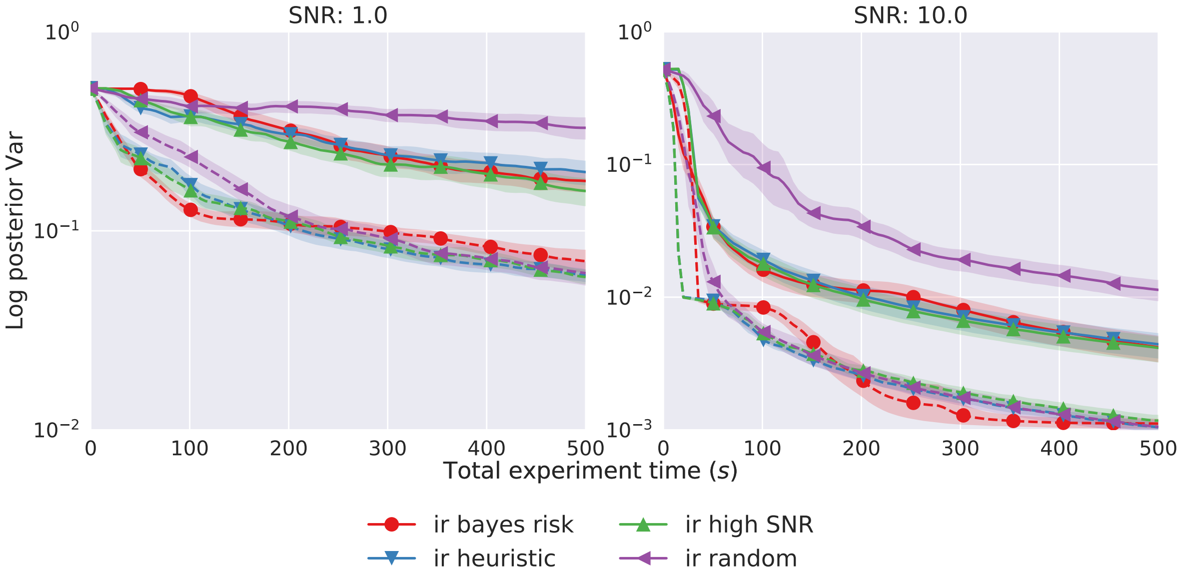


Figure 8.25: Covariance plot of $M(\infty)$(dashed) and T_1(solid) estimates for heuristic online experimental design under uninformative identical priors for $M(\infty)$ and T_1 when optimized with respect to total experiment time. Performance is equivalent to high SNR Bayes risk.

8.3 Learning the SNR Simultaneously

Thus far the value of the variance σ^2 has been assumed known. For a given experimental system, conditions and sample the noise is an *i.i.d.* fixed value of the instrumentation and is characterized before an experiment. Alternatively, the noise σ^2 may be treated as an unknown model parameter in the particle filter and assigned a prior. Now σ^2 will be estimated simultaneously with $M(\infty)$ and T_1. As noted by Chaloner and Verdinelli, D-optimal(information gain) designs are highly sensitive to the variance whereas A-optimal(Bayes risk) have an innate robustness for linear model functions [18]. Unfortunately, this does not hold true for nonlinear models in which case both the Bayes risk and information gain will depend on the observed data.

Simulations are performed with fixed magnetization of $M(\infty) = 1.0$ and varying σ to obtain SNRs of 1 and 10. The same priors for T_1 and $M(\infty)$ as in the previous section are used, while assigning a uniform prior for $\pi(\sigma) = U(10^{-6}, 10)$. For brevity, only simulations where both $M(\infty)$ and T_1 are estimated with time-weighted experiment design heuristics are performed. Results are shown in figure 8.26. The MIS Bayes risk significantly outperforms the high-SNR risk in the low-SNR regime. This enhanced performance is attributed to the increased nonlinearity of the posterior distribution as a consequence of estimating σ^2.

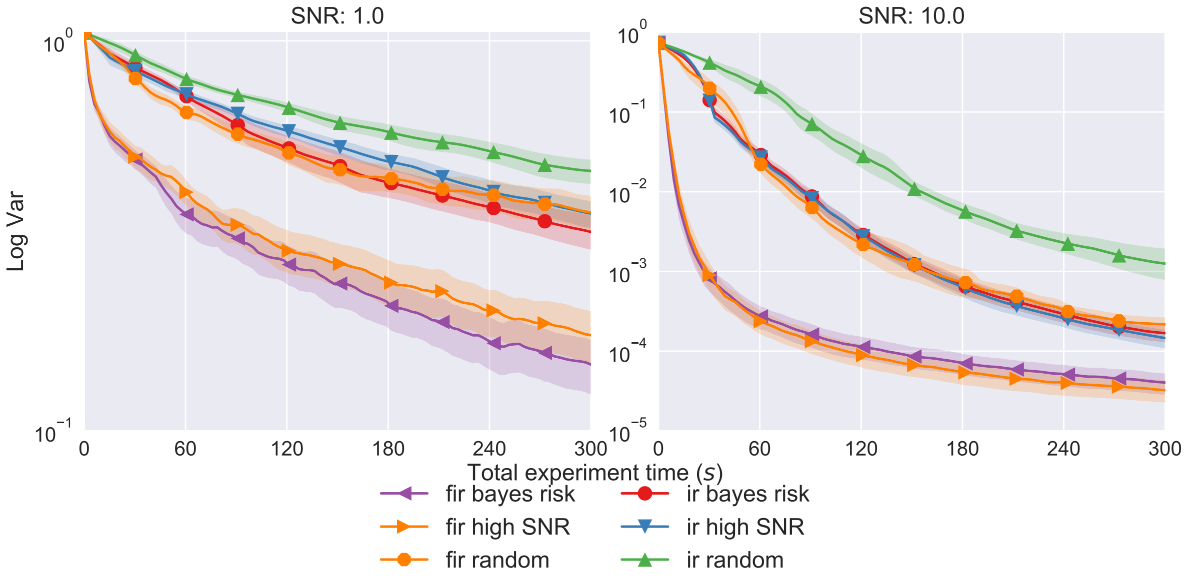


Figure 8.26: Simulations comparing total covariance with identity matrix Q with prior over σ^2.

8.4 Q Matrix Selection

For all design heuristics in the previous sections an identity weighting matrix Q. In practice when estimating the T_1 the magnetization $M(\infty)$ is not of interest and it is not desired to design experiments that account for this parameter. In this case choose $Q' = \text{diag}([0, 1])$. The performance of estimating $M(\infty)$ and T_1 with the two different weighting matrices is compared in figure 8.27. The performance of T_1 estimation with Q' is slightly enhanced, while for $M(\infty)$ it performs significantly worse. From an implementation perspective it has been found that it is often better to include some weighting for all parameters to aid the particle filter in convergence to realistic particle locations.

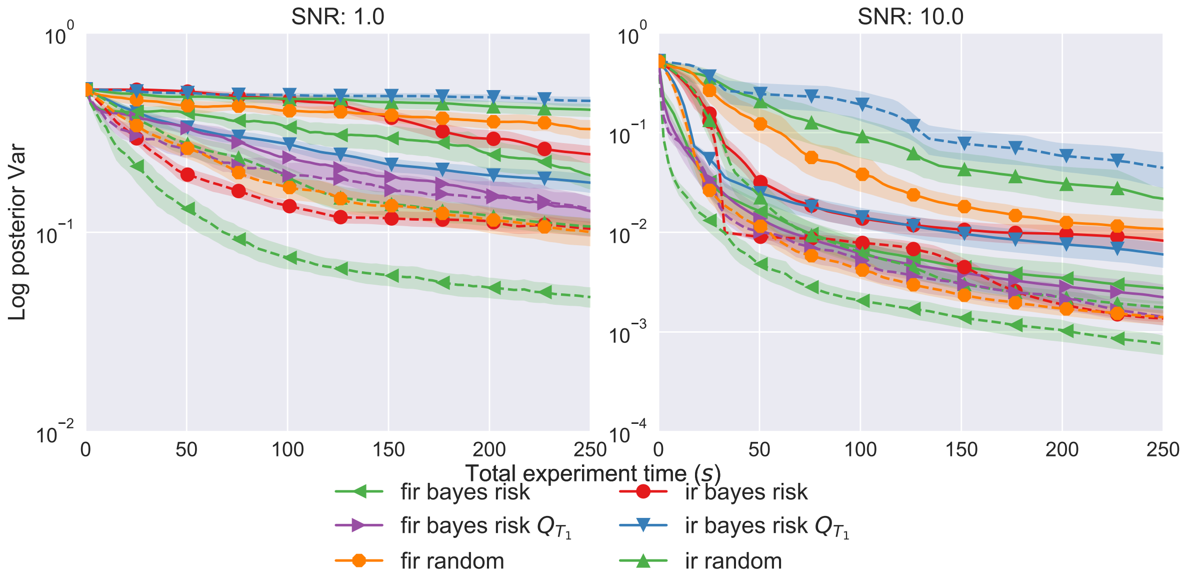


Figure 8.27: Covariance plot of $M(\infty)$(dashed) and T_1(solid) estimates for heuristic online experimental design under uninformative identical priors for $M(\infty)$ and T_1 when optimized with respect to total experiment time for different weighting matrices.

8.5 Optimal Batched Experiment Design

Optimal designs of relaxation estimates have been considered in detail by Weiss *et. al.* within a frequentist least-squares estimate framework for batched fixed interval experiment design [89]. They consider polynomial designs of the form

$$\tau_i = \tau_1 + (\tau_n - \tau_1)\left(\frac{i-1}{n-1}\right)^r \qquad i = 1, 2, 3, \cdots, n \tag{8.13}$$

Where n is the number of predetermined experiments and the experiments τ_i are determined by the initial experiment τ_1, final experiment τ_n, and power law spacing $r > 0$. They design experiments to minimizes the estimated least squares standard deviation $\sigma_{\hat{T}_1}$

$$\frac{\sigma_{\hat{T}_1}}{T_1} = \frac{\sigma}{M(\infty)} G\left(\frac{\tau_1}{T_1}, \cdots, \frac{\tau_1}{T_1}\right), \tag{8.14}$$

Where

$$G = \sqrt{\frac{\beta_1}{\beta_1\beta_3 - \beta_2^2}}. \tag{8.15}$$

With dimensionless β's and experiment times $t_i = \tau_i/\hat{T}_1$ and $w = \tau_{th}/\hat{T}_1$.

$$\beta_1 = \sum_i \left(1 - \left(2 - e^{-w}\right) e^{-t_I}\right)^2, \tag{8.16a}$$

$$\beta_2 = \sum_i \left(1 - \left(2 - e^{-w}\right) e^{-t_i}\right) \left(-2t_i + (w + t_i)e^{-w}\right) e^{-t_i}, \tag{8.16b}$$

$$\beta_3 = \sum_i \left(-2t_i + (w + t_i)e^{-w}\right)^2 e^{-2t_i}. \tag{8.16c}$$

The error estimates are based on a linearized error propagation of the expected variance of a least squares fit. It is important to consider how their methods differ from the methods presented here. First note that the predicted standard deviation of the relaxation parameter $\frac{\sigma_{\hat{T}_1}}{T_1}$ found in equation (8.14) has a simple linear dependence on the SNR. The methods of Weiss are based on high SNR approximations and do not hold in the low SNR regime where the MIS Bayes risk should hold.

Another significant difference is found in the quantity G which depends on a point estimate of T_1. Their method is not naturally able to account for a distribution of prior knowledge over the T_1 parameter. Weiss et. al attempt to correct for this by taking the most conservative set of experiments

$$\max_{T_1 \in \left[T_{1_A}, T_{2_B}\right]} \frac{\sigma_{\hat{T}_1}}{T_1}(\tau_1, \tau_n, T_1), \tag{8.17}$$

over some interval determined by their prior knowledge $[T_{1_A}, T_{1_B}]$. This method of incorporating prior knowledge is unable to incorporate nonuniform priors, and the heuristic of selecting the most conservative experiment settings will on average waste experimental resources. The argument for the most conservative experiment is that it provides a bound on the maximum expected error given an interval knowledge for T_1. One can easily see how this interpretation becomes problematic if a gamma distributed prior over T_1 is considered. Such a prior's tails will extend unbounded, and consequently, the most conservative experiment heuristic would select times that are also unbounded even though the probability of such a value occurring is asymptotically small. A more natural formalism might choose the most conservative experiment over some credible region. If worrying about significant deviation of outliers from the Bayes risk, fewer consecutive experiments should be performed with more frequent experiment using the current state of knowledge. Due to the nonlinearity of the IR and FIR model functions sequential design is expected to result in improved performance [18]. This is not possible within the framework of Weiss, which is unable to accurately incorporate previous experiments into the experimental design process. Taitelbaum *et al.* attempt to incorporate this by performing a two stage design. Where the first stage of the design is used to determine an estimate of T_1 and the second stage is designed with this estimate under a credible region assumption. The reliance on batched experimental designs precludes the design of arbitrary unstructured experiments and forces the use of structured experiments such as equation (8.13) due to the computational complexity of optimizing such an experiment. The framework outlined in chapter 7 incorporates all of these features naturally and allows the design of the optimal experiment given the current state of knowledge within a consistent framework.

Comparing the performance of the Bayesian experimental designs over those listed by Taitelbaum [78] for a FIR experiment with $\tau_{th} = 2T_{1_B}$ and varying priors defined for the interval $T_1 \in U\left(T_{1_A}, T_{1_B}\right)$ which are parameterized by the unitless $\alpha = T_{1_B}/T_{1_A}$, and designed with the equation (8.13) using $r = 1$. Bayesian experiment design is performed as described in section 7.3 for SNRs of 1 and 10, and $M(\infty) \in U(0.5, 3.0)$ with $Q_{M(\infty)} = 0$ and $Q_{T_1} = 1$ using the Bayes risk, high SNR Bayes risk and the greatest risk. The greatest risk is an adhoc SMC form of equation (8.17) which recalling equation (7.19) selects the batch of experiments that minimizes the risk over the maximum risk model parameter's particle

$$GR = \max_i \sum_{j=1}^{n_{sam}} \sum_{k=1}^{n_m} Q_{kk} \sum (\mathbf{x}_{ik} - \hat{\mathbf{x}}_{jk})^2. \tag{8.18}$$

The designed parameters are found in table 8.1, it is clear that there are significant differences between the Bayesian designs and Taitelbaum's and that there is a strong dependence on the SNR as depicted in figure 8.28. In the low SNR regime the Bayes risk prefers to perform a very narrow sweep, whereas in the high SNR regime it performs an experiment sweep more similar to that predicted by Taitelbaum et. al. This is believed to be due to increasing sensitivity to optimal measurements at high SNR as shown in figure 8.1. The effect of a true T_1 far from the mean T_1 would be significant in the high SNR regime and would induce poor experimental performance. The optimal experiment is therefore to perform a larger sweep to account for these outliers relaxation times, whereas in the low SNR regime the outliers have a much less drastic effect on performance and it is therefore best to optimize the experimental design for the expected T_1 value.

For reference, designs are also provided for the IR experiment under the same priors as in table 8.2. Figure 8.29 depicts the calculated Bayes risk for the designed optimal experiments. It is clear that the Bayes risk is expected to outperform the design of Taitelbaum with the high SNR and MIS risk converging at high SNR.

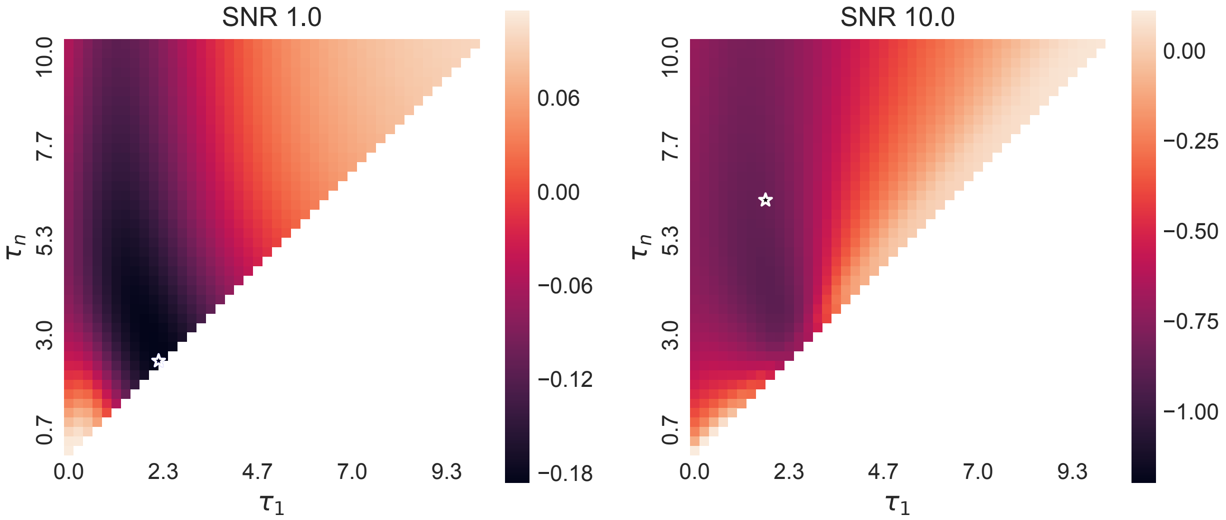


Figure 8.28: Heatmap of $n = 8$ experiment Bayes risk and $\alpha = 5.0$. The white and black star marks the location of the minimum risk. For low SNR it is clear that algorithm prefer to have little or no sweep range.

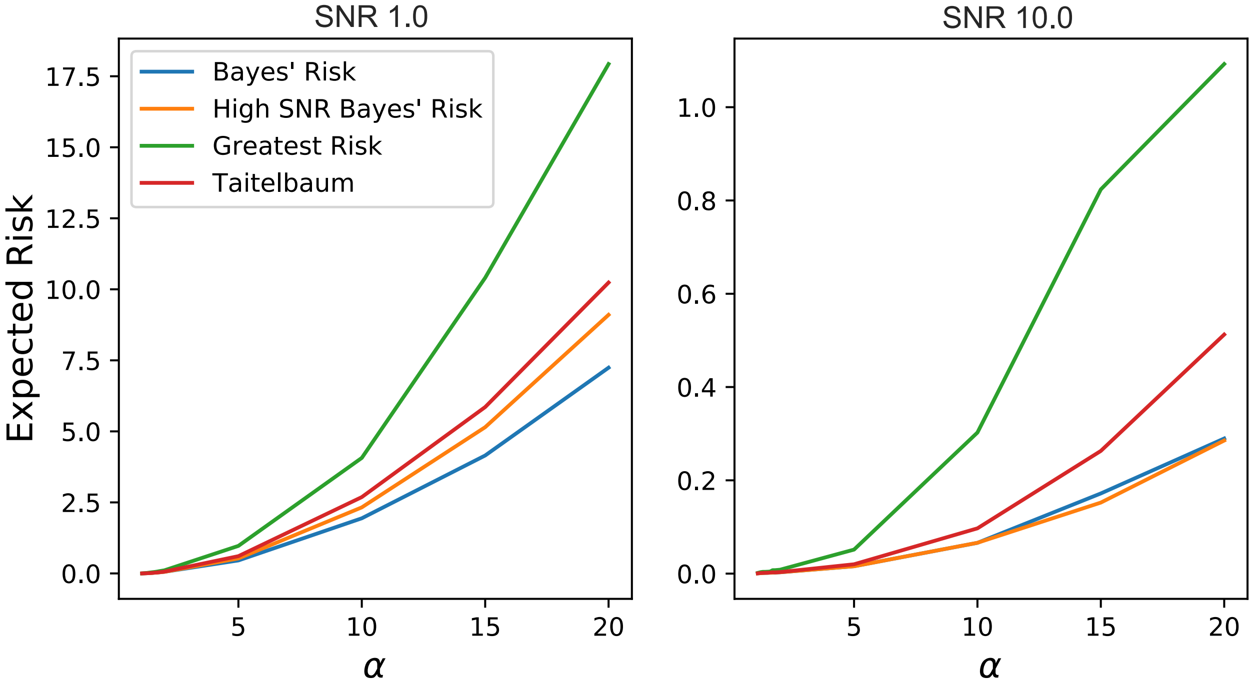


Figure 8.29: Expected posterior variance of optimal designs found in table 8.1 as a function of α.

		Taitelbaum			Bayes Risk			High SNR Bayes Risk			Greatest Risk		
SNR	α	T_{1_A}	T_{1_B}	$\langle\sigma^2_{T_1}\rangle$	T_{1_A}	T_{1_B}	$\langle\sigma^2_{T_1}\rangle$	T_{1_A}	T_{1_B}	$\langle\sigma^2_{T_1}\rangle$	T_{1_A}	T_{1_B}	$\langle\sigma^2_{T_1}\rangle$
1.0	1.10	0.50	0.87	0.0009	0.76	0.76	0.0009	0.76	0.79	0.0009	0.70	0.72	0.0024
	1.20	0.48	1.01	0.0033	1.09	1.09	0.0033	1.09	1.11	0.0033	0.61	1.16	0.0085
	1.30	0.46	1.16	0.0074	1.18	1.18	0.0071	1.13	1.15	0.0070	0.76	1.28	0.0174
	1.40	0.46	1.27	0.0118	1.16	1.16	0.0112	1.13	1.16	0.0111	0.52	1.78	0.0291
	1.50	0.45	1.40	0.0183	1.15	1.15	0.0173	1.12	1.17	0.0170	1.01	1.07	0.0401
	1.60	0.45	1.51	0.0266	1.17	1.17	0.0248	1.14	1.17	0.0244	1.00	1.06	0.0499
	1.70	0.44	1.66	0.0333	1.22	1.22	0.0307	1.19	1.25	0.0306	0.87	1.39	0.0631
	1.80	0.45	1.73	0.0423	1.20	1.20	0.0388	1.20	1.23	0.0387	0.93	1.82	0.0797
	2.00	0.43	2.01	0.0640	1.31	1.31	0.0568	1.31	1.34	0.0558	1.18	1.21	0.1087
	5.00	0.38	4.41	0.6076	2.40	2.40	0.4572	1.81	4.47	0.5286	1.56	4.28	0.9689
	10.00	0.55	7.22	2.6868	4.31	4.43	1.9402	2.91	8.04	2.3248	3.03	8.04	4.0662
	15.00	0.68	9.72	5.8562	6.13	6.30	4.1512	3.75	11.75	5.1474	5.11	11.75	10.4060
	20.00	0.77	12.09	10.2384	7.97	7.97	7.2412	4.65	15.28	9.1022	5.76	15.28	17.9300
10.0	1.10	0.50	0.87	0.0006	0.72	0.76	0.0005	0.63	0.94	0.0004	0.67	0.83	0.0015
	1.20	0.48	1.01	0.0012	0.71	0.87	0.0010	0.59	1.13	0.0009	0.68	0.90	0.0030
	1.30	0.46	1.16	0.0016	0.69	1.01	0.0013	0.59	1.26	0.0013	0.66	1.11	0.0038
	1.40	0.46	1.27	0.0018	0.64	1.26	0.0015	0.59	1.42	0.0015	0.64	1.39	0.0038
	1.50	0.45	1.40	0.0019	0.64	1.42	0.0016	0.59	1.55	0.0017	0.72	1.47	0.0041
	1.60	0.45	1.51	0.0023	0.64	1.57	0.0020	0.59	1.71	0.0020	0.67	1.68	0.0044
	1.70	0.44	1.66	0.0028	0.70	1.66	0.0024	0.61	1.83	0.0022	0.67	2.01	0.0073
	1.80	0.45	1.73	0.0026	0.70	1.79	0.0022	0.64	1.97	0.0024	0.82	2.09	0.0074
	2.00	0.43	2.01	0.0032	0.71	2.17	0.0028	0.65	2.24	0.0029	0.91	1.91	0.0083
	5.00	0.38	4.41	0.0200	1.66	4.41	0.0158	1.47	4.41	0.0158	2.05	4.41	0.0515
	10.00	0.55	7.22	0.0970	3.49	8.31	0.0660	2.89	8.31	0.0661	3.61	8.31	0.3024
	15.00	0.68	9.72	0.2630	5.20	11.95	0.1719	3.98	11.95	0.1523	6.58	11.95	0.8237
	20.00	0.77	12.09	0.5125	7.07	15.74	0.2896	4.79	15.74	0.2853	7.76	15.74	1.0925

Table 8.1: Computed optimal parameters for FIR experimental initial (τ_1) and final (τ_N) relaxation delays for $N = 8$, $\tau_{th} = 2T_{1_B}$ for varying α. $\langle\sigma^2_{T_1}\rangle$ is the expected posterior covariance of T_1 for the given experiment parameters. The optimal parameters designed by Taitelbaum *et. al.* [77] are independent of SNR.

		Bayes Risk			High SNR Bayes Risk			Greatest Risk		
SNR	α	T_{1_A}	T_{1_B}	$\langle \sigma^2_{T_1} \rangle$	T_{1_A}	T_{1_B}	$\langle \sigma^2_{T_1} \rangle$	T_{1_A}	T_{1_B}	$\langle \sigma^2_{T_1} \rangle$
1.0	1.10	0.81	0.83	0.0008	0.79	0.81	0.0008	0.63	0.99	0.0025
	1.20	0.97	0.97	0.0032	0.94	0.97	0.0032	0.80	0.85	0.0084
	1.30	1.03	1.03	0.0069	0.98	1.01	0.0068	0.59	1.23	0.0171
	1.40	0.92	0.95	0.0114	0.90	0.92	0.0112	0.64	1.33	0.0285
	1.50	1.02	1.02	0.0176	0.99	1.02	0.0172	0.72	1.40	0.0350
	1.60	0.98	1.00	0.0237	0.98	1.00	0.0233	0.81	1.51	0.0516
	1.70	1.10	1.10	0.0298	1.07	1.13	0.0296	0.99	0.99	0.0629
	1.80	1.07	1.07	0.0379	1.04	1.07	0.0372	1.07	1.31	0.0761
	2.00	1.16	1.19	0.0523	1.16	1.19	0.0519	1.06	1.12	0.1042
	5.00	2.24	2.24	0.4547	1.73	3.72	0.5063	1.60	4.36	0.9193
	10.00	4.07	4.18	1.7826	2.79	8.02	2.0613	3.02	8.02	4.2336
	15.00	6.15	6.15	4.0134	3.87	12.13	4.8453	4.57	12.13	9.6652
	20.00	8.07	8.07	7.2786	4.61	15.91	8.8911	5.07	15.91	17.3395
10.0	1.10	0.72	0.74	0.0005	0.61	0.97	0.0004	0.67	0.85	0.0013
	1.20	0.75	0.78	0.0008	0.59	1.13	0.0008	0.66	1.01	0.0024
	1.30	0.69	1.02	0.0011	0.57	1.31	0.0011	0.67	1.24	0.0031
	1.40	0.64	1.26	0.0013	0.57	1.44	0.0012	0.59	1.44	0.0032
	1.50	0.68	1.36	0.0016	0.60	1.57	0.0015	0.65	1.57	0.0039
	1.60	0.61	1.59	0.0017	0.58	1.70	0.0016	0.56	1.86	0.0044
	1.70	0.66	1.64	0.0019	0.60	1.79	0.0018	0.72	1.73	0.0058
	1.80	0.67	1.85	0.0021	0.64	1.97	0.0021	0.67	1.82	0.0049
	2.00	0.74	2.00	0.0026	0.64	2.22	0.0025	0.80	2.12	0.0094
	5.00	1.68	4.46	0.0154	1.42	4.46	0.0142	2.26	4.46	0.0503
	10.00	3.54	8.15	0.0579	2.72	8.15	0.0532	4.72	8.15	0.2824
	15.00	4.67	11.50	0.1399	3.50	11.50	0.1307	6.00	11.50	0.5544
	20.00	6.21	15.30	0.2412	4.44	15.30	0.2182	8.43	15.30	1.9555

Table 8.2: Computed optimal parameters for IR experimental initial (τ_1) and final (τ_N) relaxation delays for $N = 8$ and varying α. $\langle \sigma^2_{T_1} \rangle$ is the expected posterior covariance of T_1 for the given experiment parameters.

The average performance for an experiment batch of $n = 8$ is simulated and the ratio of the final to initial variance for α from [77] is plotted in figure 8.30. For simulations true T_1 values were randomly sampled from the uniform interval defined by α, while a fixed prior of $\pi_{M(\infty)} = U(0.5, 3.0)$ is used for the magnetization. For comparison results are also included for sequential design procedure. It is clear that the Bayes risk outperforms the design of Taitelbaum for both batched and sequential designs. Performance is especially enhanced for low SNR measurements, where the designs of Taitelbaum are expected to perform poorly. The conservative risk performs worse than the Bayes risk as expected. The superior performance of the Bayes risk is attributed to its ability to properly account for the full prior distribution over all parameters, rather then incorporating it in an ad-hoc manner as in the frequentist experiment design formalism.

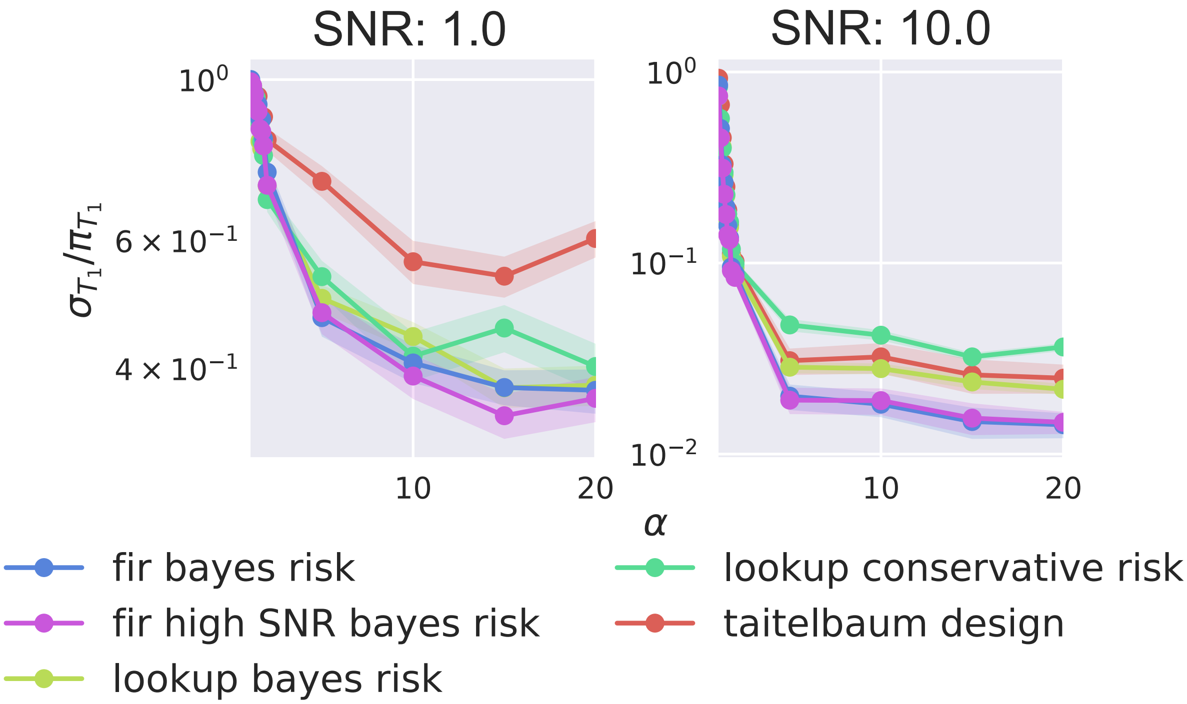


Figure 8.30: Ratio of initial to final σ_{T_1} for varying α and $n = 8$ experiments.

The performance of the batched designs for fixed α and varying true relaxation rate T_1 are considered in order to study the experiment design performance when the true parameter value is near the prior boundaries. Simulation results are shown for $\alpha = 10$ in figure 8.31. Performance is enhanced for Bayesian heuristics compared to the Taitelbaum strategy when the true T_1 is close to the maximum of the prior. The adaptive strategies significantly outperform the Taitelbaum design for all true relaxation values, while the fixed Bayes risk design is less consistent.

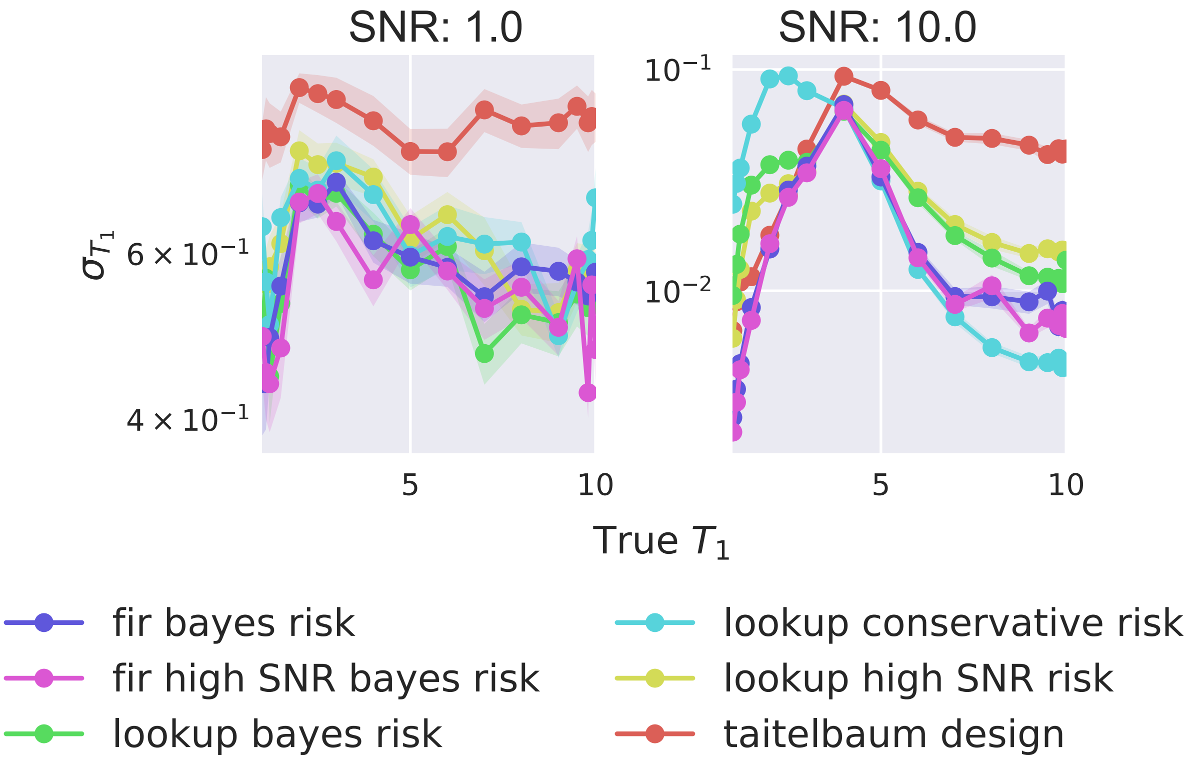


Figure 8.31: Simulation of batched design strategies for $n = 8$ and fixed $\alpha = 10$ as a function of the true simulation relaxation value T_1.

Chapter 9

Future Work and Conclusion

Bayesian parameter estimation provides an intuitive formalism for interpreting the results of scientific data analysis, allowing one to apply prior knowledge to data modelling [30]. While analytically intractable outside of trivial cases, the development of modern numerical methods has allowed the resurgence of Bayesian parameter estimation and placed it in contention with "traditional" frequentist fitting methods. Particle filtering is particularly suited for analyzing experimental results as it is a sequential process that does not require a full resampling step with the acquisition of new data as is the case with RWMCMC methods. In chapter 6 the problem of T_1 estimation that we consider throughout this thesis is introduced. This estimation problem is crucial to industries such as NMR oil logging and medical imaging.

It has been demonstrated how these methods allow questions such as "given the data we currently have, where should an experiment be performed next to maximize the future knowledge?" to be formalized. By asking this question at each experiment step a sequential online experiment design may be performed with the aim of minimizing the time or some other experiment cost. Within the framework of particle filtering, this question is expensive for any continuous outcome likelihood function. In chapter 7 Monte Carlo integration techniques exploiting the inherent distributions available to the particle filter were developed. For the specific case of additive white noise several heuristic methods that achieve equivalent performance to the MIS Bayes risk on the problem of interest were given.

In chapter 8 sequential experiment design algorithms were applied to T_1 inversion. When compared with traditional linear and random experiment strategies, the greedy sequential experiment design strategies achieve near a half order of magnitude improvement in the per experiment estimation of T_1, improving to near a full order of magnitude in the experiment total time domain when the FIR experiment procedure is applied in conjunction with estimation of the magnetization. As the computational cost of such methods is large the high SNR Bayes risk and FI heuristic have been provided as alternative design strategies. These approximate(heuristic) strategies have equivalent performance to the MIS Bayes risk but only apply to the experiment models we have considered here, whereas the MIS Bayes risk applies to any model with a likelihood function, albeit at enhanced computational cost. Finally, it is shown that by including the noise variance as a parameter to be estimated the sequential design strategies have a further improvement over traditional strategies.

The developed Bayesian experimental design strategies are then compared with the earlier works of Weiss and Taitelbaum which apply frequentist design strategies to the task of designing batched optimal inversion recovery experiments [89, 78]. For fixed priors it is demonstrated that the Bayesian methods have superior performance and that sequential strategies outperform batched designs. Furthermore, it is noted that the advantage of sequential estimation strategies is the ability to terminate experiments after a specified certainty has been obtained.

Throughout our analysis the counterintuitive result that the high SNR risk typically equal or better than the MIS Bayes risk even in the low SNR regime has been observed. The only case in which the MIS Bayes risk had superior performance was when a prior was placed over σ^2, which is not expected for nonlinear models. In figure 8.1 large discrepancies between the MIS and high SNR Bayes risk in the high SNR regime for wide priors on both the T_1 and M_0 were observed which is counterintuitive. This may be attributed to the normality assumptions failing for the high SNR Bayes risk. It is also possible that there is a nuanced bug somewhere in the code that remains to be found. Unfortunately only the linear model may be evaluated exactly and it agrees well with the MIS result. Note that this cause may be too simple to bring out the software error (if it exists). In our experience, the MIS code has displayed surprising robustness to implementation errors as the averaging performed tends to reduce the impact of bugs. On the other hand, it could just be that the inversion recovery model with normal additive noise is simply not 'nonlinear enough' to display enhanced performance over the high SNR risk.

The next step in verifying the performance of these online experiment design strategies is to study them experimentally. One candidate is the T_1 of the phosphorus donor in silicon at low temperatures which was studied in the first half of this thesis. It should be possible to realize a significant improvement in total calibration time of the system T_1^{DNP} which must be calibrated upon every cooldown, cutting into our limited cold time (10-12 hours). Another aspect of the system that has eluded estimation is relaxation time T_1 of the phosphorus donor which has historically been very difficult to measure; no signal decay was observed after a relaxation τ of 4.5 hours [33]. The current experiment setup has an improved SNR and hyperpolarization rates. It is conceivable that it may now be possible to measure this rate with enhanced estimation strategies.

Another avenue to explore is the application of online experiment design strategies to model selection. In conjunction with a prior defined over model parameters when several competing models are available it is possible to set a prior over the models themselves [30]. The question "What experiment should be performed to discern the true model optimally?" may be asked. Thus allowing experimental design steps to be incorporated with model selection. It should even be possible to strike a balance between model selection and parameter estimation with a set of $\mathbf{Q}$ weighting matrices.

The initial motivation for these design strategies was their application to commercial NMR oil logging [46]. The primary time constraint in such systems is the estimation of T_1 rates. The algorithms described here provide methods which may be used to expedite relaxation rate estimation.

References

[1] *Fundamentals of Semiconductors - Physics and Materials Properties — Peter YU — Springer.*

[2] HV connector source - Google Groups. https://groups.google.com/forum/#!msg/sci.electronics/9VdG

[3] *Sequential Monte Carlo Methods in Practice — Arnaud Doucet — Springer.*

[4] IEEE Guide for Field Testing and Evaluation of the Insulation of Shielded Power Cable Systems Rated 5 kV and Above. *IEEE Std 400-2012 (Revision of IEEE Std 400-2001)*, pages 1–54, June 2012.

[5] Eisuke Abe, Alexei M. Tyryshkin, Shinichi Tojo, John J. L. Morton, Wayne M. Witzel, Akira Fujimoto, Joel W. Ager, Eugene E. Haller, Junichi Isoya, Stephen A. Lyon, Mike L. W. Thewalt, and Kohei M. Itoh. Electron spin coherence of phosphorus donors in silicon: Effect of environmental nuclei. *Physical Review B*, 82(12):121201, September 2010.

[6] Daniel Abergel and Arthur G. Palmer. Approximate Solutions of the Bloch–McConnell Equations for Two-Site Chemical Exchange. *ChemPhysChem*, 5(6):787–793, June 2004.

[7] Phillip E. Allen and Douglas R. Holberg. CMOS Analog Circuit Design. http://cds.cern.ch/record/1136695, 2002.

[8] B. Andreas, Y. Azuma, G. Bartl, P. Becker, H. Bettin, M. Borys, I. Busch, M. Gray, P. Fuchs, K. Fujii, H. Fujimoto, E. Kessler, M. Krumrey, U. Kuetgens, N. Kuramoto, G. Mana, P. Manson, E. Massa, S. Mizushima, A. Nicolaus, A. Picard, A. Pramann, O. Rienitz, D. Schiel, S. Valkiers, and A. Waseda. Determination of the Avogadro Constant by Counting the Atoms in a ${̂28}\mathrm{Si}$ Crystal. *Physical Review Letters*, 106(3):030801, January 2011.

[9] null Appelt, null Wäckerle, and null Mehring. Deviation from Berry's adiabatic geometric phase in a 131Xe nuclear gyroscope. *Physical Review Letters*, 72(25):3921–3924, June 1994.

[10] J. Ballato, T. Hawkins, P. Foy, R. Stolen, B. Kokuoz, M. Ellison, C. McMillen, J. Reppert, A. M. Rao, M. Daw, S. Sharma, R. Shori, O. Stafsudd, R. R. Rice, and D. R. Powers. Silicon optical fiber. *Optics Express*, 16(23):18675–18683, November 2008.

[11] C. D. Bass, T. D. Bass, B. R. Heckel, C. R. Huffer, D. Luo, D. M. Markoff, A. M. Micherdzinska, W. M. Snow, H. E. Swanson, and S. C. Walbridge. A liquid helium target system for a measurement of parity violation in neutron spin rotation. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 612(1):69–82, December 2009.

[12] Christopher D. Bass, Tiffany D. Bass, B. R. Heckel, C. R. Huffer, Paul R. Huffman, D. Luo, D. M. Markoff, A. M. Micherdzinska, William M. Snow, H. E. Swanson, and S. C. Walbridge. A liquid helium target system for a measurement of parity violation in neutron spin rotation. *Nuclear Instruments & Methods in Physics Research Section A-Accelerators Spectrometers Detectors and Associated Equipment*, October 2010.

[13] P. Becker, H.-J. Pohl, H. Riemann, and N. Abrosimov. Enrichment of silicon for a better kilogram. *physica status solidi (a)*, 207(1):49–66, January 2010.

[14] J. B. Brask, R. Chaves, and J. Kołodyński. Improved Quantum Magnetometry beyond the Standard Quantum Limit. *Physical Review X*, 5(3):031010, July 2015.

[15] J. D. Brownridge. Device for removal of solid impurities from liquid nitrogen. *Cryogenics*, 29:70–71, 1989.

[16] Marina Brustolon. *Electron Paramagnetic Resonance: A Practitioner's Toolkit.* John Wiley & Sons, February 2009.

[17] R. Byrd, P. Lu, J. Nocedal, and C. Zhu. A Limited Memory Algorithm for Bound Constrained Optimization. *SIAM Journal on Scientific Computing*, 16(5):1190–1208, September 1995.

[18] Kathryn Chaloner and Isabella Verdinelli. Bayesian Experimental Design: A Review. *Statistical Science*, 10(3):273–304, August 1995.

[19] Y. H. Chen and S. A. Lyon. Photoluminescence and diffusivity of free excitons in doped silicon. *IEEE Journal of Quantum Electronics*, 25(5):1053–1055, May 1989.

[20] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth. On the Lambert<Emphasis Type="Italic">W</Emphasis> function. *Advances in Computational Mathematics*, 5(1):329–359, December 1996.

[21] Harald Cramér. *Mathematical Methods of Statistics.* Princeton University Press, Princeton, 1999. OCLC: 185436716.

[22] G. Feher. Electron Spin Resonance Experiments on Donors in Silicon. I. Electronic Structure of Donors by the Electron Nuclear Double Resonance Technique. *Physical Review*, 114(5):1219–1244, June 1959.

[23] G. Feher. Nuclear Polarization via "Hot" Conduction Electrons. *Physical Review Letters*, 3(3):135–137, August 1959.

[24] G. Feher and E. A. Gere. Electron Spin Resonance Experiments on Donors in Silicon. II. Electron Spin Relaxation Effects. *Physical Review*, 114(5):1245–1256, June 1959.

[25] Christopher Ferrie, Christopher E. Granade, and D. G. Cory. How to best sample a periodic probability distribution, or on the accuracy of Hamiltonian finding strategies. *Quantum Information Processing*, 12(1):611–623, January 2013.

[26] R. Fletcher. *Practical Methods of Optimization.* John Wiley & Sons, June 2013.

[27] Ray Freeman, H. D. W Hill, and R Kaptein. An adaptive scheme for measuring NMR spin-lattice relaxation times. *Journal of Magnetic Resonance (1969)*, 7(1):82–98, May 1972.

[28] Martin Fuechsle, Jill A. Miwa, Suddhasatta Mahapatra, Hoon Ryu, Sunhee Lee, Oliver Warschkow, Lloyd C. L. Hollenberg, Gerhard Klimeck, and Michelle Y. Simmons. A single-atom transistor. *Nature Nanotechnology*, 7(4):242–246, April 2012.

[29] Eiichi Fukushima. *Experimental Pulse NMR: A Nuts and Bolts Approach.* CRC Press, Boulder, Colo., 1 edition edition, October 1993.

[30] Andrew Gelman, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin. *Bayesian Data Analysis, Third Edition.* Chapman and Hall/CRC, Boca Raton, 3 edition edition, November 2013.

[31] Christopher Granade, Christopher Ferrie, Ian Hincks, Steven Casagrande, Thomas Alexander, Jonathan Gross, Michal Kononenko, and Yuval Sanders. QInfer: Statistical inference software for quantum applications. *Quantum*, 1:5, April 2017.

[32] M. A. Green. Solar cells: Operating principles, technology, and system applications. January 1982.

[33] P. Gumann, O. Patange, C. Ramanathan, H. Haas, O. Moussa, M. L. W. Thewalt, H. Riemann, N. V. Abrosimov, P. Becker, H.-J. Pohl, K. M. Itoh, and D. G. Cory. Inductive Measurement of Optically Hyperpolarized Phosphorous Donor Nuclei in an Isotopically Enriched Silicon-28 Crystal. *Physical Review Letters*, 113(26):267604, December 2014.

[34] J. Haase, N. J. Curro, and C. P. Slichter. Double Resonance Probes for Close Frequencies. *Journal of Magnetic Resonance*, 135(2):273–279, December 1998.

[35] H.-Y. Hao and H. J. Maris. Dispersion of the long-wavelength phonons in Ge, Si, GaAs, quartz, and sapphire. *Physical Review B*, 63(22):224301, May 2001.

[36] R. K Harris and R. H Newman. Choice of pulse spacings for accurate T1 and NOE measurements in NMR spectroscopy. *Journal of Magnetic Resonance (1969)*, 24(3):449–456, December 1976.

[37] Hiroshi Hasegawa. Spin-Lattice Relaxation of Shallow Donor States in Ge and Si through a Direct Phonon Process. *Physical Review*, 118(6):1523–1534, June 1960.

[38] Matthew D. Hoffman and Andrew Gelman. The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. *arXiv:1111.4246 [cs, stat]*, November 2011.

[39] Holger Haas, Rahul Deshpande, Thomas Alexander, Mike Thewalt, and D. G. Cory. Phonon Mediated Hyperpolarization of Phosphorus Donor Nuclei in Silicon. *In Preparation*, 2018.

[40] D. I. Hoult. The principle of reciprocity in signal strength calculations—A mathematical guide. *Concepts in Magnetic Resonance*, 12(4):173–187, January 2000.

[41] Ivan Hrvoic, Greg M Hollyer, and P Eng. Brief review of quantum magnetometers. *GEM Systems Technical Papers*, 2005.

[42] Ian Hincks, Thomas Alexander, Michal Kononenko, Benjamin Soloway, and David G. Cory. Adaptive Online Experiment Design with Quantum Systems. *In Preparation*.

[43] J. Järvinen, J. Ahokas, S. Sheludyakov, O. Vainio, L. Lehtonen, S. Vasiliev, D. Zvezdov, Y. Fujii, S. Mitsudo, T. Mizusaki, M. Gwak, SangGap Lee, Soonchil Lee, and L. Vlasenko. Efficient dynamic nuclear polarization of phosphorus in silicon in strong magnetic fields and at low temperatures. *Physical Review B*, 90(21):214401, December 2014.

[44] E. T. Jaynes. *Probability Theory: The Logic of Science*. Cambridge University Press, Cambridge, UK ; New York, NY, annotated. edition edition, April 2003.

[45] B. E. Kane. A silicon-based nuclear spin quantum computer. *Nature*, 393(6681):133–137, May 1998.

[46] Robert L. Kleinberg and Jasper A. Jackson. An introduction to the history of NMR well logging. *Concepts in Magnetic Resonance*, 13(6):340–342, January 2001.

[47] Donald E. Knuth. *Art of Computer Programming, Volume 2: Seminumerical Algorithms*. Addison-Wesley Professional, Reading, Mass, 3 edition edition, November 1997.

[48] W. Kohn and J. M. Luttinger. Hyperfine Splitting of Donor States in Silicon. *Physical Review*, 97(4):883–888, February 1955.

[49] Melvin Lax. Cascade Capture of Electrons in Solids. *Physical Review*, 119(5):1502–1523, September 1960.

[50] Micah Ledbetter, Kasper Jensen, Ran Fischer, Andrey Jarmola, and Dmitry Budker. Gyroscopes based on nitrogen-vacancy centers in diamond. *Physical Review A*, 86(5), November 2012.

[51] George C Levy and Ian R Peat. The experimental approach to accurate carbon-13 spin-lattice relaxation measurements. *Journal of Magnetic Resonance (1969)*, 18(3):500–521, June 1975.

[52] Adrian S. Lewis and Michael L. Overton. Nonsmooth optimization via quasi-Newton methods. *Mathematical Programming*, 141(1-2):135–163, October 2013.

[53] G. Lindblad. On the generators of quantum dynamical semigroups. *Communications in Mathematical Physics*, 48(2):119–130, 1976.

[54] Jane Liu and Mike West. Combined Parameter and State Estimation in Simulation-Based Filtering. In *Sequential Monte Carlo Methods in Practice*, Statistics for Engineering and Information Science, pages 197–223. Springer, New York, NY, 2001.

[55] G. G. Macfarlane, T. P. McLean, J. E. Quarrington, and V. Roberts. Fine Structure in the Absorption-Edge Spectrum of Si. *Physical Review*, 111(5):1245–1254, September 1958.

[56] G. D. Mahan and Ryan Woodworth. Spin-exchange scattering in semiconductors. *Physical Review B*, 78(7):075205, August 2008.

[57] H. S. W. Massey and B. L. Moiseiwitsch. The Scattering of Electrons by Hydrogen Atoms. *Physical Review*, 78(2):180–181, April 1950.

[58] D. R. McCamey, J. van Tol, G. W. Morley, and C. Boehme. Fast Nuclear Spin Hyperpolarization of Phosphorus in Silicon. *Physical Review Letters*, 102(2):027601, January 2009.

[59] Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics*, 195306.

[60] G. E. Moore. Cramming More Components Onto Integrated Circuits. *Proceedings of the IEEE*, 86(1):82–85, January 1998.

[61] G. W. Morley, D. R. McCamey, H. A. Seipel, L.-C. Brunel, J. van Tol, and C. Boehme. Long-Lived Spin Coherence in Silicon with an Electrical Spin Trap Readout. *Physical Review Letters*, 101(20):207602, November 2008.

[62] Masatoshi Nakayama and Hiroshi Hasegawa. Relaxation Effects on Donor Spins in Silicon and Germanium. *Journal of the Physical Society of Japan*, 18(2):229–239, February 1963.

[63] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, Cambridge ; New York, 10th anniversary ed. edition edition, December 2010.

[64] Albert W. Overhauser. Polarization of Nuclei in Metals. *Physical Review*, 92(2):411–415, October 1953.

[65] Giuseppe Pica, Gary Wolfowicz, Matias Urdampilleta, Mike L. W. Thewalt, Helge Riemann, Nikolai V. Abrosimov, Peter Becker, Hans-Joachim Pohl, John J. L. Morton, R. N. Bhatt, S. A. Lyon, and Brendon W. Lovett. Hyperfine Stark effect of shallow donors in silicon. *Physical Review B*, 90(19):195204, November 2014.

[66] C. Piermarocchi, F. Tassone, V. Savona, A. Quattropani, and P. Schwendimann. Exciton formation rates in GaAs/${\mathrm{Al}}_{\mathrm{x}}$${\mathrm{Ga}}_{1\mathrm{\ensuremath{-}}\mathrm{x}}$As quantum wells. *Physical Review B*, 55(3):1333–1336, January 1997.

[67] David Pines, John Bardeen, and Charles P. Slichter. Nuclear Polarization and Impurity-State Spin Relaxation Processes in Silicon. *Physical Review*, 106(3):489–498, May 1957.

[68] Kamyar Saeedi, Stephanie Simmons, Jeff Z. Salvail, Phillip Dluhy, Helge Riemann, Nikolai V. Abrosimov, Peter Becker, Hans-Joachim Pohl, John J. L. Morton, and Mike L. W. Thewalt. Room-Temperature Quantum Bit Storage Exceeding 39 Minutes Using Ionized Donors in Silicon-28. *Science*, 342(6160):830–833, November 2013.

[69] R. R. Schaller. Moore's law: Past, present and future. *IEEE Spectrum*, 34(6):52–59, June 1997.

[70] W. Schmid. Auger lifetimes for excitons bound to neutral donors and acceptors in Si. *physica status solidi (b)*, 84(2):529–540, December 1977.

[71] N. Sclar. Neutral Impurity Scattering in Semiconductors. *Physical Review*, 104(6):1559–1561, December 1956.

[72] T. Sekiguchi, M. Steger, K. Saeedi, M. L. W. Thewalt, H. Riemann, N. V. Abrosimov, and N. Nötzel. Hyperfine Structure and Nuclear Hyperpolarization Observed in the Bound Exciton Luminescence of Bi Donors in Natural Si. *Physical Review Letters*, 104(13):137402, April 2010.

[73] Peter W. Shor. Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer. *SIAM Journal on Computing*, 26(5):1484–1509, October 1997.

[74] K. N. Shrivastava. Theory of Spin–Lattice Relaxation. *physica status solidi (b)*, 117(2):437–458, June 1983.

[75] Charles P. Slichter. *Principles of Magnetic Resonance*. Springer, Berlin ; New York, 3rd enlarged and updated ed. 1990. corr. 3rd printing 1996 edition edition, March 1996.

[76] C. R. N. Staff. Intel's Gelsinger Sees Clear Path To 10nm Chips. http://www.crn.com/news/components-peripherals/208801780/intels-gelsinger-sees-clear-path-to-10nm-chips.htm, June 2008.

[77] H. Taitelbaum, J. A. Ferretti, R. G. S. Spencer, and G. H. Weiss. Two-Stage Inversion-Recovery Experiments for Measurements of T1. *Journal of Magnetic Resonance, Series A*, 105(1):59–64, October 1993.

[78] H. Taitelbaum, J. A. Ferretti, R. G. S. Spencer, and G. H. Weiss. Optimization of Two-Stage Measurements of T1. *Journal of Magnetic Resonance, Series A*, 109(2):166–171, August 1994.

[79] M. L. W. Thewalt. Details of the structure of bound excitons and bound multiexciton complexes in Si. *Canadian Journal of Physics*, 55(17):1463–1480, September 1977.

[80] D. D. Thornton and A. Honig. Shallow-Donor Negative Ions and Spin-Polarized Electron Transport in Silicon. *Physical Review Letters*, 30(19):909–912, May 1973.

[81] Christoph W. Ueberhuber. *Numerical Computation 2: Methods, Software, and Analysis*. Springer, Berlin ; New York, softcover reprint of the original 1st ed. 1997 edition edition, April 1997.

[82] Michele Vallisneri. Use and abuse of the Fisher information matrix in the assessment of gravitational-wave parameter-estimation prospects. *Physical Review D*, 77(4):042001, February 2008.

[83] Harry L. Van Trees. Classical Detection and Estimation Theory. In *Detection, Estimation, and Modulation Theory, Part I*, pages 19–165. John Wiley & Sons, Inc., 2001.

[84] M. Veldhorst, H. G. J. Eenink, C. H. Yang, and A. S. Dzurak. Silicon CMOS architecture for a spin-based quantum computer. *Nature Communications*, 8(1):1766, December 2017.

[85] M. Veldhorst, C. H. Yang, J. C. C. Hwang, W. Huang, J. P. Dehollain, J. T. Muhonen, S. Simmons, A. Laucht, F. E. Hudson, K. M. Itoh, A. Morello, and A. S. Dzurak. A two-qubit logic gate in silicon. *Nature*, 526(7573):410–414, October 2015.

[86] Anne S. Verhulst, Ileana G. Rau, Yoshihisa Yamamoto, and Kohei M. Itoh. Optical pumping of ${̂29}\mathrm{Si}$ nuclear spins in bulk silicon at high magnetic field and liquid helium temperature. *Physical Review B*, 71(23):235206, June 2005.

[87] David Wales and Jonathan Doye. Global Optimization by Basin-Hopping and the Lowest Energy Structures of Lennard-Jones Clusters Containing up to 110 Atoms. March 1998.

[88] George H Weiss and James A Ferretti. The choice of optimal parameters for measurement of spin-lattice relaxation times. III. Mathematical preliminaries for nonideal pulses. *Journal of Magnetic Resonance (1969)*, 61(3):490–498, February 1985.

[89] George H. Weiss, Raj K. Guptaj, James A. Ferretti, and Edwin D. Becker. The choice of optimal parameters for measurement of spin-lattice relaxation times. I. Mathematical formulation. *Journal of Magnetic Resonance*, 37:369–379, 1980.

[90] D. K. Wilson and G. Feher. Electron Spin Resonance Experiments on Donors in Silicon. III. Investigation of Excited States by the Application of Uniaxial Stress and Their Importance in Relaxation Processes. *Physical Review*, 124(4):1068–1083, November 1961.

[91] D. K. Wilson and G. Feher. Electron Spin Resonance Experiments on Donors in Silicon. III. Investigation of Excited States by the Application of Uniaxial Stress and Their Importance in Relaxation Processes. *Physical Review*, 124(4):1068–1083, November 1961.

[92] M. F. Wilson, D. O. Edwards, and J. T. Tough. Construction of Vycor Glass Superleaks. *Review of Scientific Instruments*, 39(1):134–134, January 1968.

[93] A. Yang, M. Steger, T. Sekiguchi, M. L. W. Thewalt, J. W. Ager, and E. E. Haller. Homogeneous linewidth of the P31 bound exciton transition in silicon. *Applied Physics Letters*, 95(12):122113, September 2009.

[94] A. Yang, M. Steger, T. Sekiguchi, M. L. W. Thewalt, T. D. Ladd, K. M. Itoh, H. Riemann, N. V. Abrosimov, P. Becker, and H.-J. Pohl. Simultaneous Subsecond Hyperpolarization of the Nuclear and Electron Spins of Phosphorus in Silicon by Optical Pumping of Exciton Transitions. *Physical Review Letters*, 102(25):257401, June 2009.

[95] Peter YU and Manuel Cardona. *Fundamentals of Semiconductors: Physics and Materials Properties*. Springer, Berlin ; New York, 4th ed. 2010 edition edition, July 2010.

[96] Yantian Zhang, Hong N. Yeung, Matthew O'Donnell, and Paul L. Carson. Determination of sample time for T1 measurement. *Journal of Magnetic Resonance Imaging*, 8(3):675–681, May 1998.

APPENDICES

Appendix A

Practical T_1 Estimation

The implementation of Bayesian parameter estimation and online experimental design relies on small tricks and optimizations to be made practical. This appendix tries to summarize these techniques.

A.1 Removing Stochastic Noise from MIS Bayes Risk

The MIS Bayes risk as described in section 7.1.1 to select the minimal risk experiment *i.e.* best experiment, which is necessarily an optimization problem, which may be solved with ones flavour of the month optimization algorithm. The MIS Bayes risk as described above is stochastic due to the requisite sampling of the outcome distribution and results in a noisy Bayes Risk utility function as seen in figure 8.1(a). As the majority of optimization algorithms assume a smooth objective function optimization of the MIS Bayes Risk will be made more difficult.

Two solutions present themselves. The naive approach is simply to increase the number of outcome particles to reduce the variance of the Bayes risk evaluation to an acceptable level for optimization algorithms. However, this solution can require many thousands of outcomes to be sampled making it very expensive to implement.

Rather this problem is approached by exploiting the fact that the noise model of interest is additive and independent of experiment choice. Recalling that in equation (7.12) a sampling step is performed to draw outcomes from the distribution for the j^{th} particle. Rewriting this step, now accounting for the evaluated experiment parameters $\mathbf{c}$

$$o_{j;\mathbf{c}} \sim \Pr(o|\mathbf{x}_j;\mathbf{c}) \qquad \forall \mathbf{x}_j \in \mathbf{X}_{sam}. \tag{A.1}$$

Now the additive normal noise model is exploited to rewrite this step as

$$\begin{aligned} \Pr(o|\mathbf{x}_j;\mathbf{c}) &= \mathscr{N}(\mathbf{x_j}, \sigma(\mathbf{x}_j)^2) && \text{(A.2)}\\ &= \mathscr{N}(f(\mathbf{x_j};\mathbf{c}), \sigma_j^2) && \text{(A.3)}\\ &= f(\mathbf{x}_j;\mathbf{c}) + \mathscr{N}(0, \sigma(\mathbf{x}_j)^2) = f(\mathbf{x}_j;\mathbf{c}) + \epsilon_j. && \text{(A.4)} \end{aligned}$$

Outcome sampling corresponds to evaluating the model function and adding some sampled noise value ϵ_j, which may depend on the sampled particle $\mathbf{x}_j$ but not the experiment to be performed $\mathbf{c}$. We define the set of sampled noise values $\epsilon = \{\epsilon_1, \cdots, \epsilon_j\}$. Now assume there is a set of possible experiments $\mathbf{C}$ with which the Bayes risk must be evaluated with, in order to determine the optimal experiment. As before, draw an initial set of model parameters $\mathbf{X}_{sam}$ and set of noise values

$$\epsilon_j \sim \mathcal{N}(0, \sigma(\mathbf{x}_j)^2) \qquad \forall \mathbf{x}_j \in \mathbf{X}_{sam}, \tag{A.5}$$

at which point $r(\pi; \mathbf{c}_l) \quad \forall \mathbf{c}_l \in \mathbf{C}$ may be evaluated. Provided the model function $f(\mathbf{x}; \mathbf{c})$ is a continuous function of $\mathbf{c}$, the Bayes risk will now be continuous allowing the Bayes risk to be used as an input function for most numerical optimization algorithms.

Note that the above caching based solution may be used for any model distribution where the model noise is independent of the experimental settings. In situations where this condition does not hold, the MIS Bayes risk may still be optimized with a combination of a large number of outcome samples to reduce risk variance and stochastic optimization routines [52].

A.2 Bayes Risk Optimization Algorithms

Online experimental design exploration within the thesis was performed with a naive grid search minimization algorithm. As the computational complexity of grid search is polynomial with the degree being proportional to the number of unknown model parameters it rapidly becomes infeasible with complex models. In practice quasi-Newton minimization methods such as Broyden–Fletcher–Goldfarb–Shanno algorithm (BFGS) [26] and L-BFGS [17] may be used to find the minimum of the Bayes risk with fewer function calls, while also allowing the parameter search spacing to be bounded. As these methods are typically susceptible to search initialization conditions and local minima they may be incorporated into a basin-hopping algorithm which attempts to find the global minimum of a function [87]. A comparison of performance and evaluation times shown in figure A.1 demonstrates the benefit of utilizing more sophisticated optimization strategies.

A.3 Rescaled Weighting (Q) Matrix

The Bayes risk evaluates the expected posterior variance weighted by the diagonal matrix $\mathbf{Q}$. This is sufficient if all estimated model parameters $\hat{\mathbf{x}}$ have similar values, which is not true in general. Consider the two parameter case $\mathbf{x} = \{x, y\}$ with true values $\mathbf{x}_t = \{x_t, y_t\}$ with Bayes risk $r(\mathbf{c}|\pi) = Q_{11} r_x + Q_{22} r_y$. If $\mathbf{Q}$ is just the identity we have $Q_{11} r_x + Q_{22} r_y = r_x + r_y$. Consider now the change of units for only x with the transformation $x' = \alpha x$. The risk will transform as $r(\mathbf{c}|\pi)_\alpha = Q_{11} \alpha^2 r_x + Q_{22} r_y S$. It is easy to see that if α is large experiments

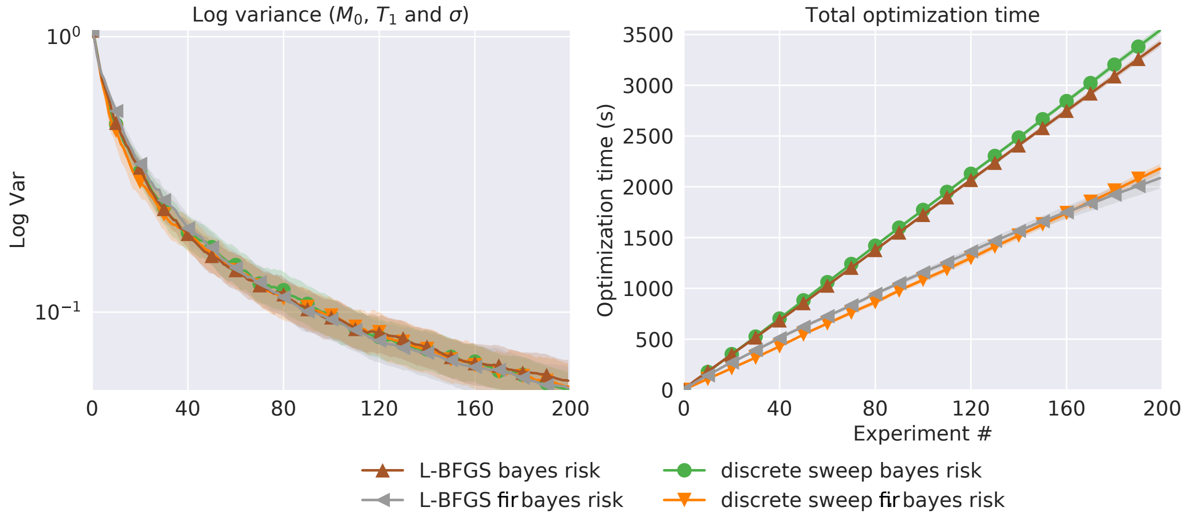


Figure A.1: Comparing the performance of grid search (300 points) to L-BFGS risk function optimization (left) Log variance plot for Bayes risk, and FIR Bayes risk online experiment design. No difference in performance is observed. (right) The average wall clock optimization time between the two optimization algorithms. Note that L-BFGS is significantly quicker.

will disproportionately be designed to minimize the variance of x and vice versa for small α, when in truth the variance of both model parameters is of equal concern. In practice parameters are often rescaled by some factor α_i, or have their mean shifted by some amount β_i

$$\Lambda(x_i) = \alpha_i(x_i + \beta_i). \tag{A.6}$$

The ideal experiment design heuristic would be invariant to these mappings. The mean, variance and posterior variance transform under this mapping as

$$\hat{x}_i \xrightarrow{\Lambda} \alpha_i(\hat{x}_i + \beta_i) \qquad \mathrm{Cov}\left[\mathbf{x}\right]_{ij} \xrightarrow{\Lambda} \alpha_i\alpha_j \mathrm{Cov}\left[\mathbf{x}\right]_{ij}$$
$$\mathbb{E}_{o_{n+1}}\left[\mathrm{Cov}\left[\mathbf{x}|o_{n+1}, O; c\right]\right]_{ij} \xrightarrow{\Lambda} \alpha_i\alpha_j\, \mathbb{E}_{o_{n+1}}\left[\mathrm{Cov}\left[\mathbf{x}|o_{n+1}, O; c\right]\right]_{ij}.$$

Recalling that the Bayes risk is

$$ri(c) = \sum_{i=1}^{N} Q_{ii}\alpha_i^2\, \mathbb{E}_{o_{n+1}}\left[\mathrm{Cov}\left[\mathbf{x}|o_{n+1}, O; c\right]\right]_{ii} \xrightarrow{\Lambda} \sum_{i=1}^{N} Q_{ii}\alpha_i^2\, \mathbb{E}_{o_{n+1}}\left[\mathrm{Cov}\left[\mathbf{x}|o_{n+1}, O; c\right]\right]_{ii}. \tag{A.7}$$

Let $Q_{ii} \to Q_{ii}/\mathrm{Cov}_{\mathbf{x}\sim\pi(x|I)}\left[\mathbf{x}\right]_{ii} \xrightarrow{\Lambda} Q_{ii}/\alpha_{ii}^2\mathrm{Cov}_{\mathbf{x}\sim\pi(x|I)}\left[\mathbf{x}\right]_{ii}$, and obtain the new quantity

$$rr(c) = \sum_{i=1}^{N} \frac{Q_{ii}\, \mathbb{E}_{o_{n+1}}\left[\mathrm{Cov}\left[\mathbf{x}|o_{n+1}, O; c\right]\right]_{ii}}{\mathrm{Cov}_{\mathbf{x}\sim\pi(x|I)}\left[\mathbf{x}\right]_{ii}}, \tag{A.8}$$

and title this the the risk ratio. Where $\text{Cov}_{\mathbf{x}\sim\pi(x|I)}\left[\mathbf{x}\right]_{ii}$ is the initial prior distribution before any observations have been made. Importantly the risk ratio is invariant under both shifting and rescaling of model parameters. It has the simple interpretation as the weighted sum of ratios between the expected posterior variance of the current prior distribution and variance of the initial prior distribution. Note that the rr is also unitless which has up until now been ignored when defining the Bayes risk and the matrix $\mathbf{Q}$. If the current prior rather than the initial prior is used poor performance may be experienced as parameters with negligible variance may contribute significant weighting to the rr if their ratios of improvement are large, and the result will be a further reduction of variance for an already negligible quantity. Note that when using the initial prior weighted quality matrix the risk improvement has the interpretation as the expected reduction in the percent variance with respect to the initial prior.

The rr makes explicit the importance of the experimentalist in selecting a proper risk weighting matrix $\mathbf{Q}$ matrix just as it is essential to choose an appropriate prior. The setting of Q_{ii} weighs the importance of improving the variance of a model parameter over the experimentalists initial state of knowledge, and the selection of a poor prior for a model parameter will result in the selection of experiments that reflect this prior.