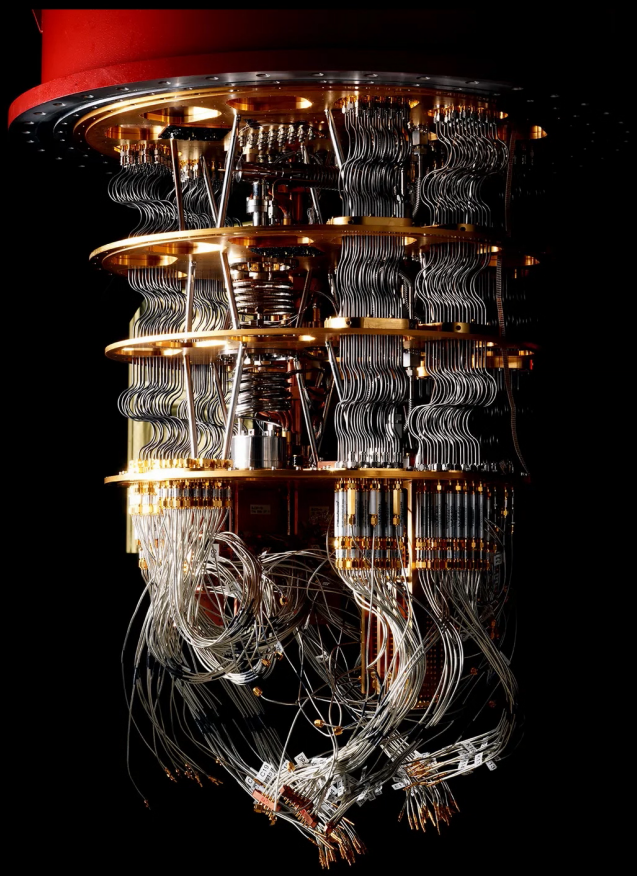
Gate Set Tomography on the Nitrogen-Vacancy Center

Zoya Valeriyivna Polshchykova



Delft University of Technology

TU Delft

Gate Set Tomography on the Nitrogen-Vacancy Center

by

Zoya Valeriyivna Polshchykova

To obtain the degree of Bachelor of Science in Applied Mathematics and Applied Physics
at Delft University of Technology

Student number:	4921607
Supervisors:	Dr. J. Borregaard
	Dr. D. de Laat
Other committee members:	Dr. E. Greplová
	Dr. M. Möller
Project Duration:	February, 2022 - July, 2022
Faculty:	Faculty of Applied Sciences and
	Faculty of Electrical Engineering,
	Mathematics & Computer Science, Delft

Cover: A cryostat at Google's quantum computing lab in Santa Barbara, California (2021) (Modified)

TU Delft

Abstract

In this thesis, gate set tomography (GST) has been conducted on the nitrogen vacancy center (NV). Gate set tomography is a protocol for characterization of logic operations (gates) on quantum computing processors. The NV's electron served as a qubit. The quantum circuits were run both experimentally as well as on an NV-simulator. GST is different from its predecessors in the sense that it estimates all aspects of the processor simultaneously, without assuming any of its parts to be ideal. However, it does assume that the gates are Markovian. This allows to analyse the gate errors more precisely via their error generators. In addition to this, the diamond norm and a measure of the amount of model violation were used to examine the results. The electron qubit can couple to the nearby nitrogen nucleus, which can cause non-Markovian dynamics. The nitrogen nucleus ($S = 1$) can be initialised using nuclear spin polarization. The effects of different initialisation procedures on GST's results were researched. Furthermore, a dynamical decoupling XY-4 echo sequence could be employed to protect the quantum state of the qubit. How the presence of the echo affected the estimated gates and their errors was probed. It was discovered that the echo was extremely good at decreasing the amount of model violation, most likely by preventing the electron from coupling to the nitrogen, which can lead to non-Markovian dynamics. Without the echo, GST's estimates were satisfactory only if the nitrogen nucleus was initialised with a high enough fidelity, at least 0.95. Another topic for further research would be to perform GST on the electron and the nitrogen system, by modelling the nitrogen nucleus as a qutrit.

Contents

1

Introduction

Near the end and at the beginning of the nineteenth century there were a couple of things "left" that physicists could not get their minds wrapped around. *"I have now put before you what I consider to be the greatest difficulty yet encountered by the molecular theory"*, said Maxwell of the discrepancy in the heat capacity between observations and physical laws in 1869 [1]. Then Planck and Einstein came up with solutions to respectively the black-body radiation problem (1900) and the photoelectric effect (1905). However, this "Old Quantum theory" was short-lived [2] and the couple of things "left" would lead to a whole new area in physics: quantum mechanics. As one revolution generates the next, so gifted the founding fathers of *quantenmechanik* [3] us not only the laws governing the universe, but also the opportunity to use these laws to create new technologies: the second quantum revolution [4]. Quantum communication, computation, simulation, metrology, and sensing technologies are being developed all over the world by thousands of scientists.

Forty years after Feynman's proposal of using quantum physics to build a more powerful kind of computer [5], the scientific community is unfortunately not (nowhere near) there. Nevertheless, progress has been made and at least it has been established which properties a quantum computer should have, by DiVincenzo's criteria [6]. Three of these are i) controlled state preparation and ii) unitary transformations of the qubits, and iii) state-specific quantum measurements. In this thesis, a powerful tool towards realizing these three criteria is considered: gate set tomography (GST).

"Gate set tomography is not just hyperaccurate, it's a different way of thinking" [7], is how its inventors like to see it, and rightly so. GST has been proposed and developed by Nielsen *et al* [8] at Sandia National Laboratories, and has been picked up by many since its birth in 2012. In Delft, it has been used to demonstrate the existence of extremely fault tolerant semiconductor qubits [9]. In principle, a quantum information processor executes three tasks: qubits are prepared in some state, then manipulated, and finally measured. These operations form the *gate set*, for which GST finds the best-fit via maximum likelihood estimation. GST's "different way of thinking" lies in that it recognizes the *inter dependency* of the gate set's elements, instead of viewing the processor as a sum of its parts, like (some of) its predecessors, state, process, and measurement tomography do. Moreover, it tells which types of errors cause its estimations to differ from ideal, such that accuracy can be taken to even further levels, closer and closer to ideal.

Nitrogen vacancy (NV) centers in diamond are one of the most advanced platforms for quantum computation so far. In such systems, one carbon atom has been replaced by a nitrogen atom and one has been kicked out to form a vacancy. The electron of the NV can be used as a qubit, as in this thesis, possessing promising properties. These include second long coherence times, and a wide working temperature range, up to room temperature. Nearby atoms can serve as qubits as well. The surrounding carbon-13 atoms, for example, hold the record for the longest dephasing time (1.93 minutes) for individually controlled qubits [10]. Furthermore, NV centers can be connected and entangled over large distances, such that quantum computers can be formed by connecting many copies of simple quantum processors. These quantum networks are suitable for investigating complex algorithms and error correction [11], and teleportation-based protocols [12].

In this thesis, first, the literature research is presented. In Chapter 2, the necessary quantum theory is discussed. As stated above, gate set tomography uses maximum likelihood estimation to optimize its

results. To this end, it employs the Levenberg-Marquardt numerical optimization algorithm, which is explained in Chapter 3. Then, in Chapter 4, gate set tomography is considered. In this work, the electron of the NV center functions as qubit, so in Chapter 5 its physical system is described. The second part of this thesis consists of GST experiments on the NV center, described in Chapter 6. The results are given in Chapter 7 and discussed in Chapter 8. Finally, this work is concluded in Chapter 9. This thesis was written as part of the double bachelor's degree in Applied Mathematics and Applied Physics at Delft University of Technology.

2

Quantum Theory

In this chapter, the quantum theory fundamental to the rest of this thesis, mainly to gate set tomography, will be presented. Gate set tomography (GST) seeks to characterize quantum information processors. Thus, to understand GST, one first needs to understand how quantum processors work. However, in physics, but not confined to physics only, it is naive to question *how* something works before figuring out *why* you want something to work. The answer to the why-question can be summarized in one sentence: quantum information processors are used to manipulate qubits. Qubits will be the first thing described in this chapter. Their mathematical description and the space where they live will be presented. Secondly, the how-question will be considered. Quantum information processing employs *circuits* of *logic gates* to manipulate qubits. The empirical observation that any description of reality is exactly this, merely a description, holds for quantum processors as well. So finally, gate set tomography's *description* of quantum processors, its *model*, and its weak and strong points will be examined. For a more detailed review, I refer to Nielsen and Chuang [13] and Nielsen *et al.* [8].

2.1. Quantum mechanics: states, Hilbert-Schmidt space and measurements

This section starts off with a short recap of qubits and Hilbert space. Then, a description of qubits commonly used in quantum information processing will be given, namely the density matrix. Gate set tomography slightly transforms this notation into so-called *superket's* and *superbra's*, which will be considered thereafter. Lastly, a useful tool for visualising one qubit and how a quantum processor manipulates it will be presented: the *Bloch sphere*.

2.1.1. Quantum states in Hilbert-Schmidt space

According to the first postulate of quantum mechanics, quantum states "live" in a complex d-dimensional Hilbert space $\mathcal{H}$, where per definition $d = 2^n$ for a system of n qubits, with n finite, or more precisely $n = 1$, in this thesis [13]:

Postulate 1. *Associated to any isolated physical system is a complete normed complex vector space with inner product (that is, a Hilbert space) known as the state space of the system. The system is completely described by its state vector, which is a unit vector in the system's state space.*

An alternative formulation for the state vector approach known by most novices at quantum mechanics (like in Griffiths [14]), employs so-called *density operators* or *density matrices*. Density matrices are more general and are more convenient when, for example, describing systems whose state is not completely known [13]. Then, the system's density operator is defined as follows [13]:

Definition 2.1. A quantum state is described by a $d \times d$ density matrix $\rho : \mathcal{H} \to \mathcal{H}$. More precisely, suppose a quantum system is in one of the states $\left|\psi_i\right\rangle$ with probabilities p_i, where $i \in \mathbb{N}$, then the density matrix is given by: $\rho \equiv \sum_i p_i \left|\psi_i\right\rangle \left\langle\psi_i\right|$.

A quantum state is called pure if $i = 1$ in the equation above and *mixed* otherwise, i.e. if it is an ensemble (sum) of pure states. A pure state can be represented by a unit vector in the Hilbert space,

in accordance with Postulate 1. Reading this postulate more carefully shows why the density matrix notation is useful when dealing with real (actual) physical systems: the postulate only holds for *isolated physical systems*. Strictly speaking, the only truly isolated physical system is the whole universe. So, as every qubit ever considered is affected by some type(s) of noise, its state is in fact a mixed one. These cannot be represented by traditional state vectors, hence the need for density matrices.

There are two necessary and sufficient conditions a density operator needs to satisfy to represent a physically valid state [13]. Its trace needs to equal one (trace-1 condition) and it needs to be a positive operator, i.e. the density matrix is positive semidefinite. The density matrices considered here are assumed to be self-adjoint, i.e. Hermitian. For more details on the density operator I refer to Nielsen and Chuang [13], as in this thesis it serves only as a way of introducing the state notation used in GST.

Hermitian matrices form a real d^2-dimensional subspace of the *Hilbert-Schmidt space* $\mathcal{B}(\mathcal{H})$, the complex d^2-dimensional vector space of $d \times d$ matrices. In this thesis, notation is slightly abused such that $\mathcal{B}(\mathcal{H})$ denotes the *real subspace* of the Hilbert-Schmidt space. It will be the qubits' living-space, as will become clear in the next section. From now on, until the Bloch sphere section, Nielsen *et al* [8] will be followed.

$\mathcal{B}(\mathcal{H})$ is endowed with an inner product: $\langle A, B\rangle \equiv \mathrm{Tr}(A^\dagger B)$. Analogous to the Dirac notation on a Hilbert space, an element of the Hilbert-Schmidt space will be denoted by a column vector $|B\rangle\rangle$, a *superket*. Likewise, elements of the isomorphic dual space will be represented by *superbra's* $\langle\langle A|$, such that $\langle\langle A|B\rangle\rangle = \mathrm{Tr}(A^\dagger B)$. Often it is useful to define a basis $\{B_i\}$ for the Hilbert-Schmidt space, satisfying the following, where $\mathbb{1}$ denotes the $d \times d$-dimensional identity map:

1. Hermiticity: $B_i = B_i^\dagger$
2. Orthonormality: $\mathrm{Tr}(B_i B_j) = \delta_{ij}$
3. Traceless for $i > 0$: $B_0 = \mathbb{1}/\sqrt{d}$ and $\mathrm{Tr}(B_i) = 0 \; \forall i > 0$.

The normalized Pauli matrices $\{\mathbb{1}/\sqrt{2}, \sigma_x/\sqrt{2}, \sigma_y/\sqrt{2}, \sigma_z/\sqrt{2}\}$ are commonly used as a basis for a single qubit. Extension to n-qubit systems is achieved by utilising n-fold tensor products of the Pauli operators and the identity operator.

2.1.2. Measurements

Now, measurements are described. In qubit systems, an outcome or measurement result can be seen as one possibility out of a discrete set of k alternatives. Furthermore, *"each outcome's probability is a linear function of the state ρ"* [8]. So, the ith outcome is represented by a dual vector $\langle\langle E_i|$. The probability of measuring this outcome when the initial state was ρ is given by $\Pr(i|\rho) = \langle\langle E_i|\rho\rangle\rangle$. Here, the E_i's are *effects*. The set of effects $\{E_i\}$ is called a *positive operator-valued measure (POVM)* and completely describes the measurement, since it provides the agent with the probability of each possible outcome. In accordance with the laws of probability, effects satisfy $\sum_i E_i = \mathbb{1}$ and $E_i > 0$.

A brief example will now be provided to illustrate the concepts explained above. Consider some qubit, with density matrix as in Eq.(2.1). It follows that it can be represented by some $|\rho\rangle\rangle \in \mathcal{B}(\mathcal{H})$. For this, ρ is written as a linear combination of the basis for the Hilbert-Schmidt space, the set of normalized Pauli matrices, in this case: $\rho = \sum_i c_i B_i$ (Eq.(2.2)). c_i are real coefficients that can be found by taking inner products: $c_i = \mathrm{Tr}(B_i^\dagger \rho)$. $|\rho\rangle\rangle$ is then given by the column vector $\mathbf{c}$ (Eq.(2.3)).

$$\rho = \begin{pmatrix} 3/4 & (1+i)/8 \\ (1-i)/8 & 1/4 \end{pmatrix} \tag{2.1}$$

$$\rho = \frac{1}{\sqrt{2}}\left(\frac{\mathbb{1}}{\sqrt{2}}\right) + \frac{1}{4\sqrt{2}}\left(\frac{\sigma_x}{\sqrt{2}}\right) - \frac{1}{4\sqrt{2}}\left(\frac{\sigma_y}{\sqrt{2}}\right) + \frac{1}{2\sqrt{2}}\left(\frac{\sigma_z}{\sqrt{2}}\right) \tag{2.2}$$

$$|\rho\rangle\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1/4 \\ -1/4 \\ 1/2 \end{pmatrix} \tag{2.3}$$

The superbra notation of effects corresponding to measuring respectively the $|0\rangle$ and $|1\rangle$ state is found in the same way. $E_0 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ and $E_1 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$ are written as $\langle\langle E_0| = \begin{pmatrix} 1/\sqrt{2} & 0 & 0 & 1/\sqrt{2} \end{pmatrix}$,

$\langle\langle E_1| = \begin{pmatrix} 1/\sqrt{2} & 0 & 0 & -1/\sqrt{2} \end{pmatrix}$. Finally, the probabilities of measuring 0 or 1 are given by the inner products:

$$p_0 = \langle\langle E_0|\rho\rangle\rangle = \mathrm{Tr}(E_0\rho) = \frac{3}{4} \tag{2.4}$$

$$p_1 = \langle\langle E_1|\rho\rangle\rangle = \mathrm{Tr}(E_1\rho) = \frac{1}{4}, \tag{2.5}$$

which can be read off the diagonals of the density matrix (Eq.(2.1)) directly.

2.1.3. The Bloch sphere

Finally, the Bloch sphere will be introduced as an aid in visualising *single* qubit states. A single qubit quantum state can be represented by a *Bloch vector* inside or on the Bloch sphere, a three-dimensional sphere with unit radius. The derivation of the Bloch vector in the case of a pure qubit, one in the state vector notation, is given in the appendix (A.1.1). When dealing with density matrices, the Bloch vector is defined as follows. As explained and showed above, every ρ can be written as

$$\rho = \frac{1}{2}(\mathbb{1} + v_x\sigma_x + v_y\sigma_y + v_z\sigma_z), \tag{2.6}$$

where the coefficient in front of the identity matrix is fixed, because of the trace-1 condition on ρ. The Bloch vector is then given by (v_x, v_y, v_z). If the qubit's state is pure, then its Bloch vector has length one and thus lies on the Bloch sphere. A mixed state's Bloch vector lies inside the Bloch sphere. An example is depicted in figure 2.1. In the next section, the Bloch sphere will function as a useful tool for visualizing operations on qubits.

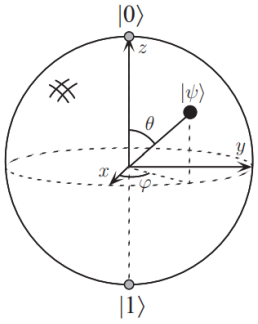


Figure 2.1: The Bloch sphere. The $|0\rangle$ and $|1\rangle$ states correspond to the Bloch sphere's intersection with the z-axis. A quantum state $|\psi\rangle$ can be represented by a Bloch vector, whose angles with the z- and $x-$ axes are θ and ϕ respectively. Figure is taken from Nielsen and Chuang [13].

2.2. Quantum computation: logic gates, circuits and processors

In this section, the manipulation of qubits via quantum information processors will be explained using a top-down approach. Firstly, processors and circuits will be described. Then, logic gates, comprising quantum circuits, will be studied through the eyes of gate set tomography. Nielsen *et al* [8] is followed in this section.

2.2.1. Processors and circuits

A quantum information processor *runs quantum circuits* by executing three steps on n qubits. First, the qubits are initialised into some state. Then, the qubits' states are changed, or manipulated, by quantum *logic gates*. The final step of a circuit is measurement of the qubits using the set of effects $\{E_i\}$. The overall circuit is a *fixed-input classical output circuit* (FI/CO); the input is "fixed" by the state preparation, and the output is classical measurement data. The middle part of this circuit is a *quantum-input quantum-output circuit* (QI/QO); the input is quantum states, and the output produced by the logic gates is quantum states. In figure 2.2, an example circuit of a 2-qubit processor is shown. Both qubits start off in the $|0\rangle$

state. Measurement is indicated by the icon of a half-circle with an arrow going through it. Gates (the $X^{1/2}, Y^{1/2}, Y$ and "crossed-circle" icons) are applied in between. Here, every step is a *circuit layer*. If only one qubit is considered, as in this thesis, every circuit layer is comprised of only one logic gate. The "sandwiched" middle step C of the circuit (the QI/QO circuit) is a sequence of quantum logic gates: $C = (\gamma_1, ..., \gamma_L)$ where L is the number of circuit layers (gates applied), and $\gamma_i \in \{G_i\}_{i=1}^{N_G}$ with $1 \leq \gamma_i \leq N_G$ is some gate. Now, logic gates will be defined mathematically.

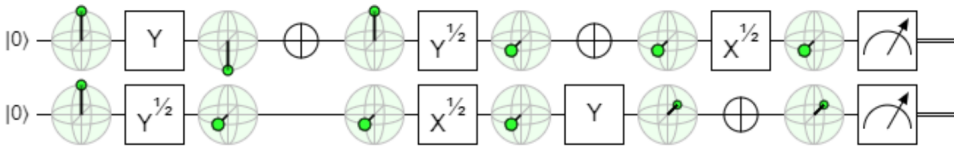


Figure 2.2: An example of a quantum circuit of a 2-qubit processor. The qubits start off in the $|0\rangle$-state. A sequence of gates is applied on them. The gates are indicated by the $X^{1/2}, Y^{1/2}, Y$ and "crossed-circle" icons. Measurement is depicted by the half-circles with an arrow going through them. The qubits' states are visualised by Bloch spheres in between every circuit layer. Figure made with [15].

2.2.2. Logic gates and their properties

In this thesis, logic gates will transform a qubit via left multiplication with its superket. Such operators are called *superoperators*. Their *transfer matrix* representation will be used:

Definition 2.2. Any superoperator $\Lambda : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H})$ can be denoted by a $d^2 \times d^2$ transfer matrix S_Λ that acts associatively on $|\rho\rangle\rangle \in \mathcal{B}(\mathcal{H})$, i.e. $\Lambda : |\rho\rangle\rangle \mapsto S_\Lambda |\rho\rangle\rangle$.

The transfer matrix for a QI/QO circuit C is denoted by $\tau(C) : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H})$, such that $\tau(C) = G_{\gamma_L} \cdots G_{\gamma_1}$. Bluntly put, gates are considered physically valid when, if the input is physically valid, the produced output is as well. Recall that a density matrix is considered physically valid if it satisfies the trace-1 and positivity conditions. So, only i) *trace-preserving* and ii) *completely positive* superoperators are considered physically valid. The first condition implies that $\mathrm{Tr}(\rho)$ is equal to 1 before and after the operation, so that the superoperator complies with the laws of probability. Secondly, a stronger condition than simple preservation of positivity, $\rho \geq 0$, is required. If a superoperator Λ acts on a component of a larger system, it has to actually preserve the positivity of that extended system, i.e. $\Lambda \otimes \mathbb{1}_{\mathcal{A}} \geq 0$ for any *auxiliary* state space $\mathcal{A}$. Such a superoperator is called completely positive. The *completely positive and trace-preserving* (CPTP) constraint is necessary *and* sufficient, but will not be assumed unless stated otherwise, to for example probe how GST performs without the constraint. Hereafter, the term "superoperator" will be used to refer to its transfer matrix.

In this thesis, five different superoperators will be estimated using gate set tomography. We will refer to the superoperators as gates, since they will be employed in quantum circuits. As only one qubit will be used, these superoperators are 4×4 matrices:

$$I = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, X^{\pi/2} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix}, Y^{\pi/2} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix},$$

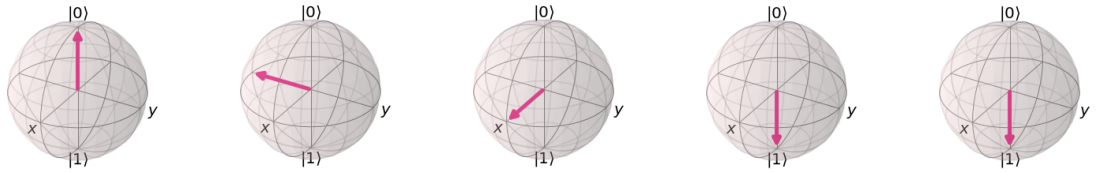
$$X^{\pi} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \text{ and } Y^{\pi} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}.$$

The gate names, excluding I, refer to the rotation axis and rotation angle of their operation on the qubit. An example is given to illustrate how applying gates looks like. Suppose a qubit in the $|0\rangle$- state (Z-basis). Its density matrix is: $\rho = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right)$, which becomes $|\rho\rangle\rangle = 1/\sqrt{2} \begin{pmatrix} 1 & 0 & 0 & 1 \end{pmatrix}^T$ in the superket notation. Applying a gate on this qubit is now easy: the superoperator is multiplied with the superket. For example, applying a $X^{\pi/2}$ gate on this qubit results in

$$|\rho'\rangle\rangle = X^{\pi/2}|\rho\rangle\rangle = 1/\sqrt{2} \begin{pmatrix} 1 & 0 & -1 & 0 \end{pmatrix}^T$$

The transfer notation makes calculating circuits' outcomes easy. The probability of an outcome i after applying some gate with transfer matrix S_Λ is: $p_i = \langle\langle E_i|S_\Lambda|\rho\rangle\rangle = \mathrm{Tr}(E_i S_\Lambda \rho)$.

The action of all five given gates on $|0\rangle$ is visualised using Bloch spheres. The initial state's vector is given in Fig. 2.3a, and the resulting states after applying one of $X^{\pi/2}$, $Y^{\pi/2}$, X^{π}, Y^{π} on the initial state is depicted in Fig.2.3b, 2.3c, 2.3d and 2.3e, respectively. The I gate is the identity gate and its resulting state is equal to the initial state.



(a) No gate or I gate. **(b)** $X^{\pi/2}$ gate. **(c)** $Y^{\pi/2}$ gate. **(d)** X^{π} gate. **(e)** Y^{π} gate.

Figure 2.3: Visualisation of the gates used in this thesis via Bloch vectors. The initial state is $|0\rangle$, depicted in 2.3a. The resulting state after applying one of the gates, indicated in the secondary captions, is depicted in the remaining figures. Figures made using Vos' "Bloch widget" [16].

Lastly, a norm used for comparing gates is given, the diamond norm [8].

Definition 2.3. If G and G_0 are gates, then the diamond norm is

$$\|G - G_0\|_\diamond = \sup_\rho \left\|(G \otimes \mathbb{1}_d)[\rho] - (G_0 \otimes \mathbb{1}_d)[\rho]\right\|_1 . \tag{2.7}$$

ρ ranges over all valid quantum states. $\otimes$ denotes the tensor product.

Since the diamond norm ranges over all possible ρ's, it cannot be computed analytically in most cases. It is mostly computed using semidefinite programming. Stated simply, the diamond norm tells how well two gates can be distinguished by looking at how different their resulting quantum states are, giving an upper bound on this "distinguishability". For example, suppose either G (with probability p) or G_0 (with probability $1-p$) is applied on some quantum state. The maximal probability of guessing correctly which one had been applied by measuring the resulting state is then given by $p_{\text{success}} = 1/2 + 1/2\left\|pG - (1-p)G_0\right\|_\diamond$ [17].

2.2.3. XY-4 echo sequence

Like systems in the classical world, no quantum system is truly isolated. Quantum systems are extremely sensitive to perturbations, caused by the system's coupling to the environment. This results in loss of quantum information: decoherence. Decoherence is an obstacle in quantum information processing, as the lifetime of information can become too short for practical computation. Dynamical decoupling (DD) methods seek to reduce perturbation of the qubit(s), by *de*-coupling the qubit from the environment. This way, the qubit's input state is preserved. DD can also be combined with gates, i.e. decoupling after every gate operation, like in this thesis. In 1950, Hahn discovered that inverting the qubit by applying a gate, inverts the perturbation Hamiltonian as well, time-reversing the perturbation and nullifying it. The XY-4 sequence considered in this thesis is an example of such an inversion sequence. In this subsection, first this mechanism is described intuitively, and then formally. To this end, a single spin-$\frac{1}{2}$ system functions as a qubit, and the environment is modelled as a spin-bath.

First, the intuitive explanation [18]. The surrounding bath-spins flip-flop randomly, creating a slowly varying magnetic field. This changes the qubit's energy levels, resulting in a random phase evolution and/or qubit flips. Now, inverting the state of the qubit via π-gates (recall that this flips a qubit's Bloch vector) inverts the magnetic field produced by the surrounding spins as well. If the pulse is symmetric in time, i.e. the qubit is as long in the non-inverted state as in the inverted one, the effect of the surrounding field is canceled. Of course this only works if the magnetic field is constant over time, i.e. the next pulse comes before the environment has had time to change significantly. This condition is satisfied by decreasing the waiting time. As pulses have finite lengths, this is reduction is limited. Furthermore, multiple pulses can be applied to "protect" the qubit spin for longer times, i.e. increase the

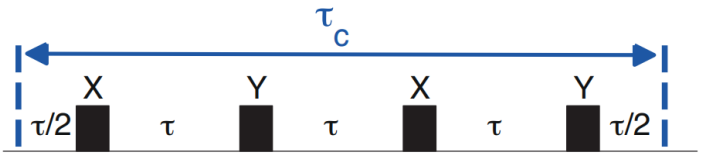


Figure 2.4: The XY-4 sequence. The input state is located at the left (blue) dashed line. After $\tau/2$, the XY-4 sequence is applied using the gates: $X = X^{\pi}$ and $Y = Y^{\pi}$, with τ in between. After $\tau/2$, the inversion of the system-environment interaction has been completed. In this work, the XY-4 has been combined with gate operations, which are applied at the blue lines. Figure adapted from Souza *et al* [19].

qubit's coherence time. However, the amount of power deposited in the system by applying gates has to be taken into account as well to avoid heating effects or damaging the sample.

Secondly, the more formal description. The total Hamiltonian of the qubit-system and its environment is [19]

$$H = H_S + H_{SE} + H_E. \tag{2.8}$$

Here, H_S is the qubit system Hamiltonian and H_E is the environment Hamiltonian. H_{SE} is system-environment interaction, given by

$$H_{SE} = \sum_k \beta_x^k S_x I_x^k + \beta_y^k S_y I_y^k + \beta_z^k S_z I_z^k, \tag{2.9}$$

where k indexes the environment spin and $\boldsymbol{I^k} = (I_x^k, I_y^k, I_z^k)$ is the corresponding spin vector operator. $\boldsymbol{S} = (S_x, S_y, S_z)$ (Eq.(A.4)) is qubit's spin vector operator. In a resonantly rotating (with the qubit) reference frame considered here, $H_S = 0$. If the energy level splitting of the qubit differs significantly from the energy level splitting of the spin-bath (i.e. every spin in the environment), H_{SE} can be approximated by a $z - z$ interaction [20] (heterenuclear system). This causes qubit dephasing (z-component), but not qubit flips ($x-$ and $y-$components). If the energy level splitting are similar, this approximation to H_{SE} cannot be made (homonuclear system). Dynamical decoupling aims to eliminate the average time-independent SE interaction Hamiltonian over some time period τ_c, i.e. $\overline{H}_{SE} = 0$, by applying some inversion sequence. If the SE interaction includes all three components (x, y, z: homonuclear), then the decoupling sequence must include rotations around at least two different axes [20].

Now, it will be showed that the XY-4 sequence (Fig.2.4) in fact eliminates the system-environment Hamiltonian by applying the following π-pulses: $[\tau/2 - X^{\pi} - \tau - Y^{\pi} - \tau - X^{\pi} - \tau - Y^{\pi} - \tau/2]$. The total sequence is then given by [19]

$$U = e^{-iH\tau/2}\left(\prod_{k=2}^{4} R_k e^{-iH\tau}\right) R_1 e^{-iH\tau/2}, \tag{2.10}$$

where H is the total Hamiltonian (Eq.(2.8)), $R_1 = R_3 = e^{-i\pi S_x}$ (the X^{π} gate) and $R_2 = R_4 = e^{-i\pi S_y}$ (the Y^{π} gate). Recall that π-pulses flip the sign of H_{SE}. Thus, the above sequence is rewritten using

$$\tilde{H}_{k=1,5} = (H_E + H_{SE}), \tag{2.11}$$

$$\tilde{H}_{k=2,3,4} = [H_E + (-1)^{k+1} H_{SE}], \tag{2.12}$$

such that $U = \prod_{k=1}^{5} e^{-i\tilde{H}_k \tau_k}$ with $\tau_{k=1,5} = \tau/2$ and $\tau_{k=2,3,4} = \tau$. Using the Baker-Campbell-Hausdorff formula,

$$\log e^A e^B \approx A + B + \frac{1}{2}[A, B] + \frac{1}{12}([A, [A, B]] + [[A, B], B]), \tag{2.13}$$

the matrix exponentials are computed explicitly. This results in zeroth and first order averages Hamiltonians: $\overline{H_0} = H_E$ and $\overline{H_1} = 0$ [19]. The second order term is non-zero, $\overline{H_2} = \tau^2/8\left[[H_e, H_{SE}], H_e - 1/3 H_{SE}\right]$, but is assumed (and higher order terms as well) small enough to conclude that H_{SE} is effectively

eliminated by the XY-4 sequence over a time period 2τ. Note that by decreasing τ the second order term can be decreased further.

Lastly, if the gates in the sequence are not perfect, pulse imperfections can accumulate and destroy the state of the qubit by creating an effective general SE interaction [20]. Initially, the XY-4 sequence was developed to combat this effect, as it applies pulses along different spatial directions. Furthermore, more exotic sequences can be created by varying the rotation angles of the applied pulses and the times between them. This, as well as imperfect echo sequences are beyond the scope of this thesis and more can be found in Souza *et al* [19] [20].

2.3. Modelling quantum processors: gate sets, error generators and gauge freedom

In the final section of this chapter, it will be explained how gate set tomography models quantum processors. First, the notion of *gate sets* is introduced. Then, it will be presented what gate estimations tell us about the modelled quantum processor. Lastly, one of the "flaws" of GST, precisely a consequence of GST's use of gate sets, is explained.

2.3.1. Gate sets and gate set models

Suppose a quantum processor can perform N_{G} distinct gates, N_ρ distinct state preparations, and N_M distinct measurements, where the m-th measurement has $N_{\mathrm{E}}^{(m)}$ distinct outcomes. These operations will be represented by

$$\begin{aligned} G_i : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H}) \text{ for } i = 1, ..., N_{\mathrm{G}}, \\ |\rho^{(i)}\rangle\rangle \text{ for } i = 1, ..., N_\rho \text{ and} \\ \langle\langle E_i^{(m)}| \text{ for } m = 1, ..., N_M \text{ and } i = 1, ..., N_{\mathrm{E}}^{(m)}. \end{aligned}$$

The notion of a *gate set* $\mathcal{G}$ is used to model and describe a quantum processor's behaviour:

Definition 2.4. A gate set $\mathcal{G} \in \mathcal{M}$ (the "matrix space") is the set of a quantum processor's *all* distinct state preparations, gates and measurements, which are defined as above:

$$\mathcal{G} = \left\{ \left\{ |\rho^{(i)}\rangle\rangle \right\} ; \{G_i\} ; \left\{ \langle\langle E_i^{(m)}| \right\} \right\}.$$

Here $\mathcal{M}$ is isomorphic to $\mathbb{R}^{N_{\mathrm{e}}}$, where $N_{\mathrm{e}} = d^4 N_{\mathrm{G}} + d^2 \left(N_\rho + \sum_{m=1}^{N_M} N_{\mathrm{E}}^{(m)} \right)$ is the total number of (real) elements in a gate set.

In the remainder of this thesis, various target gate sets will be estimated using a *gate set model* and gate set tomography, which will be explained in the next chapter. Oftentimes, the target gate set is taken to be the ideal gates a device is designed to implement. For now, a gate set model is defined:

Definition 2.5. A gate set model corresponds to a choice of a parameter space $\mathcal{P}$ and a map $W : \mathcal{P} \to \mathcal{M}$. Here $\mathcal{P}$ has dimension N_{p} and is isomorphic to $\mathbb{R}^{N_{\mathrm{p}}}$. A gate set model associates every point in parameter space with a gate set.

A gate set model is called *fully parameterized* if $\mathcal{P} = \mathcal{M}$ and $W(x) = x$, i.e. every element of every operation in the gate set is an independent parameter: $N_{\mathrm{p}}^{\mathrm{full}} = N_{\mathrm{e}}$. Oftentimes a *smaller* gate set model is defined ($N_{\mathrm{p}} < N_{\mathrm{e}}$), which for example parameterises strict subsets of $\mathcal{M}$ only, e.g. CPTP gate sets.

A *TP parameterized gate set model* has $N_{\mathrm{p}} = N_{\mathrm{p}}^{\mathrm{TP}}$ parameters, where

$$N_{\mathrm{p}}^{\mathrm{TP}} = N_{\mathrm{G}} d^2 (d^2 - 1) + N_\rho (d^2 - 1) + \left(\sum_{m=1}^{N_{\mathrm{M}}} N_{\mathrm{E}}^{(m)} \right) d^2.$$

Such gate set models enforce density matrices to have unit trace, and the gates applied on them to be trace-preserving. This results in the first row of every superoperator being $[1, 0, ..., 0]$ and the first element of every state preparation vector being equal to $1/\sqrt{d}$, because $\mathrm{Tr}(B_0) = d/\sqrt{d} = \sqrt{d}$.

A *CP parameterized gate set model* is, in comparison with the TP parameterized gate set model, a bit more challenging to define and impose, since the CP constraint is a nonlinear inequality. Here, one way to construct a mapping function W whose range is constrained to CPTP gate sets is presented. `pyGSTi`, the software used to conduct gate set tomography, uses this CPTP parameterization as well. A

CP gate is written in terms of an *error generator* L, which is also a superoperator. If G_k is the transfer matrix of the (estimated) gate and G_k^0 is the corresponding error-free CPTP operation, then we define $L = \log\left(G_k(G_k^0)^{-1}\right)$ such that

$$G_k = e^L G_k^0 \tag{2.14}$$

Error generators are modelled as *post-gate error processes*, meaning it is as if the error is applied directly after the gate. CPTP maps form a semigroup, meaning products of CPTP maps are CPTP as well. Thus, G_k can be forced to be CPTP by having e^L be a CPTP-map. Now, this is done by restricting L to be of Lindbladian form, i.e. in accordance with the Lindblad Master Equation. The Lindblad Master equation is the most general generator of Markovian dynamics in quantum systems [21]. That is: $\partial\rho/\partial t = \dot{\rho} = L[\rho]$ with L Lindbladian describes Markovian dynamics, which will be defined in the next section. For now, the Lindbladian form of L is presented:

$$L = \sum_{i=1}^{d^2} \alpha_i H_i + \sum_{j,k=2}^{d^2} \beta_{jk} S_{jk} \tag{2.15}$$

Here, H_i and S_{jk} are the following operators, acting on density matrices ρ:

$$H_i : \rho \to i[\rho, B_i], \tag{2.16}$$

$$S_{jk} : \rho \to B_j \rho B_k - \frac{1}{2}(B_k B_j \rho + \rho B_k B_j). \tag{2.17}$$

$\alpha_i \in \mathbb{R}$, β is a positive semidefinite (Hermitian) matrix, and B_i is a basis for $\mathcal{B}(\mathcal{H})$ as described in Ch.2.1. In this thesis, the basis will be formed by the normalized Pauli matrices $\{\mathbb{1}/\sqrt{2}, \sigma_x/\sqrt{2}, \sigma_y/\sqrt{2}, \sigma_z/\sqrt{2}\}$, so $d^2 = 4$. This means β is a 4×4 matrix. Beginning at 2, the indices in the second sum only go over *non-identity* elements. This Lindbladian form of the error generator can be split into four parts, such that each part corresponds to a known quantum logical error. These errors will be described in the next subsection.

The number of parameters for such a CPTP gate set model is the same as for the TP model above. With the CP model, the number of parameters is not reduced; it is the same as in the full model.

2.3.2. Error generators: small Markovian errors

The above formulation of the error generators allows us not only to impose the CPTP condition, but to also analyse how the estimated gate differs from the ideal gate. In the `pyGSTi` authors' opinion, error generators are the most useful diagnostic for gate errors. The analysis relies on the assumption that the target gate G_k^0 and the estimated gate G_k differ by a small, Markovian-type amount. First, 'Markovianity' is explained, a term for which in the scientific community *many* definitions exist. It is oftentimes used to denote the "memory-less" property of a system. This notion is formalized by the definition of Markovianity used in this thesis:

Definition 2.6. A quantum gate is Markovian if it is a solution of a master equation with generator in Lindblad form (Eq.(2.15)), i.e. it can be written in the form of Eq.(2.14) with L Lindbladian. Moreover, the generator L is time-independent.

So, maps with time-*dependent* generators are non-Markovian. The memory-less property means that, at any point in time in the future, evolution depends only on the present state and not on any previous states (the history of the system). That is, in a process changing a system's state, $\rho \to \rho'$, ρ' is completely determined by ρ. To summarize this: any CPTP map, where the CPTP constraint is implemented as above, describes Markovian dynamics. Now, an error is called *small*, if $G_k - G_k^0$ is small, meaning that terms of order $O([G_k - G_k^0]^2)$ can be neglected. A sufficient condition for this is: $\left\|G_k - G_k^0\right\|_\diamond \ll 1$, where $\|\cdot\|_\diamond$ is the diamond norm (Eq.(2.7)).

Recall the Lindbladian form of the error generator (Eq.(2.15)). We are going to split it into four terms, each corresponding to some type of error. For now, we omit the factors $1/\sqrt{2}$ in the Pauli basis, such that $\{B_1, B_2, B_3, B_4\} = \{\mathbb{1}, \sigma_x, \sigma_y, \sigma_z\} = \{\mathbb{1}, X, Y, Z\}$. We will index this set by B and the set of Pauli operators $\{X, Y, Z\}$ (so without the identity matrix) by P.

$$L[\rho] = \sum_B \alpha_B i[\rho, B] + \sum_{P,Q} \beta_{P,Q}\Big[P\rho Q - \frac{1}{2}(QP\rho + \rho QP)\Big] \tag{2.18}$$

$$= \alpha_{\mathbb{1}} i(\rho\mathbb{1} - \mathbb{1}\rho) + \sum_P \alpha_i i[\rho, P] + \sum_{P,Q,P=Q} \beta_{P,Q}\Big[P\rho Q - \frac{1}{2}(QP\rho + \rho QP)\Big] +$$

$$\sum_{P,Q,P\neq Q} \beta_{P,Q}\Big[P\rho Q - \frac{1}{2}(QP\rho + \rho QP)\Big] \tag{2.19}$$

For every Pauli matrix P, $P^2 = \mathbb{1}$. The first term above disappears and working out the third term gives:

$$L[\rho] = \sum_P \alpha_P i[\rho, P] + \sum_P \beta_P\big[P\rho P - \rho\big] + \sum_{P,Q,P\neq Q} \beta_{P,Q}\Big[P\rho Q - \frac{1}{2}(QP\rho + \rho QP)\Big].$$

Now, we are going to work out the last term. Note that since β is Hermitian, $\beta_{P,Q} = \overline{\beta_{Q,P}}$, i.e. $\mathrm{Re}(\beta_{P,Q}) = \mathrm{Re}(\beta_{Q,P})$ and $\mathrm{Im}(\beta_{P,Q}) = -\mathrm{Im}(\beta_{Q,P})$. The last term is split in distinct pairs of distinct indices, and the notation $P, Q > P$ is used to denote that only one of the pairs P, Q and Q, P is considered. We define $c_{P,Q} = \mathrm{Re}(\beta_{P,Q})$ and $a_{P,Q} = \mathrm{Im}(\beta_{P,Q})$. Lastly, $[\cdot,\cdot]$ is used to denote the commutator and $\{\cdot,\cdot\}$ denotes the anticommutator.

$$\sum_{P,Q,P\neq Q} \beta_{P,Q}\Big[P\rho Q - \frac{1}{2}(QP\rho + \rho QP)\Big]$$

$$= \sum_{R=\{P,Q|P,Q>P\}} \beta_{P,Q}\Big[P\rho Q - \frac{1}{2}(QP\rho + \rho QP)\Big] + \sum_{(P,Q,P\neq Q)\setminus R} \beta_{Q,P}\Big[Q\rho P - \frac{1}{2}(PQ\rho + \rho PQ)\Big]$$

$$= \sum_{P,Q>P} \mathrm{Re}(\beta_{P,Q})\Big[P\rho Q + Q\rho P - \frac{1}{2}(QP\rho + \rho QP + PQ\rho + \rho PQ)\Big] +$$

$$\sum_{P,Q>P} \mathrm{Im}(\beta_{P,Q}) i\Big[P\rho Q - \frac{1}{2}(QP\rho + \rho QP) - Q\rho P + \frac{1}{2}(PQ\rho + \rho PQ)\Big]$$

$$= \sum_{P,Q>P} c_{P,Q}\Big[P\rho Q + Q\rho P - \frac{1}{2}((PQ + QP)\rho + \rho(PQ + QP))\Big] +$$

$$\sum_{P,Q>P} a_{P,Q} i\Big[P\rho Q - Q\rho P + \frac{1}{2}((PQ - QP)\rho + \rho(PQ - QP))\Big]$$

$$= \sum_{P,Q>P} c_{P,Q}\Big[P\rho Q + Q\rho P - \frac{1}{2}\big\{\{P, Q\}, \rho\big\}\Big] + \sum_{P,Q>P} a_{P,Q} i\Big[P\rho Q - Q\rho P + \frac{1}{2}\big\{[P, Q], \rho\big\}\Big]$$

To summarize, we have dissected the Lindblad:

$$L = L_{\mathbb{H}} + L_{\mathbb{S}} + L_{\mathbb{C}} + L_{\mathbb{A}}$$
$$= \sum_P h_P H_P + \sum_P s_P S_P + \sum_{P,Q>P} c_{P,Q} C_{P,Q} + \sum_{P,Q>P} a_{P,Q} A_{P,Q}$$

We call each coefficient the *rate* of the corresponding error process, and $\{H_P\}$, $\{S_P\}$, $\{C_{P,Q}\}$, $\{A_{P,Q}\}$ the *elementary generators* of the mutually disjoint subspaces $\mathbb{H}, \mathbb{S}, \mathbb{C}, \mathbb{S}$, of the *error generator space* ($\mathbb{L} = \mathbb{H} \oplus \mathbb{S} \oplus \mathbb{C} \oplus \mathbb{A}$). As we have seen above, every $L \in \mathbb{L}$ can be written as a linear combination of elementary generators with real coefficients. For overview, the elementary error generators:

$$H_P[\rho] = -i[P, \rho] = -iP\rho\mathbb{1} + i\mathbb{1}\rho P \tag{2.20}$$

$$S_P[\rho] = P\rho P - \mathbb{1}\rho\mathbb{1} \tag{2.21}$$

$$C_{P,Q}[\rho] = P\rho Q + Q\rho P - \frac{1}{2}\{\{P, Q\}, \rho\} \tag{2.22}$$

$$A_{P,Q}[\rho] = i\big(P\rho Q - Q\rho P + \frac{1}{2}\{[P, Q], \rho\}\big) \tag{2.23}$$

Now, following Blume-Kohout *et al.*[22], these errors and their effect on the Bloch sphere will be described.

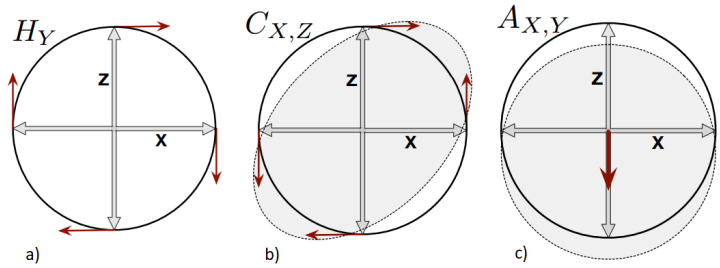


Figure 2.5: a) The 1-qubit Hamiltonian error generator H_Y acts on the $X-Z$ plane by rotating: $Z \rightarrow X$ and $X \rightarrow -Z$. **b)** The 1-qubit Correlation error generator $C_{X,Z}$ acts on the Bloch sphere by pulling it in the positive $x-$ and z-axis, and the negative $x-$ and z-axis. **c)** The active error generator $A_{X,Y}$ shifts the Bloch sphere in the negative z-direction. Figures taken from Blume-Kohout *et al.* [22]

Hamiltonian generators

The first class of errors we will consider, are *Hamiltonian generators*. Any unitary error can be written as a linear combination of d^2-1 (so 3 in the 1-qubit case) Hamiltonian generators (Eq.(2.20).a), and they span the (d^2-1)-dimensional subspace $\mathbb{H}$ of $\mathbb{L}$. We will explain them further for the 1-qubit case. Hamiltonian generators generate unitary rotations of the Bloch sphere; *coherent errors*, a legitimate and physically valid unitary error process, meaning the length of the Bloch vector of the state the gate acts on, does not get shorter (i.e. no decoherence occurs). $\vec{h} = (h_X, h_Y, h_Z)$ is a pseudovector defining the erroneous rotation, whose three 'coordinates' tell how much the rotation differs from ideal with respect to each Pauli axis (Fig.2.5.a). If G has only coherent errors, the error generator will consist entirely out of generators from $\mathbb{H}$, as the other subspaces contain non-coherent error generators. Coherent errors can often be minimized by re-calibrating control (according to $\vec{h}$), or decoupling from nearby systems acting on the 'main'-system's Hamiltonian.

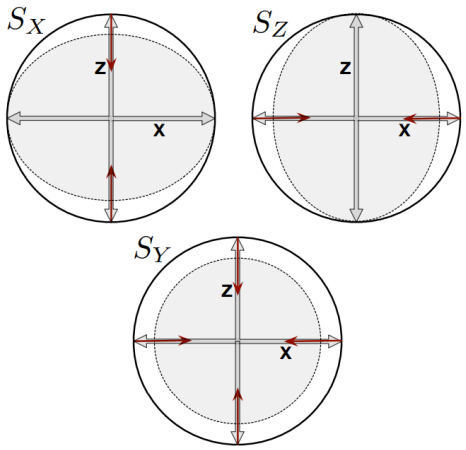


Figure 2.6: How the 1-qubit Pauli-stochastic error generators act on the $X-Z$ plane, forming ellipsoids by generating dephasing. Figure is taken from Blume-Kohout *et al.* [22]

Stochastic generators

Stochastic generators form the second class of error generators and are convex mixtures (not linear combinations!) of Hamiltonian generators, spanning $\mathbb{S}$ and $\mathbb{C}$. Without proof, we state that *any* convex mixture of unitary evolutions is generated by (1) the Hamiltonian generators and (2) some linear combination of d^2-1 *stochastic Pauli generators* (Eq.(2.21)) and $(d^2-1)(d^2-2)/2$ *Pauli correlation generators* (Eq.(2.22)) [22]. S_P terms shrink the Bloch sphere to an ellipsoid, generating dephasing towards the P axis (Fig.2.6). Dephasing along other axes than the X, Y and Z axis of the Bloch sphere is possible

by *combinations* of the S_P and the Pauli correlation $C_{P,Q}$ generators. $C_{P,Q}$ are never physically valid, being non-CP maps (they can 'inflate' the Bloch sphere, Fig.2.5.b). However, together with S_P terms they generate valid dephasing, by *modifying* S' error rates (Fig.2.7).

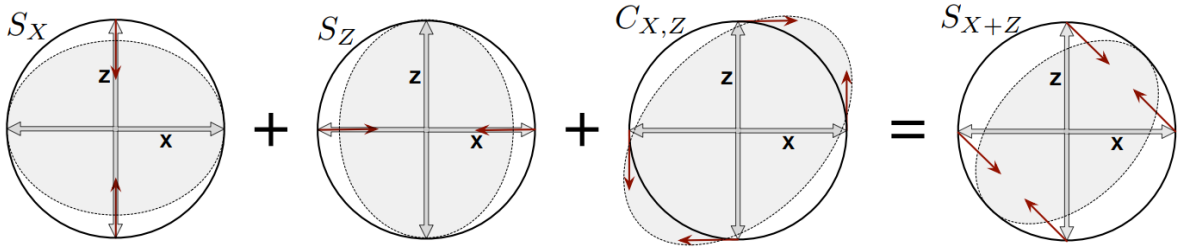


Figure 2.7: Correlation errors, like $C_{X,Z}$ are never physically valid on their own, since they make the Bloch sphere "longer" in some places. However, together with Stochastic error generators, correlation errors are physically valid (most right Bloch sphere). Figure is taken from Blume-Kohout *et al.* [22]

Stochastic errors are not only produced by mixtures of different unitary rotations (Hamiltonian generators), but can also be the consequence of *minimally disturbing measurements* whose outcomes are ignored. An example is that when a qubit's environment measures it (weakly) in the Z basis, the qubit will dephase a little (a stochastic Z error).

Active generators

Active, or antisymmetric, generators (Eq.(2.23)) are "everything that is left"; spanning the $(d^2-1)(d^2-2)/2$-dimensional complement $\mathbb{A}$ of $\mathbb{H} \oplus \mathbb{S} \oplus \mathbb{C}$. $A_{P,Q}$ generate affine shifts of the Bloch sphere (Fig.2.5.c), moving the whole Bloch sphere into the R-direction (R being the Pauli different from P and Q). Like the correlation generators, the A-type generators are not physically valid on their own. In combination with stochastic S-type generators, they generate non-unital decoherence processes. The most widely known example of this is T_1 decay, when a qubit suddenly, irreversibly, decays from a excited state into a lower state (amplitude damping) due to coupling with a large cold environment. Ampltitude damping can be seen as a combination of stochastic Pauli errors and a non-CP affine shift, shrinking and shifting the Bloch sphere (Fig.2.8). Apart form T_1 decay, active errors are not as well understood as Hamiltonian, or stochastic errors.

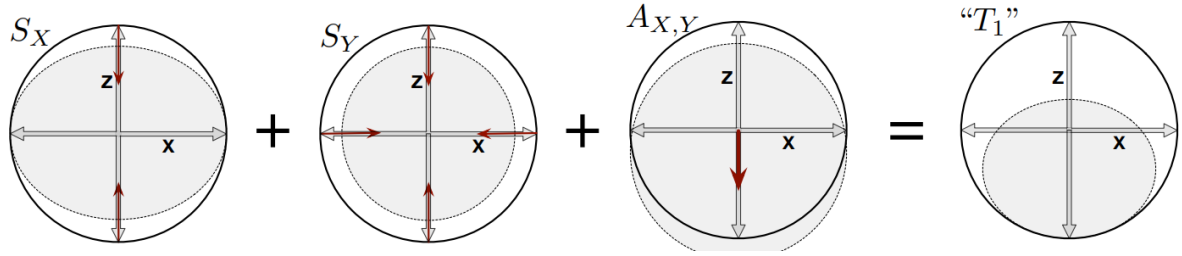


Figure 2.8: T_1 decay is a combination of stochastic and active errors. Figure is taken from Blume-Kohout *et al.* [22]

2.3.3. Gauge freedom

In the final section of this chapter, the downpoint of gate set tomography will be described shortly, which is *gauge freedom*. The representation of a gate set given in definition 2.4 is actually an *over-specification* of the physical gate set, since the preparation, gate and measurement operations that a gate set describes are actually *relational and inter-dependent*. This gives rise to gauge freedom: two (or more) descriptions of the same system, differing by a *gauge transformation*, result in the same observable physics. More precisely, it is said that at every $\vec{\theta} \in \mathcal{P}$ there are directions along which none of the predicted probabilities change. Variations along these directions, *gauge directions*, do not affect the goodness-of-fit of the gate set model to the gate (as observed probabilities do not change). However, gauge freedom makes it difficult to compare gate sets with each other, as gauge transformations do not preserve distances in most metrics used in quantum information.

Definition 2.7. A gauge freedom exists, if there is a transformation $T_M : \mathcal{M} \to \mathcal{M}$ that can be applied to $\vec{\theta} \in \mathcal{P}$ that changes $\vec{\theta}$, but not any observable probability that can be computed from the corresponding gate set $W(\vec{\theta})$.

These *matrix-space gauge transformations* give rise to gauge directions and act on gate sets as follows:

$$\begin{aligned} \langle\langle E_i^{(m)}| &\to \langle\langle E_i^{(m)}|M^{-1} \\ G_i &\to M G_i M^{-1} \\ |\rho^{(i)}\rangle\rangle &\to M|\rho^{(i)}\rangle\rangle \end{aligned} \tag{2.24}$$

Here, $M \in \mathrm{GL}(d^2)$ is any invertible superoperator, an element of the "gauge group" $\mathrm{GL}(d^2)$. Indeed, we see that nothing observable changes: $\Pr(E_i^{(m)}) = \langle\langle E_i^{(m)}|M^{-1}MG_iM^{-1}M|\rho^{(i)}\rangle\rangle = \langle\langle E_i^{(m)}|G_i|\rho^{(i)}\rangle\rangle$. Concluding, a quantum information processor does not have one *true* gate set. However, `pyGSTi` chooses the "best" representation of the gate set, when conducting gate set tomography. Further details on gauge freedom are beyond the scope of this thesis, but can be found in Nielsen *et al.* [8].

3

Numerical Optimization Theory

A popular algorithm for solving non-linear optimization problems is the Levenberg-Marquardt algorithm (LM). It was proposed by Levenberg in 1944, when it apparently was not that popular, as it had to be *rediscovered* by Marquardt in 1963. It is currently used in gate set tomography and will therefore be described in this chapter. First, two concepts will be explained: linear least-squares problems, as these appear in LM, and the Gauss-Newton method, as LM flows from this method. Then, it is explained how LM iterates from one point to the next: by looking within a trust region. Using this approach instead of the traditional line search method is what distinguishes LM from the Gauss-Newton method. Finally, LM will be presented. Nocedal and Wright [23] is followed throughout this chapter.

3.1. Linear least-squares problems and the Gauss-Newton method

In this section, firstly least-squares optimization problems are formulated. Then, a couple of ingredients used in the Gauss-Newton method will be discussed, such as the gradient and Hessian. Thereafter, linear least-squares problems will be discussed. To conclude, the Gauss-Newton method will be considered.

3.1.1. Least-squares problems

In least-squares problems, the objective function f looks as follows:

$$f(x) = \frac{1}{2}\sum_{j=1}^{m} r_j^2(x). \tag{3.1}$$

Here, every $r_j : \mathbb{R}^n \to \mathbb{R}$ is smooth function, which are called *residuals*. Furthermore, we assume $m \geq n$. The *residual vector* is defined:

Definition 3.1. The residual vector $r : \mathbb{R}^n \to \mathbb{R}^m$ is given by: $r(x) = (r_1(x), \ldots, r_m(x))^T$.

f in terms of the residual vector is then: $f(x) = \frac{1}{2}\left\|r(x)\right\|_2^2$. The gradient and Hessian of f become, where J is r's Jacobian (Def.A.1):

$$\begin{aligned}\nabla f(x) &= \sum_{j=1}^{m} r_j(x)\nabla r_j(x) = r_1(x)\begin{bmatrix}\frac{\partial r_1}{\partial x_1}\\ \vdots \\ \frac{\partial r_1}{\partial x_n}\end{bmatrix} + \cdots + r_m(x)\begin{bmatrix}\frac{\partial r_m}{\partial x_1}\\ \vdots \\ \frac{\partial r_m}{\partial x_n}\end{bmatrix} = \begin{bmatrix}\frac{\partial r_1}{\partial x_1}r_1(x) + & \ldots & + \frac{\partial r_m}{\partial x_1}r_m(x)\\ \vdots & \ddots & \vdots \\ \frac{\partial r_1}{\partial x_n}r_1(x) + & \ldots & + \frac{\partial r_m}{\partial x_n}r_m(x)\end{bmatrix}\\ &= \begin{bmatrix}\frac{\partial r_1}{\partial x_1} & \cdots & \frac{\partial r_m}{\partial x_1}\\ \vdots & \ddots & \vdots \\ \frac{\partial r_1}{\partial x_n} & \cdots & \frac{\partial r_m}{\partial x_n}\end{bmatrix}\begin{bmatrix}r_1(x)\\ \vdots \\ r_m(x)\end{bmatrix} = J(x)^T r(x)\end{aligned} \tag{3.2}$$

$$\nabla^2 f(x) = J(x)^T J(x) + \sum_{j=1}^{m} r_j(x)\nabla^2 r_j(x) \tag{3.3}$$

Note that when the gradient of f has been computed, part of the Hessian ($J(x)^T J(x)$) comes for (almost) free. From now on, $\|\cdot\|$ will be used to denote the Euclidian norm $\|\cdot\|_2$.

3.1.2. Linear least-squares problems

In this subsection, linear least-squares (LS) problems will be discussed, as they form subproblems in the considered algorithms. In linear LS problems, the residual vector is of special form: $r(x) = Jx - y$, for some matrix J and vector y (both independent of x). The objective function f becomes:

$$f(x) = \frac{1}{2}\left\|Jx - y\right\|^2, \text{ where } y = r(0). \tag{3.4}$$

The gradient and Hessian of f are:

$$\nabla f(x) = J^T(Jx - y) \text{ and } \nabla^2 f(x) = J^T J. \tag{3.5}$$

The second term in Eq.(3.3) disappears as taking second derivatives in linear problems gives 0. To find the minimizer of f, we first note that the $f(x)$ in Eq.(3.4) is convex. Recall the definition of a convex function [24]:

Definition 3.2. Let $f(x)$ be a real-valued function defined over points $x \in D \subseteq \mathbb{R}^n$, where D is either $\mathbb{R}^n$ or a convex subset of $\mathbb{R}^n$. Then, $f(x)$ is a *convex* function if and only if for any points x^1 and x^2 in D and $0 \leq \alpha \leq 1$, we have:

$$f(\alpha x^1 + (1-\alpha)x^2) \leq \alpha f(x^1) + (1-\alpha) f(x^2) \tag{3.6}$$

Now let $x^1, x^2 \in D$ and $\alpha \in [0,1]$. Then,

$$\begin{aligned} f(\alpha x^1 + (1-\alpha)x^2) &= \frac{1}{2}\left\|J(\alpha x^1 + (1-\alpha)x^2) - y\right\|^2 = \frac{1}{2}\left\|J\alpha x^1 + J(1-\alpha)x^2 - y + \alpha y - \alpha y\right\|^2 \\ &= \frac{1}{2}\left\|\alpha(Jx^1 - y) + (1-\alpha)(Jx^2 - y)\right\|^2 \leq \frac{1}{2}\left\|\alpha(Jx^1 - y)\right\|^2 + \frac{1}{2}\left\|(1-\alpha)(Jx^2 - y)\right\|^2 \\ &= \alpha^2\frac{1}{2}\left\|Jx^1 - y\right\|^2 + (1-\alpha)^2\frac{1}{2}\left\|Jx^2 - y\right\|^2 \leq \alpha f(x^1) + (1-\alpha) f(x^2), \end{aligned}$$

where the triangle inequality and the fact that for $\alpha \in [0,1]$, $\alpha^2 \leq \alpha, (1-\alpha)^2 \leq (1-\alpha)$, were used. Furthermore, recall that any local minimizer x^* of a convex function $f(x)$ ($\nabla f(x^*) = 0$) defined on a convex subset C of $\mathbb{R}^n$ is also a global minimizer. Finding x^* is thus solving $\nabla f(x^*) = 0$:

$$0 = J^T(Jx^* - y) \leftrightarrow J^T J x^* = J^T y. \tag{3.7}$$

These equations are called the *normal equations* for the linear LS objective function. The normal equations will come back in the discussion of the Gauss-Newton method.

3.1.3. The Gauss-Newton method

Recall the (standard) Newton method, where the minimizer is found by solving

$$\nabla^2 f(x_k)(x_{k+1} - x_k) = \nabla^2 f(x_k) p = -\nabla f(x_k) \text{ for } k \geq 0. \tag{3.8}$$

The Gauss-Newton method modifies the Newton method, by approximating the Hessian as $\nabla^2 f_k \approx J^T J$ (disregarding the second sum in Eq.(3.3)). This can be done because oftentimes either the $\nabla^2 r_j(x)$ or the $r_j(x)$ are relatively small, such that the first term of Eq.(3.3) is more important. Plugging $\nabla f_k = J_k^T r_k$ (Eq.(3.2)) and $\nabla^2 f_k = J^T J$ into the above equation gives the Gauss-Newton problem:

$$J_k^T J_k p_k^{\text{GN}} = -J_k^T r_k. \tag{3.9}$$

Observe now that this is exactly of the form of the normal equations in Eq.(3.7). p_k^{GN} is thus the solution of the linear LS problem

$$\min_p \frac{1}{2}\left\|J_k p + r_k\right\|^2. \tag{3.10}$$

Thus, before the minimizer of the objective function f is found, the above *subproblem* has to be solved, i.e. computing the best optimization direction. To this end, the Levenberg-Marquardt method uses a different strategy than the Gauss-Newton method: the trust region approach versus line search, respectively. Both methods are discussed in the next section.

3.2. Line search versus trust region approach

Iterative optimization methods try to find the minimizer of some objective function $f(x)$ by producing a sequence $\{x_k\}$ which terminates if it is not possible to come closer to the minimizer, or if it is close enough to the optimal solution. In deciding *how* to move from the current point x_k to the better, next point x_{k+1}, often one of two strategies is used: *line search*, or the *trust region*. The (Gauss-) Newton method and the method of Steepest Descent are the most widely known line search methods. They choose a direction p_k and search along it to find the next point x_{k+1}, i.e. solve

$$\min_{\alpha>0} f(x_k + \alpha p_k). \tag{3.11}$$

The Levenberg-Marquardt algorithm considered here, is of the second strategy type. Trust region algorithms use a *model function* m_k, an estimate for the objective function f in an area close to the current point x_k; the trust region. The model m_k is usually defined as the following quadratic function:

$$m_k(x_k + p) = f_k + p^T \nabla f_k + \frac{1}{2} p^T B_k p. \tag{3.12}$$

Here, $f_k, \nabla f_k$ and B_k are respectively a scalar, vector and matrix. B_k is the Hessian of f_k ($\nabla^2 f_k$) or an approximation of it. The difference between f and the model function is $O\left(\|p\|^2\right)$, which is small when p is small. Unlike in line search methods, where the direction p_k to go along to the next point is calculated from the current point and f, in trust region algorithms a sub-optimization problem has to be solved to find p_k:

$$\min_p m_k(x_k + p), \text{ where } x_k + p \text{ lies inside the trust region.} \tag{3.13}$$

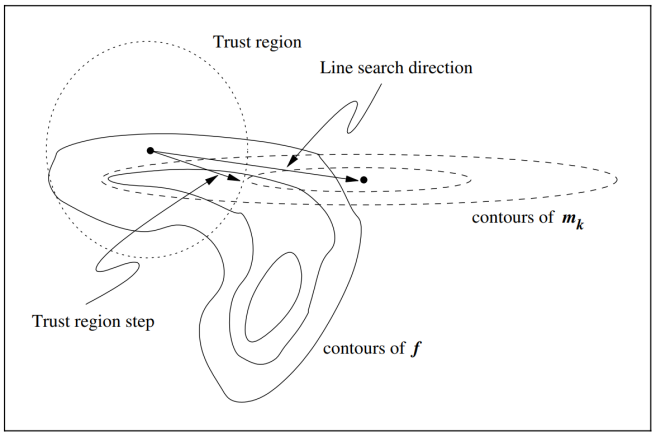


Figure 3.1: Figure illustrating the line search and trust region approach. The line search method goes along the minimizer of m_k, reducing f only a small amount (as the next point is not that much closer to f's minimizer than the current point). The trust region approach however, looks inside the dotted circle for m_k's minimizer, resulting in a greater reduction of the objective value. Figure is taken from Nocedal and Wright [23].

For example, if the found p_k does not result in a much better, so smaller, f, the trust region is probably too large, as far from x_k the model m_k may not be a good approximation of f. Accordingly, the trust region is shrinked and Eq.(3.13) is solved again. This reasoning is formalized by the algorithm below (Alg.1). Here, the ratio between the *actual reduction* (numerator) and the *predicted reduction* (denominator) is used to evaluate how well the function and its model match:

$$\rho_k = \frac{f(x_k) - f(x + p_k)}{m_k(x_k) - m_k(x_k + p_k)}. \tag{3.14}$$

The predicted reduction will always be nonnegative: either $m_k(x_k + p_k) < m_k(x_k)$ (we are minimizing), or $p_k = 0$ (trust region is too large to find a smaller objective function value). There are three distinct cases:

1. ρ_k is negative (or zero). Then $f(x_k + p_k) > f(x_k)$ (or $f(x_k + p_k) = f(x_k)$), which is rejected, as we are minimizing. Accordingly, the trust region is shrinked.

2. ρ_k is close to one. The model function and objective function agree well in the current trust-region, so in the next step it can be expanded.
3. $0 < \rho_k << 1$. The trust region is not altered.

$\hat{\Delta}$ is a bound on the step lengths in the following algorithm.

Algorithm 1 Trust region algorithm.

Given $\hat{\Delta} > 0, \Delta_0 \in (0, \hat{\Delta})$, and $\eta \in [0, \frac{1}{4})$:
for $k = 0, 1, 2, \ldots$ **do**
  Obtain p_k by (approximately solving) Eq.(3.13);
  Evaluate ρ_k in Eq.(3.14);
  if $\rho_k < \frac{1}{4}$ **then**
    $\Delta_{k+1} = \frac{1}{4}\Delta_k$
  else
    if $\rho_k > \frac{3}{4}$ and $\|p_k\| = \Delta_k$ **then**
      $\Delta_{k+1} = \min(2\Delta_k, \hat{\Delta})$
    else
      $\Delta_{k+1} = \Delta_k$;
    end if
  end if
  if $\rho_k > \eta$ **then**
    $x_{k+1} = x_k + p_k$
  else
    $x_{k+1} = x_k$;
  end if
end for

To summarize, the difference between line search and trust region methods, is as follows. Line search methods first fix p_k and then compute the distance to go along it. This is contrary to trust region algorithms, where first the maximum distance (the trust region radius) to the next point is set, and after this the direction and step are calculated. If the next value of the objective function is not satisfactory, the trust region radius is reduced and the subproblem of direction and step is solved again. Now, it can finally be explained how the Levenberg-Marquardt method solves the subproblem (Eq.(3.13)).

3.3. The Levenberg-Marquardt method

The Levenberg-Marquardt algorithm has the same subproblem (Eq.(3.10)) as the Gauss-Newton method. However, LM solves it differently by employing the trust-region method instead of line search. The subproblem (now dropping the iteration counter k) and its solution are presented in the following lemma (illustrated in Fig.3.2):

Lemma 3.1. The vector p^{LM} is a solution of the trust-region subproblem

$$\min_p \|Jp + r\|^2, \text{ subject to } \|p\| \leq \Delta,$$

if and only if p^{LM} is feasible and there is a scalar $\lambda \geq 0$ such that

$$(J^T J + \lambda I)p^{\mathrm{LM}} = -J^T r, \tag{3.15}$$

$$\lambda(\Delta - \|p^{\mathrm{LM}}\|) = 0. \tag{3.16}$$

The proof can be found in Nocedal and Wright [23]. Suppose the Gauss-Newton method gives p^{GN} as the solution of the subproblem (by applying some linear least-squares algorithm to Eq.(3.10)). The lemma gives then two options for p^{LM}:

1. If $\|p^{\mathrm{GN}}\| < \Delta$, then $p^{\mathrm{LM}} = p^{\mathrm{GN}}$ (Eq.(3.15) and Eq.(3.9) coincide as $\lambda = 0$)

2. Otherwise, $\exists \lambda > 0$ ($p^{\text{LM}} = \Delta$), such that p^{LM} solves Eq.(3.15).

If 1 is the case, we are done. Otherwise, Eq.(3.15) needs to be solved. Note that Eq.(3.15) are the normal equations for the following linear LS problem:

$$\min_p \frac{1}{2} \left\| \begin{bmatrix} J \\ \sqrt{\lambda} I \end{bmatrix} p^{\text{LM}} + \begin{bmatrix} r \\ 0 \end{bmatrix} \right\|^2 . \tag{3.17}$$

Thus, before solving the subproblem (of p^{LM}), another problem (sub-subproblem) has to be solved: finding λ. Fortunately, this is not so complicated and can be done with the rootfinding algorithm. What rests then, is to solve Eq.(3.17). Details on the rootfinding algorithm and how the Levenberg-Marquardt method can be implemented to solve Eq.(3.17) can be found in Nocedal and Wright [23].

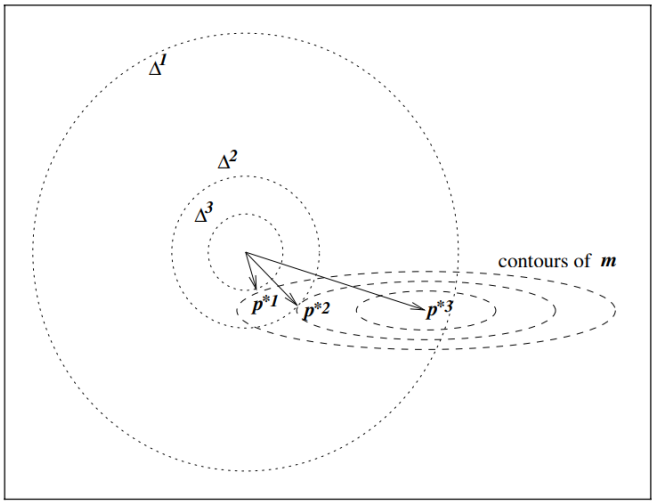


Figure 3.2: Figure illustrating Lemma 3.1. p^{*i} is the solution corresponding to radius Δ^i. If the solution lies strictly inside the trust region (for $i = 1$), then $\lambda = 0$ and $J^T J p^* = -J^T r$. Otherwise (for $i = 2, 3$: $\|p^*\| = \Delta$), $\lambda > 0$ and p^* is collinear with the negative gradient of the model function m and normal to its contours. This is seen from noting that: $\lambda p^* = -J^T J p^* - J^T r = -\nabla m(p^*)$. Figure is taken from Nocedal and Wright [23].

4

Gate Set Tomography

Quantum characterization, verification and validation (QCVV) protocols probe how qubits, logic operations, and entire quantum processors behave. The many different protocols share their essential structure. Firstly, data is collected from experiments, described by quantum circuits performed on the processor. Then, the data is analysed to learn something about (an aspect of) the researched processor's behaviour. In this thesis, one such protocol is used: gate set tomography (GST). The term "gate set" has been described shortly in Chapter 2. "Tomography" is best defined by a group at Sandia National Labs, which has researched GST since 2012: *""Tomography" means the reconstruction of a comprehensive model (of something) from many partial cross-sections or slices, each of which provides only a limited view that may be useless by itself. Tomographic techniques are distinguished by their aspiration to construct a comprehensive model for a system or component, by fitting that model to the data from many independent experiments"* [8]. In this chapter, the whole process behind gate set tomography will be explained. "Traditional" tomographies will be presented, which are merged to realize GST. First, `pyGSTi` is introduced, a software for implementing QCVV protocols.

"`pyGSTi` (Python Gate Set Tomography Implementation) *is an open-source software for modeling and characterizing noisy quantum information processors (QIPs), i.e., systems of one or more qubits"* [25], and has been developed by amongst others the group at Sandia National Labs mentioned above. In this thesis, `pyGSTi` was used to perform gate set tomography. So, some parts of this chapter have been explained through the eyes of `pyGSTi`'s implementation of GST. It will be mentioned explicitly when that is the case. The enthusiastic reader is referred to [25] and Nielsen *et al* [26] for respectively downloading `pyGSTi` and more information on its mechanism.

4.1. An introduction to QCVV: state, process and measurement tomography

In this section, standard state, process and measurement tomography will be presented. To this end, we follow Nielsen *et al* [8]. As mentioned above, GST can be seen as a unification of the three. So, understanding these protocols is one of the first steps toward understanding GST. First, some handy definitions are introduced.

Definition 4.1. A set $\{|\rho\rangle\rangle\}$ of matrices $|\rho\rangle\rangle \in \mathcal{B}(\mathcal{H})$ is informationally complete (IC) if and only if it spans the vector space $\mathcal{B}(\mathcal{H})$.

For example, the set of effects $\{E_i^{(m)}\}$ is IC if and only if it spans the entire set of effects, forming a complete dual basis for states. This also means that the probabilities $p_i^{(m)}(\rho) = \mathrm{Tr}\left[\rho E_i^{(m)}\right]$ *uniquely* identify ρ, meaning there $\nexists \rho'$, $\rho \neq \rho'$ consistent with the measurements.

Furthermore, the term "fiducial" will be used quite often; it means "accepted as a fixed basis of reference". The meaning of this will become clear from the context.

4.1.1. Quantum state tomography

Quantum state tomography is a procedure to experimentally determine the unknown quantum states of a system (Fig.4.1). "Standard" tomographies use two ingredients: experiments and linear algebra ("linear inversion"). Furthermore, they assume the existence of a known informationally complete set. In state tomography, this is the set of effects, $\{\langle\langle E_i^{(m)}|\}$. Here, $m = 1 \dots M$ with M the number of fiducial measurements and $i = 1 \dots N^{(m)}$ are the outcomes of the mth measurement.

First, the experimental part is described. Here, all the effects are listed as $\{E_j : j = 1, ..., N_{f1}\}$, where N_{f1} ($= \sum_m N^{(m)}$) is the total number of distinct measurement outcomes for *all* measurements performed. Then, every E_j is applied to the unknown ρ many times, to estimate the probability of the measurement outcome. Copies of the unknown ρ can be made as this is just "pressing a button" on the processor, i.e. the entries of the states do not have to be known to create the states.

Now, the "theoretical" part. The true probabilities are given by Born's rule: $p_j = \mathrm{Tr}\left[\rho E_j\right]$. Rewriting this with the Hilbert-Schmidt space inner product (Ch.2.1) gives $p_j = \langle\langle E_j|\rho\rangle\rangle$. Then,

$$\vec{p}_j = A|\rho\rangle\rangle, \tag{4.1}$$

where A is the known $N_{f1} \times d^2$ matrix of effect vectors:

$$A = \begin{pmatrix} \langle\langle E_1| \\ \vdots \\ \langle\langle E_{N_{f1}}| \end{pmatrix}. \tag{4.2}$$

Lastly, ρ is reconstructed via linear inversion, using the experimentally acquired estimates for the probabilities $\hat{\vec{p}}$. Two cases are considered, namely:

1. A is a square matrix and thus has an inverse;
2. The $\{\langle\langle E_i^{(m)}|\}$ form an overcomplete basis, i.e. $N_{f1} > d^2$ and A is not square.

In the first case, inverting Eq.(4.1) gives: $|\rho\rangle\rangle = A^{-1}\hat{\vec{p}}$. In the second case, a pseudo-inverse (definition A.2) is used, resulting in: $|\rho\rangle\rangle = (A^T A)^{-1} A^T \hat{\vec{p}}$.

4.1.2. Quantum process tomography

Similarly to quantum state tomography, quantum process tomography tries to reconstruct gates G (e.g. *quantum processes*) from experimentally acquired data (Fig.4.1). This requires an additional assumption, namely that a list $\{|\rho_i\rangle\rangle\}$ of N_{f2} informationally complete fiducial quantum states is available. The matrix of the column vectors of these states is defined:

$$B = \left(|\rho_1\rangle\rangle \dots |\rho_{N_{f2}}\rangle\rangle\right). \tag{4.3}$$

Like above, running experiments (of the form $E_j G[\rho_i]$; preparing ρ_i, applying G and measuring with possible outcomes $\{E_j\}$) allows the estimation of the probabilities $P_{j,i}$, where $P_{j,i} = \mathrm{Tr}(E_j G[\rho_i]) = \langle\langle E_j|G|\rho_i\rangle\rangle$. The estimated probabilities are combined into a $N_{f1} \times N_{f2}$ matrix P, such that $P = AGB$.

Like previously, if $N_{f1} = N_{f2} = d^2$, then G is simply given by $G = A^{-1}PB^{-1}$. Otherwise, if $N_{f1} > d^2$ and/or $N_{f2} > d^2$, the pseudo-inverse is used to obtain $G = (A^T A)^{-1} A^T P B^T (BB^T)^{-1}$.

4.1.3. Quantum measurement tomography

Quantum measurement tomography tries to reconstruct the POVM effects for an unknown measurement and directly follows from generalising state tomography. If $\{E_j\}$ is the unknown POVM, a vector of estimated probabilities is defined of each unknown effect, with $\{\rho_i\}$ the known fiducial IC states: $[p_j]_i = \mathrm{Tr}[\rho_i E_j] = \langle\langle E_j|\rho_i\rangle\rangle$. Using B (Eq.(4.3)) we acquire: $\vec{p}_j^{\,T} = \langle\langle E_j|B$. Finally, linear inversion delivers: $\langle\langle E_j| = \vec{p}_j^{\,T} B^{-1}$, where B^{-1} is the inverse or pseudo-inverse of B, as explained earlier.

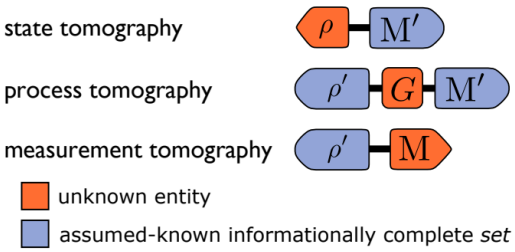


Figure 4.1: Structures of the circuits used in state, process and measurement tomography. In each protocol an unknown entity (orange) is deduced by using observed probabilities and linear-inversion, under the assumption that some informationally complete set (blue) is known. ρ's are states, M's are measurement effects, and G is a gate. Figure is taken from Nielsen *et al.* [8]

4.1.4. The pre-calibration problem in standard tomography

State, process and measurement tomography all work under the assumption that some informationally complete set is known exactly (Fig.4.1), i.e. *calibrated*. However, these assumptions are never satisfied in practice, since this calibration leads to an infinite loop of self-referentiality. For state tomography, known POVM's are needed, which can be obtained through measurement tomography, for which the results of state tomography are required. Likewise, fiducial states and measurement are necessary for process tomography. As most processors possess only a few native states and measurements, these are typically produced using quantum logic gates, that are yet to be estimated!

Gate set tomography overcomes this pre-calibration problem by not assuming *anything* about the processor's states, measurements and gates. GST treats the processor as a "black box" (Fig.4.2). That is, circuits can be run by pressing the processor's buttons, but how the buttons look exactly (in the mathematical sense), the *gate set*, is assumed unknown, to be estimated by gate set tomography. In the next section, linear GST, GST's "first step", is described.

Figure 4.2: Gate set tomography treats the studied quantum processor as a black box. Circuits can be run by pressing the processor's buttons, but how the buttons look exactly (in the mathematical sense) is unknown. Figure taken from Blume-Kohout *et al* [27].

4.2. Lineair gate set tomography

In this section, linear gate set tomography (LGST) is described as a starting point for gate set tomography. Recall the gate set $\mathcal{G}$ (Def.2.4). LGST constructs its low-precision estimate, by fitting it to data from a particular set of short circuits. As LGST does not assume anything about the gate set, it does not suffer from the above calibration problem. However, this unfortunately leads gauge freedom (Ch.2.7), which makes the produced representation of the gate set *not unique*. Another downside of LGST is that it has the same accuracy/effort scaling as process tomography.

4.2.1. The linear gate set tomography algorithm

LGST starts with the assumption that unknown IC sets of fiducial states $\{|\rho_j'\rangle\rangle\}$ and measurement effects $\{\langle\langle E_i'|\}$ can be created (by "pressing" the right buttons on the processor). The sets of states and measurements have respective dimensions $N_{f1} = N_{f2} = d^2$, and will form an unknown reference frame. Using larger sets does have practical advantages, but these are outside the scope of this thesis and the interested reader is therefore referred to Nielsen *et al* [8]. Furthermore, finite sample error (i.e.

$\hat{p} = p \pm O(1/\sqrt{N})$ in N trials) in estimated probabilities and the IC sets will be ignored in this section, so $p = \hat{p}$.

For each gate G_k to be reconstructed, a matrix P_k is defined as $P_k = AG_kB$, where $[P_k]_{i,j} = \langle\langle E'_i|G_k|\rho'_j\rangle\rangle$, similarly to process tomography. Now however, since the elements of A and B are not known, linear inversion cannot be used to solve for G_k. Instead, probabilities corresponding to the null operation (i.e. no gate), the *Gram matrix*, are measured:

Definition 4.2. The Gram matrix for the fiducial states and effects is given by $\tilde{\mathbb{1}} = AB$.

The assumption that A and B are square and invertible, which follows from informational completeness, allows us to invert the Gram matrix and solve for G_k. Multiplying both sides of $P_k = AG_kB$ with $\tilde{\mathbb{1}}^{-1}$ gives:

$$\tilde{\mathbb{1}}^{-1}P_k = B^{-1}A^{-1}AG_kB = B^{-1}G_kB$$

$$G_k = B\tilde{\mathbb{1}}^{-1}P_kB^{-1}.$$

The above can be done for every G_k. Furthermore, the ("true") native states $\rho^{(l)}$ and native measurement effects $\{E_l^{(m)}\}$ need to be reconstructed. This is achieved with the vectors of observable probabilities $\left[\vec{R}^{(l)}\right]_j = \langle\langle E'_j|\rho^{(l)}\rangle\rangle$ and $\left[\vec{Q}_l^{(m)}\right]_j = \langle\langle E_l^{(m)}|\rho'_j\rangle\rangle$. Measuring these is analogous to carrying out state tomography and measurement tomography, respectively:

$$\vec{R}^{(l)} = A|\rho^{(l)}\rangle\rangle,$$
$$\vec{Q}^{(m)T} = \langle\langle E_l^{(m)}|B.$$

Inverting above equations using the Gram matrix provides the whole gate set *up to a gauge*:

$$G_k = B\tilde{\mathbb{1}}^{-1}P_kB^{-1}, \tag{4.4}$$
$$|\rho^{(l)}\rangle\rangle = B\tilde{\mathbb{1}}^{-1}\vec{R}^{(l)}, \tag{4.5}$$
$$\langle\langle E_l^{(m)}| = \vec{Q}_l^{(m)T}B^{-1}. \tag{4.6}$$

Here, B (Eq.(4.3)) is the gauge transformation, on which the gate set representation relies. So, choosing another IC set of states leads to a different gate set, but not to different physics, as observed probabilities stay the same. The best *a priori* choice for B corresponds to the tomographer's best *a priori* guess for the fiducial states; B implicitly defines the expected gauge. Oftentimes, *a posteriori* gauge-fixing is required to reliably compute gauge-dependent quantifies. Unfortunately, that is beyond the scope of this thesis, so I refer to Nielsen *et al.* [8] for more information.

4.2.2. Creating fiducial vectors

In the section above, fiducial informationally complete sets of states $\{|\rho'_j\rangle\rangle\}$ and measurements $\{\langle\langle E'_i|\}$ were used. However, most quantum processors are able to create only *one* native state preparation, and *one* measurement. To achieve IC sets, to create more states and measurements, the unknown gates are applied to the native state and measurement. That is, *preparation and measurement fiducial QI/QO circuits*, $\{F_k\}$ and $\{H_k\}$ respectively, are constructed:

$$\langle\langle E'_i| = \langle\langle E_{t(i)}^{(m(i))}|\tau(H_{h(i)}), \tag{4.7}$$
$$|\rho'_j\rangle\rangle = \tau(F_{f(j)})|\rho^{(r(j))}\rangle\rangle. \tag{4.8}$$

Recall that τ denotes a QI/QO circuits (Ch.2.2.2). This way, every item in the IC set, indexed by i or j, is mapped to the native state index ($r(j)$) or native measurement index ($m(i)$) by its corresponding QI/QO circuit (indexed by $h(i)$ or $f(j)$). This section and the previous one are summarized in Fig.4.3.

4.2.3. Optimization procedure

Once again is repeated, that the "probabilities" described above are actually observed *frequencies* and thus estimations for the predicted probabilities. GST tries to locate the gate set that fits the observed

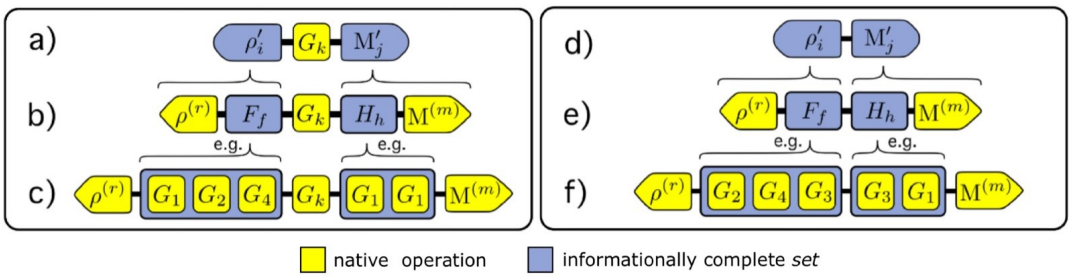


Figure 4.3: The circuits used in linear gate set tomography to estimate native operations (yellow), whilst assuming the ability to create, but not know, informationally complete sets (blue). Circuits used to estimate gates (**a)**) include informationally complete sets of states and measurements (ρ_i' and M_j'), which themselves are constructed by applying preparation and measurement fiducial circuits (F_f and H_h) to the processor's native state and measurement $\rho^{(r)}$ and $M^{(m)}$**(b)**). **c)** Unknown gates make up F_f and H_h. Analogously to this, in **d) - f)** the circuits for producing the Gram matrix are depicted, which is equivalent to sandwiching the null operation between the states and measurements. Figure taken from Nielsen *et al.* [8]

data best. To this end, LGST minimizes the squared difference of observed and predicted probabilities:

$$f(\text{gate set}) = \sum_j (p_i^{(\text{gate set})} - f_j^{(\text{observed})})^2. \tag{4.9}$$

However, using particularly least-squares is not well-motivated. Another downside of this approach is that not all circuits performed in obtaining the LGST data are included in $f(\text{gate set})$, as these depend on the choice of $\tilde{\mathbb{1}}$.

A better approach is to maximize the *likelihood function*: $\mathcal{L}(\text{gate set}) = \Pr(\text{data} \mid \text{gate set})$, for which the sum-of-squared objective function (Eq.(4.9)) is an approximation. This method views gate sets as statistical models that predict circuit probabilities, over which we numerically vary to find the one that maximizes $\mathcal{L}$; *maximum likelihood estimation* (MLE). This way, data from every circuit is used, and the final gate set really does minimize the total squared error. As we will see in the next section, GST employs MLE, whereas LGST serves only as a starting point for GST.

4.3. Long-sequence gate set tomography

Long-sequence gate set tomography (or just GST for short) is realized by modifying two aspects of linear gate set tomography. First, GST has a different experiment design, since it employs *deep circuits*. Secondly, as mentioned above, GST finds the best gate set via maximum likelihood estimation (MLE). This is done in a 2-process. Before explaining these alterations in more detail, an example motivating the use of deep circuits is given.

In linear gate set tomography, each entry of a superoperator G_k is pretty close to being a linear combination of observed probabilities. The accuracy of $\hat{G}_k$ is therefore limited by the finite sample error: $\hat{G}_k = G_k \pm O(1)/\sqrt{N}$, where N is the amount of times each circuit is performed. The use of deep circuits in long-sequence GST forces the outcome probabilities to depend more sensitively on the elements of G_k. For example, in $\Pr = \langle\langle E| G_k G_k G_k G_k |\rho\rangle\rangle$, some entries of G_k are up to four times more sensitive to changes in G_k, as in the circuit where G_k appears only once. Now, it will be explained how GST cleverly incorporates deep circuits in its experiment design.

4.3.1. Experiment design

The term *experiment* will be used to refer to an experiment design, that is a set of quantum circuits to be run, together with the data acquired from running each circuit many times. Every long-sequence GST circuit, specified by an (j, k, p, m) tuple, consists of three consecutive parts (Fig.4.4):

1. A state $|\rho_k'\rangle\rangle$ is prepared by performing a fiducial circuit on a native preparation, like in LGST (Ch.4.2.2).
2. A short circuit, a *germ* g, is repeated p times.
3. A measurement $\{\langle\langle E_i'^{(m)}|\}$ is carried out by performing a fiducial circuit followed by a native POVM measurement, like in LGST (Ch.4.2.2).

In the sandwiched step, g^p is called a *base circuit* and p a *germ power*. Like LGST, GST assumes the ability to create informationally complete sets of states and measurement out of the processor's native states and measurements. This allows performing tomography on $\tau(g_j^p)$ via estimation of probabilities $p = \langle\langle E_i'^{(m)} | \tau(g_j^p) | \rho_k' \rangle\rangle$. Errors that *commute* with $\tau(g)$, can be "separated" from $\tau(g_j^p)$, such that repetition of the germ amplifies them (they get multiplied with each other as well).

To achieve maximum accuracy, every parameter in the gate set that can be amplified, must be amplified. However, this excludes two types of parameters. Firstly, gauge parameters, as these cannot be measured at all (as they are independent of observed probabilities). Secondly, state preparation and measurement (SPAM) operations cannot be amplified, as these appear only once in each circuit. Now, we will dive into the details of steps 1, 2 and 3. It will be explained how `pyGSTi` chooses the right set of germs g_i, germ-powers p and SPAM pairs to amplify all $N_{\mathrm{p}}^{\text{non-gauge}}$ amplifiable parameters.

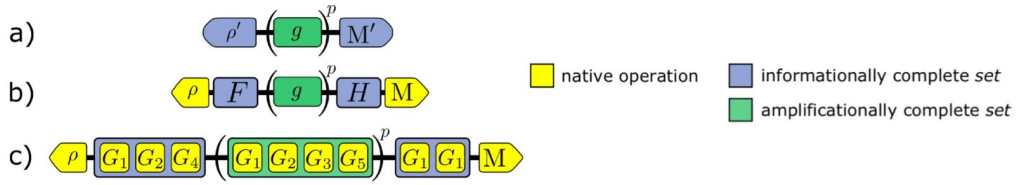


Figure 4.4: The experiment design for long-sequence GST. **a)** A circuit consists of a germ g that is repeated a p amount of times, sandwiched between an effective state preparation ρ' and an effective measurement $M' = \{E_i'\}$. **b)** The states and measurements are produced by applying a *preparation fiducial circuit* F and *measurement fiducial circuit* H from an informationally complete set, respectively after and before a native state ρ or native measurement M. **c)** The fiducials and germ consist of native operations of a gate set. In this case, the preparation fiducial circuit has depth 3, the germ has depth 4 and the measurement fiducial has depth 2. Figure taken from Nielsen *et al.* [8]

Fiducial selection

In this section the selection of informationally complete sets of effective state preparations and effective measurements will be described (selection of $\{H_k\}$ and $\{F_k\}$, Eq.(4.8), Eq.(4.7)); fiducial selection. Theoretically, any set of d^2 random circuits is IC, as it spans $\mathcal{B}(\mathcal{H})$. However, it is preferred to have "maximal" linear independency of the circuits (more uniformly IC sets decrease the prefactor in the accuracy scaling). The spectrum of the Gram matrix quantifies the amount of linear independence.

The Gram matrix will not have d^2 non-zero singular values (Def.A.3) if either $\{|\rho_j'\rangle\rangle\}$ or $\{\langle\langle E_i'\}$ is informationally *incomplete*. The fiducial set is *close* to being linearly dependent if the d^2-th largest magnitude singular value is close to zero. An optimal fiducial set, so whose elements span $\mathcal{B}(H)$ as uniformly as possible, is thus found by maximising the smallest of the top d^2 singular values of the Gram matrix over some candidate fiducial sets. In practice, a candidate set is composed of all the possible circuits of some maximal depth.

To compute the Gram matrix, estimates of the actual gates are needed (as the sets of preparation and measurements are created using gates). If the best-fit gate is close to the actual gate, then produced states and effects will be close to the expected ones, and thus close to being uniformly IC. Otherwise, fiducials are selected again, by using the new estimates of the gates (provided by GST). Thus, it is possible that multiple experiment-generation steps are necessary to find the optimal set of fiducials.

In the next section, where base circuit and germ selection will be discussed, it is assumed that fiducial selection has been successful.

Germ selection

In gate set tomography, errors in the gate set are modelled as small perturbations to the target gate set. A set of germs that amplifies all errors that *can* be amplified, is called *amplificationally complete* (AC). In this section we will derive a notion of completeness for germs, so a necessary and sufficient condition for an AC germ set.

At the beginning of this section, a short example portraying the benefit of deep circuits was given. Now, a similar example will be given, but this time to motivate the use of germs other than the gates themselves. The reason is that single gates amplify *some*, but not *all* parameters of the gate set. Suppose G_x is an exact $\pi/2$ rotation, but now around the wrong axis, i.e. $G_x = e^{-i(\pi/4)(\cos\epsilon\sigma_x + \sin\epsilon\sigma_y)}$, in which σ_x and σ_y are the Pauli operators, and ϵ is some small constant. This "tilt error" cannot be amplified by

raising G_x to some power. Take for example $p = 4$, then $G_x^4 = e^{-i(\pi)(\cos\epsilon\sigma_x + \sin\epsilon\sigma_y)} = \tilde{\mathbb{1}}$ (this equality can be verified by, for example, writing out the matrix exponential Taylor expansion). Other circuits are needed to amplify such errors. Here, G_xG_y suffices, G_y being a perfect $\pi/2$ rotation around the y-axis. G_xG_y is a rotation by $2\pi/3 + \epsilon/\sqrt{3}$, so repeating this germ a p amount of times ensures that ϵ can be estimated up to $1/(p\sqrt{N})$.

Generally, every circuit amplifies some nontrivial linear combination of gate set parameters. In choosing a set of germs, a list of candidate germs is created, i.e. all circuits shorter than some cutoff depth. The Jacobian will be the tool to identify the linear combination of parameters a candidate germ amplifies:

Definition 4.3. If $\vec{\theta} \in \mathcal{P}$ is a vector of gate set parameters, the Jabobian of $\tau(g) : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H})$ is given by:

$$\nabla_g^{(p)} := \frac{1}{p} \left.\frac{\partial[\tau(g)^p]}{\partial\vec{\theta}}\right|_{W(\vec{\theta})=\mathcal{G}}. \tag{4.10}$$

Like in LGST, where the elements of $\tau(g)$ could be accessed because IC sets of fiducials were used, the full Jacobian here is accessible. $\tau(g)$ is $d^2 \times d^2$ and N_p is the amount of parameters in the gate set, so $\nabla_g^{(p)}$ is a $d^4 \times N_\mathrm{p}$ matrix. The linear combinations of model parameters that $\tau(g)$ amplifies are indicated by its d^4 right singular vectors, the eigenvectors of the matrix $\nabla_g^{(p)T}\nabla_g^{(p)}$ [28]. The amplification is quantified by the corresponding singular values. A parameter combination is not amplified at all if it has zero singular value. Thus far, only one germ was considered. A set of N_g germs gives the following $N_\mathrm{e} \times N_\mathrm{p}$ Jacobian, where $N_\mathrm{e} = N_\mathrm{p}^\mathrm{full} = d^4N_\mathrm{G}$:

$$J^{(p)} = \begin{pmatrix} \nabla_{g_1}^{(p)} \\ \vdots \\ \nabla_{g_{N_g}}^{(p)} \end{pmatrix}. \tag{4.11}$$

In `pyGSTi`'s germ selection, the fact that base circuits of length $L >> 1/\eta$, with η the rate of stochastic noise, do not give interesting information is ignored. That is, the gates are assumed to be reversible/unitary ($\tau(g)^{-1} = \tau(g)^\dagger$), which is justified for small η. This assumption makes it possible to define the $p \to \infty$ limit in Eq.(4.11). First, the product rule and $\tau(g)^{-1} = \tau(g)^\dagger$ give:

$$\nabla_g^{(p)} = \frac{1}{p}\sum_{n=0}^{p-1} \tau(g)^n \frac{\partial\tau(g)}{\partial\vec{\theta}} \tau(g)^{p-1-n} \tag{4.12}$$

$$= \left[\frac{1}{p}\sum_{n=0}^{p-1} \tau(g)^n \nabla_g^{(1)} (\tau(g)^\dagger)^n\right] \tau(g)^{-(p-1)}. \tag{4.13}$$

With a result from representation theory, namely Schur's lemma, it is found that taking $p \to \infty$ results in:

$$\lim_{p\to\infty} \nabla_g^{(p)} = \prod_{\tau(g)} \left[\nabla_g^{(1)}\right], \tag{4.14}$$

where $\prod_{\tau(g)}$ is the projection onto the commutant of $\tau(g)$, the subspace of matrices that commute with $\tau(g)$. The critical reader that does not believe this claim without proof, and rightly so, is referred to Nielsen *et al* [8], as unfortunately representation theory is outside the scope of this project. Finally, as promised, amplificational completeness for germs can be defined:

Definition 4.4. The set of germs $\{g_i\}_{i=1}^{N_g}$ is AC if and only if the right singular rank of its Jacobian (the $p \to \infty$ limit of Eq.(4.11)), equals the total number of physically accessible (gauge-invariant) parameters in the gate set $N_\mathrm{p}^\mathrm{nongauge}$.

Thus, `pyGSTi` can build an AC set of germs by adding germs to the germ set until its Jacobian has rank $N_\mathrm{p}^\mathrm{nongauge}$. This subsection is concluded with an example [29]. Consider a set of three single-qubit

trace-preserving gates. A gauge transformation is given by $G_k \to TG_kT^{-1}$, where T is some invertible trace-preserving operator. The TP condition reduces the amount of parameters of an unknown $d^2 \times d^2$ matrix from d^4 ($= d^2d^2$) to $d^2(d^2-1)$. Now, there are 12 gauge parameters ($d = 2$) coming from T, responsible for gauge-variant parameters, and 36 (3×12) parameters coming from the three gates we are trying to estimate. This results in $36 - 12 = 24$ gauge-invariant parameters. So, according to the definition above, it is sufficient to add germs until the Jacobian of the set of germs has rank 24. More rigorous germ-selection is done by adding and removing germs, and only keeping those that lower a certain score function. `pyGSTi` uses the following one:

$$f(\{g_1 \dots g_{N_g}\}) = \frac{\mathrm{Tr}[(J^\dagger J)^{-1}]}{N_g}. \tag{4.15}$$

Base circuit selection

The final step of GST's experiment design is base circuit selection: choosing germ powers p. The goal is to map every parameter $\vec{\theta}$ in the gate set model to a probability that depends on it as $p\theta$, where p is the germ-power. To achieve this, every germ is sandwiched for multiple p's (a motivating example can be found in A.4.1), where $p(g, l) = \lfloor l/|g| \rfloor$. Here, l is the logarithmically-spaced *approximate depth* of a base circuit and $|g|$ is the depth of the germ. For example, the (single) germ $G_xG_yG_y$ has $|g| = 3$. $l = 1, m, m^2, m^3, \dots$ for some $m \in \mathbb{R}$ and Nielsen's group [8] has empirically found $m = 2$ to work reliably. For $m = 2$ we define $L := l$. So, the germ $G_xG_yG_yG_xG_yG_y$ is repeated once for $L = 8$, and twice for $L = 16$. p is made germ-dependent, because a base circuit's *overall* depth is more relevant than its power. Consider g_1 with depth 1, and g_2 with depth 3, then g_1^9 and g_2^3 both have overall depth 9. This process is illustrated in Fig. 4.5(2). In this thesis, $L_{\max} = 16$. The reasoning behind this choice can be found in A.4.2.

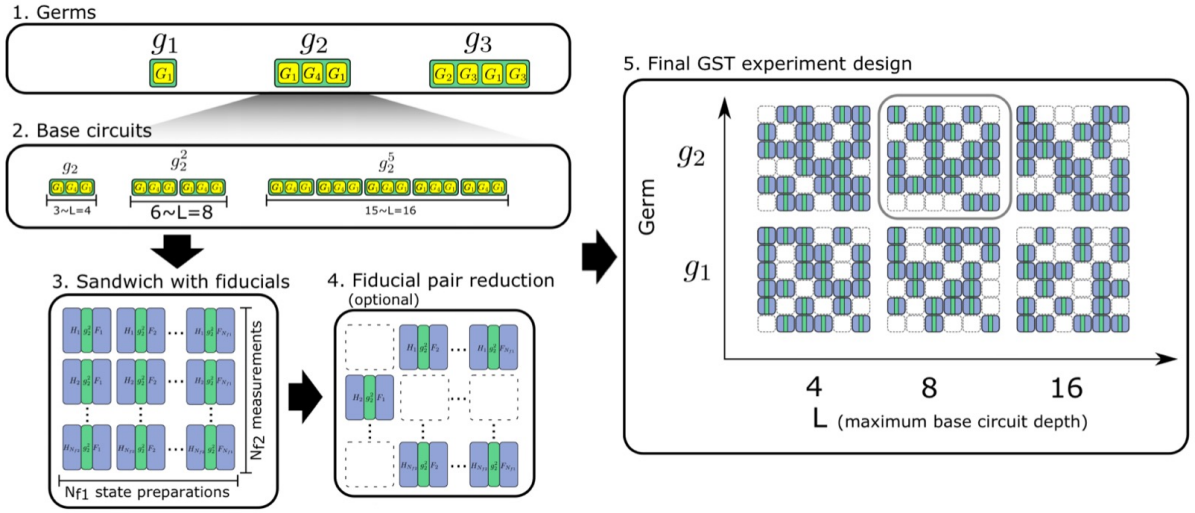


Figure 4.5: The road-map to designing a GST experiment. **Step 1.** An amplificationally complete set of *germs* is chosen, so one amplifying *all* gauge-invariant parameters in the gate set. **Step 2.** *Base circuits* are defined by choosing powers $p = \lfloor L/|g| \rfloor$, with $L = 1, 2, 4, \dots$ and $|g|$ the germ's depth. **Step 3.** Each base circuit is sandwiched between each of $N_{f1} \times N_{f2}$ fiducial pairs, defined by N_{f1} preparation and N_{f2} measurement fiducial circuits. **Step 4.** Optional *fiducial pair reduction* (FPR) can be applied to reduce the amount of circuits to be run, whilst retaining sensitivity to every base circuit's amplified parameters. **Step 5.** Visualization of a GST experiment design as a grid of plaquettes. Every plaquette is a base circuit, ordered by germ and L. Each square within a plaquette corresponds to a fiducial pair. Figure taken from Nielsen *et al.* [8]

The experiment design is completed by sandwiching every germ between every pair of fiducials for the multiple germ powers, as can be seen in figure 4.5.

Fiducial reduction

Lastly, since the elements of a gate set are interdependent, not all circuits of the experiment design need to be run to find the optimal gate set. Fiducial pair reduction (FPR) is used to eliminate unnecessary circuits. In this thesis, FPR was not conducted, as the amount of circuits run was not that large. So, the interested reader is once again referred to Nielsen *et al* [8].

4.3.2. The optimisation procedure: gate set parameters estimation

Recall the very beginning of this section. There, it was stated that two of GST's characteristics make it one of a kind: its experiment design, and its use of maximum likelihood estimation (MLE). In this subsection, the second characteristic will be explained. First, the maximum (log) likelihood function is considered. Then, the algorithm used to find the maximum is explained.

The loglikelihood function

Let $\mathcal{G}$ indicate the gate set model, s index a conducted circuit and N_s the amount of times circuit s was repeated in the GST-experiment. Additionally, let m_s be the number of outcomes of s, and the number of times outcome β_s was observed be denoted by N_{s,β_s}. The likelihood function of circuit s is derived from the probability mass function of a multinomial distribution [30]:

$$\mathcal{L}_s = \frac{N_s!}{\prod_{\beta_s}(N_{s,\beta_s}!)} \prod_{\beta_s} p_{s,\beta_s}^{N_{s,\beta_s}}. \tag{4.16}$$

Assuming that $\sum_{\beta_s} N_{s,\beta_s} = N_s$, $\mathcal{G}$ is TP (so that the probabilities sum to one: $\forall s, \sum_{\beta_s} p_{s,\beta_s} = 1$) and setting the observed *frequency* $f_{s,\beta_s} = N_{s,\beta_s}/N_s$:

$$\begin{aligned}
\log(\mathcal{L}_s) &= \log\left(\frac{N_s!}{\prod_{\beta_s}(N_{s,\beta_s}!)} \prod_{\beta_s} p_{s,\beta_s}^{N_{s,\beta_s}}\right) \\
&= \log(N_s!) - \log\left(\prod_{\beta_s}(N_{s,\beta_s}!)\right) + \log\left(\prod_{\beta_s} p_{s,\beta_s}^{N_{s,\beta_s}}\right) \\
&= \log(N_s!) - \sum_{\beta_s} \log\Big(N_{s,\beta_s}!\Big) + \sum_{\beta_s} N_{s,\beta_s} \log\Big(p_{s,\beta_s}\Big) \\
&= N_s \sum_{\beta_s} f_{s,\beta_s} \log\Big(p_{s,\beta_s}\Big) + \log(N_s!) - \sum_{\beta_s} \log\Big(N_{s,\beta_s}!\Big)
\end{aligned}$$

Here, p_{s,β_s} is the *probability* predicted by $\mathcal{G}$ of getting outcome β_s from circuit s. From now on, the factors independent of these probabilities are omitted, such that the total loglikelihood for the entire GST-experiment becomes:

$$\log(\mathcal{L}_s) = \sum_s \log(\mathcal{L}_s) = \sum_{s,\beta_s} N_s f_{s,\beta_s} \log\Big(p_{s,\beta_s}\Big). \tag{4.17}$$

Here s ranges over all the conducted circuits. If the gate set model is not TP constrained, another, more general likelihood function is used. `pyGSTi` always uses the general log $\mathcal{L}$ (A.4.3).

The probabilities p_{s,β_s} are nonlinear functions of parameters of $\mathcal{G}$, which makes maximizing the GST likelihood function nontrivial. The probabilities p_{s,β_s} *oscillate*, resulting in an extremely non-convex loglikelihood function with many local maxima, resembling an egg crate. If the CP condition is implemented, more problems are created by gauge freedom. Ignoring the CP constraint makes optimization easier, but can lead to unphysical gate sets with zero or negative likelihood. The optimisation procedure described next can reliably deal with these complications.

The optimisation algorithm

Now *a* way of how parameter-estimation can be conducted using long-sequence gate set tomography is presented. The algorithm, summarized in Algorithm 2, is used by `pyGSTi` as well. $\mathcal{D}_0$ is the full GST data set, so all the circuits of the experiment design and their outcomes. $\vec{\theta}_0$ is the initial vector of model parameters, which is oftentimes provided by linear GST. $\vec{\theta}$ is the vector of the parameters of $\mathcal{G}$; these are being optimized such that the gate set they form fits observations best. Truncate$(\mathcal{D}_0, L)$ results in a *subset* $\mathcal{D}$ of $\mathcal{D}_0$, by only including circuits with base circuit depth $\leq L$. Argmin$(S, \mathcal{G}, \mathcal{D}, \vec{\theta}_1)$ gives the $\vec{\theta}$ at which the statistic $S(\mathcal{G}(\vec{\theta}), \mathcal{D})$ is minimal, in which the local optimizer is seeded at $\vec{\theta}_1$.

Algorithm 2 Long-sequence gate set tomography

$\vec{\theta} \leftarrow \vec{\theta}_0$
for $L \in 1, 2, 4, 8, \ldots$ **do**
  $\mathcal{D} \leftarrow \text{Truncate}(\mathcal{D}_0, L)$
  $\vec{\theta} \leftarrow \text{Argmin}(\chi^2, \mathcal{G}, \mathcal{D}, \vec{\theta})$
end for
$\vec{\theta} \leftarrow \text{Argmin}(-\log \mathcal{L}, \mathcal{G}, \mathcal{D}_0, \vec{\theta})$

The optimization procedure is executed in multiple "stages": for $L = 1, 2, 4, \ldots$. In each stage, only circuits (and their outcomes) with base circuit depth smaller than the current L are used in the minimization. The starting point of the optimization is the best-fit of the previous stage. As mentioned, the first stage uses linear GST's fit as a starting point. Repeating each circuit enough times ensures finite sample fluctuations to be small, so that starting points lie in the correct basin of the next stage's objective function, thus avoiding local minima.

In every stage except the final one, the χ^2 statistic is being optimized instead of loglikelihood. It is used in the pre-final stages because, in comparison with the loglikelihood, it can be computed faster and is more well-behaved as an objective function. In addition, if the number of circuit repetitions is high enough, minimising χ^2 is more robust as a starting point for the final $\log(\mathcal{L})$ maximization in contrast to performing only $\log(\mathcal{L})$ maximizations. The weighted-sum-of-squares function of χ^2 quantifies goodness-of-fit and is calculated as the sum of contributions from each circuit s:

$$\chi^2 = \sum_s \chi_s^2 = \sum_s \sum_{\beta_s} N_s \frac{(p_{s,\beta_s} - f_{s,\beta_s})^2}{p_{s,\beta_s}} = \sum_{s,\beta_s} N_s \frac{(p_{s,\beta_s} - f_{s,\beta_s})^2}{p_{s,\beta_s}}. \tag{4.18}$$

χ^2 is a local quadratic approximation to the negative loglikelihood. Unfortunately, minimum-χ^2 estimation can be significantly *biased* by overestimating the probability of rare events. In certain GST-experiments, it can overestimate SPAM error by 100%. Luckily, optimizing the loglikelihood in the last stage resolves this issue.

Currently, the Levenberg-Marquardt optimization method (LM) described in Chapter 3 is used in all the stages, including the final step of loglikelihood minimisation [29]. Recall that LM is used for minimising *least-squares* (LS) objective functions, of the form $f(x) = 1/2 \sum_{j=1}^{m} r_j^2(x)$ (Eq.(3.1)). However, the objective functions in this GST algorithm, the loglikelihood (Eq.(4.17)) and the χ^2 proxy (Eq.(4.18)), are not of this form. This problem can be resolved by simply taking squares of the non-LS objective functions, unfortunately resulting in less nice to deal with Jacobians and Hessians. It was found that with the loglikelihood, indeed the square root is taken [31]. Furthermore, Nielsen *et al* found that this approach of "forcing" the problem into a least-squares form outperformed other optimization methods, such as conjugate gradient descent. As for the χ^2 proxy, the square root is taken as well and LM's performance was found to be acceptable. Note that χ^2 can be made LS easily by setting the probabilities in the denominator to a constant, for instance the probabilities of the previous iteration. At present, it is not certain that the LM method is the best one for GST optimization, and further research into the topic would be interesting.

A final adjustment is sometimes made to increase the reliability of optimization. Both $\log(\mathcal{L})$ (Eq.(4.17)) and χ^2 (Eq.(4.18)) have poles when probabilities are zero. χ^2 is regularized by limiting the least-squares weights to a maximum cutoff $1/p_{\min}$. This does not affect the final fit since χ^2 is a proxy (for the negative $\log(\mathcal{L})$). For $\log(\mathcal{L})$, when $p_{s,\beta_s} < p_{\min}$, it is replaced by its second-order Taylor series:

$$\log(L) \approx \sum_{s,\beta_s} N_s f_{s,\beta_s} \left[\log(p_{\min}) + \frac{1}{p_{\min}}(p_{s,\beta_s} - p_{\min}) - \frac{1}{p_{\min}^2} \frac{(p_{s,\beta_s} - p_{\min})^2}{2} \right]. \tag{4.19}$$

Here $p_{\min}$ is much less than the smallest possible non-zero frequency, for example $p_{\min} = 10^{-4}$ if each circuit is repeated 1000 times. This adaptation distorts the objective's value only for particularly bad fits when a probability is much different than its observed frequency.

Before it is explained how GST's results can be analysed, the protocol is summarized. The experiment begins with knowing which circuits need to be run, the experiment design. Every circuit is composed

of a preparation fiducial, then a germ, followed by a measurement fiducial. First, the sets of fiducials and germs are selected. The circuit list is formed by all possible combinations of fiducials and germs. In every circuit, the germ is repeated for a different amount of times, determined by the germ-power. The circuits' outcomes are collected and the best-fit gate set is found by the Levenberg-Marquardt method in maximum likelihood estimation. The optimization consists of multiple stages. The χ^2 squared statistic is optimized for increasing data sets, functioning as a proxy for the loglikelihood, which is optimized in the final stage.

4.4. Analysing GST estimates

In this section, it is discussed how gate set tomography's results can be interpreted. It is assumed that the GST optimization did not get stuck at a local maximum, but was successful in finding a global optimum. Quantification of goodness of fit will be presented. Then, gauge optimization will be considered briefly, as most metrics used in quantum computation are gauge-dependent.

4.4.1. Assumptions of the model

In the next section, it is explained how goodness of fit for GST models is quantified. If GST's gate set model fails to fit the data, *"this is strong evidence that* some *assumption of the model was violated"* [8]. Nielsen *et al* use "non-Markovianity" synonymously with "model-violation", to mean *"the observed behaviour was influenced by some internal or external* contex *variable that was not included in the model"*. It should be noted, however, that non-Markovianity in this sense is not necessarily caused by non-Markovian dynamics, as in the sense of the negation of definition 2.6. The assumptions of the model are stated for overview. The model could be violated because any of them is not satisfied:

1. The sets of preparation fiducials and measurement fiducials are informationally complete.
2. GST's optimization resulted in a global optimum, it did not get stuck in a local maximum.
3. The system's dynamics is Markovian (Def.2.6) (if the CPTP constraint is implemented).
4. The germ set is amplificationally complete.

The last assumption of AC germs is not necessary to conduct gate set tomography correctly. However, generally it is implemented, since it results in a gate set sensitive to all possible errors. As assumption 3 is the most difficult one to satisfy, a lot of model violation *"indicates high confidence in the conclusion "the Markovian model* (assumption 3) *was violated""* [8]. Unfortunately, GST does not tell the cause of model-violation, it only tells that the model has been violated.

4.4.2. Goodness of fit

Now, the validity of the GST estimate will be assessed, i.e. it will be quantified how well the GST model matches the data. A measure of non-Markovianity will be introduced.

Firstly, we consider the amount of independent degrees of freedom in the GST-experiment. Suppose a GST experiment contains N_{exp} distinct circuits, and its produced dataset is described by N_{o} free parameters, N_{o} being the number of independent circuit outcomes that can be observed. If there is just one native measurement ($N_{\text{M}} = 1$) that has $N_{\text{E}^{(1)}}$ outcomes, then:

$$N_{\text{o}} = N_{\text{exp}}(N_{\text{E}}^{(1)} - 1). \tag{4.20}$$

In the case of a single qubit supporting one native 2-outcome measurement ($N_{\text{E}}^{(1)} = 2$), as in this thesis, each circuit's data has $2 - 1 = 1$ independent degree of freedom. Simply put, if a circuit was run 1000 times, and 600 times the outcome was 0, then it can be deduced that 1 was 400 times the outcome ($N_{\text{experiments}} = N_{\text{E},1}^{(1)} + N_{\text{E},2}^{(1)}$). Fitting the *maximal model* with N_{o} parameters to the data, so where each independent outcome is assigned one probability, delivers $\log \mathcal{L}_{\text{max}}$.

If the data were produced by a Markovian gate set, then the maximal model and the GST model should fit the data equally well *if* extra free parameters in the maximal model are accounted for. From Wilks' theorem [8], it is known that if the gate set model is valid, the loglikelihood ratio statistic between the GST estimate and the maximal model is a χ^2_k random variable:

$$2(\log \mathcal{L}_{\text{max}} - \log \mathcal{L}) \sim \chi^2_k, \tag{4.21}$$

where $k = N_o - N_p^{\text{nongauge}}$ is the amount of parameters that the maximal model has more than the gate set model has non-gauge parameters. The mean and standard deviation of the χ^2_k distribution are respectively k and $\sqrt{2k}$. The observed *model violation* is quantified by the number of standard deviations by which the loglikelihood ratio surpasses the expected value, mean, of χ^2_k:

$$N_\sigma := \frac{2(\log \mathcal{L}_{\max} - \log \mathcal{L}) - k}{\sqrt{2k}}. \tag{4.22}$$

If $N_\sigma \leq 1$, then the GST model fits the data extremely well. Note that in this case, N_σ can become negative as well. If $N_\sigma >> 1$, there is reason to *assume* that the Markovian model was violated. So, a high N_σ does not follow necessarily from non-Markovian behaviour, but is a strong indicator of its occurrence during the experiment. Furthermore, N_σ generally increases linearly with the number of times each circuit is repeated (a measure of *sensitivity*), meaning more sensitive experiments have increased N_σ's per definition.

Lastly, it is interesting to consider each circuit's contribution to the loglikelihood. When, for example, one suspects one particular gate to be (mostly) responsible for model-violation. For each circuit s in a Markovian model, $2(\log \mathcal{L}_{\max,\,s} - \log \mathcal{L}_s) \sim \chi^2_k$, where k is approximately the number of independent outcomes of a single circuit, so $k \approx 1$ in a 1-qubit 2-outcomes experiment. Using a temperature scale, each circuit's model violation can then be indicated in a grid similar to figure 4.5.5.

4.4.3. Gauge optimization

Finally, to conclude this chapter, something on gauges and gauge optimization is said. Recall that a gate set's representation is *not* unique, as gauge transformations do not change circuit outcomes (Ch.2.3.3). However, most metrics in quantum computing, like the diamond norm, require a concrete representation of the gate set, so one in specific gauge. Thus, when GST presents its estimates, a gauge in which to represent them needs to be chosen. GST does this in *gauge optimization*, by selecting the gauge in which its results are closest to ideal, *"to make the gates look as good as possible"* [8]. `pyGSTi` offers a lot of methods for gauge optimization, for example the Nelder-Mead and the conjugate gradient method, which are all derivative free. The methods simply go over possible gauges to find the one that minimizes the Frobenius distance between the estimated and ideal gate set:

Definition 4.5. The *weighted* sum of squared Frobenius distances between gate sets $\hat{\mathcal{G}}$ and $\mathcal{G}$ is defined as

$$f(\hat{\mathcal{G}}, \mathcal{G}) = \sum_{i=1}^{N_\rho} \alpha_i \left| \hat{\rho}^{(i)} - \rho^{(i)} \right|^2 + \sum_{i=1}^{N_G} \beta_i \left| \hat{G}_i - G_i \right|^2 + \sum_{m=1}^{N_M} \sum_{i=1}^{N_E^{(m)}} \gamma_{m,i} \left| \hat{E}_i^{\,(m)} - E_i^{(m)} \right|^2.$$

Here, the weights $\alpha_i, \beta_i, \gamma_{m,i} \in \mathbb{R}$ and $|\cdot|$ is the Frobenius norm, $\|X\|_F \equiv \sqrt{\text{Tr}(XX^\dagger)}$.

Gauge freedom is an interesting topic an sich, and will unfortunately not be treated further in this thesis. As per usual, the reader is referred to Nielsen *et al* [8] for more details.

5

The NV Center in Diamond

The quest for systems that might be used as hardware in quantum information processing (QIP) has led researchers to consider diamond as a candidate due to its favourable properties, like high Debye temperature and stable 'colour centers' [32]. Nitrogen defects are one of the more than 100 of these colour centers in diamond. The nitrogen-vacancy (NV^-) center was in 1997 the first *single* colour center to be detected. The use of the NV^- center as a solid state spin qubit for QIP [33] is one of the reasons why its discovery is considered a breakthrough. This and other implementations, like the use of NV^- centers in quantum optical networks, have resulted in the NV^- center and its applications being widely researched. Moreover, the electronic spin and nuclear spins of the NV^- center have been demonstrated suitable for implementation in quantum registers, systems consisting of multiple qubits, at room temperature [33]. In this thesis however, all experimental methods have been performed at 3.8 Kelvin. The NV center is found in both the neutral charge (NV^0) and the negatively-charged (NV^-) state. Their electronic band structure was proposed in 1977 by Loubser and Van Wyk through their research of the electron paramagnetic resonance (EPR) signal of the NV^- center. Five electrons are responsible for the filling of the molecular orbitals of NV^0. Two of these are from the nitrogen atom and the other three are inherent to the carbon atoms adjacent to the vacancy. NV^0 is able to transition into NV^- only via capturing the additional electron from a nearby donor [33], a local charge trap. Since the NV^0 center has not been able to demonstrate the suitable properties of the NV^- center, most research, as well as this thesis, utilises the latter state (hereafter denoted NV).

In this chapter, first the physical and electronic structure of the NV center will be described. Then, the coupling between its electron, which will be the qubit considered in this thesis, and the nitrogen nucleus, will be discussed.

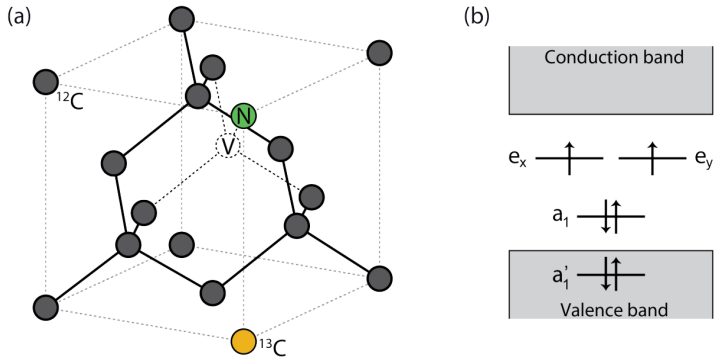


Figure 5.1: (a) The cubic unit cell of a diamond lattice with an NV center. The nitrogen atom (green) and a vacancy (white) have replaced two ^{12}C atoms (grey). The ^{13}C isotope (yellow) replaces 1.1% of the ^{12}C atoms. **(b)** The molecular orbitals (a_1', a_1, e_x and e_y) of the NV^- center showed in the ground-state filling, occupied by six electrons (the arrows). Figures are taken from Bradley [34].

5.1. The physical structure and electronic structure of the NV center

The physical and electronic structure of the NV center has been determined by various experiments as well as through the performance of *ab initio* calculations. For a detailed description of this ongoing research process of the NV center I refer to Doherty *et al* [33].

At the basis of the crystal structure of the NV center lies the standard diamond structure; two interpenetrating face centered cubic lattices. However, in the NV center, two carbon atoms have been substituted by a nitrogen-vacancy pair in the [111] direction. That is, one carbon atom is removed, and one has been replaced by a nitrogen atom. The NV lattice structure is depicted in Fig.5.1(a).

There are four possible ways in which the six electrons of the NV can occupy the molecular orbitals (Fig.5.2). Two of these are triplet states (^{3}A$_2$, ^{3}E,) and the other two are singlet states (^{1}E, ^{1}A$_1$). Recall that a triplet state has spin quantum number $S = 1$, meaning there are three (hence triplet) possible spin component values ($m_s = -1, 0, 1$). The singlet states have overall spin quantum number $S = 0$, resulting in only one spin value of $m_s = 0$. The four ways of distributing the six electrons differ in the amount of electrons they have on each molecular orbital (MO). The NV has four MO's, of which three (a_1, e_x and e_y) are deep within the diamond band gap, so far from both the conduction band and the valence band. The last one ($a_1^{'}$) is completely inside the valence band. This can be seen in Fig.5.1(b).

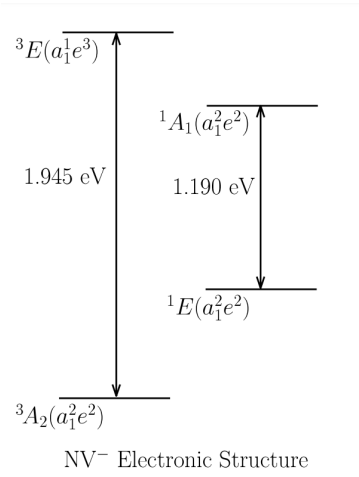


Figure 5.2: The electronic structure of the NV$^-$ center. The triplet ground and excited states ^{3}A$_2$ and ^{3}E, as well as the intermediate singlet states ^{1}A$_1$ and ^{1}E are depicted. The electron occupation of the MO's is indicated between brackets. The corresponding energies of the transitions between the levels are shown. Figure is taken from Doherty *et al* [33].

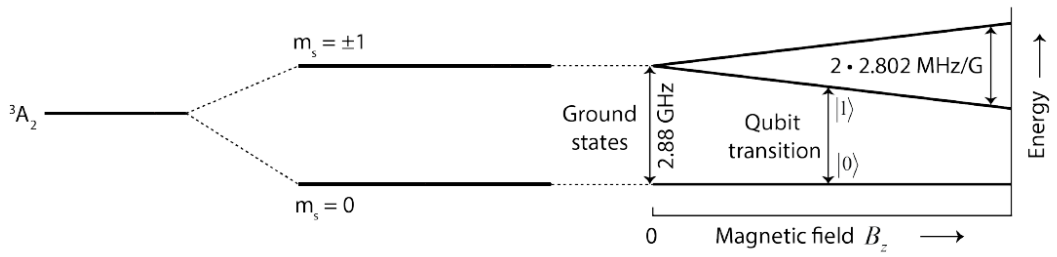


Figure 5.3: The spin-triplet ground state ^{3}A$_2$ is separated into the $m_s = 0$ level and degenerate $m_s = \pm 1$ levels with a zero-field-splitting of 2.88 GHz (no magnetic field present). Applying a magnetic field splits the $m_s = \pm 1$ states (Zeeman splitting). A qubit is then defined in the $m_s = \{0, -1\} = \{|0\rangle, |1\rangle\}$ basis. Figure adapted from Bradley [34].

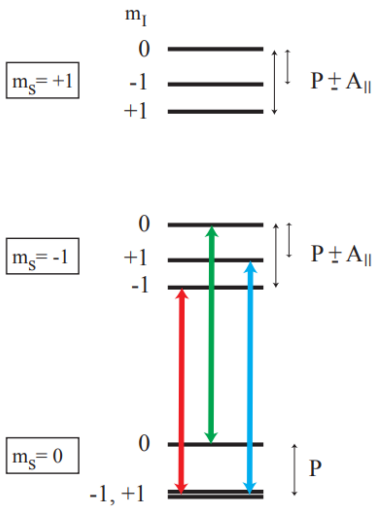
The electronic state that will be considered in this thesis, is the triplet ^{3}A$_2$ ground state. Here, four electrons occupy the two molecular orbitals (MO's) a_1 and $a_1^{'}$. Each one of the two remaining electrons occupies one of the two remaining MO's (Fig.5.1(b)), which are degenerate (with the same energy). These two unpaired electrons on the same energy levels result in a spin-triplet ($S = 1$) state. The $m_s = \pm 1$ are degenerate and are separated by a zero-field-splitting (in the absence of a magnetic field) of 2.88

GHz. Applying a magnetic field lifts the degeneracy between the $m_s = \pm 1$ states, i.e. splits them into distinguishable states (Zeeman splitting). Now an electron-spin qubit can be defined, which is mostly, also in this thesis, done in the $m_s = \{0, -1\} = \{|0\rangle, |1\rangle\}$ basis (Fig.5.3).

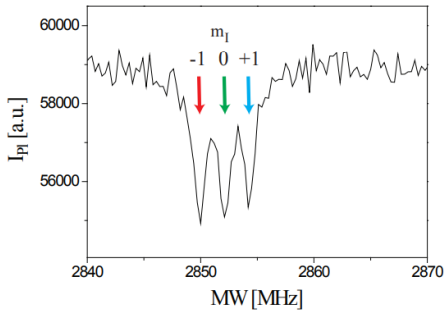
5.2. The nitrogen nucleus

Needless to say, the electron considered in the NV center is not an isolated system, but interacts in varying degrees with the surrounding subsystems. The physical system having most influence on the qubit is the ^{14}N nucleus [35], which will be described in this section.

Like the electron spin, the nitrogen nuclear spin m_I is a spin triplet with possible values $m_I = -1, 0, 1$ [36]. Quadrupole interaction lifts the degeneracy of $m_I = 0$ and $m_I = \pm 1$ in the $m_s = 0$ state by a value $P \approx 4.9$ MHz. Furthermore, the hyperfine interaction in the $m_s = \pm 1$ states splits the $m_I = \pm 1$ states, such that the $m_I = 0$ and $m_I = \pm 1$ states now differ $P \pm A_{||}$ with $A_{||} \approx 2.2$ MHz. This is shown in figure 5.4a. In figure 5.4b the NV electron-qubit transition between the two states of the qubit can be seen (so $m_s = 0 \leftrightarrow m_s = -1$). The red, green and blue lines correspond to the three nuclear spin states. Using radio frequency (RF) pulses, the nitrogen nucleus can be *initialised*, via nuclear spin polarization, into one of the three spin states, such that it is not an incoherent mixture of the three spin states anymore. Consequently, the fluorescence of the electron qubit transition shows only one dip, instead of three like in figure 5.4b. For more details on the polarization of the nitrogen nucleus I refer to Steiner [36].



(a) The energy levels of the NV electron qubit (defined between $m_s = -1$ and $m_s = 0$). Due to the nitrogen nuclear spin m_I coupling to the electron, the decay/excitation of the qubit depends on the spin state of the nitrogen. The nitrogen spin is also a spin triplet ($m_I = -1, 0, 1$). The red, green and blue vertical lines indicate the possible transitions of the qubit, now dependent on m_I. P and $A_{||}$ indicate by which frequencies the m_I states differ from each other.



(b) A closeup of the $m_s = 0 \leftrightarrow m_s = -1$ qubit transition. The intensity of the transition depends on the spin state of the nitrogen nucleus.

Figure 5.4: The NV electron couples to the nitrogen nucleus. Figures are taken from Steiner [36].

The NV center's dynamics is best summarized by its Hamiltonian, which can be found in the appendix (App.A.3).

6
Method

In this chapter, it will be described which gate sets were used for conducting gate set tomography, and how GST was actually implemented. GST was applied on three different gate sets, whose gates all act on one qubit, the electron of the NV center. First, the gate sets will be described. The circuits were run experimentally or by an NV simulator. Secondly, the experimental conditions that could be varied to research their effects on the outcomes of GST will be presented. These include different ways of initialising the nitrogen nucleus of the NV center, as well as the presence of an XY-4 echo (Ch.2.2.3). Lastly, it will be explained how GST was performed using `pyGSTi`.

6.1. The three researched gate sets

In this thesis, three different gate sets were investigated using gate set tomography. In this section they will be presented and explained. Every gate set's preparation and measurement fiducials (which are the same sets here), as well as their germs, are given in Table 6.1a, 6.1b and 6.1c. Note that all three gate sets have the same set of fiducial sequences. As stated in Ch.4.3.1, for a set of fiducials to be informationally complete (IC) in the 1-qubit Hilbert-Schmidt space (of dimension four), at least four different fiducials are necessary. However, one would like for the fiducials to be "as linearly independent" as possible. The fiducial sets of Blume-Kohout *et al* [29] were used, but without the $\emptyset$ (no gate applied; an ideal identity gate without duration, informally). It was not possible to implement the $\emptyset$-gate in our sequences, probably due to a reading error in the parser used to import the lists of circuits. Blume-Kohout *et al* chose this set, if $\emptyset$ is included, for their fiducials, as it ideally prepares the six stabilizer states, i.e. the eigenstates of the Pauli X, Y and Z operators.

Furthermore, germ selection (Ch.4.3.1) was conducted, which resulted in the amplificationally complete (AC) germ set of Gate Set A. For Gate Set B, the germ set found by Blume-Kohout *et al* [29] was used, an AC germ set as well. The germ set in Gate Set C is not AC, but its results will be examined anyway.

6.1.1. The NV centers

It was possible to run the circuits either experimentally or to simulate them using an NV simulator. Circuits corresponding to Gate Set A and C were run both ways, whereas Gate Set B's circuits were only simulated. The NV center at the Taminiau Lab at QuTech was used in the experiments. The NV simulator was controlled by a postdoc at this research group. The simulator is currently under development and during this thesis it was not possible to get the process matrices of the gates (the superoperators in Ch.2.2.2) out directly. So, the simulations were similar to the experiments in the sense that both do not have access to the true gates. In the future, if the process maps can be accessed directly, it can be assessed how well GST works, by comparing its estimates to the direct gates.

The electron spin of the NV is manipulated by microwave (MW) driving pulses (i.e. the gates) at the associated transition frequency between the initial and the desired state [34]. The qubit is measured via single-shot read-out using lasers. Different optical transitions are associated to different spin states. A laser pulse is applied that is resonant with the $m_s = 0$ ($|0\rangle$) state, such that only if the electron is in

this state, the electron is excited and photons are detected. For more details on the read-out and the experimental setup I refer to Bradley [34].

6.1.2. Experimental conditions

There were two experimental conditions that could be varied to research their effects on the results of gate set tomography. First, the circuits could be run with an XY-4 echo after every applied gate (Germ Sets A and C). Furthermore, the nitrogen nucleus could be initialised in different ways. For the actual experiments, the nitrogen could be either not initialised (mixed state), or initialised into either one state ($m_I = 0$ state) or into one of its three spin states ($m_I = -1, 0, 1$ states). In the simulations, the nitrogen could be initiliased into the $m_I = 0$ state with a fidelity of $p = 0, p = 0.95$ or $p = 1$. Initialisation fidelity $p = 0$ means that the nitrogen is in a completely mixed state. The nitrogen spin states are described in chapter 5.2. An overview of every gate set's experimental conditions can be found in Table 6.2a, 6.2b.

Using the germs and fiducials, an experiment design (Ch.4.3.1) composed of all circuits to be run, was created for each gate set. The maximum base circuit depth was $L = 16$. The qubit was measured in the Z-basis. Every circuit was run 1000 times, and the amount of times a 0 and 1 (corresponding to respectively $|0\rangle$ and $|1\rangle$ in Fig.2.1) were measured, was recorded (in an `.npy`-file). With these lists of outcomes, the gate sets could be estimated using `pyGSTi`.

Table 6.1: The gate sets researched in this thesis. The preparation and measurement fiducials (fiducials for short) are the same for each gate set.

(a) Gate set A, estimating three gates: I, $X^{\pi/2}$ and $Y^{\pi/2}$.

	Fiducials	Germs
1	$X^{\pi/2}$	I
2	$Y^{\pi/2}$	$X^{\pi/2}$
3	$X^{\pi/2}X^{\pi/2}$	$Y^{\pi/2}$
4	$X^{\pi/2}X^{\pi/2}X^{\pi/2}$	$IIIIX^{\pi/2}$
5	$Y^{\pi/2}Y^{\pi/2}Y^{\pi/2}$	$X^{\pi/2}Y^{\pi/2}$
6		$IIIIX^{\pi/2}Y^{\pi/2}$
7		$IIIIY^{\pi/2}X^{\pi/2}$
8		$IIY^{\pi/2}X^{\pi/2}X^{\pi/2}X^{\pi/2}$
9		$X^{\pi/2}X^{\pi/2}Y^{\pi/2}Y^{\pi/2}X^{\pi/2}Y^{\pi/2}$
10		$IIX^{\pi/2}X^{\pi/2}X^{\pi/2}Y^{\pi/2}$

(b) Gate set B, estimating three gates: I, $X^{\pi/2}$ and $Y^{\pi/2}$.

	Fiducials	Germs
1	$X^{\pi/2}$	I
2	$Y^{\pi/2}$	$X^{\pi/2}$
3	$X^{\pi/2}X^{\pi/2}$	$Y^{\pi/2}$
4	$X^{\pi/2}X^{\pi/2}X^{\pi/2}$	$X^{\pi/2}Y^{\pi/2}$
5	$Y^{\pi/2}Y^{\pi/2}Y^{\pi/2}$	$X^{\pi/2}Y^{\pi/2}I$
6		$X^{\pi/2}IY^{\pi/2}$
7		$X^{\pi/2}II$
8		$Y^{\pi/2}II$
9		$X^{\pi/2}X^{\pi/2}IY^{\pi/2}$
10		$X^{\pi/2}Y^{\pi/2}Y^{\pi/2}I$
11		$X^{\pi/2}X^{\pi/2}Y^{\pi/2}X^{\pi/2}Y^{\pi/2}Y^{\pi/2}$

(c) Gate set C, estimating five gates: I, $X^{\pi/2}$, $Y^{\pi/2}$, X^{π} and Y^{π}.

	Fiducials	Germs
1	I	I
2	$X^{\pi/2}$	$X^{\pi/2}$
3	$Y^{\pi/2}$	$Y^{\pi/2}$
4	$X^{\pi/2}X^{\pi/2}$	X^{π}
5	$X^{\pi/2}X^{\pi/2}X^{\pi/2}$	Y^{π}
6	$Y^{\pi/2}Y^{\pi/2}Y^{\pi/2}$	

6.2. Running pyGSTi

A parser made it very easy to perform gate set tomography. Besides the list of circuits and their outcomes mentioned above, one additional thing was needed. A *target model* had to be made for every gate set. This model ran the same circuits as the corresponding gate set using ideal gates. This made it possible to compare, using various metrics, the ideal (target) model and the estimated one, computed by `pyGSTi`.

On top of specifying the circuits, their outcomes and the target model, the constraints on the estimated model had to be chosen. In this thesis, every model was run both TP and CPTP constrained (Ch.2.3.1). The reasoning behind this choice is well-explained by Nielsen *et al* [8]: *"If a significantly better non-CP fit is found, then the two estimates should be examined carefully: either the CPTP fit got trapped in a local*

maximum, or significant non-Markovian dynamics occurred during the experiment.". By running the model with both constraints, it could be assessed whether one of these events had occurred. The second option is caused by the fact that CPTP constraint models capture Markovian dynamics only, whereas TP also captures more general evolutions (Ch.2.3.2). So, in the presence of significant non-Markovian dynamics, the non-CP model would produce less model violation, i.e. a better fit.

Table 6.2: Tables elaborating on the gate sets researched in this thesis. In Table 6.2a, for every gate set can be seen whether an XY-4 was used during the circuits and whether the circuits were run experimentally (or exp. for short), were simulated (sim. for short), or both. In Table 6.2b, the different ways of initialising the nitrogen nucleus can be seen for the experimental circuits and simulated circuits, respectively.

(a)

Gate set	Amount of estimated gates	Exp. or sim.	Presence XY-4 echo
A	3	Both	Present
B	3	Simulated only	Absent
C	5	Both	Absent

(b)

	Experiments	Simulations
Initialisation states	mixed, $m_I = -1, 0, 1$	-
Initialisation fidelity into $m_I = 0$ state	-	$p = 0$ (mixed), $0.95, 1$

7

Results

In this chapter, the analyses provided by conducting gate set tomography (GST) on the three different gate sets (Table 6.1a, 6.1b, 6.1c) will be presented and interpreted. Firstly, since Gate Set C does not have an amplificationally complete (AC) germ set, the validity of its results will be investigated. As stated in the methods section, some circuits (Gate Set A and C) have been run experimentally and via using a simulator, whereas one circuit set (Gate Set B) has only been simulated. It will be analysed how well the simulated experiments are able to imitate real experiments, as this is essentially what simulations ought to do. The simulations will be included in the analyses if they are considered valid in this sense. Then, the effect of different initialisation procedures of the nitrogen nucleus will be analysed. Lastly, the consequences of the presence of an XY-4 echo (Ch.2.2.3) will be investigated. Three concepts will aid analysis of the results: the diamond norm (Eq.(2.3)), the level of model violation N_σ (Eq.(4.22)), and error generators (Ch.2.3.2).

The diamond distances (Eq.(2.3)) between the ideal gates and i) the gates produced by the TP models and ii) the gates estimated by CPTP models, turned out (mostly) higher for the TP models. The TP optimization is performing worse probably due to its larger parameter space (this claim could be checked by running GST on a fully parameterized model; one with the maximal parameter space (Ch.2.3.1). So, from the diamond distances it cannot be concluded that any of the CPTP fits got stuck in a local maximum. As the only purpose of the TP diamond distances was to assess whether this had happened, they have served their purpose and will not be analysed further. The TP diamond distances for all the gate sets can be found in the appendix (App.A.5).

7.1. Gate set C: an amplificationally incomplete germ set

In this section, it will be explored whether conducting GST, whilst not using an AC set of germs, influences the results significantly. To this end, Gate Set C (Table 6.1c), whose circuits were run without an XY-4 echo (Ch.2.2.3), will be considered.

The diamond distances between the five ideal gates (Eq.(2.7)) to the gates estimated by i) running GST on experimentally acquired data and by ii) running GST on simulated data can be seen in tables 7.1a and 7.1b, in which both models are CPTP constrained. Furthermore, the diamond distances between i) and ii) have been calculated (Tab.7.2). Here, the mixed state (experiments) was compared to initialisation fidelity $p = 0$ (the completely mixed state) in the simulations. Initialisation in the experiments was compared to initialisation fidelity $p = 0.95$ in the simulations, as the probability of successful initialisation, in reality, lies around $0.90 - 0.95$ [35]. If the nitrogen is not initialised, the estimates for both models are far from ideal, and moreover: far from each other. To make this claim more concrete, consider the largest diamond distance ($X^{\pi/2}$ gate in Tab.7.2). If one of $X^{\pi/2}_{\text{experimental}}$ and $X^{\pi/2}_{\text{simulated}}$ gates was applied, then the maximum probability of reporting correctly which one of the two was applied is 0.58 ($0.315419/4 + 0.5$; formula below Def.2.3). If the gates were the same, this would be 0.5. Thus, the diamond distances indicate that the simulation and the experiments produce differing gate estimations if the nitrogen is not initialised.

Now consider the case where the nitrogen has been initialised. Then, the diamond distances from the ideal gates to the i) experimental gates and ii) simulated are low. Moreover, the estimates are not far

from each other. However, the decision is made to not investigate this gate set further. There are two main reasons for this decision. Firstly, it is preferred that the simulation is able to mimic reality well for all different initialisation procedures of the nitrogen (and not only when it has been initialised with a high fidelity); right now, it has not proved itself to work reliably. Secondly, eventually the gate error generators will be analysed. The results of such analyses are well grounded if the germ set is sensitive to all possible errors, which is not the case with this non-AC gate set.

Table 7.1: 1/2-diamond distances ($\cdot 10^2$) from the experimental and simulated models to the ideal gates (CPTP Germ Set C, without echo). The model is indicated in the secondary captions. The five estimated gates are I, $X^{\pi/2}$, $Y^{\pi/2}$, X^{π} and Y^{π}. The experimental circuits were run either initialising the nitrogen nucleus into the $m_I = 0$ state, or not (mixed state). In the simulations, the nitrogen nucleus was initialised into the $m_I = 0$ state with different fidelities ($p = 0, 0.95, 1$).

(a) Experimental model.

	mixed state	$m_I = 0$
I	17.979	0.5164
$X^{\pi/2}$	26.1779	2.3881
$Y^{\pi/2}$	27.3567	2.0028
X^{π}	12.3189	3.7754
Y^{π}	10.0384	4.0837

(b) Simulated model.

	$p = 0$	$p = 0.95$	$p = 1$
I	5.221	0.2979	0.2343
$X^{\pi/2}$	26.55	0.6722	0.5211
$Y^{\pi/2}$	23.8462	0.4865	0.4788
X^{π}	10.1032	0.8147	0.6741
Y^{π}	2.3561	0.7201	0.6529

Table 7.2: 1/2-diamond distances ($\cdot 10^2$) from the simulated models to the experimental models (Germ Set C, CPTP constrained). The five estimated gates are I, $X^{\pi/2}$, $Y^{\pi/2}$, X^{π} and Y^{π}. The experiments were conducted by initialising the nitrogen nucleus to either the mixed state or the $m_I = 0$ state. The mixed state is compared with initialisation fidelity $p = 0$ (mixed state) in the simulation and the $m_I = 0$ state is compared with the simulation's initialisation fidelity set to $p = 0.95$.

	mixed state	$m_I = 0$ state
I	18.4688	0.5278
$X^{\pi/2}$	31.5419	2.841
$Y^{\pi/2}$	19.9483	2.2073
X^{π}	21.3397	3.4841
Y^{π}	9.5017	3.5489

7.2. Gate sets A and B: effect of nitrogen initialisation and XY-4 Echo

In this section, it will be studied how the nitrogen nucleus initialisation and XY-4 echo (Ch.2.2.3) affect the results of gate set tomography. First, it will be established that simulations of gate sets with amplificationally complete germ sets are valid, i.e. produce results significantly close to experimentally produced ones. For this, Gate Set A, with an AC germ set, will be considered, as it was run both experimentally and on an NV-simulator. To evaluate the effect of the nitrogen nucleus initialisation, the (simulated only) Gate Set B will be considered, as it was run without XY-4 echo. Then, to research the consequences of the XY-4 echo on GST's results, the (simulated) Gate Sets A and B will be compared to each other.

7.2.1. Gate set A: validity of simulations

To verify the validity of simulated data when regarding a gate set with an amplificationally complete germ set, Gate Set A (Table 6.1a, with XY-4 echo) will be considered. From the significantly low diamond distances in tables 7.3a and 7.3b, it is concluded that both the experiments and the simulations produce good estimations. To assess how well they match, the diamond distances between the experimental and the simulated gates is calculated (Tab.7.4). The mixed state experiments are compared to nitrogen nucleus initialisation fidelity 0 in the simulations. Initialisation fidelity $p = 0.95$ (into $m_I = 0$) is compared to all three initialisation states in the experiments ($m_I = 0, -1, +1$), to probe whether the simulator is perhaps better at simulating another state than the one it is supposed to. In terms of produced estimations, i.e. from the diamond distances for the mixed state and initialisation into $m_I = 0$, it is concluded that the simulations are able to mimic reality well enough. The largest diamond distance is 0.0345 ($Y^{\pi/2}$ gate), which translates to a probability of 0.51 of distinguishing the simulated and

experimental gate correctly. Furthermore, it is concluded the simulator is not significantly better at simulating another initialisation state ($m_I = +1, -1$) than the one it is supposed to ($m_I = 0$). The greatest difference in diamond distances is for the $Y^{\pi/2}$ gate: 1.5% ($m_I = 0$ vs $m_I = +1$).

The N_σ in the simulations is very low (Table 7.5a), most likely because the NV simulator is not able to reenact all true dynamics: an $N_\sigma << 1$ is rarely seen, except in artificially simulated data [8]. The N_σ of the experiments (Table 7.5b) indicates a good fit as well (`pyGSTi` rates it 4 out of 5). In terms of model violation, the difference in N_σ's of the simulations and experiments is not considered significant. However, it does indicate that the simulator assumes too ideal dynamics, hence the practically absent model violation. It is advised to tweak the NV simulator's dynamics so that it matches reality better.

Recall the amplificationally *in*complete Gate set C (Tab.6.1c), discussed in the previous section. Whereas Gate set A's simulations are reliable even if the nitrogen is not initialised, that was not the case with Gate set C. Unfortunately, it cannot be known whether Gate set C's shortcoming is caused by its non-AC germ set, of by the fact that its circuits did not use the XY-4 echo. During this work, it was not possible to run additional (AC) circuits experimentally without an XY-4 echo. Henceforth, it is assumed that the simulations produce reliable results if the germ set is AC, even if no echo is employed. However, it is advised to investigate this statement more carefully: it needs to be examined whether extrapolation of this conclusion to circuits without XY-4 echo is truly justified.

To summarize, it is concluded that the NV-simulator produces acceptable results. Its estimates are significantly close to estimates produced by the experiments. The discrepancy in the amounts of model violation of simulations versus experiments is not considered to impact further comparisons significantly. However, in the future, the validity of the NV-simulator should certainly be investigated more closely by comparing simulations and experiments without an XY-4 echo.

Table 7.3: 1/2-diamond distances ($\cdot 10^2$) from Gate set A's (with echo) simulated and experimental CPTP models to the three gate ideal model. The model is indicated in the secondary captions. The three estimated gates are I, $X^{\pi/2}$ and $Y^{\pi/2}$. The simulations were done for different fidelities of initialising the nitrogen nucleus into the $m_I = 0$ state ($p = 0$ (mixed), 0.95, 1), corresponding to different columns. In the experiments, the nitrogen was either not initialised (mixed state), or initialised into one of its three spin states ($m_I = 0, -1, +1$).

(a) Simulated.

	$p = 0$	$p = 0.95$	$p = 1$
I	0.3562	0.4482	0.457
$X^{\pi/2}$	0.3314	0.1237	0.1337
$Y^{\pi/2}$	0.1948	0.2574	0.2729

(b) Experimental.

	mixed	$m_I = 0$	$m_I = -1$	$m_I = +1$
I	0.2541	0.2703	0.6535	0.5565
$X^{\pi/2}$	0.2236	2.8794	3.4729	1.8653
$Y^{\pi/2}$	0.3151	4.0757	5.1134	2.5685

Table 7.4: 1/2-diamond distances ($\cdot 10^2$) from the simulated models to the experimental models (Germ Set A, both with echo, CPTP). The three estimated gates are I, $X^{\pi/2}$ and $Y^{\pi/2}$. The experiments were conducted by initialising the nitrogen nucleus to one of four states: the mixed state (not initialised), the $m_I = 0$ state, the $m_I = -1$ state and the $m_I = +1$ state. The mixed state is compared with initialisation fidelity $p = 0$ in the simulation. The three remaining states ($m_I = 0$, m_1 and p_1 state) are compared with the simulation's initialisation fidelity set to $p = 0.95$.

	mixed state	$m_I = 0$	$m_I = -1$	$m_I = +1$
I	0.5603	0.6468	0.9579	0.8825
$X^{\pi/2}$	0.4136	2.8389	3.4315	1.8434
$Y^{\pi/2}$	0.3233	3.845	4.8816	2.3502

Table 7.5: The N_σ for all models of Gate Set A (3 gates, with echo). Its circuits were simulated, as well as actually run ("experimental"). In the simulations, the nitrogen was initialised into the $m_I = 0$ state with different fidelities: $p = 0, 0.95, 1$. In the experiments, the nitrogen was initialised into one of three states (the $m_I = 0$, m_1 and p_1 state), or not. For all models, gate set tomography was performed TP constrained and CPTP constrained.

(a) Simulated.

	$p = 0$	$p = 0.95$	$p = 1$
TP	-12.7442	-17.2156	-17.822
CPTP	-12.6776	-17.0892	-17.5733

(b) Experimental.

	mixed	$m_I = 0$	$m_I = -1$	$m_I = +1$
TP	16.8616	9.48565	13.6271	9.9985
CPTP	17.3721	9.96056	14.3721	10.2937

7.2.2. Gate set B: effect of nitrogen initialisation without XY-4 echo

Now that the NV-simulator has been deemed to produce acceptable results, the effect of the nitrogen initialisation can be studied using Gate Set B, which was simulated only (without XY-4 echo).

In Table 7.6, the values of N_σ can be seen. They do not differ significantly for the TP vs CPTP models. However, for $p = 0$, the N_σ indicates high levels of model violation. In that case, the nitrogen is in a mixture of its three spin-states (Ch.5.2). From Wolf *et al* [37], it is known that non-Markovian dynamics can arise from an environment which is in a mixture of states (like the nitrogen nucleus). Furthermore, models with an initialised nitrogen ($p = 0.95$ and $p = 1$) do not have such high N_σ's. So, it is safe to assume that a mixed nitrogen nucleus causes significant model violation. As one of GST's assumptions, namely of Markovian dynamics, is likely not satisfied for $p = 0$, it does not necessarily produce reliable results. This is seen in Tab.7.7: the estimated gates are far from ideal. For future studies, it is advised to include the nitrogen nucleus as a qutrit in the model, to assess whether the amount of model violation decreases (as non-Markovian dynamics should decrease).

For $p = 0.95$ and $p = 1$, the diamond distances indicate that the estimates are close to ideal (Table 7.7). The low values of N_σ indicate low model violation. So, these cases will now be researched further using gate error generators (Ch.2.3.2). The pie charts in figures 7.1a and 7.1b show what percentage of the error generator is captured by each subspace. The Hamiltonian subspace (H) is pink, the stochastic subspace (S) is green. The (yellow) active and correlation subspaces are combined (A/C). Comparing the error generator sub space percentages of $p = 0.95$ and $p = 1$ is justified, because their error generators are of the same order of magnitude (their diamond distances are very close). If the nitrogen nucleus is initialised with a fidelity of 100%, all stochastic and A/C errors disappear. This is expected, as the dynamics of the electron coupled to the nitrogen is of stochastic form, i.e. a convex combination of Hamiltonians (the electron's and nitrogen's). Thus, if the nitrogen is initialised completely ($p = 1$), the stochastic errors disappear. The A/C errors are not physically valid on their own, but always come in combination with stochastic errors. So, they vanish together with the stochastic errors. What is left, are Hamiltonian errors (unitary errors), most likely caused by a non-ideal experimental setup (the applied microwaves are not at perfect resonance, for example).

In short, if no echo is used, the produced estimates in the case of a mixed nitrogen nucleus ($p = 0$) are not good, probably due to non-Markovian dynamics occurring. If the nitrogen is initialised with a fidelity of at least $p = 0.95$, the amount of model violation is low and the produced estimates are close to ideal. Furthermore, initialising the nitrogen completely ($p = 1$) removes all stochastic errors in the estimates. This is expected, as the electron's coupling to the nitrogen is of stochastic form.

Table 7.6: The values of N_σ for the simulated Gate Set B model (without echo).

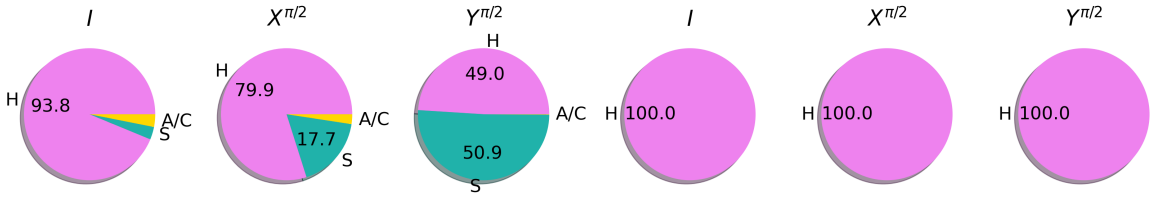
	$p = 0$	$p = 0.95$	$p = 1$
TP	793.29	-10.9863	-20.1789
CPTP	716.795	-10.713	-19.9957

Table 7.7: 1/2-diamond distances ($\cdot 10^2$) from the CPTP simulated model to the three gate ideal model, using the outcomes of GST on Germ Set B (without echo). The three estimated gates are I, $X^{\pi/2}$ and $Y^{\pi/2}$. The nitrogen nucleus was initialised into the $m_I = 0$ state with three different fidelities: $p = 0, 0.95$ and $p = 1$.

	$p = 0$	$p = 0.95$	$p = 1$
I	97.6537	0.2323	0.2249
$X^{\pi/2}$	24.0437	0.1837	0.1091
$Y^{\pi/2}$	22.7691	0.0951	0.0363

7.2.3. Gate set A versus B: effect of XY-4 echo

In one of the previous sections, it was established that, when an amplificationally complete germ set is used in gate set tomography, the NV-simulator produces realistic results. So, comparing two (only) simulated data sets is justified, which is what will be done in this section to study the XY-4 echo (Ch.2.2.3). The (simulated, CPTP) Gate Set A (Table 6.1a), whose circuits were run with an XY-4 echo, will be compared to Gate Set B (Table 6.1b), whose circuits were run without the echo. These gate sets are not equal: they have different sets of germs. However, it is assumed that this will not influence



(a) Nitrogen nucleus initialisation fidelity $p = 0.95$.

(b) Nitrogen nucleus initialisation fidelity $p = 1$.

Figure 7.1: Pie charts indicating what percentage of the error generator is captured by the three subspaces (pink: Hamiltonian errors "H", yellow: active/correlation errors "A/C", green: stochastic errors "S"). The gates I, $X^{\pi/2}$ and $Y^{\pi/2}$ are CPTP model fits of Gate Set B (simulated, no echo). The nitrogen nucleus initialisation fidelity is indicated in the secondary captions.

`pyGSTi`'s gate set tomography results significantly, as both germ sets are amplificationally complete and thus should capture all possible errors.

First, the nitrogen nucleus initialisation fidelity $p = 0$ will be considered. Recall the N_σ for the simulated CPTP model's of Gate Set A (Table 7.5a) and Gate Set B (Table 7.6). When an XY-4 echo is used, the N_σ turns out much lower than when no echo is used: -12.6776 versus 716.795. In the previous section, it was concluded that the non-initialised nitrogen nucleus is most likely the cause of the high model violation. Hence, it is likely that the presence of non-Markovian dynamics is insignificant if an echo is used. Thus, even when the nitrogen nucleus is not initialised, the echo most likely prevents non-Markovian dynamics. The discrepancy between the two gate sets in this case is seen in their large diamond distances (Tab.7.8).

Now, consider the $p = 0.95$ and $p = 1$ cases. Here, with and without echo, there is no significant model violation (Tab.7.5a and 7.6), as both N_σ's being ≤ 1 *"indicates an extremely good fit that appears completely trustworthy"* [8]. The N_σ's also do not differ significantly. The difference between the diamond distances produced by the two gate sets (tables 7.3a 7.7) is non-significant: it is maximally 0.0023, i.e. 0.2% (for the I gate, $p = 1$). Concluding, the quality of the estimated gates does not depend very much on whether an XY-4 echo is used or not, if the nitrogen nucleus is initialized with at least a fidelity of $p = 0.95$. However, it is interesting to see whether the types of errors causing the estimations to deviate from ideal depend on the XY-4 echo. So, now the error generators of the $p = 0.95$ and $p = 1$ cases will be researched. The error generators of both models are of the same order of magnitude (as their diamond distances are also of the same size).

If the nitrogen nucleus is initialized with $p = 1$, then for both models, the error generator lies for 100% in the Hamiltonian subspace for all estimated gates. However, if the initialization fidelity is $p = 0.95$, then the contribution of each subspace differs in the two models. How the error generator is distributed over the three subspaces in the case of an XY-echo is visualised in Fig.7.2. Comparing this to Fig.7.1a, it can be seen that the XY-4 echo gets rid of (almost) all stochastic and active/correlation errors. Apparently, the echo prevents the electron from coupling to the nitrogen nucleus at all. This claim is supported by the fact that even if the nitrogen is not initialised at all ($p = 0$), the error generators of the estimated gates (with XY-4 echo) are totally captured by the Hamiltonian sub space (99.9%, 99.9%, 99.5% for the I, $X^{\pi/2}$, $Y^{\pi/2}$ gates, respectively).

To conclude this chapter, it was found likely that the XY-4 echo prevents non-Markovian dynamics from taking place, even if the nitrogen nucleus is in a completely mixed state. Furthermore, it was found that the XY-4 echo eliminates stochastic errors for all researched initialisation fidelities, even for $p = 0$. So, the XY-4 echo is extremely good at preventing the electron from coupling to the nitrogen, as hoped.

Table 7.8: 1/2-diamond distances ($\cdot 10^2$) between Gate Set A (with echo) to Gate Set B (without echo), both CPTP and simulated. The three estimated gates are I, $X^{\pi/2}$ and $Y^{\pi/2}$. Note that Gate Set A and B have the same fiducials, but different germs. The simulations were done for different initialisation fidelities of the nitrogen nucleus ($p = 0, 0.95, 1$), corresponding to different columns.

	$p = 0$	$p = 0.95$	$p = 1$
I	97.6831	0.3689	0.3061
$X^{\pi/2}$	23.7244	0.159	0.1187
$Y^{\pi/2}$	22.6081	0.2712	0.2859

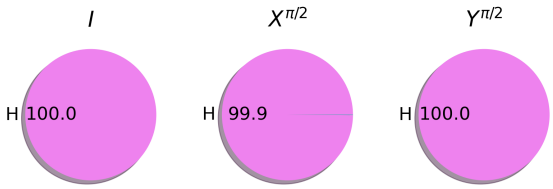


Figure 7.2: Pie charts indicating what percentage of the error generator is captured by the three subspaces (pink: Hamiltonian errors "H", yellow: active/correlation errors "A/C", green: stochastic errors "S"). The gates I, $X^{\pi/2}$ and $Y^{\pi/2}$ are CPTP model fits of Gate Set A (simulated, with echo). The nitrogen nucleus is initialised with a fidelity of $p = 0.95$.

8

Discussion and Outlook

8.1. XY-4 echo

Gate sets A (Tab.6.1a) and C (Tab.6.1c), found in Chapter 6.1, both employ the XY-4 echo (Ch.2.2.3). However, the XY-4 echo was not included in the circuits' sequences. In GST's perspective, to some extent this can be considered cheating. After all, one of GST's unique characteristics is that it does not assume any applied gate to be perfectly ideal. Now however, by not including the XY-4 echo, we pretended as if it was perfect (an exact identity gate, but of longer duration). So, Gate set A's results were analysed again, but now including the XY-4 echo in the circuit sequences. This posed a couple of problems.

First, it was not possible to compare the initial data set (so with *implicit* echo) to a model with *explicit* echo, because of the way a GST experiment is designed. Recall that GST repeats germs according to their approximate germ depth (Ch.4.3.1). The echo was made explicit in the target model by including the XY-4 echo after every gate in the circuit. This however, changed the depths of the target model's germ and fiducial sets, which resulted in a different experiment design, i.e. the target model circuits differed from the input (run) circuits. As `pyGSTi` does not consider *how* a circuit looks, but considers its position in the circuit list, the target model and the experimental model circuits did not match anymore. Solving this problem, i.e. forcing `pyGSTi` to look at the components of a circuit instead of its position on the list can be considered a topic for future development. It would make communication between researchers, with for example the same circuits, but in a different order, easier.

New circuit lists had to be made and run on the simulator. Recall the XY-4 sequence, $X^{\pi}Y^{\pi}X^{\pi}Y^{\pi}$, and that the time between the gate and the first X^{π} is $\tau/2$, whereas the time between the gates in the sequence is τ (Ch.2.2.3). Thus, we tried to implement an additional gate in the target model, G_{τ}, whose target gate is the identity I gate. It could be added to the circuits to represent the $\tau/2$ and τ "waiting" times. However, this proved to be more difficult than it sounds, as `pyGSTi` did not want to accept two gates that both have the identity gate as target. So, instead of modelling the echo with this G_{τ} gate, after every $X^{\pi/2}$ and $Y^{\pi/2}$ gate the XY-4 sequence ($IX^{\pi}IIY^{\pi}IIX^{\pi}IIY^{\pi}I$) was applied. Moreover, the original identity gate was swapped for the XY-4 sequence itself (so $I \rightarrow IX^{\pi}IIY^{\pi}IIX^{\pi}IIY^{\pi}I$). Now, circuits employing this representation of the echo were run and analysed. Whereas previously the longest germ had depth 6, now it had depth 78 (13×6, 12 being the XY-4 sequence's depth). For all the germs to be repeated at least once in the circuits, the analysis had to be conducted until at least $L = 124$. Unfortunately, it was decided that the benefits of such an analysis did not outweigh the downside, the extremely long running-time.

In the future, gate set tomography on the XY-4 echo can be conducted, modelling it as an identity gate: $X^{\pi}Y^{\pi}X^{\pi}Y^{\pi} \rightarrow I'$. As mentioned in Ch.2.2.3, sometimes imperfections in dynamical decoupling sequences can destroy the qubit's state instead of protecting it. GST could be useful to research this topic. However, now the individual X^{π} and Y^{π} gates will not be estimated; only the whole echo. This is caused by the fact that, in this formulation of the echo, the gates are not generated by a time-*in*dependent Lindbladian L anymore (Ch.2.15), because of the different waiting times between the gates. The single gates do not satisfy the assumption of Markovian dynamics (Def.2.6) anymore, and are therefore not captured by the model well (unlike the whole I' echo). An advantage of this approach is that the running time of `pyGSTi` will not increase significantly.

8.2. Error bars

Another thing the attentive reader might have noticed, is the lack of error bars in this thesis. The quantities used for analysing GST's results, like diamond distance and percentages of the error generator subspaces, are not produced with "traditional" error bars. These quantities are gauge-dependent, and the gates used to compute them simply do not come with conventional uncertainties. The topic of error bars in gate set tomography is, according to its developers, definitely not a closed chapter yet, and in need of further research and development. How far it is now can be found in, like always, Nielsen *et al* [8].

8.3. Outlook

In the results chapter (Ch.7), a few suggestions for further research on gate set tomography on the NV center were given. Here, the suggestions, as well as a few additions are presented.

First, the parser for conducting `pyGSTi` used in this thesis needs to be adjusted, such that the null sequence (no gates: $\emptyset$) can be added to the sets of germs and fiducials. Secondly, the adequacy of the NV-simulator has to be researched if no XY-4 is used, by conducting GST on an experimental and simulated data set without XY-4 echo. In this thesis, two gate sets with differing germ sets were compared to investigate the effect of the XY-4 echo. However, needless to say, the comparison would be more well grounded if the gate sets had the same germ set. Hence, it is advised to carry out such an analysis in the future. When it becomes possible to get the process maps of the gates directly out of the NV center simulator, the "direct" gates and the gates estimated by GST can be compared to investigate GST's performance. Moreover, currently, it is not known whether the Levenberg-Marquardt method is the best method for gate set tomography optimization. Further research into this topic could be considered.

As mentioned in the results section, it would be interesting to conduct gate set tomography on the NV center, where the electron functions as qubit and the nitrogen nucleus is modelled as a qutrit. This way, it truly can be probed whether the electron coupling to the nitrogen nucleus was mostly responsible for non-Markovian dynamics. Ideally, the estimated gates would be tensor products of the individual systems' gates. Adding too many qubits to the system quickly becomes unmanageable for `pyGSTi`. So, instead of adding qubits/qutrits from the same NV center, it would be interesting to perform GST on qubits from different NV centers. This could aid the development of quantum networks composed of multiple NV centers. Continuing this line of thinking, the effects of echo's acting on multiple qubits could be investigated then as well.

9 Conclusion

In this thesis, gate set tomography has been conducted on the nitrogen vacancy center in diamond, whose electron served as a qubit. The effects of the surrounding nitrogen nucleus on GST's results have been researched. How the presence of a dynamical decoupling XY-4 sequence affects GST has been investigated as well. To this end, GST has been conducted on three different gate sets. Some circuits were only simulated by the NV center simulator, and some have been run experimentally as well. The nitrogen nucleus ($S = 1$) has three spin states: $m_I = -1, 0, +1$. In the experiments, it was initialised into one of these, or not at all (mixed state initialisation). In the simulations, the nitrogen could be initialised into the $m_I = 0$ state with different fidelities: $p = 0, 0.95$ and $p = 1$, $p = 0$ meaning the nitrogen is in a completely mixed state.

The gate set with a non-amplificationally complete germ set was run experimentally and on the simulator, both times without echo. It was found that the experiments and the simulator did not agree very well if the nitrogen was not initialised (mixed state and $p = 0$, respectively). Initialisation (into $m_I = 0$) in the experiments was compared to initialisation fidelity $p = 0.95$ in the simulations. In this case, the results did agree.

A gate set with an amplificationally complete germ set was also simulated and conducted experimentally, but now with an XY-4 echo. It was found that the simulated and experimental results agreed for all different initialisation fidelities of the nitrogen nucleus. To verify the simulator's results without an XY-4 echo further research is needed: from the first gate set (without an AC germ set), it cannot be concluded that the simulation does not work for $p = 0$, as this shortcoming may be caused by its non-AC germ set. Despite of this, the simulator's results were assumed suitable for further analysis.

A simulated (only) gate set's results, without XY-4 echo, were analysed to investigate the effects of different nitrogen nucleus initialisation fidelities. If the nitrogen was not initialised ($p = 0$), it was found very likely that significant non-Markovian dynamics took place, violating the GST model. If the nitrogen was initialised ($p = 0.95, 1$), no significant model violation was detected. Furthermore, the full initialisation ($p = 1$) removed all stochastic type errors of the gates' error generators. This was excepted, as the coupling of the electron to the nitrogen is of this form.

Lastly, two simulated gate sets, one without and one with an XY-4 echo, were compared. For $p = 0$, it was discovered that the amount of model violation decreased enormously if an XY-4 echo was used. Thus, it is likely that the echo combats non-Markovian dynamics very well. For $p = 0.95$ and $p = 1$, the estimates of the two gate sets did not differ significantly. However, for $p = 0.95$, the error generators of the gate set with echo were captured for 100% by the Hamiltonian subspace, whereas without echo some errors laid in the stochastic sub space. Moreover, even for $p = 0$, the errors were 100% Hamiltonian errors in the presence of an echo. So, it was concluded that the echo prevents the electron from coupling to the nitrogen nucleus extremely well.

9.1. Acknowledgements

First and foremost, I would like to thank my supervisors Dr. Johannes Borregaard and Dr. David de Laat. Without them and their help, this project would not have been possible, and I am privileged to have had them as my supervisors. Johannes' extensive knowledge on quantum mechanics has been

indispensable. David's maths explanations, of whatever small or big concept, have taught me a lot. Dr. Jiwon Yun has been essential to this thesis, as he is the mastermind behind the NV center and its simulator who ran all the circuits. His proactive attitude towards the encountered problems has helped to achieve progress, and I am very thankful for that. Without David Vos, who just finished his bachelor's in Computer Science, this project would not have gotten as far as it has. I am sure he has a bright future ahead, be it related to quantum mechanics or any other field. I would like to thank Dr. Erik Nielsen and his team for creating gate set tomography and helping clarify a couple of questions. Furthermore, I am grateful to Dr. Eliška Greplová and Dr. Matthias Möller for being a member of my thesis committee.

Last but not least, I am forever grateful to my parents, for their endless efforts, support and patience.

Bibliography

[1] Richard P Feynman. *Statistical mechanics: a set of lectures*. CRC press, 2018.

[2] James L Park. "The concept of transition in quantum mechanics". In: *Foundations of physics* 1.1 (1970), pp. 23–33.

[3] Max Born and Pascual Jordan. "Zur quantenmechanik". In: *Zeitschrift für Physik* 34.1 (1925), pp. 858–888.

[4] Jonathan P Dowling and Gerard J Milburn. "Quantum technology: the second quantum revolution". In: *arXiv preprint quant-ph/0206091* (2002).

[5] Richard P Feynman. "Simulating physics with computers". In: *Feynman and computation*. CRC Press, 2018, pp. 133–153.

[6] David P DiVincenzo and Daniel Loss. "Quantum information is physical". In: *Superlattices and Microstructures* 23.3-4 (1998), pp. 419–432.

[7] Gregory AL White. "Gate set tomography is not just hyperaccurate, it'sa different way of thinking". In: *Quantum Views* 5 (2021), p. 60.

[8] Erik Nielsen et al. "Gate set tomography". In: *Quantum* 5 (2021), p. 557.

[9] Xiao Xue et al. "Quantum logic with spin qubits crossing the surface code threshold". In: *Nature* 601.7893 (2022), pp. 343–347.

[10] HP Bartling et al. "Coherence and entanglement of inherently long-lived spin pairs in diamond". In: *arXiv preprint arXiv:2103.07961* (2021).

[11] CE Bradley et al. "Robust quantum-network memory based on spin qubits in isotopically engineered diamond". In: *arXiv preprint arXiv:2111.09772* (2021).

[12] SLN Hermans et al. "Qubit teleportation between non-neighbouring nodes in a quantum network". In: *Nature* 605.7911 (2022), pp. 663–668.

[13] Michael A Nielsen and Isaac Chuang. *Quantum computation and quantum information*. 2002.

[14] David J Griffiths and Darrell F Schroeter. *Introduction to quantum mechanics*. Cambridge university press, 2018.

[15] *Quirk*. `https://algassert.com/quirk`. 2022.

[16] David Vos. "Gate Set Tomography for Nitrogen-Vacancy Systems". In: (2022).

[17] Thomas Theurer et al. "Quantifying operations with an application to coherence". In: *Physical review letters* 122.19 (2019), p. 190405.

[18] QuTech. *Operations on NV center qubits*. `https://www.qutube.nl/quantum-computer-12/operations-on-nv-center-qubits-175`. 2022.

[19] Alexandre M Souza, Gonzalo A Alvarez, and Dieter Suter. "Effects of time-reversal symmetry in dynamical decoupling". In: *Physical Review A* 85.3 (2012), p. 032306.

[20] Alexandre M Souza, Gonzalo A Álvarez, and Dieter Suter. "Robust dynamical decoupling". In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 370.1976 (2012), pp. 4748–4769.

[21] Daniel Manzano. "A short introduction to the Lindblad master equation". In: *Aip Advances* 10.2 (2020), p. 025106.

[22] Robin Blume-Kohout et al. "A taxonomy of small Markovian errors". In: *arXiv preprint arXiv:2103.01928* (2021).

[23] Stephen Wright, Jorge Nocedal, et al. "Numerical optimization". In: *Springer Science* 35.67-68 (1999), p. 7.

[24] Aardal, Van Iersel, and Janssen. *Lecture Notes AM2020 Optimization*. 2020.

[25] Erik Nielsen et al. *pyGSTi*. `https://doi.org/10.5281/zenodo.6363115`. 2022.

[26] Erik Nielsen et al. "Probing quantum processor performance with pyGSTi". In: *Quantum science and technology* 5.4 (2020), p. 044002.

[27] Robin Blume-Kohout et al. "Robust, self-consistent, closed-form tomography of quantum logic gates on a trapped ion qubit". In: *arXiv preprint arXiv:1310.4492* (2013).

[28] Biswa Datta. *Numerical methods for linear control systems*. Vol. 1. Academic Press, 2004.

[29] Robin Blume-Kohout et al. "Demonstration of qubit operations below a rigorous fault tolerance threshold with gate set tomography". In: *Nature communications* 8.1 (2017), pp. 1–13.

[30] Fetsje Bijma, Marianne Jonker, and Aad van der Vaart. "An introduction to mathematical statistics". In: *An Introduction to Mathematical Statistics*. Amsterdam University Press, 2017.

[31] Erik Nielsen. Personal communication. July 6, 2022.

[32] Fedor Jelezko and Jörg Wrachtrup. "Single defect centres in diamond: A review". In: *physica status solidi (a)* 203.13 (2006), pp. 3207–3225.

[33] Marcus W Doherty et al. "The nitrogen-vacancy colour centre in diamond". In: *Physics Reports* 528.1 (2013), pp. 1–45.

[34] CE Bradley. "Order from Disorder: Control of Multi-Qubit Spin Registers in Diamond". PhD thesis. Delft University of Technology, 2021.

[35] Jiwon Yun. Personal communication. June 1, 2022.

[36] Matthias Steiner. "Towards diamond-based quantum computers". MA thesis. 2009.

[37] Michael Marc Wolf et al. "Assessing non-Markovian quantum dynamics". In: *Physical review letters* 101.15 (2008), p. 150402.

[38] MHMA Abobeih. "From atomic-scale imaging to quantum fault-tolerance with spins in diamond". PhD thesis. Delft University of Technology, 2021.

[39] TNE Greville. "Some applications of the pseudoinverse of a matrix". In: *SIAM review* 2.1 (1960), pp. 15–22.

A

Appendix A

A.1. Quantum mechanics

A.1.1. Bloch sphere

Suppose a qubit in the following *superposition*: $\left|\psi\right\rangle = \alpha\left|0\right\rangle + \beta\left|1\right\rangle$, $\alpha, \beta \in \mathbb{C}$. Using their polar form, we rewrite $\alpha = \alpha_0 e^{i\phi_0}, \beta = \beta_1 e^{i\phi_1}$. Multiplying the equation by a *global phase factor* $e^{-i\phi_0}$ results in an identical state, from an observational point of view (as for any measurement operator M_m, for probabilities for outcome m we have: $\left\langle\psi\right|e^{-i\phi_0}M_m^\dagger M_m e^{i\phi_0}\left|\psi\right\rangle = \left\langle\psi\right|M_m^\dagger M_m\left|\psi\right\rangle$):

$$e^{-i\phi_0}\left|\psi\right\rangle = e^{-i\phi_0}\alpha_0 e^{i\phi_0}\left|0\right\rangle + e^{-i\phi_0}\beta_1 e^{i\phi_1}\left|1\right\rangle \tag{A.1}$$

$$\left|\psi\right\rangle = \alpha_0\left|0\right\rangle + \beta_1 e^{i\phi}\left|1\right\rangle, \tag{A.2}$$

where $\phi = \phi_1 - \phi_0$. Now, the laws of probability tell us that $|\alpha_0|^2 + \left|\beta_1\right|^2 = 1$, so we use: $\alpha_0 = \cos\theta, \beta_1 = \sin\theta$. θ and ϕ define a point on a three-dimensional sphere with unit radius, the Bloch sphere (Fig.2.1). Furthermore, we know that $\left|\psi(\theta = 0, \phi)\right\rangle = \left|0\right\rangle$ and $\left|\psi(\theta = \pi, \phi)\right\rangle = \left|1\right\rangle$, i.e.:

$$\left|\psi\right\rangle = \cos\frac{\theta}{2}\left|0\right\rangle + e^{i\phi}\sin\frac{\theta}{2}\left|1\right\rangle, \tag{A.3}$$

where $\theta \in [0, \pi]$ and $\phi \in [0, 2\pi]$.

A.1.2. Spin operators

The spin-$\frac{1}{2}$ operators (the Pauli matrices) are given by, where i is the imaginary unit:

$$S_x = \frac{1}{2}\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}, S_y = \frac{1}{2i}\begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix}, S_z = \frac{1}{2}\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}, \tag{A.4}$$

The spin-1 operators ($S = 1$) are:

$$S_x = \frac{1}{\sqrt{2}}\begin{pmatrix}0 & 1 & 0\\ 1 & 0 & 1\\ 0 & 1 & 0\end{pmatrix}, S_y = \frac{1}{\sqrt{2}i}\begin{pmatrix}0 & 1 & 0\\ -1 & 0 & 1\\ 0 & -1 & 0\end{pmatrix}, S_z = \begin{pmatrix}1 & 0 & 0\\ 0 & 0 & 0\\ 0 & 0 & -1\end{pmatrix}. \tag{A.5}$$

A.2. Mathematics

Definition A.1. If r is a function from $\mathbb{R}^n$ to $\mathbb{R}^m$, then its Jacobian is the $m \times n$ matrix of first partial derivatives, where $\nabla r_j(x), j = 1, \ldots, m$ is the gradient of r_j:

$$J(x) = \left[\frac{\partial r_j}{\partial x_i}\right]_{\substack{j=1,\ldots,m\\ i=1,\ldots,n}} = \begin{bmatrix}\nabla r_1(x)^T\\ \vdots\\ \nabla r_m(x)^T\end{bmatrix} = \begin{bmatrix}\frac{\partial r_1}{\partial x_1} & \cdots & \frac{\partial r_1}{\partial x_n}\\ \vdots & \ddots & \vdots\\ \frac{\partial r_m}{\partial x_1} & \cdots & \frac{\partial r_m}{\partial x_n}\end{bmatrix} \tag{A.6}$$

A.3. The NV Hamiltonian

The NV electron spin's dynamics are influenced by the subsystems surrounding it. That is, the electron spin couples to the other spins, which can be seen in figure A.1. Here, the electron spin couples (purple line) to the host ^{14}N spin. It also couples to the ^{13}C spins surrounding the NV (red lines). Furthermore, the nuclear spins couple to each other (green and yellow lines). All these interactions are best summarized by the system's Hamiltonian.

The overall Hamiltonian is given by the sum of the Hamiltonian of each subsystem and the Hamiltonians of the interactions between the subsystems [38]:

$$H = H_e + H_N + H_c + H_{eN} + H_{ec} + H_{cc} + H_{Nc} \tag{A.7}$$

Here, H_e is the NV electron spin ground state Hamiltonian, H_N is the Hamiltonian of the ^{14}N nuclear spin, H_c is the ^{13}C spins' Hamiltonian, H_{eN} is the interaction between the electron spin and the ^{14}N nuclear spin, H_{ec} is the interaction between the electron spin and the ^{13}C nuclear spins, H_{cc} is the coupling between the ^{13}C spins and finally, H_{Nc} is the coupling between the ^{14}N and ^{13}C spins. In describing each term of the Hamiltonian (Eq.(A.7)), we will follow Abobeih [38].

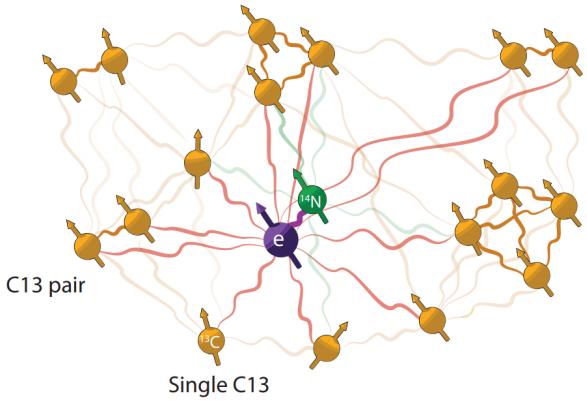


Figure A.1: This figure shows how electron and nuclei of the NV center interact with eacher via coupling. The electron qubit couples to the nitrogen nucleus (purple line) and the surrounding carbon atoms (red lines). The nitrogen nucleus couples to the carbon atoms as well (green lines). Lastly, the carbons couple to each other, where the strength of the interaction depends on their distance and angle with respect to the applied magnetic field. This interaction and its strength is indicated by the yellow lines. Figure is taken from Abobeih [38].

Electron spin

The spin triplet NV electron spin ground state has the following Hamiltonian:

$$H_e = \Delta_{\mathrm{ZFS}} S_z^2 + \gamma_e \left(\mathbf{B} \cdot \mathbf{S}\right) \tag{A.8}$$

Here, $\Delta_{\mathrm{ZFS}} \approx 2.88$ GHz is the zero field splitting, $\gamma_e \approx 2.8$ MHz/G is the electron gyromagnetic ratio. $\mathbf{B} = (B_x, B_y, B_z)$ is the magnetic field and $\mathbf{S} = (S_x, S_y, S_z)$ are the electron spin-1 operators (Eq.(A.5)).

Nitrogen spin

The Hamiltonian of the ^{14}N nuclear spin and its coupling to the electron spin is given by:

$$H_N = -Q_N I_{N,z}^2 + \gamma_n \left(\mathbf{B} \cdot \mathbf{I_N}\right) \tag{A.9}$$

$$H_{eN} = \mathbf{S} \cdot \mathbf{A_N} \cdot \mathbf{I_N} \approx A_{||} \hat{S}_z \hat{I}_{N,z} \tag{A.10}$$

Here, $Q_N \approx 4.98$ MHz is the quadrupole splitting separating the nitrogen-spin $m_I = 0$ and $m_I = \pm 1$ states, discussed in Ch.5.2. $\gamma_n = 0.3077$ kHz/G is the nitrogen-spin gyromagnetic ratio. $\mathbf{I_N} = (I_{N,x}, I_{N,y}, I_{N,z})$ are the spin-1 operators of the ^{14}N nuclear spin and $\mathbf{A_N}$ is the hyperfine tensor for the electron-^{14}N interaction. An approximation in H_{eN} is made, as due to the large zero-field splitting of the electron spin, the hyperfine interaction perpendicular to the NV axis (so the $x-$ and $y-$axes) can be neglected: the secular approximation [36].

Carbon-13 spins
The ^{13}C spins and their interaction with the electron are written as:

$$H_c = \sum_i \gamma_c \mathbf{B} \cdot \mathbf{I}_i \tag{A.11}$$

$$H_{ec} = \sum_i \mathbf{S} \cdot \mathbf{A}_i \cdot \mathbf{I}_i \tag{A.12}$$

Here, $\gamma_c = 1.0705$ kHz/G is the carbon-spin gyromagnetic ratio, $\mathbf{I}_i = (I_{i,x}, I_{i,y}, I_{i,z})$ are the spin-$\frac{1}{2}$ operators (Eq.(A.4)) for the ^{13}C spins and $\mathbf{A}_i$ are the hyperfine tensors for the electron-^{13}C interaction.

Nuclear-nuclear interactions
Lastly, the nuclear spins couple to each other:

$$H_{cc} = \sum_{i,j} \mathbf{I}_i \cdot \mathbf{C_{ij}} \cdot \mathbf{I}_j \tag{A.13}$$

$$H_{Nc} = \sum_j \mathbf{I}_N \cdot \mathbf{C_{N,j}} \cdot \mathbf{I}_j \tag{A.14}$$

$\mathbf{C_{i,j}}$ and $\mathbf{C_{N,j}}$ are the tensors for ^{13}C-^{13}C and ^{14}N- ^{13}C nuclear-nuclear interaction.

A.4. Gate set tomography

Definition A.2. An $m \times n$ matrix A of rank $r > 0$ can be written as the product $A = BC$, where B is $m \times r$ and C is $r \times n$ and both have rank r. The pseudo-inverse of A is given by $A^\dagger = C^T(CC^T)^{-1}(B^TB)^{-1}B^T$ [39].

Definition A.3. Let $\tilde{\mathbb{1}}$ be an $N_{f1} \times N_{f2}$ matrix. If $\sigma \in \mathbb{R}$, $\sigma \geq 0$ and $\mathbf{u}$ and $\mathbf{v}$ are nonzero N_{f1}- and N_{f2}-vectors, respectively, such that:

$$\tilde{\mathbb{1}}\mathbf{v} = \sigma\mathbf{u} \text{ and } \tilde{\mathbb{1}}^{\mathrm{T}}\mathbf{u} = \sigma\mathbf{v}, \tag{A.15}$$

then σ is called a singular value of $\tilde{\mathbb{1}}$.

A.4.1. Multiple germ powers

Here, an example motivating the repetition of germs for multiple germ powers p is given. Suppose g to be a single unitary $x-$rotation by $\pi/2$, G_x, but with an over-rotation by angle ϵ, so that it actually rotates by $\theta = \pi/2 + \epsilon$. If we measure $\langle\langle E_i|\tau(O)|\rho'_j\rangle\rangle$ with $\tau(O) = G_x^p$ for $i = 1 \ldots N_{f1}$, $j = 1 \ldots N_{f2}$ and every circuit is measured N times, then:

$$p\theta \bmod 2\pi = p\pi/2 + p\epsilon \pm \frac{\alpha}{\sqrt{N}} \tag{A.16}$$

$$\theta \bmod 2\pi/p = \pi/2 + \epsilon \pm \frac{\alpha}{p\sqrt{N}}, \tag{A.17}$$

for some constant α. As p increases, it becomes harder to differentiate between angles differing by $2\pi/p$ ("branches"), as this factor becomes smaller. So, to make differentiation between branches possible, the germ is repeated for different powers p.

A.4.2. Choosing the maximum germ depth

The choice of the maximum L depends on three factors. Firstly, increasing L yields more precision as errors are amplified more. However, larger L presents more circuits to run and analyse and thus enlarges the experiment. Lastly, L should not be increased beyond the point where decoherence and stochastic errors dominate. Suppose each gate has rate of decoherence η, then circuits of depth $>> 1/\eta$ ($L > O(1)/\eta$) produce the same equilibrium state and tell us nothing interesting. This maximum circuit depth is called L_η. However, if the rate of stochastic errors varies from gate to gate, then it might be

valuable to let L be germ-dependent, motivated by prior knowledge of the physical system's stochastic noise behaviour. A maximum L is thus chosen from the set of "maximum depths" $\{L_i = 2^{i-1}\}_{i=1}^{N_1}$ such that $L_{\max} = \max_i L_i = L_{N1} \geq L_\eta$. Logarithmically-spaced L are used so that a single set L_i can be used for *all* germs. If instead a set of logarithmically-spaced powers was used, there may be wasted circuits (with depth greater than $L_{\max}$) or short germs that are not repeated the maximum amount of times.

A.4.3. General loglikelihood function

In the case of non-TP gate sets, the likelihood (Eq.(4.17)) is modified. Suppose K Poisson-distributed event probabilities $\{n_j\}$ (random variables) with their own rate parameter λ_j:

$$\Pr(n_j) = \frac{e^{-\lambda_j}\lambda_j^{n_j}}{n_j!}. \tag{A.18}$$

The random variable $X \coloneqq \sum_j n_j$ is Poisson distributed with rate parameter $\lambda \coloneqq \sum_j \lambda_j$:

$$\Pr(X = N) = \frac{e^{-\lambda}\lambda^N}{N!}. \tag{A.19}$$

The n_j are independent, so the probability that a specific set $\{n_j\}_{j=1}^K$ is observed is given by the product of n_j's:

$$\Pr(\{n_j\}) = \prod_j \Pr(n_j) = e^{-\lambda}\prod_j \frac{\lambda_j^{n_j}}{n_j!}. \tag{A.20}$$

From conditional probability follows:

$$\Pr(\{n_j\}|X = N_0) = \frac{\Pr(\{n_j\} \cap \{X = N_0\})}{\Pr(X = N_0)}. \tag{A.21}$$

Calculating the numerator now:

$$\Pr(\{n_j\} \cap \{X = N_0\}) = \begin{cases} \Pr(\{n_j\}) & \sum_j n_j = N_0 \\ 0 & \sum_j n_j \neq N_0 \end{cases} \tag{A.22}$$

The conditional probability becomes:

$$\begin{aligned} \Pr(\{n_j\}|X = N_0) &= \left(e^{-\lambda}\prod_j \frac{\lambda_j^{n_j}}{n_j!}\right)\left(\frac{N_0!}{e^{-\lambda}\lambda^{N_0}}\right) \\ &= \frac{N_0!}{n_1!\ldots n_K!}\left(\frac{\prod_j \lambda_j^{n_j}}{\lambda^{N_0}}\right) \\ &= \frac{N_0!}{n_1!\ldots n_K!}\prod_j \left(\frac{\lambda_j}{\lambda}\right)^{n_j} \\ &= \frac{N_0!}{n_1!\ldots n_K!}\prod_j \left(\frac{\lambda_j}{\sum_k \lambda_k}\right)^{n_j} \end{aligned}$$

This is a multinomial distribution over N_0 trials, with event probabilities $p_j = \frac{\lambda_j}{\sum_k \lambda_k}$. If the gate set is not TP, the total number of counts is not necessarily equal to the observed number of counts. If it is assumed that the predicted number of counts is Poisson-distributed, the above equations exactly describe a GST experiment, with $\{n_j\} = \{p_{s,\beta_s}\}_{\beta_s}$. The *rate* of observing outcome β_s is then $\lambda_{s,\beta_s} \coloneqq N_s p_{s,\beta_s}$. So:

$$\begin{aligned} \mathcal{L}_s &= \prod_{\beta_s} \frac{\lambda_{s,\beta_s}^{N_{s,\beta_s}} e^{-\lambda_{s,\beta_s}}}{N_{s,\beta_s}!} \\ \log \mathcal{L}_s &= \sum_{\beta_s} N_{s,\beta_s} \log\Big(N_s p_{s,\beta_s}\Big) - N_s p_{s,\beta_s} - \log\Big(N_{s,\beta_s}!\Big) \end{aligned} \tag{A.23}$$

The total loglikelihood becomes, again omitting the factor independent of p_{s,β_s}, like in the main body:

$$\log \mathcal{L} = \sum_s \log \mathcal{L}'_s = \sum_{s,\beta_s N_s} N_{s,\beta_s} \log\Big(p_{s,\beta_s}\Big) - N_s p_{s,\beta_s} \tag{A.24}$$

A.5. Results

Table A.1: 1/2-diamond distances ($\cdot 10^2$) from the Gate set A (with echo) simulated and experimental TP models to the three gate ideal model. The model is indicated in the secondary captions. The three estimated gates are I, $X^{\pi/2}$ and $Y^{\pi/2}$. The simulations were done for different initialisation fidelities of the nitrogen nucleus ($p = 0, 0.95, 1$) into the $m_I = 0$ state, corresponding to different columns.

(a) Simulated.

	$p = 0$	$p = 0.95$	$p = 1$
I	0.37	0.4531	0.4673
$X^{\pi/2}$	1.6837	1.4739	1.4829
$Y^{\pi/2}$	1.5012	1.6575	1.6793

(b) Experimental.

	mixed	$m_I = 0$	$m_I = -1$	$m_I = +1$
I	0.3524	0.3385	0.6942	0.5585
$X^{\pi/2}$	1.6737	4.1737	4.7231	3.2938
$Y^{\pi/2}$	1.8772	5.5405	6.6652	3.8105

Table A.2: 1/2-diamond distances ($\cdot 10^2$) from the TP simulated model to the three gate ideal model, using the outcomes of GST on Germ Set B (without echo). The three estimated gates are I, $X^{\pi/2}$ and $Y^{\pi/2}$. The nitrogen nucleus was initialised into the $m_I = 0$ state with three different fidelities: $p = 0, 0.95$ and $p = 1$.

	$p = 0$	$p = 0.95$	$p = 1$
I	102.5643	0.2473	0.2355
$X^{\pi/2}$	20.5605	1.4232	1.5492
$Y^{\pi/2}$	19.5706	1.3681	1.4606

Table A.3: 1/2-diamond distances ($\cdot 10^2$) from the experimental and simulated models to the ideal gates (TP Germ Set C, without echo). The model constraint is indicated in the secondary captions. The five estimated gates are I, $X^{\pi/2}$, $Y^{\pi/2}$, X^{π} and Y^{π}. The circuits were run either initialising the nitrogen nucleus into the $m_I = 0$ state, or not (mixed). In the simulations, the nitrogen was initialised into the $m_I = 0$ state with different fidelities ($p = 0, 0.95, 1$).

(a) Experimental model.

	mixed state	$m_I = 0$ state
I	99.0067	0.5766
$X^{\pi/2}$	31.2822	6.63
$Y^{\pi/2}$	31.5554	6.8583
X^{π}	19.9804	10.8493
Y^{π}	23.8169	10.6076

(b) Simulated model.

	$p = 0$	$p = 0.95$	$p = 1$
I	94.8903	0.3406	0.227
$X^{\pi/2}$	28.0941	1.0362	1.1058
$Y^{\pi/2}$	43.4257	0.5381	1.0332
X^{π}	21.7444	2.4342	1.4528
Y^{π}	19.2569	1.0878	1.4545