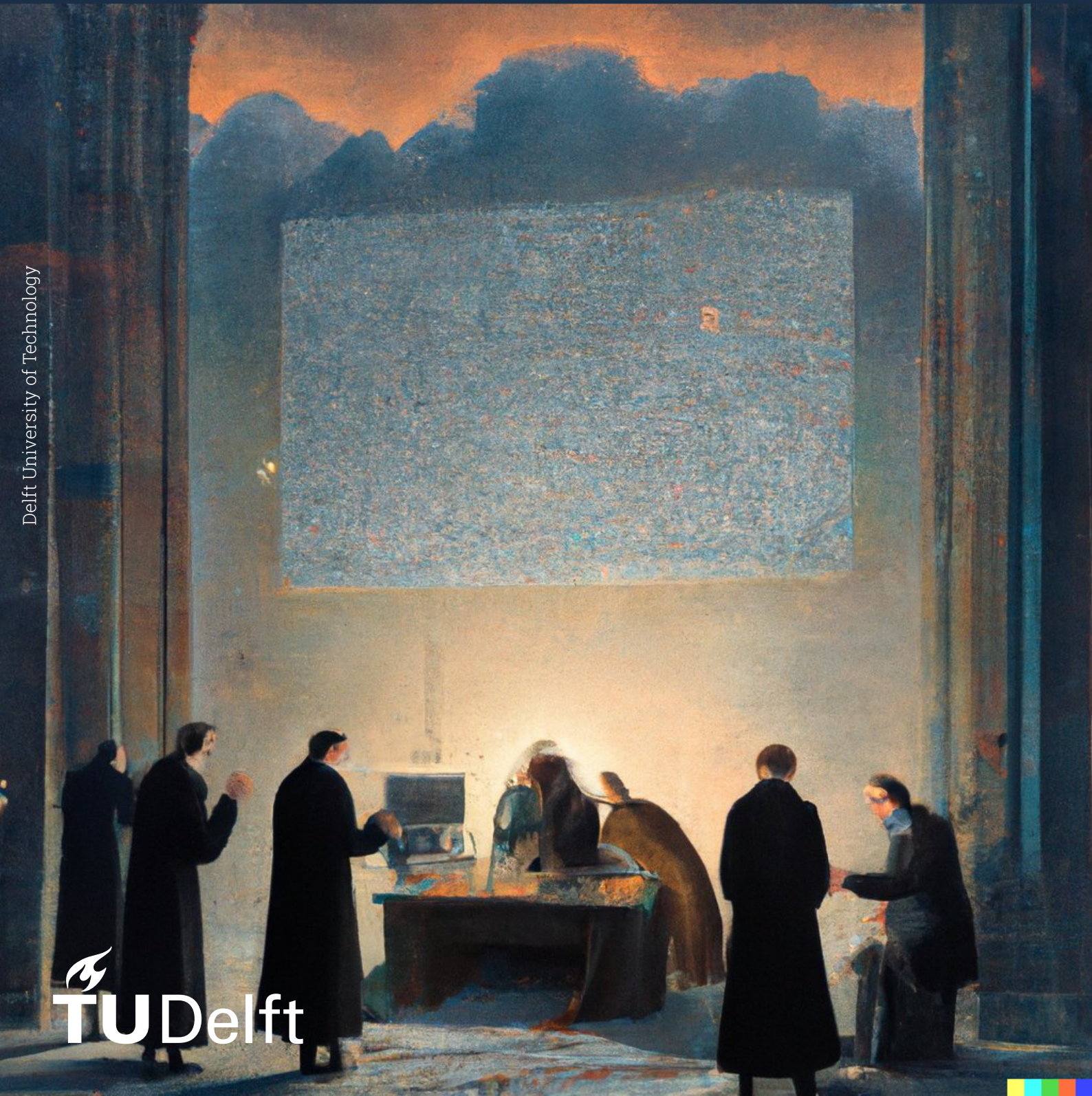


Characterizing qutrits and their noise environments

Master's Thesis
Joseph Barreto

Delft University of Technology



Characterizing qutrits and their noise environments

by

Joseph Barreto

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Thursday March 11, 2021 at 3:00 PM.

Student number:	5041422	
Project duration:	July 2020 – February, 2021	
Thesis committee:	Prof. Johannes Borregaard,	TU Delft, supervisor
	Prof. Leo Di Carlo,	TU Delft
	Prof. Barbara Terhal,	TU Delft
	Prof. David DiVincenzo,	RWTH Aachen, TU Delft

Cover:	"Operating a quantum computer" as rendered by DALL-E 2, sourced from Iftach Yakar @QuantumYakar
Style:	TU Delft Report Style, with modifications by Daan Zwaneveld

An electronic version of this thesis is available at <http://repository.tudelft.nl/>.

Preface

Before diving into the report, I would like to spend a few words in thanking several individuals who, in various ways, supported me throughout my thesis.

I would like to thank my supervisor, Johannes, for his guidance throughout the project. Between starting his professorship and dealing with the tumults of the pandemic, I am grateful for the time he devoted to me each and every week, and I am proud to have seen the group grow so much in its first year.

It is not an overstatement to say that I owe my sanity to my girlfriend and my reliable group of friends, known on the street as the "CuteBits". Working on a theoretical project from a tiny apartment, I spent most of my waking hours trapped within the same four walls. From the Tichu League to climbing, coffee walks to snowball fights, you rightfully pulled me from my laptop on many an occasion, and have helped me more than you know.

Lastly, to my parents and brother, who supported me in my decision to move abroad and whose Zoom-enabled support has been steadfast.

Delft, March, 2021

Abstract

The standard toolbox of modeling and characterizing quantum systems comes with a standard set of assumptions as well. The two-level approximation replaces a many-level system with a qubit, and the Markovian approximation assumes an environment with a short memory. In this thesis, these assumptions are relaxed, and the dynamics of a single qutrit are reconstructed from a complete set of measurement data, using maximum-likelihood estimation (MLE) to self-consistently infer a set of state-preparation and measurement parameters (SPAM), along with a time-dependent process map. The process map can then be used to quantify the non-Markovianity of the qutrit evolution. The SPAM parameters and process maps produced by the MLE framework are compared to ground-truth simulations, with good agreement found in all cases studied. A Markovian example, the amplitude and phase damping channel, and a non-Markovian example, two transmons with a static coupling, are investigated. With its ability to directly capture higher level effects such as leakage errors, and also to detect non-completely positive evolution due to entanglement with the environment, this framework improves upon existing characterization algorithms with the purpose of encouraging future experimental work with qutrits.

Contents

Preface	i
Abstract	ii
1 Motivation	1
2 Prior Work	3
2.1 Qutrits	3
2.2 Markovianity	4
3 Background	5
3.1 Qubits and Qutrits	5
3.2 Time Evolution	7
3.2.1 Quantum Operations	7
3.2.2 Open Quantum Systems	7
3.2.3 Time Evolution as a Quantum Operation	8
3.2.4 Recipe for Obtaining Lindblad Form	10
3.2.5 Alternate Derivation	11
3.2.6 Markovianity	11
3.2.7 Quantifying Markovianity	12
3.2.8 Markovianity - Loose Ends	13
3.3 Quantum Measurement	14
3.4 Characterization	15
3.4.1 Quantum State Tomography	15
3.4.2 Quantum Process Tomography	16
3.4.3 Gate Set Tomography	18
3.4.4 Randomized Benchmarking	18
3.5 Superconducting Quantum Hardware	19
3.5.1 Transmons	19
3.5.2 Asymmetric Transmon Hamiltonian	21
3.5.3 MIT Device	21
3.5.4 Full Hamiltonian	23
3.5.5 Two-Qubit Gates	23
3.5.6 ZZ-Interaction	23
3.5.7 Controlled-Z Gate	25
4 Results	28
4.1 Estimating SPAM Errors	28
4.1.1 State Preparation	28
4.1.2 Measurement	32
4.1.3 Measuring in Other Bases	33
4.1.4 Caveat	34
4.1.5 Maximum Likelihood Estimation	34
4.2 Verifying the Estimation	35
4.3 Estimating the Quantum Operation	39
4.3.1 Maximum Likelihood Estimation	40
4.3.2 Amplitude and Phase Damping	40
4.3.3 Transmon Device Evolution	46
4.4 Measuring Markovianity	50
4.4.1 Amplitude and Phase Damping	50
4.4.2 Transmon Device Evolution	52

4.5	Analysis	57
5	Conclusion	59
5.1	Future Directions and Potential Improvements	59
5.2	Summary	60
	References	61
A	Quantum Maps	66
A.1	Vectorization	66
A.2	Vectorized Lindblad Equation	67
A.3	Choi Matrix	68
A.4	Pauli Transfer Matrix	69
A.5	Canonical Lindblad Form	70
B	Linear Algebra	71
B.1	Singular Value Decomposition	71
B.2	Diagonalization	71
B.3	Spectral Theorem	71
B.4	Cholesky Decomposition	72
B.5	Matrix Norms	72
B.5.1	Trace Distance	72
B.5.2	Spectral Norm	72
B.5.3	Diamond Norm	73
C	Rotating Frames	74
C.1	Interaction Picture Device Hamiltonian	75
D	Additional Datasets	76
D.1	Coherent Error Scale = 0.01	76
D.2	Coherent Error Scale = 0.05	79
D.3	Coherent Error Scale = 0.25	80

1

Motivation

The goal of this section is to motivate *why* this work is worthwhile and relevant to field of quantum computing, and to that end, I have written the section in a fairly self-contained and semi-technical manner.

Quantum computing is rapidly emerging as one of the most promising technologies of our time. The current state of hardware has been dubbed “noisy intermediate-scale” [49] to denote that though quantum error-correction and fault-tolerance are at least a few breakthroughs away, researchers have managed to scale quantum devices to tens of qubits, with 100-qubit systems on the horizon. While novel qubit designs and alternative hardware paradigms remain intense fields of research in the lab, industry players have joined the fray, pouring resources into the full-stack development and system engineering approach to quantum computers, encompassing quantum compilers, software frameworks, user interfaces, and cloud access. Specialized variational algorithms for practical problems in chemistry and optimization that can faithfully run on imperfect hardware have proliferated in the last two years, and informative yet efficient characterization and benchmarking techniques are in high demand. Clearly, quantum computing is here to stay, and while its proponents work diligently towards scalable systems that can offer true quantum advantage, it is equally important to get as much as we can from quantum computers as they stand today. For both the functionality of quantum error-correction, which can only offer noise suppression below a certain maximum threshold, and for the execution of deeper quantum circuits, it is critical to understand and mitigate quantum “bugs”, encompassing the loss of quantum information from the system via decoherence and imperfections in qubit control, gate design, and system calibration. Though it may seem inconsequential, when it comes to their hardware’s performance, quantum engineers care very much about the difference between 99% and 99.9%.

In modeling and characterizing any physical system, one always makes assumptions to get a solid, if oversimplified, grasp of observed phenomena. Once understood, these assumptions can be progressively relaxed to bring one’s understanding closer to reality at the price of increased complexity. Attending to the details of quantum computers is no different, and even in systems that are generally well-understood, we must be meticulous to diagnose remaining sources of error. Two assumptions that are commonly made in treating quantum computers will be relaxed throughout this work. The first assumption is known as Markovianity, which relates to certain properties of the noise experienced by a system of qubits in a realistic “lab” setting. The second assumption refers to the word “qubit” itself, which denotes the two logical states 0 and 1 and only approximates the physical system engendering the qubit. These two concepts are now elaborated.

At a high level, the distinction between Markovian and non-Markovian noise is whether or not the qubit environment has a memory. Interaction between the system (the qubit device of interest) and the environment (everything else surrounding it) should be suppressed as much as possible, except, of course, when we want to measure the system and extract information. This interaction can never be completely removed, and the two are always coupled to some extent. If information leaks from the system to the environment in the Markovian way, that information will be very quickly scrambled into myriad environmental degrees of freedom and be effectively irretrievable. Future states of the system will only depend on the particular dynamics occurring at some present time, and not on what happened in the past. If, instead, this information leakage demands a non-Markovian description, it means that

the environment does not scramble what it receives from the system so quickly, and may even import some of that information *back* into the system at a future time. Thus, the time-evolution of a state in the present can, in theory, depend on its state at some time(s) in the past, which has not been “forgotten”. It is no surprise that Markovian noise is much simpler to deal with, and many physical models and characterization techniques make this assumption in their treatment of quantum devices [2, 57]. While this has been fruitful in developing quantum hardware so far, in the quest for high-fidelity systems, we will need to treat noise processes from the more general standpoint of non-Markovianity to improve our understanding and control of quantum hardware.

Another common, and arguably more fundamental, assumption made about qubits is that they really *are* qubits. When people rack their brains about Schrodinger’s cat and how wavefunctions collapse, they are confronted with the duality of dead and alive, or 0 and 1. To the end-users of quantum hardware, qubits obey this dichotomy, but to the physicist, they are often free to take on any number of values. Qubits with three possible states should actually be called qutrits (0, 1, 2), and generally for d states they are called q -dits (0, 1, 2, ..., d). The number of allowed states depends on the physical system in which the qubit is encoded. Though there are interesting protocols and uses for d -level systems [9, 17, 52, 58], quantum computing has typically relied on analogies and comparisons with classical computing, and, thus, most research utilizes qubits. Therefore, excursions out of the computational space defined by the 0 and 1 states represent “forbidden” transitions in the logical operation of the system as a computer, and must be suppressed by careful design and control. These excursions are known as leakage errors, and if some information is lost to these non-computational levels, then it is even possible for that information to re-enter the computational space at a later time. Thus, leakage errors could also result in non-Markovian behavior [56].

In light of these two properties, the goal of this thesis is to directly characterize the time evolution of a quantum system comprised of a qutrit while subject to potentially non-Markovian noise. The methods and code developed are then applied within the context of superconducting hardware. It is my hope and intention that by characterizing a system in this general way, I can help shed light on how and when non-computational states contribute to the failure rate of quantum logic gates, and furthermore, that I can promote the more frequent use of qutrits in quantum information protocols and experiments.

2

Prior Work

In this section, we will summarize some previous work closely related to the topics just introduced. In particular, we focus on experimental protocols using qutrits, and Markovianity measures applied to real hardware.

2.1. Qutrits

Recent years have seen the emergence of more elaborate experiments and proposals that harness states beyond the two-level approximation. Qudits offer many computational advantages, including shortened gate compilations [23], improved error-correction schemes [52, 17, 4], and more secure quantum cryptography [16]. In the realm of superconducting hardware, several groups have produced in-depth characterizations of individual qutrit systems. In 2010, Bianchetti et al. [7] demonstrated quantum state tomography for a complete basis of qutrit states, with an average state fidelity of 95%. Their device was a transmon embedded in a coplanar microwave resonator. In 2015, Peterer et al. [46] performed a more detailed analysis of the energy decay and phase coherence of the first five levels of a transmon embedded in a 3D cavity. While they do not perform any kind of tomography, their analysis of the dominant decay channels is quite informative for the simulations performed later in this thesis.

Going beyond the characterization of the higher states, in 2011 Fedorov et al. [23] reduce the gate depth of a Toffoli gate on 3 transmon qubits by diabatically moving to the avoided crossing between $|11\rangle$ and $|20\rangle$. This “hides” the occupation of the middle qubit in the state $|11x\rangle$ so that only the state $|011\rangle$ picks up a conditional phase at one point during the gate decomposition of the Toffoli. Using the $|2\rangle$ directly in this way, and also conventionally to perform a CZ gate, the authors achieved a Toffoli gate of fidelity 65% with 3 two-qubit/qutrit gates and 2 single-qubit gates, as opposed to 6 two-qubit and 10 single-qubit gates in the usual all-qubit decomposition. Other proposals exist for the realization of qutrit-assisted 3-qubit gates in superconducting hardware [8].

While the above example merely uses the $|2\rangle$ as an intermediary state to aid qubit logic, we now consider two papers that go even further, executing gates and algorithms explicitly designed for qutrit-based processors. In Wu et al. [61], a SWAP gate between the $|0\rangle$ and $|2\rangle$ states of a 3D transmon is demonstrated, where the control waveform for the gate is found by solving a nonlinear optimization problem that minimizes the distance to the target operation and the occupation of the non-computational $|3\rangle$ state. The authors validate their result by applying the gate up to 21 times, fitting the diagonal elements in the Z-basis of the parametrized process to the actual diagonal probabilities. The authors admit that this does not amount to real process tomography because they prepare neither a complete input set, nor measure in a complete basis. They simply hope that the data is enough to constrain the process. In Blok et al. [9], the authors demonstrate a unitary operation that can scramble quantum information in a five-qutrit transmon-based processor. They use the scrambling unitary in a teleportation scheme and achieve a success fidelity greater than 0.5, confirming the scrambling property of their gate. They also perform full process tomography for the scrambling gate and its sub-component, the C-SUM gate, and provide a detailed analysis of the control, coherence, and cross-talk of their qutrit device.

2.2. Markovianity

Now that we have concretely established the utility of qutrits, we turn to several examples of quantum processes that have been characterized from the standpoint of Markovianity. We defer the theoretical discussion of how to quantify Markovianity (section 3.2.6), and instead focus on experimental papers that actually implemented some measure thereof.

Some early Markovianity experiments used photonic platforms [54, 39], showing that the coupling between the polarization (system) and frequency (environment) degrees of freedom of a photon could give rise to controllable non-Markovian dynamics. In these papers, the non-Markovianity is quantified using the trace-distance measure of Bruer et al. [13]. More recently, efforts have shifted towards superconducting platforms. In 2018, Pokharel et al. [47] investigate the use of dynamical decoupling sequences to preserve various initial states over time, with data from IBM and Rigetti devices. A non-monotonic decay of the fidelity indicates non-Markovianity. Interestingly, they find that decoupling compensates for the non-Markovian environment of IBM devices, while it apparently *causes* non-Markovianity to appear in Rigetti devices. In 2019, Morris et al. [43] characterize a 4-qubit IBM device by determining to what extent a gate “remembers” the gate that preceded it, a clearly non-Markovian effect. Their method is agnostic to the noise source and represents only an indirect measure of non-Markovianity based on the context-dependence of a select set of accessible gates. In 2020, the same authors [60] follow up with a complete characterization of non-Markovianity using the recently proposed process tensor framework of Pollock et al. [48]. They decompose a process of fixed length into a set of “basis” processes, which can express both Markovian and non-Markovian evolutions. Using IBM’s hardware, they find a non-zero correlation between state preparation and measurement sequences that are separated by, among other things, a completely depolarizing channel that should prevent such correlations. To explain the persistent correlations, they note that information must have entered the environment (an adjacent qubit) with some memory, and then re-entered the system after the depolarization. Lastly, also in 2020, Urrutia et al. [55] explore the nonconvexity of quantum maps by making non-Markovian channels out of Markovian ones, and vice versa, using both trace distance and Choi matrix-based measures of non-Markovianity. Their methods (for photonic qubits) are similar to those of this work, but they focus heavily on comparing Markovianity metrics, while we highlight one metric in particular as a use case of our methods.

To my knowledge, there has been no explicit characterization of qutrits through the lens of Markovianity to date. The methods in this work are used in the paper “Lindblad Tomography of a Superconducting Qubit”, which is currently in preparation [53].

3

Background

DiVincenzo [21] puts forth a set of seven criteria that any physical realization of a quantum computer should satisfy.

1. A scalable, physical system with well-characterized qubits
2. The ability to initialize to a simple fiducial state
3. Decoherence times much longer than the gate times
4. A universal set of quantum gates
5. A qubit-specific measurement capability
6. The ability to interconvert stationary and flying qubits
7. The ability to faithfully transmit flying qubits

The explicit inclusion of the $|2\rangle$ state relates to point 1, as it helps us understand when our qubits might deviate from a two-level space. The direct characterization of the time-evolution naturally connects to points 3 and 4. If we characterize the system when left idle, only subject to environmental effects, then we learn about potential sources of decoherence. If we characterize the time-evolution during some external drive or gate trajectory, then we probe the ability to implement the gates of some chosen universal set with high fidelity. Furthermore, as we will see later, imperfections in the state initialization and measurement operators can also be included in the process characterization, which incorporates points 2 and 5. Points 6 and 7 are not as relevant, as they pertain to quantum communication and quantum key distribution, which are important for connecting quantum networks. These criteria provide context for which aspects of quantum computing our results may inform.

This section will provide various pieces of theory that are needed to understand the results presented in the following section. The natural flow of quantum computation generally has three steps: state preparation, time evolution, and measurement. Thus, the background information and theory in the following sections will adhere to this ordering. Mainstream characterization methods for these steps are then explained. Lastly, a summary of superconducting device theory is provided.

3.1. Qubits and Qutrits

Any quantum state may be expressed as a density matrix, which is Hermitian, positive-semidefinite, and has trace equal to 1. For a single qubit, the matrix is 2×2 , and the elements of the Pauli group for 1 qubit form a basis for all 2×2 Hermitian matrices¹, of which the density matrices are a subset.

$$X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad Y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} \quad Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \quad (3.1)$$

In this basis, the density matrix ρ may be expressed as

$$\rho = \frac{1}{2}(I + \vec{r} \cdot \vec{\sigma}) = \frac{1}{2} \begin{bmatrix} 1 + r_z & r_x - ir_y \\ r_x + ir_y & 1 - r_z \end{bmatrix} \quad (3.2)$$

¹If the expansion coefficients are allowed to be complex (which is not physical), then the Paulis span *all* 2×2 matrices.

The vector of Pauli operators is $\vec{\sigma} = (X, Y, Z)$, and the Bloch vector components r_i are the expectation values of the Pauli operators P_i of the state ρ : $r_i = \text{Tr}(P_i \rho)$. Since X, Y , and Z are all traceless, this construction guarantees that ρ is a valid density matrix. Being Hermitian, the Paulis can be exponentiated to generate rotations that live in $\text{SU}(2)$. The mathematical form for qutrits is similar, but the basis is formed by the Gell-Mann matrices, which are the generators of $\text{SU}(3)$:

$$\lambda_1 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \lambda_2 = \begin{bmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \lambda_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \lambda_4 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \quad (3.3)$$

$$\lambda_5 = \begin{bmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{bmatrix} \quad \lambda_6 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \quad \lambda_7 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{bmatrix} \quad \lambda_8 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{bmatrix} \quad (3.4)$$

Generally, one can extend this type of construction for any dimension d , forming $\text{SU}(2)$ -type Paulis out of all 2-dimension subspaces, to get a set of generators for $\text{SU}(d)$. For other basis choices in higher dimensions, see Bertlmann and Kramer (2008) [6]. Three $\text{SU}(2)$ subspaces are easily identifiable within the Gell-Mann matrices, being the subspaces spanned by $\{|0\rangle, |1\rangle\}$, $\{|1\rangle, |2\rangle\}$, and $\{|0\rangle, |2\rangle\}$. The Z operators for the 12 and 02 subspaces are obtained by linear combinations of λ_3 and λ_8 . The density matrix of a qutrit has the following form

$$\begin{bmatrix} \rho_{00} & \rho_{01} & \rho_{02} \\ \rho_{01}^* & \rho_{11} & \rho_{12} \\ \rho_{02}^* & \rho_{12}^* & \rho_{22} \end{bmatrix} \quad (3.5)$$

where the main diagonal contains real numbers, and the off-diagonals are complex. Thus there are 8 free parameters, after accounting for the trace constraint. Any qutrit may then be expressed in terms of an 8-dimensional generalized Bloch vector $\vec{\lambda}$ [20]:

$$\rho = \frac{1}{3}(I + \sqrt{3}\vec{r} \cdot \vec{\lambda}) = \begin{bmatrix} 1 + \frac{\sqrt{3}}{2}(\langle\lambda_8\rangle + \sqrt{3}\langle\lambda_3\rangle) & \frac{3}{2}(\langle\lambda_1\rangle - i\langle\lambda_2\rangle) & \frac{3}{2}(\langle\lambda_4\rangle - i\langle\lambda_5\rangle) \\ \frac{3}{2}(\langle\lambda_1\rangle + i\langle\lambda_2\rangle) & 1 + \frac{\sqrt{3}}{2}(\langle\lambda_8\rangle - \sqrt{3}\langle\lambda_3\rangle) & \frac{3}{2}(\langle\lambda_6\rangle - i\langle\lambda_7\rangle) \\ \frac{3}{2}(\langle\lambda_4\rangle + i\langle\lambda_5\rangle) & \frac{3}{2}(\langle\lambda_6\rangle + i\langle\lambda_7\rangle) & 1 - \sqrt{3}\langle\lambda_8\rangle \end{bmatrix} \quad (3.6)$$

The isomorphism between $\text{SU}(2)$ and $\text{SO}(3)$ allows for a convenient visualization of single qubit states on the Bloch sphere, but qutrits lack a simple, low-dimensional representation. This is evident in the following parametrization for a pure qutrit [20]:

$$|\psi\rangle = \sin \frac{\xi}{2} \cos \frac{\theta}{2} |0\rangle + e^{i\phi_{01}} \sin \frac{\xi}{2} \sin \frac{\theta}{2} |1\rangle + e^{i\phi_{01}} \cos \frac{\xi}{2} |2\rangle \quad (3.7)$$

Orthogonal states are not anti-aligned, but correspond to states of maximum opening angle $2\pi/3$ on the 7-dimensional unit sphere S_7 [20]. Interestingly, though all pure states still lie on the surface of this hypersphere, unlike qubits, not all vectors that lie on such a sphere correspond to physical states [6]. It is instructive to also ask what the X and Z qutrit operators should be.

$$Z = \begin{bmatrix} 1 & 0 & 0 \\ 0 & e^{i\pi/3} & 0 \\ 0 & 0 & e^{i2\pi/3} \end{bmatrix} \quad X = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \quad (3.8)$$

These are generally known as the $d = 3$ clock and shift operators [28]. The clock operator is the complex exponential of a matrix whose diagonal keeps track of a clock that increments in unit time steps, and the shift operator is a cyclic permutation of the basis elements by 1 step in space. These kinds of operators are no longer idempotent, but become the identity when raised to the power d . Note that only the qubit X and Z operators are Hermitian. For qutrits, they are of course still unitary, but they do not generate rotations upon exponentiation (they are not elements of the generalized Gell-Mann basis).

Since qutrits are three-level systems, there are three distinct ways that their level structure may be ordered. If one starts with a two-level system, then a third level $|e\rangle$ could be added below, above, or in between the existing two levels. These configurations are called V, ladder, and lambda systems,

respectively, due to the appearance of transitions through the added level. Since the levels throughout this report are energy states, we care about sequential transitions, and so we are always dealing with a ladder system (in which case the level labels are 0, 1, 2, not 0, e , 1).

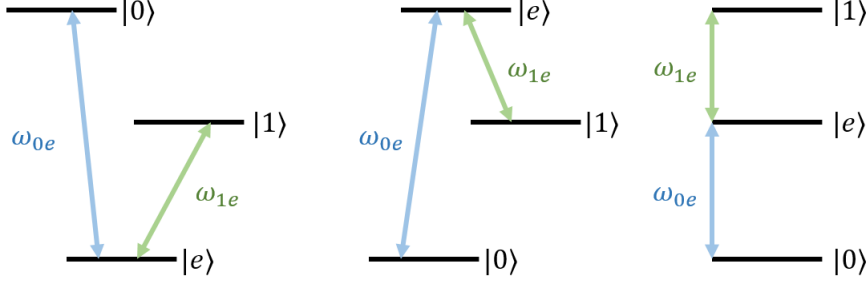


Figure 3.1: V, Λ , and ladder structures for three-level systems.

3.2. Time Evolution

3.2.1. Quantum Operations

The most general map we can apply to a quantum state is called a quantum operation, and we can figure out the general structure of such an operation from some natural expectations of its behavior. The operation obviously should be linear, because quantum mechanics is a linear theory. The operation should also map physical states to physical states, which means that if the input is a valid density matrix, so too should be the output. A valid density matrix is Hermitian, has trace equal to 1, and is positive-semidefinite, so all three properties must be preserved by the action of the operation. Thus, quantum operations are Hermiticity-preserving, trace-preserving (TP), and completely positive (CP). Note that we say not just positive, but *completely* positive. An operator is positive if it preserves the positivity of its inputs, so we at least require this property. However, we also require that if the system is coupled with some environment, and the action of the operation on this environment is trivial (i.e. it does what it is supposed to do on the system, and the identity elsewhere), the resulting output density matrix of the universe must still be positive semi-definite. This stronger property—called complete positivity—is what we actually need. By using the Choi isomorphism [38], these requirements can be used to derive the following structure for a CPTP quantum operation:

$$\mathcal{E}(\rho) = \sum_i K_i \rho K_i^\dagger \quad (3.9)$$

$$\sum_i K_i^\dagger K_i = I \quad (3.10)$$

where the matrices $\{K_i\}$ are the Kraus operators of the quantum operation. See A.3 for more on the Choi matrix. Quantum operations are able to describe unitary gates (where the Kraus set only has one element), and also any other physical but non-unitary evolution, such as the evolution of open quantum systems. The number of Kraus operators needed to describe a channel is $N \leq d^2$, where d is the Hilbert space dimension.

3.2.2. Open Quantum Systems

Let H_S denote the Hilbert space of the *system*, which we have control over, and H_E the Hilbert space of the *environment*, which we do not. Together, $H_S \otimes H_E$ is assumed to be a closed system, and the evolution of a closed system is always given by unitary evolution according to its total Hamiltonian, which may or may not be time dependent. Experimentally, we typically only measure the system and not the environment (or at least, we assume that the outcomes of measurements on the environment are lost), which means our effective view of just the system must average over the environment, which operationally is represented as

$$\rho_S(t) = \text{Tr}_E[U_{t,0}(\rho_S(0) \otimes \rho_E(0))U_{t,0}^\dagger] \quad (3.11)$$

where $\rho_A(t)$ denotes a density matrix in a Hilbert space labeled A at some time t (S for system, E for environment). The unitary $U_{t,0}$ evolves the joint state in $H_S \otimes H_E$ from some initial time (here taken to be $t = 0$) to some final time t . Then we trace out the environment and obtain the state of just the system. Our effective view of the system is no longer unitary. The action of the above process can be worked out further if we assume that the environment begins in a pure state:

$$\rho_S(t) = \sum_k \langle e_k | U_{t,0} (\rho_S(0) \otimes |e_0\rangle\langle e_0|) U_{t,0}^\dagger | e_k \rangle \quad (3.12)$$

$$= \sum_k \langle e_k | U_{t,0} | e_0 \rangle \rho_S(0) \langle e_0 | U_{t,0}^\dagger | e_k \rangle \quad (3.13)$$

$$= \sum_k M_k(t) \rho_S(0) M_k(t)^\dagger \quad (3.14)$$

where $\{|e_k\rangle\}$ denotes an arbitrarily chosen basis for the environment. Projecting out the environment part of the unitary results in the Kraus operators for the effective evolution of just the system. Since we trace out the environment, it may be that two very different environments lead to the same effective dynamics of the system. After taking the partial trace over the environment, if there is more than one Kraus operator in the expansion, then the output of the quantum channel looks mixed, and there must have been some kind of entanglement with the environment that was traced over (assuming the state of the system was pure initially).

3.2.3. Time Evolution as a Quantum Operation

If we vectorize (A.1) the quantum operator-sum, we get the following linear equation for $\text{vec}(\rho) = \vec{\rho}$:

$$\text{vec}(\mathcal{E}(\rho)) = \sum_i \text{vec}(K_i \rho K_i^\dagger) \quad (3.15)$$

$$= \sum_i (K_i^* \otimes K_i) \vec{\rho} \quad (3.16)$$

$$= \Phi \vec{\rho} \quad (3.17)$$

where Φ is the matrix form of the superoperator $\mathcal{E}$ in the vectorized representation. If we vectorize the particular Kraus sum representing time evolution from 0 to t , we obtain a linear equation for the state of the system

$$\vec{\rho}(t) = \Phi_t \vec{\rho}(0) \quad (3.18)$$

where the vectorized notation will be suppressed unless stated otherwise. Everything we want to know about our dynamics is encoded in the mathematical properties of Φ_t . Since the underlying quantum map is valid, the superoperator is also valid, in the sense that Φ_t preserves the Hermiticity of input density matrices, and is also CPTP. Now we will assume that the inverse of Φ_t exists for all times², which allows us to say two things. First, we can define a two-parameter family of maps by the relation [12]

$$\Phi_{t,s} = \Phi_t \Phi_s^{-1} \quad t \geq s \geq 0 \quad (3.19)$$

where $\Phi_{t,0} = \Phi_t$, and $\Phi_{t,s} = \Phi(t, s)$. Since the inverse exists for all positive times, we can write

$$\Phi_{t,0} = \Phi_{t,s} \Phi_{s,0} \quad (3.20)$$

Now, $\Phi(t, s)$ may not be CP or even P because it depends on the inverse of a CP map Φ_s^{-1} , which may not be P or CP. Thus, we can define two forms of divisibility which will be useful in framing Markovianity later on:

- If $\Phi(t, s)$ is CP for all $t \geq s \geq 0$, then the map is *CP-divisible*.

²A process would not be invertible if two distinct initial quantum states evolved to the same final state in finite time, instead of asymptotically [27].

- If $\Phi(t, s)$ is P for all $t \geq s \geq 0$, then the map is *P-divisible*.

Next, we can use the existence of the inverse to obtain a time-local equation for the evolution of the system:

$$\dot{\rho}(t) = (\dot{\Phi}_t)\rho(0) = \dot{\Phi}_t\Phi_t^{-1}\rho(t) \equiv \mathcal{K}(t)\rho(t) \quad (3.21)$$

The operator $\mathcal{K}(t)$ generates the time evolution of the system, so the evolution operator Φ can be written as

$$\Phi(t, s) = \mathcal{T} \exp \left(\int_s^t dt' \mathcal{K}(t') \right) \quad (3.22)$$

and it can be shown that under the following conditions on $\mathcal{K}(t)$

$$[\mathcal{K}(t), A]^\dagger = \mathcal{K}(t)A^\dagger \quad \text{Hermiticity-preserving} \quad (3.23)$$

$$\text{Tr}_S\{\mathcal{K}(t)A\} = 0 \quad \text{trace-preserving} \quad (3.24)$$

that $\mathcal{K}(t)$ has the following form (derived from the time-convolutionless projection operator technique [14, 15])

$$\mathcal{K}(t)\rho_S = -i[H_S(t), \rho_S] + \sum_i \gamma_i(t)[A_i(t)\rho_S A_i^\dagger(t) - \frac{1}{2}\{A_i^\dagger(t)A_i(t), \rho_S\}] \quad (3.25)$$

We will henceforth refer to Eq. 3.25 as the time-dependent master equation (TDME). This structure guarantees Hermiticity and trace preservation, but does *not* guarantee positivity or the semigroup property. However, if we stipulate that the rates are positive and constant for all times

$$\gamma_i(t) = \gamma_i \geq 0 \quad (3.26)$$

and that H_S, A_i are also time-independent, then we recover the celebrated Lindblad form for a quantum dynamical semigroup:

$$\mathcal{L}\rho = -i[H_S, \rho_S] + \sum_i \gamma_i[A_i\rho_S A_i^\dagger - \frac{1}{2}\{A_i^\dagger A_i, \rho_S\}] \quad (3.27)$$

in which case we rename $\mathcal{K}(t) \rightarrow \mathcal{L}$, and A_i are the Lindblad jump operators. The GKSL theorem states that any $\mathcal{L}$ of this form is the generator of a semigroup of completely-positive quantum dynamical maps [15]. This is easily seen by solving for the evolution

$$\Phi(t, s) = \exp(\mathcal{L}(t - s)) \quad (3.28)$$

If the positive rates are allowed to vary in time, then we lose the semigroup property but still retain complete positivity³, because in each timestep the evolution will adopt the Lindblad form, in which the GKSL theorem will apply for small enough dt (this is called time-dependent Markovian).

In summary, though $\Phi(t, 0)$ is CP by construction from the quantum channel, if we try to decompose the evolution as a composition $\Phi(t, s)\Phi(s, 0)$, we find that one of the constituent processes is not necessarily also CP due to the inverse inside of $\Phi(t, s)$. We concluded that the process $\Phi(t, s)$ is still CP if the rates in the TDME are positive, that is, that the process is CP-divisible. If the following weaker condition holds

$$\sum_i \gamma_i(t) |\langle n|A_i(t)|m \rangle|^2 \geq 0 \quad (3.29)$$

then the resulting dynamics need only be P-divisible.

³Positive rates are sufficient to guarantee positive evolution, but determining necessary conditions is an open problem [12].

3.2.4. Recipe for Obtaining Lindblad Form

Before we continue on to Markovianity, it will be very insightful to first understand how the Lindblad form might be obtained from our physical knowledge of a system, and not just general mathematical theorems. Furthermore, experimental data is often fitted to the Lindblad form in a phenomenological way, and it would be satisfying to *derive* it from some Hamiltonian (i.e. a microscopic description of the system). Here I will summarize the main steps in this procedure [38], which will naturally introduce the concept of Markovianity. First, denote the total Hamiltonian and its von Neumann equation by

$$H = H_S + H_E + V \quad (3.30)$$

$$\frac{d\rho}{dt} = -i[H, \rho] \quad (3.31)$$

We move to the interaction picture to focus on V :

$$H_0 = H_S + H_E \quad (3.32)$$

$$\tilde{\rho} = e^{iH_0 t} \rho e^{-iH_0 t} \quad (3.33)$$

$$\tilde{V} = e^{iH_0 t} V e^{-iH_0 t} \quad (3.34)$$

$$\frac{d\tilde{\rho}}{dt} = -i[\tilde{V}, \tilde{\rho}] \quad (3.35)$$

From here on, the tildes will be suppressed. We assume that at $t = 0$, $\rho(0) = \rho_S(0) \otimes \rho_E(0)$, such that the system and environment are initially separable. We also assume that the environment begins in a thermal state

$$\rho_E(0) = \frac{e^{-\beta H_E}}{\mathcal{Z}} \quad (3.36)$$

which helps some terms drop out in the details, and $\mathcal{Z}$ is the partition function. Lastly, we assume that the interaction (in the lab frame) is linear in the bosonic operators b_k

$$V = \sum_{\alpha, k} g_{\alpha, k} M_{\alpha} b_k^{\dagger} + g_{\alpha, k}^* M_{\alpha}^{\dagger} b_k \quad (3.37)$$

which is natural because most environments of interest are bosonic (i.e. phonons, photons). The M_{α} terms are system operators. With all of this, the Nakajima-Zwanzig approach yields an exact equation for the reduced dynamics of the system

$$\frac{d}{dt} \mathcal{P} \rho = \int_0^t dt' \mathcal{P} \mathcal{V}_t G(t, t') \mathcal{V}_{t'} \mathcal{P} \rho(t') \quad (3.38)$$

where the projection superoperator $\mathcal{P}$ causes the state to be projected into a separable form where the environment doesn't "move".

$$\mathcal{P} \rho(t) = \rho_S(t) \rho_E(0) = \text{Tr}_E(\rho(t)) \rho_E(0) \quad (3.39)$$

This is reminiscent of the idea that bath is so large that it never drifts far from its initial state due weak interaction with a tiny system, but the use of a projector keeps this step exact. $G(t, t')$ is a Green's function that depends on the projection of $\mathcal{V}$ (the superoperator form of the interaction picture V) onto the space orthogonal to $\mathcal{P}$. This equation is nonlocal in time for $\rho(t)$, and now some approximations are in order.

Born approximation We parametrize the interaction by writing $V \rightarrow \epsilon V$, and then assume that ϵ is small enough that the bath is barely affected by its coupling to the system. This means the Green's function becomes $G(t, t') \approx 1$ if the whole equation is to be second-order in ϵ .

Markov approximation We assume that the dynamics of the environment are much faster than the system. No future state of the system could depend on itself at previous times, because any record of previous states is “forgotten” by the environment. Conceptually, this is the core assumption of Markovianity. In this derivation, it means that we replace $\rho(t')$ with $\rho(t)$ in Eq. 3.38, and change variables $s = t - t'$ and integrate from 0 to ∞ .

Applying the projectors and expanding commutators to Eq. 3.38, we get

$$\frac{d}{dt}\rho_S = \int_0^\infty \text{Tr}_E[V(t)\rho_S(t)\rho_E V(t-s) - \rho_S(t)\rho_E V(t-s)V(t)] + \text{h.c.} \quad (3.40)$$

Plugging in the form of V yields the Bloch-Redfield master equation [14, 38]⁴. Working through Eq. 3.40 may sometimes result in time-dependent terms that would prevent the final derivation of a Lindblad equation, in which case we apply our final approximation.

Secular approximation Time-dependent terms can be dropped using the secular approximation (also known as the rotating-wave approximation), if it holds that the oscillation of those terms is fast with respect to the strength of the system-environment coupling. The omission of fast terms is not new, since we already assumed some timescale for the system to justify the existence of a faster bath timescale, which was then used to justify Markovianity.

With this, we have our recipe for obtaining a Lindblad master equation. Just compute V in the interaction picture, plug in to Eq. 3.40, and drop the time-dependent terms. Once the Lindbladian of a system is known, it is easy to solve for the generator of the time evolution via vectorization (see A.2).

3.2.5. Alternate Derivation

The Lindblad form can also be derived in a quicker, albeit less fundamental way. The Nakajima-Zwanzig method ultimately flows from the unitary evolution of the combined system-environment, but if we just looked at the system, we assume that the time evolution has the form of a quantum operation:

$$\rho(t)dt \approx \rho(t + \delta t) = \sum_m K_m(\delta t)\rho(t)K_m^\dagger(\delta t) \quad (3.41)$$

This Kraus set causes infinitesimal time evolution and only depends on δt . We choose one Kraus operator to have the particular form

$$K_0 = I + \delta t(-iH + G) + \mathcal{O}(\delta t^2) \quad (3.42)$$

where H is an arbitrary Hermitian matrix, and G is unknown. The other operators are redefined as $K_m = \sqrt{\delta t}L_m + \mathcal{O}(\delta t)$ so that the resulting equation will be first-order in t . Plugging in these definitions, we find that we can only obtain the Lindblad form if we assume that the unknown G is

$$G = -\frac{1}{2} \sum_{m \neq 0} L_m^\dagger L_m \quad (3.43)$$

which allows us to normalize the Kraus operators. Since we didn't need to enforce our three big assumptions (Born, Markov, secular), evidently these assumptions must already be built in to the ansatz of infinitesimal time evolution in Eq. 3.41.

3.2.6. Markovianity

By this point, we have seen Markovianity from two different angles. In the first case, we saw that assuming positive rates in the TDME led to CP-divisible dynamics. In the second, if we assume that the environment's dynamics are very fast compared to the system, it rapidly loses track of any information it receives from the system, and forgets it before it can leak back into the system. Since these two apparently different assumptions led to the Lindblad master equation, they must *both* be expressions

⁴This shows how to get the Redfield equation from the projection-operator approach, but it also follows more readily just by plugging in the von-Neumann equation into itself and making the same approximations.

of Markovianity. The form of the Kraus operators that generate infinitesimal time evolution in the “alternate” proof completes the connection, because if you recall that the most general completely-positive evolution takes the form of a quantum operation, then the ansatz of Eq. 3.41 is exactly the statement of CP-divisibility.

To be precise, these two views are necessary conditions that must hold in their respective proofs if we want to obtain Markovian dynamics, and certainly there were other pieces needed in both derivations along the way. At the very least, we can say that the memory of the environment is definitely related to the positivity of the time-dependent rates. We can summarize these views like so

1. A system undergoes Markovian dynamics if it is weakly coupled to an environment that quickly forgets information it obtains about the system.
2. Markovian dynamics follow from positive rates in the time-dependent master equation.⁵

Ostensibly, a *non*-Markovian environment is one that negates these points:

1. Non-Markovian environments do not evolve that quickly, and some information that leaks into the environment may remain coherent long enough to then leak back into the system at some later point in time. The evolution of $\rho(t)$ depends on its past.
2. Non-Markovian dynamics correspond to time-dependent master equations with at least one rate that becomes negative at least once during some evolution.

3.2.7. Quantifying Markovianity

Equipped with how to define non-Markovianity, we now turn to the problem of quantifying it. Point 2 appears more easily dealt with, since it already is a mathematical statement. If we could fit experimental data to the TDME, then we could extract the rates and directly observe any negativity. There is a problem with this, and it is a familiar sight in quantum mechanics: there exists a unitary degree of freedom in how the rates α_i and operators A_i are defined. The same TDME may be written in different operator bases, leading to different conditions on the rates. The resolution is to fix the basis, and the authors of Hall et al. [27] do this by diagonalizing a “decoherence” matrix⁶. They use its eigenvalues as canonical rates in the TDME, and its eigenvectors to form an *orthogonal* set of decoherence channels, i.e. a set of canonical A_i operators. Having explained their approach, this is not the method we will use, because the Kraus sets we will extract from our data later are a sequence of macroscopic time evolutions, from which the infinitesimal evolution would first need to be derived (see A.5 for more details about the canonical Lindblad form). Instead, we turn to point 1 and face the daunting challenge of measuring how well our environment “remembers” things. We will follow the approach of Breuer et al. [13], and begin with the concept of distinguishability.

We first must recall some properties of the trace distance (B.5.1)

$$D(\rho_1, \rho_2) = \frac{1}{2} \|\rho_1 - \rho_2\| \quad (3.44)$$

The trace distance is bounded between 0 and 1, and is invariant under unitary transformations. Importantly, the trace distance is also contractive under CPTP maps

$$D(\Lambda\rho_1, \Lambda\rho_2) \leq D(\rho_1, \rho_2) \quad (3.45)$$

meaning that two distinct input states can only become closer after the action of the map Λ . Equality is achieved when Λ is unitary, and the two states just rotate together. It is known that if Alice wants to perform a single measurement to determine whether Bob sent her ρ_1 or ρ_2 (which are sent with equal probability), then her maximal success probability is [24]

$$P_{max} = \frac{1}{2}(1 + D(\rho_1, \rho_2)) \quad (3.46)$$

This makes sense, because if the trace distance is 0, then Alice can do no better than just guessing randomly. If it is nonzero, then in principle there is some kind of difference between the states that

⁵CP evolution does *not* imply positive rates. There is at least one example of a system with CP evolution and negative rates [27].

⁶The decoherence matrix is the process matrix after removing its first row and column.

Alice could use to her advantage. If the distance is 1, then the states are orthogonal, and Alice only needs to make 1 clever measurement to tell them apart. So, if the trace distance can bias Alice's success probability, then we can interpret the trace distance as the "distinguishability" between the two states.

With this analogy, if we take two states $\rho_1(s)$ and $\rho_2(s)$ and evolve them in time, we can play the role of Alice and ask how distinguishable $\rho_1(s+t)$ and $\rho_2(s+t)$ are. If the evolution is given by a CPTP map, then the trace distance must contract, and our ability to distinguish them must at best stay the same, or otherwise get worse over time. In terms of Φ , we have

$$D(\Phi_{t,s}(\rho_1(s)), \Phi_{t,s}(\rho_2(s))) \leq D(\rho_1(s), \rho_2(s)) \quad (3.47)$$

If the time evolution is CP-divisible, then at each moment in time our states undergo dynamics that can only monotonically decrease their trace distance. In the language of memory, we imagine that whatever information might differentiate them is slowly leaking from the system to the environment where it is forgotten, never to return.

But what if the evolution is non-Markovian? Earlier, under very general conditions we wrote

$$\Phi_{t,0} = \Phi_{t,s} \Phi_{s,0} \quad (3.48)$$

It is possible that the evolution cannot be divided into CP steps ad infinitum (the composition with $\Phi_{s,0}$ is what keeps the whole process CP). When the CP condition is violated, the trace distance is no longer required to contract, and the distinguishability between the two states could actually *increase*. Though the full evolution from 0 out to time t is CP, and the trace distance cannot end higher than it started, the decay can be non-monotonic, with periodic resurgences. At such times, at least one canonical rate must be negative, and the environment leaks information back into the system. Thus, Markovianity in the language of the trace distance becomes

1. An evolution is Markovian if the trace distance between any two pair of states monotonically decays over time.
2. An evolution is non-Markovian if there exists a pair of input states whose trace distance does not monotonically decay over time.

The quantity of interest, then, is the rate of change of the trace distance over time

$$\sigma(t, \rho_{1,2}(0)) = \frac{d}{dt} D(\rho_1(t), \rho_2(t)) \quad (3.49)$$

and we wish to compute the total increase of distinguishability over the whole time-evolution. The measure defined by Bruer et al. [13] is

$$\mathcal{N}(\Phi) = \max_{\rho_{1,2}(0)} \int_{\sigma > 0} dt \sigma(t, \rho_{1,2}(0)) \quad (3.50)$$

where the time integration is extended over all time-intervals (a_i, b_i) in which σ is positive. The optimal (i.e. most non-Markovian) pair of input states is the one that maximizes this integral. It can be shown [15] that these two states are always orthogonal and lie on the boundary of the state space. If the evolution is Markovian, then there is no domain for the integral because σ is always negative, so $\mathcal{N}(\Phi) = 0$.

The trace distance is an intuitive quantity to work with, and is readily obtained from experimental data. Thus, it is this metric that we will apply to the simulations later on. One downside, however, is that the integration depends on the total evolution time, making the metric unbounded and difficult to compare across different implementations (i.e. the protocol would need to be standardized before a comparison could be made).

3.2.8. Markovianity - Loose Ends

The picture of Markovianity presented here is satisfying, but we comment here on a few things that were brushed under the rug. It turns out that the contractive property also holds for processes that are only positive, but not completely positive. Furthermore, if the classical definition of Markovianity is translated into the quantum domain, then the condition obtained for the transition probabilities of the so-called Chapman-Kolmogorov equation takes on exactly the same form as the P-divisibility condition

for the canonical rates. Thus, it is argued by Breuer [15] that quantum Markovianity corresponds to P-divisible, and not CP-divisible, evolution.

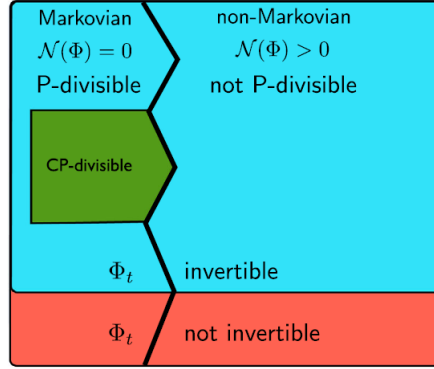


Figure 3.2: Relations between divisibility classes and Markovianity. The angled boundary indicates non-convex spaces. [15]

Not all are in agreement with this, for Rivas et al. [51] decide to quantify Markovianity by maximally entangling the system with an ancilla, and then measuring that entanglement over time. Non-Markovianity occurs whenever this entanglement increases, because it can only ever decrease under local CPTP operations. Thus, their metric is explicitly based on CP, rather than P, evolution. It has even been shown that a CP-divisible evolution could still result in memory effects given an initially entangled environment state, suggesting that Markovianity should not be based on CP-divisibility, and by extension, the trace distance [42].

In reality, there is no consensus in the literature on a single operational definition for Markovianity (perhaps this is evident from the many approaches already described). There are a multitude of interpretations and as many quantifiers, and they are not all equivalent. Theorists have devised metrics that base Markovianity on the Bures distance, Fisher information, Choi matrix, Bloch volume, and added isotropic noise. To make matters more complicated, Hall et al. [27] provided a specific, simple model with completely positive evolution, yet an eternally negative canonical rate. They showed that only some of these various metrics were able to detect its non-Markovianity. Fortunately, if the jury is out on a universal definition, then experimentalists should simply choose the definition they find most accessible, and make clear their choice in their work.

3.3. Quantum Measurement

Here we briefly review the formalism of quantum measurement [44]. The measurement postulate states that any quantum measurement can be described by a set of operators $\{M_m\}$ where

- The index m refers to each possible measurement outcome in consideration
- The probability of obtaining result m is

$$p_m = \text{Tr}(M_m \rho M_m^\dagger) = \text{Tr}(M_m^\dagger M_m \rho) \quad (3.51)$$

- The state of the system after obtaining result m is

$$\rho_m = \frac{M_m \rho M_m^\dagger}{\text{Tr}(M_m \rho M_m^\dagger)} \quad (3.52)$$

- The completeness relation is satisfied

$$\sum_m M_m^\dagger M_m = I \quad (3.53)$$

This postulate gives a recipe for computing measurement probabilities and results, but if we only care about the probabilities alone (i.e. we don't have access to the post-measurement state), then we discard

bullet 3 above and see that all defined quantities depend on the set $\{M_m\}$ through the terms $M_m^\dagger M_m$ which we redefine as E_m . Though seemingly simple, this redefinition produces a positive-operator valued measure (POVM) described by the set $\{E_m\}$. Given only some POVM $\{E_m\}$, we can obtain a measurement set that gives the same probabilities via $M_m = \sqrt{E_m}$ (which is allowed because E_m is positive semi-definite). Going the other direction, we also note that there are infinitely many measurement sets that obey the same POVM, because we can define $M_m \rightarrow U_m M_m$ for some unitary U_m , and then form the same POVM $E_m = M_m^\dagger U_m^\dagger U_m M_m = M_m^\dagger M_m$. To be clear, these two measurement sets $\{M_m\}$ and $\{U_m M_m\}$ will result in *different* post-measurement states, but for POVMs, we assumed no interest in the states to begin with. Since probabilities are real numbers in the range $[0, 1]$, and since bullet point 2 must yield a probability for *any* input state ρ , we see that E_m must be a positive semidefinite matrix, and therefore the POVM operators are Hermitian.

Ideally, the measurements we perform in the lab are projective measurements (PVM). In this case, the set of measurement operators M_m is the same as the POVM E_m , because $(|m\rangle\langle m|)^2 = |m\rangle\langle m|$. In reality, real measurements take finite time, during which noise processes corrupt our model. We must use the experimental data to find not a PVM, but a POVM that best reproduces the measurement statistics.

It is interesting to note that with this definition of POVMs, there is a set of “dilation” theorems regarding state preparation, evolution, and measurement.

- Every mixed state is a pure, possibly entangled state, in a larger Hilbert space (purification).

$$\rho_S = \text{Tr}_A[|\psi\rangle\langle\psi|_{SA}] \quad (3.54)$$

- Every CPTP map is a unitary operation in a larger Hilbert space (Stinespring dilation [38]).

$$\mathcal{E}(\rho_S) = \text{Tr}_A[U(\rho_S \otimes |0\rangle\langle 0|_A)U^\dagger] \quad (3.55)$$

- The measurement statistics of any POVM can be equivalently thought of as arising from projective measurement in a larger Hilbert space (Naimark’s Theorem [45]).

$$p_i = \text{Tr}[(I \otimes |i\rangle\langle i|)(U|\psi\rangle\langle\psi|_{SA}U^\dagger)] \quad (3.56)$$

If the PVM consists of $M_i = I \otimes |i\rangle\langle i|$, then the POVM can be obtained by finding the action of the PVM on the reduced state of the system, if the outcome i was obtained on the ancilla/environment

$$\rho_S = \text{Tr}_A[I \otimes |i\rangle\langle i|(U|\psi\rangle\langle\psi|_{SA}U^\dagger)I \otimes |i\rangle\langle i|] \quad (3.57)$$

$$= \langle i|U|a\rangle|s\rangle\langle s|\langle a|U^\dagger|i\rangle \quad (3.58)$$

where the total initial state was assumed to be separable. The operators $\langle a|U^\dagger|i\rangle\langle i|U|a\rangle$ constitute a POVM. If the result of the measurement i is forgotten, we take a sum over all outcomes, which recovers the Kraus operator sum. So a quantum operation is effected whenever we measure with a POVM, but toss out the result. This is intuitive, because one way of framing the effect of an environment is to imagine that it performs some measurement on the system, which we never have access to. Thus, we can only assume the environment’s result is thrown out, and we take a sum over all outcomes. The evolution of our system is then modeled by a quantum operation.

3.4. Characterization

Until now, we have discussed how to model the preparation, evolution, and measurement of quantum systems in the abstract, symbolically expressing them in terms of conditions that hold in an ideal, theoretical picture. In the lab, nature only provides us with measurement outcomes. The purpose of any characterization is to use these outcomes to estimate numerical representations (matrices) of the real world that also conform to our abstractions (operators). In this section, we discuss four algorithms commonly used to reconstruct quantum objects based solely on experimentally accessible data.

3.4.1. Quantum State Tomography

Quantum state tomography (QST) is an algorithm whose purpose is to use a set of measurement outcomes to reconstruct an unknown quantum state ρ . To understand QST, we can use the vectorized

density matrix $|\rho\rangle\rangle$ (see A.1). Projecting two vectorized matrices onto each other is defined through the Hilbert-Schmidt inner product $\langle\langle\sigma|\rho\rangle\rangle = \text{Tr}\{\sigma^\dagger\rho\}/d$, where the division is for normalization. In this notation, quantum state tomography is simply [26]

$$p_j = \langle\langle E_j|\rho\rangle\rangle = \sum_k \langle\langle E_j|k\rangle\rangle \langle\langle k|\rho\rangle\rangle \quad (3.59)$$

where the vectorized operator basis elements are $\{|k\rangle\rangle\}$. Conveniently, the Born Rule for obtaining the probabilities is equivalent to computing the inner product between a POVM operator and a density matrix. The operators E_j are elements of a tomographically-complete POVM whose measurement results are sufficient to constrain ρ . Since ρ is $d \times d$ and Hermitian, it has d^2 real parameters, and our POVM will need d^2 operators. For qubits, if we choose as our POVM the set $\{|0\rangle\langle 0|, |1\rangle\langle 1|, |+\rangle\langle +|, |+\rangle\langle +| + |+\rangle\langle +i|\}$, this will be sufficient to constrain the unknown ρ :

$$\langle\langle E_1|\rho\rangle\rangle = \text{Tr}(|0\rangle\langle 0|\rho)/2 = \rho_{00} \quad (3.60)$$

$$\langle\langle E_2|\rho\rangle\rangle = \text{Tr}(|1\rangle\langle 1|\rho)/2 = \rho_{11} \quad (3.61)$$

$$\langle\langle E_3|\rho\rangle\rangle = \text{Tr}(|+\rangle\langle +|\rho)/2 = \frac{1}{2}(\rho_{00} + \rho_{01} + \rho_{10} + \rho_{11}) \quad (3.62)$$

$$\langle\langle E_4|\rho\rangle\rangle = \text{Tr}(|+\rangle\langle +i|\rho)/2 = \frac{1}{2}(\rho_{00} + i\rho_{01} - i\rho_{10} + \rho_{11}) \quad (3.63)$$

The $\{|0\rangle\langle 0|, |+\rangle\langle +|, |+\rangle\langle +i|\}$ measurement data determines the projections of the state ρ onto the axes of the Bloch sphere, and $|1\rangle\langle 1|$ allows for trace normalization ⁷ If desired, the full set of $\{|0\rangle\langle 0|, |1\rangle\langle 1|, |+\rangle\langle +|, |-\rangle\langle -|, |+\rangle\langle +i|, |-\rangle\langle -i|\}$ can be used to overconstrain the system, which can help compensate for some experimental imperfections ⁸. Actually, that data is necessarily obtained, because we cannot choose to measure, say, just the projector $|+\rangle\langle +|$, but only to measure in the X basis. The smaller POVM simply represents one possible minimally-constraining subset of the measurement data. Other than this difference—choosing states to prepare in “measurement tomography” and POVMs to measure with in state tomography—the two procedures are identical, which is evident in the symmetry of the Born rule:

$$p_i = \text{Tr}(E_i\rho) \quad (3.64)$$

Mathematically, M_i and ρ are just two Hermitian PSD matrices; physics dictates their distinction. Returning to Eq. 3.59, since the POVM is assumed to be known, matrix inversion provides the result

$$A^{-1}|\hat{p}\rangle\rangle = |\hat{\rho}\rangle\rangle \quad (3.65)$$

where $A_{jk} = \langle\langle E_j|k\rangle\rangle$ is known. In practice, maximum-likelihood estimation is used instead for this step [30]. See Lundeen et al. [40] for an example of measurement tomography, and section 4.1.2 for more details on POVM reconstruction for qutrits.

3.4.2. Quantum Process Tomography

Quantum process tomography (QPT) goes one step further than QST, where now both a set of input states and a POVM are presumed known in advance, while some quantum channel or logic gate separating the preparation and measurement in time is unknown. Quantum process tomography for a process G is

$$p_{ij} = \langle\langle E_j|G|\rho_i\rangle\rangle \quad (3.66)$$

$$= \sum_{k,l} \langle\langle E_j|k\rangle\rangle \langle\langle k|G|l\rangle\rangle \langle\langle l|\rho_i\rangle\rangle \quad (3.67)$$

⁷The implication is that the same number of trials is used for each basis, so in at least one basis we need to count both kinds of outcomes. Counting the (-1)-type outcomes in each basis would implicitly be done by counting the (+1)-type outcomes, if the total number of trials is known beforehand. If the trace is normalized in this way, then ρ only has $d^2 - 1$ constraints.

⁸One can imagine a scenario where the 4 states of the minimal POVM all lie very close to each other, such that they are barely linearly independent. One would need a large amount of data to resolve the different measurements, and this can be remedied somewhat by using the full POVM. Of course, any imperfections in the POVM will still be falsely attributed to the state.

Since $\langle\langle E_j | G | \rho_i \rangle\rangle = \text{Tr}(E_j G(\rho_i))$, we see that in QPT, we perform QST for each ρ_i in a complete set of inputs. By determining the action of G on the operator basis (which can be written as linear combinations of the ρ_i), we can determine how any state expressed in that basis will transform under G . The same projectors of the POVM $\{E_j\}$ can be used as the initial states $\{\rho_i\}$. The superoperator G can be easily found by vectorizing the Kraus form for the process:

$$\text{vec}(\mathcal{E}(\rho)) = \sum_i (K_i^* \otimes K_i) \text{vec}(\rho) \quad (3.68)$$

If we then choose an operator basis $\{B_j\}$ to express $\{K_i\}$ in, we get

$$K_i = \sum_j B_j a_{ji} \quad (3.69)$$

$$\text{vec}(\mathcal{E}(\rho)) = \sum_i \left(\left(\sum_k B_k a_{ki} \right)^* \otimes \left(\sum_j B_j a_{ji} \right) \right) \text{vec}(\rho) \quad (3.70)$$

$$= \sum_i \sum_{j,k} a_{ji} a_{ki}^* (B_k^* \otimes B_j) \text{vec}(\rho) \quad (3.71)$$

$$= \sum_{j,k} [a a^\dagger]_{j,k} (B_k^* \otimes B_j) \text{vec}(\rho) \quad (3.72)$$

Note that in writing a_{ji} , we are assuming that the basis operators have been assembled into a row vector, and the expansion coefficients are along the columns of the matrix a . The process matrix is $\chi = a a^\dagger$, and the matrix elements are $\chi_{jk} = \frac{1}{d^2} \sum_i \text{Tr}(B_j^\dagger K_i) \text{Tr}(B_k^\dagger K_i)^*$. The process matrix is unique once the basis is fixed, and is Hermitian positive semidefinite. Thus, χ has $d^4 - d^2$ free parameters, which are fixed by obtaining $d^2 \times d^2$ measurement probabilities from the complete POVM and input sets. Therefore, QPT can also be framed in terms of χ . Both QST and QPT can be solved with linear inversion (LI) or maximum likelihood estimation (MLE), but only MLE allows for physicality constraints to be easily enforced. [26].

It is well-known that the Kraus form is not unique. A new, equally valid, Kraus form for the same operation $\mathcal{E}(\cdot)$ can be obtained via a certain linear combination of the old Kraus form:

$$M_i = \sum_j K_j u_{ji} \quad (3.73)$$

This is easily seen by substitution

$$\mathcal{E}(\rho) = \sum_i M_i \rho M_i^\dagger \quad (3.74)$$

$$= \sum_i \sum_{j,n} (u_{ji} K_j) \rho (u_{ni}^* K_n^\dagger) \quad (3.75)$$

$$= \sum_{j,n} \sum_i u_{ji} u_{in}^\dagger K_j \rho K_n^\dagger \quad (3.76)$$

If u is a unitary matrix, then $\sum_i u_{ji} u_{in}^\dagger = [u u^\dagger]_{j,n} = \delta_{j,n}$, and we obtain

$$\mathcal{E}(\rho) = \sum_j K_j \rho K_j^\dagger \quad (3.77)$$

which is the same as the original operation. Since there are an infinite number of unitary matrices from which we can take our expansion coefficients, the form of the quantum operation is very malleable. It would be hard to tell just by looking whether or not two different sets $\{K_i\}$ and $\{M_i\}$ might be unitarily related. Since the process map is unique, if the two are really equivalent, then their vectorized forms would have to be equal. Thus, one must be careful not to draw too many conclusions from the Kraus sets alone.

If one chooses to use the Pauli basis instead of the natural basis, then the map G is known as the Pauli transfer matrix (A.4), which has its own set of handy properties, but it is not relevant for our purposes.

3.4.3. Gate Set Tomography

In QST, we assumed perfect measurements, in measurement tomography, we assumed perfect input states, and in QPT, we assumed both to determine an unknown process. These perfect aspects are also faulty in practice, and will pollute the estimates of the unknown quantities to some extent. In particular, it is contradictory in QPT to characterize some presumably faulty gate from a gate set, while at the same time using presumably perfect gates from the *same* set to prepare ideal inputs and measurement pre-rotations during the characterization. All the methods we have seen so far are subject to state preparation and measurement (SPAM) errors. Gate set tomography (GST) [11] self-consistently estimates a gate set along with a fixed starting state (vacuum) and measurement POVM (Z-basis):

$$\mathcal{G} = \{|\rho\rangle\rangle, \langle\langle E|, G_0, G_1, G_2 \dots\} \quad (3.78)$$

where one gate G_0 is the null gate, which does nothing for no time. A set of SPAM gates F_i is used for prepare other input states from $|\rho\rangle\rangle$, and to rotate into other bases for $\langle\langle E|$. These SPAM gates must themselves be composed from elements of the gate set:

$$F_i = G_1^{f,i} \circ G_2^{f,i} \circ \dots \quad (3.79)$$

where $\{G_k^{f,i}\}$ is the specific subset of $\{G\}$ needed to compile the specific SPAM gate F_i . Measurement statistics are obtained for all possible SPAM choices and gates

$$p_{ijk} = \langle\langle E|F_i G_j F_k|\rho\rangle\rangle \quad (3.80)$$

where $\langle\langle E|$ projects onto outcome m for all output counts of type m . To enforce physicality constraints, MLE is used to find a parametrization of all unknown quantities $\mathcal{G}$ that is most likely to reproduce the observed data. GST can also be solved via linear inversion (LGST), and it is suggested that although the output of the inversion may not be physical, the closest physical model to LGST can provide a good seed for the MLE routine [26]. While QPT will spread the error of one faulty gate over a whole gate set, GST will correctly attribute imperfections to the faulty gate alone [26]. Likewise, if the only imperfection is in $|\rho\rangle\rangle$ or $\langle\langle E|$, QPT will spread that error over the gates, while GST will not.

3.4.4. Randomized Benchmarking

QST, QPT, and GST all provide detailed information about quantum hardware, but come with the price of inefficiency: they all scale exponentially with the system size. Randomized benchmarking [34] provides an estimate of the average Clifford gate error⁹ in some quantum device (as opposed to the native gate set). The procedure involves choosing a random string of Clifford gates, applying the inverse of that string, and then measuring the output.

$$C_1 \circ C_2 \circ \dots \circ C_k \circ C_k^\dagger \circ \dots \circ C_2^\dagger \circ C_1^\dagger |0\rangle \quad (3.81)$$

If the gates were implemented perfectly, the string should work out to the identity. In reality, the process on the device becomes

$$C_1 \circ \Lambda \circ C_2 \circ \Lambda \circ \dots \circ \Lambda \circ C_k \circ \Lambda \circ C_k^\dagger \circ \Lambda \circ \dots \circ \Lambda \circ C_2^\dagger \circ \Lambda \circ C_1^\dagger \circ \Lambda |0\rangle \quad (3.82)$$

where noise processes prevent perfect cancellation. A sequence of lengths is chosen, and many Clifford strings of each length are measured and averaged over. The final probability decays exponentially from 1 to 0.5, and the decay parameters can be used to estimate the gate fidelity (the random Cliffords “twirl” the noise channel into a depolarizing channel with an error probability that depends on the gate fidelity [26, 34]). The strength of RB is also its weakness: because it provides a single number to quantify the gate fidelity, it is agnostic to the underlying error source, and may even underestimate the error strength in certain cases [26]. Moreover, it is argued by Proctor et al. [50] that the average gate infidelity, being representation-dependent, is ill-defined, while the output of the RB protocol is not. Therefore, whatever RB measures, it cannot be the gate infidelity so often ascribed to it.

⁹Clifford gates are gates that normalize the Pauli group, i.e. $CPC = P$.

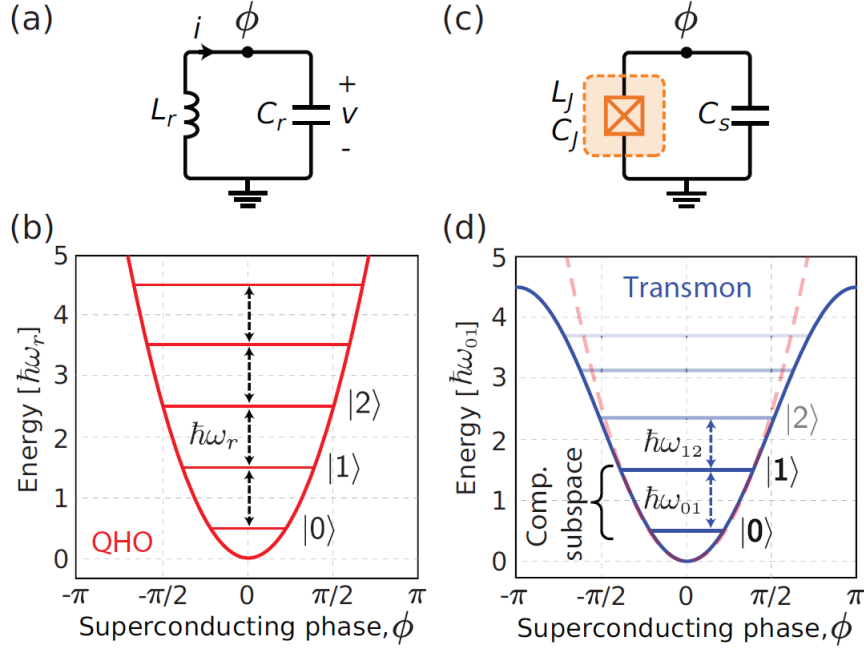


Figure 3.3: (a) Quantum harmonic oscillator. (b) Harmonic level spectrum. (c) Capacitively-shunted Josephson junction. (d) Anharmonic spectrum. [36]

3.5. Superconducting Quantum Hardware

So far, our discussions have been completely general. If characterization provides the tools to conform our theories to nature *ex post facto*, then modeling is the reverse: the construction of theories that may predict what nature does. The most ubiquitous and well-studied hardware platform for quantum computing utilizes superconducting qubits, and it is this paradigm that we will work within. The theory and modeling of superconducting devices will be introduced in the following sections.

3.5.1. Transmons

The workhorse of most superconducting quantum computers is the transmon qubit, which is a Josephson junction (JJ) shunted by a capacitor. One first writes out the classical Lagrangian for a capacitor in parallel with a JJ, performs a Legendre transform to obtain a classical Hamiltonian, and then canonically quantizes the conjugate variables to obtain the following expression

$$H = 4E_C n^2 - E_J \cos \phi \quad (3.83)$$

The behavior of the circuit is determined by the relative energy scale E_J/E_C , the ratio of the Josephson and capacitor energies. Specifically, the transmon is a particular operating regime where the Josephson energy dominates: $E_J \gg E_C$. This is achieved by using a large shunt capacitance, which reduces the charging energy scale via $E_C = e^2/2C$. The transmon is an improved version of an earlier design known as the Cooper-pair box (CPB). The CPB charge qubit was designed such that $E_J \leq E_C$ and, thus, its dominant noise channel came from charge fluctuations. Over time, it was found that flux noise was easier to mitigate, so designs shifted towards transmons. As one goes deeper into the transmon regime, one obtains better and better resistance to flux noise (smoother energy levels) and the price of reduced anharmonicity (poorly defined qubit space).

The reason we add a JJ is apparent from Fig. 3.3. For a qubit to be well-defined, its Hilbert space must be separated from surrounding states of the system that are not needed to encode quantum computation. The quantum harmonic oscillator is as far as you can get from this, because every energy level is equidistant, so for example, incident photons intended to drive gates via the $\omega_{0 \leftrightarrow 1}$ transition could just as well cause dynamics in the higher levels. The JJ adds much-needed non-linearity to the level structure, so that $\omega_{1 \leftrightarrow 2} \neq \omega_{0 \leftrightarrow 1}$, and so the qubit space is better-defined.

In quantizing the system, all the flux and charge fields are promoted to operators: $\phi_i \rightarrow \hat{\phi}_i$ and $\dot{\phi}_i \rightarrow \hat{\dot{\phi}}_i \equiv \hat{n}_i$, and then are made conjugate to each other by requiring $[\phi_i, n_i] = i$. The charge operator $\hat{n}$ encodes the difference in the number of Cooper pairs across the junction (i.e. the number of pairs that have tunneled). The phase or flux operator $\hat{\phi}$ encodes the phase difference of the superconducting condensate across the junction. By comparison with the quantum harmonic oscillator, if we interpret the phase and charge variables as position and momentum, we can introduce the raising and lowering operators

$$\hat{\phi} \propto b^\dagger + b \quad (3.84)$$

$$\hat{n} \propto i(b^\dagger - b) \quad (3.85)$$

To treat the transmons as qudits, we simply truncate the operators to be $d \times d$ shaped-matrices. For qutrits, $d = 3$, and we can write

$$b = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & \sqrt{2} \\ 0 & 0 & 0 \end{bmatrix} \quad \hat{\phi} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & \sqrt{2} \\ 0 & \sqrt{2} & 0 \end{bmatrix} \quad (3.86)$$

$$b^\dagger = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & \sqrt{2} & 0 \end{bmatrix} \quad \hat{n} = i \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & -\sqrt{2} \\ 0 & \sqrt{2} & 0 \end{bmatrix} \quad (3.87)$$

With these operators, the transmon Hamiltonian can be rewritten in a more familiar basis. The cosine term can be Taylor expanded to fourth-order, and after substituting in the b -operators, all non-energy-preserving terms are dropped (their first-order contributions to state-shifts must be small for this to be valid). The resulting approximate Hamiltonian for the transmon in the basis of creation and annihilation operators is

$$H_T = (\sqrt{8E_J E_C} - E_C) b^\dagger b - \frac{E_C}{2} b^\dagger b^\dagger b b \quad (3.88)$$

$$H_T/\hbar = \omega b^\dagger b - \frac{\alpha}{2} b^\dagger b^\dagger b b \quad (3.89)$$

This Hamiltonian (a.k.a a Duffing oscillator) is now diagonal in the number basis. The anharmonicity is defined as the difference in the energy level spacing between the second and first levels, to the first and zeroth levels: $\alpha \equiv (E_2 - E_1) - (E_1 - E_0)$.

3.5.2. Asymmetric Transmon Hamiltonian

The asymmetric transmon is one variant of the transmon, which replaces the JJ with two JJs of different Josephson energies in parallel. The presence of a loop allows the qubit's energy to respond to magnetic fields threading the loop. The asymmetric transmon was introduced to suppress the flux response of the symmetric transmon such that it could have reduced flux noise, but still retain sufficient tunability for gate operations and fabrication variation compensation. The Hamiltonian of an asymmetric flux-tunable transmon qubit is given by

$$H_{q_i} = 4E_C n_i^2 - E_{J\Sigma} \sqrt{\cos \phi_{e,i} + d^2 \sin \phi_{e,i} \cos \phi_i} \quad (3.90)$$

where $E_{J\Sigma} = E_{J1} + E_{J2}$ and $d = (\gamma - 1)/(\gamma + 1)$ is the junction asymmetry parameter, with $\gamma = E_{J2}/E_{J1}$, and $\phi_{e,i}$ is the external magnetic flux applied to qubit i . In the limit $d = 0$, the Hamiltonian reduces to that of a symmetric split transmon

$$H = 4E_C n^2 - 2E_J |\cos \phi_e| \cos \phi \quad (3.91)$$

and when $|d| \rightarrow 1$ (by making one junction's Josephson energy much larger than the other), the term dependent on the external flux ϕ_e approaches 1, and the whole system loses its flux-sensitivity and returns to a single-junction transmon.

3.5.3. MIT Device

The physical device we have in mind in this report is a linear chain of 3 Xmons developed by Prof. Will Oliver's group at MIT, as described in Kjaergaard et al. [33]. Xmons are theoretically described just like transmons, but they are fabricated in a particular cross-shape that allows for direct capacitive coupling to a drive line, flux bias line, readout line, and other Xmons (earlier transmons had couplings mediated by 2D and 3D cavities, or bus resonators). In the MIT device, each Xmon is capacitively coupled to its nearest neighbors and also to a dedicated drive line and a shared readout resonator. Each Xmon also has a dedicated flux bias line for tunability. The Hamiltonian for such a system is

$$H = \sum_i H_{q_i} + \sum_{(i,j)} H_{c_{ij}} + H_r + \sum_i H_{r,q_i} \quad (3.92)$$

where we assume that there is no driving and that the flux biases are parked at some operating point(s). The first term is the sum of the individual transmon Hamiltonians, the second term is the nearest-neighbor capacitive coupling of the transmons ($(i, j) \in [(1, 2), (2, 3)]$), the third term is the readout resonator's energy, and the fourth term is each transmon's coupling to the readout resonator. On top of this, there may be thermal bath effects such as relaxation and absorption by both the transmons and resonators, requiring an open system treatment.

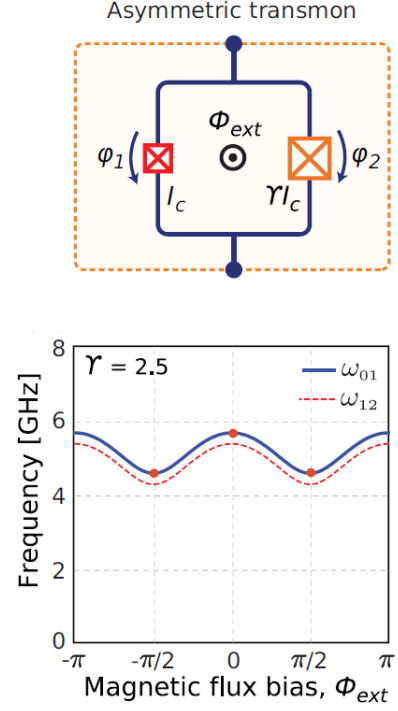


Figure 3.4: (top) Asymmetric transmon circuit. (bottom) Tunable spectrum of the asymmetric transmon. [36]

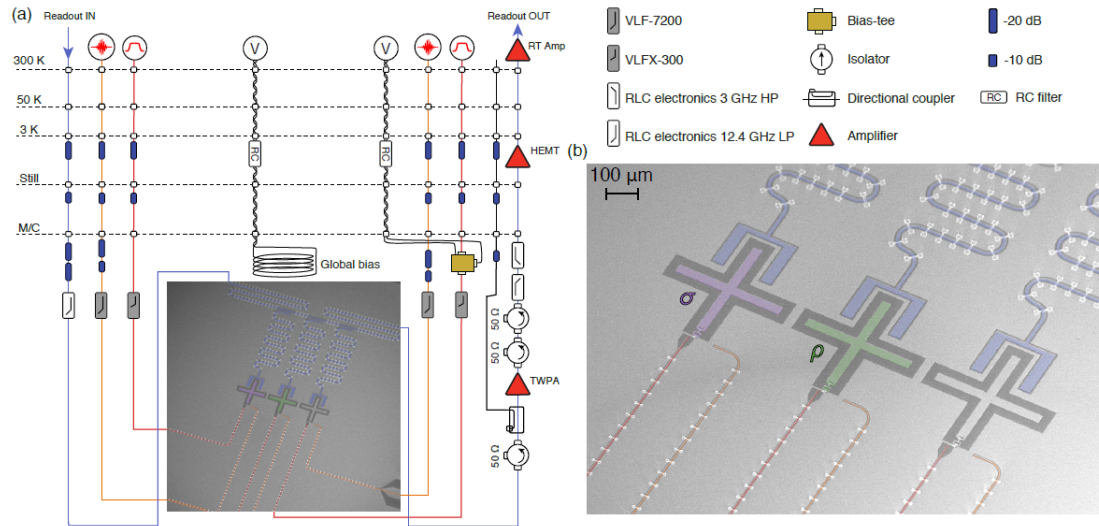


Figure 3.5: Three-transmon device from MIT [33]. (a) Schematic of control and readout electronics. (b) SEM micrograph of device.

Parameter	Qubit 1 (σ , target)	Qubit 2 (ρ , instruction)
Idling frequency, $\omega_i/2\pi$	4.748 GHz	4.225 GHz
Anharmonicity, $\eta/2\pi$	−175 MHz	−190 MHz
Coupling strength, $g/2\pi$	10.6 MHz	
Readout resonator frequency, $f_i/2\pi$	7.251 GHz	7.285 GHz
Junction asymmetry	1:5	1:10

Figure 3.6: Device parameters from Kjaergaard et al. [33].

Coupling two transmons with a capacitor gives rise to an interaction term in the Hamiltonian proportional to the operator $\hat{n}_1\hat{n}_2$; in the case we are considering here, the coupling is

$$H_{c_{ij}} = H_{ij}^{int} = 4e^2 \frac{C_g}{C_1 C_2} n_1 n_2 \equiv g n_1 n_2 \quad (3.93)$$

$$= g(i(b_1^\dagger - b_1))(i(b_2^\dagger - b_2)) \quad (3.94)$$

$$= -g(b_1^\dagger b_2^\dagger - b_1 b_2^\dagger - b_1^\dagger b_2 + b_1 b_2) \quad (3.95)$$

In these kinds of calculations, we usually move to the rotating frame (Appendix ??) to focus on the dynamics created by couplings we impose and remove “bare” dynamics due to the natural precession of the system. Moving to this frame causes terms that do not conserve energy (such as terms with products of “like” operators, e.g. dagger-dagger) to oscillate rapidly relative to terms of unlike operators. Since their fast movement averages out over the slower timescale of the other terms, we drop from the expression (a.k.a the rotating-wave approximation). This results in the coupling

$$H_{r,qi} \approx g(b_1 b_2^\dagger + b_1^\dagger b_2) \quad (3.96)$$

We note that since these operators are purely-off diagonal, the interaction's effect, generally, is to drive transitions between different energy states, and so such an interaction is called *transverse* because in the qubit approximation, the operators $\hat{\phi}$ and $\hat{n}$ become the Pauli X and Y operators, which are "transverse"

to the qubit's quantization axis, Z . The Hamiltonian for the readout resonator is simply that of a QHO

$$H_r = \hbar\omega_r a^\dagger a \quad (3.97)$$

A transmon can be coupled to a resonator in precisely the same way it can be coupled to another transmon. Therefore, the transverse coupling is

$$H_{r,q_i} \approx g(ab^\dagger + a^\dagger b) \quad (3.98)$$

where the a operators belong to the resonator, and the b operators to the transmon.

3.5.4. Full Hamiltonian

We are interested in two-qubit gates and couplings, and so we focus on a subset of the aforementioned device Hamiltonian. Specifically, consider just two adjacent transmons coupled by a capacitor. The Hamiltonian for a circuit comprised solely of transmons and capacitive couplings between them is [63]

$$H/\hbar = \sum_i \omega_i b_i^\dagger b_i - \frac{\alpha_i}{2} b_i^\dagger b_i^\dagger b_i b_i + \sum_{(i,j)} g_{ij} (b_i^\dagger b_j + b_i b_j^\dagger) \quad (3.99)$$

where i indexes the transmons and (i, j) refers to every coupled pair of transmons, ω_i is the lowest transition energy of the transmon, and α is its anharmonicity. In our case, there is a "left" transmon (1) and a "right" transmon (2)

$$H/\hbar = \omega_1 b_1^\dagger b_1 - \frac{\alpha_1}{2} b_1^\dagger b_1^\dagger b_1 b_1 + \omega_2 b_2^\dagger b_2 - \frac{\alpha_2}{2} b_2^\dagger b_2^\dagger b_2 b_2 + g(b_1^\dagger b_2 + b_1 b_2^\dagger) \quad (3.100)$$

both of which will be treated as qutrits, so the joint Hilbert space is spanned by the states $|\psi\rangle \in \{|0\rangle, |1\rangle, |2\rangle\} \times \{|0\rangle, |1\rangle, |2\rangle\}$.

3.5.5. Two-Qubit Gates

The physical interaction engineered between qubits in a device can be utilized to perform a 2-qubit logic gate. In superconducting hardware, two-qubit gates typically come in two flavors. If the qubits are flux-tunable, magnetic fields can be used to bias them towards certain level-crossings, resulting in state-dependent interactions. If the qubits are fixed-frequency, then carefully designed microwave drives can, in some transformed basis (e.g via a Schrieffer-Wolff transform) give rise to an effective two-qubit coupling. The drawback of using flux control is that it adds another channel by which noise can degrade the fidelity of the system. Furthermore, flux control requires deviating from sweet-spots, exposing qubits to more dephasing. All-microwave control can be used to circumvent these issues, and many gates have been designed in this way to avoid flux-related issues (cross-resonance, microwave-activated CPHASE, bSWAP, etc. [36]), arguably at the expense of greater complexity. Gate schemes involving tunable couplers and hybrid designs of fixed-tunable qubits exist as well. Nonetheless, flux-based gates remain standard in many superconducting systems, and in the next section the physics of the ZZ interaction and how it can be used to generate a CZ gate will be presented.

3.5.6. ZZ-Interaction

To see the ZZ-interaction, we need to understand the effect of the coupling capacitor, which can be treated perturbatively. Recall the equation for the second-order energy corrections [63]:

$$E_n^{(2)} = \sum_{n \neq m} \frac{\langle n|V|m\rangle \langle m|V|n\rangle}{E_n^{(0)} - E_m^{(0)}} \quad (3.101)$$

If we take the ground state $|00\rangle$ to have 0 energy, then we need to compute the new energies of the three states $\{|01\rangle, |10\rangle, |11\rangle\}$ in the presence of the capacitor, so $V = g(a^\dagger b + b^\dagger a)$. The single-occupation states only can couple to each other via the exchange of 1 photon, while the state $|11\rangle$ can couple to both $|20\rangle$ and $|02\rangle$, which lie outside the qubit space. The energies of the $|0\rangle, |1\rangle, |2\rangle$ states for a single transmon are $0, \omega, 2\omega - \alpha$. The perturbed energies are given below:

State	Unperturbed	Perturbations	Approximate Values
$ 01\rangle$	$\omega_{01}^{(0)}$	g^2/Δ	215 kHz
$ 10\rangle$	$\omega_{10}^{(0)}$	$-g^2/\Delta$	-215 kHz
$ 11\rangle$	$\omega_{01}^{(0)} + \omega_{10}^{(0)}$	$\frac{2g^2}{\alpha_1 + \Delta} + \frac{2g^2}{\alpha_2 - \Delta}$	352 kHz

Table 3.1: Table of energy corrections. It is interesting to note that if the anharmonicities are equal and opposite, then the $|11\rangle$ perturbation vanishes because the $|11\rangle$ level is pushed by equal and opposite amounts by the surrounding levels. This can be used to exactly cancel the ZZ interaction [37].

In the table, $\Delta/\hbar \equiv \omega_{01}^{(0)} - \omega_{10}^{(0)}$, the detuning between the two transmons' 0-th order lowest transition energies, and α_i is the 0-th order anharmonicity of transmon i . Note that the combined second-order perturbation to the sum $\omega_{01} + \omega_{10}$ is 0. Also, there are no first-order shifts because the perturbation is purely off-diagonal in the basis of unperturbed states. Without considering the $|2\rangle$ level, the state $|11\rangle$ would not experience a shift because it doesn't couple to other computational states. This shift means that $\omega_{11} \neq \omega_{01} + \omega_{10}$, and there is a conditional shift only if both transmons are excited. We define ζ to quantify this shift:

$$\zeta = \omega_{11} - \omega_{01} - \omega_{10} \quad (3.102)$$

$$= (\omega_{11}^{(0)} + \omega_{11}^{(2)}) - (\omega_{01}^{(0)} + \omega_{01}^{(2)}) - (\omega_{10}^{(0)} + \omega_{10}^{(2)}) \quad (3.103)$$

$$= \omega_{11}^{(2)} \quad (3.104)$$

Since the coupling is weak, the computational states are only slightly perturbed from their zeroth-order forms. They are approximately still the correct stationary states, so their Hamiltonian is approximately diagonal, with small energy shifts that can be computed up to some order:

$$H/\hbar \approx \begin{bmatrix} \omega_{00} & 0 & 0 & 0 \\ 0 & \omega_{01} & 0 & 0 \\ 0 & 0 & \omega_{10} & 0 \\ 0 & 0 & 0 & \omega_{01} + \omega_{10} + \zeta \end{bmatrix} \quad (3.105)$$

If one wants to obtain this result more rigorously, we refer to the derivation in Ku. et al [37]. There, the authors investigate the ZZ interaction's effect on the cross-resonance gate. Their device is a transmon coupled to a capacitively-shunted flux qubit, with a coupling resonator. Via a Schrieffer-Wolff transform (to remove the resonator from the effective dynamics), one obtains a Hamiltonian of the form [62]:

$$H_{eff} = \tilde{\omega}_{10}Z_1 + \tilde{\omega}_{01}Z_2 + \zeta Z_1Z_2 \quad (3.106)$$

$$= \tilde{\omega}_{10}Z_1 + (\tilde{\omega}_{01}I + \zeta Z_1)Z_2 \quad (3.107)$$

$$= Z_1(\tilde{\omega}_{10}I + \zeta Z_2) + \tilde{\omega}_{01}Z_2 \quad (3.108)$$

where $\tilde{\omega}$ indicates a dressed frequency. This result is transferrable in the sense that it was also derived for two statically-coupled anharmonic systems. We see that the ZZ interaction causes a frequency shift to one qubit whenever the other's state is known. If one qubit is an equal superposition, then the other qubit's precession looks like a sum of the two possible frequencies, and they become entangled.

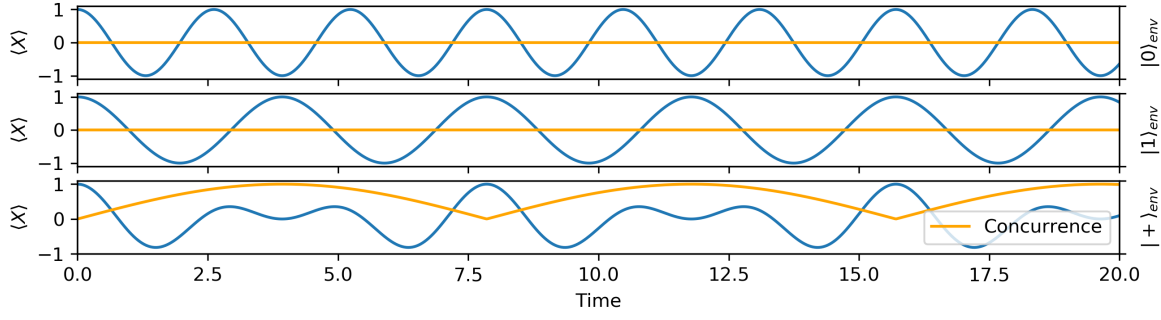


Figure 3.7: Evolution of $\langle \sigma_x \rangle$ over time, for three different environment states. Non-zero concurrence in the third plot indicates entanglement due to the ZZ-coupling (Eq. 3.105).

3.5.7. Controlled-Z Gate

Now we are ready to understand the design of a CZ-gate using the ZZ-interaction. One can slowly bias the flux line of one transmon (in a coupled pair) to cause the $|11\rangle$ energy level to approach an avoided crossing with the $|20\rangle$ level. The presence of this higher level affects the phase accrued by the state as it undergoes this adiabatic trajectory. Though the states $|01\rangle$ and $|10\rangle$ accrue different phases themselves, these may be transformed away with appropriate single-qubit Z-rotations, which admit a virtual (i.e. in software) implementation. However, if this trajectory is not perfectly adiabatic, the state of the qubit may “leak” out of the logical subspace and into a higher level, such as $|2\rangle$. Furthermore, even without leakage, the coupling that facilitates this gate is “always on” and contributes to spurious entanglement between coupled qubits via state-dependent Larmor frequency shifts. Tuning the flux away from the CZ operating point cannot turn off this interaction anymore than it can “turn off” the capacitors: the ZZ interaction can only be suppressed.

To understand the flux-tuning, we can diagonalize the system’s Hamiltonian over a range of different flux tunings. Though both transmons are flux-tunable, this example only requires tuning one of them; from here on out, it is assumed that the “left” transmon (1) is tuned, while the “right” transmon is parked at 0 flux bias. The resulting level structure in energy-flux space is plotted in Fig. 3.8. Observe that as the flux on transmon 1 is increased, the $|20\rangle$ level drops down and “pushes” against the $|11\rangle$ level. Normally, without coupling, these energy levels would simply cross unaffected, as the systems could be treated independently of each other. With coupling, the levels cannot cross, and the strength of the coupling determines how sharply the levels repel each other. Practically, what matters is that the energy of the $|11\rangle$ level is no longer equal to the sum of the $|10\rangle$ and $|01\rangle$ levels. We can use this deviation to perform the required gate. First, the flux bias is slowly tuned to the avoided crossing, and after some time, the bias is turned off by returning along the same trajectory. Assuming the system started in some fixed energy state, then the adiabatic theorem states that the system will remain in this state, even as the energy of that state slowly changes in time. Thus, the total phase accrued by some state $|ij\rangle$ is

$$\theta_{ij}(l(\tau)) = \int_0^\tau \omega_{ij}(l(\tau)) d\tau \quad (3.109)$$

where $l(\tau)$ is some trajectory in flux-time space, and τ is the total gate time. If sufficiently adiabatic, then the unitary we obtain in the qubit space is

$$U(l) = \begin{bmatrix} e^{i\theta_{00}(l)} & 0 & 0 & 0 \\ 0 & e^{i\theta_{01}(l)} & 0 & 0 \\ 0 & 0 & e^{i\theta_{10}(l)} & 0 \\ 0 & 0 & 0 & e^{i\theta_{11}(l)} \end{bmatrix} \quad (3.110)$$

This is just the time-evolution generated by the Hamiltonian argued at the end of the previous section. It is important to note that this avoided crossing is the first to occur in the level diagram (when moving out from 0 flux) for the 2-qubit level subspace, so that we do not have to worry about moving through other operating points while performing the CZ. Earlier, we defined the difference in phase accrued by

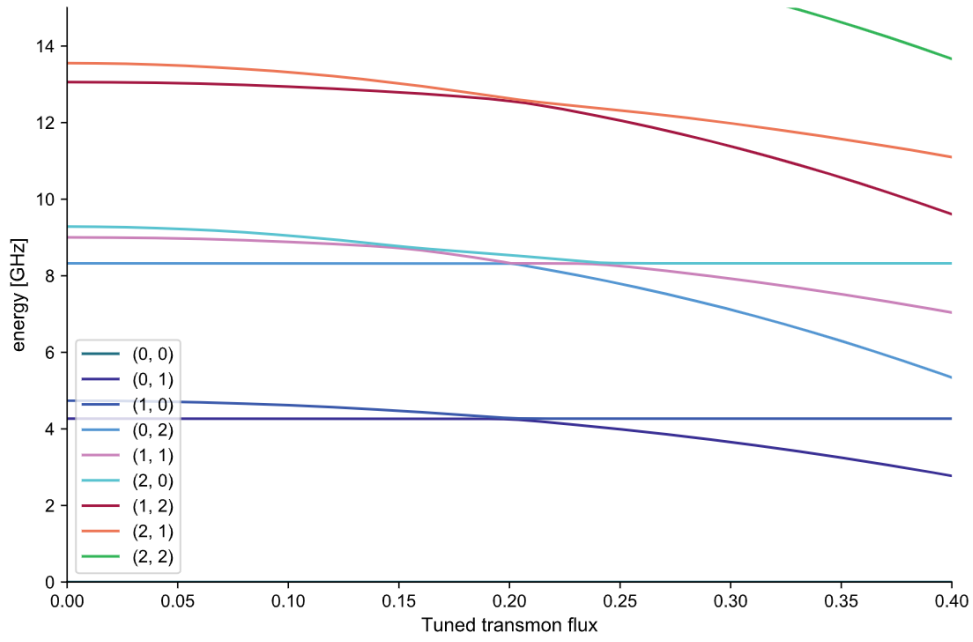


Figure 3.8: Level structure for coupled transmons, as the flux on the first transmon is tuned.

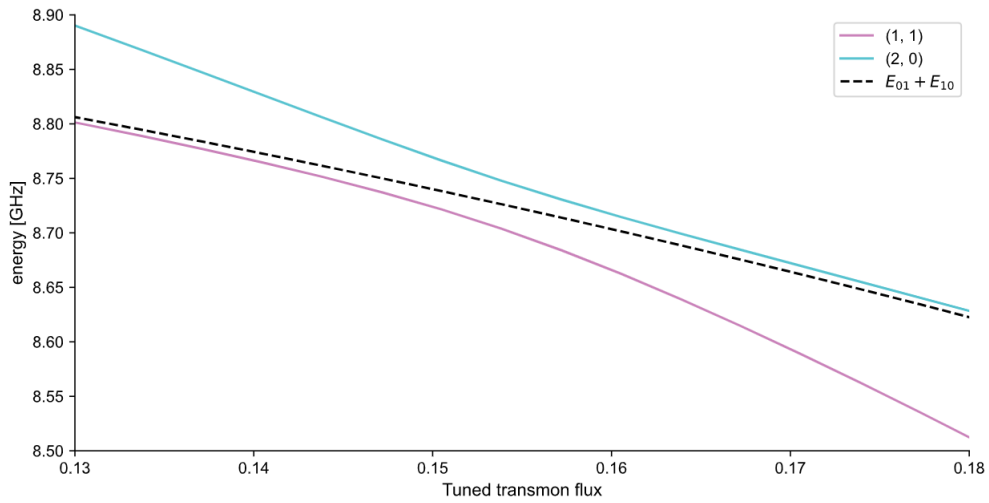


Figure 3.9: Zoomed level structure near relevant avoided crossing of $|20\rangle$ and $|11\rangle$, with dashed line showing the unperturbed level.

the $|11\rangle$ state relative to its single-occupation constituents

$$\zeta = \omega_{11} - (\omega_{10} + \omega_{01}) \quad (3.111)$$

Whatever this ζ rate is, we can define our phase trajectory such that the total phase *difference* accrued is π . With this, the above matrix becomes

$$U(l) = \begin{bmatrix} e^{i\theta_{00}(l)} & 0 & 0 & 0 \\ 0 & e^{i\theta_{01}(l)} & 0 & 0 \\ 0 & 0 & e^{i\theta_{10}(l)} & 0 \\ 0 & 0 & 0 & e^{i(\pi+\theta_{01}(l)+\theta_{10}(l))} \end{bmatrix} \quad (3.112)$$

No phase is accrued by the ground state $|00\rangle$ if we set it to be our reference, and redefine the other frequencies accordingly. We can transform away the next two elements on the diagonal with single-qubit Z-rotations

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \otimes \begin{bmatrix} 1 & 0 \\ 0 & e^{-i\theta_{01}(l)} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & e^{-i\theta_{01}(l)} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & e^{-i\theta_{01}(l)} \end{bmatrix} \quad (3.113)$$

$$\begin{bmatrix} 1 & 0 \\ 0 & e^{-i\theta_{10}(l)} \end{bmatrix} \otimes \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & e^{-i\theta_{10}(l)} & 0 \\ 0 & 0 & 0 & e^{-i\theta_{10}(l)} \end{bmatrix} \quad (3.114)$$

Compiling $U(l)$ with these extra gates yields the CZ gate.

$$\text{CZ} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} \quad (3.115)$$

Of course, choosing π was deliberate, and we could just as well desire any controlled phase ϕ , which can be achieved in the same way

$$\text{CZ}(\phi) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & e^{-i\phi} \end{bmatrix} = \exp(-i\frac{\phi}{4}(Z \otimes Z - I \otimes Z - Z \otimes I)) \quad (3.116)$$

A similar method can be used to generate an iSWAP gate from the avoided level crossing of $|01\rangle$ and $|10\rangle$, though it involves direct, diabatic oscillations across the crossing.

$$\text{iSWAP} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & -i & 0 \\ 0 & -i & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (3.117)$$

Both CZ and iSWAP can be used to generate entangled states. Figures 3.8 and 3.9 were generated using with a Python library called `scqubits`[35], which is an open-source toolbox for simulating superconducting circuits. This Python model is further investigated in section 4.3.3, where the Hamiltonian built in `scqubits` is then simulated in QuTiP (Quantum Toolbox in Python) [32].

4

Results

With all of the background elaborated, we now begin the main body of this report. The system in question will be a single qutrit, and we seek to reconstruct its evolution from a set of measurement data. In some experiment, one would first prepare a complete set of input states and immediately measure them in a complete basis—this characterizes the “null” process. Then, any further experiments (i.e. tomographies) could be run as desired. Before fitting the results of those experiments, one would first use the null process data to estimate the initial ground state, measurement POVM, and rotation pulses via maximum likelihood estimation, constituting a self-consistent set that describes all SPAM parameters. We will assume that the SPAM parameters do not change significantly throughout the following experiments, since the null process data is obtained just prior to executing them. Thus, once the SPAM parameters are known, they can be fixed in subsequent fits.

Rather than implementing a particular gate or quantum circuit, the experiment we simulate is simply the idle evolution of the qutrit, subject only to environmental noise. The motivation for this task is simply that qutrits are relatively unexplored by experimentalists, and one would first like to characterize their idle behavior before moving on to more complex situations like time-dependent drives and logic gates. A total duration time T and resolution Δt is chosen, and quantum process tomography is performed at each timestep $n\Delta t$. The resulting set of processes can be used to probe the Markovianity of the noise environment by applying the trace distance metric described in 3.2.7.

We first apply this method to the prototypical Markovian evolution of a qutrit subject to amplitude and phase damping, and confirm that the results match our expectations. Next, we tackle the device model described in section 3.5.3. The device consists of a linear chain of three capacitively coupled transmons, and we simulate the idle evolution of two of them. Modeling both transmons as qutrits, and treating one as the system and the other as the environment, we extract a set of idle processes over time, and infer properties of the qutrit-qutrit coupling by quantifying the Markovianity of the evolution.

In Fig. 4.1, the blue tiles represent the mathematical structure of open quantum systems. The green, yellow, and orange tiles represent the experiments described in this report that are necessary to reconstruct the evolution.

4.1. Estimating SPAM Errors

4.1.1. State Preparation

In the qutrit experiments discussed here, we require the ability to prepare pure input states to characterize the measurement POVM. In the MIT device, transitions between $|0\rangle$ and $|2\rangle$ require high power and are assumed to be forbidden, so we will rely on X and Y rotations within the $\{|0\rangle, |1\rangle\}$ and $\{|1\rangle, |2\rangle\}$ subspaces for state preparation. And since experimentalists do not typically characterize gates involving the $|2\rangle$ level, we will explicitly introduce a coherent error parameter that controls the degree of over/under rotation that occurs during pulses in the $\{|1\rangle, |2\rangle\}$ space. Single-qubit gates with high fidelity ($> 99.9\%$) are consistently achievable with current hardware, so these are assumed perfect [3, 33].

Specifically, of the Gell-Mann matrices, λ_1, λ_2 and λ_3 are the usual Paulis, λ_4 and λ_5 are forbidden (so they must be compiled from other gates), and λ_6 and λ_7 are accessible but faulty. The faulty rotations

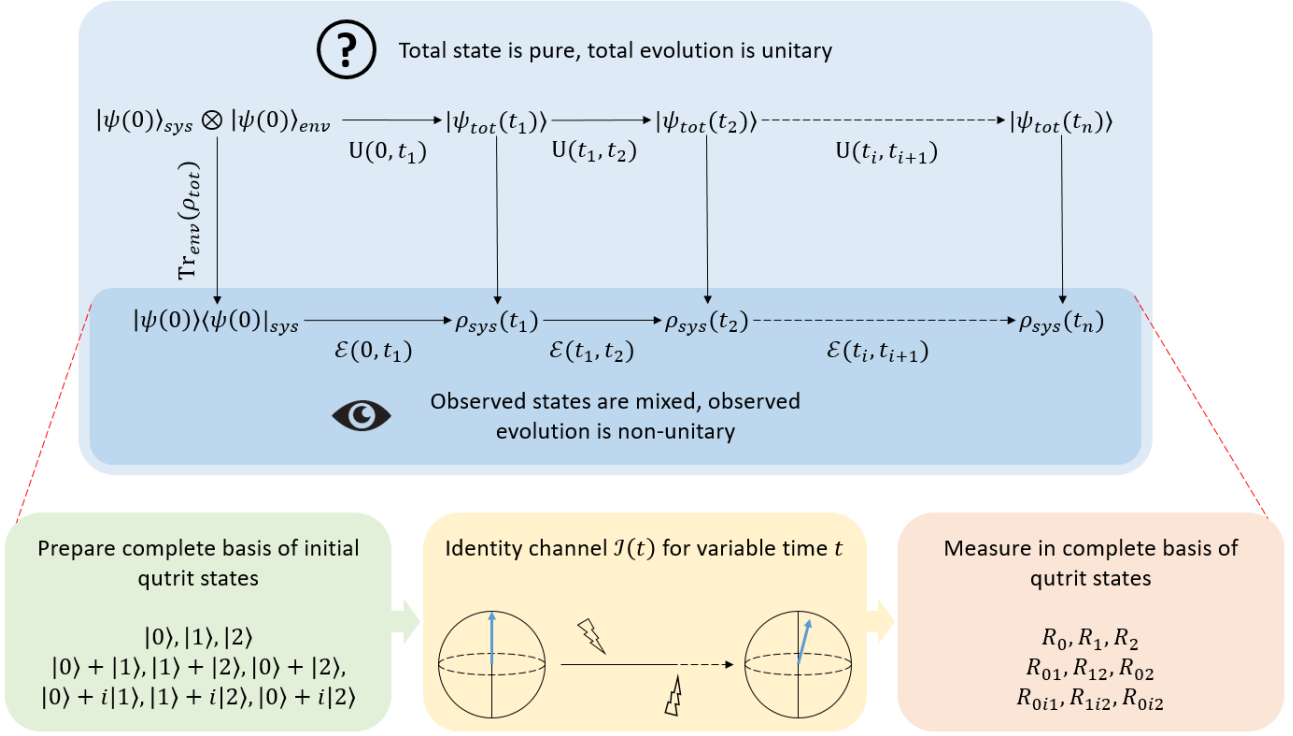


Figure 4.1: Schematic of open quantum systems and characterization pipeline.

are assumed to have the following form

$$R_x^{12}(\theta) = e^{-\frac{i}{2}(\theta + \epsilon_x)\lambda_6} \approx \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos((\theta + \epsilon_x)/2) & -i \sin((\theta + \epsilon_x)/2) \\ 0 & -i \sin((\theta + \epsilon_x)/2) & \cos((\theta + \epsilon_x)/2) \end{bmatrix} \quad (4.1)$$

$$R_y^{12}(\theta) = e^{-\frac{i}{2}(\theta + \epsilon_y)\lambda_7} \approx \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos((\theta + \epsilon_y)/2) & -\sin((\theta + \epsilon_y)/2) \\ 0 & \sin((\theta + \epsilon_y)/2) & \cos((\theta + \epsilon_y)/2) \end{bmatrix} \quad (4.2)$$

$$(4.3)$$

We need to construct gate sequences for preparing a complete set of states. They are “complete” in the sense that they constitute a minimal set of states whose measurement statistics are sufficient to constrain the matrix elements of the qutrit POVM. For a qubit, one such complete set is $\{|0\rangle, |1\rangle, |+\rangle, |+i\rangle\}$ (3.4.1). By analogy, then, one possible complete set for a qutrit is

$$\{|0\rangle, |1\rangle, |2\rangle, |+01\rangle, |+i01\rangle, |+12\rangle, |+i12\rangle, |+02\rangle, |+i02\rangle\} \quad (4.4)$$

where + and +i indicate states that lie along X and Y-type axes of sub-Bloch spheres. The below table contains the gate sequences $\{\tilde{R}_i\}$ needed to prepare the complete set. The tilde again refers to the fact that some of the rotations are faulty.

State	Gate Sequence	Restricted Gate Sequence
$ 0\rangle$	I	I
$ 1\rangle$	X_{π}^{01}	X_{π}^{01}
$ 2\rangle$	$\tilde{X}_{\pi}^{02}$	$\tilde{X}_{-\pi}^{12} X_{\pi}^{01} \tilde{X}_{\pi}^{12}$
$ +01\rangle$	$Y_{\pi/2}^{01}$	$Y_{\pi/2}^{01}$
$ +i01\rangle$	$X_{\pi/2}^{01}$	$X_{\pi/2}^{01}$
$ +12\rangle$	$\tilde{Y}_{\pi/2}^{12} X_{\pi}^{01}$	$\tilde{Y}_{\pi/2}^{12} X_{\pi}^{01}$
$ +i12\rangle$	$\tilde{X}_{\pi/2}^{12} X_{\pi}^{01}$	$\tilde{X}_{\pi/2}^{12} X_{\pi}^{01}$
$ +02\rangle$	$\tilde{Y}_{\pi/2}^{02}$	$\tilde{X}_{-\pi}^{12} Y_{\pi/2}^{01} \tilde{X}_{\pi}^{12}$
$ +i02\rangle$	$\tilde{X}_{\pi/2}^{02}$	$\tilde{X}_{-\pi}^{12} X_{\pi/2}^{01} \tilde{X}_{\pi}^{12}$

Table 4.1: Table of state preparation sequences. Tildes indicate explicitly faulty rotations.

To unpack this table a bit, first note that the usual qubit 01 subspace rotations require no modification from their usual forms. Any state preparation involving the $|2\rangle$ state is faulty by assumption. Preparing $(+)$ -type states requires rotating about a Y -axis, and $(+i)$ -type states require rotations about an X -axis. We always begin with the vacuum state $|0\rangle$, but the sub-Bloch sphere that contains $|+12\rangle$ and $|+i12\rangle$ does not include the vacuum, so we must first use a π -pulse (X_{π}^{01}) to access this subspace. The restricted gate sequence column is the result of not being able to directly drive rotations in the 02 subspace. As an example, consider the gate sequence responsible for converting $|0\rangle \leftrightarrow |2\rangle$. If we apply this to a general state, we get

$$\tilde{X}_{\pi}^{02} |\psi\rangle = \tilde{X}_{-\pi}^{12} X_{\pi}^{01} \tilde{X}_{\pi}^{12} (a |0\rangle + b |1\rangle + c |2\rangle) \quad (4.5)$$

$$= \tilde{X}_{-\pi}^{12} X_{\pi}^{01} (a |0\rangle + b |\tilde{2}\rangle + c |\tilde{1}\rangle) \quad (4.6)$$

$$= \tilde{X}_{-\pi}^{12} (a |1\rangle + b |\tilde{2}\rangle + c |\tilde{0}\rangle) \quad (4.7)$$

$$= a |\tilde{2}\rangle + b |1\rangle + c |\tilde{0}\rangle \quad (4.8)$$

Indeed, $|2\rangle$ and $|0\rangle$ are swapped (albeit imperfectly), but since the error is in the rotation angle and is assumed to remain constant after calibration and over the course of the subsequent experiment, it cancels out on the $|1\rangle$ state, leaving it unchanged as it should, since $|1\rangle$ is not accessible via 02-type gates. Below, we give similar proofs for the preparations of the other restricted states $|+02\rangle$ and $|+i02\rangle$.

$$\tilde{Y}_{\pi/2}^{02} |\psi\rangle = \tilde{X}_{-\pi}^{12} Y_{\pi/2}^{01} \tilde{X}_{\pi}^{12} (a |0\rangle + b |1\rangle + c |2\rangle) \quad (4.9)$$

$$= \tilde{X}_{-\pi}^{12} Y_{\pi/2}^{01} (a |0\rangle + b |\tilde{2}\rangle + c |\tilde{1}\rangle) \quad (4.10)$$

$$= \tilde{X}_{-\pi}^{12} \left(\frac{a}{\sqrt{2}} (|0\rangle + |1\rangle) + b |\tilde{2}\rangle + \frac{c}{\sqrt{2}} (|\tilde{0}\rangle - |\tilde{1}\rangle) \right) \quad (4.11)$$

$$= \frac{a}{\sqrt{2}} (|0\rangle + |\tilde{2}\rangle) + b |1\rangle + \frac{c}{\sqrt{2}} (|\tilde{0}\rangle - |2\rangle) \quad (4.12)$$

$$\tilde{X}_{\pi/2}^{02} |\psi\rangle = \tilde{X}_{-\pi}^{12} X_{\pi/2}^{01} \tilde{X}_{\pi}^{12} (a |0\rangle + b |1\rangle + c |2\rangle) \quad (4.13)$$

$$= \tilde{X}_{-\pi}^{12} X_{\pi/2}^{01} (a |0\rangle + b |\tilde{2}\rangle + c |\tilde{1}\rangle) \quad (4.14)$$

$$= \tilde{X}_{-\pi}^{12} \left(\frac{a}{\sqrt{2}} (|0\rangle + i |1\rangle) + b |\tilde{2}\rangle + \frac{c}{\sqrt{2}} (|\tilde{0}\rangle - i |\tilde{1}\rangle) \right) \quad (4.15)$$

$$= \frac{a}{\sqrt{2}} (|0\rangle + i |\tilde{2}\rangle) + b |1\rangle + \frac{c}{\sqrt{2}} (|\tilde{0}\rangle - i |2\rangle) \quad (4.16)$$

By conjugating the desired pulse with π -pulses in the 12 space, $|1\rangle$ and $|2\rangle$ are exchanged, and whichever gate we have chosen for the 01 space (applied in the middle) gets mapped to the 02 space, as desired [9]. In Table 4.1, there are four faulty rotations: $\{\tilde{X}_{\pi}^{12}, \tilde{X}_{-\pi}^{12}, \tilde{X}_{\pi/2}^{12}, \tilde{Y}_{\pi/2}^{12}\}$. Since the second is the adjoint of the

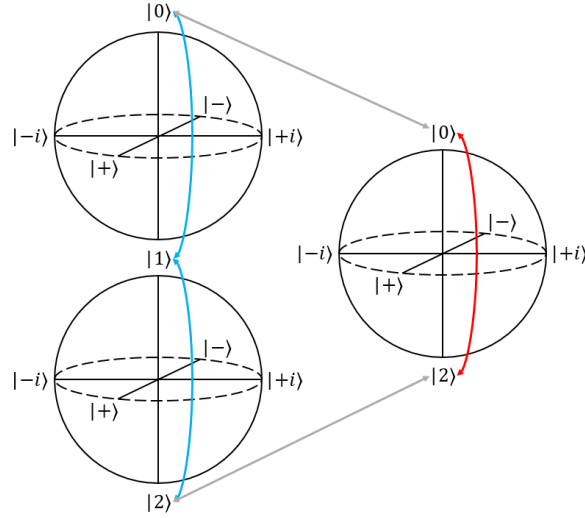


Figure 4.2: A mental picture for the qutrit space as considered in this work. Note: this image is not meant to be a faithful representation of $SU(3)$. Red indicates a forbidden transition.

first, there are three unique faulty rotations:

$$R_x^{12}(\pi) \equiv \tilde{X}_{\pi}^{12} = e^{-\frac{i}{2}(\pi+\epsilon_1)\lambda_6} \quad (4.17)$$

$$R_x^{12}(\pi/2) \equiv \tilde{X}_{\pi/2}^{12} = e^{-\frac{i}{2}(\pi/2+\epsilon_2)\lambda_6} \quad (4.18)$$

$$R_y^{12}(\pi/2) \equiv \tilde{Y}_{\pi/2}^{12} = e^{-\frac{i}{2}(\pi/2+\epsilon_3)\lambda_7} \quad (4.19)$$

Next, we also would like to account for imperfect vacuum state preparation, where the initial $|0\rangle$ that we start with is not pure, but mixed:

$$\rho_{\text{vac}} = a|0\rangle\langle 0| + b|1\rangle\langle 1| + c|2\rangle\langle 2| \quad (4.20)$$

If we assume that ground states are prepared via thermalizing with a cold environment, then we can calculate the mixed state probabilities using the Boltzmann distribution [31]

$$P_i = \frac{1}{\mathcal{Z}} g_i \exp(-E_i/k_B T) \quad (4.21)$$

where the state degeneracy g_i is 1 for our purposes, and $\mathcal{Z} = \sum_i P_i$ is the partition function. Using the QuTiP transmon model 3.5.3, and assuming equilibrium with a $T = 50\text{mK}$ environment (typical range for the effective temperature is 50-130mK [31], including hot quasiparticle noise), the following state energies and occupations are obtained

State	Energy	$P_i(\%)$
$ 00\rangle$	0 GHz	97.29
$ 01\rangle$	4.27 GHz	1.62
$ 10\rangle$	4.74 GHz	1.03
$ 02\rangle$	8.32 GHz	0.03
$ 11\rangle$	9.00 GHz	0.02
$ 20\rangle$	9.28 GHz	0.01
$ 12\rangle$	13.06 GHz	$\sim 10^{-4}$
$ 21\rangle$	13.55 GHz	$\sim 10^{-4}$
$ 22\rangle$	17.61 GHz	$\sim 10^{-6}$

Table 4.2: Thermal state populations of the two-qutrit transmon system.

Natural reset (i.e. waiting) is slow and can be limited by hot photons from higher dilution fridge stages [31, 22]. Active reset protocols use calibrated drive tones to pump out photons into lossy resonators, quickly taking the system to its ground state. The best of these protocols can achieve ground state initialization with excited state population ($|1\rangle$ and higher) less than 1% [41], corresponding to an effective temperature of 40mK in our model. Though such low excited state populations can be achieved, in the following simulations we will be very pessimistic and assume that the excited state parameters b and c are both on the order of 10^{-1} . This demonstrates the robustness of the simulation to extremely faulty measurement schemes, and is “intentionally bad” in the sense that very small values for b and c would produce data nearly indistinguishable from a perfect system, and we *want* to make sure that the simulation can actually detect and estimate the faulty parameters.

4.1.2. Measurement

We also need to characterize the POVM of the measurement process. $\tilde{E}_i$ will refer to the actual set (POVM), and E_i to the ideal set (PVM):

$$E_0 = |0\rangle\langle 0| \quad E_1 = |1\rangle\langle 1| \quad E_2 = |2\rangle\langle 2| \quad (4.22)$$

The complete set of states chosen above will allow us to reconstruct the matrix elements of the qutrit POVM. To justify its choice, first recall that the probability of obtaining a measurement outcome for some state ρ is given by the Born rule:

$$p_i = \text{Tr } E_i \rho \quad (4.23)$$

These probabilities may be estimated via many repeated measurements, giving $\hat{p}_i = \frac{m_i}{N}$. In the limit of an infinite number of experiments, the estimate $\hat{p}_i$ approaches the true value p_i . If the density matrices we prepare are basis elements of this space, then the measured probabilities are estimates of the matrix elements of E_i . Thus, we only need to choose a set of ρ s that can fix all the matrix elements. The natural basis $\{|i\rangle\langle j|\}$ is one such operator basis—the Paulis are another—but since both bases contain some traceless elements, neither as a whole constitute a set of physical states, and we need to pick some density matrices such that linear combinations thereof form the chosen basis.

For a qubit, the projectors $\text{Tr}(\tilde{E}|0\rangle\langle 0|) = \tilde{E}_{00}$ and $\text{Tr}(\tilde{E}|1\rangle\langle 1|) = \tilde{E}_{11}$ are fine as is, so the diagonal matrix elements are immediately obtained. For the others, we can use $|+\rangle\langle +|$ and $|-\rangle\langle -|$:

$$\text{Tr}(\tilde{E}|+\rangle\langle +|) = \frac{1}{2}(\tilde{E}_{00} + \tilde{E}_{01} + \tilde{E}_{10} + \tilde{E}_{11}) \quad (4.24)$$

$$\text{Tr}(\tilde{E}|\tilde{M}|+i\rangle\langle +i|) = \frac{1}{2}(\tilde{E}_{00} + i\tilde{E}_{01} - i\tilde{E}_{10} + \tilde{E}_{11}) \quad (4.25)$$

Normally the POVM elements are Hermitian (because they are PSD), so we could assume that $\tilde{E}_{ij} = \tilde{E}_{ji}^*$, but this most likely won't be true just based on the experimental data alone (these are $\tilde{E}$, not E). Instead, we could first subtract the diagonal elements $\tilde{E}_{00}$ and $\tilde{E}_{11}$, and then proceed as below

$$\text{Tr}(\tilde{E}|+\rangle\langle +|) - \frac{1}{2}(\tilde{E}_{00} + \tilde{E}_{11}) = \frac{1}{2}(\tilde{E}_{01} + \tilde{E}_{10}) \equiv a \quad (4.26)$$

$$\text{Tr}(\tilde{E}|\tilde{M}|+i\rangle\langle +i|) - \frac{1}{2}(\tilde{E}_{00} + \tilde{E}_{11}) = \frac{1}{2}(i\tilde{E}_{01} - i\tilde{E}_{10}) \equiv b \quad (4.27)$$

$$a - ib = \tilde{E}_{01} \quad (4.28)$$

$$a + ib = \tilde{E}_{10} \quad (4.29)$$

In this way, we can find the off-diagonals as well. We can easily extend this to the qutrit case. Since we have decomposed the qutrit space into two Bloch spheres, the complete set of states is two copies of what we just saw.

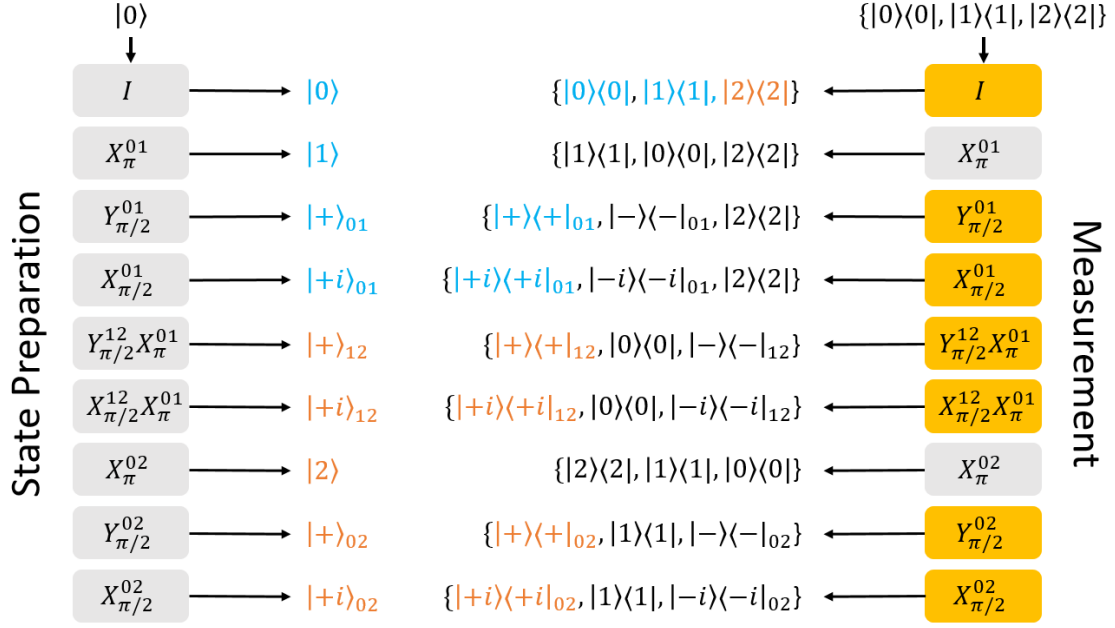


Figure 4.3: Cyan states and projectors are complete for qubit measurement and state tomography, respectively. Orange states and projectors (in addition to cyan) are complete for qutrit measurement and state tomography, respectively. Gold gates applied to the base POVM produce new POVMs that contain at least one necessary projector for completeness.

$ \psi\rangle$	$\text{Tr}(\tilde{E} \psi\rangle \langle \psi)$
$ 0\rangle$	$\tilde{E}_{00}$
$ 1\rangle$	$\tilde{E}_{11}$
$ 2\rangle$	$\tilde{E}_{22}$
$\frac{1}{\sqrt{2}}(0\rangle + 1\rangle)$	$\frac{1}{2}(\tilde{E}_{00} + \tilde{E}_{01} + \tilde{E}_{10} + \tilde{E}_{11})$
$\frac{1}{\sqrt{2}}(0\rangle + i 1\rangle)$	$\frac{1}{2}(\tilde{E}_{00} + i\tilde{E}_{01} - i\tilde{E}_{10} + \tilde{E}_{11})$
$\frac{1}{\sqrt{2}}(1\rangle + 2\rangle)$	$\frac{1}{2}(\tilde{E}_{11} + \tilde{E}_{12} + \tilde{E}_{21} + \tilde{E}_{22})$
$\frac{1}{\sqrt{2}}(1\rangle + i 2\rangle)$	$\frac{1}{2}(\tilde{E}_{11} + i\tilde{E}_{12} - i\tilde{E}_{21} + \tilde{E}_{22})$
$\frac{1}{\sqrt{2}}(0\rangle + 2\rangle)$	$\frac{1}{2}(\tilde{E}_{00} + \tilde{E}_{02} + \tilde{E}_{20} + \tilde{E}_{22})$
$\frac{1}{\sqrt{2}}(0\rangle + i 2\rangle)$	$\frac{1}{2}(\tilde{E}_{00} + i\tilde{E}_{02} - i\tilde{E}_{20} + \tilde{E}_{22})$

Table 4.3: Complete set of states and theoretical form of their measurement probabilities.

The measurement outcomes of a complete set of states are sufficient to solve for the matrix elements of the POVM. Since each POVM matrix has nine degrees of freedom (3 real parameters on the diagonal, and 3 complex off-diagonals of a Hermitian matrix), we must prepare at minimum nine input states to create nine data constraints to fix those elements. For each input state, a set of measurements is performed in the Z basis, yielding a measurement record that consists of some number of 0s, 1s, and 2s. All of the 0 counts contribute to fixing the elements of $\tilde{E}_0$, and likewise for the rest (if we obtain outcome i , then by post-selection we have measured with the operator $\tilde{E}_i$).

4.1.3. Measuring in Other Bases

To perform process tomography, we need to measure in a complete basis as well. To do this, we can just apply the same set of state preparation rotations R_i that we found above to the POVM. We will need as many different rotations as necessary to guarantee that among the operators of the rotated POVMs, the same 9 projectors appear. This is best summarized by the following graphic: Only the gold gates in

Fig. 4.3 are actually applied before measurement when simulating our procedure ¹. Once we apply the same rotations to both the ground state and the POVM to ensure completeness, we can express the tomography of the null process as

$$p_{ijk} = \text{Tr}(R_j E_k R_j^\dagger R_i |0\rangle \langle 0| R_i^\dagger) \quad (4.30)$$

which is the probability of getting result k from the i th input state measured the basis accessed by the j th rotation of the POVM. Of course, the model we have built assumes that the ground state is a thermal state, the rotations involving $|2\rangle$ have a coherent (systematic) error, and the POVM is unknown. The probabilities actually obtained are then

$$\tilde{p}_{ijk} = \text{Tr}(\tilde{R}_j \tilde{E}_k \tilde{R}_j^\dagger \tilde{R}_i |\tilde{0}\rangle \langle \tilde{0}| \tilde{R}_i^\dagger) \quad (4.31)$$

In QPT, the χ matrix has $d^4 - d^2$ free parameters, which is 72 for qutrits. There are 9 input states and 9 necessary projectors, yielding 81 constraints. Including all 15 unique measurement projectors (2 X-type states $\times$ 2 Y-type states $\times$ 3 sub-Bloch spheres, + 3 Z-basis projectors), yields 135 constraints in total. In the next section, we will use MLE to (over)constrain χ .

4.1.4. Caveat

Referring to the theoretical form of the probabilities (Eq. 4.31), assume for a moment that the only imperfection lies in the POVM

$$\tilde{p}_{ijk} = \text{Tr}(\tilde{E}_k R_j^\dagger R_i |0\rangle \langle 0| R_i^\dagger R_j) \quad (4.32)$$

$$= \text{Tr}(C(E_k) R_j^\dagger R_i |0\rangle \langle 0| R_i^\dagger R_j) \quad (4.33)$$

where C is some process that corrupts the k th ideal into the k th faulty POVM operator. In the vectorized notation, we have

$$\tilde{p}_{ijk} = \langle \langle E_k | G_C^\dagger G_R^{ij} | 0 \rangle \rangle \quad (4.34)$$

where G_C is the vectorized form of C , and G_R^{ij} is the vectorized form of the two rotations applied to the initial vacuum state. In general, it is impossible to distinguish whether the error stems from the POVM, or, if $[G_R^{ij}, G_C] = 0$, it originates in a faulty initial vacuum state. It could also be misconstrued as some error in the rotations G_R^{ij} , but a SPAM-insensitive procedure like randomized benchmarking could help detect this case. This ambiguity is intrinsic to all characterization techniques (it is a form of gauge freedom [26]), since we can never separate preparation and measurement from each other.

4.1.5. Maximum Likelihood Estimation

In section 3.4, we discussed different algorithms for extracting information about our quantum states and gates. All of these involve fitting experimental data to some model; the more expressive methods generally involve solving some kind of nonlinear optimization. One such optimization, called maximum likelihood estimation (MLE), is frequently used for its ability to naturally incorporate physicality constraints and overcomplete datasets.

In any experiment, we do not have access to the true, underlying probability distribution that governs a set of values obtained from the measurement of a random variable. At most, we have a mathematical model with some number of free parameters for the distribution, and a dataset that the model must conform to. The goal of maximum likelihood estimation is to find the parametrization of the model that is *most likely* to have produced the observed data. For our purposes, the unknown distribution is parametrized by the set of all POVM matrix elements, the thermal state parameters a, b, c , and the coherent error parameters $\epsilon_1, \epsilon_2, \epsilon_3$. The distribution itself comes down to choosing a likelihood function. We will simply define it as the probability to obtain the observed outcomes as independent trials ².

$$\mathcal{L} = \prod_{t \in \text{trials}} [(p_0)^{m_0} (p_1)^{m_1} (p_2)^{m_2}]_t = \prod_{s,b} (\text{Tr} \tilde{E}_{0,b} \tilde{\rho}_s)^{m_0} (\text{Tr} \tilde{E}_{1,b} \tilde{\rho}_s)^{m_1} (\text{Tr} \tilde{E}_{2,b} \tilde{\rho}_s)^{m_2} \quad (4.35)$$

¹So technically in our MATLAB routine, the set of initial rotations contains 9 gates, but the set of measurement rotations only contains 7, but we refer to both as $\{R_i\}$.

²One can also derive from the central limit theorem a likelihood function in the form of least-squares minimization [26]

where t enumerates the trials used to build up the measurement results. A trial consists of choosing an input state and a measurement pre-rotation, indexed by s and b , respectively. The number of outcomes of type i observed in a trial is m_i . Since the logarithm is monotonic, we can make the likelihood function more “well-behaved” by replacing it with the log-likelihood

$$\log \mathcal{L} = \sum_{s,b} m_0 \log(\text{Tr } \tilde{E}_{0,b} \tilde{\rho}_s) + m_1 \log(\text{Tr } \tilde{E}_{1,b} \tilde{\rho}_s) + m_2 \log(\text{Tr } \tilde{E}_{2,b} \tilde{\rho}_s) \quad (4.36)$$

Without the log, the likelihood $\mathcal{L}$ would numerically approach zero, since $0 \leq p_i \leq 1$ and the number of shots per trial is large. We use MATLAB’s `fmincon` solver with the default interior-point algorithm to minimize the negative log-likelihood, subject to the following constraints:

$$a + b + c = 1 \quad (4.37)$$

$$0 \leq a \leq \gamma_a \quad 0 \leq b \leq \gamma_b \quad 0 \leq c \leq \gamma_c \quad (4.38)$$

$$|\epsilon_1| \leq \gamma_4 \quad |\epsilon_2| \leq \gamma_5 \quad |\epsilon_3| \leq \gamma_6 \quad (4.39)$$

$$\sum_i \tilde{E}_i = I \quad (4.40)$$

The initial state must be normalized, and γ_i is an upper bound that should be chosen based on one’s rough knowledge of the physical system (i.e. ground and excited state populations after some initialization procedure, and gate fidelity of faulty rotations). The operators $\tilde{E}_i$ must constitute a valid POVM and sum to the identity. The positive semidefinite property of $\tilde{E}_i$ is implicitly enforced because we parametrize it in terms of a Cholesky decomposition (B.4):

$$\tilde{E}_i = LL^\dagger \quad (4.41)$$

$$L = \begin{bmatrix} t_1 & 0 & 0 \\ t_2 + it_3 & t_4 & 0 \\ t_5 + it_6 & t_7 + it_8 & t_9 \end{bmatrix} \quad (4.42)$$

where all three sets of 9 t -parameters are optimized along with $a, b, c, \epsilon_1, \epsilon_2, \epsilon_3$ ³.

In nonlinear optimization, we must also provide an initial guess for each unknown parameter. The closer the guess is to the true optimum, the better our chances are of not getting stuck in a local minimum. If the hardware is good enough, guessing $a = 1, b = c = \epsilon_1 = \epsilon_2 = \epsilon_3 = 0$ should be fairly close to reality. We initialize $\tilde{E}_i$ with data from trials where each of the initial states is prepared, and then measured in the Z -basis—that is, we use a subset of the data that represents measurement tomography. The function `NearestSPD` from the MATLAB library QETLAB (Quantum Entanglement Theory LABoratory) then finds the closest PSD matrix to each estimated $\tilde{E}_i$. At this point, they do not sum to the identity, which is sufficient because we only need a rough guess.

In the discussion of how to construct complete states and bases, we verified our choices by showing how to fix all the unknown matrix elements essentially by performing experiments that simulate operator projections. This is not to say that MLE literally is doing basic sums and differences under the hood, but just that the data fed to MLE is sufficient to uniquely constrain the values of interest in some way. In the next section, we evaluate how well the MLE methodology performs.

4.2. Verifying the Estimation

The goal of this section is to benchmark how well the MLE routine can estimate the SPAM parameters described so far. We generated three data sets, where each data set consists of 100 samples of parameter settings of the form `[a, b, c, e1, e2, e3, nip]`. NIP stands for “near-ideal” POVM, which is generated by adding Gaussian noise ($\sigma = 0.1$) to the ideal PVM. To ensure PSD, we again use `NearestSPD`. To enforce the trace constraint, we can reset one element of the POVM in the following way to correct this

$$\tilde{E}_k \rightarrow I - \tilde{E}_i - \tilde{E}_j \quad i \neq j \neq k \quad (4.43)$$

Of course, the difference between two PSD matrices may not be PSD, so we must use `NearestSPD` again, and iterate the two fixes via the following pseudocode

³MATLAB’s “chol” function returns an upper triangular matrix instead.

```

GENERATE_NIP():
nip = PVM + Gaussian noise
while nip is not valid:
nip[0] = I - nip[1] - nip[2]
nip[0] = NearestSPD(nip[0])
nip[1] = I - nip[0] - nip[2]
nip[1] = NearestSPD(nip[1])
nip[2] = I - nip[0] - nip[1]
nip[2] = NearestSPD(nip[2])

```

This loop typically takes 1-5 iterations to produce a valid POVM, but rarely can loop forever (reason unknown). This case can be easily caught and the process restarted. The result is a valid ground-truth POVM that we want to estimate.

The parameters b and c are both chosen uniformly between 0 and 0.1, and $a = 1 - b - c$. The three data sets differ in their scaling of the coherent errors, where $\epsilon_1, \epsilon_2, \epsilon_3$ are all sampled from a normal distribution centered at 0, with standard deviations $\sigma = 0.01, 0.05$, and 0.1 for the three data sets, respectively. Then, each parameter set is used to generate 5000 shots per rotation per initial state, and these shots are used to create an estimated parameter set $[a', b', c', e1', e2', e3', nip']$ with `fmincon` (primes indicate estimates). The pseudocode for the whole procedure is

```

FIT_DATA(sigma):
[a,b,c,e1,e2,e3,nip] = sample_parameters(sigma)
thermal_state = a|0><0| + b|1><1| + c|2><2|
Rs = rotations(e1,e2,e3)
initial_states = complete_states(Rs, thermal_state)
shots = get_shots(nip, initial_states, shots=1000, Rs);
[a',b',c',e1',e2',e3',povm'] = extract_params_from_shots(shots);

```

We will examine the second dataset, where $\sigma = 0.05$, which corresponds to a coherent error of roughly 2.9 degrees. All errors are computed with bounded norms that have been normalized to $[0, 1]$. All stated percentages refer to this range, scaled by 100 (i.e. we are not computing relative percent errors that may exceed 100%). This convention is chosen to avoid situations where the relative error is abnormally high, i.e. if the true value of a parameter is 0, then any nonzero MLE output for that parameter technically would have infinite relative error.

In Fig. 4.4, the y -axis is the trace distance (B.5.1) between the estimated thermal state $\rho(a', b', c')$ and the actual thermal state $\rho(a, b, c)$, and the x -axis is the trace distance from the actual thermal state to the ideal case $\rho(1, 0, 0)$. The plot shows how the estimation error varies as a function of the non-ideality of the simulated, ground-truth data. We see that the 100 samples span a range of 0 to 0.2, noting that 1 is the maximum possible trace distance between any two states. Over this range, the estimation error is roughly constant, with an average value of just below 10^{-2} , so the thermal parameters discovered by maximizing the likelihood result in initial states only 1% different from their true forms.

In Fig. 4.5, the y -axis is the spectral norm (B.5.2) of the difference between the estimated rotation $U(\epsilon')$ and the actual rotation $U(\epsilon)$, and the x -axis is the spectral norm of the difference between the actual rotation $U(\epsilon)$ and the ideal rotation $U(0)$. For a difference of unitaries, the spectral norm is bounded between 0 and 2⁴, so the norms have been divided by 2 so that the error range is $[0, 1]$. We see that the estimation error ranges from roughly 10^{-2} to 10^{-4} at all values of the true deviation from the ideal case, with an average value of around 0.1% for all three parameters. Recall that the three faulty rotations are

$$R_x^{12}(\pi + \epsilon_1) \quad (4.44)$$

$$R_x^{12}(\pi/2 + \epsilon_2) \quad (4.45)$$

$$R_y^{12}(\pi/2 + \epsilon_3) \quad (4.46)$$

The data appears clumped because the coherent errors that label the datasets are Gaussian-distributed.

In Fig. 4.6, the y -axis is the diamond norm (B.5.3) of the difference between the estimated POVM and the actual POVM, and the x -axis is the diamond norm of the difference between the actual POVM and

⁴This was confirmed numerically, and also analytically in B.5.2.

the ideal PVM. The diamond norm is a bounded metric used to compare quantum channels. As we saw in section 3.3, a POVM defines a set of Kraus operators, where a measurement result lets us post-select and choose which measurement operator to apply. If the measurement is lost, we must take a weighted sum over the operators with their respective probabilities, which is a quantum channel. If we convert the POVM operators to measurement operators via $\sqrt{E_i} \rightarrow M_i$, we can then use the diamond norm to compare the estimated and actual POVMs via their Kraus sets $\{M_i\}$. The average error comes out to around 7%, and is stable across the range of near-ideal-POVMs. If one simply averages over the spectral norm of the differences between the estimated and actual POVMs, then the average error is 0.0324 for this dataset. Although this measure seems more optimistic, it is harder to interpret because it is only bounded from below by 0. We used the stated metrics because they are upper bounded as well, so their outputs can be interpreted as percentages of some maximum possible distance. Furthermore, the trace distance and diamond norm ⁵ represent “worst-case” situations, where a semidefinite program finds either the measurement or the state that maximizes the distinguishability ⁶.

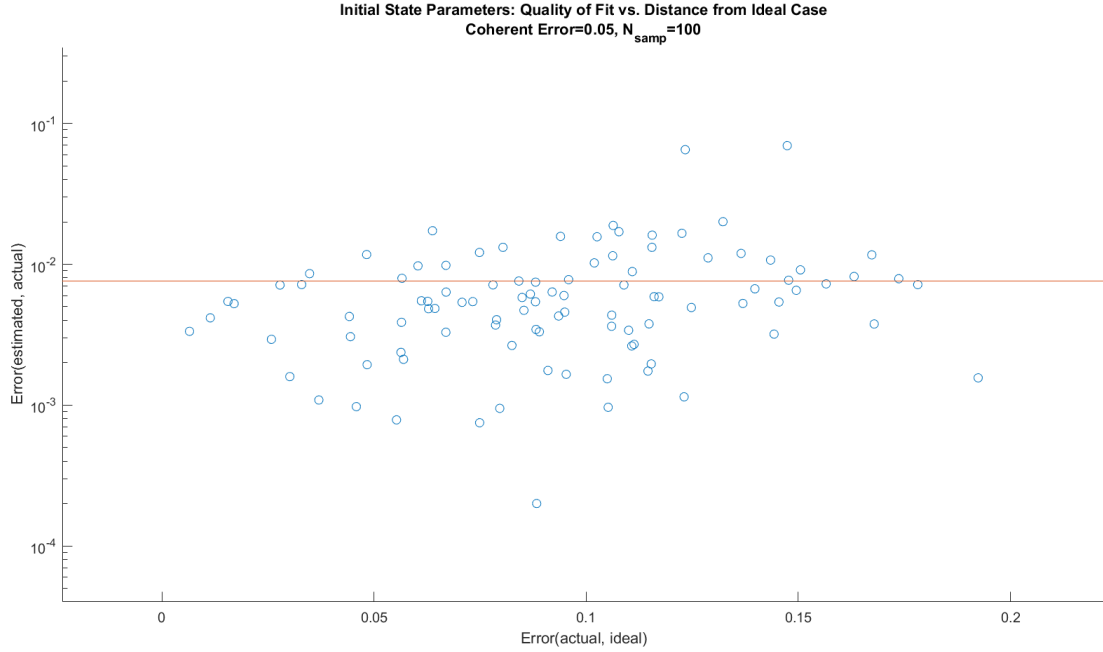


Figure 4.4: The y -axis is the trace distance between the estimated and actual initial states, and the x -axis is the trace distance between the actual and ideal initial states. Symbolically, this is $D(\rho', \rho)$ vs. $D(\rho, |0\rangle\langle 0|)$

⁵For small coherent errors, the diamond norm is very nearly the same as the spectral norm.

⁶There are simpler ways of computing the trace distance, at least, without convex optimization.

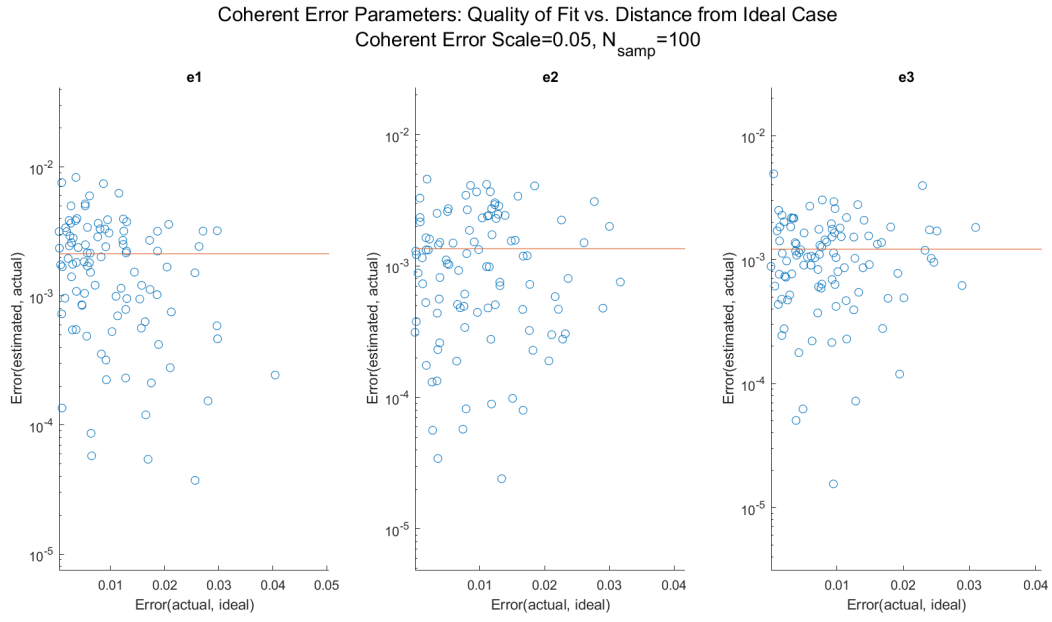


Figure 4.5: Each subplot shows the estimation error for one of the three coherent parameters ϵ . The y -axis is the spectral norm of the difference between the estimated and actual rotations, and the x -axis is the spectral norm of the difference between the actual and ideal rotations. Symbolically, this is $\|U(\epsilon') - U(\epsilon)\|_2$ vs. $\|U(\epsilon) - U(0)\|_2$.

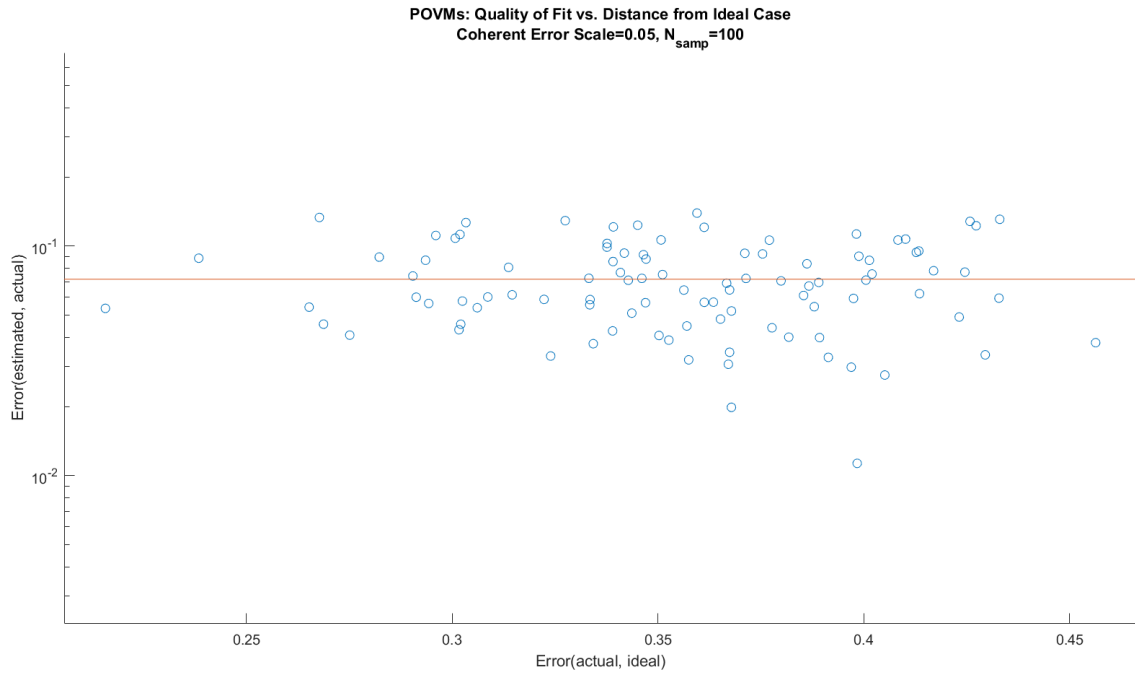


Figure 4.6: With the POVMs converted to quantum channels, the y -axis is the diamond norm between the estimated and actual POVM, and the x -axis is the diamond norm between the actual POVM and ideal PVM. Symbolically, this is $\|\mathcal{E}\{\sqrt{E'_i}\} - \mathcal{E}\{\sqrt{E_i}\}\|_\diamond$ vs. $\|\mathcal{E}\{\sqrt{E_i}\} - \mathcal{E}\{|i\rangle\langle i|\}\|_\diamond$.

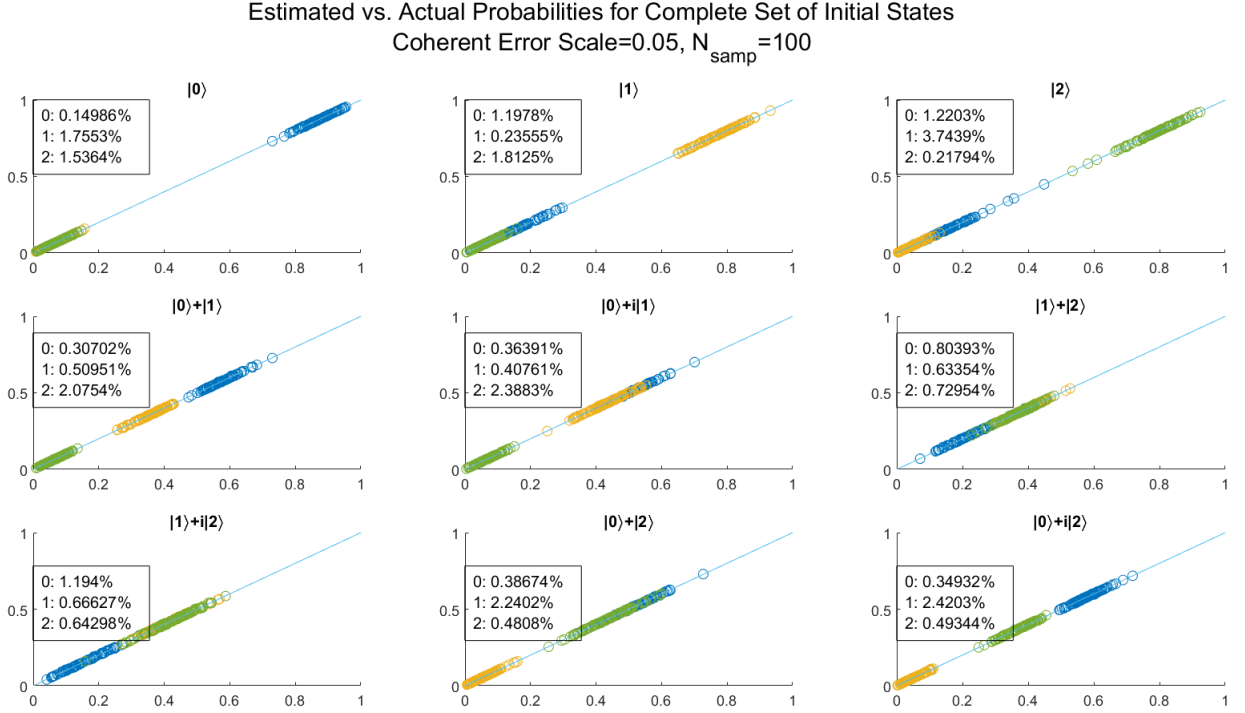


Figure 4.7: Probabilities for each measurement outcome (0/1/2) for each initial qutrit state, MLE output (y) vs. ground truth (x). Thus, we are plotting $\text{Tr}(\hat{M}_i' \hat{\rho}_j')$ vs. $\text{Tr}(\hat{M}_i \hat{\rho}_j)$, where i indexes the color (0:blue, 1:yellow, 2:green), and j the subplot. These results would be sufficient to perform measurement tomography.

Lastly, in Fig. 4.7, we note that the fit quality of all the above parameters is best ascertained by the measurement statistics obtained by each sampled model. That is, even if the estimated parameters deviate from their true values, the estimated parameter set should agree quite well with the actual set in terms of measurement probabilities, because this is the only quantity that MLE could access. Each color in each plot corresponds to a different measurement outcome—the top row of plots indicates that blue is for $|0\rangle$, yellow for $|1\rangle$, and green for $|2\rangle$. The ideal initial state before measurement is the title of each subplot. To keep the plots readable, we only show the results for measurement without pre-rotation (otherwise there are 63 possible plots). If MLE were perfect, plotting the estimated probability against its ground truth should follow the line $y = x$, and we see that to within less than 1% percent, this is indeed the case. Deviating from the above plots, here we *are* computing the relative percent error. Thus, whenever the true probability is close to 0, then the error is amplified not because the fit itself is worse, but because the scale of the data is simply small. This explains certain legend entries that exceed 1% error.

We refer the reader to Appendix ??, where plots from the $\sigma = 0.01$ (0.6 degrees) and $\sigma = 0.25$ (14 degrees) datasets are tabulated. The performance of MLE in these regimes is comparable to the data shown in this section (2.9 degrees). With the SPAM set $[a', b', c', e1', e2', e3', \text{nip}']$ known, in the next section we fix these parameters and use MLE to estimate the process matrix over time.

4.3. Estimating the Quantum Operation

In this section, we apply the same MLE method to extract the process matrix of some quantum evolution. We first generate a set of ground-truth SPAM parameters: $a = 0.9894$, $b = 0.0105$, $c = 1.3352e-4$, $e1 = -0.0853$, $e2 = 0.1398$, $e3 = -0.0605$, and near-ideal POVM

$$\begin{array}{lll} 0.9386 + 0.0000i & 0.0802 + 0.0574i & 0.0300 + 0.0471i \\ 0.0802 - 0.0574i & 0.1055 + 0.0000i & 0.0923 - 0.0262i \end{array}$$

```

0.0300 - 0.0471i    0.0923 + 0.0262i    0.0910 + 0.0000i

0.0034 + 0.0000i   -0.0425 - 0.0327i    0.0001 + 0.0026i
-0.0425 + 0.0327i    0.8443 + 0.0000i   -0.0260 - 0.0307i
0.0001 - 0.0026i   -0.0260 + 0.0307i    0.0019 + 0.0000i

0.0580 + 0.0000i   -0.0377 - 0.0247i   -0.0302 - 0.0496i
-0.0377 + 0.0247i    0.0502 + 0.0000i   -0.0663 + 0.0568i
-0.0302 + 0.0496i   -0.0663 - 0.0568i    0.9071 + 0.0000i

```

and obtain a set of estimates via the method of the preceding section $a' = 0.9910$, $b' = 0.0090$, $c' = 0$, $e1' = -0.0856$, $e2' = 0.1433$, $e3' = -0.0604$, and estimated POVM

```

0.9366 + 0.0000i    0.0808 + 0.0565i    0.0309 + 0.0468i
0.0808 - 0.0565i    0.1060 + 0.0000i    0.0932 - 0.0266i
0.0309 - 0.0468i    0.0932 + 0.0266i    0.0927 + 0.0000i

0.0042 + 0.0000i   -0.0437 - 0.0321i   -0.0001 + 0.0024i
-0.0437 + 0.0321i    0.8449 + 0.0000i   -0.0267 - 0.0293i
-0.0001 - 0.0024i   -0.0267 + 0.0293i    0.0024 + 0.0000i

0.0592 + 0.0000i   -0.0371 - 0.0244i   -0.0309 - 0.0492i
-0.0371 + 0.0244i    0.0491 + 0.0000i   -0.0666 + 0.0559i
-0.0309 + 0.0492i   -0.0666 - 0.0559i    0.9049 + 0.0000i

```

These parameter sets are fixed for the remainder of this report. The true parameters are used in the data generation, and the estimated parameters are used in MLE.

4.3.1. Maximum Likelihood Estimation

To estimate the a quantum process, the log-likelihood from earlier can be reused, but the density matrices are now time-dependent.

$$\log \mathcal{L}(t) = \sum_{s,b} m_0 \log(\text{Tr} \tilde{E}_{0,b} \tilde{\rho}(t)_s) + m_1 \log(\text{Tr} \tilde{E}_{1,b} \tilde{\rho}(t)_s) + m_2 \log(\text{Tr} \tilde{E}_{2,b} \tilde{\rho}(t)_s) \quad (4.47)$$

where

$$\rho(t) = \sum_i K_i(t) \rho(0) K_i^\dagger(t) \quad (4.48)$$

and $\{K_i(t)\}$ is the quantum channel that evolves the system from $t = 0$ out to some time t . For qutrits, the set of Kraus maps consists of 9 complex-valued 3×3 matrices, which only must obey the trace constraint

$$\sum_i K_i^\dagger(t) K_i(t) = I \quad (4.49)$$

which is passed as a nonlinear constraint to MLE. The only unknowns are the matrix elements of each $K_i(t)$, all other parameters in the likelihood are fixed.

4.3.2. Amplitude and Phase Damping

For our first case study, we can apply the MLE fitting to simulated data from a qutrit undergoing both amplitude and phase damping channels. To recap, for a qubit, the amplitude damping (AD) channel is

$$K_1^{AD} = \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{1-p_{AD}} \end{bmatrix} \quad (4.50)$$

$$K_2^{AD} = \begin{bmatrix} 0 & \sqrt{p_{AD}} \\ 0 & 0 \end{bmatrix} \quad (4.51)$$

and the phase damping channel (PD) is

$$K_1^{PD} = \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{1-p_{PD}} \end{bmatrix} \quad (4.52)$$

$$K_2^{PD} = \begin{bmatrix} 0 & 0 \\ 0 & \sqrt{p_{PD}} \end{bmatrix} \quad (4.53)$$

where p indicates the probability that the respective term occurs in the channel. When the two channels occur simultaneously, we can compose the maps

$$\mathcal{E}_{AD}(\mathcal{E}_{PD}(\rho)) = \sum_{i,j} K_i^{AD} K_j^{PD} \rho (K_j^{PD})^\dagger (K_i^{AD})^\dagger \quad (4.54)$$

It turns out that these channels commute, so we are free to take products of the individual channel operators to form the operators for the combined channel (AP):

$$K_1^{AP} = K_1^{PD} K_1^{AD} = \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{1-p_{AD}}\sqrt{1-p_{PD}} \end{bmatrix} \quad (4.55)$$

$$K_2^{AP} = K_1^{PD} K_2^{AD} = \begin{bmatrix} 0 & \sqrt{p_{AD}} \\ 0 & 0 \end{bmatrix} \quad (4.56)$$

$$K_3^{AP} = K_2^{PD} K_1^{AD} = \begin{bmatrix} 0 & 0 \\ 0 & \sqrt{1-p_{AD}}\sqrt{p_{PD}} \end{bmatrix} \quad (4.57)$$

$$K_4^{AP} = K_2^{PD} K_2^{AD} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \quad (4.58)$$

The fourth element is 0 because first the second AP operator leaves only 0 or $|0\rangle$, and then the second PD operator sends $|0\rangle$ to 0. Since they commute, the products could be taken in reverse, and a different Kraus set would be obtained (where the fourth element is not 0). This is fine, because we know the Kraus form is not unique. The probabilities are usually taken to be exponential decays

$$p_{AD} = 1 - e^{-t/T_1} \quad p_{PD} = 1 - e^{-t/T_\phi} \quad (4.59)$$

because expanding the action of the channel for small time will lead to a first-order differential equation, i.e. the Lindblad master equation [18]. Plugging in $t = 0$ yields the identity map, and $t = \infty$ yields the ground state, so the limits check out as well. Many different quantum channels may give rise to the same map—recall that for phase damping, the channel may be also rewritten as a random phase flip:

$$K_1^{PD} = \sqrt{\alpha} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad (4.60)$$

$$K_2^{PD} = \sqrt{1-\alpha} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \quad (4.61)$$

where $\alpha = (1 + \sqrt{1-p})/2$. The unitary freedom of the Kraus form shows that a continuous dephasing process actually has a discrete interpretation in terms of random I or Z gates. Complete dephasing $t = \infty \rightarrow p = 1$ corresponds to either doing nothing or applying a phase flip with equal probability.

The amplitude and phase damping channels for qutrits are similar [59]. For AD, we add a decay operator for each transition between $i \rightarrow j$ that we want to allow in our system, with some probability. For PD, we add a dephasing operator for each state in the system, with some probability. In Peterer et al. (2015) [46], it is experimentally and numerically confirmed for a transmon embedded in a 3D cavity, that the dominant decay channels are between sequential levels—i.e. we should include decay from $1 \rightarrow 0$ and $2 \rightarrow 1$, but that $2 \rightarrow 0$ can be neglected. This also connects well with how we have decomposed the qutrit space into 01 and 12 Bloch spheres. Therefore, our qutrit AD Kraus set is

$$K_1^{AD} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \sqrt{1-p_{11}} & 0 \\ 0 & 0 & \sqrt{1-p_{12}} \end{bmatrix} \quad K_2^{AD} = \begin{bmatrix} 0 & \sqrt{p_{11}} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad K_3^{AD} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & \sqrt{p_{12}} \\ 0 & 0 & 0 \end{bmatrix} \quad (4.62)$$

and the PD Kraus set is

$$K_1^{PD} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \sqrt{1-p_{21}} & 0 \\ 0 & 0 & \sqrt{1-p_{22}} \end{bmatrix} \quad K_2^{PD} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \sqrt{p_{21}} & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad K_3^{PD} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \sqrt{p_{22}} \end{bmatrix} \quad (4.63)$$

where there are different probabilities to control the different state-dependent operators, and different T_1 and T_ϕ parameters for both the 01 and 12 subspaces. Explicitly, we have

$$p_{11} = 1 - e^{-t/T_1^{01}} \quad (4.64)$$

$$p_{12} = 1 - e^{-t/T_1^{12}} \quad (4.65)$$

$$p_{21} = 1 - e^{-t/T_\phi^{01}} \quad (4.66)$$

$$p_{22} = 1 - e^{-t/T_\phi^{12}} \quad (4.67)$$

The first subscript refers to the channel (1 for AD, 2 for PD), and the second to the subspace (1 for 01, 2 for 12). The two channels commute as in the qubit case, so we can compose them in whichever way produces more null Kraus maps, which is to apply the amplitude damping first. The joint Kraus set is

$$K_1^{AP} = K_1^{PD} K_1^{AD} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \sqrt{(1-p_{11})(1-p_{21})} & 0 \\ 0 & 0 & \sqrt{(1-p_{12})(1-p_{22})} \end{bmatrix} \quad (4.68)$$

$$K_2^{AP} = K_2^{PD} K_1^{AD} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \sqrt{p_{21}(1-p_{11})} & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad (4.69)$$

$$K_3^{AP} = K_3^{PD} K_1^{AD} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \sqrt{p_{22}(1-p_{12})} \end{bmatrix} \quad (4.70)$$

$$K_4^{AP} = K_1^{PD} K_2^{AD} = \begin{bmatrix} 0 & \sqrt{p_{11}} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad (4.71)$$

$$K_5^{AP} = K_2^{PD} K_2^{AD} = 0 \quad (4.72)$$

$$K_6^{AP} = K_3^{PD} K_2^{AD} = 0 \quad (4.73)$$

$$K_7^{AP} = K_1^{PD} K_3^{AD} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & \sqrt{p_{12}(1-p_{21})} \\ 0 & 0 & 0 \end{bmatrix} \quad (4.74)$$

$$K_8^{AP} = K_2^{PD} K_3^{AD} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & \sqrt{p_{12}p_{21}} \\ 0 & 0 & 0 \end{bmatrix} \quad (4.75)$$

$$K_9^{AP} = K_3^{PD} K_3^{AD} = 0 \quad (4.76)$$

We choose realistic decay parameters for our simulations $T_1^{01} = 84$, $T_1^{12} = 41$, $T_\phi^{01} = 126$, $T_\phi^{12} = 52$, using

$$1/T_2 = 1/2T_1 + 1/T_\phi \quad (4.77)$$

to get the pure dephasing times from the T_2 times listed in Petersen et al. [46] (all units in microseconds). Both the Kraus operators (Fig. 4.8) and the process matrix (Fig. 4.9) for the qutrit amplitude-and-phase damping channel with these decay parameters are shown below.

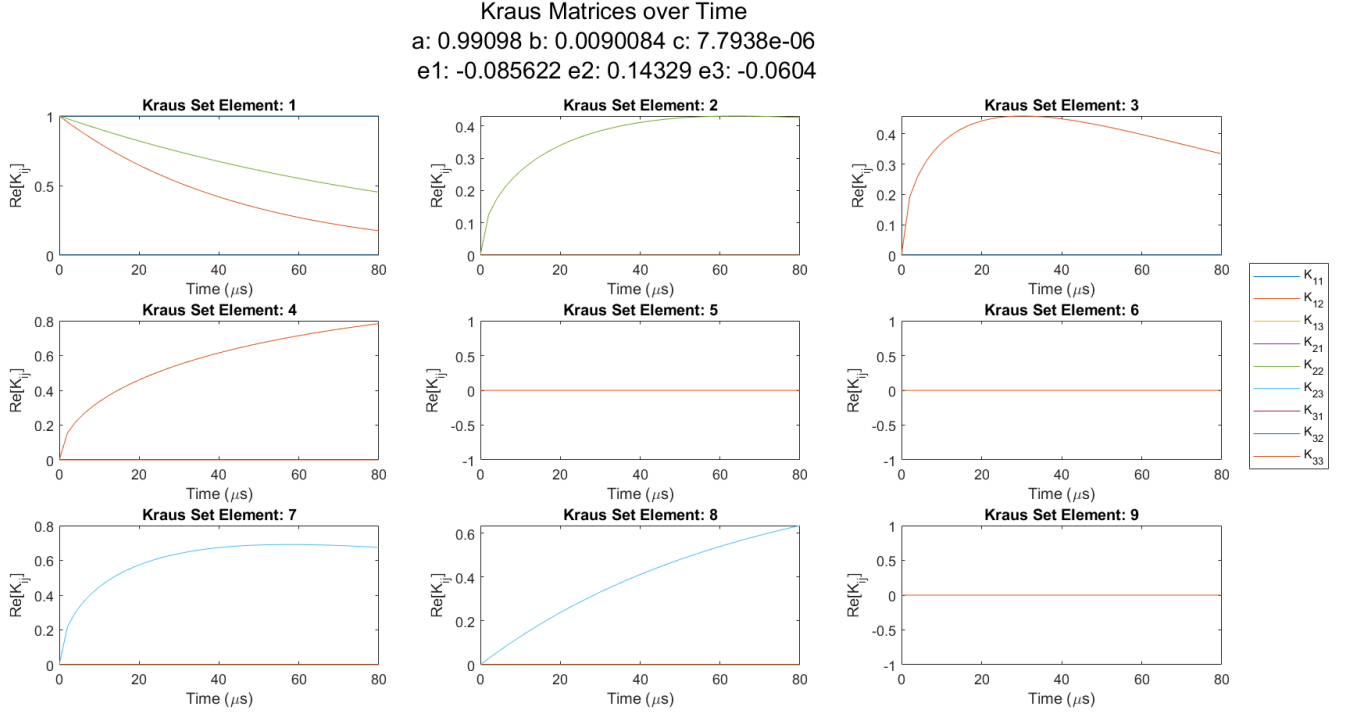


Figure 4.8: Ground-truth Kraus matrices for qutrit AP channel over time.

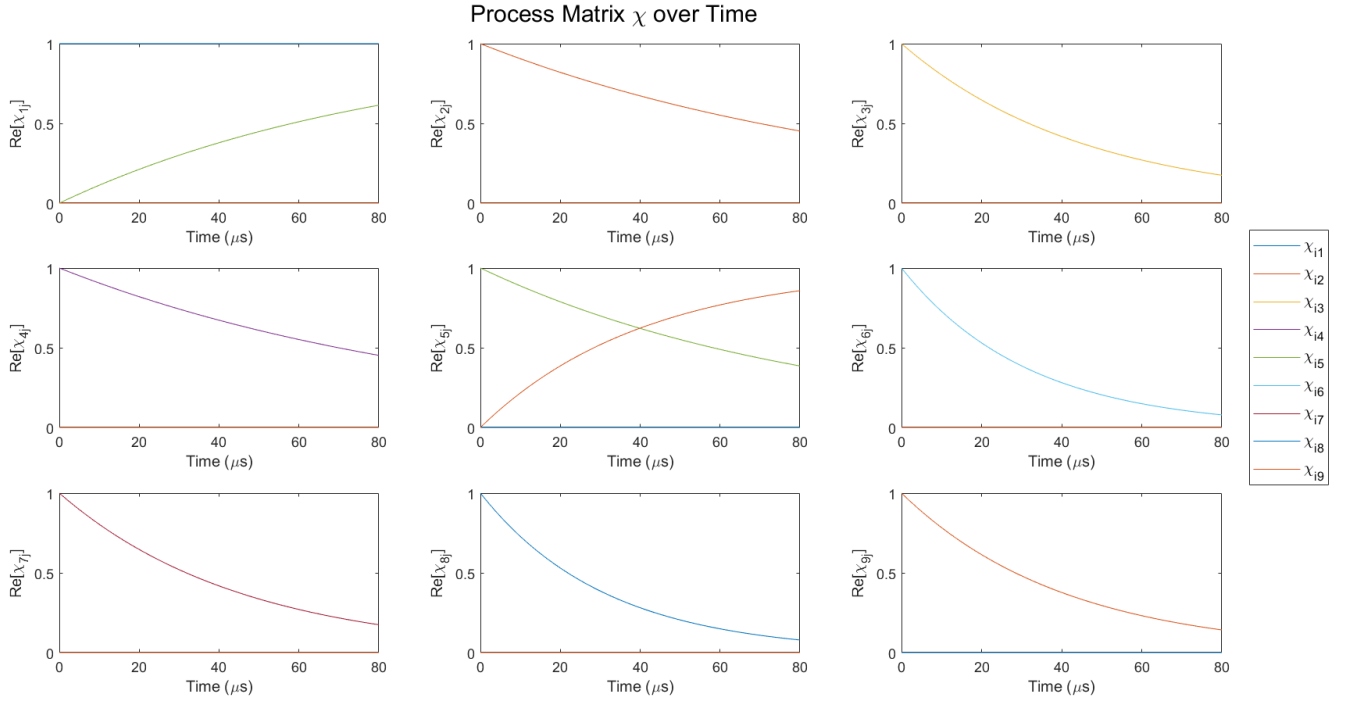


Figure 4.9: Ground-truth process matrix elements for qutrit AP channel over time.

The goal is to reconstruct figures 4.8 and 4.9 from faulty data. The true (unprimed) parameters are used to generate 10^4 shots per initial state per measurement rotation, and the AP channel takes

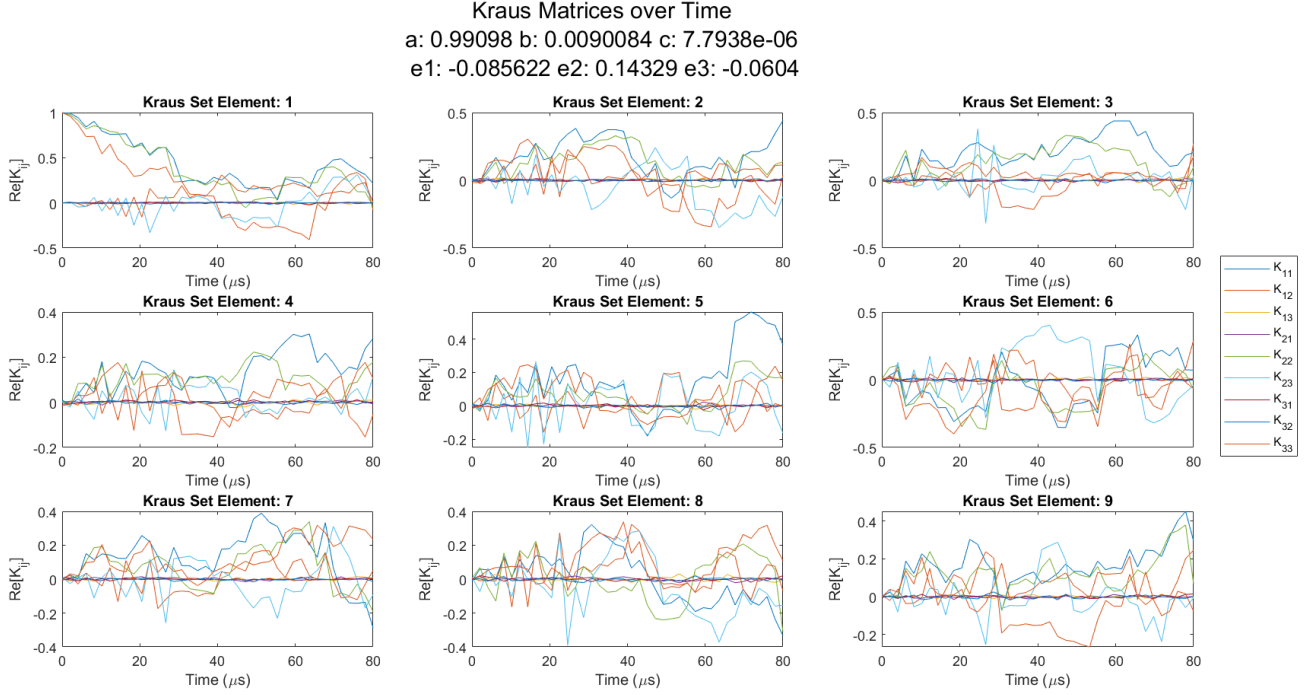


Figure 4.10: MLE Kraus matrices for qutrit AP channel over time.

the system from $t = 0$ to $t = t'$ in between. Using MLE, we estimate the process at 40 times equally spaced between 0 and 80 microseconds. Fig. 4.10 shows the reconstructed Kraus matrices, and Fig. 4.11 shows the reconstructed process matrix over time. Only the real component of the processes and Kraus matrices are shown.

Each timestep represents a separate nonlinear optimization problem to solve. Ideally, we should use an initial point that is as close as possible to what we believe the true optimum is. On the other hand, we want the method to work for any process, so we can make two assumptions:

1. If the time evolution is smooth, then for small enough time intervals, the process map should not change very much from t to $t + dt$. We can use the output of the previous MLE to seed the initial point of the current MLE.
2. The very first MLE should be initialized to the identity map, since for small enough time, all maps appear close to the identity.

It is worth stressing that the sets of measurement counts must be obtained with a fine-enough time resolution that the dynamics really do appear smooth. If the system has some fast components that are undersampled, the output of one process fit may not be very close to the process of the next timestep, resulting in a potentially poor initialization. The caveat here is that since each fit is initialized from the prior fit, if even one fit is erratic, it could potentially throw off all subsequent fits.

From Fig. 4.10, the unitary freedom of the Kraus form is now quite apparent. The Kraus set found for each time by MLE does not “know” about any of the other simulations (other than the initialization just described), and in principle each output could be in another basis. As a result, the Kraus maps over time look very unstable, but when the process map is formed from the Kraus set, the evolution remains smooth, proving that the Kraus maps indeed correspond to the same evolution despite their irregularity.

There is one additional point to clarify, regarding the form of the initial identity process. There are many ways of creating such a process. One way might take the Kraus set to be a set of 8 matrices, all of which are 0, and the last matrix is the 3×3 identity matrix. Alternatively, each matrix could be $\frac{1}{3}$ of the identity matrix. In general, each matrix could be some multiple of the identity, provided that the trace constraint is satisfied. It is conceivable that one form might seed the resulting optimization better than another. In the case of the AP channel, if we set $t = 0$ we can determine exactly what “kind” of identity channel we should start with, and use that for the first initialization. From Eq. 4.68, we see that the

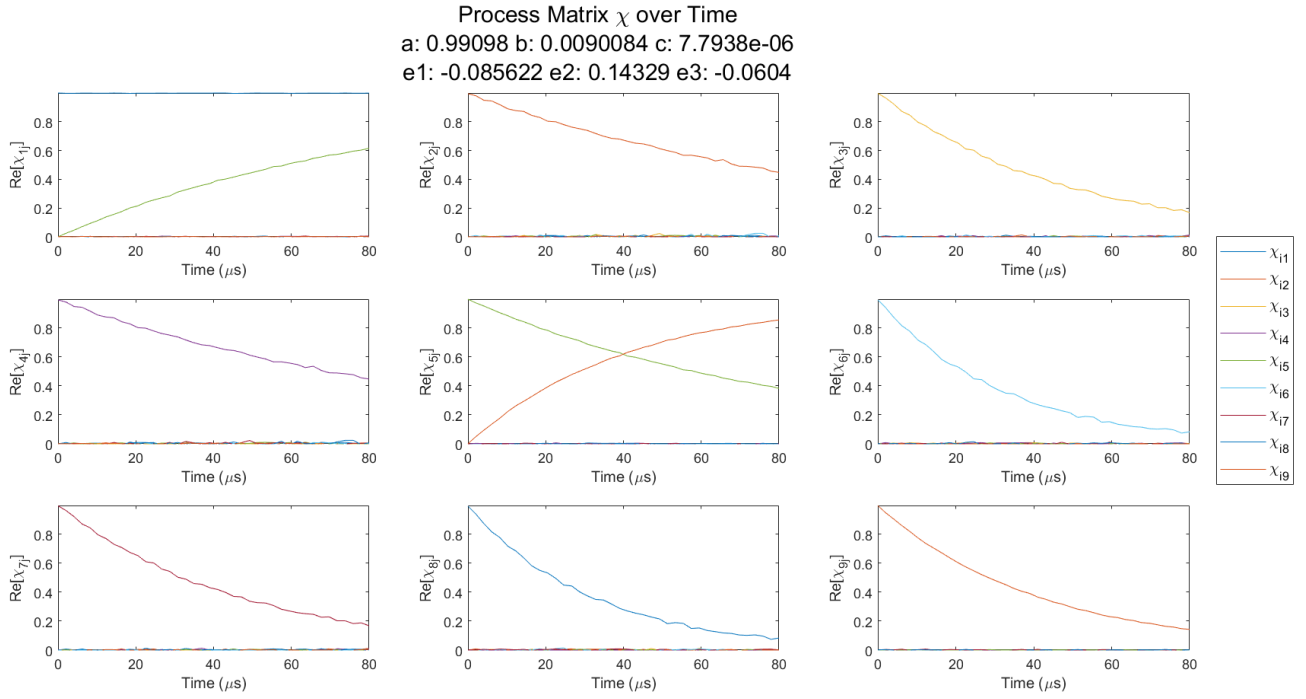


Figure 4.11: MLE process matrix elements for qutrit AP channel over time.

first option we considered (just one nonzero Kraus matrix) is the correct choice. In the absence of prior knowledge regarding the Kraus set—such as the next case study—one may need check empirically which form produces the most trustworthy result.

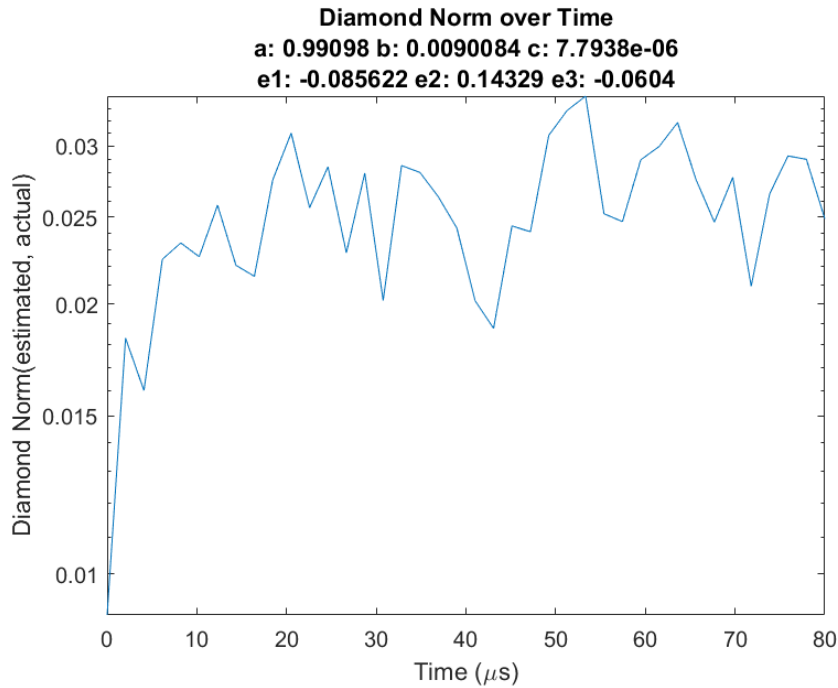


Figure 4.12: Diamond norm between estimated and actual AP processes over time.

Finally, in Fig. 4.12, we confirm the accuracy of the process estimation. The diamond distance

between the MLE estimate and ground-truth processes for the AP channel (ladder system) is less than 4% over the simulated time. This inaccuracy is partially due to estimation error in the MLE routine (i.e. the optimality of the minimum), and partially due to the difference between the estimated SPAM parameters and their true values, tabulated in section 4.3.

4.3.3. Transmon Device Evolution

The purpose of the previous example was to validate the accuracy of the MLE routine when the ground-truth channel was known, so we could have something to compare the fits to. In reality, any experimental system that we model as closed will still interact with its environment to some extent, so all datasets obtained will obey open system dynamics. Since the line between system and environment is always arbitrarily chosen ⁷, we do not have the luxury of tracing over known, unitary dynamics. Instead, we can directly estimate a series of Kraus maps from a set of measurements taken over time, and trust that the results reflect the true, effective process that the system undergoes.

In this section, we focus on the transmon device from MIT previously explained in section 3.5.3, which was simulated in `scqubits`. The system will be one transmon, and its environment will be another transmon that it is capacitively coupled to, both modeled as qutrits. This is physically motivated by current processor designs, which consist of transmon lattices. The same complete basis of states used above was created in Python, and prepared for the left (tunable) qutrit. The right qutrit was prepared in a fixed state, from the set $\{|0\rangle, 1/\sqrt{2}(|0\rangle + |1\rangle), |1\rangle, 1/\sqrt{2}(|1\rangle + |2\rangle), |2\rangle\}$. The joint evolution is simulated using the `mesolve` function in QuTiP, which evolves the two-qutrit state under the unitary evolution of the device Hamiltonian, subject to a set of noise channels passed in a Lindblad collapse operators. In the following simulations, there are two collapse operators, a 3×3 annihilation operator for each qutrit, which model excitation loss. Both operators are scaled by the same rate 0.001, which was arbitrarily chosen based on the resulting timescale of the dynamics. The system is evolved for a set of times, and the right qutrit is traced out of the final state at each time. The resulting density matrix of the left qutrit is transferred to MATLAB, where it is measured in the same set of bases as usual. The counts are then used by MLE to reconstruct the reduced evolution of the left qubit, which is generally non-unitary due to the partial trace over the coupling between the system and environment qutrits, resulting in non-Markovianity. Though superconducting circuits are usually analyzed in the interaction picture, simulating the interaction Hamiltonian in QuTiP resulted in some technical difficulties, so the lab-frame Hamiltonian was used in the end. See section C.1 for the analytical form of the interaction picture Hamiltonian. The Hamiltonian is simulated for two distinct flux regimes. The first is when the flux is roughly $\Phi_{ext} = 0.15$ (Figs. 4.13-4.17), tuning the system to the avoided level crossing that gives rise to the ZZ-interaction. The second is when the flux is 0 on the left transmon, suppressing the coupling (Fig. 4.18). In the coming plots, again, only the real component of the processes are shown.

Figures 4.13-4.17 contain a lot of information, but a few main points may be gleaned. First, recall from section 3.4.2

$$\chi_{jk} = \frac{1}{d^2} \sum_i \text{Tr}(B_j^\dagger K_i) \text{Tr}(B_k^\dagger K_i)^* \quad (4.78)$$

$$= \frac{1}{d^2} \sum_i \langle\langle B_j | K_i \rangle\rangle \langle\langle K_i | B_k \rangle\rangle \quad (4.79)$$

which just expresses the elements of the process matrix in terms of the elements of the vectorized Kraus operator K_i in the natural basis (the process matrix is the transfer matrix in this basis, up to a basis reordering) Vectorizing a projector associates two indices with one

$$\text{vec}(|m\rangle\langle n|) = |n\rangle \otimes |m\rangle \equiv |B_j\rangle \quad (4.80)$$

where there is a one-to-one mapping from the projective indices m, n to the vectorized index j : $j \in [1, 9]$ corresponds to basis element $|\lfloor (j-1)/3 \rfloor \rangle \langle j-1 \bmod 3|$. In the plots, the row index refers to the output element, and the column index to the input element. We see that the dynamics is divided between fast and slow oscillating terms. The slow terms occur in elements $(1, 5, 9) \times (1, 5, 9)$, which correspond to how the projectors $|0\rangle\langle 0|, |1\rangle\langle 1|, |2\rangle\langle 2|$ are mapped into themselves. These are the qutrit occupations that undergo decay due to the amplitude damping, in addition to some resonant exchange with the

⁷Technically, there never is an environment, and the whole universe is always the system.

environment, which causes oscillations that are slow with respect to the unitary dynamics. All of the fast terms originate from coherences in superpositions that rotate quickly in the lab frame. These coherences are also affected by the coupling, which manifests as a slow-varying envelope over the high-frequency fringes. Generally, the unitary dynamics also are aliased, depending on the time-resolution of the data. Though this will produce lower apparent frequencies, as long as they still remain much higher than the coupling dynamics, there will be no ambiguities.

As an example, consider Fig. 4.13, specifically the three orange curves of the “diagonal” plots, i.e. elements χ_{19} , χ_{59} , and χ_{99} . Respectively, these terms describe how population in $|2\rangle$ is sent to $|0\rangle$, $|1\rangle$, and $|2\rangle$. The results match our expectations, because we see that in the χ_{19} plot, the decay constantly increases the population in $|0\rangle$ via decay from $|2\rangle$, in addition to a very low oscillation from the resonant exchange between $|20\rangle$ and $|02\rangle$. This exchange is weak because it is a two-photon exchange which must be mediated by a one-photon coupling. In χ_{59} and χ_{99} , the $|2\rangle$ is directly swapped via $|20\rangle \leftrightarrow |11\rangle$, giving rise to strong oscillations induced by the flux-bias. The oscillations in χ_{59} do not completely reach zero for short times (hundreds of nanoseconds) because decay from $|2\rangle$ is temporarily populating $|1\rangle$ on the way to $|0\rangle$. In χ_{99} , population can only return via the coupling, hence the overall decay in time. Similar analysis holds for the other elements, plots, and environment states. Though the coherences are harder to intuit due to their precession, notice that the plots mirrored across the plot-diagonal show similar time evolution. This is because terms like $|i\rangle\langle j|$ and $|j\rangle\langle i|$ undergo identical dynamics, since the process takes states to states: $\mathcal{E}(\rho^\dagger) = \mathcal{E}(\rho)^\dagger \rightarrow \mathcal{E}(|i\rangle\langle j|) = \mathcal{E}(|j\rangle\langle i|)$. Throughout all of this, remember that there is also residual inaccuracy due to the estimated SPAM parameters differing from their true values by some small amount, though this should be minimal.

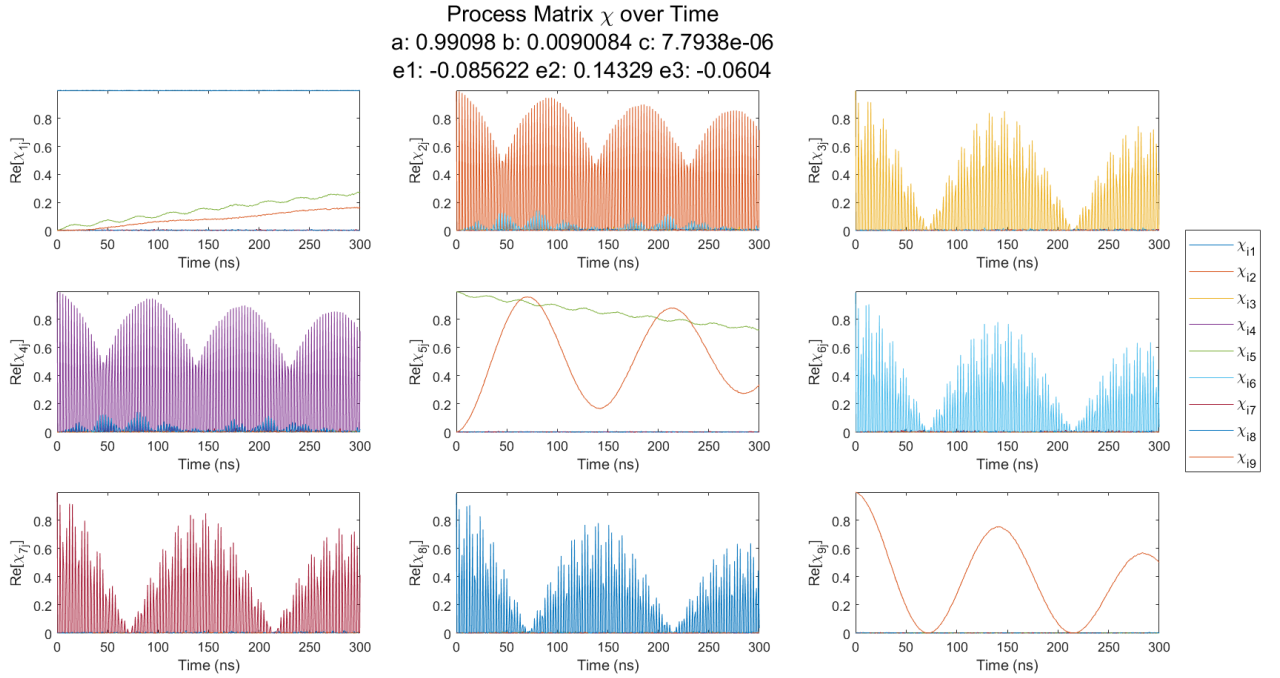


Figure 4.13: Process map over time for left qutrit, right qutrit in $|0\rangle$ state.

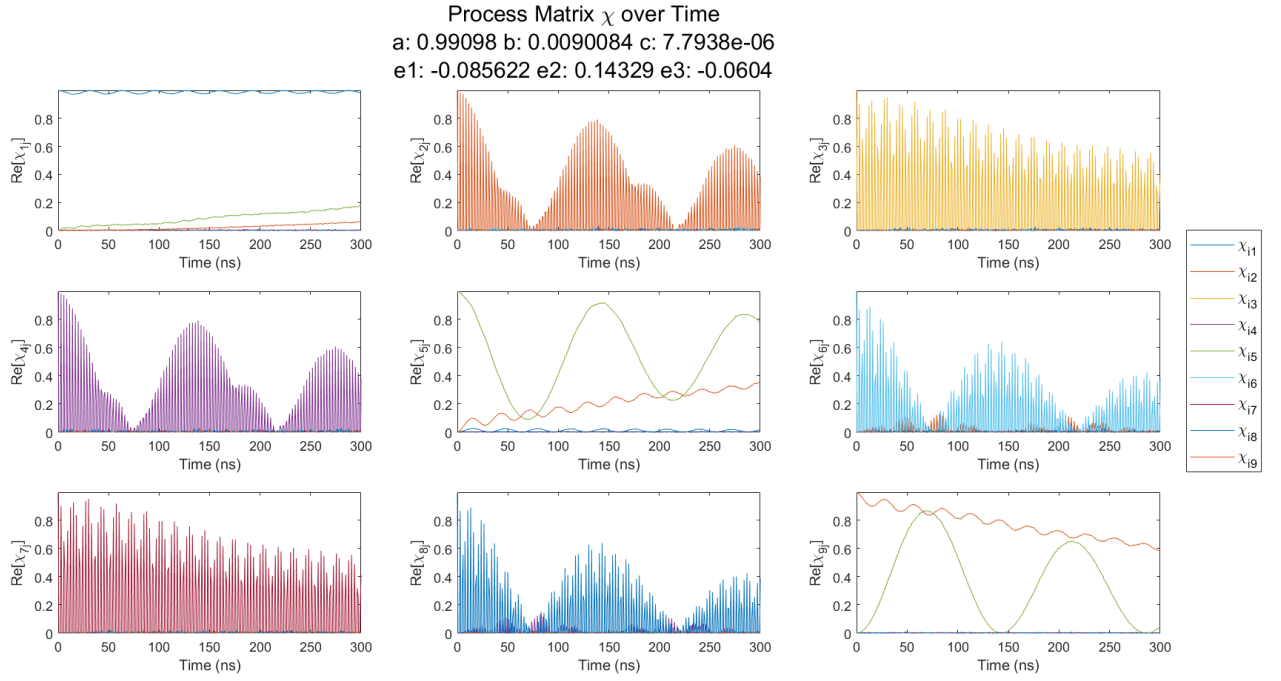


Figure 4.14: Process map over time for left qutrit, right qutrit in $|1\rangle$ state.

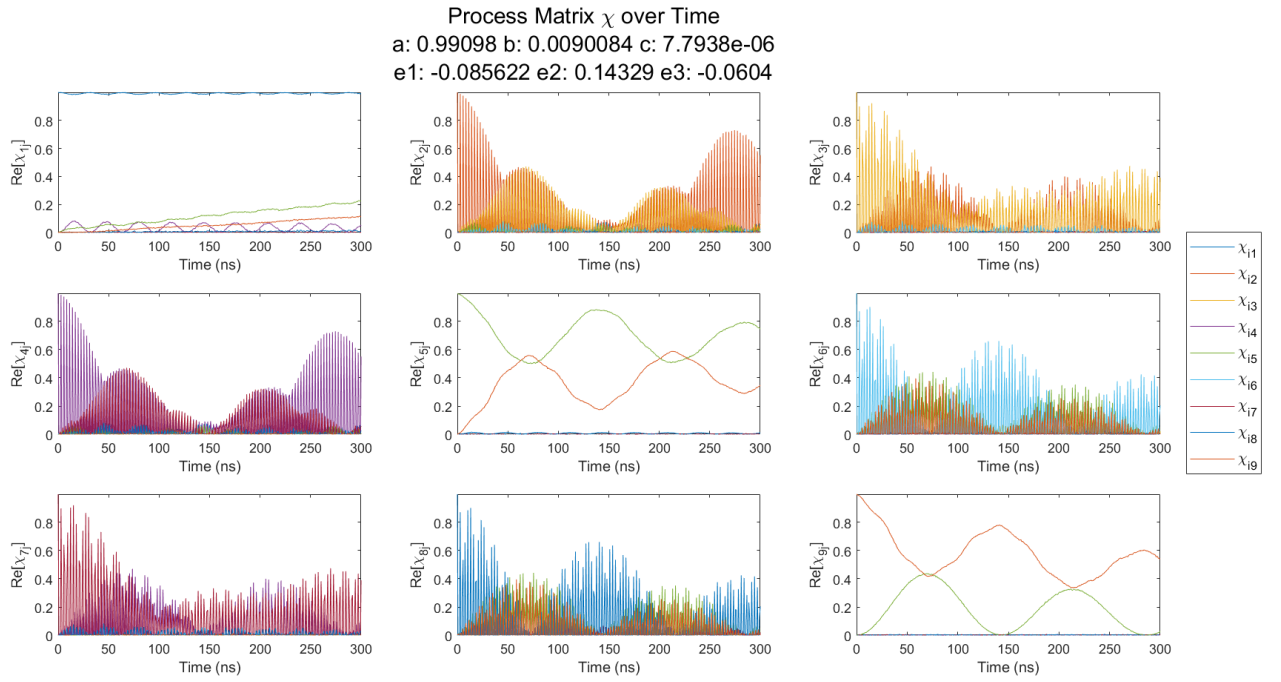


Figure 4.15: Process map over time for left qutrit, right qutrit in $|+\rangle_{01}$ state.

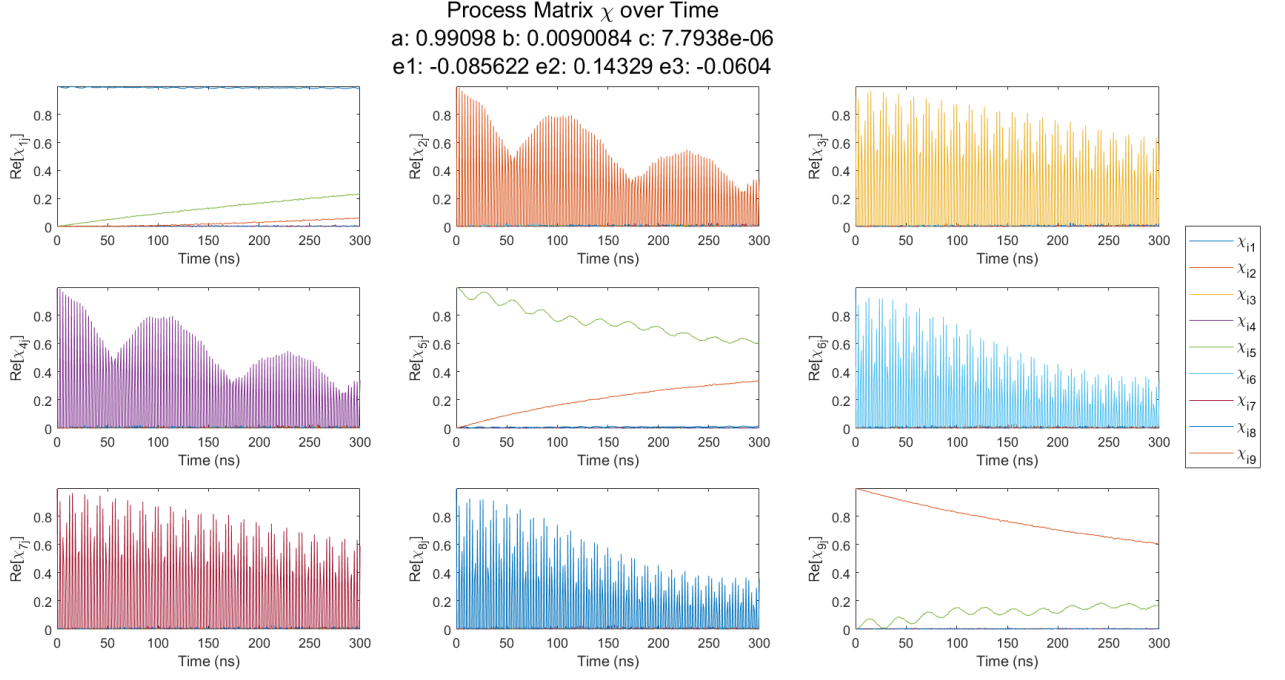


Figure 4.16: Process map over time for left qutrit, right qutrit in $|2\rangle$ state.

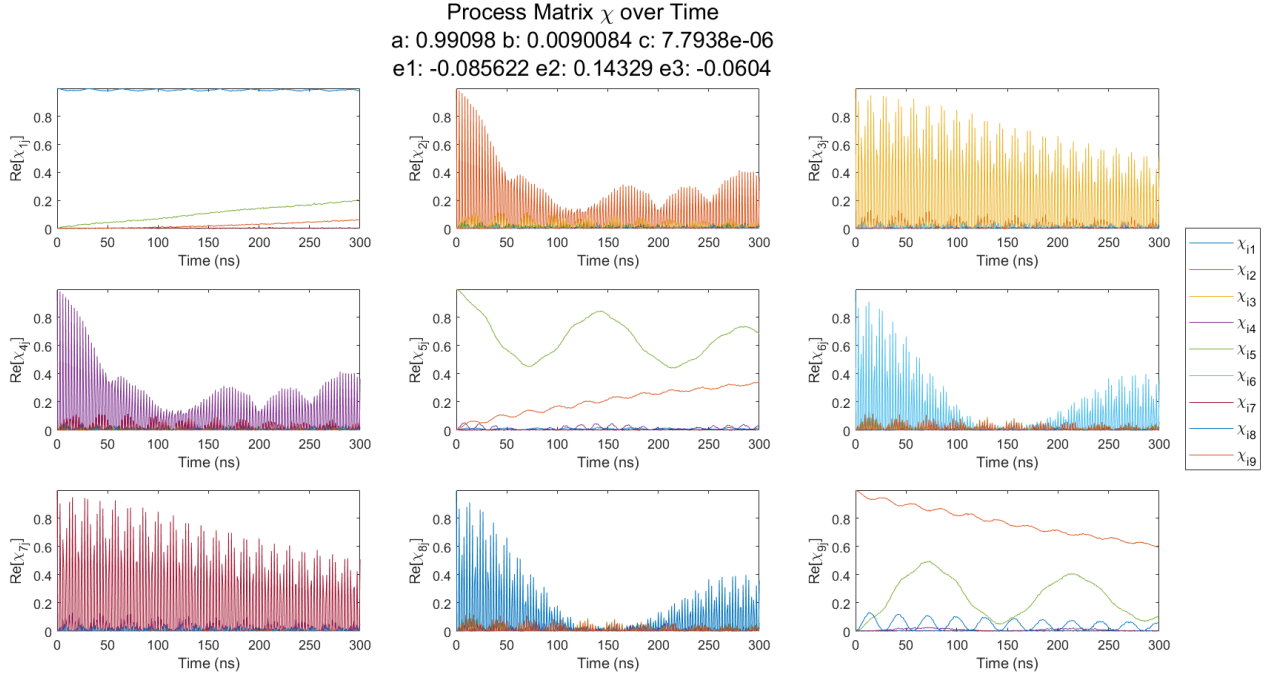


Figure 4.17: Process map over time for left qutrit, right qutrit in $|+\rangle_{12}$ state.

Lastly, we consider an example where the flux bias is 0, so that the ZZ-interaction is turned off. In this regime, the state of the environment has little observable effect, so all the dynamics are very similar. Thus, we only show the results for the case of an environment in the $|0\rangle$ state, in Fig. 4.18.

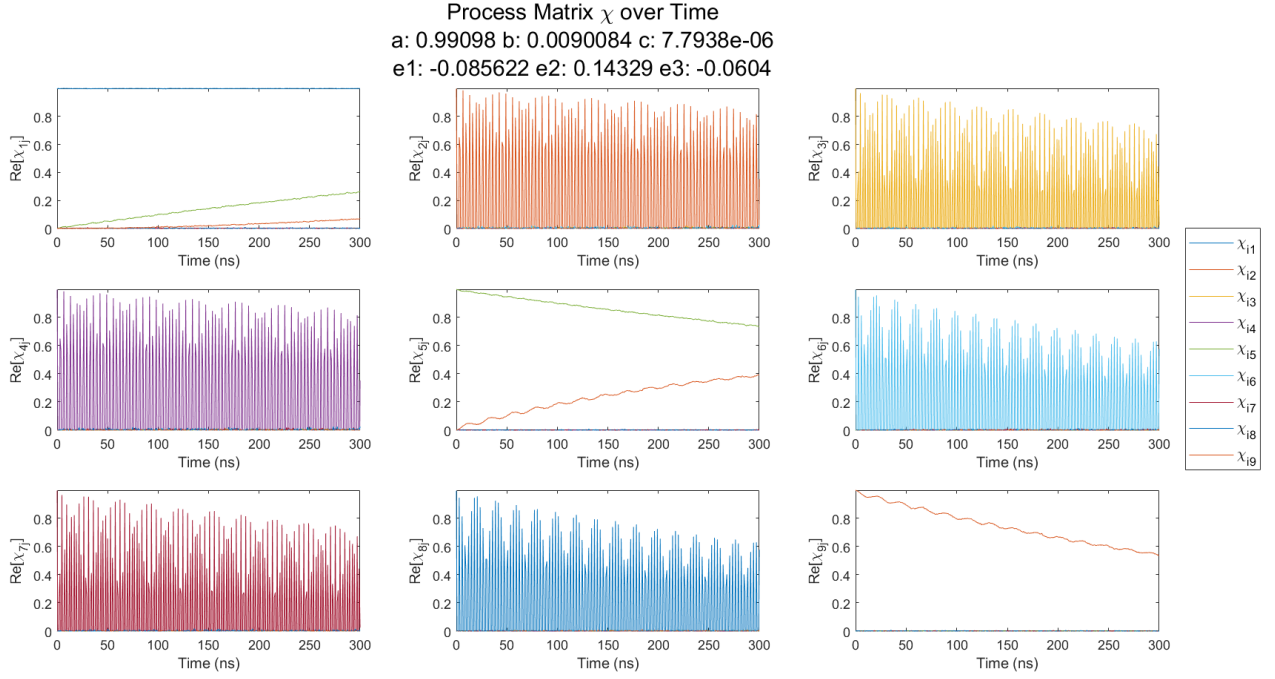


Figure 4.18: Process map over time for left qutrit, right qutrit in $|0\rangle$ state (flux off).

The residual non-Markovianity due to the always-on ZZ-coupling described in section 3.5.6 is small enough that it cannot be detected from these simulations, which were performed in the lab frame. Additional simulations in a rotating frame were attempted, but difficult to interpret due to the comparative smallness of 350kHz to the dynamics in the interaction picture ($g \sim 10\text{MHz}$). Having established that the dynamics of a qutrit may be successfully reconstructed via MLE, we can now move on to analyzing the dynamics themselves.

4.4. Measuring Markovianity

Since there are infinitely many ways of purifying a quantum operation (i.e. via a Stinespring dilation), knowing the channel does not give any explicit information about the environment. However, certain properties of the system-environment interaction may be inferred from the estimated quantum operations. One such quantity is the Markovianity of the evolution, which can be computed via the method of Bruer et al. [13].

4.4.1. Amplitude and Phase Damping

For our first example, we use the same AP channel found in section 4.3.2. Two random pure states are selected⁸ as initial states, and the extracted process is applied for each moment in time. The trace distance between these two states over time is then evaluated, and this is looped over 10^4 randomly chosen pairs of initial states. The non-Markovianity is assessed by finding the pair whose trace distance increases the most throughout the evolution. The discrete derivative of the array of trace distances is computed, and summed over all positive entries, yielding the non-Markovianity. This is the number in the titles of the plots in this section. If no such pair shows an increasing trace distance, then the evolution is Markovian, and the value is 0.

⁸The sampling is Haar-random, and is performed by a function from the QETLAB library.

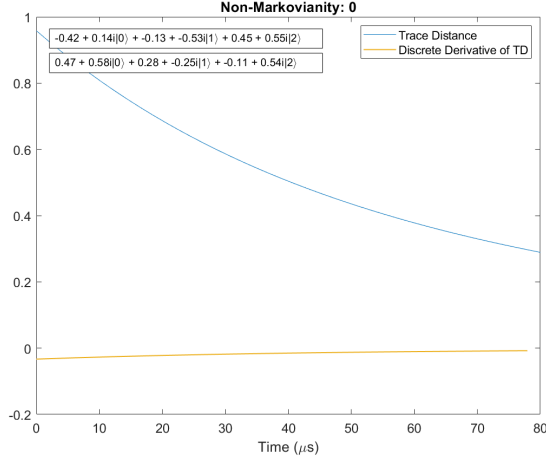


Figure 4.19: Trace distance over time for pair that displays the “most” non-Markovianity, using the Λ -system ground-truth AP channel. Since the channel is Markovian, the pair is effectively random.

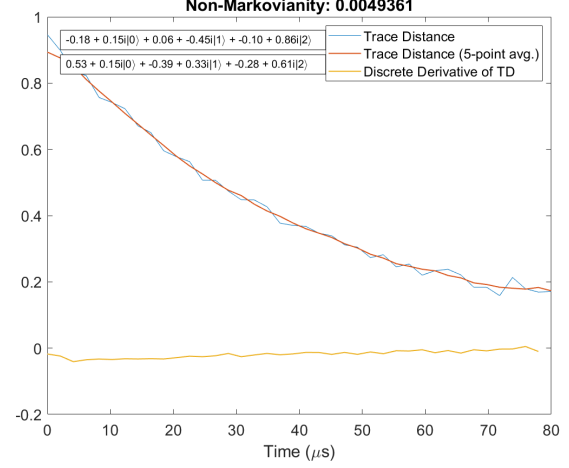


Figure 4.20: Trace distance over time for pair that displays the “most” non-Markovianity, using the Λ -system AP channel found by MLE. Since the channel is Markovian, the pair is effectively random.

The two qutrit pure states to the left of the legend are the pair (ρ_1 and ρ_2) whose trace distance increased the most out of the batch of 10^4 samples. In cases with true Markovian evolution, the optimal pair should essentially be random, since all pairs produce the same value, namely, 0. In Fig. 4.20, noise in the reconstructed process maps causes small amounts of fictitious non-Markovianity to appear, which is sensed by the metric. To remedy this, one can first fit the reconstructed processes to some functional form, such as the product of a decaying exponential and a sinusoid, in order to extract an analytic model from the MLE routine. Here, we simply use a moving average window to smooth the data, which improves the consistency of the pair found via sampling across multiple runs of the script.

Now, you might have noticed that the plot captions say Λ -system. The above plots were actually simulated for a system identical to the one described in section 4.3.2, except that the $|2\rangle \rightarrow |1\rangle$ decay channel was replaced with a $|2\rangle \rightarrow |0\rangle$ channel of the same strength, forming a Λ -system instead of a ladder system. Below, we show the plots for the ladder system originally described.

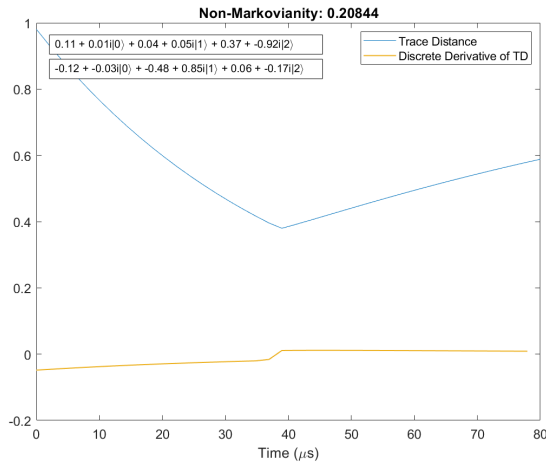


Figure 4.21: Trace distance over time for pair that displays the most non-Markovianity, using the ground-truth ladder system AP channel.

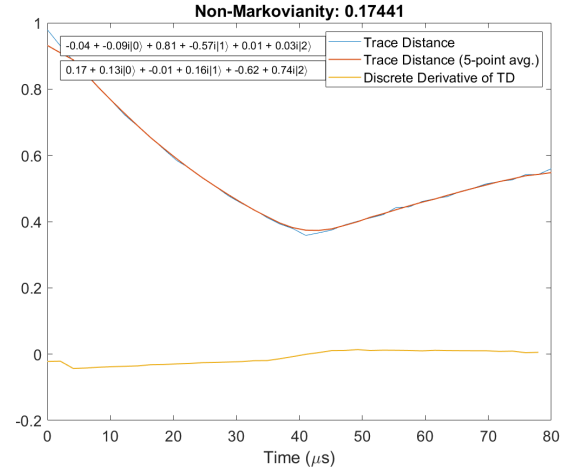


Figure 4.22: Trace distance over time for pair that displays the most non-Markovianity, using the ladder system AP channel found by MLE.

Here, we observe an unexpected result. Amplitude and phase damping channels are prototypical Markovian evolutions—they are useful precisely because they can be added to Lindblad master equations, and phenomenologically they describe observed data in experiments quite well. However, it is clear from

the data that when the initial density matrices are close to $|1\rangle$ and $|2\rangle$, respectively, the evolution is non-Markovian, because the trace distance begins to increase again around $t = 40$. The states become mixed and approach each other, but their purity returns because there are two fixed points, with $|1\rangle \rightarrow |0\rangle$ and $|2\rangle \rightarrow |1\rangle$. This behavior apparently comes from the amplitude damping component of the channel, where we assumed that decay could only occur sequentially, so from $2 \rightarrow 1$ and $1 \rightarrow 0$. Since the Kraus form represents a convex sum over different things that can happen to the system, in the way we have rewritten it, technically *either* of the decays can occur, but not one after another, since it describes *macroscopic* time evolution. In reality, when Kraus operators are imported into the Lindblad equation, they are expanded for small time and are applied infinitesimally as collapse operators. The interpretation of the Lindblad equation is that throughout the unitary evolution, it is as if we are constantly measuring with the collapse operators with some rate, and indeed, the *iterated* application of the channel would cause $|1\rangle$ to decay to $|0\rangle$, because the iteration allows for the composition of Kraus operators (such as $|2\rangle \rightarrow |1\rangle \rightarrow |0\rangle$). For the sake of confirming that a Markovian evolution is recognized as such, we modified the simulation after noticing this edge case, and presented the Λ -system results first, which hopefully was not too confusing.

For the curious reader, the reason why the dip occurs at $t \approx 40$ is because the trace distance is computing the sum of the absolute values of a matrix of the form (if using $|1\rangle$ and $|2\rangle$ as the initial states)

$$\begin{bmatrix} 1 - a(t) & 0 & 0 \\ 0 & a(t) - (1 - b(t)) & 0 \\ 0 & 0 & -b(t) \end{bmatrix} \quad (4.81)$$

where $a(t)$ and $b(t)$ are the populations of the $|1\rangle$ and $|2\rangle$ states over time. The trace distance is

$$\frac{1}{2}(1 - a(t) + |a(t) - (1 - b(t))| + b(t)) \quad (4.82)$$

At $t = 0$ and $t = \infty$, the trace distance is 1, but at some intermediate time, the remaining population in $|1\rangle$ exactly cancels what has decayed from $|2\rangle$ causing the middle term to be 0. The trace distance then is just $b(t)$ at this time, which is less than 1, proving that the trace distance must decay and then revive. The value $t \approx 40$ depends on the two decay times, and it turns out that

$$e^{-t/82} + e^{-t/41} = 1 \quad (4.83)$$

is satisfied if $t = 39.8$.

4.4.2. Transmon Device Evolution

In this section, we apply the trace-distance measure to the transmon process maps found in section 4.3.3. The plots appear in the same order as before, with the environment qutrit in the state $\{|0\rangle, |1\rangle, |+\rangle_{01}, |2\rangle, |+\rangle_{12}\}$. Deviating from the random search used above, the pairs of states in this section are found by solving a separate optimization problem using `fmincon`, where the objective to maximize is the Markovianity defined via the trace distance, and the free parameters are the matrix elements of pair of pure states, $|\psi\rangle_1$ and $|\psi\rangle_2$. The initialization is random, which is sufficient for our purposes, but in general, the best approach is to first obtain a coarse random sampling over initial pairs (like in the previous section), followed by a nonlinear optimization at each random initial point. Combining sampling and optimization allows one to accurately probe each local minimum, without getting stuck in any particular one.

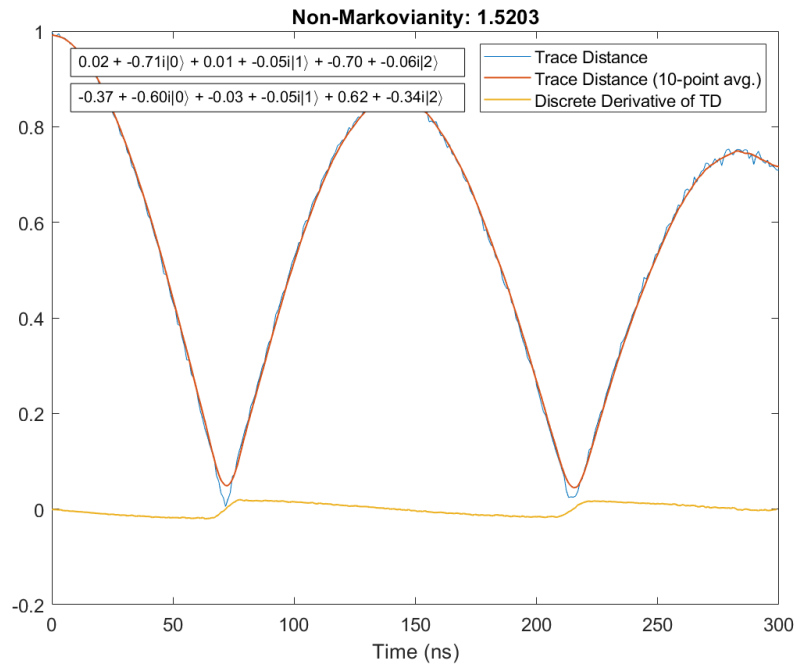


Figure 4.23: Trace distance over time for left qutrit, right qutrit in $|0\rangle$ state.

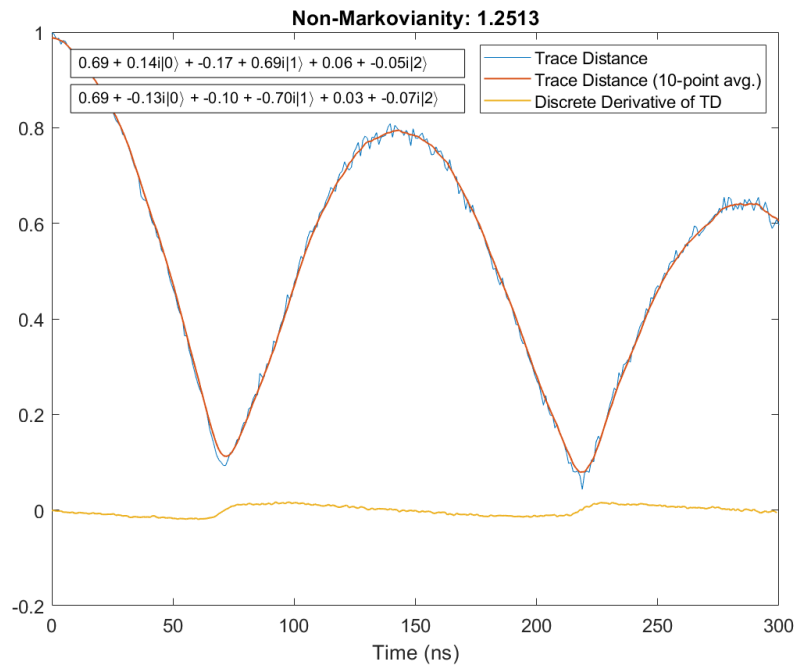


Figure 4.24: Trace distance over time for left qutrit, right qutrit in $|1\rangle$ state.

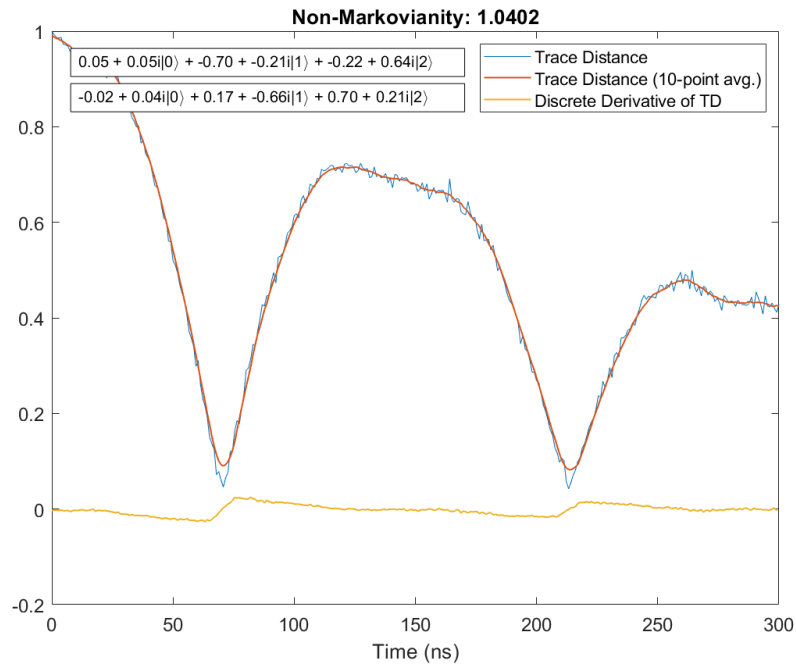


Figure 4.25: Trace distance over time for left qutrit, right qutrit in $|+\rangle_{01}$ state.

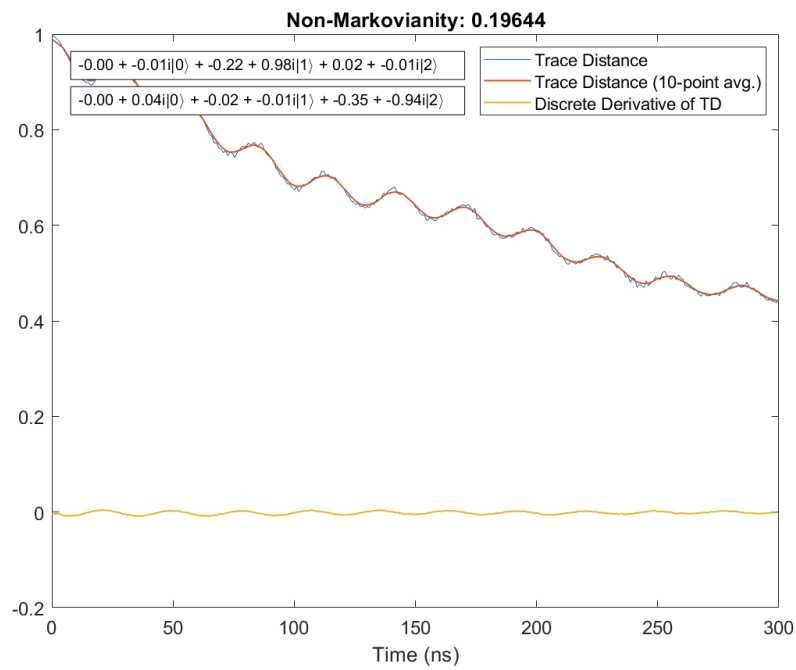


Figure 4.26: Trace distance over time for left qutrit, right qutrit in $|2\rangle$ state.

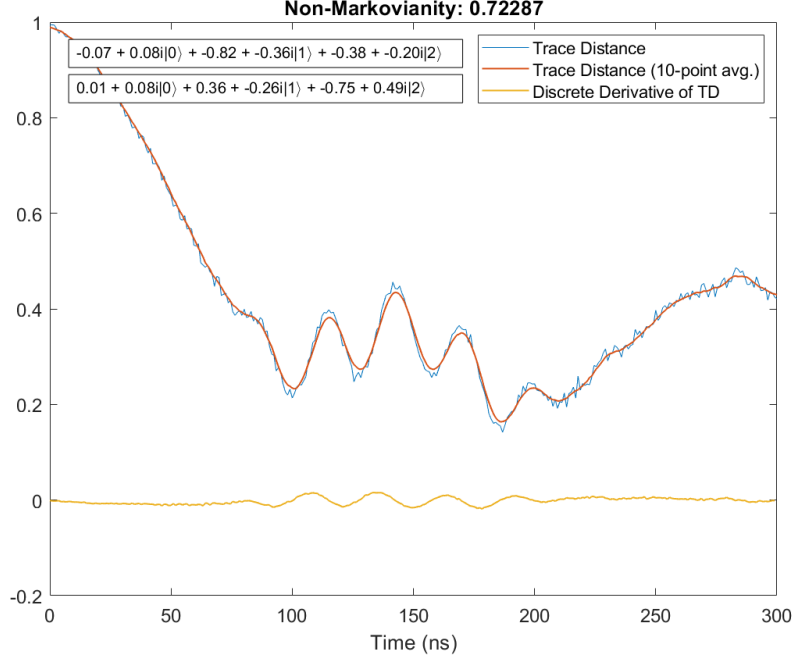


Figure 4.27: Trace distance over time for left qutrit, right qutrit in $|+\rangle_{12}$ state.

$ \psi\rangle_{env}$	$ \psi\rangle_1$	$ \psi\rangle_2$
$ 0\rangle$	$0.50 0\rangle + 0.00 1\rangle + 0.49 2\rangle$	$0.50 0\rangle + 0.00 1\rangle + 0.50 2\rangle$
$ 1\rangle$	$0.50 0\rangle + 0.50 1\rangle + 0.01 2\rangle$	$0.49 0\rangle + 0.50 1\rangle + 0.01 2\rangle$
$1/\sqrt{2}(0\rangle + 1\rangle)$	$0.01 0\rangle + 0.53 1\rangle + 0.46 2\rangle$	$0.00 0\rangle + 0.46 1\rangle + 0.53 2\rangle$
$ 2\rangle$	$0.00 0\rangle + 1.00 1\rangle + 0.00 2\rangle$	$0.00 0\rangle + 0.00 1\rangle + 1.00 2\rangle$
$1/\sqrt{2}(1\rangle + 2\rangle)$	$0.01 0\rangle + 0.80 1\rangle + 0.18 2\rangle$	$0.01 0\rangle + 0.20 1\rangle + 0.80 2\rangle$

Table 4.4: Pairs of input states that maximized the increase of their trace distance over time, while evolving in the presence of an environment qutrit $|\psi\rangle_{env}$, subject to the two-transmon device Hamiltonian. The state weights are probabilities.

We organize the results of the above plots, listing the pairs of initial states that maximized the non-Markovianity in their respective cases in Table 4.4. Each state is expanded in terms of its probabilities over the qutrit basis. In each case, the optimal pair of states lies primarily within one of the qubit subspaces, with little weight in the excluded state. These states are plotted in Fig. 4.28. The order of the Bloch spheres, from top to bottom, is the same as the row order of the tables. For example, the top Bloch sphere contains the optimal pair of states when the environment is in the $|0\rangle$ state, which have virtually no weight in the $|1\rangle$ state, so the 02-Bloch sphere is shown. The red and blue dots are the projections of the orange and green vectors, respectively, which help denote where the states actually lie within the spheres.

The particular states listed Table 4.4 allow us to connect the theory of the Markovianity metric with our physical intuition based on the device Hamiltonian. Recall that the metric wants to find states that, assuming they start distinguishable (say, on opposite sides of a sub-Bloch sphere), will evolve to become nearly identical, and then later recover some of their initial distinction. This will occur when the joint system becomes entangled, such that tracing over the environment produces a mixed state. When the two initial states become maximally entangled in their own evolutions, their reduced forms will be maximally mixed, and the trace distance between them will be 0, or at least small. When the entanglement oscillates away, the states become distinguishable again. Also, recall that the interaction Hamiltonian couples states that differ by one photon, and that the trace distance between $|i\rangle$ and $|j\rangle$ is 1, if $i \neq j; i, j \in (0, 1, 2)$. The level structure of the system (Fig. 3.8) at $\phi_e = 0.15$, when the coupling is turned on, indicates four main near-resonances, listed in order of decreasing strength (roughly): $|11\rangle \leftrightarrow |20\rangle$, $|01\rangle \leftrightarrow |10\rangle$, $|12\rangle \leftrightarrow |21\rangle$, $|11\rangle \leftrightarrow |02\rangle$. The first is much stronger than the other three, and the stronger the interaction, the more completely the two states are swapped. Lastly, remember that the amplitude damping in the simulation causes *all* initial states to approach $|0\rangle$ eventually.

In the first case, the right qutrit (environment) is $|0\rangle$, and the most entanglement occurs if the system states are orthogonal superpositions of $|0\rangle$ and $|2\rangle$, so that during the strongest interaction, the Bell states $|00\rangle \pm |11\rangle$ are formed, resulting in identical mixed states near $t = 70\text{ns}$.

In the second case, a similar situation arises, but the relevant entangled states are $|01\rangle \pm |20\rangle$. The situation is complicated by the $|11\rangle \rightarrow |02\rangle$ transition, which is evident in Fig. 4.24, whose trace distance does not dip as low as the first case. Extra kinds of entanglement, with their own timescales, will generally prevent the system states from becoming maximally mixed at the same time.

When the environment is $1/\sqrt{2}(|0\rangle + |1\rangle)$, this picture is even more complicated, and we see that the system states struggle to ever be far from each other (the flattened fringes in Fig. 4.25), which is presumably due to the states always being entangled via some component of the interaction. Interestingly, superpositions of $|1\rangle$ and $|2\rangle$ now comprise the preferred pair.

The fourth case is apparently quite different, looking like a small non-Markovian oscillation added to a Markovian decay. The chosen pair is $|2\rangle$ (which doesn't evolve at all) and $|1\rangle$. The non-Markovianity is weak because the system has no way to access the strongest exchange, and ultimately relies on the weaker coupling $|12\rangle \leftrightarrow |21\rangle$, where the reduced state is able to slightly approach $|2\rangle$, causing small resurgences.

In the last case, the environment is in the state $1/\sqrt{2}(|1\rangle + |2\rangle)$. We would then expect the $|2\rangle$ component to contribute the small oscillations we just saw, while the $|1\rangle$ component allows access to the strong exchange, causing large resurgences like in the first three plots. Fig. 4.27 shows that this is indeed the case. We can also see this via Fig. 4.28, where in the first three cases, the pairs all lie within their respective equators, but in the fourth, the states lie along the z-axis; it then makes sense that for the fifth case, the pairs lie between the equator and the poles.

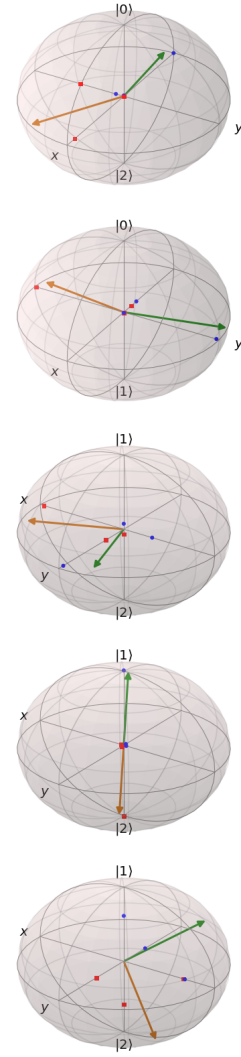


Figure 4.28: States of Table 4.4

Overall, the dynamics are most non-Markovian when the environment is in the qubit subspace, which is evident in the strongest exchange term $|11\rangle \leftrightarrow |20\rangle$. Paralleling our previous analysis, we also show the non-Markovianity for the 0 flux case, when the environment is $|0\rangle$.

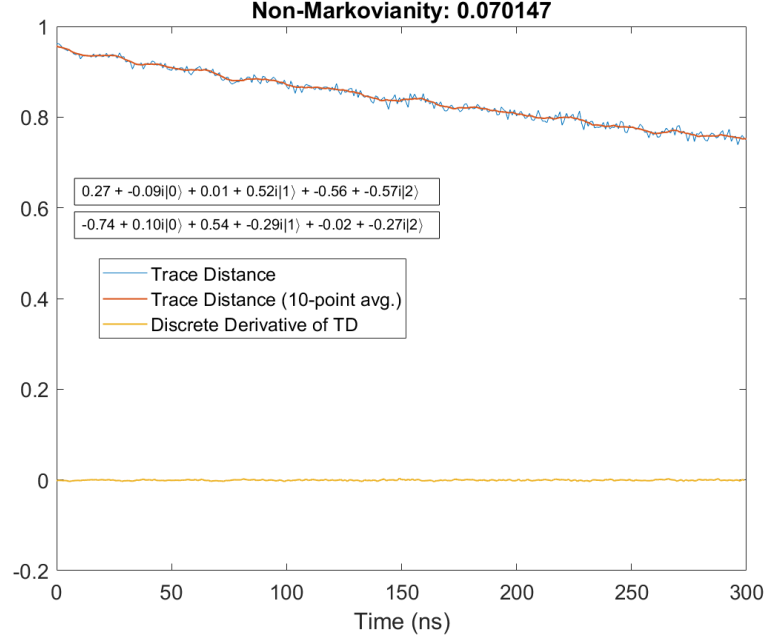


Figure 4.29: Trace distance over time for left qutrit, right qutrit in $|1\rangle$ state (flux off).

As expected, the evolution is mostly Markovian, but not completely, due to the always-on coupling. The goal of all this analysis was to provide a convincing link between actual, physical systems of interest, and the conclusions one could draw by using the MLE framework and Markovianity metric. This further substantiates the correctness of the methods described in this report, and illustrates their use.

4.5. Analysis

Our work contrasts with the methods in section 3.4 in several important ways. The first step, fitting the POVM, is something between QPT and GST, because though we characterize the null process, we end up with a self-consistent parameterization of the SPAM procedure, where the error is neither attributed entirely to the state preparation or measurement. Though it is still possible that such error could be spread incorrectly across these parameters, using a physically informed model (i.e. initial state is not totally unknown but mixed, rotations are not totally unknown but just are over/under-rotations) should constrain the optimizer and limit the number of extraneous models that can mimic the data. The second step, fitting the process map, is indeed equivalent to QPT, but performing it over the course of an individual process—as opposed to characterizing a gate in its entirety, as is the usual case—provides crucial information regarding the CP-divisibility of the dynamics, which in turn illuminates the Markovianity of the system-environment interaction. Importantly, none of the methods in section 3.4, in their standard incarnations, can probe non-Markovian processes.

Furthermore, we note that fixing the POVM parameters in the subsequent process estimation, as opposed to simultaneously optimizing over them *and* the process as in GST, guarantees a lower-order optimization problem. In the best case (Gaussian likelihood based on Pauli transfer matrix), GST must deal with a 10th order polynomial [26]. Neither approach is very efficient, however, the per-trial time of our method is quite fast. Each batch of SPAM parameters and snapshot of process parameters was estimated in a few seconds. Obtaining the overall dynamics of some process could take a while, depending on the desired temporal resolution. In essence, our method requires solving many inexpensive MLE instances.

The validity of our assumption that the SPAM parameters may be fixed in the process estimation entirely depends on the stability of the experimental device over time. For realistic device parameters,

running 1000 shots takes a time on the order of $O(1)$ seconds, assuming preparation and measurements take a few microseconds, and that gate times are negligible, taking nanoseconds. There are 3^2 and 3^4 experiments to estimate qutrit states/POVMs and processes, respectively. The total timescale should be upper bounded by several minutes per resolved timestep, whereas qubit T_1/T_2 lifetimes—and by extension, their environments—typically fluctuate on the scale of hours. Therefore, our assumption is justified if one takes up to around 60 snapshots of the time-evolution. This simply demonstrates that realistic timescales do not completely prohibit our procedure; of course, a more careful analysis is needed once a device has actually been selected, and these timescales better quantified.

Conclusion

5.1. Future Directions and Potential Improvements

Having completed and evaluated the framework presented herein, there are several avenues for exploration, and also many potential improvements.

The experimental data used to estimate the process maps could also be fit directly to a Lindblad master equation. If the evolution was already found to be sufficiently Markovian, then the reconstructed evolutions under the process map and Lindblad ME should be very similar. Extra features present in the process map evolution, but absent in the fitted master equation evolution, would represent non-Markovian dynamics. This approach has previously been demonstrated for single[29] and two-qubit[53] cases in a solid-state setting, but the qutrit case is unexplored.

Regarding further evaluation of the framework, different ansatzes for errors in the initial state and rotation matrices could be incorporated in the MLE objective function, and instead of the idle process, the dynamics of quantum gates should be investigated. Other models known to be non-Markovian, such as the damped-Jaynes Cummings Hamiltonian, or more complicated circuit QED device models, would be worth simulating as well. With the time-dependent process in hand, other metrics of non-Markovianity may be utilized, such as the negative canonical rate approach of Hall et al. [27]. Importantly, applying the various metrics listed in section 3.2.6 to real data could cast the deciding vote on which to prefer, basing their quality in terms of practicality, similar to the analysis of Uriri et al. [55].

Regarding the MLE procedure itself, the likelihood function could also be expressed in a Gaussian form, instead of a product of independent probabilities. Furthermore, Bayesian mean estimation could be used instead of MLE entirely, which remedies the conundrum that MLE might produce estimates involving probabilities equal to 0, which are unphysical¹ and impossible to obtain error bars for [10]. Speaking of error bars, when applying the MLE framework to real-world data, there will no longer be any ground-truth to reference, so the error of the estimated quantities will need to be considered differently. This is actually a subtle issue [26], because the variance of the MLE estimate due to sampling error eludes the standard Hessian-based method, due to the lack of asymptotic normality². One way to assess sampling error is to use a Monte-Carlo approach, which is essentially what was performed in this report—i.e. generating many datasets within different error regimes, and evaluating the fit quality as a function of that error. We did not, however, vary the sampling error, choosing instead to leave it built-in to the procedure, and quantifying the effects of systematic deficiencies in the SPAM parameters.

Lastly, it would be remiss to not discuss the issue of scaling up, though any method based on quantum process tomography does not bode well in this regard. However, there do exist bootstrapping methods such as the pairwise perturbative ansatz (PAPA) [25], which trades accuracy for polynomial efficiency, but still remains descriptive. PAPA expresses an n -qubit process in terms of an ansatz built out of two-qubit processes, which are then informed via the tomographic method of choice, which could be QST, GST, or a 2-qutrit extension of this work.

¹In short, it should be impossible to say something occurs with 0 probability based on only a finite number of samples.

²We intentionally bias the output of MLE to be physical, which violates certain statistical assumptions.

5.2. Summary

In this report, we built, validated, and then utilized a tomographic framework to reconstruct the dynamics of a qutrit in time, self-consistently estimating the SPAM parameters of the system via maximum-likelihood estimation. The time-dependent process map was then used to quantify the Markovianity of an evolution based on the behavior of the trace distance between certain initial states. Both Markovian and non-Markovian examples were studied, via the qutrit amplitude and phase damping channel, and a system of two coupled superconducting qutrits, respectively.

Potential uses include pinpointing when and how a qubit state undergoes a leakage error (when elements of the process matrix take the system out of the computational subspace); crosstalk errors due to spurious entanglement (when the evolution becomes non-Markovian while an adjacent qubit/trit is driven); and the strength and type of the dominant noise channels (when the decay and/or oscillation of certain process elements fluctuate); all estimated in a manner consistent with the inherent SPAM conditions (which can be informed based on one's prior understanding of the system).

We hope that this framework represents a step towards more careful device characterization, as hardware progresses from the NISQ regime to fault-tolerance, and also promotes the use of qutrits as a computational resource in their own right, which may lead to more powerful demonstrations of quantum advantage.

References

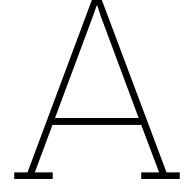
- [1] Dorit Aharonov, Alexei Kitaev, and Noam Nisan. “Quantum Circuits with Mixed States”. In: *arXiv:quant-ph/9806029* (June 1998). arXiv: quant-ph/9806029. URL: <http://arxiv.org/abs/quant-ph/9806029> (visited on 02/09/2021).
- [2] Robert Alicki, Daniel A. Lidar, and Paolo Zanardi. “Internal consistency of fault-tolerant quantum error correction in light of rigorous derivations of the quantum Markovian limit”. en. In: *Physical Review A* 73.5 (May 2006), p. 052311. ISSN: 1050-2947, 1094-1622. DOI: [10.1103/PhysRevA.73.052311](https://doi.org/10.1103/PhysRevA.73.052311). URL: <https://link.aps.org/doi/10.1103/PhysRevA.73.052311> (visited on 02/11/2021).
- [3] R. Barends et al. “Superconducting quantum circuits at the surface code threshold for fault tolerance”. en. In: *Nature* 508.7497 (Apr. 2014), pp. 500–503. ISSN: 0028-0836, 1476-4687. DOI: [10.1038/nature13171](https://doi.org/10.1038/nature13171). URL: <http://www.nature.com/articles/nature13171> (visited on 07/16/2020).
- [4] Francesco Battistel, Boris M. Varbanov, and Barbara M. Terhal. “A hardware-efficient leakage-reduction scheme for quantum error correction with superconducting transmon qubits”. In: *arXiv:2102.08336 [quant-ph]* (Feb. 2021). arXiv: 2102.08336. URL: <http://arxiv.org/abs/2102.08336> (visited on 02/18/2021).
- [5] Giuliano Benenti and Giuliano Strini. “Computing the distance between quantum channels: Usefulness of the Fano representation”. In: *Journal of Physics B: Atomic, Molecular and Optical Physics* 43.21 (Nov. 2010). arXiv: 1004.4110, p. 215508. ISSN: 0953-4075, 1361-6455. DOI: [10.1088/0953-4075/43/21/215508](https://doi.org/10.1088/0953-4075/43/21/215508). URL: <http://arxiv.org/abs/1004.4110> (visited on 02/09/2021).
- [6] Reinhold A Bertlmann and Philipp Krammer. “Bloch vectors for qudits”. en. In: *Journal of Physics A: Mathematical and Theoretical* 41.23 (June 2008), p. 235303. ISSN: 1751-8113, 1751-8121. DOI: [10.1088/1751-8113/41/23/235303](https://doi.org/10.1088/1751-8113/41/23/235303). URL: <https://iopscience.iop.org/article/10.1088/1751-8113/41/23/235303> (visited on 01/15/2021).
- [7] R. Bianchetti et al. “Control and Tomography of a Three Level Superconducting Artificial Atom”. en. In: *Physical Review Letters* 105.22 (Nov. 2010), p. 223601. ISSN: 0031-9007, 1079-7114. DOI: [10.1103/PhysRevLett.105.223601](https://doi.org/10.1103/PhysRevLett.105.223601). URL: <https://link.aps.org/doi/10.1103/PhysRevLett.105.223601> (visited on 07/16/2020).
- [8] T. Bækkegaard et al. “Realization of efficient quantum gates with a superconducting qubit-qutrit circuit”. en. In: *Scientific Reports* 9.1 (Dec. 2019), p. 13389. ISSN: 2045-2322. DOI: [10.1038/s41598-019-49657-1](https://doi.org/10.1038/s41598-019-49657-1). URL: <http://www.nature.com/articles/s41598-019-49657-1> (visited on 01/15/2021).
- [9] M. S. Blok et al. “Quantum Information Scrambling in a Superconducting Qutrit Processor”. In: *arXiv:2003.03307 [quant-ph]* (Mar. 2020). arXiv: 2003.03307. URL: <http://arxiv.org/abs/2003.03307> (visited on 07/25/2020).
- [10] Robin Blume-Kohout. “Optimal, reliable estimation of quantum states”. In: *New Journal of Physics* 12.4 (Apr. 2010). arXiv: quant-ph/0611080, p. 043034. ISSN: 1367-2630. DOI: [10.1088/1367-2630/12/4/043034](https://doi.org/10.1088/1367-2630/12/4/043034). URL: <http://arxiv.org/abs/quant-ph/0611080> (visited on 07/18/2020).
- [11] Robin Blume-Kohout et al. “Robust, self-consistent, closed-form tomography of quantum logic gates on a trapped ion qubit”. In: *arXiv:1310.4492 [quant-ph]* (Oct. 2013). arXiv: 1310.4492. URL: <http://arxiv.org/abs/1310.4492> (visited on 07/18/2020).
- [12] Heinz-Peter Breuer. “Foundations and measures of quantum non-Markovianity”. en. In: *Journal of Physics B: Atomic, Molecular and Optical Physics* 45.15 (Aug. 2012), p. 154001. ISSN: 0953-4075, 1361-6455. DOI: [10.1088/0953-4075/45/15/154001](https://doi.org/10.1088/0953-4075/45/15/154001). URL: <https://iopscience.iop.org/article/10.1088/0953-4075/45/15/154001> (visited on 11/09/2020).

- [13] Heinz-Peter Breuer, Elsi-Mari Laine, and Jyrki Piilo. "Measure for the Degree of Non-Markovian Behavior of Quantum Processes in Open Systems". en. In: *Physical Review Letters* 103.21 (Nov. 2009), p. 210401. ISSN: 0031-9007, 1079-7114. DOI: [10.1103/PhysRevLett.103.210401](https://doi.org/10.1103/PhysRevLett.103.210401). URL: <https://link.aps.org/doi/10.1103/PhysRevLett.103.210401> (visited on 10/06/2020).
- [14] Heinz-Peter Breuer and F. Petruccione. *The theory of open quantum systems*. OCLC: ocm49872077. Oxford ; New York: Oxford University Press, 2002. ISBN: 978-0-19-852063-4.
- [15] Heinz-Peter Breuer et al. "Colloquium : Non-Markovian dynamics in open quantum systems". en. In: *Reviews of Modern Physics* 88.2 (Apr. 2016), p. 021002. ISSN: 0034-6861, 1539-0756. DOI: [10.1103/RevModPhys.88.021002](https://doi.org/10.1103/RevModPhys.88.021002). URL: <https://link.aps.org/doi/10.1103/RevModPhys.88.021002> (visited on 10/07/2020).
- [16] D. Bruß and C. Macchiavello. "Optimal Eavesdropping in Cryptography with Three-Dimensional Quantum States". en. In: *Physical Review Letters* 88.12 (Mar. 2002), p. 127901. ISSN: 0031-9007, 1079-7114. DOI: [10.1103/PhysRevLett.88.127901](https://doi.org/10.1103/PhysRevLett.88.127901). URL: <https://link.aps.org/doi/10.1103/PhysRevLett.88.127901> (visited on 01/29/2021).
- [17] Earl T. Campbell. "Enhanced Fault-Tolerant Quantum Computing in d -Level Systems". en. In: *Physical Review Letters* 113.23 (Dec. 2014), p. 230501. ISSN: 0031-9007, 1079-7114. DOI: [10.1103/PhysRevLett.113.230501](https://doi.org/10.1103/PhysRevLett.113.230501). URL: <https://link.aps.org/doi/10.1103/PhysRevLett.113.230501> (visited on 01/29/2021).
- [18] Paola Cappellaro. *Quantum Theory of Radiation Interactions*, MIT 22.51, *Lecture Notes*. 2012. URL: https://ocw.mit.edu/courses/nuclear-engineering/22-51-quantum-theory-of-radiation-interactions-fall-2012/lecture-notes/MIT22_51F12_Notes.pdf.
- [19] Carlton Caves. *Quantum Information Theory*, UNW Phys 572.001, *Lectures* 17-20. 2014. URL: <http://info.phys.unm.edu/~caves/courses/qinfo-f14/syllabus.html>.
- [20] A. Checinska and K. Wodkiewicz. "Noisy Qutrit Channels". In: *arXiv:quant-ph/0610127* (Nov. 2006). arXiv: quant-ph/0610127. URL: <http://arxiv.org/abs/quant-ph/0610127> (visited on 09/30/2020).
- [21] David P. DiVincenzo. "The Physical Implementation of Quantum Computation". en. In: *Fortschritte der Physik* 48.9-11 (2000). eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/1521-3978%28200009%2948%3A%2F9%3E3.0.CO%3B2-E>, pp. 771–783. ISSN: 1521-3978. DOI: [https://doi.org/10.1002/1521-3978\(200009\)48:9/11<771::AID-PROP771>3.0.CO;2-E](https://doi.org/10.1002/1521-3978(200009)48:9/11<771::AID-PROP771>3.0.CO;2-E). (Visited on 01/10/2021).
- [22] D.J. Egger et al. "Pulsed Reset Protocol for Fixed-Frequency Superconducting Qubits". en. In: *Physical Review Applied* 10.4 (Oct. 2018), p. 044030. ISSN: 2331-7019. DOI: [10.1103/PhysRevApplied.10.044030](https://doi.org/10.1103/PhysRevApplied.10.044030). URL: <https://link.aps.org/doi/10.1103/PhysRevApplied.10.044030> (visited on 01/16/2021).
- [23] A. Fedorov et al. "Implementation of a Toffoli gate with superconducting circuits". en. In: *Nature* 481.7380 (Jan. 2012), pp. 170–172. ISSN: 0028-0836, 1476-4687. DOI: [10.1038/nature10713](https://doi.org/10.1038/nature10713). URL: <http://www.nature.com/articles/nature10713> (visited on 01/26/2021).
- [24] C. A. Fuchs and J. van de Graaf. "Cryptographic distinguishability measures for quantum-mechanical states". In: *IEEE Transactions on Information Theory* 45.4 (May 1999). Conference Name: IEEE Transactions on Information Theory, pp. 1216–1227. ISSN: 1557-9654. DOI: [10.1109/18.761271](https://doi.org/10.1109/18.761271).
- [25] L. C. G. Govia et al. "Bootstrapping quantum process tomography via a perturbative ansatz". en. In: *Nature Communications* 11.1 (Dec. 2020), p. 1084. ISSN: 2041-1723. DOI: [10.1038/s41467-020-14873-1](https://doi.org/10.1038/s41467-020-14873-1). URL: <http://www.nature.com/articles/s41467-020-14873-1> (visited on 07/16/2020).
- [26] Daniel Greenbaum. "Introduction to Quantum Gate Set Tomography". en. In: *arXiv:1509.02921 [quant-ph]* (Sept. 2015). arXiv: 1509.02921. URL: <http://arxiv.org/abs/1509.02921> (visited on 07/16/2020).
- [27] Michael J. W. Hall et al. "Canonical form of master equations and characterization of non-Markovianity". en. In: *Physical Review A* 89.4 (Apr. 2014), p. 042120. ISSN: 1050-2947, 1094-1622. DOI: [10.1103/PhysRevA.89.042120](https://doi.org/10.1103/PhysRevA.89.042120). URL: <https://link.aps.org/doi/10.1103/PhysRevA.89.042120> (visited on 11/13/2020).

- [28] Erik Hostens, Jeroen Dehaene, and Bart De Moor. “Stabilizer states and Clifford operations for systems of arbitrary dimensions and modular arithmetic”. en. In: *Physical Review A* 71.4 (Apr. 2005), p. 042315. issn: 1050-2947, 1094-1622. doi: [10.1103/PhysRevA.71.042315](https://doi.org/10.1103/PhysRevA.71.042315). url: <https://link.aps.org/doi/10.1103/PhysRevA.71.042315> (visited on 01/15/2021).
- [29] M Howard et al. “Quantum process tomography and Linblad estimation of a solid-state qubit”. en. In: *New Journal of Physics* 8.3 (Mar. 2006), pp. 33–33. issn: 1367-2630. doi: [10.1088/1367-2630/8/3/033](https://doi.org/10.1088/1367-2630/8/3/033). url: <https://iopscience.iop.org/article/10.1088/1367-2630/8/3/033> (visited on 07/16/2020).
- [30] Daniel F. V. James et al. “Measurement of qubits”. en. In: *Physical Review A* 64.5 (Oct. 2001), p. 052312. issn: 1050-2947, 1094-1622. doi: [10.1103/PhysRevA.64.052312](https://doi.org/10.1103/PhysRevA.64.052312). url: <https://link.aps.org/doi/10.1103/PhysRevA.64.052312> (visited on 01/19/2021).
- [31] X.Y. Jin et al. “Thermal and Residual Excited-State Population in a 3D Transmon Qubit”. en. In: *Physical Review Letters* 114.24 (June 2015), p. 240501. issn: 0031-9007, 1079-7114. doi: [10.1103/PhysRevLett.114.240501](https://doi.org/10.1103/PhysRevLett.114.240501). url: <https://link.aps.org/doi/10.1103/PhysRevLett.114.240501> (visited on 01/16/2021).
- [32] J. R. Johansson, P. D. Nation, and Franco Nori. “QuTiP 2: A Python framework for the dynamics of open quantum systems”. In: *Computer Physics Communications* 184.4 (Apr. 2013). arXiv: 1211.6518, pp. 1234–1240. issn: 00104655. doi: [10.1016/j.cpc.2012.11.019](https://doi.org/10.1016/j.cpc.2012.11.019). url: <http://arxiv.org/abs/1211.6518> (visited on 02/11/2021).
- [33] M. Kjaergaard et al. “A Quantum Instruction Set Implemented on a Superconducting Quantum Processor”. In: *arXiv:2001.08838 [quant-ph]* (Jan. 2020). arXiv: 2001.08838. url: <http://arxiv.org/abs/2001.08838> (visited on 07/16/2020).
- [34] E. Knill et al. “Randomized Benchmarking of Quantum Gates”. In: *Physical Review A* 77.1 (Jan. 2008). arXiv: 0707.0963, p. 012307. issn: 1050-2947, 1094-1622. doi: [10.1103/PhysRevA.77.012307](https://doi.org/10.1103/PhysRevA.77.012307). url: <http://arxiv.org/abs/0707.0963> (visited on 02/03/2021).
- [35] Jens Koch and Peter Groszkowski. *scqubits*. <https://github.com/scqubits>. 2019. url: <https://github.com/scqubits>.
- [36] Philip Krantz et al. “A Quantum Engineer’s Guide to Superconducting Qubits”. In: *Applied Physics Reviews* 6.2 (June 2019). arXiv: 1904.06560, p. 021318. issn: 1931-9401. doi: [10.1063/1.5089550](https://doi.org/10.1063/1.5089550). url: <http://arxiv.org/abs/1904.06560> (visited on 10/14/2019).
- [37] Jaseung Ku et al. “Suppression of Unwanted Z Z Interactions in a Hybrid Two-Qubit System”. en. In: *Physical Review Letters* 125.20 (Nov. 2020), p. 200504. issn: 0031-9007, 1079-7114. doi: [10.1103/PhysRevLett.125.200504](https://doi.org/10.1103/PhysRevLett.125.200504). url: <https://link.aps.org/doi/10.1103/PhysRevLett.125.200504> (visited on 01/12/2021).
- [38] Gabriel T Landi. *Quantum Information and Quantum Noise*. en. 2018. url: <http://www.fmt.if.usp.br/~gtlandi/quantum-information-and.html>.
- [39] Bi-Heng Liu et al. “Experimental control of the transition from Markovian to non-Markovian dynamics of open quantum systems”. en. In: *Nature Physics* 7.12 (Dec. 2011), pp. 931–934. issn: 1745-2473, 1745-2481. doi: [10.1038/nphys2085](https://doi.org/10.1038/nphys2085). url: <http://www.nature.com/articles/nphys2085> (visited on 11/27/2020).
- [40] J. S. Lundeen et al. “Tomography of quantum detectors”. en. In: *Nature Physics* 5.1 (Jan. 2009), pp. 27–30. issn: 1745-2473, 1745-2481. doi: [10.1038/nphys1133](https://doi.org/10.1038/nphys1133). url: <http://www.nature.com/articles/nphys1133> (visited on 01/29/2021).
- [41] P. Magnard et al. “Fast and Unconditional All-Microwave Reset of a Superconducting Qubit”. en. In: *Physical Review Letters* 121.6 (Aug. 2018), p. 060502. issn: 0031-9007, 1079-7114. doi: [10.1103/PhysRevLett.121.060502](https://doi.org/10.1103/PhysRevLett.121.060502). url: <https://link.aps.org/doi/10.1103/PhysRevLett.121.060502> (visited on 01/16/2021).
- [42] Simon Milz et al. “Completely Positive Divisibility Does Not Mean Markovianity”. In: *Physical Review Letters* 123.4 (July 2019). arXiv: 1901.05223, p. 040401. issn: 0031-9007, 1079-7114. doi: [10.1103/PhysRevLett.123.040401](https://doi.org/10.1103/PhysRevLett.123.040401). url: <http://arxiv.org/abs/1901.05223> (visited on 12/22/2020).

- [43] Joshua Morris, Felix A. Pollock, and Kavan Modi. “Non-Markovian memory in IBMQX4”. In: *arXiv:1902.07980 [quant-ph]* (Feb. 2019). arXiv: 1902.07980. URL: <http://arxiv.org/abs/1902.07980> (visited on 12/22/2020).
- [44] Michael A. Nielsen and Isaac L. Chuang. *Quantum computation and quantum information*. en. 10th anniversary ed. Cambridge ; New York: Cambridge University Press, 2010. ISBN: 978-1-107-00217-3.
- [45] M. G. A. Paris. “The modern tools of quantum mechanics”. en. In: *The European Physical Journal Special Topics* 203.1 (Apr. 2012), pp. 61–86. ISSN: 1951-6401. DOI: [10.1140/epjst/e2012-01535-1](https://doi.org/10.1140/epjst/e2012-01535-1). URL: <https://doi.org/10.1140/epjst/e2012-01535-1> (visited on 01/20/2021).
- [46] Michael J. Peterer et al. “Coherence and Decay of Higher Energy Levels of a Superconducting Transmon Qubit”. In: *Physical Review Letters* 114.1 (Jan. 2015). arXiv: 1409.6031, p. 010501. ISSN: 0031-9007, 1079-7114. DOI: [10.1103/PhysRevLett.114.010501](https://doi.org/10.1103/PhysRevLett.114.010501). URL: <http://arxiv.org/abs/1409.6031> (visited on 07/25/2020).
- [47] Bibek Pokharel et al. “Demonstration of Fidelity Improvement Using Dynamical Decoupling with Superconducting Qubits”. en. In: *Physical Review Letters* 121.22 (Nov. 2018), p. 220502. ISSN: 0031-9007, 1079-7114. DOI: [10.1103/PhysRevLett.121.220502](https://doi.org/10.1103/PhysRevLett.121.220502). URL: <https://link.aps.org/doi/10.1103/PhysRevLett.121.220502> (visited on 02/01/2021).
- [48] Felix A. Pollock et al. “Non-Markovian quantum processes: Complete framework and efficient characterization”. en. In: *Physical Review A* 97.1 (Jan. 2018), p. 012127. ISSN: 2469-9926, 2469-9934. DOI: [10.1103/PhysRevA.97.012127](https://doi.org/10.1103/PhysRevA.97.012127). URL: <https://link.aps.org/doi/10.1103/PhysRevA.97.012127> (visited on 11/05/2020).
- [49] John Preskill. “Quantum Computing in the NISQ era and beyond”. In: (2018). DOI: [10.22331/q-2018-08-06-79](https://arxiv.org/pdf/1801.00862.pdf). URL: <https://arxiv.org/pdf/1801.00862.pdf>.
- [50] Timothy Proctor et al. “What randomized benchmarking actually measures”. In: *Physical Review Letters* 119.13 (Sept. 2017). arXiv: 1702.01853, p. 130502. ISSN: 0031-9007, 1079-7114. DOI: [10.1103/PhysRevLett.119.130502](https://doi.org/10.1103/PhysRevLett.119.130502). URL: <http://arxiv.org/abs/1702.01853> (visited on 07/20/2020).
- [51] Ángel Rivas, Susana F. Huelga, and Martin B. Plenio. “Entanglement and Non-Markovianity of Quantum Evolutions”. en. In: *Physical Review Letters* 105.5 (July 2010), p. 050403. ISSN: 0031-9007, 1079-7114. DOI: [10.1103/PhysRevLett.105.050403](https://doi.org/10.1103/PhysRevLett.105.050403). URL: <https://link.aps.org/doi/10.1103/PhysRevLett.105.050403> (visited on 11/09/2020).
- [52] S. Rosenblum et al. “Fault-tolerant detection of a quantum error”. en. In: *Science* 361.6399 (July 2018), pp. 266–270. ISSN: 0036-8075, 1095-9203. DOI: [10.1126/science.aat3996](https://doi.org/10.1126/science.aat3996). URL: <https://www.sciencemag.org/lookup/doi/10.1126/science.aat3996> (visited on 01/26/2021).
- [53] Gabriel O Samach et al. “Lindblad tomography of a superconducting quantum processor”. In: *arXiv preprint arXiv:2105.02338* (2021).
- [54] Jian-Shun Tang et al. “Measuring non-Markovianity of processes with controllable system-environment interaction”. en. In: *EPL (Europhysics Letters)* 97.1 (Jan. 2012), p. 10002. ISSN: 0295-5075, 1286-4854. DOI: [10.1209/0295-5075/97/10002](https://doi.org/10.1209/0295-5075/97/10002). URL: <https://iopscience.iop.org/article/10.1209/0295-5075/97/10002> (visited on 11/27/2020).
- [55] S. A. Uriri et al. “Experimental investigation of Markovian and non-Markovian channel addition”. en. In: *Physical Review A* 101.5 (May 2020), p. 052107. ISSN: 2469-9926, 2469-9934. DOI: [10.1103/PhysRevA.101.052107](https://doi.org/10.1103/PhysRevA.101.052107). URL: <https://link.aps.org/doi/10.1103/PhysRevA.101.052107> (visited on 11/06/2020).
- [56] Joel J Wallman, Marie Barnhill, and Joseph Emerson. “Robust characterization of leakage errors”. en. In: *New Journal of Physics* 18.4 (Apr. 2016), p. 043021. ISSN: 1367-2630. DOI: [10.1088/1367-2630/18/4/043021](https://doi.org/10.1088/1367-2630/18/4/043021). URL: <https://iopscience.iop.org/article/10.1088/1367-2630/18/4/043021> (visited on 01/10/2021).
- [57] Joel J Wallman and Steven T Flammia. “Randomized benchmarking with confidence”. en. In: *New Journal of Physics* 16.10 (Oct. 2014), p. 103032. ISSN: 1367-2630. DOI: [10.1088/1367-2630/16/10/103032](https://doi.org/10.1088/1367-2630/16/10/103032). URL: <https://iopscience.iop.org/article/10.1088/1367-2630/16/10/103032> (visited on 02/11/2021).

- [58] Yuchen Wang et al. “Qudits and high-dimensional quantum computing”. In: *Frontiers in Physics* 8 (Nov. 2020). arXiv: 2008.00959, p. 589504. ISSN: 2296-424X. DOI: [10.3389/fphy.2020.589504](https://doi.org/10.3389/fphy.2020.589504). URL: <http://arxiv.org/abs/2008.00959> (visited on 01/29/2021).
- [59] Hai-Rui Wei, Bao-Cang Ren, and Fu-Guo Deng. “Geometric measure of quantum discord for a two-parameter class of states in a qubit-qutrit system under various dissipative channels”. In: *Quantum Information Processing* 12.2 (Feb. 2013). arXiv: 1208.0102, pp. 1109–1124. ISSN: 1570-0755, 1573-1332. DOI: [10.1007/s11128-012-0458-8](https://doi.org/10.1007/s11128-012-0458-8). URL: <http://arxiv.org/abs/1208.0102> (visited on 09/30/2020).
- [60] Gregory A. L. White et al. “Experimental non-Markovian process characterisation and control on a quantum processor”. In: *arXiv:2004.14018 [quant-ph]* (Apr. 2020). arXiv: 2004.14018. URL: <http://arxiv.org/abs/2004.14018> (visited on 11/05/2020).
- [61] Xian Wu et al. “High-fidelity software-defined quantum logic on a superconducting qudit”. In: *arXiv:2005.13165 [quant-ph]* (June 2020). arXiv: 2005.13165. URL: <http://arxiv.org/abs/2005.13165> (visited on 07/25/2020).
- [62] Xing Xiao and Yan-Ling Li. “Protecting qutrit-qutrit entanglement by weak measurement and reversal”. In: *The European Physical Journal D* 67.10 (Oct. 2013). arXiv: 1311.4692, p. 204. ISSN: 1434-6060, 1434-6079. DOI: [10.1140/epjd/e2013-40036-3](https://doi.org/10.1140/epjd/e2013-40036-3). URL: <http://arxiv.org/abs/1311.4692> (visited on 11/27/2020).
- [63] Gengyan Zhang. “Tunable Coupling and Its Applications in Circuit Quantum Electrodynamics”. en. PhD thesis. Princeton, 2018.



Quantum Maps

A.1. Vectorization

The vectorization operation is extremely useful in expressing the action of a superoperator via matrix multiplication. If an operator takes vectors to vectors, then a superoperator takes matrices to matrices. There are many forms a superoperator might take, but the particular form we are interested is

$$\mathcal{E}(\rho) = \sum_i A_i \rho B_i \quad (\text{A.1})$$

Vectorization is defined via the relation

$$\text{vec}(|i\rangle\langle j|) = |i\rangle|j\rangle = |i, j\rangle \equiv |m\rangle \quad (\text{A.2})$$

We see that because it takes two-index objects (matrices) to one-index objects (vectors), vectorized operators live in a Hilbert space double the size of the space in which their kets live. where the double-angle brackets remind us that the vector has double the dimension we might otherwise expect from just a single bracket. We like to think of vectors as having only one free index, so the indices i and j are subsumed into some new index, m . If we are not using the natural outer-product basis, then the notation for a matrix A is just

$$\text{vec}(A) = |A\rangle \quad (\text{A.3})$$

Vectorization is the same as stacking the columns of a matrix

$$\text{vec} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a \\ c \\ b \\ d \end{bmatrix} \quad (\text{A.4})$$

It is also possible to vectorize by stacking the rows of the matrix, and though the two conventions are equivalent, it is important to always know which is being employed. The inner product between two operators has a natural vectorized form:

$$\text{Tr}(A^\dagger B) = \text{vec}(A)^\dagger \text{vec}(B) = \langle\langle A|B \rangle\rangle \quad (\text{A.5})$$

In the same way that we normalized vectors, one can divide by the dimension, d , in the above inner product (i.e. if you want the norm of I to be 1). There are some neat identities involving vectorization, which are listed below. Interestingly, vectorizing the identity gives a maximally entangled state

$$|I\rangle\rangle = \sum_i \text{vec}(|i\rangle\langle i|) = \sum_i |ii\rangle \quad (\text{A.6})$$

For the product of three operators,

$$\text{vec}(ABC) = (C^T \otimes A)\text{vec}(B) \quad (\text{A.7})$$

The expansion of a matrix into the outer-product basis

$$A = \left(\sum_i |i\rangle\langle i| \right) A \left(\sum_j |j\rangle\langle j| \right) \quad (\text{A.8})$$

$$= \sum_{i,j} |i\rangle\langle i| A |j\rangle\langle j| \quad (\text{A.9})$$

$$|A\rangle\rangle = \sum_{i,j} (|j\rangle\langle j| \otimes |i\rangle\langle i|) |A\rangle\rangle \quad (\text{A.10})$$

Matrix transposition becomes

$$A^T = \left(\sum_i |i\rangle\langle i| \right) A^T \left(\sum_j |j\rangle\langle j| \right) \quad (\text{A.11})$$

$$= \sum_{i,j} |i\rangle\langle j| A |i\rangle\langle j| \quad (\text{A.12})$$

$$|A^T\rangle\rangle = \sum_{i,j} (|j\rangle\langle i| \otimes |i\rangle\langle j|) |A\rangle\rangle \quad (\text{A.13})$$

So we have seen two different ways a superoperator can act, either via the “quantum operator” form, or the vectorized matrix form. These are summarized below, for some superoperator S in the outer-product basis [19]:

$$S(\cdot) = \sum_{i,j,k,l} S_{ij,kl} |i\rangle\langle j| \cdot |k\rangle\langle l| \quad (\text{A.14})$$

$$S(\cdot) = \sum_{a,b} S_{ab} |a\rangle\rangle\langle\langle b| \cdot \quad (\text{A.15})$$

The first action looks like the usual quantum operator form, where a matrix is dropped in the middle of two linear operators. If we want our basis to have a product state form, we must use the natural basis, as opposed to the Pauli basis, which is equally valid but does not admit a nice two-to-one index correspondence. The second action is the result of first vectorizing the operator basis and the operator input, and then applying S like a matrix. In terms of matrix elements, the following relations can be obtained between the two actions

$$S_{ab} = \langle a|S|b\rangle = \langle i,j|S|k,l\rangle = \text{vec}(|i\rangle\langle j|)^\dagger \cdot S \cdot \text{vec}(|k\rangle\langle l|) \quad (\text{A.16})$$

$$= \text{Tr} \left[|j\rangle\langle i| S (|k\rangle\langle l|) \right] \quad (\text{A.17})$$

$$= \langle i|S(|k\rangle\langle l|)|j\rangle \quad (\text{A.18})$$

$$= S_{ij,kl} \quad (\text{A.19})$$

A.2. Vectorized Lindblad Equation

One very important use of vectorization is that it gives the Lindblad master equation the form of a linear system [18].

$$\dot{\rho} = (\mathcal{H} + \mathcal{G})\rho \quad (\text{A.20})$$

$$\mathcal{H} = -i(H \otimes I + I \otimes H) \quad (\text{A.21})$$

$$\mathcal{G} = \sum_i L_i^* \otimes L_i - \frac{1}{2} (I \otimes (L_i^\dagger L_i) + (L_i^T L_i^*) \otimes I) \quad (\text{A.22})$$

where $\mathcal{H}$ is the vectorized unitary evolution from the Hamiltonian, and $\mathcal{G}$ is the vectorized Lindbladian. The solution is simply

$$\rho(t) = \exp[(\mathcal{H} + \mathcal{G})t]\rho(0) \quad (\text{A.23})$$

A.3. Choi Matrix

Similar to how the vectorization map relates vectors to matrices, the Choi matrix provides a way of uniquely associating quantum channels (i.e. Kraus sums) to matrix superoperators. We have some abstract channel $\Phi : \mathcal{L}(\mathcal{X}) \rightarrow \mathcal{L}(\mathcal{Y})$; in the specific case of density matrices, the output space is equal to the input space (because they are square matrices), so $\mathcal{X} = \mathcal{Y}$, but generally Φ is just some superoperator that takes linear operators on $\mathcal{X}$ to linear operators on $\mathcal{Y}$ (these spaces are Hilbert spaces). The Choi matrix is defined as

$$J(\Phi) = [I \otimes \Phi](|\omega\rangle\langle\omega|) \quad (\text{A.24})$$

where $|\omega\rangle$ is the maximally entangled state

$$|\omega\rangle = \sum_{i=0}^{d-1} |ii\rangle \quad (\text{A.25})$$

Thus, the Choi matrix is the density matrix of a maximally entangled state, half of which is passed through a channel whose effect is given by Φ . There is a choice of which “half” to run through the channel: I have seen the “right”, or second half, more commonly used in the literature, so I will keep that convention here. We can expand out the Choi matrix:

$$J(\Phi) = [I \otimes \Phi] \left(\sum_{i,j=0}^{d-1} |ii\rangle\langle jj| \right) \quad (\text{A.26})$$

$$= [I \otimes \Phi] \left(\sum_{i,j=0}^{d-1} |i\rangle\langle j| \otimes |i\rangle\langle j| \right) \quad (\text{A.27})$$

$$= \sum_{i,j=0}^{d-1} |i\rangle\langle j| \otimes \Phi(|i\rangle\langle j|) \quad (\text{A.28})$$

without assuming anything else about Φ , this is as far as we can go. Now, let's assume that Φ is a valid CPTP quantum channel, which results in the operator-sum form for Φ :

$$\Phi(\rho) = \sum_i K_i \rho K_i^\dagger \quad (\text{A.29})$$

With this, we can proceed with the Choi matrix:

$$J(\Phi) = \sum_{i,j=0}^{d-1} |i\rangle\langle j| \otimes \sum_m K_m (|i\rangle\langle j|) K_m^\dagger \quad (\text{A.30})$$

$$= \sum_m I \otimes K_m \left(\sum_{i,j=0}^{d-1} |ii\rangle\langle jj| \right) I \otimes K_m^\dagger \quad (\text{A.31})$$

$$= \sum_m \left(\sum_i I \otimes K_m |ii\rangle \right) \left(\sum_j \langle jj| I \otimes K_m^\dagger \right) \quad (\text{A.32})$$

Now, the operator $I \otimes K_m$ has a block diagonal form:

$$I \otimes K_m = \begin{bmatrix} [K_m] & & \\ & \ddots & \\ & & [K_m] \end{bmatrix} \quad (\text{A.33})$$

Moreover, it is acting on a maximally entangled ket, which in d dimensions has a 1 every $d + 1$ spaces down its column, and 0s elsewhere. Acting the block matrix on this vector produces a new vector where

the n th column of K_m ($1 \leq n \leq d$) appears in the n th subcolumn of the vector:

$$\sum_i^{d-1} I \otimes K_m |ii\rangle = \begin{bmatrix} [\text{col}_1] \\ [\text{col}_2] \\ \vdots \\ [\text{col}_d] \end{bmatrix} \quad (\text{A.34})$$

so that the columns are stacked. This is exactly the process of vectorization. With this, the Choi matrix takes on a simple form:

$$J(\Phi) = \sum_m \text{vec}(K_m) \text{vec}(K_m)^\dagger \quad (\text{A.35})$$

This makes explicit that the Choi matrix is Hermitian (which also obviously follows from the fact that it is a density matrix). So, to get the Choi matrix from the Kraus operators, we just compute the above sum over outer products. To get the Kraus operators from the Choi matrix, we can use the spectral theorem to decompose the Hermitian $J(\Phi)$:

$$J(\Phi) = \sum_m \lambda_m |m\rangle\langle m| \quad (\text{A.36})$$

The Kraus operator K_m , then, is just the un-vectorized form of $|m\rangle$, and the number of Kraus operators is the rank of the spectral decomposition. The Kraus operators found in this way are the “canonical” Kraus operators. Since the rank of a matrix can never be larger than the space in which its columns live, this shows why one never needs more than d^2 Kraus operators to describe a process.

We just saw how to obtain a set of canonical Kraus operators from the Choi matrix, but the construction of the Choi matrix itself may have seemed arbitrary. In fact, there is a very intuitive way of seeing $J(\Phi)$. First, we should express Φ in terms of its process matrix

$$\Phi(\rho) = \sum \chi_{ij} P_i \rho P_j \quad (\text{A.37})$$

where the operator basis P is chosen to be the Pauli basis. Plugging in, we get

$$J(\Phi) = \sum_{i,j=0}^{d-1} |i\rangle\langle j| \otimes \left(\sum_{mn} \chi_{mn} P_m |i\rangle\langle j| P_n \right) \quad (\text{A.38})$$

$$= \sum_{m,n} \chi_{mn} \left(\sum_{i,j=0}^{d-1} |i\rangle\langle j| \otimes P_m |i\rangle\langle j| P_n \right) \quad (\text{A.39})$$

$$= \sum_{m,n} \chi_{mn} \left(I \otimes P_m |\omega\rangle\langle\omega| \right) \left(\langle\omega| I \otimes P_n \right) \quad (\text{A.40})$$

$$= \sum_{m,n} \chi_{mn} |B_m\rangle\langle B_n| \quad (\text{A.41})$$

where we used the fact that applying each of the four single qubit Paulis $I \otimes P_i$ to the maximally entangled Bell state $|\omega\rangle$ produces the four Bell basis states. Thus, the Choi matrix is the process matrix after it has been rotated into the Bell basis by the Bell preparation unitary $U = H_1 \text{CNOT}_{12}$.

A.4. Pauli Transfer Matrix

The Pauli transfer matrix R_Λ just says what happens if you pass a Pauli matrix through some process, and then project it back into the basis of Pauli matrices:

$$\langle\langle j | R_\Lambda | k \rangle\rangle = \frac{1}{d} \text{Tr}(P_j \Lambda(P_k)) \quad (\text{A.42})$$

Trace-preservation holds if the first row is $[1, 0, 0, 0]$, and CP holds if the corresponding Choi matrix is positive semidefinite [26]. The PTM can be expressed in terms of the process matrix by just plugging in for Λ :

$$\langle\langle j | R_\Lambda | k \rangle\rangle = \langle\langle j | \sum_{m,n} \chi_{m,n} P_n^* \otimes P_m | k \rangle\rangle = \frac{1}{d} \sum_{m,n} \chi_{mn} \text{Tr}(P_j P_m P_k P_n) \quad (\text{A.43})$$

A.5. Canonical Lindblad Form

The various Markovianity metrics (distance measures, volume measures, entanglement, Choi-matrix, etc.) may all be unified under the following unifying description of non-Markovianity. In the time-dependent master equation, the canonical form is one such that the Lindblad collapse operators $L_k(t)$ form an orthonormal basis set of traceless operators, and in this canonical form, the evolution at a time t is non-Markovian if and only if the canonical decoherence rates are negative. If they are positive, then the evolution is Markovian. These decoherence rates are uniquely determined and are invariant under any unitary transformation. Such a unitary transformation will, however, change the $L_k(t)$ operators, and this is actually how the decoherence rates are obtained. Essentially, one starts with the Kraus form for the time evolution of the system

$$\dot{\rho} = \Phi_t(\rho) = \sum_k A_k(t) \rho B_k^\dagger(t) \quad (\text{A.44})$$

Then we decompose the Kraus operators into a complete set of $N = d^2$ basis operators, which are all Hermitian, orthogonal, and traceless (except for the element G_0 which is assumed to be $I/\sqrt{d}$). This yields the unique decomposition

$$\dot{\rho} = \sum_{i,j=0}^{N-1} c_{ij} G_i \rho G_j \quad (\text{A.45})$$

which is basically the same procedure by which the process matrix is obtained. The decoherence matrix d is defined as the $N-1 \times N-1$ matrix whose elements $d_{ij} = c_{ij}$ for $i, j = 1, 2, \dots, N$. After some manipulation, the equation can be written as

$$\dot{\rho} = C\rho + \rho C^\dagger + \sum_{i,j=1}^{N-1} d_{ij} G_i \rho G_j \quad (\text{A.46})$$

where $C = \frac{c_{00}}{d} + \sum_i \frac{c_{i0}}{\sqrt{d}} G_i$. Already, the canonical form is beginning to emerge, and using some trace constraints, the time-dependent Lindblad form can be obtained. The elements $d_{ij}(t)$ are the rates of the collapse terms, and since the matrix is unitary, we can diagonalize d :

$$d_{ij} = \sum_k U_{ik} \gamma_k U_{jk}^* \quad (\text{A.47})$$

where the eigenvalues γ_k are real but not necessarily positive. The canonical collapse operators are $L_k(t) = \sum_{i=1}^{N-1} U_{ik}(t) G_i$. From the above, we see that it is very important to first rewrite the master equation in canonical form (i.e. unitarily transform the collapse operators) to obtain the correct canonical rates $\gamma_k(t)$.

B

Linear Algebra

B.1. Singular Value Decomposition

The SVD theorem states that any $m \times n$ matrix can be decomposed into the following form:

$$M = U\Sigma V^\dagger \quad (\text{B.1})$$

where U is a unitary $m \times m$ matrix, Σ is a diagonal $m \times n$ matrix, and V is a unitary $n \times n$ matrix. In quantum information, we usually have that the input and output spaces are equal, so that all three matrices are square and have the same size. The rank of a matrix is the number of nonzero singular values in Σ .

B.2. Diagonalization

A matrix is diagonalizable if it is similar to a diagonal matrix:

$$M = PDP^{-1}$$

A matrix M can be diagonalized iff the sum of the dimensions of its eigenspaces is equal to the total dimension of the space in which M lives. This means that the eigenvectors of M form a basis for the space.

B.3. Spectral Theorem

A matrix is *normal* if it commutes with its adjoint

$$MM^\dagger = M^\dagger M \quad (\text{B.2})$$

A quick result is that if M is Hermitian ($M = M^\dagger$), it is automatically normal. Unitary operators ($M^\dagger = M^{-1}$), positive operators ($M = NN^\dagger$), and skew-Hermitian operators ($M^\dagger = -M$) are also normal. If a matrix is normal, then it is diagonalizable via a unitary matrix P :

$$M = PDP^{-1} = PDP^\dagger \quad (\text{B.3})$$

$$= \begin{bmatrix} \vec{p}_1 & \dots & \vec{p}_n \end{bmatrix} D \begin{bmatrix} \vec{p}_1^\dagger \\ \vdots \\ \vec{p}_n^\dagger \end{bmatrix} \quad (\text{B.4})$$

$$= \sum_i^n d_i |p_i\rangle\langle p_i| \quad (\text{B.5})$$

The columns of the matrix P form an orthonormal eigenbasis for the space that M lives in. A Hermitian M has only real d_i . The singular values are the eigenvalues, and the rank of the matrix is the number of nonzero eigenvalues.

B.4. Cholesky Decomposition

Any Hermitian, positive-definite matrix M has a Cholesky decomposition into the product of a lower triangular matrix and its adjoint.

$$M = LL^\dagger \quad (\text{B.6})$$

$$L = \begin{bmatrix} l_{11} & & 0 \\ \vdots & \ddots & \\ l_{n1} & \dots & l_{nn} \end{bmatrix} \quad (\text{B.7})$$

This is very useful for parameterizing density matrices—although they are positive semi-definite, you can always just add a very small multiple of the identity to your matrix so that a solver will find a Cholesky decomposition, with only a negligible change to your data. In MATLAB, the output of `chol(M)` is upper triangular, so the decomposition convention is $M = L^\dagger L$ instead.

B.5. Matrix Norms

B.5.1. Trace Distance

The trace distance (also known as the Schatten $p = 1$ norm) expresses the difference between two states ρ and σ in terms of their distinguishability, which is based on the probability of determining whether one was given ρ or σ based on a single measurement with some optimal POVM $\{P, 1 - P\}$.

$$D(\sigma, \rho) = \frac{1}{2} \|\rho - \sigma\|_1 \quad (\text{B.8})$$

$$= \frac{1}{2} \text{Tr} \left[\sqrt{(\rho - \sigma)^\dagger (\rho - \sigma)} \right] \quad (\text{B.9})$$

$$= \frac{1}{2} \sum_i |\lambda_i| \quad (\text{B.10})$$

$$= \max_P \text{Tr}[P(\rho - \sigma)] \quad (\text{B.11})$$

The third equality, the sum over the absolute value of the eigenvalues of $\rho - \sigma$, follows from the fact that ρ and σ are Hermitian. The success probability of distinguishing the states is

$$p_{\text{succ}} = \frac{1}{2} (1 + D(\sigma, \rho)) \quad (\text{B.12})$$

The trace distance is bounded between 0 and 1. The trace distance between ρ and σ is also half the Euclidean distance between the two Bloch vectors of the states [44].

B.5.2. Spectral Norm

In MATLAB, the `norm` function computes the spectral norm, also known as the induced norm, which uses a vector p -norm to construct an operator norm.

$$\|A\|_p = \sup_{x \neq 0} \frac{\|Ax\|_p}{\|x\|_p} \quad (\text{B.13})$$

where x is some vector, and $\|\cdot\|_p$ is the p -norm

$$\|x\|_p = \left(\sum_i |x_i|^p \right)^{1/p} \quad (\text{B.14})$$

If $p = 2$, the vector norm is simply related to the inner product in Hilbert space

$$\|x\|_2 = \sqrt{\langle x|x \rangle} \quad (\text{B.15})$$

If we want to compute the spectral norm of the difference between two unitary matrices U and V (as we did in section 4.2), we can obtain an upper bound by maximizing over pure states $|x\rangle$:

$$\|U - V\|_2 = \sup_{|x\rangle \neq 0} \frac{\|(U - V)|x\rangle\|_2}{\sqrt{\langle x|x\rangle}} \quad (\text{B.16})$$

$$= \sup_{|x\rangle \neq 0} \sqrt{\langle x|(U^\dagger - V^\dagger)(U - V)|x\rangle} \quad (\text{B.17})$$

$$= \sup_{|x\rangle \neq 0} \sqrt{\langle x|2I - (V^\dagger U + U^\dagger V)|x\rangle} \quad (\text{B.18})$$

$$= \sup_{|x\rangle \neq 0} \sqrt{2 - \langle x|V^\dagger U + U^\dagger V|x\rangle} \quad (\text{B.19})$$

Since we are using the same unitary with two different angles, we can say $U = U(\phi)$, $V = U(\phi + \delta)$, so $V^\dagger U = U(-\delta)$.

$$\langle V^\dagger U + U^\dagger V \rangle = \langle U(-\delta) + U(\delta) \rangle \quad (\text{B.20})$$

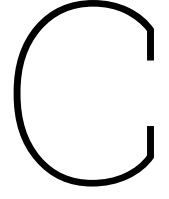
To maximize the norm, we need the variational term to be as negative as possible. Once $|x\rangle$ is rotated, we can get as most -1 from projecting it back onto itself. This extremum could occur, for example, if the unitary is such that $U(\delta)|0\rangle = -|-\rangle$, so that $U(-\delta)|0\rangle = -|+\rangle$. In both cases, the projection back onto $|0\rangle$ yields -1 . The upper bound, over all pure states and rotations $U(\delta)$ is $\sqrt{2 - (-2)} = 2$. This is the bound stated in section 4.2.

B.5.3. Diamond Norm

We have seen that the trace distance describes the probability of successfully distinguishing two quantum states using one (optimal) measurement. Analogously, the diamond distance [1] describes the probability of distinguishing whether channel $\mathcal{E}_1$ or $\mathcal{E}_2$ happened to some state ρ , by finding the input ρ that maximizes the distinguishability of the output states [5]. In general, entanglement can bias this and help our success probability, so we add an ancilla space in which the channels act trivially, and maximize over density matrices in the larger space.

$$\|\mathcal{E}_1 - \mathcal{E}_2\|_\diamond = \max_{\rho} \|(\mathcal{E}_1 \otimes I)[\rho] - (\mathcal{E}_2 \otimes I)[\rho]\|_1 \quad (\text{B.21})$$

with everything vectorized, so that the difference of channels makes sense. The diamond norm is bounded between 0 and 2.



Rotating Frames

Whenever we are given a Hamiltonian that consists of a “bare” term (a sum of uncoupled systems) plus an interaction term (some set of operators that couples the systems), it is useful to obtain an effective Hamiltonian that transforms away some or all of the unitary evolution (which nature put there) to focus on the dynamics from the coupling alone (which we put there).

$$H = H_0 + H_{int} \quad (C.1)$$

If the unitary evolution of H is given by $U(t)$, then recall that the operator itself evolves according to

$$i\hbar \frac{d}{dt} |\psi(t)\rangle = H |\psi(t)\rangle \quad (C.2)$$

$$i\hbar \frac{d}{dt} U(t) |\psi(0)\rangle = H U(t) |\psi(0)\rangle \quad (C.3)$$

$$i\hbar \frac{d}{dt} U(t) = H U(t) \quad (C.4)$$

where $U(t)$ propagates a state forward in time:

$$U(t) = \mathcal{T} \exp \left(-\frac{i}{\hbar} \int_0^t H(t') dt' \right) \quad (C.5)$$

and the time-ordering is necessary because $H(t')$ may not commute with itself at different times [18]. If we imagine the time evolution of some state $R(t) |\psi(t)\rangle$, we can write

$$i\hbar \frac{d}{dt} (R(t) |\psi(t)\rangle) = i\hbar R(t) \frac{d}{dt} |\psi(t)\rangle + i\hbar \left(\frac{d}{dt} R(t) \right) |\psi(t)\rangle \quad (C.6)$$

$$= R(t) H R(t)^\dagger R(t) |\psi(t)\rangle + i\hbar \dot{R}(t) R(t)^\dagger R(t) |\psi(t)\rangle \quad (C.7)$$

$$= (R(t) H R(t)^\dagger + i\hbar \dot{R}(t) R(t)^\dagger) R(t) |\psi(t)\rangle \quad (C.8)$$

$$= H_{\text{eff}}(t) R(t) |\psi(t)\rangle \quad (C.9)$$

where $R(t)$ is some unitary matrix, and $\dot{R}(t)$ denotes its time derivative. So whatever the state $R(t) |\psi(t)\rangle$ is, it evolves according to the effective Hamiltonian H_{eff} . In general, H is time-dependent, but even when it is static, H_{eff} will still be time-dependent. This procedure is known as “moving to a rotating frame”, and whether or not it is useful to us entirely depends on some clever choice for $R(t)$. When $R(t) = U_0(t)^\dagger$, then that means it will exactly cancel the unitary evolution $U_0(t)$ that would normally occur due to H_0 .

$$H_{\text{eff}} = U_0^\dagger(t) H U_0(t) + i\hbar U_0^\dagger(t) \dot{U}_0(t) \quad (C.10)$$

$$(C.11)$$

If we are lucky and H_0 is time-independent, then $U_0 = \exp(-iH_0t/\hbar)$.

$$H_{\text{eff}} = U_0^\dagger(t)(H_0 + H_{\text{int}})U_0(t) + i\hbar U_0^\dagger(t)\dot{U}_0(t) \quad (\text{C.12})$$

$$= U_0^\dagger(t)H_0U_0(t) + U_0^\dagger(t)H_{\text{int}}U_0(t) - H_0 \quad (\text{C.13})$$

$$= U_0^\dagger(t)H_{\text{int}}U_0(t) \quad (\text{C.14})$$

When $R(t) = U_0(t)^\dagger$, the frame we have rotated to is known as the interaction picture, since in this frame states only evolve according to a transformed version of H_{int} .

C.1. Interaction Picture Device Hamiltonian

Recalling the device Hamiltonian from section 3.5.3, it will be easier to view the dynamics if we go to the interaction picture:

$$H_0 = H_1 + H_2 = \omega_1 b_1^\dagger b_1 - \frac{\alpha_1}{2} b_1^\dagger b_1^\dagger b_1 b_1 + \omega_1 b_1^\dagger b_1 - \frac{\alpha_2}{2} b_2^\dagger b_1^\dagger b_2 b_2 \quad (\text{C.15})$$

$$H_{\text{int}} = g(b_1^\dagger b_2 + b_1 b_2^\dagger) \quad (\text{C.16})$$

First, we can rewrite the transmon terms in the qutrit approximation

$$H_{1,2} = H_{\text{trans}} = \omega b^\dagger b - \frac{\alpha}{2} b^\dagger b^\dagger b b \quad (\text{C.17})$$

$$= \omega b^\dagger b - \frac{\alpha}{2} b^\dagger (b b^\dagger - 1) b \quad (\text{C.18})$$

$$= \omega b^\dagger b - \frac{\alpha}{2} ((b^\dagger b)^2 - b^\dagger b) \quad (\text{C.19})$$

$$= (\omega - \frac{\delta}{2}) b^\dagger b + \frac{\delta}{2} (b^\dagger b)^2 \quad (\text{C.20})$$

$$= \begin{bmatrix} 0 & 0 & 0 \\ 0 & \omega & 0 \\ 0 & 0 & 2\omega + \delta \end{bmatrix} \quad (\text{C.21})$$

where the anharmonicity has been redefined as $-\alpha = \delta$ and shifts the harmonic part. The, we just multiply matrices

$$U_0(t) = \exp(-i(H_1 + H_2)t) \quad (\text{C.22})$$

$$= \begin{bmatrix} 1 & 0 & 0 \\ 0 & e^{-i\omega_1 t} & 0 \\ 0 & 0 & e^{-i(2\omega_1 + \delta_1)t} \end{bmatrix} \otimes \begin{bmatrix} 1 & 0 & 0 \\ 0 & e^{-i\omega_2 t} & 0 \\ 0 & 0 & e^{-i(2\omega_2 + \delta_2)t} \end{bmatrix} \quad (\text{C.23})$$

$$H_{\text{eff}} = gU_0^\dagger(t)(b_1^\dagger b_2 + b_1 b_2^\dagger)U_0(t) \quad (\text{C.24})$$

$$= g \begin{bmatrix} 0 & e^{-i\omega_1 t} & 0 \\ 0 & 0 & \sqrt{2}e^{-i(\delta_1 + \omega_1)t} \\ 0 & 0 & 0 \end{bmatrix} \otimes \begin{bmatrix} 0 & 0 & 0 \\ e^{i\omega_2 t} & 0 & 0 \\ 0 & \sqrt{2}e^{i(\delta_2 + \omega_2)t} & 0 \end{bmatrix} \quad (\text{C.25})$$

$$+ g \begin{bmatrix} 0 & 0 & 0 \\ e^{i\omega_1 t} & 0 & 0 \\ 0 & \sqrt{2}e^{i(\delta_1 + \omega_1)t} & 0 \end{bmatrix} \otimes \begin{bmatrix} 0 & e^{-i\omega_2 t} & 0 \\ 0 & 0 & \sqrt{2}e^{-i(\delta_2 + \omega_2)t} \\ 0 & 0 & 0 \end{bmatrix} \quad (\text{C.26})$$

Since H_1 and H_2 act on different Hilbert spaces, they commute with each other and also with system-2 and system-1 type operators, respectively. The terms $b^\dagger b^\dagger$ and bb were preemptively dropped because at this step, their higher frequencies would average to 0 when looking at the dynamics of the slower frequencies. It is cumbersome to simulate this Hamiltonian in QuTiP because the function `mesolve` prefers time-dependent operators of the form $c_i(t)H_i$, where all of the time-dependence is in the coefficient. Thus, the qutrit matrices above must be split into sums of the required form. The qubit case does not encounter this issue.

D

Additional Datasets

Additional POVM datasets are shown here. Beyond the four kinds of plots explained in section 4.2, two more are provided in this appendix. Thus, for the $\sigma = 0.01$ and $\sigma = 0.25$ datasets, there are 6 plots, and for $\sigma = 0.05$, there are two more in addition to what was shown earlier. The extra plots are the direct comparison of the thermal and coherent parameters, between their estimated and actual values. Ideally, these data would follow the line $y = x$, and indeed we observe that to be the case. There are two exceptions, however.

When the coherent error scale is small, the probabilities are relatively unaffected regardless of how good the MLE estimate is. As a result, for $\sigma = 0.01$, there appears to be no correlation between the estimate and its ground-truth. Moving up to $\sigma = 0.05$, this correlation has mostly recovered, and by $\sigma = 0.25$, we find very close agreement. Systematically observing such trends over different error scales provides an empirical way of gauging the performance and sensitivity of the MLE routine.

The second exception is present in all σ datasets, and manifests as a small overestimation of the first thermal parameter a , at the expense of slightly underestimating b and c , to preserve the trace of ρ . The reason for this bias is unknown, but ultimately, we only care about the reconstructed probabilities, so to that end, this bias apparently has little effect.

D.1. Coherent Error Scale = 0.01

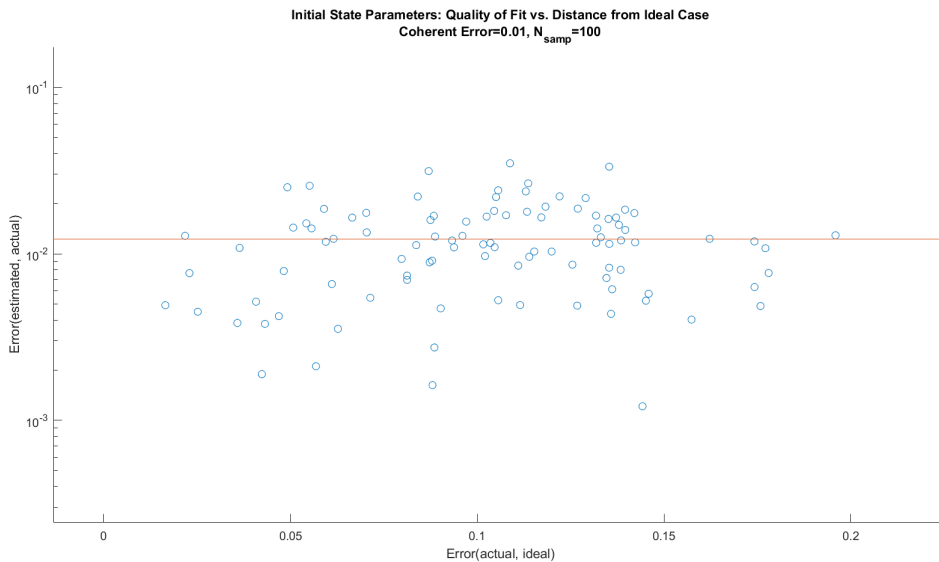


Figure D.1: The y-axis is the trace distance between the estimated and actual initial states, and the x-axis is the trace distance between the actual and ideal initial states. Symbolically, this is $D(\rho', \rho)$ vs. $D(\rho, |0\rangle\langle 0|)$

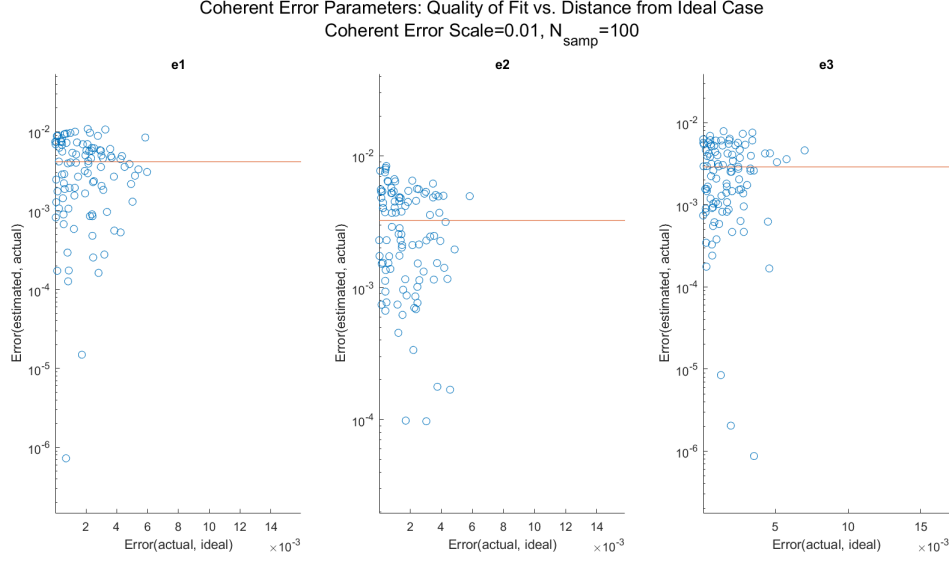


Figure D.2: Each subplot shows the estimation error for one of the three coherent parameters ϵ . The y-axis is the spectral norm of the difference between the estimated and actual rotations, and the x-axis is the spectral norm of the difference between the actual and ideal rotations. Symbolically, this is $\|U(\epsilon') - U(\epsilon)\|_2$ vs. $\|U(\epsilon) - U(0)\|_2$.

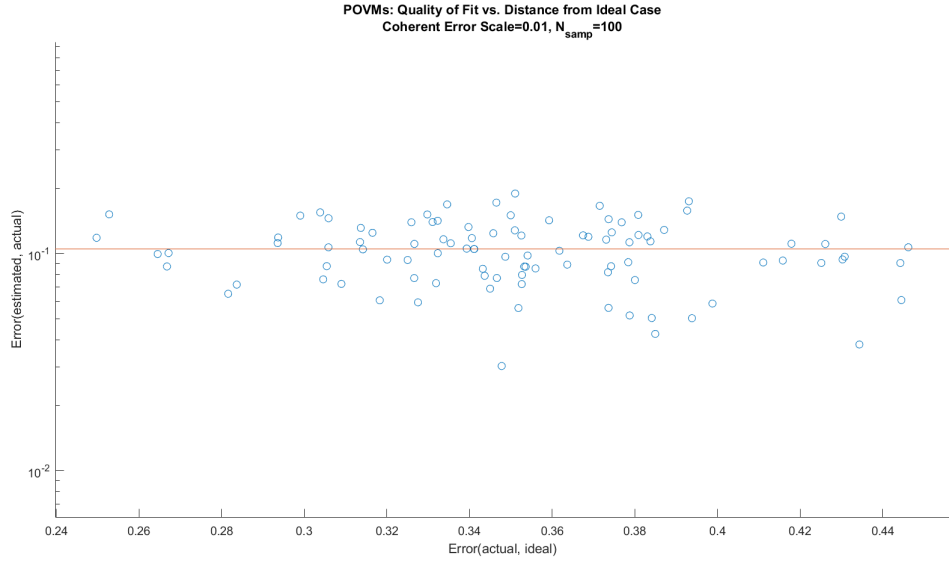


Figure D.3: With the POVMs converted to quantum channels, the y-axis is the diamond norm between the estimated and actual POVM, and the x-axis is the diamond norm between the actual POVM and ideal PVM. Symbolically, this is $\|\mathcal{E}\{\sqrt{E_i'}\} - \mathcal{E}\{\sqrt{E_i}\}\|_\diamond$ vs. $\|\mathcal{E}\{\sqrt{E_i}\} - \mathcal{E}\{|i\rangle\langle i|\}\|_\diamond$.

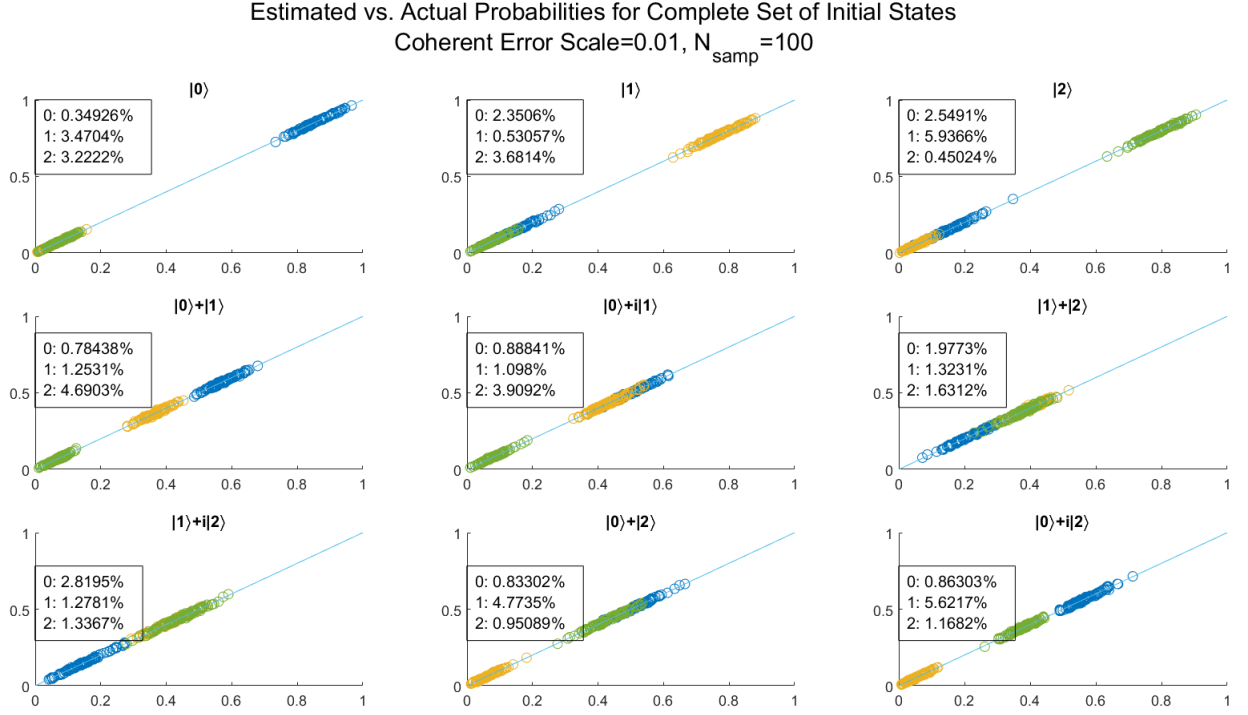


Figure D.4: Probabilities for each measurement outcome (0/1/2) for each initial qutrit state, MLE output (y) vs. ground truth (x). Thus, we are plotting $\text{Tr}(\hat{M}_i' \hat{\rho}_j')$ vs. $\text{Tr}(\hat{M}_i \hat{\rho}_j)$, where i indexes the color (0:blue, 1:yellow, 2:green), and j the subplot. These results would be sufficient to perform measurement tomography.

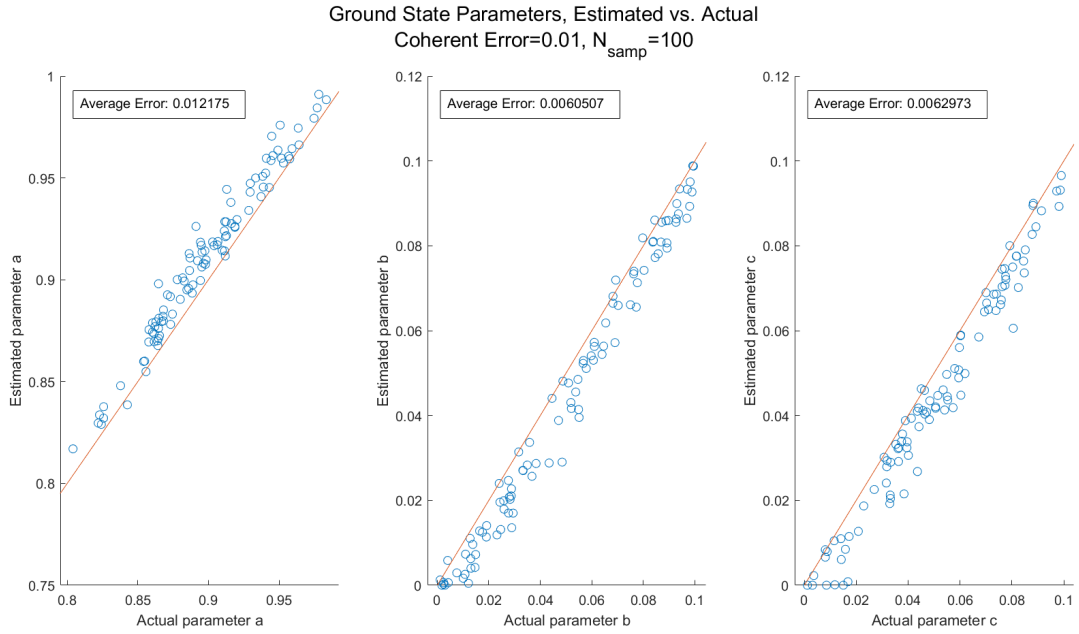


Figure D.5: Direct comparison of estimated and actual thermal parameters.

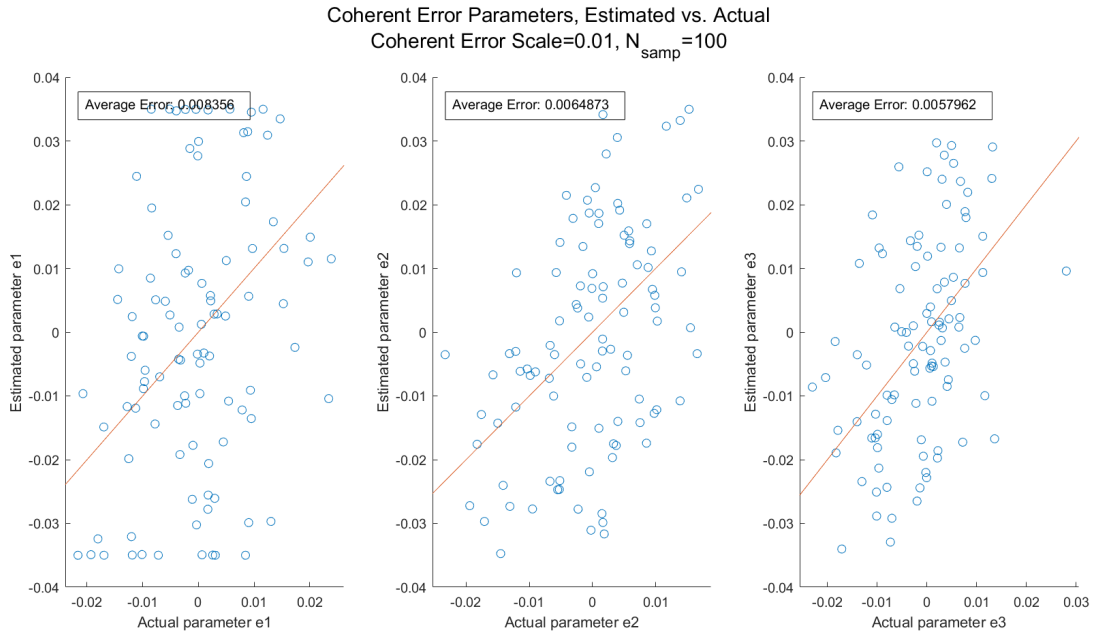


Figure D.6: Direct comparison of estimated and actual coherent error parameters.

D.2. Coherent Error Scale = 0.05

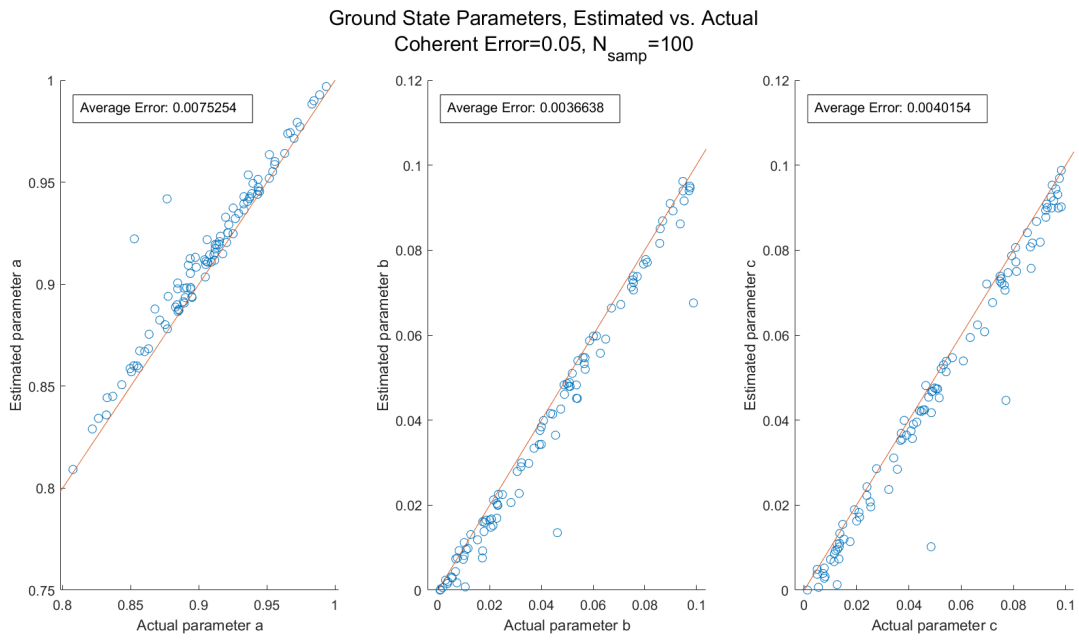


Figure D.7: Direct comparison of estimated and actual thermal parameters.

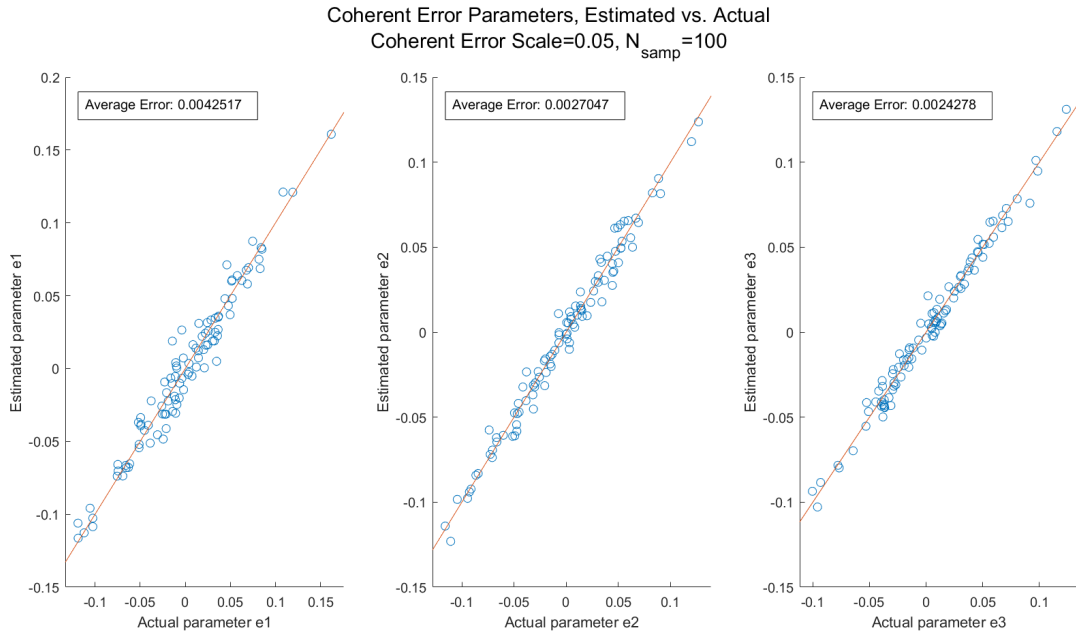


Figure D.8: Direct comparison of estimated and actual coherent error parameters.

D.3. Coherent Error Scale = 0.25

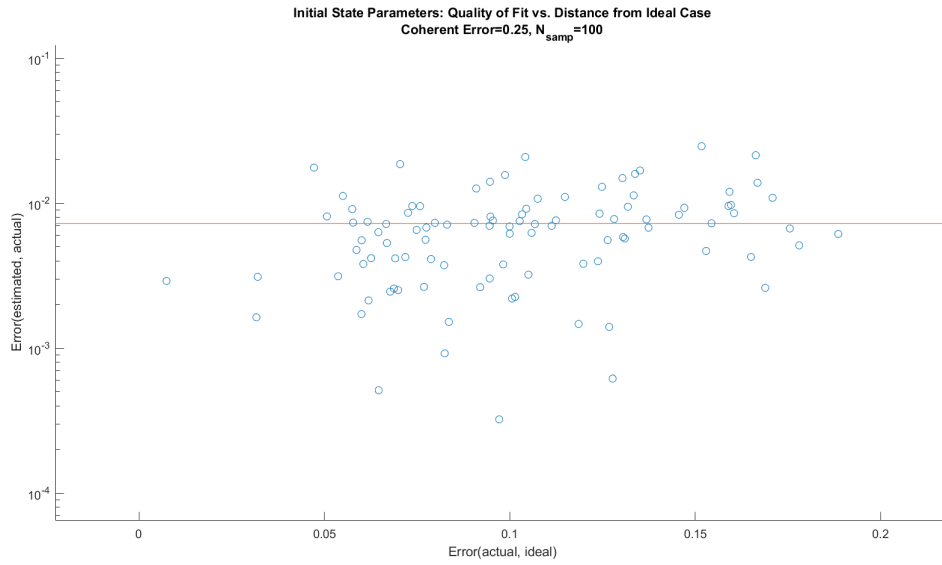


Figure D.9: The y-axis is the trace distance between the estimated and actual initial states, and the x-axis is the trace distance between the actual and ideal initial states. Symbolically, this is $D(\rho', \rho)$ vs. $D(\rho, |0\rangle\langle 0|)$

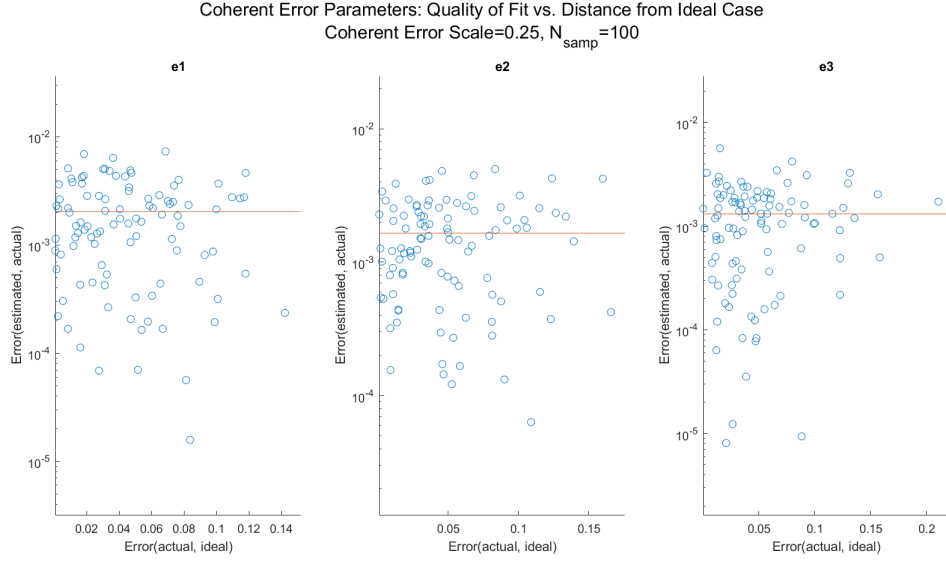


Figure D.10: Each subplot shows the estimation error for one of the three coherent parameters ϵ . The y-axis is the spectral norm of the difference between the estimated and actual rotations, and the x-axis is the spectral norm of the difference between the actual and ideal rotations. Symbolically, this is $\|U(\epsilon') - U(\epsilon)\|_2$ vs. $\|U(\epsilon) - U(0)\|_2$.

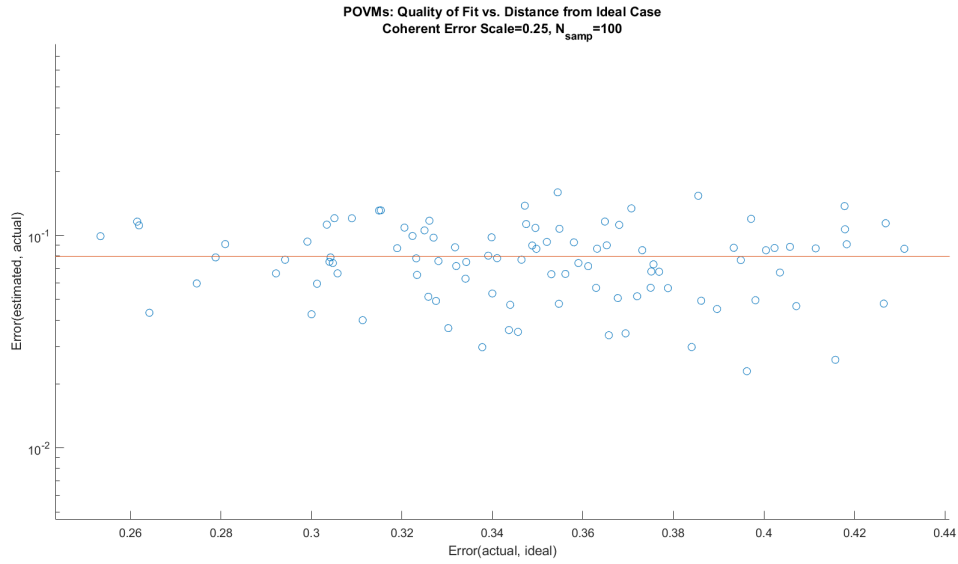


Figure D.11: With the POVMs converted to quantum channels, the y-axis is the diamond norm between the estimated and actual POVM, and the x-axis is the diamond norm between the actual POVM and ideal PVM. Symbolically, this is $\|\mathcal{E}\{\sqrt{E_i'}\} - \mathcal{E}\{\sqrt{E_i}\}\|_\diamond$ vs. $\|\mathcal{E}\{\sqrt{E_i}\} - \mathcal{E}\{|i\rangle\langle i|\}\|_\diamond$.

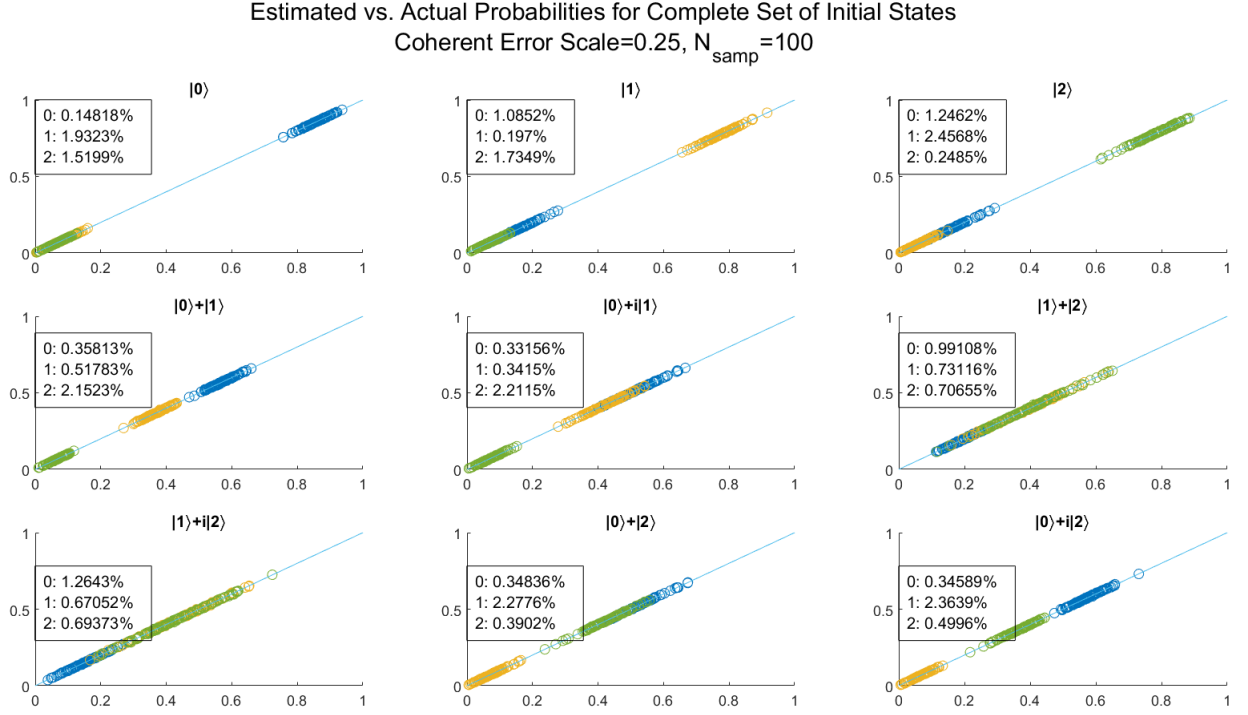


Figure D.12: Probabilities for each measurement outcome (0/1/2) for each initial qutrit state, MLE output (y) vs. ground truth (x). Thus, we are plotting $\text{Tr}(\hat{M}_i' \hat{\rho}_j')$ vs. $\text{Tr}(\hat{M}_i \hat{\rho}_j)$, where i indexes the color (0:blue, 1:yellow, 2:green), and j the subplot. These results would be sufficient to perform measurement tomography.

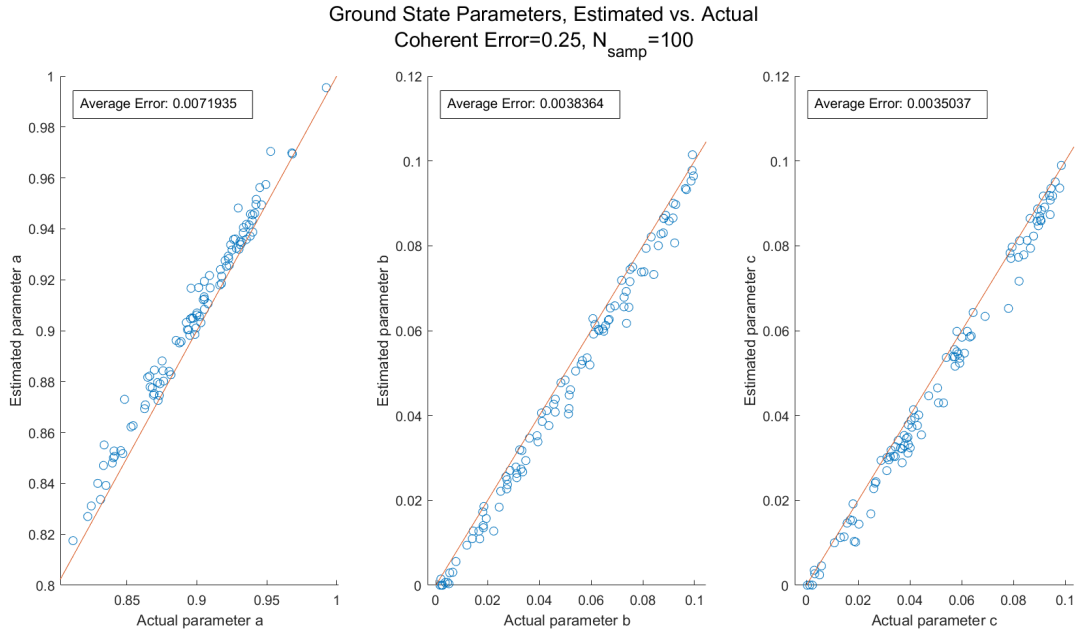


Figure D.13: Direct comparison of estimated and actual thermal parameters.

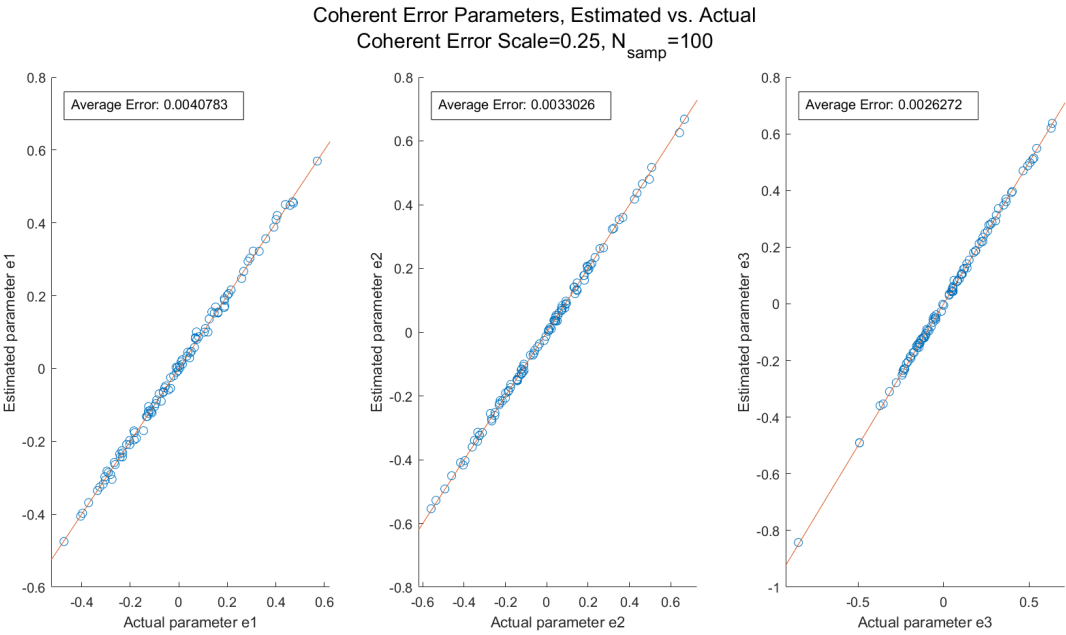


Figure D.14: Direct comparison of estimated and actual coherent error parameters.