

DiSuQ - Discovering Superconducting Quantum Circuits

Master Thesis



Author: Shahrukh Chishti (Student ID: 7385095)

Supervisor: Univ.-Prof. Dr. Tommaso Calarco

Co-Supervisor: Dr. Felix Motzoi

Department of Physics

University of Cologne

March 12, 2025

Abstract

Superconducting circuits are practical infrastructure to host quantum information technology.

In this thesis, superconducting circuits are studied as quantum system with several degrees of freedom. Various numerical methods are assessed to develop a computational framework for efficient simulation and optimization on these quantum systems.

In particular, a graph theoretic framework that describe general superconducting circuits. It is combined with automatic differentiation to optimize parameters within a given circuit topology.

This methodology is applied to rederive and/or rediscover optimal parameter regime of known circuits. Examples include popular superconducting qubits of Transmon, shunted-Flux, Zero-Pi. Additionally, an entirely new qubit "Prismmon" is proposed, which may have enhanced operating qualities compared to its predecessors.

Finally, a critical problem in the Hamiltonian representation of this methodology is identified. It limits numerical study of quantum systems, and guide its research.

Contents

1	Introduction	1
1.1	Objective	2
1.2	Result & Outcome	3
1.3	Outline	3
2	Theory	5
2.1	Superconducting Circuits	6
2.1.1	Harmonic Oscillator	6
2.1.2	Superconductivity	7
2.1.3	Flux Quantization	8
2.2	Circuit Modelling	9
2.2.1	Lumped Circuit	9
2.2.2	Lagrangian-Hamiltonian formulation	11
2.2.3	Node-Branch formalism	12
2.2.4	Spanning Tree	12
2.2.5	Quantization	13
2.2.6	Modal Transform	15
2.3	Diagonalization	16
2.4	Optimization	18
2.4.1	Computation Graph	19
2.4.2	Backpropagation	20
3	Analysis	23
3.1	Asymptotic Analysis	23
3.1.1	Transmon - Charge Insensitivity	24
3.2	Basis Representation	28
3.2.1	Single Mode Example	29
3.2.2	Dominant Operator basis	30
3.3	Middle Region	32
4	Examples	34
4.1	C-shunted profiling	35
4.1.1	Spectrum profiling problem	35
4.1.2	Multiple Solutions	37
4.1.3	Effective optimizer	39
4.2	Zero-Pi degeneracy	40
4.2.1	Robustness model	42
4.2.2	Degeneracy objective	43

4.2.3	Optimizer comparison	44
4.2.4	Optimal Solution	46
4.3	Prismon symmetry	47
4.3.1	Minima Minimization	50
4.3.2	Collapse configuration	51
4.3.3	Results	52
5	Conclusion	55
A	Newtonian Optimization	56
A.1	BFGS	56
A.2	Error Accumulation	57
B	Spanning Tree - Gauge invariance	58
C	Eigenvalue derivates	59
D	Benchmarks	59
D.1	Framework	59
D.2	Computational	60
E	Circuit Bounds	60
F	Shunted-Qubit profile : Optimization	61
	References	64

1 Introduction

Computation procedure, deriving definition from its roots in logic system, is manipulation of symbols to achieve a solution.

The idea is to arrange a set of rules allowed by the logic, into more complex blocks. Subsequently, these modules are combined through chains, hierarchies and recursions etc., acting step by step. The final assembly of operations is termed as algorithm.

Quantum computation, simply put, is a computation model utilising dynamics of quantum realm. This physics is consistent and reproducible on statistical averages. Superconducting circuits are nano-sized circuits that demonstrate quantum computing at scales higher than any other manifestations. Qubits with high coherence times provide longer computation session for quantum codes. As well as, robust qubits benefits error correcting codes, significantly. For example, surface code distances defining the size of a logical qubit could be lowered with robust physical qubits [1]. Superconducting quantum circuits capable of sustaining coherent states are described as artificial atoms. Having defined the energy state to computational state correspondence, a memory qubit is manifested on these *atoms*.

Besides, the potential utility of such superconducting circuits is diverse. For example, in quantum metrology, superconducting sensors could measure structural fractures and brain signals, or in fast signal processing or magnetic confinement etc. In principle, they inherit all the practical applications of superconductivity [2].

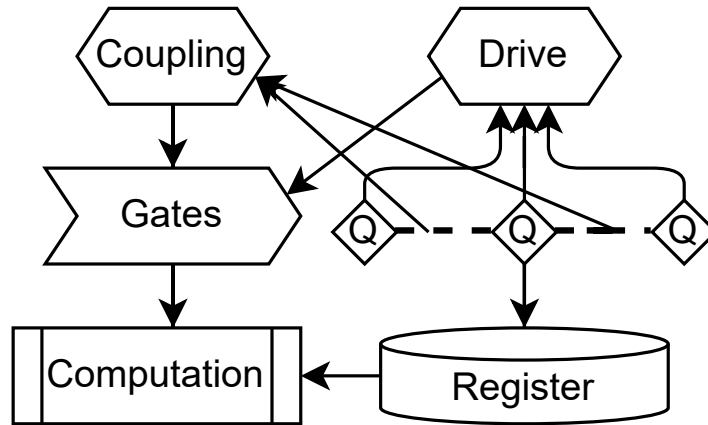


Figure 1: Qubit(Q) operation architecture

Memory circuits maintain a couple of particular states in a sustainable way. Corresponding information state is assigned to those energy states. Oscillators are great candidate to sustain energy content, as their states are bound.

Fine-tuning coupling circuit is an important aspect of large scale circuit. Consider, a circuit with many oscillators(harmonic or anharmonic), coupled to each other locally, but in repetitive manner. Such arrangement are ubiquitous in surface codes etc. In these networks, chaotic fluctuations could emerge from the collective behaviour of oscillators. As is known from bifurcation theory of non-linear systems, these phase transition are sensitive on parameters.

1.1 Objective

The main goal of this master thesis, is to formulate superconducting circuits and build a numerical and computational techniques to solve various design features. Subsequently, it is utilized to study various circuits as case examples, as well as, to discover their optimal designs.

A circuit is a network of electric/superconducting elements connected to each other in various patterns. Further, each element needs to be specified with respective parameter.

In prevalent literature, discovering superconducting circuits implies searching for novel designs and parameters of its elements that demonstrate defined requirement fig.2.

- Ansatz : Intuitive design suggested by a reduced model
- Full connected : All possible connections are parameterized with circuit components [3]
- Population Swarm : Generating huge number of candidates and subsequent filtering [4]

In this work, discovery implies optimization of parameter values of all elements of a fixed pre-defined network. Thus, the set of all parameters describe the space to search for a circuit, that satisfies requirements optimally.

Despite diverse utilisation of superconductors, this framework is centred on designing quantum computing processors. However, in principle it could be extended to other scenarios, with little difficulty.

From a mathematical perspective, this framework is divided into three separate problems :

- Network Modeling
- Diagonalization
- Optimization

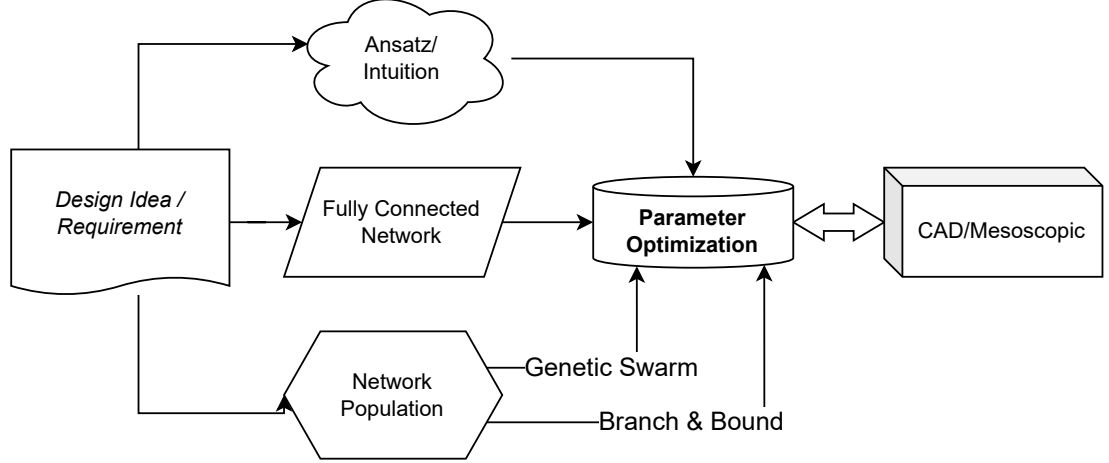


Figure 2: Routes to design superconducting circuits.

Superconducting circuits could also be designed for analog quantum computers or hybrid digital-analog[5].

In this thesis, "experiments" is interchangeably used for "optimization". The proposition of example problems certainly makes them equivalent.

The scope of this work is also related to the search of quantum materials[6], as superconducting circuits are finite degree of freedom quantum materials. These problems belong to the same class of algorithmic complexity, as their underlying mathematics is same.

1.2 Result & Outcome

The fundamental proposition of this thesis is demonstration of symbolic back-propagation on eigendecomposition of circuit Hamiltonian. It is demonstrated to speed up the calculations. To our knowledge, it is the first demonstration with backpropagation algorithm. These methods are compiled in a simulation-optimization library for superconducting circuits, *DiSuQ*.

In addition, novel methods to discover physics in superconducting circuits are presented successfully with prevalent examples.

The shunted-Flux qubit is forced to a given spectrum, Zero-Pi is explored for degeneracy from purely numerical requirement. A proposed, and most complicated of them all, Prismon is also identified for degeneracy.

Simulations related to appropriate representation describe a technical flaw.

1.3 Outline

Majority of the work in this project, is related to development, improvement and demonstration of computational/mathematical techniques. It is a culmination

of existing formulation, programming and algorithms to study superconducting circuits.

Section(2) discuss underlying physics of superconducting currents and circuit analysis. It describe methods of formulation that convert circuit into numbers. Further, circuit discovery is represented as an optimization problem. Various algorithms, constraints & objective functions of optimization, mostly in relation to qubit design are discussed.

Section(3) present application of the framework for a variety of simulation and optimization tasks. Examples review several known results from established research, and draws comparison between numerical & analytic results. This demonstrate the proposition of utilizing gradient-based methods for circuit optimization. Problems related to improvement in numerical simulation of quantum Hamiltonians is identified and discussed. Finally, it reveals an essential shortcoming in existing methodology to pursue quantum computing, that is open-ended.

2 Theory

Classical electric circuits have been revolutionizing technology for the last 100 years. They are utilized as functional transforms based on input-output paradigm, such as transformers in power transmission, photo-multipliers in metrology and in signal processing.

A state of a classical circuit is described by the measurement of currents flowing through the branch and voltages at the nodes with respect to the ground node. In order to describe a superconducting circuit in a computable form, it first need a mathematical description. These circuits are a network of quantized degree of freedoms, and their description is equivalent to any other finite sized quantum system, for example coupled oscillators, nucleon-nucleon interaction etc.

The dynamics of superconducting circuits is at the interplay of classical electric circuit principles, and additional mathematical structure from quantum mechanics and constraints from superconductivity. A long standing objective in the study of superconducting circuits is to develop analytic theorems for simplified interpretation.

The physics of superconducting circuits described here, builds the core of *DiSuQ*. However, at different stages there are several different ways to formulate a circuit Hamiltonian.

For example, conditions such as, parameter regime for circuit elements and choice of basis, mode representation etc. In sec.3.2, these separately defined condition are explained to be inter-related.

Despite, the purpose of these different approaches is to render accurate results for vast diversity of circuits. Each stage of making a choice brings disparity on class of problems to be solved efficiently. Formulation methodologies is discussed in each subsection, and design principles of *DiSuQ* are asserted.

The numerical analysis of circuit based on quantized degrees of freedom requires Hamiltonian description. In this thesis, scope of quantum mechanics is restricted to time-independed Schrodinger equation.

This chapter is divided into four sections, that should be separate subjects in themselves. It begins with the possiblity of macroscopic coherent states enabled by superconductivity on electric circuits. A general circuit network based on superconducting electric currents is formulated. This formulation renders a Hamiltonian, that contains all the physics of circuit. Diagonalizing this Hamiltonian solves the circuit dynamics. Finally, objective functions defined on the resulting states & spectrum achieves for the discovery of an optimal circuit.

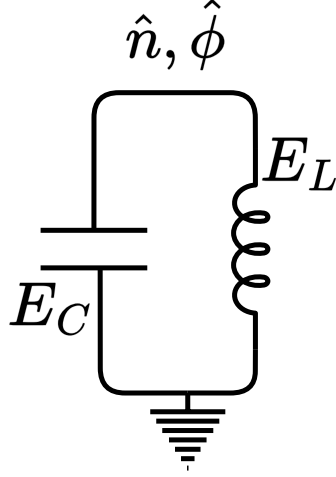


Figure 3: Quantized energy levels and wavefunction in charge basis. Oscillator circuit in n, ϕ phase space

2.1 Superconducting Circuits

2.1.1 Harmonic Oscillator

Harmonic oscillator are the simplest models available to understand dynamics of free particle with a quadratic potential. It can be shown that any complicated potential $U(x)$, can locally $x = a$, be approximated as a second degree polynomial; $U(x) = U(a) + U'(a)(x-a) + U''(a)(x-a)^2 + \mathcal{O}(2)$, usually upto necessary accuracy. In this way, harmonic oscillator can describe locally trapped particles. Tunneling suppressed locally trapped wavelet would be its quantum analogue.

In electric circuits, a harmonic oscillator comprises of an ideal capacitor (C) and inductor (L) as depicted in . It is a single node system, with respect to the ground. The energy of the system oscillates between capacitor in the form of charge deposition at plates and the magnetic flux of the coil.

The ground node acts as a reference frame, that assigns absoluteness to voltage measurement on a node.

It is suggestive to treat electrostatic capacitance energy as kinetic (T_C) and the magnetic inductance as potential (U_L) energy.

Consider the voltage at the node as $V(t)$, and its time integral as fluxoid variable $\Phi(t)$. The Lagrangian of this system is derived using characteristic current-voltage relations

$$\begin{aligned} \Phi(t) &= \int V(t) dt \quad V = L \frac{dI}{dt}, \quad I = C \frac{dV}{dt} \\ T_C &= \frac{1}{2} C \dot{\Phi}^2 \quad U_L = \frac{1}{2L} \Phi^2 \\ \mathcal{L} &= T_C - U_L = \frac{1}{2} C \dot{\Phi}^2 + \frac{1}{2L} \Phi^2 \end{aligned}$$

The variable conjugate to fluxoid $\frac{d\mathcal{L}}{d\Phi} = C\dot{\Phi}$ is Q , the charge on capacitors. Thereby with mass equivalent to capacitance, the kinetic and potential treatment is justified. Effectively it is a conjugate pair $(Q, \Phi) \equiv (p, q)$.

Under Legendre transformation, oscillator Hamiltonian describes total energy in phase-space (Q, Φ)

$$H = Q\dot{\Phi} - \mathcal{L} = \frac{Q^2}{2C} + \frac{\Phi^2}{2L} \quad (1)$$

Aforementioned Hamiltonian(and conjugate pair) derivation is a direct consequence of Lagrangian(fluxoid) description. And could be extended to any general circuit satisfying all conservation laws(Kirchoff, Flux Quantization), by balancing separate description on each node.

Having defined a Hamiltonian of the system, classical equation of motion could be derived as $\frac{dq}{dt} = \frac{dH}{dp}$, $\frac{dp}{dt} = -\frac{dH}{dq}$.

Extending the Hamiltonian(eq.1) to a quantized phase-space needs restructuring of circuit variables as operators($\hat{Q}, \hat{\Phi}$), that acts on a Hilbert space(\mathcal{H}). This is a space of all possible quantized states the system could achieve, and the transition are fundamentally discrete. Further, the uncertainty principle on the conjugate pair requires $[Q, \Phi] = i\hbar$. This completes a Quantum Harmonic Oscillator(QHO).

To note, QHO dynamics involve quantized current states and not necessarily superconducting. There is no additional constraint from superconductance.

To be accurate, the Hilbert space in any numerical treatment is finite $\mathcal{H}^{n_{trunc}}$, unlike the postulate of QM. Therefore, the uncertainty principle is modified with the truncation number(n_{trunc})[7]. Convergence is achieved in asymptotic limit $\tilde{n}_{trunc} \rightarrow \infty$.

2.1.2 Superconductivity

Superconductivity was discovered in 1911, describing a set of associated electromagnetic phenomena, such as zero resistance and magnetic expulsion. Since then there have been a variety of superconductivity discovered, based on temperature, phase transition, materials etc.

Quantum field theory(QFT) gives an analytic description of superconductivity (type-I), through BCS-BEC theory. A pair of opposite-spin, anti-moving electrons are weakly attracted to each other, via s-wave scattering. At low enough temperature, this local interaction intensifies and electron-pair combines, called Cooper pairing(BCS). Cooper pairs are bosonic in nature, and can occupy a macroscopic state called Bose-Einstein condensate(BEC). Ideally, this is a gapped ground state $\rho \exp\{i\theta\}$ and displays superfluidity resulting zero resistance.

However, Cooper pairs have a charge of $-2e$ and interact with electromagnetic

field(EM) in their dynamics. At this point, the gauge invariant EM field interact with BEC wavefunction phase(θ), via Anderson-Higgs mechanism [8].

The equation of motion derived from action minimization of BEC+EM model results in London's equation :

- Meissner Effect : Expulsion of magnetic field \vec{B} , from the bulk

$$\left(\frac{n_s}{m} - \frac{1}{2}\nabla^2\right)\vec{B} = 0$$

- Persistent Current : Surface currents \vec{j} that balances external flux

$$\nabla \times \vec{j} = \frac{n_s}{m}\vec{B} \quad (2)$$

, n_e is superfluid density of BEC, m is mass of electron-pair. The net magnetic field decays exponentially inside a superconductor and similarly the superconducting current resides only on the surface($\sim 16nm$).

In modern implementation of quantum computers, Aluminum based Type-I superconducting wires are utilized. They have a superconducting gap and critical temperature of 1.2K [9]. It remains an open question on demonstration of high-temperature superconducting circuits [10][11].

2.1.3 Flux Quantization

Consider a closed curve (γ) of a superconducting ring(??, whose condensate wavefunction $\Psi = \sqrt{\rho}\exp\{i\theta\}$. The current \vec{j} through a superconductor in response(linear) to external field \vec{A} , is described by minimal coupling.

$$\begin{aligned} \vec{j} &= \frac{1}{2m_e} \left[\Psi^*(-i\hbar\nabla)\Psi - \Psi(-i\hbar\nabla)\Psi^* - 2e\vec{A}|\Psi|^2 \right] \\ \Rightarrow \vec{j} &= \frac{\hbar}{2m_e} \left(\nabla\theta - \frac{2e}{\hbar}\vec{A} \right) \rho \end{aligned}$$

In the bulk of superconductor current density is negligible(eq.2). Integrating \vec{j} with a line element $d\vec{l}$ over the curve,

$$0 = \oint \vec{j} \cdot d\vec{l} \Rightarrow \oint \nabla\theta d\vec{l} = -\frac{2e}{\hbar} \oint \vec{A} \cdot d\vec{l}$$

under Stokes theorem($d\vec{l} \rightarrow d\vec{S}$), and by definition of vector potential $\nabla \times \vec{A} = \vec{B}$ results,

$$\Rightarrow \Delta\theta = -\frac{2e}{\hbar} \oint \vec{B} \cdot d\vec{S} \Rightarrow 2\pi n = \frac{2e}{\hbar} \Phi_\gamma = \Phi_\gamma/\Phi_0 \quad (3)$$

Since, $\Delta\theta$ is quantized in $2\pi n$, $n \in \mathbb{Z}$ to have a singular wavefunction, it conditions a topological quantization on total magnetic flux Φ_γ through the loop. $\Phi_0 = h/e2$ is the magnetic flux quantum, that is a physical normalizing constant.

Constraint equation(eq.3) is of purely superconductor physics origin. The direct relation between magnetic flux across the loop and the phase of wavefunction, is an evidence of Anderson-Higgs mechanism.

2.2 Circuit Modelling

The goal to discover superconducting circuits is ultimately related to its simulation by numerical methods. In any case, the identification and verification of novel features is as good as its proximity to reality.

In this section, the approach to quantization of the harmonic oscillator is extended to general multi-node superconducting circuits. In addition, non-linear Josephson junction and loop-flux quantization is included.

A multi-node circuit is a network of nodes represented as a graph. The branches/edges of this network contains one of the several electric-magnetic components, each with its special current-voltage characteristic.

This picture is a layout of an actual physical circuit. Specially assigned ground is a reference point for voltage measurement on any node $\Delta V = V_{node} - V_{ground}$.

The method of circuit modeling discussed in this section is numerically implemented in *DiSuQ*. It is subsequently, utilized in all the simulation and optimization related to this project.

Hamiltonian formulation is, first identifying a set of degrees of freedom of a given electric network. Their collective dynamics should describe all the physics of the overall circuit. Secondly, the network connection in conjugation with the components on them should be exploited for chances of simplification of the model. The transformed degree of freedom define the state of a circuit, and describe the energy of each component.

2.2.1 Lumped Circuit

Physically, a superconducting circuit is micro/nanometer network of wires, cavities, junctions, connectors etc. Depending on the shape, arrangement, size and composition of these physical, an electric component is physically manifested. The design characteristic specify the characteristic parameters of the components.

In principle, the dynamics of current and voltages at different points describe the complete picture of the circuit. However, these solid state devices contain macroscopic amount degree of freedom. That is, each electron/Cooper-pair/impurities contribute to the development of EM field, the potential energy

and finally to the overall wavefunction.

Simulation on this level is computationally challenging, though FEM methods are applicable [12]. And such a large scale simulation is a drawback for fast-prototyping. This physical piece of specially crafted integrated chip is reduced to a Lumped model.

It is an encapsulation of a component's physics into a unit with overall EM fields contained within, such that current and voltage across it specify the aggregate potential energy. This reduction neglects local contribution based on exact geometry. These minor effects are spurious non-local interaction, sharp edges, inhomogeneity etc.

As well, noise induced by thermal coupling, charge and flux fluctuations are similarly included with a macroscopic model [13].

Reduced circuit have nodes that could be measured for a voltage with respect to a ground, and branches that contain a lumped component, that allows charge in response to the nodes across.

Besides electric components, control mechanism of a circuit is of importance. Magnetic flux through a loop Φ_{ext}^{loop} and a bias voltage V_{node}^{bias} at a node are external knobs to a superconducting circuit.

For the clarity of description, loop flux is threaded to a minimally participating inductor branch. It is practical to assume to a finite loop-level inductance, however at the cost of additional node.

In the examples sec.4, peripherals attached to a qubit is avoided. They usually include readout resonator, activation pulse, cavity photonics, node bias voltage, source, grounding etc. Supporting architecture is necessary for the functioning of a qubit, however, it is separable from qubit dynamics. So, only a lumped circuit model on the states of a qubit is analysed.

In some analytical formalism, ground node is assigned a degree of freedom as per convenience. Subsequent transformation as in sec.2.2.6, usually discard it.

In this work, the state of a multi-node circuit is defined by a set of variables $\{\Phi_b\}$. As before sec.2.1.1, for each branch b , a position-type variable $\Phi_b = \int V_b(t)dt$. Combined with its derivative it sets up the phase space $\{(\dot{\Phi}_b, \Phi_b)\}$ for Lagrangian formulation. Fixing the ground, the circuit has N_n nodes and N_b branches.

Current characteristics and the energy content of the lumped components are defined with branch current I and voltage V , as:

- Inductance(L) : $V = L\dot{I}$, $\mathcal{U}_L = \frac{1}{2}LI^2 \rightarrow \frac{\Phi^2}{2L}$
- Junction(E)[14] : $V = I_c \sin 2\pi\Phi/\Phi_0$, $\mathcal{U}_J = E_J(1 - \cos 2\pi\Phi/\Phi_0)$
- Capacitance(C) : $V = Q/C$, $\mathcal{T}_C = \frac{1}{2}CV^2 \rightarrow \frac{C\dot{\Phi}^2}{2}$

Inverse capacitance is equivalent to the mass of a particle. And inverse inductance is similar to the strength of a linear restoring force.

Node bias voltage is a constant additional voltage applied to a node by external mechanism.

2.2.2 Lagrangian-Hamiltonian formulation

In the branch description, each branch b is an individual degree of freedom represented by $(\dot{\Phi}_b, \Phi_b)$. The Lagrangian include all the energies contained in different components of the circuit, as $\mathcal{L} = \mathcal{T} - \mathcal{U}$

$$\mathcal{L} = \sum_b C_b \dot{\Phi}_b^2 / 2 - \Phi_b^2 / 2L_b - E_J \cos \Phi_b \quad (4)$$

The Hamiltonian in branch formalism is derived via Legendre transformation of the Lagrangian:

$$\frac{\partial \mathcal{L}}{\partial \dot{\Phi}_b} = C_b \dot{\Phi}_b \equiv Q_b \quad \Rightarrow \quad H = \sum_b Q_b \dot{\Phi}_b - \mathcal{L} \quad (5)$$

$$H = \sum_b Q_b^2 / 2C_b + \Phi_b^2 / 2L_b + E_J \cos \Phi_b \quad (6)$$

In addition, the seemingly independent branch variables need to satisfy two conservation laws-Kirchoff current conservation on a node and loop flux quantization constraints, respectively :

$$\sum_{b \in \text{node}} Q_b = Q_{\text{node}} \quad \sum_{b \in \text{loop}} \Phi_b = \Phi_{\text{ext}}^{\text{loop}} + 2n\pi\Phi_0 \quad (7)$$

This Hamiltonian already contains the necessary physics to describe the dynamics of the circuit. The definition of branch charge Q_b is derived as a conjugate variable. It happens to have a physical meaning with respect to the capacitance relation.

The volatge-current picture suits the experimental setting because of physical measurement available to nodes and branches. On the other hand, the charge-phase variable suits the theory. It provides an abstract picture for quantization and wavefunction analysis, agnostic on circuit parameters and units. The translation dictionary between the two picture :

$$\vec{V}_n = C_n^{-1} \vec{Q}_n \quad \vec{I}_b = L_b^{-1} \vec{\Phi}_b \quad (8)$$

It is important to note that voltages and currents V_n, I_b , would be equivalently quantized and discrete upon measurement.

2.2.3 Node-Branch formalism

In general, a circuit usually have many parallel branches and higher connections compared to the number of nodes $N_b > N_n$. Because of conservation conditions, it is a redundant description. Defining degree of freedom on each node (Q_n, Φ_n) results in simplification [15]. It is alternatively respresented as vectors $\vec{Q}_n, \vec{\Phi}_n$ of N_n length.

Kinetic term in the branches are redefined in node formalism, ensuring the total energy content:

$$C_B = \text{diag}(C_b), b = 1, 2, \dots \quad \sum \frac{Q_n^2}{2C_N} \equiv \sum \frac{Q_b^2}{2C_B} \quad Q_i - Q_j = Q_b, b = \langle i, j \rangle \quad (9)$$

Node capacitance C_N matrix ($N_n \times N_n$), is a symmetric matrix. Diagonal terms C_N^{nn} are sum of all capacitance connected to that node n . And off-diagonal terms C_N^{ij} are negative sum of capacitances connecting each i & j .

The node description of branch flux variables Φ_b , needs to incorporate the flux quantization condition eq.3.

The problem is that a loop flux is not defined on branch individually, it is shared by all the branches of the loop. The rules of sharing are defined by the flux quantization. It is a condition on the aggregate of all the flux differences. Therefore, a branch is chosen, such that the total external flux Φ_{ext} being cast on that loop and the integer quantization. This is termed as the closure branch, it has to be unique for each loop.

2.2.4 Spanning Tree

A spanning tree S is defined as a subgraph of the circuit network N , such that each node has a unique path to the ground. The difference between $N - S = C$ is the set of closure branches. Closure branch is then a unique representative to each loop, such that $\Phi_{ext}^{loop} + n\Phi_0$ is appended to their branch fluxoid component. This algebraic balance closes each loop flux quantization constraint, independently.

The branch flux is expressed in node flux variables, under loop flux quantization, $\Phi_{\langle i, j \rangle} = \Phi_i - \Phi_j + \mathbf{cl}_b^{loop}(\Phi_{ext}^{loop} + n\Phi_0)$, $\mathbf{cl}_b^{loop} = 1$ if $\langle i, j \rangle$ is the closure of loop, otherwise $\mathbf{cl}_b^{loop} = 0$.

Spanning tree S for a defined network is non-unique and there are many legit ways to assign/distribute Φ_{ext} . Its a gauge choice as any choice of closure among branches is physically equivalent, B.

In this work, closure branches are chosen on a Josephson-junction branch. Such that quantization flux($n\Phi_0$) is non-participatory because of cosine periodicity. Assignment of offset flux variables to linearly inductive components is

incomplete under finite-dimension representation of fluxoid operators.

Importantly, assignment of closure branch on capacitive branches modifies its contribution as, $\sim (\dot{\Phi}_i - \dot{\Phi}_j + \dot{\Phi}_{ext})^2/2C_b$. The standard circuit quantization method would not hold, as inconsistencies arises due to gauge dependent terms, [16]. However, time varying external flux is beyond the scope of this work.

Subsequently, the branch inductance matrix $L_B = \text{diag}(L_b)$ is assigned the directional connection convention. The branch-node $N_b \times N_n$ matrix R_{bn} , for each branch $b = \langle n, n' \rangle$, $R_{bn} = 1$ and $R_{bn'} = -1$. It redefines inductance suitable to node description of flux variables Φ_n and equivalent to branch description eq.10 Hamiltonian.

$$H_U = \sum_{b=\langle i,j \rangle} \frac{1}{2L_b} (\Phi_i - \Phi_j)^2 + E_{J_b} \left(1 - \cos \frac{2\pi}{\Phi_0} (\Phi_i - \Phi_j + \mathbf{cl}_b^{loop} \Phi_{ext}) \right) \quad (10)$$

$$\vec{\Phi}_b := R_{bn} \vec{\Phi}_n \Rightarrow \vec{\Phi}_b^\top L_B^{-1} \vec{\Phi}_b = \vec{\Phi}_b^\top R_{bn}^{-1\top} L_N^{-1} R_{bn}^{-1} \vec{\Phi}_b \quad (11)$$

Unfortunately, the junction term is non-separable with node-branch transformation because of the cosine term. Its parameter, the junction energy E_{J_b} are not transformed in a node description. However, it is possible to expand cosine in the even powers of $\vec{\Phi}_b$, and subsequently implement the aforementioned routine. The length of expansion is obviously limited by the error in specification. This approach is adapted by quCAT [17].

2.2.5 Quantization

The node Hamiltonian is quantized by assigning a Hilbert space \mathcal{H} to each physical node of the circuit. The quantized states ψ_n of that node resides in \mathcal{H} . Each variable is lifted to an operator acting as a transformation on \mathcal{H} .

The quantized Hamiltonian in vectorized notation, discarding the constant junction energy term:

$$H_{LC} + H_J = \frac{1}{2} \vec{Q}_n^\top C_N^{-1} \vec{Q}_n + \frac{1}{2} \vec{\Phi}_n^\top L_N^{-1} \vec{\Phi}_n + \sum_b E_{J_b} \cos \frac{2\pi}{\Phi_0} (\Phi_i - \Phi_j + \mathbf{cl}_b^{loop} \Phi_{ext})$$

Quantization assign an additional axiomatic requirement on the conjugate pair: $[Q_n, \Phi_n] = i\hbar$. The equivalent condition is the Fourier transform, under $(\hbar \rightarrow 1) : \mathcal{F}^\dagger Q_n \mathcal{F} = \Phi_n$.

The cosine term in H_J could be expanded with Euler's identity in terms of

displacement operators \mathcal{D} ,

$$\cos \frac{2\pi}{\Phi_0} (\Phi_i - \Phi_j + \mathbf{c}_b^{loop} \Phi_{ext}) = \frac{1}{2} (\mathcal{D}_{ij}^+ \exp \left\{ \iota \frac{2\pi}{\Phi_0} \Phi_{ext} \right\} + \mathcal{D}_{ij}^- \exp \left\{ -\iota \frac{2\pi}{\Phi_0} \Phi_{ext} \right\})$$

$$\mathcal{D}^\pm = \exp \{ \pm \iota \Phi_i \} \exp \{ \mp \iota \Phi_j \}$$

Immediate step post-quantization is basis assignment. The abstractly defined wavefunction ψ and operators O acting on \mathcal{H} require numerical representation. The basis $|e_i\rangle$ is an orthonormal set of unit vectors that span the \mathcal{H} .

$$|\psi\rangle = \sum_i^n \psi_i |e_i\rangle \quad \hat{O} = \sum_{ij}^n o_{ij} |e_i\rangle \langle e_j|$$

Under finite memory of computer, a complete \mathcal{H} could not be realized and the vectors(ψ) and matrices(\hat{O}) are truncated at a length, $n := n_{trunc}$. Truncation size should allow faithful representation of low-energy dynamics.

Depending on the diagonalized representation of an operator, that basis could be given a corresponding definition. For example, a charge basis(q) diagonalize the charge operator(\hat{Q}), flux basis(ϕ) diagonalize flux operator($\hat{\Phi}$) and an oscillator basis(o) diagonalize the combination of two.

In other words, the basis is constructed by the eigenvectors of the respective operators. Operator-basis notation is defined as - \hat{O}_{basis} :

$$Q |q\rangle = 2q |q\rangle \quad \Phi |\phi\rangle = \phi |\phi\rangle \quad H_o |n\rangle = \hbar\omega(n + 1/2) |n\rangle$$

Therefore, the numerical representation is a diagonal matrix:

$$Q_q := 2 \begin{bmatrix} -n & & \\ & \ddots & \\ & & n \end{bmatrix} \quad \Phi_\phi := \Phi_0 \begin{bmatrix} -\phi_0 & & \\ & \ddots & \\ & & \phi_0 \end{bmatrix}$$

Besides, coherent state basis is also frequently chosen.

Having defined the basis, subsequent operators such as $\Phi_q, Q_\phi, \Phi_o, Q_o$ etc., are derived from their physical relation to the diagonalized operators.

Charge states have a natural uniform grid, the integer number of Cooper pairs. There is no continuum in between because of charge quantization. Unlike, the fluxoid variables is assigned an artificial grid, usually uniform.

In $DiSuQ$, the conjugate operators, that is the non-commuting pair are derived via Fouier transformation(\mathcal{F}). This construction respect the uncertainty

principle, even under finite representation. For example:

$$Q_\phi = \mathcal{F}^\dagger \Phi_\phi \mathcal{F} \quad \Phi_q = \mathcal{F}^\dagger Q_q \mathcal{F}$$

2.2.6 Modal Transform

Numerical representation of operators require a basis in \mathcal{H} . However, a second, more network-inspired transformation improves the numerical representation of the Hamiltonian operator.

The transformation \mathcal{R} attempts to maximize the circuit description as a collection of anharmonic oscillators. It does so by segregating the subspace of nodes associated to inductor branches. The inductance matrix is thus block diagonalized:

$$\begin{aligned} L^{-1} &= \begin{bmatrix} L_O^{-1} & 0 \\ 0 & 0 \end{bmatrix} & C^{-1} &= \begin{bmatrix} C_O^{-1} & C_{OJ}^{-1} \\ C_{OJ}^{-1} & C_J^{-1} \end{bmatrix} \\ C^{-1} &= \mathcal{R} C_N^{-1} \mathcal{R}^{-1} & L^{-1} &= \mathcal{R} L_N^{-1} \mathcal{R}^{-1} \\ \mathcal{R} \vec{\Phi}_n &\rightarrow \vec{\Phi} = [\vec{\Phi}_O \ \vec{\Phi}_J] & \mathcal{R} \vec{Q}_n &\rightarrow \vec{Q} = [\vec{Q}_O \ \vec{Q}_J] \end{aligned}$$

The division of transformed variables are termed as oscillator(O) and Josephson(J) modes. Subsequently, Q, Φ are termed as "mode" operators and $\vec{Q}, \vec{\Phi}$ as vector of these operators.

$$\begin{aligned} H &= H_{LC} + H_J = H_O + H_{int} + H_J \\ H_O &= \frac{1}{2} \left(\vec{Q}_O^\top C_O^{-1} \vec{Q}_O \right) + \frac{1}{2} \left(\vec{\Phi}_O^\top L_O^{-1} \vec{\Phi}_O \right) \\ H_{int} &= \frac{1}{2} \left(\vec{Q}_O^\top C_{OJ}^{-1} \vec{Q}_J \right) & H_J &= \frac{1}{2} \left(\vec{Q}_J^\top C_J^{-1} \vec{Q}_J \right) \end{aligned}$$

Turning to basis assignment, oscillator modes are defined in oscillator/number basis. The Josephson modes are defined in the charge basis, as there are no inductive element In this way, flux variable representation is minimized in charge basis.

The original formulation considers bifurcation of Josephson modes, further into Josephson and "island modes". However, it is again described in a charge basis and the segregation might not be straightforward [18].

Geometric heuristic on circuit modes are crucial in Hamiltonian simplification. And different approach to circuit formulation takes different perspective app.D.2. Consideration of circuit parameters and their dominance might alter the heuristic, as well. As of now, there isn't sufficient comparative evidence to single out a unified theory. Importantly, geometric transformation mean that, the exact value

of the physical parameters should not change transformation \mathcal{R} .

In theory, transformation \mathcal{R} is an invertible unit norm (preserves the commutator), inductance matrix is reduced into disjoint subspaces. However, it is not a unique transformation, as the subspace L_O could be rotated separately.

In analytic methods sec.4.2, the transformation is constructed with integers and in positive/negative pair. It simplifies the calculation and make the transformation interpretable.

In this work, full alignment is requested to R , such that L_O is a diagonal matrix. The strict requirement has a definite diagonalizing transformation (upto permutation). And the weightage maybe not be integer.

It is a noted drawback, as the transformation could include more oscillator modes in the description of junction energy. Displacement operators in oscillator basis have inconsistent response towards truncation sec.3.2.

Finally, numerical construction of the Hamiltonian requires choice of bases(charge:q, oscillator:o) on each mode. The numerical construction of these operators is adapted from Kerman [18].

In general, the mode(A_i)-mode(B_j) interaction(O_{ij}) is a matrix multiplication with a vector of operators:

$$\vec{A}^\dagger O \vec{B} = \sum_{ij} O_{ij} A_i^\dagger B_j \quad \vec{A}, \vec{B} \in \{\vec{\Phi}, \vec{Q}\} \quad (12)$$

Finally, the Hamiltonian is expressed in a composite basis of the size, $\prod_i n_{trunc}^i$, resulting from the truncation of individual mode.

In this thesis, prevalent convention of the literature [14][19] is adopted:

- Charge, Flux operators are in reduced terms : $n = Q/2e, \phi = \Phi/\Phi_0$
- External flux $\Phi_{ext} \in [0, 1]$ is alternatively expressed in reduced unit ϕ
- Units of capacitance and inductance are converted to energy : $E_C = e^2/2C, E_L = \frac{\Phi_0^2}{4\pi^2 L}$, to compare against E_J

2.3 Diagonalization

Proceeding through previous chain of calculations and choices, the circuit Hamiltonian is derived. It is a truncated representation of the Hamiltonian as a Hermitian matrix, that acts as a energy operator in \mathcal{H} for the underlying circuit, in a mixed & composite basis.

The time independent dynamics of the circuit, is set of energy-state pair $\{E_i, |\psi_i\rangle\}$. The eigenvalue problem $H |\psi_i\rangle = E_i |\psi_i\rangle$, is solved by diagonalizing the Hamiltonian $H = \sum_i E_i |\psi_i\rangle \langle \psi_i|$, eigenvector is in the same basis as H .

It is the most called subroutine in the proposed methods of numerical discovery. Therefore, this pursuit would be only be as good as diagonalization.

Locally Optimal Block Preconditioned Conjugate Gradient(LOBPCG) is a numerical algorithm to solve eigenvalue problem for Hermitian matrices. It is widely used in modern programming modules and utilised in *DiSuQ* , for Hamiltonian diagonalization.

In principle, LOBPCG is an iterative gradient descent method, that minimize the Rayleigh quotient $\rho(H, x) = x^\dagger H x / x^\dagger x$, to approximate a candidate x vector to optimal eigenvector.

Gradient of the following Lagrangian \mathcal{L} , in combination with its historic aggregate directs along X_\perp^i .

$$\mathcal{L}(X, \Lambda) = \frac{1}{2}(X^\dagger H X)^\top - \frac{1}{2}(X^\dagger X - \mathbf{1})\Lambda^\top$$

It results in the next update $X^{i+1} = c_1 X^i + c_2 X_\perp^i$. The coefficients c_1, c_2 are evaluated by Rayleigh-Ritz method [20], minimizing the Rayleigh quotient. Eigenvectors are usually bundled in a bloc $[x_i] = X$ of dimension $n \times k$, $k \leq n$. It is efficiently implemented on parallel computers.

Efficient convergence of LOBPCG is maintained by the preconditioning operations, prior to each operation fig.4. The original problem is reformulated to suit finite-arithmetic on computation.

Unfortunately, in prevalent python implementation, controller on the size of block is unavailable. It makes bulk of the calculation redundant, as quantum configurations are studied in low-energy regime. In fact, the underlying physical theory breakdown for higher energy levels. So, an improvement on current utilization of LOBPCG would be to limit the size of the block to lowest levels(~ 4).

Eigenvalue calculation with limited computational resource requires efficient utilization of memory and computation. As the size of Hamiltonian increases exponentially with each node, the memory explodes. As well as, low truncation level of individual mode reduces the accuracy of circuit modeling, sec.3.2.

Fundamentally any algorithm involving matrices implement summation and multiplication. The usage of parallel distributed computing tremendously improves the processing time, usually at the cost of memory (app.D.2).

Hierarchical diagonalization proposed by Kerman[18], is an effective method to decompose the network into sub-modules on the basis of network repetition,. It is particularly effective in designing couplers mediating interaction between multiple qubits. It is implemented in scQubit[21].

2.4 Optimization

To address the proposition of this thesis, so far, the superconducting circuit half is mathematically formulated. The discovery of these circuits is formulated in terms of maximization and/or minimization of set of features. The procedure of optimization is automated discovery, as illustrated in other relevant studies such as [22][23] [24].

In this work, the optimization is carried over circuit parameters $\{E_C, E_L, E_J\}$ etc., to minimize the loss function built with physical features of the circuit.

Typical features to be optimized in a superconducting circuit are related to its low-energy eigenspectrum. Such as, for qubits, the coherence time T_1, T_2 and anharmonicity α are desired to be maximal, simultaneously. Similarly, for couplers implementing gate operations, robustness of the spectrum profile is important [4].

In addition to the loss functions, optimization might be requested on a constrained subspace of the circuit's full parameter space. In *DiSuQ*, circuit parameters are assigned following constraints, via computational methods:

- Domain Boundary : For an engineering problem, the circuit parameters have practical bounds. Besides, null parameters might result in gradient instability(2.4.2). App.E suggest typical bounds based on current experiments.

Defining a bound function with unconstrained domain wraps the parameter as a bounded variable. Sigmoid function $\sigma(x) = \theta \in (0, 1), x \in \mathbb{R}$, stagnate gradients at boundary to avoid gradient instability.

$$\frac{d\mathcal{L}}{dx} = \frac{d\mathcal{L}}{d\theta} \frac{d\sigma(x)}{dx} \quad \lim_{x \rightarrow \pm\infty} \frac{d\sigma(x)}{dx} = 0$$

- Symmetry : Many ansatz on circuit design require symmetry on elements (usually diagonal), or couplers on identical sub-modules. It adds novel features in the spectrum profile, without raising parameter space. With the advantage of symbolic computation, sharing a variable across multiple element equalizes them for any subsequent computation. This is synchronization could be generalized beyond identity, as well.

These preconditioning allow unconstrained optimization algorithms to update circuit on underlying parameter space(x). Otherwise, explicit constraints in terms of Lagrangian multipliers [25] or other symmetry enforcing objective [4] in the loss function are specified. This is usually undesirable, since empirical weight adjustment is required to balance the scale of primary objective and constraint loss.

In order to explore the parameter space $\{\theta_i\}$ for the optimal value of loss function $\mathcal{L}[\theta]$, efficient gradient based methods are implemented. This assumes a locally smooth and maximally convex loss function.

Gradient-based optimization improves the solution iteratively, identifying a better solution with lower loss in the neighborhood of current state. A starting point is explicitly provided.

Locally optimal solution is in the direction of steepest descent identified with the gradient of loss function over parameters $x^{i+1} - x^i \propto -\frac{\partial \mathcal{L}}{\partial \theta_i}$. The search terminates at $\nabla \mathcal{L} = 0$, considered as optimal.

This technique is highly successful and explored in the context of machine learning. The problem of discovering/optimizing novel circuits is apparently not much different.

Despite, the roots of a polynomial are smoothly dependent on its coefficients, root-finding algorithms are usually ill-conditioned. A small variation in the coefficients results in significant variation of the roots. Smooth dependence fails at duplicate roots.

Fortunately, eigenvalue problem for Hermitian matrices is proven to be well-conditioned [26]. Therefore, the variation in the eigenspectrum of the Hamiltonian due to optimization updates would not be out-of-bounds.

2.4.1 Computation Graph

Computation graph is a directed graph of symbolic placeholders resulting from elementary calculations, usually addition and multiplication of matrices. It is utilized to manifest whole chain of calculating the circuit Hamiltonian.

The leaf nodes are static values (operators) or variables (circuit parameters). Subsequently each compound operation such as inverse, Rayleigh quotient, LOBPCG fig.4 etc. are encoded as a subgraph in terms of addition and multiplication.

In the present context, it terminates at the spectrum and subsequently with loss function. The actual calculation is executed by traversing from the leaves to the root, termed as forward pass. The computation graph is pre-compiled for efficient execution on machine-infrastructure. The tolerance based termination of LOBPCG is implemented via dynamics graph in *DiSuQ*.

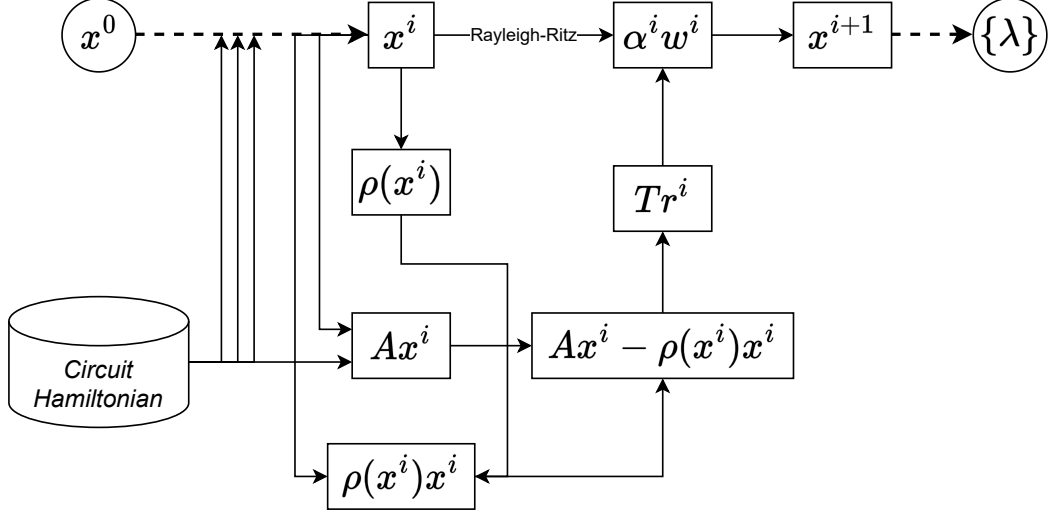


Figure 4: Computation graph of preconditioning segment of LOBPCG algorithm for eigenvector calculation, $Ax = \lambda x$. x^0 is iterated dynamically until tolerance termination. Every node represent either summation, multiplication, coefficient calculation. Rayleigh-Ritz subroutine is encapsulated. Circuit Hamiltonian data is shared at each iteration.

Feature	Static	Dynamic
Computation deployment on distributed processors	faster	difficult
Optimization(scheduling, redundancy etc.)	sufficient	limited
Indefinite length of calculation	pre-allocation	runtime

Table 1: Feature comparison between static and dyanmics computation graphs[27]

2.4.2 Backpropagation

The chain rule of derivative expands the gradient of the loss function \mathcal{L} , under summation of indices as,

$$H|i\rangle = \lambda_i|i\rangle \quad \nabla_{\theta}\mathcal{L} = \frac{\partial\mathcal{L}}{\partial\lambda_k} \frac{\partial\lambda_k}{\partial H_{ij}} \frac{\partial H_{ij}}{\partial\theta_l} \quad (13)$$

The computation graph sets the stage for efficient gradient calculation. Backpropagation is an algorithm to calculate derivative of nodes on a computation graph [28]. Backpropagation, similar to forward pass runs in opposite direction, from the terminal loss node to the leaves of variables.

The algorithm is simply keeping the track the value of loss gradient with respect to the intermediate step. It is efficiently stored on its node in the computation graph. In the backward pass, each node recieves the gradient from the preceeding node via chain rule [28].

Calculation of analytic gradient providing infinitesimally local gradient. For-

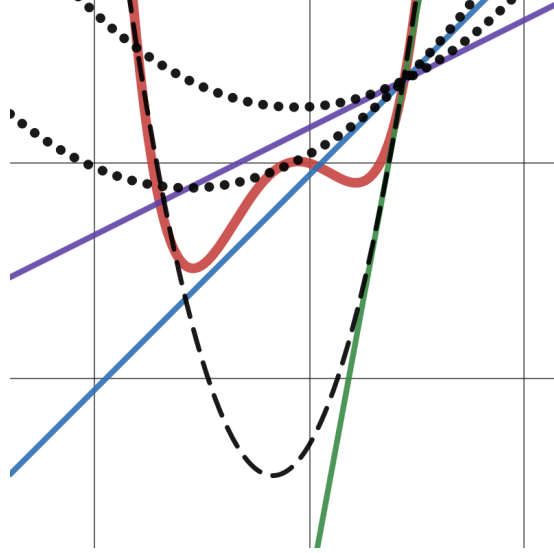


Figure 5: Forward difference method calculation at intersection point. Objective function(red) is estimated with parabola at the intersection point. Green line depicts analytic gradient, fitting dashed parabola. Blue and purple secant use forward value to calculate gradient, fitting dotted parabolas estimate.

ward difference method require local adaptation for better estimate, raising overheads and tolerances. [29]

Computationally, backpropagation is equally expensive as forward pass, required by forward difference method. So, at the same cost it is superior for gradient calculation.

Backpropagation treats iterative steps of LOBPCG just like any other computation graph, fig.4, in the calculation of $\frac{\partial \lambda_k}{\partial H_{ij}}$. Despite efficient convergence, this computational graph is not static as its truncation is dependent on numerical tolerance of approximate eigen-solution, subjected to $\|X^0 - X^*\|$. So, the gradient of eigen(value/vector) on Hamiltonian $\frac{d\lambda_k}{dH_{ij}}$ is accumulated on this chain of iteration; similar to RNN's backpropagation through time [30]. Tolerance verification is out of the scope of backpropagation. This method requires smoothly differentiable operations. Procedures discrete in nature such as sorting are non-differentiable via chain rule. Its a drawback of the present implementation, that curb better full compilation of the computational graph, since LOBPCG requires explicit sorting [20].

Gradient-based optimization iteratively provide locally optimal solution. However, in the process of calculating the gradient or a solution, at any stage, it might breakdown for numerical or algorithmic reasons [31]:

- Vanishing and Exploding gradients : Gradients of recurrent routines ($\frac{\partial x^{i+1}}{\partial x^i}$ in LOBPCG) get multiplied over at each step, on backpropagation. If unchecked, it could grow up or shrink down beyond precision.

-
- Anti-directional : In Newtonian optimization app.A, the loss surface is assumed convex, that is Hessian > 0 . A negative curvature would reverse the direction against steepest descent.
 - Local minima and Saddle point : Sub-optimal solution with $\frac{\partial \lambda}{\partial \theta} = 0$, would truncate gradient-based optimizer.

In the examples sec.3, diverse range of optimizers are implemented, each with special advantage against the aforementioned failures.

In conclusion, rationale to implement a memory-extensive computation graph and a backpropagation on it, originates from the requirement of a faster search. It is needed to accurately calculate the gradient of the Hamiltonian eigen-spectrum, possibly in a large parameter space. Such that, the sensitive gradient information would feed the LBFGS. Importance of this Hessian based gradient-descent optimizer relies on the nature of the problem. Unlike a black-box function, the optimization is known to be a smooth problem, A strong convexity would further embolden the performance.

Computing LBFGS over LOBPCG, is nesting a gradient descent optimization over another. It is termed as Bilevel optimization problem[32]. For aforementioned reasons, this problem in general is technically and computationally challenging[33].

The proposition of this thesis rests with the successful demonstration of this bi-level optimization for superconducting circuits.

3 Analysis

In the previous section, Hamiltonian formulation of a superconducting circuit is prescribed. It ends on the numerical description of the operator quantization. There are two critical question to this approach:

- Why adopt a numerical procedure to diagonalization/optimization?
- Which numerical basis to choose for operator representation ?

The first, existential question explores the importance of numerical methods in the context of superconducting circuits. In the prevalent analysis, analytic solution to a circuit Hamiltonian is provided. Subsequently, an optimal configuration is identified/justified based on the required features.

Mostly, such an approach requires a special variable transformation, that could reduce/decouple/aggregate the degrees of freedom. It is motivated by symmetry in the network design. For example, centroid-reflection in Zero-Pi network fig.15, an array in Fluxonium etc. Thereby, covering a small subspace of circuit networks and losing generality.

Secondly, analytic solution is applicable usually under certain approximation, that simplifies the differential equation. That is, its mathematical guarantee is limited to a subspace of the circuit parameters. Asymptotic reduction in the case of a transmon qubit in the following example sec.3.1, highlights the shortcomings of approximation methods.

Theoretically speaking, numerical methods doesn't constraint the problem space neither in network nor in parameter values.

However, the versatility of the numerical methods is dependent on the appropriate representation of the Hamiltonian, that contains all the physics. It leads to the second question - the challenge to identify an appropriate basis sec.3.2.

It turns out that, the subspace of interest is the middle region of parameter domain. Here, the asymptotic approximation is inadequate and basis selection is increasingly inappropriate. It links the two question, as the both the approaches to circuit exploration are failing in the middle region - analytic more than numerical. Section3.3 describes a systematic comparison and its practical importance.

3.1 Asymptotic Analysis

Asymptotic approximation assumes an extreme position in circuit parameters to dominate certain terms over others, thereby simplification by reduction. Conventional analytic methods such as statistical aggregation, quasi-potential[14], WKB [34], etc. are the examples.

In the context of superconducting physics, energy spectrum or eigenstates could be approximated with known functions(summation series). The scaling is usually derived taking ratio of two parameters, or extending it to multiple independent ratios(sec.4.2).

Perturbative method is a bridge between asymptotic solution and the numerically computed results. Starting at the boundary additional terms are included upon departing the extremal subspace to maintain the accuracy. Calculation of these additional terms usually gets exponentially complex, and is equivalent to matrix diagonalization.

Major drawback of this analysis is that its limited in the space of coverage and complexity of the network. And its extreme solution are impractical to fabricate, or the lumped modeling of the circuit dismantle due to a different physics.

In the following sec.3.1.1, transmon regime improving the coherence time is studied. This example demonstrate inevitability of numerical computation and asymptotic analysis acts as a feasibility check.

3.1.1 Transmon - Charge Insensitivity

Transmon is a simple implementation of a qubit with Josephson junctions sec.???. Historically, it is the first example where a known circuit was improved by altering its parameter regime [34].

In this section, a sequence of simulation on transmon demonstrate improvement of spectrum robustness against charge noise. This improvement is observed to be smoothly progressing. Smooth variation of circuit feature is an important assumption in the numerical techniques discussed before sec.2.4 and demonstrated in later examples.

A single node circuit, with a junction and a shunt capacitance on parallel branches is termed as a Cooper pair box(CPB).

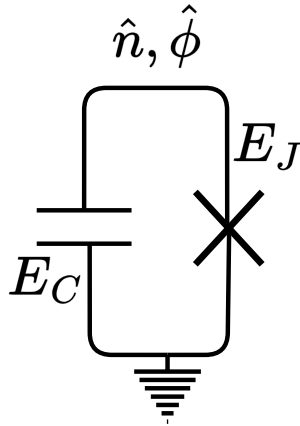


Figure 6: CPB/Transmon circuit supports anharmonic potential

The Hamiltonian of circuit fig.6, is described in phase space of (n, ϕ)

$$H = 4E_C(n - n_g)^2 - E_J \cos \phi \quad (14)$$

n_g is the offset node charge sec.2.2.1, as an external parameter to the circuit. E_J is junction energy, and $E_C = e^2/2(C_J + C_S)$, combining junction and shunt capacitance in parallel. In practical scenario, junction energy E_J could also be tuned by substituting it with a SQUID loop [14].

Replacing the inductor with a junction in harmonic oscillator sets up levels of an an-harmonic oscillator fig.6. That is, in contrast to equi-spaced levels of a harmonic oscillator $\omega_{i+1} - \omega_i = \sqrt{8E_L E_C}$, sec.2.1.1.

Anharmonicity (α), is a crucial requirement in qubit control. To prepare a qubit in a defined state, excitation pulse must exact transition energy. Typically, more pulses is desired, to have faster manipulation in a single coherence session.

However, shorter pulses introduces larger uncertainties in drive frequency and excitation might shoot a qubit to undesirable non-computational levels. Assigning an offset(in either direction), $(E_2 - E_1) = (E_1 - E_0) - \alpha$, reduces the probability.

Therefore, minimum pulse time $\tau_p \propto \text{mod } \alpha^{-1}$ assigns an upper bound on the code length.

Charge dispersion is defined as variation in spectrum, with respect to change in node charge, $\frac{\partial E_{ij}}{\partial n_g}$. In general, faulty fabrication introduce substrate impurities. This inhomogeneity accumulates charge, builds random electric fields and effectively fluctuates the node charge.

The operating parameter of a CPB is $E_J/E_C \leq 1$ ([35]). This configuration builds degenerate states at "sweet spots", $n_g = \pm 1/2$. The half integer minima has a potential restoring the circuit state against the charge fluctuations fig.???. However, the local minima is unprotected against large fluctuations. This problem is addressed by shifting the parameters to transmon regime, $E_J \gg E_C$ [34].

Transmon Hamiltonian eq.14 is effectively a Schrodinger equation $\dot{\phi} \equiv n$, a second order ODE with cosine potential. Analytically, in the phase basis, exact solutions are in the span of Mathieu functions [36].

The three important features are large fluctuations corresponding to the peak-to-peak difference ϵ_i , local fluctuation at CPBs "sweet spot" and the anharmonicity $\alpha = E_{01} - E_{12}$, are analytically derived [34]:

$$E_i(n_g) \simeq E_i(n_g = 1/4) - \frac{\epsilon_i}{2} \cos 2\pi n_g \quad \epsilon_i = E_i(n_g = 1/2) - E_i(n_g = 0)$$

$$\epsilon_i \propto E_C \left(\frac{E_J}{2E_C} \right)^{\frac{2m+3}{4}} \exp\left\{-\sqrt{8E_J/E_C}\right\}, \quad E_J/E_C \gg 1$$

$$\frac{\partial E_{01}}{\partial n_g} \simeq \pi \epsilon_1 \sin 2\pi n_g$$

Because of the shift in operating regime, the transmon circuit exhibit following:

- Charge insensitivity : $\frac{\partial E_{01}}{\partial n_g}$ falls exponentially with E_J/E_C fig.7. The sinusoidal term is reminder of the CPB "sweet spot".
- Large fluctuation resilience : ϵ_i decays exponentially in $\sqrt{E_J/E_C}$
- Anharmonicity decreases asymptotically : $\alpha \simeq -E_C$ at degeneracy point $n_g = 1/2$

Therefore, there exists a window for sufficient anharmonicity and resilience, called the *transmon regime*.

This regime identifies a trade-off between statistical uncertainty from physical impurities against quantum uncertainty of operating/controlling the qubit. In this way, it directly relates computation capacity to fabrication precision.

Importantly, the Mathieu functions are not exactly solvable and the derived results assumes a WKB approximation [34]. Therefore, in non-asymptotic parameter space, that is the transmon regime, a numerical simulation is eventually necessary for verification of the ansatz.

These simulation fig.7,fig.8, thus verifies the analytic assumptions, that are made to support the simplest workable superconducting qubits. Computationally, diagonalization is carried at each grid point in $n_g, E_J/E_C$.

In principle, this trade-off regime could be achieved as an optimal point for the two features, via efficient optimization methods. Following examples, identify the efficient methods to optimization for much complicated circuits, utilizing the smoothness quality of the problem demonstrated, here.

In theory [34], transmon regime is justified by analytic means. That is, with increasing E_J/E_C , peak-to-peak variation ϵ_i scales exponentially while anharmonicity α scales with power law. Thereby, the gain in T_2 would eventually over-run the rise in pulse interval τ_p . Overall, higher the E_J/E_C ratio higher the operation density.

Unfortunately, this analytic method relies on the asymptotic approximation, $E_J \gg E_C$. It only provides a qualitative analysis, that establish the existence of such regime. However, for practical realization this ratio could not be indefinitely large. This method is incapable of determining at what value and beyond, such an asymptotic approximation is valid.

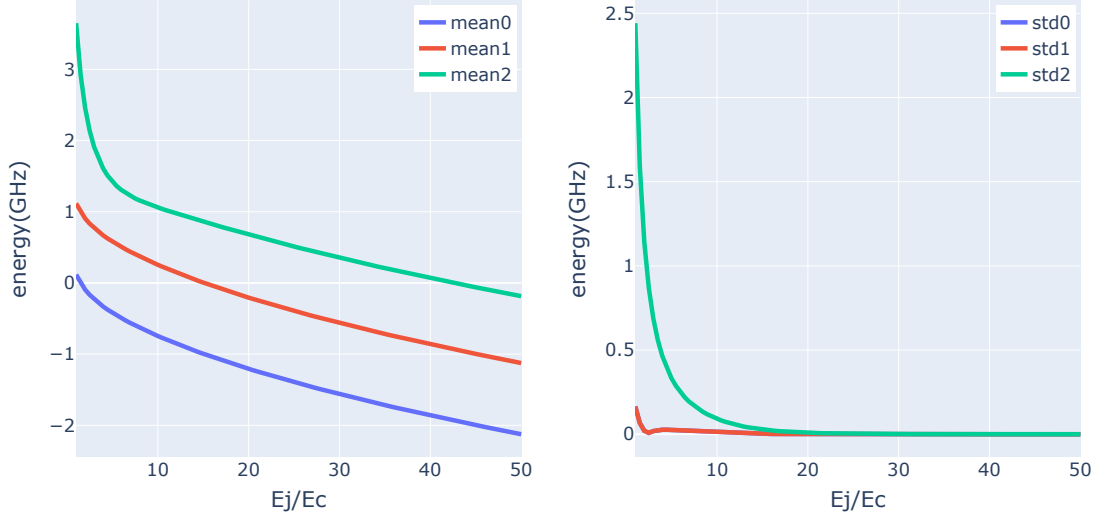


Figure 7: Comparison of anharmonicity and charge sensitivity as a function of E_J/E_C (x-axis). Mean(left) and standard deviation(deviation) of relative energy level at $n_g \in [-2, 2]$; $E_0/E_{01}(n_g)$ (blue), $E_1/E_{01}(n_g)$ (red), $E_1/E_{01}(n_g)$ (green). The gaps between successive energy levels equalize $\alpha to 0$, in the mean. The standard deviation representing $\epsilon_i \rightarrow 0$, exponentially.

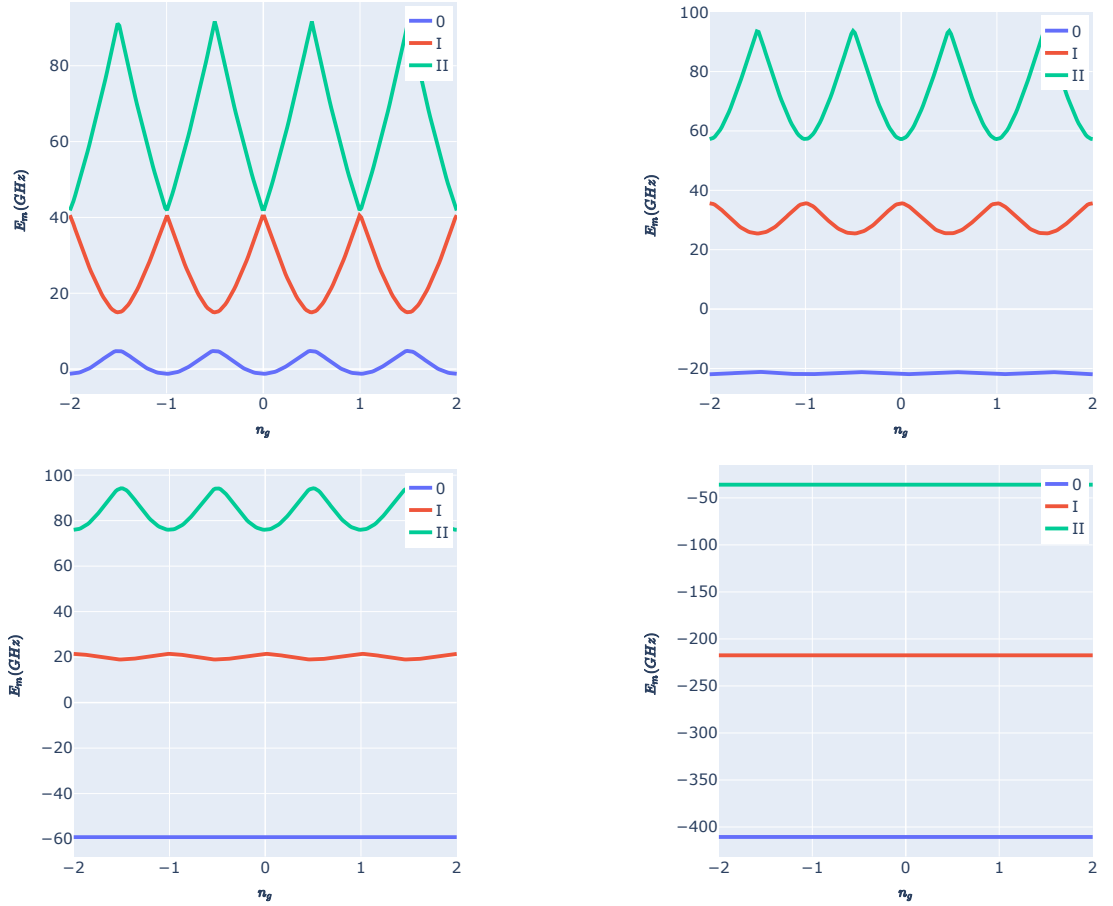


Figure 8: Energy levels(0,Ist,IInd) of transmon circuit as a function of n_g , for different E_J/E_C ratio, clockwise(from top left):1, 5, 10, 50

Consider a schematic of this method - two objective functions $f(x), g(x)$, such that f/g is maximized; similar to T_2, τ_p . Further, only their asymptotic scaling behaviour is known,

$$\begin{aligned} \lim_{x \rightarrow \infty} f(x), g(x) = \infty \quad \lim_{x \rightarrow \infty} f(x) \sim \exp\{x\}, \quad g(x) \sim \sqrt{x} \\ \Rightarrow \lim_{x \rightarrow \infty} f(x)/g(x) \sim \exp\{x\}/\sqrt{x} = \infty \end{aligned}$$

There are two independent functions f, g being compared on the same point x , that not be mathematically rigorous. Their scale of asymptote could differ, as they might be derived from different approximations. In the case of transmon, namely, the WKB and the quartic cosine expansion [34]. In order to identify a shared asymptotic region, asymptotic trend for $f \& g$ must be numerically justified, independently fig.11. Atleast, one trend regression calculation in the best case. Further, identifying a lower bound of this region is effectively searching the inner space.

This argument could be generalized to any analytic method employing asymptotes.

3.2 Basis Representation

Numerical representation of Hamiltonian is based on the choice of basis, such as charge, flux, number/oscillator etc. In principle, any unitary transformation in the Hilbert space, that preserve the norm of the wavefunction could shift representation from one basis to other ($\Sigma \rightarrow \Omega$):

$$|\psi\rangle_{\Sigma} \rightarrow U^{\dagger} |\psi\rangle_{\Sigma} = |\psi\rangle_{\Omega} \quad U^{\dagger} H U U^{\dagger} |\psi\rangle = \lambda U^{\dagger} |\psi\rangle \quad U U^{\dagger} = \mathbf{1}$$

Another characteristic of describing the H matrix is the truncation limit on its individual modes. Final size of the matrix is the product of individual mode limits. This restriction discourage equivalence between basis transformation, since, the transformation $\Sigma \rightarrow \Omega$ is never complete.

Computational constraint are inevitable by the exponential rise of matrix size with increasing nodes in the circuits. It is expected that in the asymptotic limit $n_{trunc} \rightarrow \infty$, choice of the basis shouldn't matter. However, for practical reasons its important to compare the convergence rates for different bases.

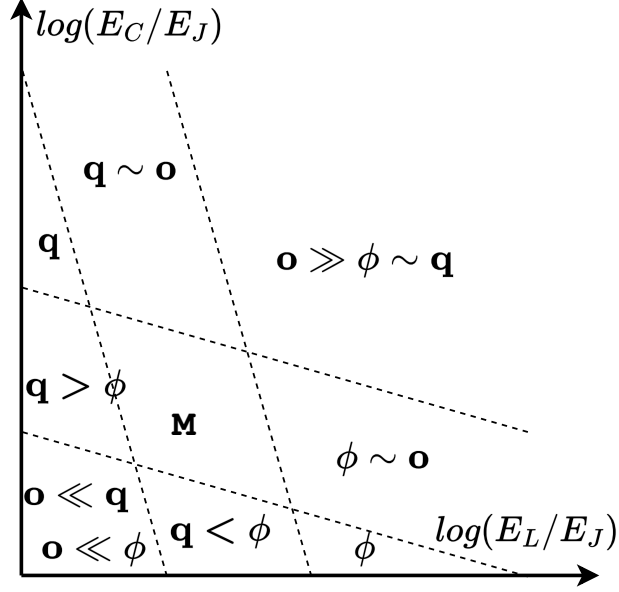


Figure 9: Schematic of parameter regime in log-scale, for anharmonic oscillator. Transition lines demarcate phase boundaries of dominating parameter. Each transition line mark a constant oscillator frequency, $2 \log \omega - \log 8 = \alpha \log E_c + \beta \log E_l$. Slopes of transition lines (α, β) specify power scaling of respective dimension. Inner lines mark lower ω ; Outer lines mark higher ω . 8 regions are specified with relatively effective basis (oscillator(o), charge(q), flux(ϕ)). Unspecified basis doesn't converge. Inequalities compares convergence rate η . Middle region(M) is ambiguous, all basis converge. Axes intersection doesn't necessarily mean origin.

3.2.1 Single Mode Example

Consider, a single node anharmonic oscillator :

$$H = 4E_C n^2 + \frac{1}{2} E_L \phi^2 - E_J \cos \phi \quad (15)$$

Each of these component operators act on the same Hilbert space. Depending on the parameter regime (E_C, E_L, E_J) , fig.9, the dominating element dictates the spectrum λ_k of the overall Hamiltonian. Its a simple demonstration of Weyls' inequality for Hermitian matrices - higher values of circuit parameters coefficients (E_C, E_L, E_J) raises the limit of the spectrum

$$\lambda_k(4E_C n^2) + \lambda_k(\frac{1}{2} E_L \phi^2) + \lambda_k(-E_J \cos \phi) \leq \lambda_k(H) \quad k \in \{1, 2 \dots n\} \quad (16)$$

Therefore, in the extremal cases the bases should be compatible to dominating operators fig.9.

The sensitivity of eigen-spectrum with respect to interaction between different states, that is, the entries in the Hamiltonian matrix H_{ij} , determines the variation.

In the first order, $\frac{\partial \lambda_k}{\partial H_{ij}} = \frac{C_{oij}}{\prod_{i \neq k}(\lambda_i)}$, C_{oij} is the cofactor matrix app.C. To minimize the sensitivity, the basis representation should be maximally diagonal and/or sparse, such that $C_{oij} = 0$.

Therefore, the regime space fig.9, could be suggested based on diagonality/s-parsity - oscillator:H,E; charge:G,D; flux:I,F. It is corroborated in the truncation simulation, fig.10.

The remainder of interlinking space could possibly be non-conventional bases. Nevertheless, it might smoothly connect optimal basis of one dominant regime to another.

Preliminary simulation, fig.10, suggest several quantitative characteristics on the eigenspectrum error on truncation:

- Oscillator basis converge via oscillation around the limit
- Oscillator basis have larger basin of convergence, particularly in the middle region 'M'
- Charge basis converge via exponential decay
- Higher energy states have slower convergence
- Rate of convergence is regime dependent

Apparently and unfortunately, even in the practical parameter regime, the convergence rate could be impractically slow and results in stagnation under numerical precision of floating point fig.10.

3.2.2 Dominant Operator basis

Qualitatively, dominant operator analysis suggest that choice of appropriate basis is parameter dependent, fig10. It need not be choice on the known discrete bases - charge, flux; but could be relaxed smoothly as in oscillator basis. Similarly, there might additional bases relaxation between inductance & junction, or capacitance & junction etc. Identifying such a basis, that its truncated representation still carries complete information on low-energy subspace, is a clearly defined problem for further research.

Observations from single mode oscillator example could only be loosely extended to general multi-mode circuits. Despite, quantization of mode operators carried post variable transformation, is independent of each other. Additional inter-mode coupling C_{OJ} builds an interaction, that should be accounted in the analysis described in the previous example.

Theory on truncation is hard and largely unexplored [37]. Method of truncation is also subjective to the numerical method of operator construction. There

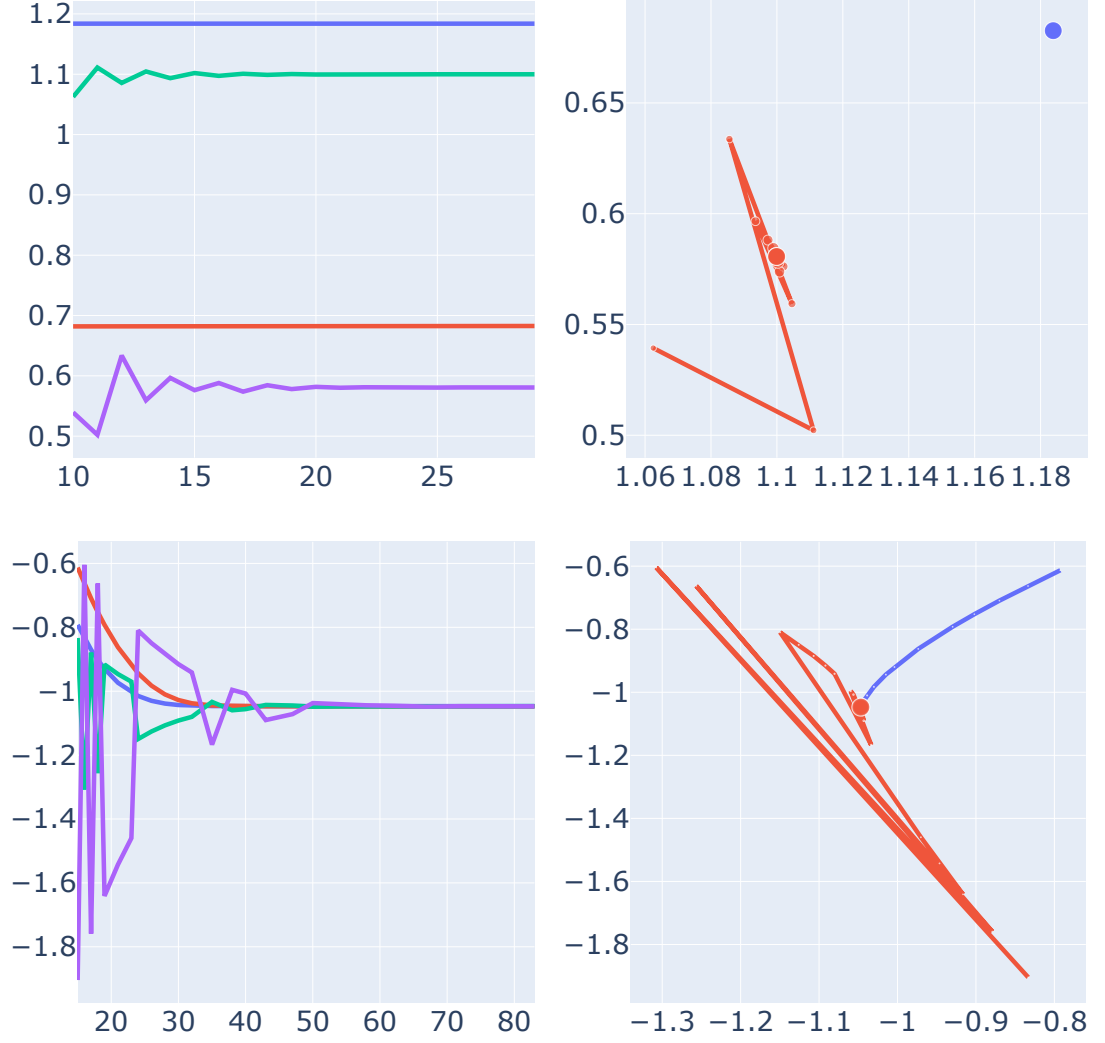


Figure 10: Convergence of two anharmonic oscillator eq.15(top: $E_L = 1, E_C = 3, E_J = 10$, bottom: $E_L = .1, E_C = 10^{-4}, E_J = 10$) transition energies on basis truncation. Left: X-axis: n_{trunc} , Y-axis: Transition spectra E_{01} : charge(blue), oscillator(green) and E_{12} : charge(red), oscillator(purple). Right: X-axis: E_{01} , Y-axis: E_{12} ; Bases: oscillator(red), charge(blue). Energy units in GHz.

are usually two different ways to represent an operator. Firstly, in the trivial diagonalizing basis where the operator is a diagonal. Secondly, via finite size transformation from other known basis. For example, a charge operator in charge basis could be constructed:

- Natural construction : $Q_q = \sum_{i=-n_{trunc}}^{n_{trunc}} 2i |i\rangle \langle i|$
- Diagonalization transformation : $Q_q = D^\dagger Q_o D \quad D^\dagger Q_o D \equiv diagonal$

Clearly, truncating in the diagonalizing basis is better than transformation from truncated oscillator. The error is expectedly maximal for transformation from flux basis, $Q_\phi, D = \mathcal{F}$

Empirically, as in fig.10, the rate of convergence $\eta(U; E_C, E_L, E_J) \in \mathbb{C}$ could be hypothesized to describe the truncation error $\epsilon_n = \|\lambda_k(H_n) - \lambda_k(H_{n\infty})\|$ as $\epsilon_n \propto \exp\{-\eta n\}$.

In the modal transformation sec.2.2.6, the \mathcal{R} matrix is held constant over complete parameter range. It transform could also be made parameteric over circuit parameters. Such that, the slow converging operators should be assigned minimal weights. For example, the displacement operator are described with \mathcal{R} transformation.

Thereby, a better theory for error due to truncation could control experiments scientifically, improving the consistency over various study.

3.3 Middle Region

Middle region could be defined as the subspace in the parameter space of a fixed circuit network, that is enclosed between the asymptotics. That is, none of the circuit elements E_J, E_C, E_L are dominating.

The choice of basis problem and weak convergence of numerical solution with limited computation, restricts faithful probe in this subspace. Simultaneously, being rich enough for potential application of superconducting circuits. As well as, keeping it practically feasible for fabrication.

The transmon example in sec.3.1.1, compares the relative growth of two features where the direction of monotonicity coincide. Generally, the features of interests in circuit design are ascending in opposite directions. Similarly, in this case as well, asymptotic analysis could only identify the trends and provide a qualitative guarantee of the existence of an optimal solution in the middle region. The quantitative result, obviously requires calculations and the definition of cost function 11. So, in the case multi-objective circuit exploration, middle region is inevitable.

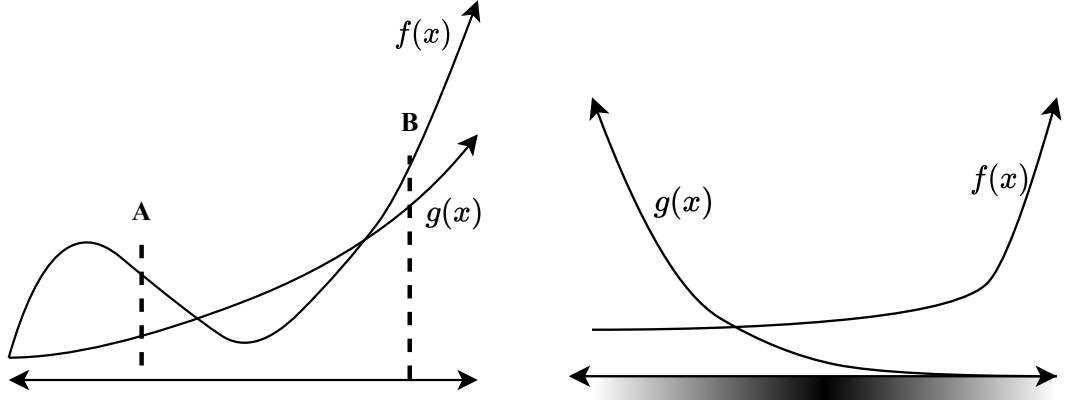


Figure 11: Schematic of multiobjective features f, g varying over parameter x . Left: Common direction of trend. Point A : $g(x)$ conforms its asymptotic trend. Point B : $f(x)$ conforms its asymptotic trend. Validity of asymptotic comparison is beyond B. Right: Anti-direction of feature's trend. Optimal lies in the middle region.

Because of inter-coupled modes even asymptotic ansatz might unavailable to beginwith. Thereby, adopting a numerical framework is more general.

Mathematically, perturbative solution out of the asymptotic approximation to the circuit dyanmics is computationally equivalent to numerical methods with matrices. However, the latter is computationally favourable.

It is expected that with sufficiently large computers, the numerical calculations would become more reliable in the middle region. Further, modern many-body solver such as Tensor-Network DMRG methods support the numerical approach to this problem [38]. The attempt to encroach the middle region should be a primary question guiding future research.

4 Examples

Superconducting circuits are effectively a finite system of massive particles with linear and sinusoidal potential arranged in a local pattern of interaction. The network, that is fixed in these examples, assigns topological features to the multi-body potential. Moreover, these coupling are two-body interaction over circuit branch. The purpose of circuit design is to shape this interactive potential. And exploring the parameter space is simply a fine-tuning over the network-assigned potential landscape.

In this section, discovery of superconducting circuits is demonstrated with three examples of multi-node qubits.

As discussed in sec.2.4, the optimization over circuit parameters identifies as discovery of new physics by numerical methods. So, a successful demonstration also validates the effectiveness of techniques previously discussed in sec.2.2.1,2.3,2.4, for this special problem.

Following examples are simple, such that, a single loss objective directs their search; symmetry constraints are inbuilt. Practical usage would involve much complicated multi-objective loss functions and constraints. It would also require careful supervision with optimizers' hyperparameters.

These experiments also provide benchmarking reference to future optimization task, such as choice of hyperparameters, initial points, truncation etc.

Optimization problem defined on eigenspectrum is fairly non-convex. Implementing convex methods on bi-level non-convex optimization makes this problem non-trivial.

In the following experiments, each optimizer is given multiple instances of initialization. Initialization grid is distributed normally around a known empirical solution. Generally, more instances cover larger search space, identifying more local minima. This increases the probability of identifying a global solution.

Running multiple instances multiplies the computational time proportionally and is effectively a brute-force method. However, for low dimensional search space it is comparatively better than momentum methods [39], that require careful adjustment over hyperparameters. As the dimensionality increases exponentially higher number of points are needed to cover the search space, maintaining the same density.

The C-shunted qubit is a regression problem manipulating circuit spectrum to achieve the given predefined values over the external flux profile. There is a likelihood to identify multiple solutions.

Zero-Pi circuit and its proposed "next-gen" Prismon network are explored to discover circuits with high degeneracy, from purely numerical methods. While,

the Zero-Pi qubit have known analytic results, numerical discovery should confirm the same. Simultaneously, paving the method applicable to Prismon qubit.

4.1 C-shunted profiling

Capacitively shunted flux qubits are considered to be improvement on transmon circuits. Additional shunted junctions on the parallel branch, provide a 2-d dimensional potential that improves coherence times [40]. Consequently, they have been found to be a better candidate for quantum annealing [41].

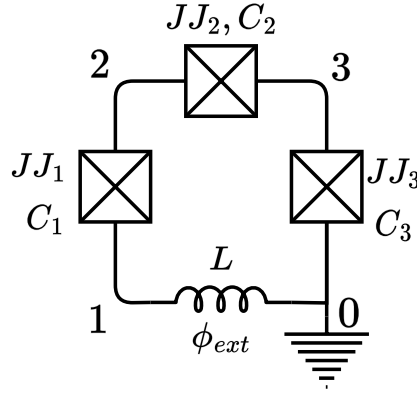


Figure 12: Capacitively shunted flux qubit, with loop flux attached on L .

External loop flux is induced explicitly through an inductor L . Its minimal contribution to the Hamiltonian is kept constant for all the experiments.

Hamiltonian in the node variable description, C_N^{-1} is node capacitance matrix sec.2.2.1:

$$H = \frac{\Phi_1^2}{2L} + \frac{1}{2} \vec{Q}_n^\top C_N^{-1} \vec{Q}_n - \frac{1}{2} (E_{J_1}(\mathcal{D}_{21}^+ + \mathcal{D}_{21}^-) + E_{J_3}(\mathcal{D}_{03}^+ + \mathcal{D}_{03}^-)) \\ - \frac{1}{2} E_{J_2}(\mathcal{D}_{32}^+ \exp\{\iota\phi_{ext}\} + \mathcal{D}_{32}^- \exp\{-\iota\phi_{ext}\})$$

Composed primarily of capacitance and junctions, each node is assigned a charge basis. The simple loop design and the lack of inductors makes any transformation \mathcal{R} non-improving.

4.1.1 Spectrum profiling problem

In this example problem, C-shunted flux qubit is tuned to achieve a specific spectrum profile over external flux. The method of tuning is optimization. Assigning a strict spectrum profile on the shunted qubit, gives a direct control on its operation.

Another important idea explored in this experiment is uniqueness of spectrum. That is, how many circuits share same spectrum profile of Ist and IInd transition over a flux range.

Understanding variation of a spectrum over to fluctuations in circuit parameters. gives freedom to engineering providing alternative. Multiple solutions for the same requirement is helpful in driving the operating parameters out of region of unstability. It could be helpful to prevent spurious chaos [42][43].

On purpose, the loss objective focus only on 3 points on the flux profile $\phi_{ext} = 0, \pi, 2\pi$. They are the extremal points and completely define the respective spectrum. Since the network topology is simple in connection and is unchanging during the optimization, the potential and subsequently the spectrum retains its shape.

Transition spectrum profile E_{ij}^* is taken from the SCILLA experiemnt [4]. It is expected that the results are not exactly reproduced because of the inclusion of inductance and range of exploration.

To study the patterns in loss function and identify diversity in potential solutions, the parameter space is kept in relatively higher E_J, E_C range, and is same for each element of its kind app.F.

Typical for regression problems, a mean square loss(MSE) is assigned against the target spectrum E_{01}^*, E_{12}^*

$$\mathcal{L}[\{\theta\}] = \sum_{\phi=0,\pi,2\pi} \|E_{01}^*(\phi) - E_{01}[\theta](\phi)\|^2 + \sum_{\phi=0,\pi,2\pi} \|E_{12}^*(\phi) - E_{12}[\theta](\phi)\|^2 \quad (17)$$

$$\frac{\partial \mathcal{L}}{\partial E_0} = 2 \sum_{\phi=0,\pi,2\pi} (E_{01}^*(\phi) - E_{01}[\theta](\phi)) \frac{\partial \mathcal{L}}{\partial E_2} = 2 \sum_{\phi=0,\pi,2\pi} (E_{12}^*(\phi) - E_{12}[\theta](\phi)) \quad (18)$$

$$\frac{\partial \mathcal{L}}{\partial E_1} = -2 \sum_{\phi=0,\pi,2\pi} (E_{01}^*(\phi) - E_{01}[\theta](\phi)) - 2 \sum_{\phi=0,\pi,2\pi} (E_{12}^*(\phi) - E_{12}[\theta](\phi)) \quad (19)$$

$$Hess(\mathcal{L})_{E_0, E_1} = 6 \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \quad (20)$$

The gradient is then backpropagated to circuit parameters sec.2.4. Effectively, condition on $\phi = 0$ and $\phi = 2\pi$ are identical, as the Hamiltonian is 2π -periodic in external flux.

The gradient of the MSE function is linear in error. Besides, for Newtonian methods the Hessian is a constant matrix, eq.17 Thus, the updates are linear and so the loss function is well-behaved (non-divergent) at the minima. Nevertheless, this derivation ($\frac{\partial \mathcal{L}}{\partial \lambda}, \frac{\partial^2 \mathcal{L}}{\partial \lambda^2}$) is with respect to the energy spectrum and doesn't ensure the behaviour of the overall loss function over the parameter space. Subsequent derivatives in the chain are governed by the eigenvalue calculations and circuit

description sec.2.4. Also, the constant Hessian matrix renders Ist order(Adam2) and IInd order(lrBFGS) similar. It is a pattern recurring in the loss distribution fig.14,app.F.

The MSE minimization is conducted in the full parameter space and pairwise subspaces. Optimization experiments are controlled with following specification:

- Symmetry constraint is defined with variable sharing, such that $E_{JJ1} = E_{JJ3}$, $E_{C1} = E_{C3}$.
- A special point "init" close to known minima is empirically fixed 10.
 - Fixed point common to all the pairwise subspaces.
 - Centroid of the normal distribution for starting point of different optimizers

Assigning identical initialization to each, optimization methods belonging to different class of algorithms are compared:

optimizer	class	hyperparameters
Adam2 [44]	Gradient descent(Ist order)	step-size : .01, iterations : 5
lrBFGS app.A	Newton(IInd order)	step-size : .1, iterations : 5
Nelder Mead [45]	Simplex minimization	tolerance : 10^{-6}
LBFGS app.A	Newton + Line search	tolerance : 10^{-6}

Table 2: Optimization algorithms in C-shunted profiling

4.1.2 Multiple Solutions

The primary result of this experiment is the demonstration of convergence of the regression problem for circuit Hamiltonian eq.4.1. Importantly, it is shown with both, gradient-based (Ist & IInd order) and non-gradient optimizers. The gradients calculated for the eigenvalues, via backpropagation. It has significant computation speed up app.D.2.

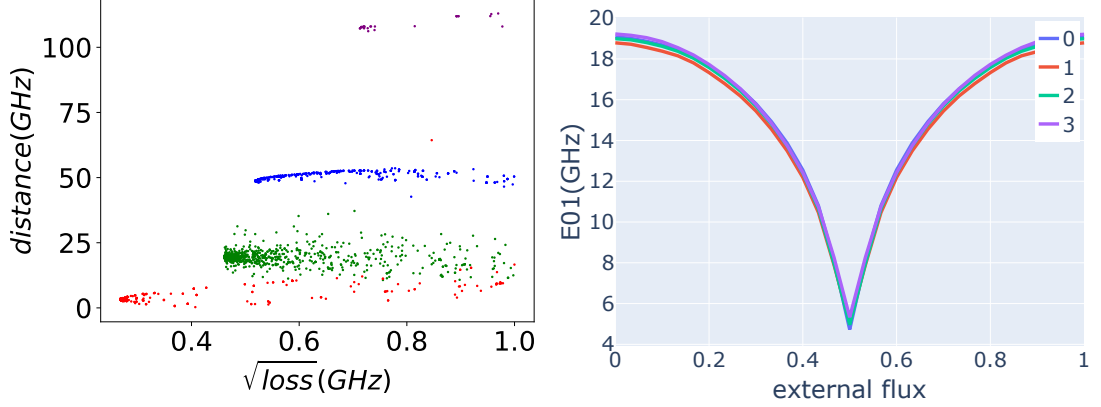


Figure 13: Left: Distance between optimization points and init compared to loss values at those points. Right: 1st excitation profile over external flux for 4 cluster centers. Circuit numbering corresponds to respective colors. Optimization points data adapted from distribution in fig.14

Simple K-means clustering consolidate the scattered points into circuits representative of local minima [46].

circuit	JJ_1	JJ_2	JJ_3	C_1	C_2	C_3	I
init	150	50	150	20	.5	20	800.0
0	119.8	50.0	119.8	2.2	.5	2.2	800
1	150.0	50.0	150.0	17.8	.43	17.8	800
2	136.6	45.7	136.6	20.0	.5	20.0	800
3	150.0	50.0	150.0	96.7	.6	96.7	800

Table 3: Local minima for shunted-Flux qubit profiling, from all the subspace optimization. Circuits 0 – 3 are cluster centers of circuits with loss < 1, fig.13. "init" is the empirical reference point. Energy units are in GHz

There are four distinct local minima of this regression problem:

- C-shunted flux design is robust to large scale fluctuation in circuit parameters.
- Middle branch J2,C2 are the most significant parameters, with least variation.
- Side junction and shunt capacitances are correlated among non-maximal

Searching large parameter space is significantly worse compared to restricting the optimization to pairwise subspaces. High dimensionality exponentially increases the search space domain. For example, maintaining a fixed density, a line needs N points, a square N^2 and so on, with higher grids. Each parameter individually contributes to a higher number of local minima. Further, the interplay

of different parameters builds more saddle points. So, the convexity assumption of the problem is degraded [47].

The order of magnitude of the optimal loss values for subspace optimization $(JJ_1, JJ_2) \sim 10^0$, $(C_1, C_2) \sim 10^3$, $(JJ_1, C_1) \sim 10^2$, have significant separation fig.28. Thus, the full space is optimization is largely, dominated by the (C_1, C_2) subspace.

The spectrum profile of various local minima, with similar MSE loss, are almost overlapping fig.13. It justifies the relaxation on choosing just the extremal, $\phi = 0, \pi, 2\pi$ as unique points for the regression over complete profile.

The loss manifold over capacitances largely have slow gradients. It results from the definition of $E_C \propto 1/C$ and range of the target values E^* . A better parameter scale for initialization would be on capacitance instead of E_C .

4.1.3 Effective optimizer

Comparison of loss distribution for different optimizers fig.14, is a helpful reference to design future experiments:

- Adam2, is an inefficient optimizer for all the subproblems. It has high variance, higher loss and weak convergence, even in the most ideal convex loss profile (JJ_1, JJ_2) , fig.fig.28.
- In (C_1, C_2) subspace, where Adam2 and lrBFGS are equivalent in loss distribution, it is probable that overall Hessian is constant. The improvement in LBFGS is mostly from the line-search.
- Low variance in Nelder-Mead and LBFGS, consistently in each sub-experiment suggest that they lead to same minima. Additionally, fractional MSE loss suggest, that the two optimizer are successful in finding the global minima.

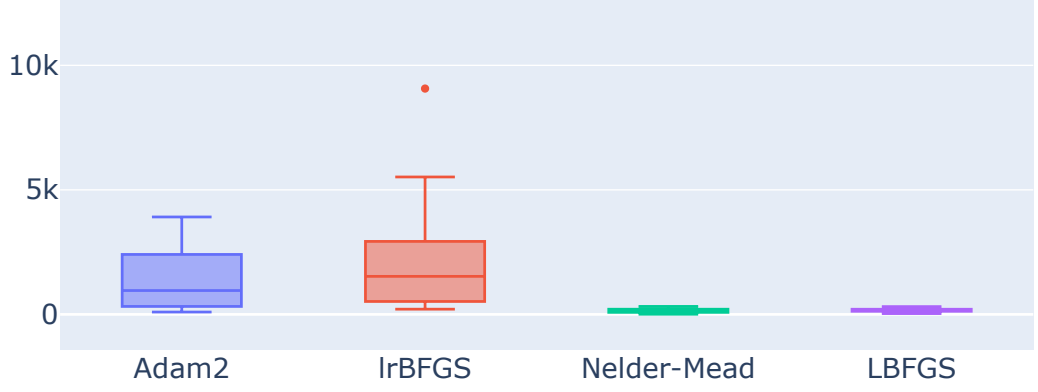


Figure 14: C-shunted profiling $\text{loss}(GH z^2)$ distribution over multiple initialization on full parameter space : JJ_1, JJ_2, C_1, C_2 . Infimum loss are 93.7, 209.1, 5.9, 31.5, respectively. Normal distribution of 50 starting point, centered at init, tab.3.

4.2 Zero-Pi degeneracy

Zero-Pi qubit was proposed to achieve an inbuilt continuous-variable error-correction, against thermal and Hamiltonian noise [48]. The error-correction is termed topological in reference to non-Abelian anyons [49], that is self-restoring with exponential suppression against perturbations. And similarly, utilize on the symmetry of the components to achieve state protection. The robustness of state coherence emerges from the circuit design and further benefits any algorithmic error-correction written on top of it. This terminology also alludes to topological insulators, that are highly resilient to magnetic disturbance, in terms of resistivity [50].

The circuit is composed of two alternating rings, consisting of four nodes. Each ring contains identical elements, a shunted Josephson junction, an inductor and a capacitor, fig.15.

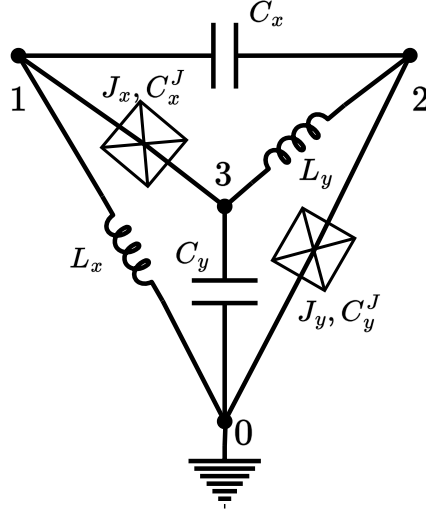


Figure 15: Zero-Pi circuit in a tetrahedral representation. Inductance: $L_x = L_y$, Capacitance: $C_x = C_y$, Junction: $J_x = J_y$, Shunt: $C_x = C_y$.

Identical parameters manifest symmetry in the circuit. This puts many constraint on a low parameter space. Their dynamics is thus expected to exhibit rich topological character. It should also make their design challenging.

Node 0, here, is assigned as a ground. This is a choice of gauge, supported by symmetric geometry without affecting the spectrum and dynamics. And benefits the numerical simulation by reducing the Hilbert space, as ground node is reference to be fixed.

In order to exploit the symmetry constraints to reduce the parameter space, following variable transformation is assigned to the circuit nodes [51]. Expressing the circuit in the node fluxoid variables ϕ_i , and resurrecting the ground node (ϕ_0), momentarily :

$$\begin{bmatrix} \varphi \\ \theta \\ \chi \\ \Omega/2 \end{bmatrix} = \begin{bmatrix} 1 & -1 & -1 & 1 \\ -1 & -1 & -1 & 1 \\ -1 & -1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \phi_0 \\ \phi_1 \\ \phi_2 \\ \phi_3 \end{bmatrix}$$

This "ridge" transformation maximally decouples the nodes, incorporating their connection sharing.

One of the modes Ω , is completely decoupled from the system. It has equal contribution to each node such that, it has no representation in the Hamiltonian. Geometrically it is similar to the ground node, attached to every other node of the circuit. Again, by the geometry, node 0 and node 3 are also interchangeable and result in just a label transposition in the "ridge" transformation.

Fixing the mode Ω , the subsystem displays subsequent redundancy in mode χ . It decouples as an independent oscillatory mode whose Hilbert space is separable.

In this case, the decoupling relies on symmetry of the elements and separability would be violated for minor variations [51].

Finally, Hamiltonian expressed in the two modes φ, θ in flux basis:

$$H = -2E_{C_J}\partial_\varphi^2 - 2E_{C_\Sigma}\partial_\varphi^2 - 2E_J \cos \theta \cos \varphi - \phi_{ext}/2 + E_L\varphi^2 + 2E_J \quad (21)$$

This a two particle system in a 2 dimensional potential, with a quadratic potential in φ and a coupling of φ & θ via sinusoidal potential. Coupling term is 2π periodic in θ, φ , and has a control on the offset with ϕ_{ext} .

4.2.1 Robustness model

In the Bloch-Redfield model of decoherence [13], there are two timescales T_1, T_2 under which the qubit statistics are coherent and assumed to sustain their quantum information.

T_1 describe the state relaxation period, while T_2 define the phase relaxation of a state in superposition state.

The bit-flip relaxation rate Γ is related to thermal fluctuations, shortening $T_1 \propto \Gamma^{-1}$. It depends on the transition energy ΔE between the two levels, as $\Gamma \propto \exp\{-\Delta E/k_B T\}$. Since $\Delta E = E_0 - E_1 < 0$, it is desired to minimize the energy difference for exponential suppression of bit-flip error.

The pure dephasing error emerge from the fluctuation in the qubit frequency $\omega_q = (E_1 - E_0)/\hbar$. The inability to maintain constant phase difference with the rotating frame depolarizes the qubit over time, most significant at the superposition state [19]. This requires $\frac{\partial \omega_q}{\partial \phi_{ext}} \rightarrow 0$. In fact, any other systemic perturbation like node charge n_g could fluctuate the transition energy, as well.

For topological protection, Zero-Pi is required to operate in a regime facilitating the aforementioned requirements.

The computational states of the Zero-Pi qubit are the two lowest energy degenerate states. The robustness of these wavefunctions to dexcitation and dephasing stems from their localization patterns.

In general, the low-energy wavefunction seeks region of lower potential. Along $\theta = 0, \pi$, the symmetric potential(alternated by $\cos \phi$), separated by a potential barrier attempts to localize the wavefunctions $|\psi_g\rangle, |\psi_e\rangle$, in the two "ridges".

By choosing significantly different effective masses $E_{C_J} \gg E_{C_\Sigma}$, the tunneling along θ is reduced in comparison to φ direction.

Further, the oscillator extent in the θ direction is reduced in $E_J \gg E_{C_\Sigma}$ regime. This is equivalent to setting up a high potential barrier.

Localization implies decoupling (non-overlap $\langle \psi_g | \psi_e \rangle = 0$) and subsequently degeneracy, as the two identical ridges could contain anti-symmetric configurations.

Delocalization in ϕ is promoted to stabilize the symmetric potential against the perturbation, such that fluctuations render the wavefunction configuration invariant.

Having enabled a relative easier tunneling along φ , the oscillation extent in the that direction is increased, $E_L \ll E_{C_J}$. Additionally, the parabolic potential should be weakened against the cosine potential, $E_L \ll E_J$. This minimizes the relative asymmetry in the two ridges under ϕ_{ext} fluctuations.

At $\phi_{ext} = \pi$, the two minima, quadratic and cosine in φ coincide. This improves the symmetry

4.2.2 Degeneracy objective

The objective of this experiment is to numerically identify if such a network could exhibit degeneracy. And, further to calculate the circuit parameters that maximizes it. Optimal solution is expected to fall under the suggested analytical regime. This method is agnostic to any physics of localization and learns solely from the Hamiltonian and its diagonalization.

The loss function for this problem is describes relative transition energy to Ist and IInd excited level:

$$\mathcal{D}[\phi] = -\log_{10} \frac{E_2 - E_0}{E_1 - E_0} \quad \mathcal{L} = -\mathcal{D}[\phi = \pi] \quad (22)$$

$$\frac{\partial \mathcal{L}}{\partial E_0} = \frac{E_1 - E_2}{(E_1 - E_0)(E_2 - E_0)} \quad \frac{\partial \mathcal{L}}{\partial E_1} = \frac{E_2 - E_0}{E_1 - E_0} \quad \frac{\partial \mathcal{L}}{\partial E_2} = -\frac{E_1 - E_0}{E_2 - E_0} \quad (23)$$

The derivative of the loss function seems to diverge, approaching degeneracy $E_1 \rightarrow E_0$ in eq.23. It would result in exploding gradient, sec.2.4. However, in LBFGS optimizer the Hessian inverse scales quadratically as $(E_1 - E_0)^2$ and it neutralizes the gradient. Convergence is linear in the vicinity of minima.

The essential symmetry in Zero-Pi circuit elements in manifested through variable sharing, as described earlier sec.2.4.

A noted difficulty in the physical implementation of Zero-Pi circuits is high inductance, $E_J \ll E_{C_J}$ [51],[52]. To solve engineering requirement bounds on circuit parameters are assigned within practically achieved limits app.E.

From the excitation spectra, the crest and troughs suggest possibility of non-convex profile of eigenvalues. The optimization is carried with two optimizers, a gradient-based LBFGS and a pattern-finding Nelder-Mead. Both methods cover the disadvantage of each other.

In the numerical simulation, the circuit is numerically represented in the charge basis, with truncation charge states $q_i \in \{-7, \dots, 0, \dots, 7\}$ for each node $i = 1, 2, 3$.

The choice of charge basis, is motivated by low E_L and consequently lower contribution of inductive energy $\sim E_L \Phi^2$. It is expected to curb numerical inconsistency of representing a Φ operator in charge basis.

Secondly, to describe the de/localization of wavefunctions as in previous analysis, it makes sense to describe them in a canonical basis. Also, the "ridge" transformation is just a rotation in the subspace. It wouldn't distort the localised patterns describing degeneracy in the wavefunction distribution.

Interestingly, the Kerman transformation prescription sec.2.2.6, is to represent this circuit in two oscillator mode and a Josephson mode. However, Josephson operators D^\pm representation in oscillator basis is

It would become important to consider oscillator modes in the case of asymmetric circuit elements, that disfavor χ mode decoupling.

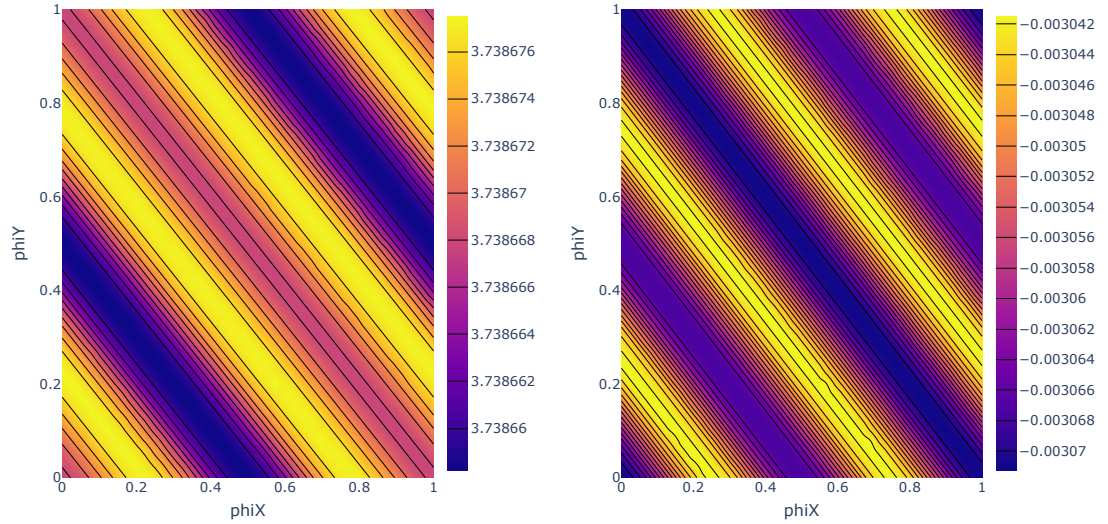


Figure 16: Zero-Pi transition spectrum(E_{01} :left, E_{12} :right) of Zero-Pi circuit, over external flux profile threaded to L_x & L_y . $E_L = 10.0$, $E_C = 50.0$, $E_J = 10.0$, $E_{C_J} = 100.0$. Energy units in GHz.

Exploiting $\phi_x^{ext} + \phi_y^{ext}$ symmetry fig.16, subsequently external flux is profiled only on L_x . In fact, this agile switching could be helpful in manipulation, without any additional overhead.

4.2.3 Optimizer comparison

The distribution of final loss describe LBFGS as a better optimizer. And is able to achieve a complete degeneracy. Their variance is higher suggesting different local minima that are found.

The Nelder-Mead solution have relatively lower degeneracy. Lower variation similarly suggest higher universal convergence, from different starting points.

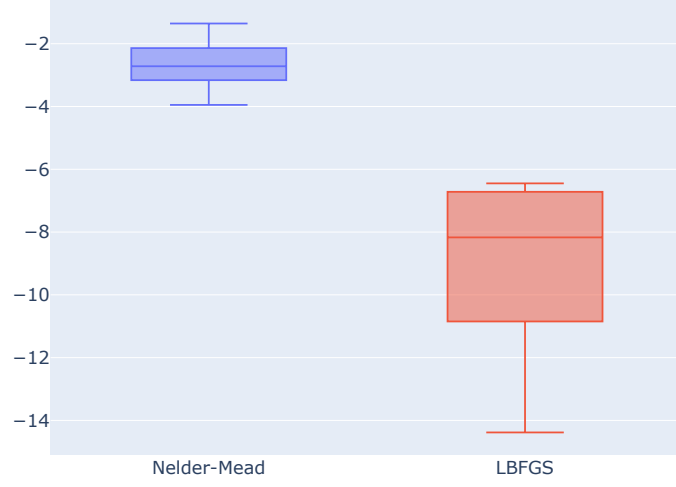


Figure 17: Loss distribution of Zero-Pi degeneracy maximization, on parameter space L_x, C_x, J_x, C_x^J . 15 starting point, normally distributed around init 4, truncated at bounds app.E. Extreme point $\mathcal{D} = \infty$ is excluded from the plot.

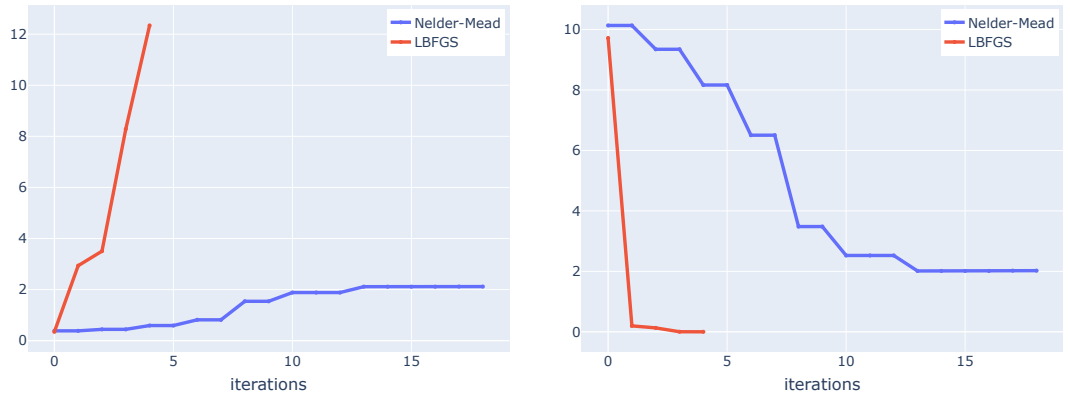


Figure 18: Optimization from sample circuit for Nelder-Mead & LBFGS, at degeneracy point. LBFGS shows truncation at 4th step. Left: Degeneracy improvement. Right: Equivalently, Ist transition energy(E_{01} in GHz) minimizes.

LBFGS display high convergence rate fig.18 and achieves higher degeneracy fig.17 compared to Nelder-Mead. It has almost monotonic descent that suggest that the loss function is mostly convex, and doesn't suffer from misdirection sec.2.4.

Nelder-Mead is slower in convergence and stagnates for a long duration before making any improvement. It makes decision on truncation difficult, both for length and undecidability.

Despite Nelder-Mead runs higher iteration, LBFGS line-search executes multiple eigen-value calculation for a single step. But it is considered as a subroutine of a single step, as the local position to calculate derivatives stays the same.

optimizer - circuit	E_L	E_C	E_J	E_{C_J}
init	10.0	5.0	10.0	50.0
Nelder-Mead - optimal	17.6	2.03	30.0	3.98
LBFGS - optimal	22.01	.38	30.0	27.06
Nelder-Mead - sample	12.7	3.76	29.75	4.4
LBFGS - sample	19.68	0.39	30.0	20.49

Table 4: Zero-Pi: Optimal circuits display highest degeneracy among all instances, for respective algorithms. Sample circuits are solution of optimization, choosing a random initial point. Init point marks the center of initialization point distribution.

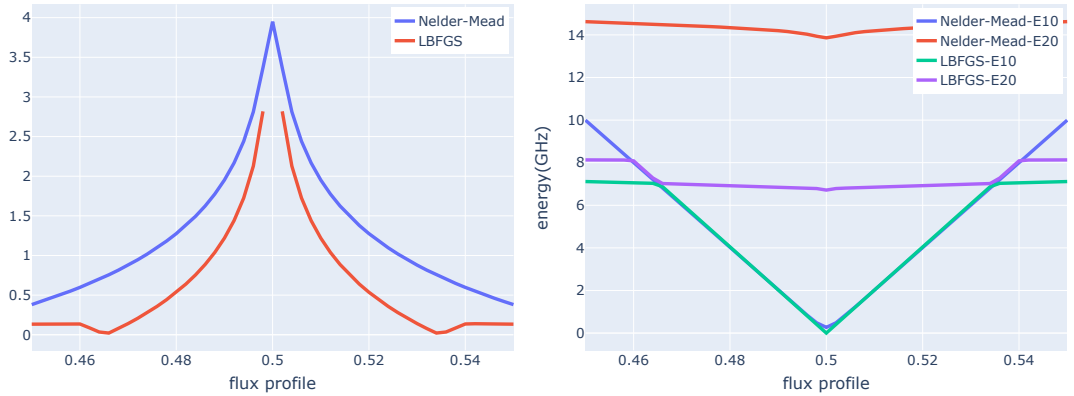


Figure 19: Optimal circuits achieved by Nelder-Mead and LBFGS, in the vicinity of degeneracy point. Left: Degeneracy profile. LBFGS has a singularity at $\phi = \pi$. Right: Transition spectrum.

4.2.4 Optimal Solution

The optimal solutions obtained from the two optimizers clearly discriminate between the conditions of tunneling suppression and delocalization for Zero-Pi robustness.

A ratio comparison of the circuit parameters describe status of optimal results against analytic expectation. Each ratio describe the parameter regime for a particular physical reasons requested previously in the analysis.

Nelder-Mead optimal shows weaker suppression of tunneling, that explains lower degeneracy $\mathcal{D} \sim 4$. It also fails in delocalisation, evidently $\frac{\partial \omega_{01}}{\partial \phi_{ext}} \neq 0$ at $\phi = \pi$, is limiting as a line

LBFGS optimal exhibit strong suppression of tunneling and achieves degeneracy $\mathcal{D} = \infty$. It is expected to have weak delocalization in φ direction. Importantly, the inductance is well under practical bounds.

physics	condition	LBFGS	Nelder-Mead
tunneling	$E_{C_J} \gg E_{C\Sigma}$	72.21	2.96
tunneling	$E_J \gg E_C$	78.95	14.78
delocalization	$E_J \gg E_L$	1.36	1.76
delocalization	$E_{C_J} \gg E_L$	1.23	.23

Table 5: Analytic conditions for Zero-Pi degeneracy. Ratio of condition's parameter comparison for optimal solutions of respective algorithms. Analytic methods suggest the ratio $\gtrsim 100$ [51].

The loss function is aimed to minimize just the $\Delta E = E_1 - E_0$ and is agnostic to magnetic perturbations to $\frac{\partial \omega_{01}}{\partial \phi_{ext}} \neq 0$. This explain the defeat of delocalization in both the cases, significantly more in Nelder-Mead fig.20.

A careful design of multi-objective loss function would tackle both the conditions simultaneously. It should result optimal solution in the expected analytic regime.

However, it is tough to numerically calculate the magnetic susceptibility $\frac{\partial \omega_{01}}{\partial \phi_{ext}}$. As degeneracy increases the derivative might be undefinable subjected to smallness of $d\phi$. Symbolic differentiation could be speculated, instead.

Despite, the optimal solution show expected degeneracy, their spectrum profile displays surprising linearity in the vicinity of $\phi = \pi$.

The wavefunction distribution shows decoupled regions along one of the diagonals, fig.20. It corresponds to the θ direction, from the rotation of variables. The little delocalization (stretched out probability density) for LBFGS-optimal, in the φ direction, would improve with additional perturbation suppression terms in the loss function.

4.3 Prismon symmetry

The topological protection featured in Zero-Pi circuit originates from its symmetric geometry. High network symmetry decouple several degree of freedom. Thus, rich spectrum features are built with lesser parameters.

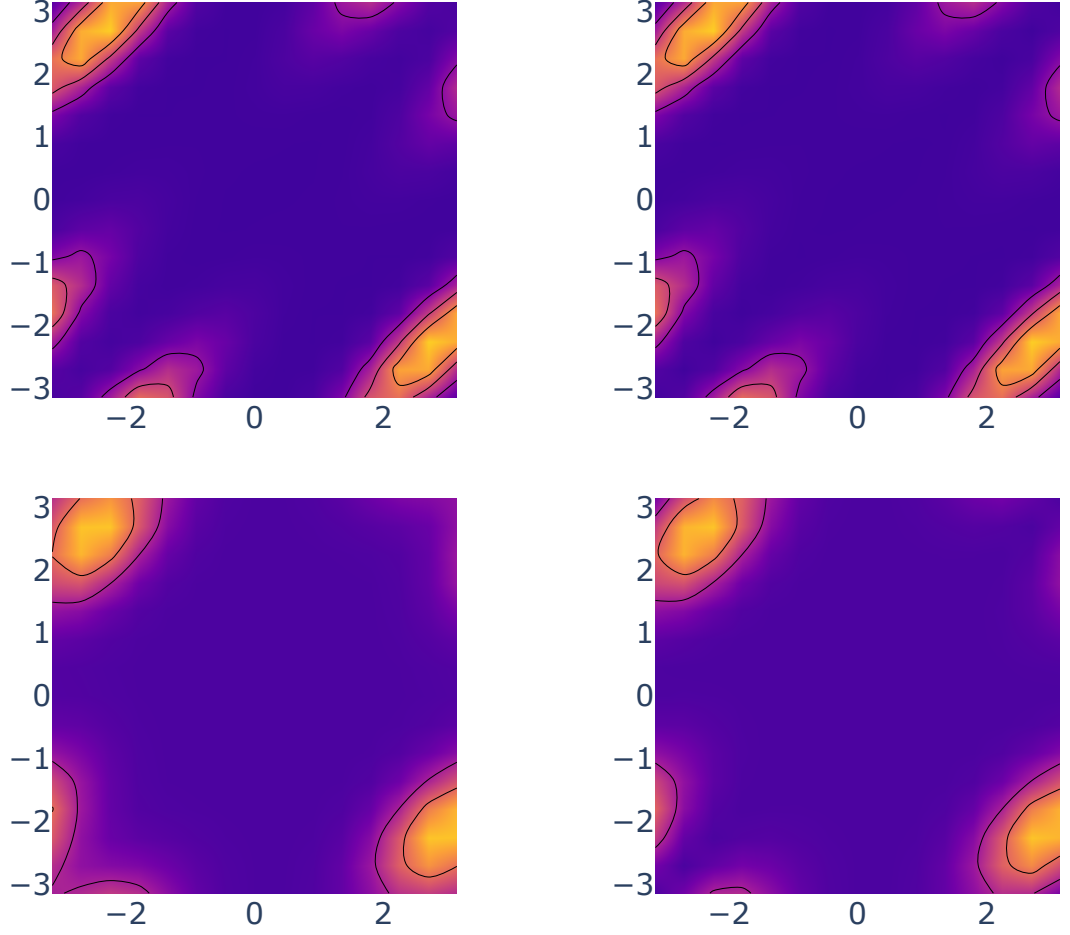


Figure 20: Ground(left), Excited(right) $|\psi|^2$ of optimal solutions of LBFGS(top) and Nelder Mead(bottom). Wavefunction ψ represented in flux basis, post Fourier transformation. X-axis: Node $\phi_0 \in [-\pi, \pi]$. Y-axis: Node $\phi_1 \in [-\pi, \pi]$. Node $\phi_2 = 0$ is fixed. This circuit corresponds to low E_L , 4(b) solution in [51].

Following abstractions are inspired from the Zero-Pi, and potentially extended to different circuits:

- Each node is joined by only three elements, L, C, J
- Same elements have symmetric characteristic energy
- Placement of ground node has symmetric combinations, such as $0 \equiv 1 \equiv 2 \not\equiv 3$

The class of symmetric regular polyhedra generally satisfies these properties.

The next simplest geometry outside the class of polyhedra, a prism is hypothesized for a topological qubit: Prismon fig.21. Prismon network, is composed of 2 triangular loops, separated by 3 square loops, along the length of the prism.

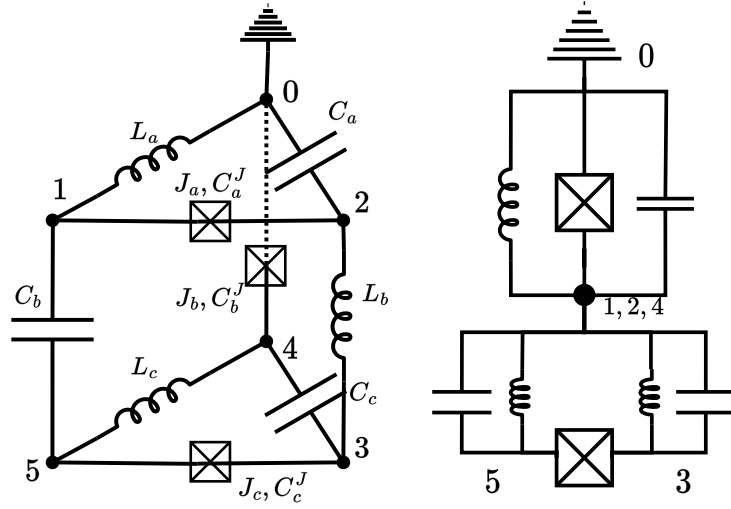


Figure 21: Prismon circuit(left). Each node vertex is joined by identical elements. Placement of ground has symmetric choices; here, its junction-opposite. Hypothetical collapsed configuration(right), making node 0 as anharmonic oscillator.

The Hamiltonian of the Prismon, in the node coordinates $\vec{\phi}_{node}$; referencing ground to null, external flux is split between J_a, J_c :

$$\begin{aligned}
 H = & 4E_{C_J} ((n_1 - n_0)^2 + (n_4 - n_2)^2 + (n_3)^2) \\
 & - E_J (\cos(\phi_1 - \phi_0 + \phi_{ext}/2) + \cos(\phi_4 - \phi_2 + \phi_{ext}/2) + \cos(\phi_3)) \\
 & + 4E_C ((n_0 - n_4)^2 + (n_3 - n_2)^2 + (n_1)^2) \\
 & + \frac{1}{2}E_L ((\phi_0)^2 + (\phi_1 - \phi_2)^2 + (\phi_3 - \phi_4)^2)
 \end{aligned}$$

The spectrum profile of Prismon, fig.22, has more variation compared to Zero-Pi, fig.16, respectively for E_{01} & E_{12} . There are two additional minima at around $\phi = \pi/2$ and $\phi = 3\pi/2$, symmetrically opposite about minima at $\phi = \pi$. It

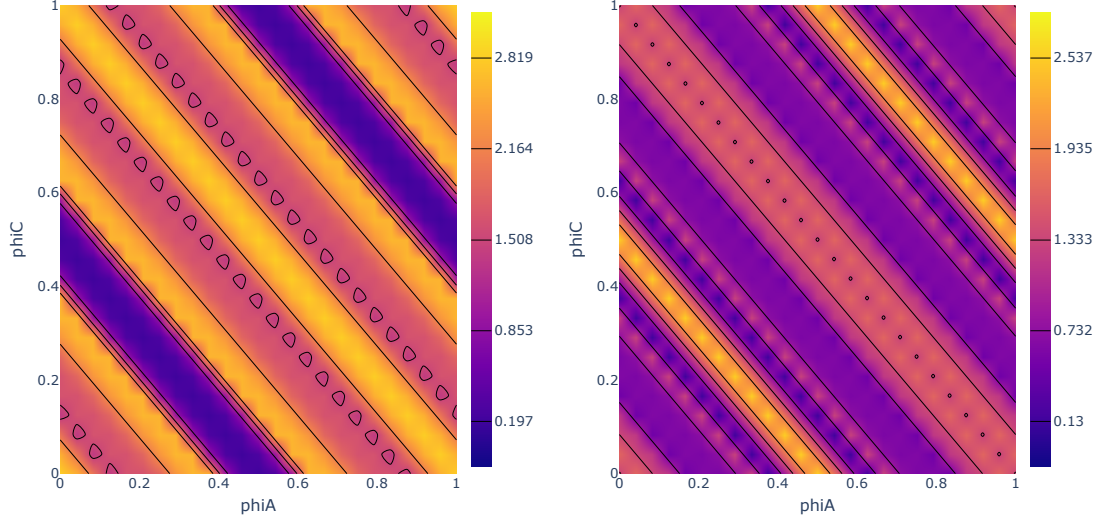


Figure 22: Primon transition spectrum(E_{01} :left, E_{12} :right) of Zero-Pi circuit, over external flux profile threaded to L_a & L_c . $E_L = 10.0$, $E_C = 50.0$, $E_J = 10.0$, $E_{C_J} = 100.0$. Energy units in GHz.

might be possible to probe and manipulate these segregation for protected higher computational states.

Subsequently, as in Zero-Pi, with $\phi_a^{ext} + \phi_c^{ext}$ symmetry in fig.22, external flux profiled is sufficiently reduced to L_a .

4.3.1 Minima Minimization

In this section, the objective is to demonstrate degeneracy at the symmetric point $\phi = \pi$. It is suggestive from the minima of the spectrum at $\phi_A + \phi_C = \pi$, that the network could support a degeneracy at that point. The goal of the optimization is then to maximize its depth.

However, the Primon is bulkier compared to Zero-Pi, with 5 nodes, despite, the parameter space is same. And simplifying transformation of the circuit nodes might be not be straightforward.

So, in the absence of analytic simplification, demonstration of degeneracy in such a circuit is achieved by searching the parameter space for such a property.

Another motivation to discover topological qubits, is weak controllability of existing Zero-Pi [53]. For the reasons that establish high robustness, state manipulation similarly becomes challenging.

By any means, degeneracy is not a sufficient condition for topological protection, but a necessary one. Subsequent modal analysis of wavefunction, with respect to potential profile should prove gain in coherence time.

Primon is numerically modeled in a mixed basis. Since, it is unclear how to segregate modes from each from algebraic transformation, Kerman transforma-

tion is applied apriori eq.24.

There are three oscillator modes truncated at 4th level $|o_i\rangle \in \{0, 1, 2, 3\}$, $i = 1, 2, 3$, and two Josephson modes in charge basis, truncated at 5 Cooper pairs $|j_i\rangle \in \{-5, \dots, 5\}$ sec.2.2.6.

$$\sqrt{2}\vec{\phi} = \mathcal{R}\vec{\phi}_{node} \quad \mathcal{R} = \begin{bmatrix} 0 & 0 & \sqrt{2} & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (24)$$

4.3.2 Collapse configuration

Principally, the Kerman transformation express the circuit in terms of reduced non-linear oscillators. A single node anharmonic oscillator has three branch with a capacitor, inductor and junction, respectively [54].

Prismon could realize six non-linear oscillators as a subsystem, assuming certain constraints on the other nodes.

For example, if $\phi_1 = \phi_2 = \phi_4$ in a particular state, then the nodes 1, 2, 4 could be collapsed and a non-linear oscillator emerge at node 0. The state of J_a reaches lowest energy. The subsequent network is a circuit with two harmonic oscillator, coupled by a junction. The system exhibit identical dynamics if node 4 represents the anharmonic oscillator and nodes 0, 5, 3 collapse.

Similarly, situating anharmonic oscillators on other nodes result in different subsequent circuits. These subsystems have different branch pairs coupled via capacitor or inductor. However, all other collapsed prismon have symmetric counterpart from its adjascent node along the length of the prism.

In total, there are three pairs of symmetric configurations. These configurations are achieved only under their respective constraints. Symmetry renders equal energy to these constrained states.

The two conditions of a symmetric pair could not be smoothly connected, unless all the nodes collapse. Since, they involve the complementary set of variables. The switching of constraints act as a barrier, in between a pair of degenerate states. The energy difference between these configuration determine the suppression of tunneling between two degenerate states.

In order to achieve such configurations, the corresponding energies should be minimized. For example, in previous example of node 0 anharmonic oscillator, the collapse of J_a should be spontaneous compared to other configurations.

It is important to realize that such constrained state might not necessarily occupy the lowest levels. However, the circuit parameters could be arranged to

bias one anharmonic node versus the other two. And, also to overall bring down the energy levels of constrained states.

Therefore, the proposed configuration is a good enough reason to probe the degeneracy in a highly symmetric circuit, like this.

4.3.3 Results

The two optimizers Nelder-Mead & LBFGS are implemented to reduce the loss function of degeneracy for Prismon eq.22. The importance of Nelder-Mead is more than the previous example of Zero-Pi. Since, loss function is expected to be more non-convex, with the addition of more loops. Consequently, the results fig.23, record relative improvement of Nelder-Mead fig.17.

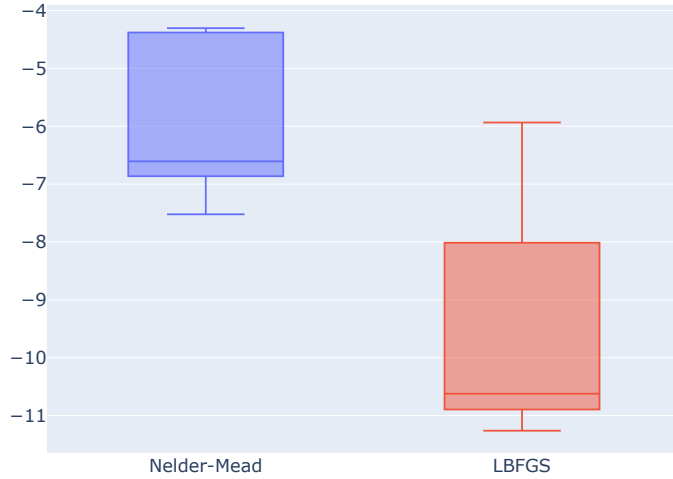


Figure 23: Loss distribution of Prismon degeneracy maximization, on parameter space L_a, C_a, J_a, C_a^J . 15 starting point, normally distributed around init 6, truncated at bounds app.E

The two optimal solution are non-local, that is, there is sufficient difference in their parameter state. As well as, none of the parameters are dominating on each other.

The LBFGS-optimal is unexplored in E_J, E_{C_J} with its proximity to the centroid init. High density of minima could restrict subspace exploration.

optimizer - circuit	E_L	E_C	E_J	E_{C_J}
init	10.0	5.0	10.0	50.0
Nelder-Mead - optimal	12.78	9.41	5.63	28.86
LBFGS - optimal	5.35	21.91	10.77	48.69

Table 6: Prismon: Optimal circuits display highest degeneracy among all instances, for respective algorithms. Init is the centroid of the distribution of the initial points.

In the previous example of Zero-Pi, degeneracy was exactly achieved $\mathcal{D} = \infty$ at a single point in the flux profile $\phi = \pi$. However, PrismoN LBFGS solution doesn't display exact degeneracy, though high $\mathcal{D} = 11.26$ and also has peak at $\phi = 0$ with $\mathcal{D} = 5.71$, taken from best instance of LBFGS.

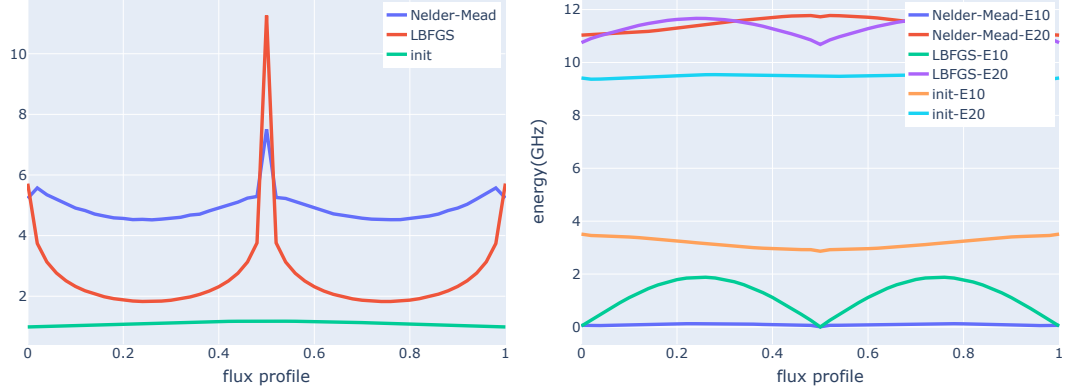


Figure 24: Optimal circuits achieved by Nelder-Mead and LBFGS, over full range of ϕ^{ext} . Top: Degeneracy profile. Bottom: Transition spectrum.

An interesting development in the PrismoN degeneracy optimization is the presence of another peak at $\phi = 0$ fig.24.

It is, however, a little inwards for the Nelder-Mead optimal. This implies that it could be manipulated along the ϕ_{ext} axis. Importantly, the minor peak is not mentioned in the loss function. Its growth is correlated to the $\phi = \pi$ peak.

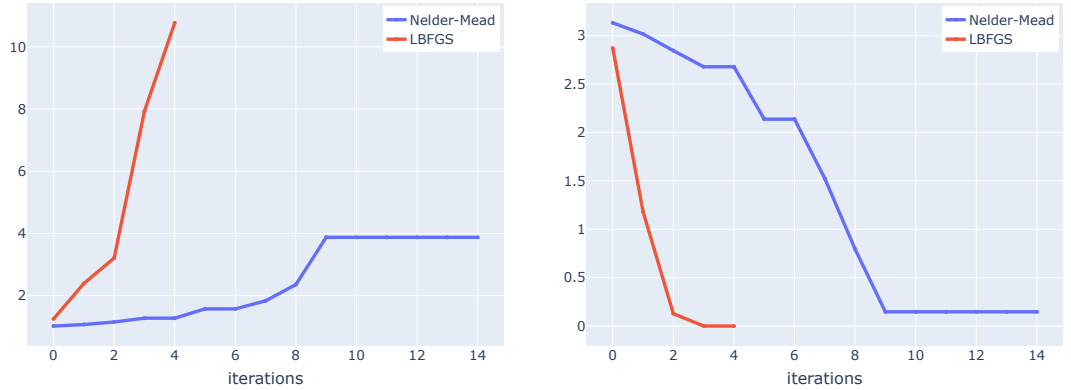


Figure 25: Optimization from sample circuit for Nelder-Mead & LBFGS, at degeneracy point. LBFGS shows truncation at 2nd step. Left: Degeneracy improvement. Right: Equivalently, 1st transition energy (E_{01} in GHz) minimizes.

As before, the convergence of LBFGS is linear near local minima. And the Nelder-Mead has poor convergence, with prominent stagnation. It could be improved with adaptive hyperparameters [55].

Finally, the evidence of symmetric configurations could be checked with the fulfilment associated with their conditions.

The ground $|\psi_g\rangle$ and the excited $|\psi_e\rangle$ state of the optimal-LBFGS circuit is yield following expectation values of the charge variable on the physical nodes:

$$\begin{aligned}\langle\psi_g|Q_0|\psi_g\rangle &= \langle\psi_g|Q_1|\psi_g\rangle = \langle\psi_g|Q_5|\psi_g\rangle = 0 \\ \langle\psi_e|Q_0|\psi_e\rangle &= \langle\psi_e|Q_1|\psi_e\rangle = \langle\psi_e|Q_5|\psi_e\rangle = 0 \\ \langle\psi_g|Q_i|\psi_g\rangle &= -\langle\psi_e|Q_i|\psi_e\rangle \quad i = 2, 3, 4\end{aligned}$$

It is evident that the calculated lowest levels don't fulfill the collapse conditions. The circuit attains a different configuration suitable to the choice of optimal parameters 6.

Despite, reaching a high metric \mathcal{D} , it could not be ascertained if the Prismo could show complete degeneracy for the lowest two levels. Not necessarily on the expected collapsed configuration. In any case, the limited parameter regime of optimizers' initialization rules out the disproval of the hypothesis.

The Prismo surely contain rich features, and is too complicated to be solved analytically. And a small set of parameters could manipulate a lot of its features.

Despite, this preliminary work on automated discovery remain inconclusive. It could be assured with larger experiment and diverse initialization grid that scales the confidence. Secondly, to push the constrained configurations to lower levels, special regimes and initialization should be biased into the optimizer, for guidance. As well as, loss function enforcing the constraints, eq.4.3.3

5 Conclusion

In the present NISQ era, the demand for better quantum computing infrastructure is ever-increasing.

The quantum systems with few (2–10) degrees of freedom, lie in the nightmare zone for analysis. It is not simple enough for basic transformation/approximation, nor large enough for statistical field treatment. Interestingly, this zone is also rich enough for applications and small enough to manover.

In the lack of simplification strategy, the proposition of this thesis, is to explore these systems numerically. Optimization techniques inspired from machine learning provide a systematic alternative to the brute-force approach.

Further, aspects of transmon analysis is argued to be a counterexample for the asymptotic regime to study superconducting circuits. And thereby, unfit for practical implementation, in general.

Superconducting circuits are formulated as system of particles in a quadratic/cosine potential. These degrees of freedom are coupled by various electric components, residing over the superconducting wire segments. Their interaction is constrained by a macroscopic flux quantization condition and current conservation. The equation of motion is solved by numerically diagonalizing the Hamiltonian in a truncated representation.

Efficient computation techniques are combined to simulate & optimize circuit dynamics. These sub-routines culminated into a dedicated programming library, *DiSuQ*.

This framework is then iteratively utilized by optimization methods for different problems. The shunted-Flux qubit is identified with multiple solutions to match a fixed spectrum profile, Zero-Pi is operating regime is found numerically and the "Prismon" exhibits the required degeneracy. The objective of automated discovery is successfully demonstrated through these examples.

The appropriateness of numerical discovery rests on the encoding of circuit physics. Empirical results suggest parameter specific representation of the truncated Hamiltonian.

The utilized toolbox and the techniques discussed in this thesis would support the engineering of next generation qubit & control designs.

A Newtonian Optimization

A Newton-based method utilizes both first & second order derivative at the iteration point [56]. It assumes a second degree polynomial h_i at each i th step and estimates the direction of local minima. Locally calculated gradient $\nabla f(x_i)$ and Hessian $\nabla^2 f(x_i)$ uniquely determine the second degree polynomial.

$$f(x + \Delta x) \sim f(x) + \Delta x \nabla f(x) + \frac{1}{2} \Delta^2 x \nabla^2 f(x) + \mathcal{O}(2)$$

$$h_i(\Delta x) = f(x) + \Delta x \nabla f(x) + \frac{1}{2} \Delta^2 x \nabla^2 f(x) \quad \frac{\partial h_i(\Delta x)}{\partial \Delta x} = \nabla f(x) + \Delta x \nabla^2 f(x)$$

$$\Delta x = -\nabla^2 f(x)^{-1} \nabla f(x) \Rightarrow x_{i+1} = x_i - \alpha \nabla^2 f(x_i)^{-1} \nabla f(x)$$

The choice of step size α , varies with different algorithms.

Exact line-search method enforces a strong condition. It modifies the multi-dimensional search into a single parameter minimization problem. Most of the line-search methods are able to bypass local minima, as well.

The rate of convergence ($f_i = f(x_i)$) for Quasi-Newton method, under convex curvature is super-linear(or Q-superlinear) [57]. For $C > 0, \alpha > 1$

$$\frac{|f_{n+1} - f|}{|f_n - f|^\alpha} = C$$

, $\alpha = 2$ is defined as quadratic convergence.

A less restrictive definition R-linear, allows oscillatory behaviour of the convergent series. It is defined with auxiliary series as the upper bound. It is helpful in stochastic and noisy convergence of BFGS [58].

A.1 BFGS

BFGS is a popular implementation for Newtonian optimization that speeds up the calculation of Hessian; the L-BFGS is its low memory version [56]. Instead of online evaluation, the Hessian is estimated from the stored history. The algorithm is stated as follows:

1. Initialization: starting at x_0 , $H_0 = 1$
2. Search direction p_k : calculate ∇f_k , $H_k p_k = -\nabla f_k$
3. Line search : $\alpha_k = \text{argmin}_\alpha f(x_k + \alpha p_k)$
4. Update parameters : $\Delta x_k = \alpha_k p_k \Rightarrow x_{k+1} = x_k + \Delta x_k$
5. Estimate Hessian : $H_{k+1} = H_k + \frac{y_k y_k^\top}{y_k^\top \Delta x_k} - \frac{H_k \Delta x_k \Delta x_k^\top H_k^\top}{\Delta x_k^\top H_k \Delta x_k}$
 $y_k = \nabla f_{k+1} - \nabla f_k$

"lrBFGS" is a Newtonian optimization, that takes a fixed step size α in place of linear search. It is constructed as a benchmark variant against line search.

A.2 Error Accumulation

In BFGS A.1, Hessian estimation exploits the knowledge of the gradient of loss function. The computational error(ϵ) in gradient evaluation accumulates into Hessian matrix, resulting in larger deviation in the search direction. Subsequently, the drift from the ideal(exact calculation of gradient) snowballs with every iteration.

In the following, semi-qualitative growth of the drift(δ_k) is analysed, $\tilde{x}_k = x_k \pm \delta_k$. For simplicity, assume a linear order correction to the error-induced trajectory(\tilde{x}_k) of optimization. And an upper bound on the computation error of the gradient as, $\|\nabla f(x_k) - \tilde{\nabla} f(x_k)\| < \epsilon$.

The L-Lipschitz gradient function assigns an upper bound on the gradient deviation, combining computational and historical errors:

$$\|\nabla f(x_k) - \nabla f(x_k \pm \delta_k)\| < L\delta_k \Rightarrow \|\nabla f(x_k) - \tilde{\nabla} f(\tilde{x}_k)\| < \epsilon + L\delta_k$$

Consider the following definition of deviation(linearized) in the updates of BFGS, $\tilde{x}_k, \tilde{H}_k, \tilde{y}_k, \tilde{\nabla} f_k$:

$$\begin{aligned} d\Delta x_k &= (x_{k+1} - x_k) - (\tilde{x}_{k+1} - \tilde{x}_k) & dH_k &= H_k - \tilde{H}_k & d\nabla f_k &= \nabla f(x_k) - \tilde{\nabla} f(\tilde{x}_k) \\ dy_k &= (\nabla f_{k+1} - \nabla f_k) - (\tilde{\nabla} f_{k+1} - \tilde{\nabla} f_k) & \|\tilde{y}_k\| &< 2\epsilon + L(\delta_{k+1} + \delta_k) \end{aligned}$$

Differential expansion of BFGS updates A.1 results in the linearized deviation $d\Delta x_k \sim \delta_{k+1} + \delta_k$:

$$\begin{aligned} d\Delta x_k &= -\alpha_k \left[\frac{\partial H_k^{-1} \cdot \nabla f_k}{\partial H_k} dH_k + \frac{\partial H_k^{-1} \cdot \nabla f_k}{\partial \nabla f_k} d\nabla f_k \right] \\ dH_{k+1} &= \frac{\partial H_{k+1}}{\partial \Delta x_k} d\Delta x_k + \frac{\partial H_{k+1}}{\partial y_k} dy_k + \frac{\partial H_{k+1}}{\partial H_k} dH_k \end{aligned}$$

To identify the maximal deviation due to computation error, an upper bound on the $\|\delta_k\|$ is estimated. For simplification, the complicated tensor derivatives are reduced to respective norms. Subsequent expansion of the norm under multiplication and addition is bounded via Cauchy-Schwartz and Triangle inequalities,

respectively.

$$\begin{aligned}
\|d\Delta x_k\| &< \left[\|H_k^{-1}\|^2 \|\nabla f_k\| + \|H_k^{-1}\| \left\| \frac{\partial H_k^{-1}}{\partial y_{k-1}} \right\| \right] \left[\left\| \frac{\partial H_k}{\partial \Delta x_{k-1}} \right\| \|d\Delta x_{k+1}\| \right. \\
&\quad \left. + \left\| \frac{\partial H_k}{\partial y_{k-1}} \right\| \|dy_{k-1}\| + \left\| \frac{\partial H_k}{\partial H_{k-1}} \right\| \|dH_{k-1}\| \right] \quad (25) \\
&\quad + \left[\|H_k^{-1}\| + \|H_k^{-1}\|^2 \|\nabla f_k\| \left\| \frac{\partial H_k}{\partial y_{k-1}} \right\| \right] \|d\nabla f_k\|
\end{aligned}$$

Substituting the starting point into the error accumulation of the updates(eq.25):

1. $\delta_0 = 0, \quad \|d\nabla f_k\| < \epsilon, \quad dH_0 = 0, \quad \Delta x_0 = \delta_1 \Rightarrow \|\Delta x_0\| < \epsilon, \quad \|dy_0\| < \epsilon$
2. $\|d\nabla f_1\| < \epsilon + L\epsilon, \quad \|dH_1\| < \left(\left\| \frac{\partial H_1}{\partial \Delta x_0} \right\| + \left\| \frac{\partial H_1^{-1}}{\partial y_0} \right\| \right) \epsilon \Rightarrow \|dy_1\| < 2\epsilon + \|d\Delta x_1\| L$

$$\begin{aligned}
&\Rightarrow \|d\Delta x_1\| \\
&< \|H_1^{-1}\| \epsilon \left[2 + L + \left\| \frac{\partial H_k^{-1}}{\partial y_{k-1}} \right\| \left\| \frac{\partial H_k}{\partial \Delta x_{k-1}} \right\| \right. \\
&\quad \left. + \|\nabla f_k\| \|H_1^{-1}\| \left((2 + L) \left\| \frac{\partial H_k}{\partial y_{k-1}} \right\| + \left\| \frac{\partial H_k}{\partial \Delta x_{k-1}} \right\| \right) \right] \quad (26)
\end{aligned}$$

Computation of error bounds for higher updates is intractable. However, the deviation bounds are linear in computational error- $\|d\Delta x_k\|, \|dH_k\| < \mathcal{O}(\epsilon)$. In this analysis, the leading order contribution due to convexity is exponential:

$$\|dH_k\|, \|d\Delta x_k\| < \epsilon L^k \prod_{i=1}^k \|H_k^{-1}\|^2 \|\nabla f_k\| \left\| \frac{\partial H_k}{\partial y_{k-1}} \right\|, \quad k > 1 \quad (27)$$

B Spanning Tree - Gauge invariance

Gauge invariance implies that the Hamiltonian developed from either choice of closure branch are unitarily equivalent.

Unitary transformation $T = e^{iQ\phi_{ext}}$, acts a translation operator to the eigenbasis of flux operator $T|\phi\rangle \rightarrow |\phi + \phi_{ext}\rangle$. Subsequently, operators defined diagonally in the fluxoid eigen-basis could be assigned an offset ϕ_{ext} :

$$\begin{aligned}
\hat{O} &= \sum_{\phi} f(\phi) |\phi\rangle \langle \phi| \quad \hat{O} |\phi\rangle = f(\phi) |\phi\rangle \\
\Rightarrow T\hat{O}T^\dagger &= \sum_{\phi} f(\phi) T |\phi\rangle \langle \phi| T^\dagger \Rightarrow \sum_{\phi} f(\phi) |\phi + \phi_{ext}\rangle \langle \phi + \phi_{ext}| \\
&\Rightarrow \sum_{\phi} f(\phi - \phi_{ext}) |\phi\rangle \langle \phi|
\end{aligned}$$

In the finite dimensional representation, the last step remains incomplete with open boundaries, $\phi \in [-\phi_{max}, \phi_{max}]$.

Applying this transformation to the i^{th} node of the circuit Hamiltonian results in a shift of origin : $T_i \hat{\phi}_i^n T_i^\dagger \rightarrow (\hat{\phi}_i - \phi_{ext})^n$, for any polynomial of order n . Independently, each node fluxoid variables have a flexible origin/offset under $T(\vec{\phi}_{ext}) = T_1 \otimes T_2 \dots$. And since the transformation is unitary $T^\dagger T = \mathbf{1}$, the Hamiltonian spectrum is invariant.

Consider a general circuit loop of N elements. Each branch $\langle i, i+1 \rangle; i \in \{0, \dots (N-1)\}$ under modulo N , is assigned a branch flux $\phi_i - \phi_{i+1}$, with closure branch $\langle k, k+1 \rangle$ additionally appended with loop flux ϕ_{ext} . Assuming counter-clock ordering, ϕ_{ext} could be transferred to the next the branch $\langle k-1, k \rangle$ under $T_k(\phi_{ext}) : \phi_k \rightarrow \phi_k - \phi_{ext}$.

Iteratively, each branch could be assigned as the closure branch, again with invariant spectrum. This substitution method, without the loss of generality, utilizing a unitary transformation renders a gauge freedom in the choice of closure branch and spanning tree, overall.

C Eigenvalue derivates

Eigenvalues(λ_k) have smooth dependence on the Hamiltonian matrix(H). Its derivative could be representated in terms of co-factor matrices(Co_{ij}), Expanding the determinant of H :

$$\prod_k \lambda_k = \det|H| = \sum_{ij} H_{ij} Co_{ij} \Rightarrow \frac{\partial \lambda_k}{\partial H_{ij}} = \frac{Co_{ij}}{\prod_{i \neq k} \lambda_i}$$

D Benchmarks

D.1 Framework

At present there are several implementation on superconducting circuit modeling. Following table provides qualitative comparison:

	backend	diagonalization	modes	specialization
quCAT [17]	dense	Polynomial	oscillator	visualization
circuitQ [59]	symbolic	Lanczos	periodicity	simulation
scQubit [21]	symbolic	Hierarchical	geometric	analysis
DiSuQ	symbolic	Lanczos	oscillator	optimization

Table 7: Libraries differ in implementation in terms of data structure and computation(backend), digonalization algorithm of their Hamiltonian, transforamtion of nodes(modes) for various utiltlities. Definition of modes should be referred to original implementation.

D.2 Computational

Comparison between *pyTorch* and *numpy* on computational benchmark is conducted for the ubiquitous diagonalization problem. A fluxonium circuit Hamiltonian in increasing truncation of oscillator basis is diagonalized.

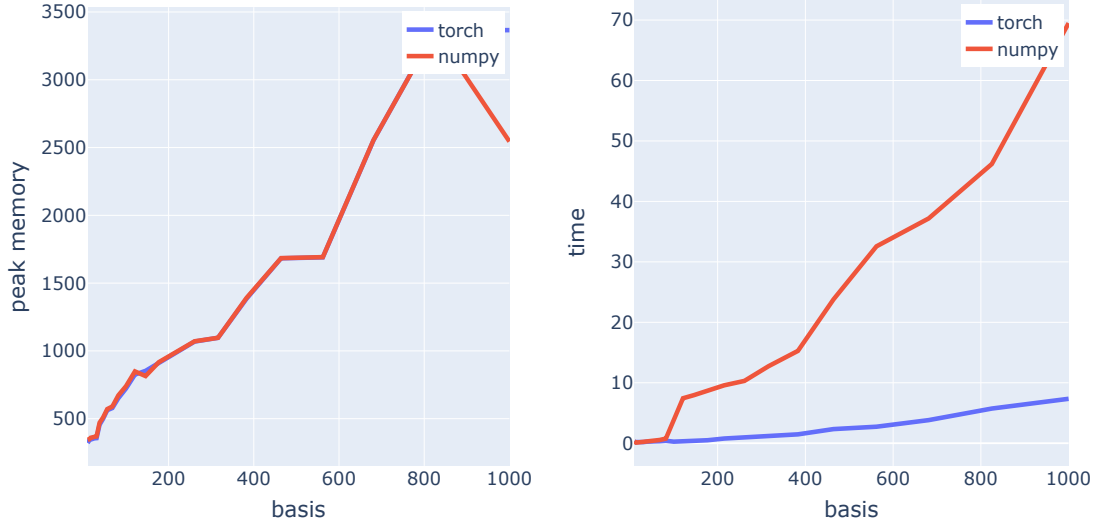


Figure 26: Diagonalization problem comparison between torch and numpy. Left: Memory(MB) consumption at peak. Right: Processing time(seconds) excluding initialization. X-axis marks that size of truncated Hamiltonian.

pyTorch displays significant advantage because the diagonalizing algorithm LOBPCG have faster operation on multi-core distributed computers [60].

LOBPCG 2.3 is compiled with BLAS(low level subroutine for faster linear algebraic operations) which is system specific. So the results via *scipy*, *numpy* might differ slightly from that of *pyTorch*.

E Circuit Bounds

Following tables provides optimization bounds on circuit elements, for C-shunted profiling 4.1 and Zero-Pi, Prismon degeneracy maximization 4.2, 4.3:

units	capacitance	inductance	junction	junction-shunt
lower	.387	1.635	5.0	3.874
upper	38.74	32.69	25.0	154.96

Table 8: Values defined in energy units GHz, $E = \hbar\omega$

units	capacitance	inductance	junction	junction-shunt
lower	.5	5	5.0	.125
upper	50	100	25.0	20

Table 9: Values defined in SI units, fF, nH, GHz, fF respectively

F Shunted-Qubit profile : Optimization

The optimization of Shunted Qubit(4.1)

parameter	lower	upper	static
L			800
JJ_1	10^{-6}	1200	150
JJ_2	10^{-6}	1200	50
C_{11}	10^{-6}	2500	20
C_{12}	10^{-6}	2500	.5

Table 10: Parameter space specification. Energy units are in GHz.



Figure 27: Loss distribution on parameter subspaces : (JJ_1, C_1) , (C_1, C_2) & (JJ_1, JJ_2) , in order.

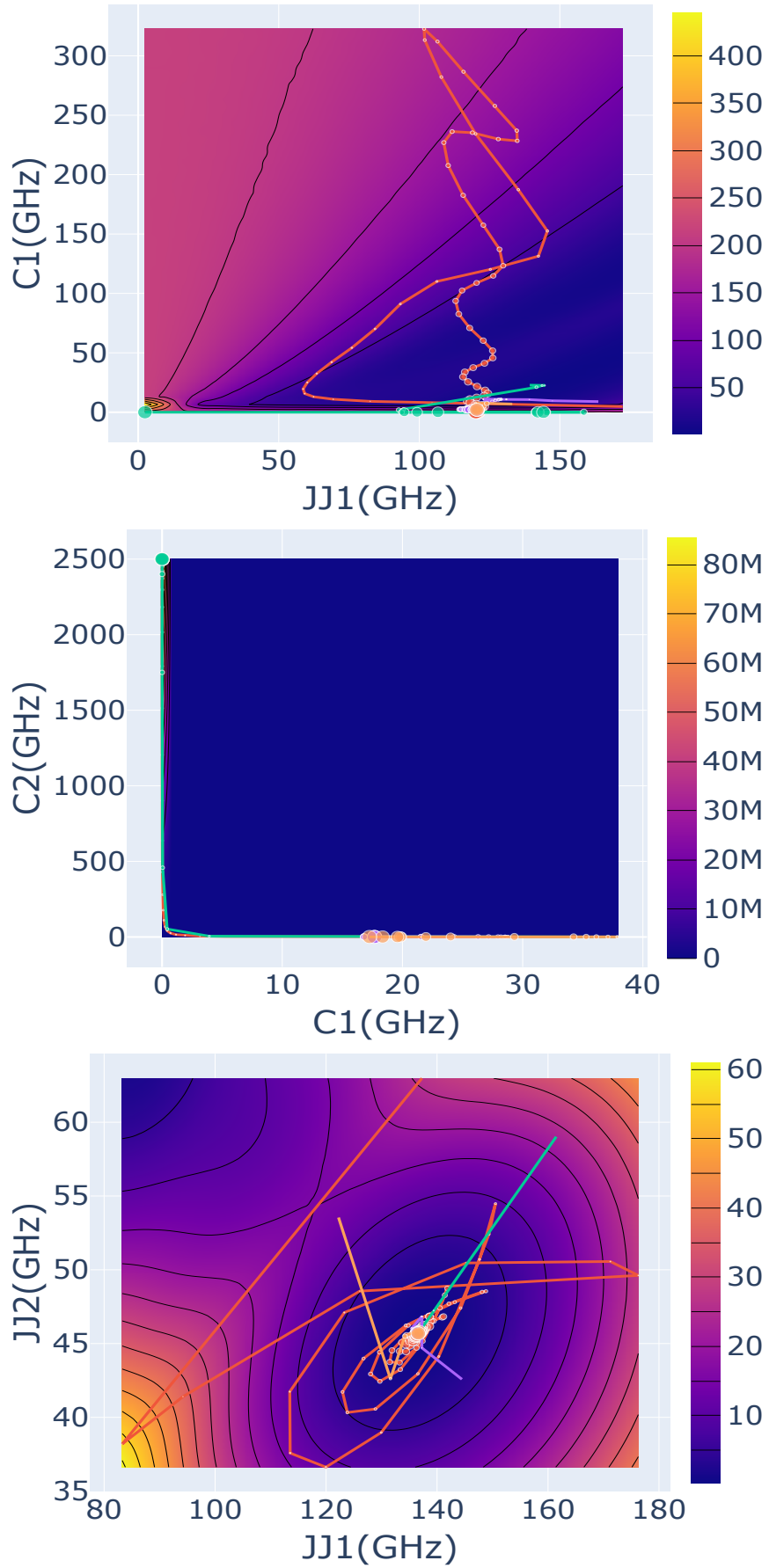


Figure 28: Trajectories depict optimization iteration, in the direction of increasing dot size. Best performing instance is taken for each algorithm: Adam2(red), lrBFGS(green), Nelder-Mead(violet), LBFGS(orange).

References

- [1] Google Quantum AI. Suppressing quantum errors by scaling a surface code logical qubit. *Nature*, 614(7949):676–681, Feb 2023.
- [2] Bernd Seeber. *Handbook of Applied Superconductivity, Volume 2*, volume 2. CRC press, 1998.
- [3] F. A. Cárdenas-López, J. C. Retamal, Xi Chen, G. Romero, and M. Sanz. Resilient superconducting-element design with genetic algorithms, 2023.
- [4] Tim Menke, Florian Häse, Simon Gustavsson, Andrew J. Kerman, William D. Oliver, and Alán Aspuru-Guzik. Automated discovery of superconducting circuits and its application to 4-local coupler design, 2020.
- [5] Jing Yu, Juan Carlos Retamal, Mikel Sanz, Enrique Solano, and Francisco Albarrán-Arriagada. Superconducting circuit architecture for digital-analog quantum computing. *EPJ Quantum Technology*, 9(1):1–35, 2022.
- [6] Valentin Stanev, Kamal Choudhary, Aaron Gilad Kusne, Johnpierre Paglione, and Ichiro Takeuchi. Artificial intelligence for search and discovery of quantum materials. *Communications Materials*, 2(1):105, 2021.
- [7] H. A. Buchdahl. Concerning a kind of truncated quantized linear harmonic oscillator. *American Journal of Physics*, 35(3):210–218, 1967.
- [8] Philip W. Anderson. Higgs, anderson and all that. *Nature Physics*, 11(2):93–93, Feb 2015.
- [9] John F. Cochran and D. E. Mapother. Superconducting transition in aluminum. *Phys. Rev.*, 111:132–142, Jul 1958.
- [10] L. Petit, H. G. J. Eenink, M. Russ, W. I. L. Lawrie, N. W. Hendrickx, S. G. J. Philips, J. S. Clarke, L. M. K. Vandersypen, and M. Veldhorst. Universal quantum logic in hot silicon qubits. *Nature*, 580(7803):355–359, apr 2020.
- [11] M. Lyatti, M. A. Wolff, I. Gundareva, M. Kruth, S. Ferrari, R. E. Dunin-Borkowski, and C. Schuck. Energy-level quantization and single-photon control of phase slips in $\text{yba}_2\text{cu}_3\text{o}_{7-x}$ nanowires. *Nature Communications*, 11(1):763, Feb 2020.

-
- [12] Svetlomir Stavrev, Francesco Grilli, Bertrand Dutoit, Nadia Nibbio, Emmanuel Vinot, Iszabela Klutsch, Gerard Meunier, Pascal Tixador, Yifeng Yang, and Elena Martinez. Comparison of numerical methods for modeling of superconductors. *IEEE transactions on magnetics*, 38(2):849–852, 2002.
 - [13] A.G. REDFIELD. The theory of relaxation processes* *this work was started while the author was at harvard university, and was then partially supported by joint services contract n5ori-76, project order i. In John S. Waugh, editor, *Advances in Magnetic Resonance*, volume 1 of *Advances in Magnetic and Optical Resonance*, pages 1–32. Academic Press, 1965.
 - [14] S.E. Rasmussen, K.S. Christensen, S.P. Pedersen, L.B. Kristensen, T. Bækkegaard, N.J.S. Loft, and N.T. Zimmer. Superconducting circuit companion—an introduction with worked examples. *PRX Quantum*, 2:040204, Dec 2021.
 - [15] Uri Vool and Michel Devoret. Introduction to quantum electromagnetic circuits. *International Journal of Circuit Theory and Applications*, 45(7):897–934, jun 2017.
 - [16] Xinyuan You, J. A. Sauls, and Jens Koch. Circuit quantization in the presence of time-dependent external flux. *Phys. Rev. B*, 99:174512, May 2019.
 - [17] Mario F Gely and Gary A Steele. QuCAT: quantum circuit analyzer tool in python. *New Journal of Physics*, 22(1):013025, jan 2020.
 - [18] Andrew J. Kerman. Efficient numerical simulation of complex josephson quantum circuits, 2020.
 - [19] P. Krantz, M. Kjaergaard, F. Yan, T. P. Orlando, S. Gustavsson, and W. D. Oliver. A quantum engineer's guide to superconducting qubits. *Applied Physics Reviews*, 6(2):021318, jun 2019.
 - [20] Jed A. Duersch, Meiyue Shao, Chao Yang, and Ming Gu. A robust and efficient implementation of LOBPCG. *SIAM Journal on Scientific Computing*, 40(5):C655–C676, jan 2018.
 - [21] Sai Pavan Chitta, Tianpu Zhao, Ziwen Huang, Ian Mondragon-Shem, and Jens Koch. Computer-aided quantization and numerical analysis of superconducting circuits, 2022.
 - [22] Mario Krenn, Mehul Malik, Robert Fickler, Radek Lapkiewicz, and Anton Zeilinger. Automated search for new quantum experiments. *Phys. Rev. Lett.*, 116:090405, Mar 2016.

-
- [23] Boyuan Chen, Kuang Huang, Sunand Raghupathi, Ishaan Chandratreya, Qiang Du, and Hod Lipson. Discovering state variables hidden in experimental data, 2021.
 - [24] Giuseppe Carleo, Ignacio Cirac, Kyle Cranmer, Laurent Daudet, Maria Schuld, Naftali Tishby, Leslie Vogt-Maranto, and Lenka Zdeborová. Machine learning and the physical sciences. *Rev. Mod. Phys.*, 91:045002, Dec 2019.
 - [25] Dimitri P Bertsekas. *Constrained optimization and Lagrange multiplier methods*. Academic press, 2014.
 - [26] Stanley C. Eisenstat and Ilse C. F. Ipsen. Three absolute perturbation bounds for matrix eigenvalues imply relative bounds. *SIAM Journal on Matrix Analysis and Applications*, 20(1):149–158, 1998.
 - [27] Moshe Looks, Marcello Herreshoff, DeLesley Hutchins, and Peter Norvig. Deep learning with dynamic computation graphs, 2017.
 - [28] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning representations by back-propagating errors. *nature*, 323(6088):533–536, 1986.
 - [29] Charles C. Margossian. A review of automatic differentiation and its efficient implementation. *WIREs Data Mining and Knowledge Discovery*, 9(4):e1305, 2019.
 - [30] Paul J Werbos. Backpropagation through time: what it does and how to do it. *Proceedings of the IEEE*, 78(10):1550–1560, 1990.
 - [31] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.
 - [32] Benoît Colson, Patrice Marcotte, and Gilles Savard. An overview of bilevel optimization. *Annals of operations research*, 153:235–256, 2007.
 - [33] Amirreza Shaban, Ching-An Cheng, Nathan Hatch, and Byron Boots. Truncated back-propagation for bilevel optimization. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 1723–1732. PMLR, 2019.
 - [34] Jens Koch, Terri M. Yu, Jay Gambetta, A. A. Houck, D. I. Schuster, J. Majer, Alexandre Blais, M. H. Devoret, S. M. Girvin, and R. J. Schoelkopf. Charge-insensitive qubit design derived from the cooper pair box. *Phys. Rev. A*, 76:042319, Oct 2007.

-
- [35] V Bouchiat, D Vion, P Joyez, D Esteve, and M H Devoret. Quantum coherence with a single cooper pair. *Physica Scripta*, 1998(T76):165, jan 1998.
- [36] Chris Brimacombe, Robert M. Corless, and Mair Zamir. Computation and applications of mathieu functions: A historical perspective. *SIAM Review*, 63(4):653–720, 2021.
- [37] Masanori Hanada, Junyu Liu, Enrico Rinaldi, and Masaki Tezuka. Estimating truncation effects of quantum bosonic systems using sampling algorithms, 2023.
- [38] Agustin Di Paolo, Thomas E. Baker, Alexandre Foley, David Sénéchal, and Alexandre Blais. Efficient modeling of superconducting quantum circuits with tensor networks. *npj Quantum Information*, 7(1):11, Jan 2021.
- [39] Ilya Sutskever, James Martens, George Dahl, and Geoffrey Hinton. On the importance of initialization and momentum in deep learning. In *International conference on machine learning*, pages 1139–1147. PMLR, 2013.
- [40] J. Q. You, Xuedong Hu, S. Ashhab, and Franco Nori. Low-decoherence flux qubit. *Phys. Rev. B*, 75:140515, Apr 2007.
- [41] Yuichiro Matsuzaki, Hideaki Hakoshima, Yuya Seki, and Shiro Kawabata. Quantum annealing with capacitive-shunted flux qubits. *Japanese Journal of Applied Physics*, 59(SG):SGI06, feb 2020.
- [42] Davide Stirpe, Juuso Manninen, and Francesco Massel. Semiclassical dynamics of a superconducting circuit: chaotic dynamics and fractal attractors, 2023.
- [43] Christoph Berke, Evangelos Varvelis, Simon Trebst, Alexander Altland, and David P. DiVincenzo. Transmon platform for quantum computing challenged by chaotic fluctuations. *Nature Communications*, 13(1):2495, May 2022.
- [44] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2017.
- [45] S. Singer and J. Nelder. Nelder-Mead algorithm. *Scholarpedia*, 4(7):2928, 2009. revision #91557.
- [46] Aristidis Likas, Nikos Vlassis, and Jakob J Verbeek. The global k-means clustering algorithm. *Pattern recognition*, 36(2):451–461, 2003.

-
- [47] Razvan Pascanu, Yann N Dauphin, Surya Ganguli, and Yoshua Bengio. On the saddle point problem for non-convex optimization. *arXiv preprint arXiv:1405.4604*, 2014.
- [48] Peter Brooks, Alexei Kitaev, and John Preskill. Protected gates for superconducting qubits. *Phys. Rev. A*, 87:052306, May 2013.
- [49] Chetan Nayak, Steven H. Simon, Ady Stern, Michael Freedman, and Sankar Das Sarma. Non-abelian anyons and topological quantum computation. *Reviews of Modern Physics*, 80(3):1083–1159, sep 2008.
- [50] Klaus Von Klitzing. The quantized hall effect. *Reviews of Modern Physics*, 58(3):519, 1986.
- [51] Joshua M. Dempster, Bo Fu, David G. Ferguson, D. I. Schuster, and Jens Koch. Understanding degenerate ground states of a protected quantum circuit in the presence of disorder. *Phys. Rev. B*, 90:094518, Sep 2014.
- [52] András Gyenis, Pranav S. Mundada, Agustin Di Paolo, Thomas M. Hazard, Xinyuan You, David I. Schuster, Jens Koch, Alexandre Blais, and Andrew A. Houck. Experimental realization of a protected superconducting circuit derived from the $0-\pi$ qubit. *PRX Quantum*, 2:010339, Mar 2021.
- [53] Agustin Di Paolo, Arne L Grimsmo, Peter Groszkowski, Jens Koch, and Alexandre Blais. Control and coherence time enhancement of the $0-i/i$ qubit. *New Journal of Physics*, 21(4):043002, apr 2019.
- [54] Daoquan Zhu, Tuomas Jaako, Qiongyi He, and Peter Rabl. Quantum computing with superconducting circuits in the picosecond regime. *Physical Review Applied*, 16(1), jul 2021.
- [55] Fuchang Gao and Lixing Han. Implementing the nelder-mead simplex algorithm with adaptive parameters. *Computational Optimization and Applications*, 51(1):259–277, Jan 2012.
- [56] Ciyu Zhu, Richard H Byrd, Peihuang Lu, and Jorge Nocedal. Algorithm 778: L-bfgs-b: Fortran subroutines for large-scale bound-constrained optimization. *ACM Transactions on mathematical software (TOMS)*, 23(4):550–560, 1997.
- [57] J. E. Dennis and Jorge J. Moré. A characterization of superlinear convergence and its application to quasi-newton methods. *Mathematics of Computation*, 28(126):549–560, 1974.

-
- [58] Yuchen Xie, Richard H. Byrd, and Jorge Nocedal. Analysis of the bfgs method with errors. *SIAM Journal on Optimization*, 30(1):182–209, 2020.
- [59] Philipp Aumann, Tim Menke, William D Oliver, and Wolfgang Lechner. Circuitq: an open-source toolbox for superconducting circuits. *New Journal of Physics*, 24(9):093012, sep 2022.
- [60] A Dziekonski, M Rewiński, Piotr Sypek, A Lamecki, and Michał Mrozowski. Gpu-accelerated lobpcg method with inexact null-space filtering for solving generalized eigenvalue problems in computational electromagnetics analysis with higher-order fem. *Communications in Computational Physics*, 22(4):997–1014, 2017.