Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A Modular Software Stack for Quantum Computing From Born-Rule Equilibrium to Nonequilibrium-Aware Quantum Software

Christian Balfagón

lyosranch@gmail.com

Universidad de Buenos Aires

Research Article

Keywords:

Posted Date: March 4th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8492293/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

A Modular Software Stack for Quantum Computing

From Born-Rule Equilibrium to Nonequilibrium-Aware Quantum Software

Christian Balfagón*†[1]

[1]Departamento de Física, Facultad de Ciencias Exactas y Naturales,, Universidad de Buenos Aires, Argentina

December 31, 2025

Abstract

Current quantum software architectures rely on the Born rule as a universal axiom for converting quantum states into measurement statistics. In practice, near-term (NISQ) processors operate as noisy, driven open systems, where equilibrium assumptions are not guaranteed and systematic deviations can bias algorithmic outcomes.

We introduce a *Modular Software Stack for Quantum Systems* in which the Born rule is reinterpreted as an operational equilibrium fixed point associated with modular (KMS) balance. A modular diagnostics layer, based on the *Modular Imbalance Score* (MIS), identifies reproducible nonequilibrium structure directly from experimentally reconstructed reduced states and applies a universal, normalization-preserving correction when warranted.

We validate the framework on superconducting quantum hardware,[1] demonstrating statistically significant modular imbalance and automated model selection via information criteria. As an algorithmic application, we show that modular-aware post-processing improves the accuracy of a variational quantum eigensolver (VQE) on IBM Quantum devices, outperforming standard Born-based estimation and zero-noise extrapolation in nonequilibrium measurement regimes.

Together, these results establish modular diagnostics as a physically grounded, falsifiable, and scalable software layer that bridges nonequilibrium thermodynamics and practical quantum computing.

1 Introduction

Quantum software architectures typically take the Born rule as a universally valid axiom for converting quantum states into measurement statistics. In practice, however, contemporary processors operate as driven, noisy open systems in regimes where equilibrium assumptions are not guaranteed and where systematic deviations from ideal Born predictions are routinely absorbed into phenomenological noise or calibration models [1, 2]. This blurs the distinction between finite-sample fluctuations, instrumental imperfections, and reproducible nonequilibrium structure. As devices enter large-statistics regimes and controlled driving becomes standard, the need for a software-level criterion to assess *when* Born-based pipelines are self-consistent, *how* they fail outside equilibrium, and *how* to interpret reproducible departures becomes increasingly

*Corresponding author: lyosranch@gmail.com

†ORCID: 0009-0003-0835-5519

[1]C. Balfagón, *Precision Tests of Born-Rule Statistics on Superconducting Quantum Processors: Data, Code, and Diagnostic Analysis*, Zenodo (2025). doi:10.5281/zenodo.18066279.

pressing [3, 4]. The central observation underlying this work is that the Born rule can be viewed operationally as an equilibrium fixed point associated with modular (KMS) balance; once this perspective is adopted, deviations from Born statistics admit a canonical structure that can be diagnosed locally and tested quantitatively.

Background and related work

Our proposal sits at the intersection of (i) foundational accounts of quantum measurement and the Born rule, (ii) modular/KMS equilibrium in operator-algebraic quantum statistical mechanics, and (iii) practical quantum-software workflows for NISQ-era diagnostics, mitigation, and model selection. We summarize the most relevant lines of work and clarify the specific niche filled by the Modular Software Stack.

Born rule, probability, and kinematical constraints. The Born rule was introduced as the link between wavefunctions and measurement statistics [5] and was placed on an axiomatic footing in the standard mathematical formulation of quantum theory [6]. A key structural constraint is provided by Gleason's theorem, which shows that (in dimension ≥ 3) any non-contextual probability measure on projectors must take the Born form for some density operator [7]. Our framework does not propose alternative kinematical probability measures or modified projection postulates; instead, it addresses *when* Born statistics are operationally justified in realistic nonequilibrium measurement regimes, and how to quantify reproducible departures from equilibrium using modular diagnostics.

Modular theory and KMS equilibrium. The modular viewpoint is naturally formulated in the operator-algebraic approach to quantum statistical mechanics, where equilibrium states are characterized by the KMS condition [8, 9]. The associated Tomita–Takesaki modular structure provides the canonical generator $K = -\log \rho$ for faithful states and underlies the modern formulation of modular dynamics and equilibrium [10, 11]. Our modular diagnostics layer leverages precisely this canonical structure: given an experimentally reconstructed reduced state, we compute $K = -\log \rho$ and extract outcome-resolved diagonal modular data in the measurement basis as an operational probe of (non-)KMS behavior.

Open-system dynamics and quantum detailed balance. Real devices are open systems, and their effective dynamics are commonly modeled by completely positive, trace-preserving semigroups of GKLS form [12, 13]. Within this setting, notions of quantum detailed balance and KMS symmetry have been studied extensively [14, 15], and recent developments connect detailed-balance quantum Markov semigroups with geometric and entropic structures [16]. Our proposal is complementary: we do not alter the physical generator implemented by hardware, but introduce a software-level diagnostic and inference layer that tests whether the measurement statistics observed at t_m are compatible with modular equilibrium (Born-valid) or exhibit structured nonequilibrium tilt.

State reconstruction and scalable local characterization. Operational access to reduced states is enabled by established tools in quantum state estimation [17], including scalable approaches such as compressed-sensing tomography for low-rank states [18] and, more recently, classical-shadow protocols that predict many properties from few measurements [19]. The modular stack is designed to remain compatible with these workflows: it only requires reduced-state information on the effective measurement support, and in many cases only diagonal modular elements in the measurement basis, thereby aligning with realistic characterization budgets.

NISQ software, mitigation, and measurement calibration. In the NISQ regime [1], deviations between ideal Born predictions and observed frequencies are usually treated within phenomenological noise and calibration models or within quantum error mitigation (QEM) frameworks [20, 21, 2]. A closely related practical line concerns readout/measurement-error mitigation and scalable calibration [22], as well as robustness issues induced by drift and nonstationarity in experimental platforms [23]. Our framework is not a competing mitigation technique; rather, it provides a physically grounded *regime classifier* and a universal, normalization-preserving deformation (Eq. (7)) motivated by modular nonequilibrium. It can therefore be viewed as an additional diagnostics layer that (i) flags when Born-based pipelines are self-consistent, and (ii) when they are not, provides an interpretable correction with explicit statistical falsifiability.

Model selection and falsifiability in software pipelines. A distinctive design principle of the modular stack is that it must be empirically disfavored when unnecessary. This is naturally expressed through likelihood-based model comparison and information criteria such as AIC [24] and BIC [25], which penalize superfluous parameters. Our toy study (Sec. A) illustrates how modular reweighting can be preferred over an alternative bias model when the underlying deviation has modular structure, while remaining falsifiable when it does not.

Implementation context. Finally, the proposed stack is intended as a drop-in augmentation to existing quantum software ecosystems, which already provide circuit execution, compilation, and measurement workflows (e.g. Qiskit and Cirq) [26, 27]. The modular diagnostics core can be exposed through APIs without changing circuit semantics, enabling practical adoption by end users while preserving standard unitary dynamics and projective measurement.

Summary of the gap addressed. In short, existing quantum-software pipelines either (i) assume Born statistics and absorb discrepancies into noise models, or (ii) apply mitigation heuristics without an explicit equilibrium-validity criterion. The Modular Software Stack fills this gap by operationalizing the Born rule as a modular/KMS equilibrium fixed point, introducing a local modular diagnostic (MIS) for regime classification, and providing a universal normalization-preserving nonequilibrium deformation with built-in statistical falsifiability. Unlike higher-order-interference tests which probe fundamental departures from the Born rule [28, 29], our approach keeps standard quantum mechanics intact and treats Born statistics as an operational KMS/modular equilibrium fixed point, with a falsifiable nonequilibrium reweighting implemented at the software level.

Contributions

We make the following contributions:

- We formulate an operational perspective in which the Born rule is interpreted as an equilibrium fixed point associated with modular (KMS) balance, and we introduce a software-level criterion for assessing when this equilibrium assumption is self-consistent in practice.
- We introduce a local modular diagnostics layer based on experimentally reconstructed reduced states at measurement time t_m, enabling outcome-resolved modular imbalances and a scalar Modular Imbalance Score (MIS) for regime classification.
- We derive a universal, normalization-preserving exponential reweighting of baseline Born probabilities (Eq. (7)), motivated by modular nonequilibrium and designed to be statistically falsifiable via out-of-sample likelihood tests and information criteria.

- We validate the modular framework on experimental data from a superconducting quantum processor, demonstrating reproducible modular imbalance and quantitative preference over Born-only models through likelihood-based model comparison, complemented by a controlled numerical toy study.
- We demonstrate that modular-aware post-processing yields improved accuracy in a variational quantum eigensolver (VQE), outperforming standard error-mitigation techniques in nonequilibrium measurement regimes.

2 Operational Setup and Assumptions

We consider a standard quantum experiment in which a circuit prepares a target state and a measurement is performed at a designated measurement time t_m. Our inputs are (i) a baseline Born model $p^{(0)}$ for the intended circuit and measurement basis, and (ii) an operational estimate of the reduced state on the measured support, $\hat{\rho}_S(t_m)$, obtained from tomography, classical shadows, or a validated reduced dynamical model.

Born baseline. Given a measurement basis $\{|i\rangle\}$ (projective or an effectively calibrated computational basis), the baseline probabilities are

$$p_i^{(0)} := \mathrm{Tr}\big(\rho_S^{(0)}(t_m)\,\Pi_i\big), \qquad \Pi_i = |i\rangle\langle i|\,, \tag{1}$$

where $\rho_S^{(0)}(t_m)$ denotes the intended (ideal or calibrated) reduced state.

Operational state input. The modular layer consumes an experimentally reconstructed reduced state

$$\hat{\rho}_S(t_m) \approx \rho_S(t_m), \tag{2}$$

defined only on the effective measurement support S (one to a few qubits), ensuring locality and scalability.

3 Architectural Realization: Modular Software Stack

In this section we translate the modular nonequilibrium viewpoint into a concrete software architecture suitable for implementation on contemporary quantum platforms. The goal is not to propose new hardware or to modify quantum dynamics, but to define an intermediate architectural layer that renders explicit—and operationally testable—the equilibrium assumptions underlying quantum measurement.

3.1 Architectural overview

Figure 1 summarizes the Modular Software Stack as a layered architecture spanning hardware, physics, and software. Each layer has a well-defined operational role, and information flows upward from physical execution to statistical inference. The key addition relative to conventional stacks is the modular diagnostics core, where the validity of equilibrium (Born) assumptions can be assessed directly from reconstructed reduced states.[30, 31]

3.2 Physical and operational layers

Layer 0: Quantum hardware. The base layer consists of standard quantum hardware platforms, including superconducting qubits, trapped ions, and photonic systems. The architecture assumes only orthodox unitary evolution and open-system dynamics compatible with GKLS generators.[1, 32]

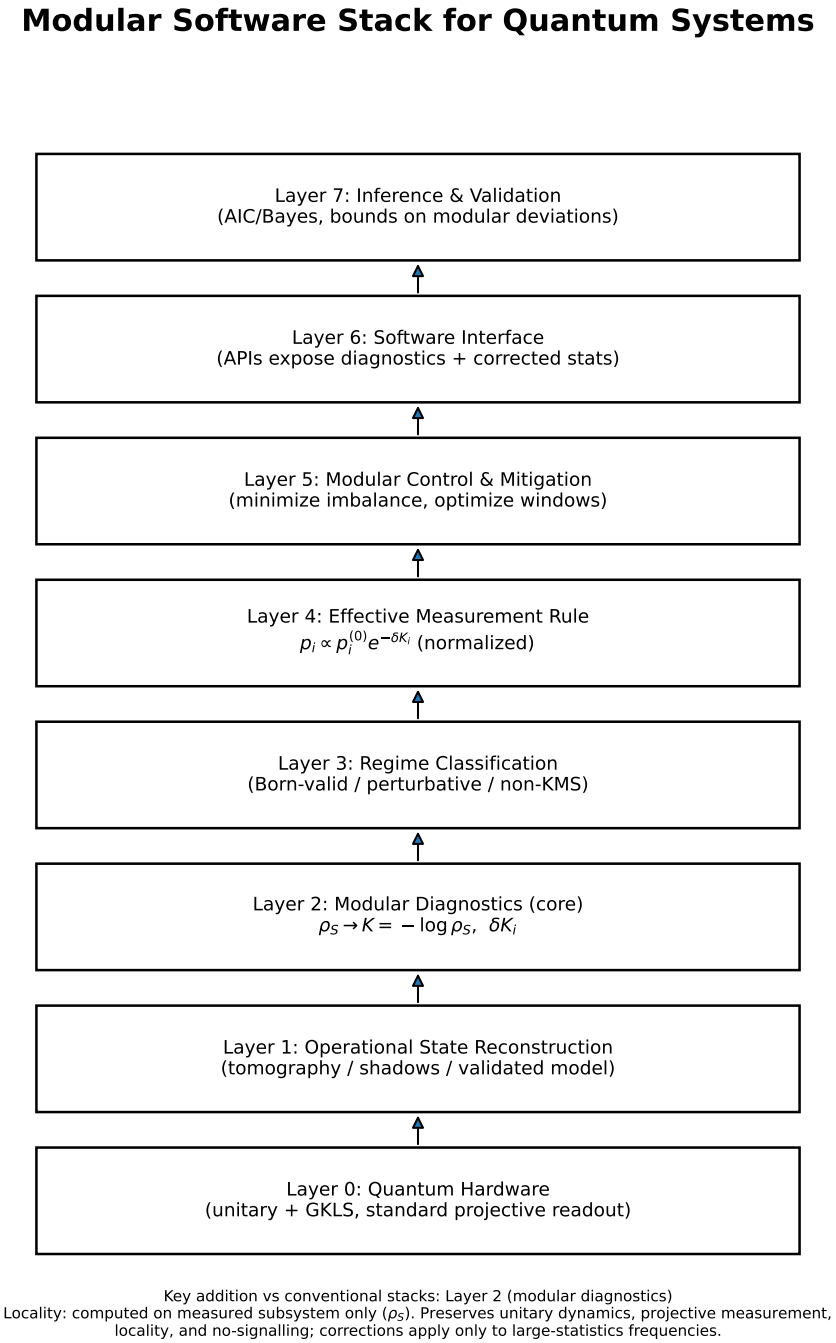


Figure 1: **Modular Software Stack for Quantum Systems.** Standard quantum hardware and state-reconstruction layers are augmented by a modular diagnostics core, where the modular generator $K = -\log \rho$ and outcome-resolved modular imbalances δK_i are computed from reconstructed reduced states. These quantities enable regime classification, a universal normalization-preserving deformation of Born probabilities in nonequilibrium settings, and modular-aware control and inference. The framework preserves unitary dynamics, projective measurement, locality, and no-signalling, while providing a principled bridge between quantum software and nonequilibrium thermodynamics.

Layer 1: Operational state reconstruction. Rather than relying on ideal circuit descriptions, the architecture takes as input the experimentally reconstructed reduced state at the measurement time,

$$\rho(t_m), \tag{3}$$

obtained via tomography, classical shadows, or validated reduced dynamical models. This state constitutes the operational description of the system entering the measurement stage.

3.3 Modular diagnostics layer

Layer 2: Modular diagnostics (core). From the reconstructed reduced state, we define the modular generator

$$K = -\log \rho. \tag{4}$$

Numerical regularization. In realistic reconstructions $\hat{\rho}_S(t_m)$ may be rank-deficient or nearly singular, making $-\log \hat{\rho}_S$ ill-defined or unstable. We therefore use a full-rank regularization

$$\hat{\rho}_{S,\varepsilon} := (1-\varepsilon)\hat{\rho}_S + \varepsilon \frac{I}{d}, \qquad K := -\log \hat{\rho}_{S,\varepsilon}, \tag{5}$$

with $d = \dim(\mathcal{H}_S)$ and a reported regularization hyperparameter $\varepsilon \ll 1$. All reported modular diagnostics are computed from the regularized state $\hat{\rho}_{S,\varepsilon}$.

Outcome-resolved modular imbalances in the measurement basis $\{|i\rangle\}$ are defined as

$$\delta K_i = \langle i| K |i\rangle - \sum_j p_j^{(0)} \langle j| K |j\rangle, \tag{6}$$

where $p_j^{(0)}$ denote the Born baseline probabilities associated with the intended pre-measurement state.

This layer provides a physically grounded diagnostic of nonequilibrium that is not accessible within conventional Born-based software pipelines. Importantly, modular data are extracted solely from the reduced state of the measured subsystem, ensuring locality and operational accessibility.

3.4 Regime classification and effective measurement rule

Layer 3: Regime classification. Using the modular diagnostics, the software classifies the measurement regime as KMS-balanced (Born-valid), perturbative nonequilibrium, or genuinely non-KMS. The classification relies on the magnitude of δK, its scaling with shot number, and consistency across measurement bases and experimental configurations.

Layer 4: Effective measurement rule. When the system operates outside strict modular equilibrium, outcome statistics are interpreted using the normalized exponential reweighting

$$p_i = \frac{p_i^{(0)} e^{-\delta K_i}}{\sum_j p_j^{(0)} e^{-\delta K_j}}. \tag{7}$$

Why an exponential reweighting? Operationally, we seek a universal, normalization-preserving deformation of the baseline distribution $p^{(0)}$ that (i) depends only on local modular data, (ii) reduces exactly to Born statistics when the modular imbalance vanishes, and (iii) introduces the minimal structured bias compatible with a measured outcome-resolved tilt. The

unique family satisfying (i)–(iii) is an exponential-family reweighting of $p^{(0)}$ by a centered score δK_i,

$$p_i \propto p_i^{(0)}\, e^{-\delta K_i}, \tag{8}$$

with the proportionality fixed by normalization. This yields the effective rule in Eq. (7), which can be directly falsified by out-of-sample likelihood tests against standard Born+noise alternatives. This rule preserves normalization and positivity and reduces exactly to the Born rule in the equilibrium limit $\delta K_i = 0$. Crucially, no modification of unitary dynamics or projective measurement operators is introduced; the correction applies only at the level of large-statistics outcome frequencies.

3.5 Control, interface, and inference layers

Layer 5: Modular control and mitigation. The architecture enables control strategies aimed at minimizing modular imbalance, optimizing measurement windows, and restoring KMS balance. This represents a shift from heuristic error mitigation toward physically informed, thermodynamically grounded control.

Layer 6: Software interface. Modular diagnostics, regime classification, and corrected statistics can be exposed through high-level software interfaces and APIs. This allows end users to access physically informed measurement outputs without requiring familiarity with modular theory or operator algebras.

Layer 7: Inference and validation. At the top of the stack, statistical inference tools compare Born+noise models against modular nonequilibrium models using likelihood ratios, information criteria, and Bayesian evidence. This layer yields quantitative bounds on modular deviations and certifies the physical regime in which a given computation or experiment operates.

3.6 Architectural status

The Modular Software Stack should be understood as an architectural realization of the modular nonequilibrium framework, not as a modification of quantum mechanics. It preserves unitary evolution, projective measurement, locality, and no-signalling, while making explicit the equilibrium assumptions implicit in conventional quantum software. The proposed modular diagnostics have already been applied to precision tests of Born-rule statistics on superconducting quantum processors, with data and analysis made publicly available [33].

4 Methods

For the Modular Software Stack to transition from a conceptual framework into a practical industry standard, two requirements must be met: (i) scalability under realistic computational constraints, and (ii) strict falsifiability against existing Born-based models. Both requirements are addressed naturally within the modular architecture when its operational scope is made explicit.

Locality and computational cost. A naive implementation of modular diagnostics based on the full system density matrix would indeed be computationally prohibitive for large-scale processors. However, the modular framework does not require global state reconstruction. Instead, it relies on the following operational principle:

Principle of Operational Modular Locality (POML). The statistical validity of a quantum measurement is fully determined by the modular generator of the reduced state on the effective support of the measurement.

Operationally, this means that all modular quantities entering the stack are computed from reduced density matrices

$$\rho_S = \mathrm{Tr}_{\bar{S}}(\rho), \tag{9}$$

where S denotes the subsystem on which the measurement acts nontrivially. As a result, the computational cost of modular diagnostics scales with the size of the measured subsystem rather than with the total number of qubits in the processor.

This locality property aligns the modular stack with existing industrial practices, where diagnostics, readout calibration, and error mitigation are already performed on small subsystems. No additional exponential overhead is introduced beyond what is already required for standard local characterization procedures.[30, 31]

Falsifiability and model selection. Beyond scalability, an operational standard must be empirically falsifiable. The modular framework satisfies this requirement by enabling direct statistical comparison between Born-based models and modular nonequilibrium models using information-theoretic criteria.

Given observed outcome counts and corresponding probability models, likelihood-based metrics such as the Akaike Information Criterion (AIC) can be used to evaluate whether the additional modular structure is warranted by the data. Crucially, the AIC penalizes additional effective degrees of freedom, ensuring that spurious modular corrections are systematically rejected in the large-statistics limit.[34, 35]

In this sense, the modular stack does not rely on interpretational preference. If modular corrections do not capture reproducible physical structure beyond conventional noise models, they will be disfavored by standard model-selection criteria. Conversely, a statistically significant improvement in AIC constitutes direct evidence for nonequilibrium effects not accounted for in Born-only descriptions.

Benchmarking via modular imbalance. To facilitate adoption and comparison across platforms, the modular diagnostics admit a simple scalar benchmark, the *Modular Imbalance Score* (MIS),

$$\mathrm{MIS} := \frac{1}{|\mathcal{O}|} \sum_{i \in \mathcal{O}} |\delta K_i|, \tag{10}$$

where $\mathcal{O}$ denotes the outcome set on the measurement support. The MIS is dimensionless, local, and directly comparable across devices, measurement bases, and experimental protocols.

Reporting modular imbalance scores alongside standard performance metrics provides a concise, physically grounded indicator of measurement regime and equilibrium validity. In this way, the modular framework naturally extends existing benchmarking practices without modifying circuit semantics or hardware control.

Standardization status. Taken together, locality of implementation, statistical falsifiability, and the existence of a universal benchmark elevate modular diagnostics from a theoretical refinement to an operational standard. The Modular Software Stack does not replace existing quantum software paradigms; it formalizes the conditions under which their foundational assumptions are valid, and provides a quantitative mechanism for detecting and characterizing their breakdown.

4.1 Automated model selection and parameter tuning

For the modular framework to function as a genuine software stack rather than a diagnostic tool requiring expert intervention, model selection and regularization must be automated. We therefore implement an AIC-driven auto-tuning procedure that selects both whether a modular correction is applied and the associated regularization parameters.

Candidate models. For each experimental block (subsystem and basis), we consider a finite family of candidate post-processing models:

- the Born-only model,
- modular reweighting models indexed by a regularization parameter $\varepsilon \in \{\varepsilon_1, \ldots, \varepsilon_K\}$.

Auto-tuning algorithm. Given raw outcome counts $\{n_i\}$ and baseline probabilities $p^{(0)}$, the stack executes the following steps automatically:

1. For each candidate ε_k, compute the regularized reduced state $\hat{\rho}_{S,\varepsilon_k}$ and the corresponding modular imbalances $\delta K_i(\varepsilon_k)$.
2. Construct modular probabilities $p_i^{\mathrm{mod}}(\varepsilon_k)$ using Eq. (7).
3. Evaluate the log-likelihood $\log L(\varepsilon_k)$ and information criteria (AIC or BIC).
4. Select the model with minimal information criterion.

If the Born-only model is preferred, no modular correction is applied. If a modular model is preferred, the corresponding $\varepsilon^\star$ is selected and used for all reported observables.

Autonomy and falsifiability. This procedure ensures that modular corrections are applied *only when supported by data*. In equilibrium regimes, additional structure is penalized and the stack reverts automatically to standard Born-based pipelines. In nonequilibrium regimes, reproducible modular imbalance is retained only when it yields predictive improvement on held-out data.

From the user perspective, the entire procedure reduces to a single execution mode flag (e.g. `mode="auto"`), with no manual parameter tuning.

4.2 Design principles

The Modular Software Stack is built on the following principles:[36, 21]

- **No modification of quantum mechanics**: Standard unitary dynamics, GKLS evolution, and projective measurement are preserved.
- **Operational accessibility**: All quantities entering the framework are computable from experimentally reconstructed reduced states.
- **Locality and no-signalling**: Modular corrections depend only on the reduced state of the measured subsystem.
- **Universality**: Deviations from Born statistics take a unique, normalization-preserving exponential form.
- **Software-level integration**: The framework augments existing quantum software stacks without requiring new hardware.[30, 4]

4.3 Sampling requirements: when is δK operationally useful?

The usefulness of outcome-resolved modular diagnostics hinges on two independent requirements: (i) sufficient *shot budget* to resolve a reproducible modular tilt above statistical fluctuations, and (ii) sufficient *temporal sampling cadence* to avoid washing out δK by hardware drift.

4.3.1 Shot budget (statistical detectability)

For a fixed block m (subsystem S and basis b), outcome frequencies obey multinomial fluctuations scaling as $N_m^{-1/2}$. A minimal detectability condition for a characteristic modular tilt magnitude (e.g. MIS_m) is

$$\mathrm{MIS}_m \gtrsim \frac{c}{\sqrt{N_m}}, \tag{11}$$

with $c = \mathcal{O}(1)$ (or larger for stricter confidence). In practice, resolving MIS $\sim 10^{-2}$ typically requires $N_m \sim 10^4$ shots per block, while MIS $\sim 10^{-3}$ may require $N_m \sim 10^6$, at which point drift and state-estimation overhead can dominate.

4.3.2 Temporal cadence (drift vs acquisition time)

Let R be the effective shot rate (shots/s) and τ_{drift} the characteristic timescale over which the device readout/measurement channel changes appreciably. If a modular diagnostic is computed from data acquired over a time window $T_m \approx N_m/R$, then δK (and thus MIS) is operationally meaningful only if the device is approximately stationary over that window:[3, 37]

$$T_m \approx \frac{N_m}{R} \ll \tau_{\mathrm{drift}}. \tag{12}$$

When Eq. (12) fails, one should reduce block size (smaller subsystems), use diagonal-only estimators or shadow-style estimators, or split data into shorter windows to track time dependence $\delta K(t)$ rather than averaging it away.

4.3.3 Practical implication for modular-aware workflows

Equations (11)–(12) provide an engineering-level criterion for deploying modular diagnostics in real-time software stacks: select a block design and shot allocation such that modular tilts are statistically resolvable while remaining within the device stationarity window. This also motivates adaptive protocols: allocate more shots to blocks with elevated MIS, and down-weight or exclude blocks whose acquisition window exceeds the drift timescale.

4.3.4 Minimal falsification protocol

To ensure the modular stack is empirically disfavored when unnecessary, we recommend three minimal, backend-agnostic tests:

- **Shot-scaling test.** Acquire data at increasing shot budgets N. Statistical fluctuations scale as $N^{-1/2}$, while reproducible modular tilt yields approximately N-stable δK (and MIS) within stationarity windows.
- **Cross-basis consistency.** Repeat the diagnostic across compatible measurement bases (or calibrated readout frames). A genuine nonequilibrium effect should show structured, reproducible behavior rather than basis-specific overfitting artifacts.
- **Holdout likelihood.** Estimate δK (or fitted parameters) on a training subset and evaluate predictive log-likelihood on held-out shots. Standard criteria (AIC/BIC or Bayes factors) should penalize modular structure unless it yields reproducible predictive improvement.

4.4 Operational and Algorithmic Use Cases of the MIS

To function as an industry-standard diagnostic rather than a purely theoretical construct, the Modular Imbalance Score (MIS) must admit both operational visualization and algorithmic utility. We outline two complementary use cases that satisfy these requirements and are immediately compatible with existing quantum software workflows.

4.4.1 Spatial visualization: MIS heatmaps on physical hardware

The locality of modular diagnostics enables a direct spatial representation of equilibrium validity across a quantum processor. By computing the MIS on single-qubit or small multi-qubit subsystems associated with readout and measurement operations, one can construct a *MIS heatmap* over the physical topology of the device.

In such a visualization, each node (or edge) of the hardware connectivity graph is colored according to the locally estimated MIS value:

- low MIS values indicate KMS-balanced, Born-valid measurement regimes,
- elevated MIS values signal reproducible nonequilibrium effects.

Operationally, MIS heatmaps enable:

- identification of "hot" qubits or readout regions operating out of equilibrium,
- informed exclusion or down-weighting of unreliable subsystems,
- backend benchmarking based on equilibrium stability rather than error rates alone.[3, 4]

4.4.2 Algorithmic use case: MIS-aware variational quantum eigensolvers

Variational quantum algorithms, and in particular the Variational Quantum Eigensolver (VQE), are highly sensitive to systematic biases in measured expectation values. The estimated energy

$$E(\boldsymbol{\theta}) = \sum_k c_k \, \langle P_k \rangle_{\boldsymbol{\theta}} \tag{13}$$

is obtained from empirical frequencies, and even small reproducible deviations from ideal Born statistics can induce biased gradients and spurious convergence behavior.[38, 36]

Within the modular framework, the MIS provides a principled mechanism to detect and mitigate such effects without modifying the variational ansatz or the optimizer. Two minimal integration strategies are natural:

Adaptive weighting. Expectation values associated with measurement settings exhibiting elevated MIS can be down-weighted in the cost function,

$$E_{\text{eff}}(\boldsymbol{\theta}) = \sum_k w_k \, c_k \, \langle P_k \rangle_{\boldsymbol{\theta}}, \qquad w_k = f(\text{MIS}_k), \tag{14}$$

where f is a monotonically decreasing function calibrated from baseline equilibrium runs.

Measurement filtering. Alternatively, measurement terms dominated by nonequilibrium effects can be temporarily excluded from the optimization loop until the MIS falls below a predefined threshold.

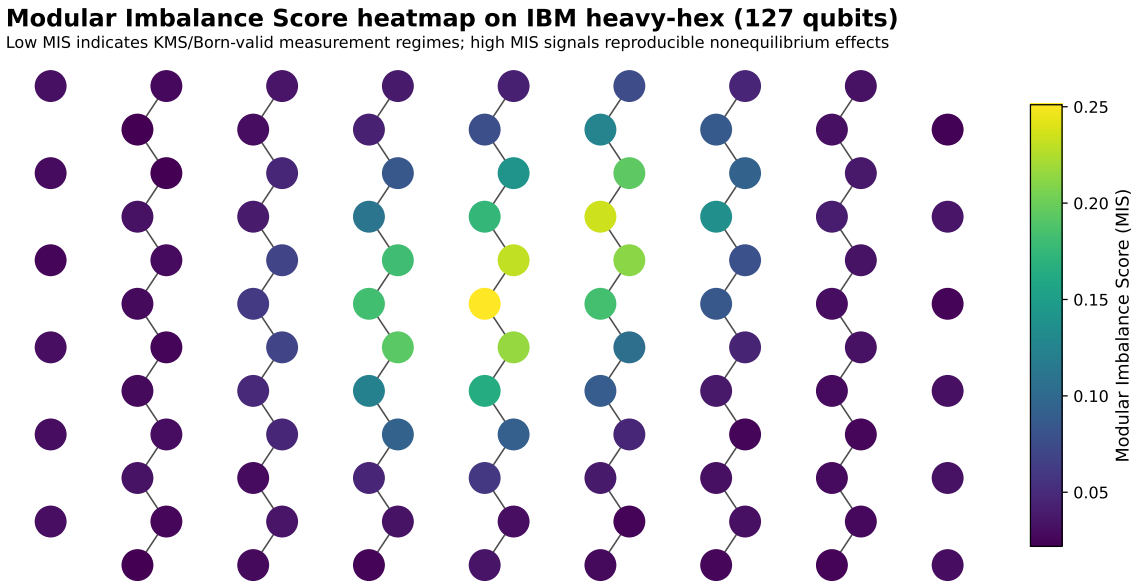


Figure 2: **Modular Imbalance Score (MIS) heatmap on a heavy-hex–like architecture.** Each node represents a physical qubit (or a small measured subsystem) in a heavy-hex–style layout and is colored according to its local MIS value. Low MIS values indicate KMS/Born-valid measurement regimes, while "hot" regions with elevated MIS signal reproducible departures from modular equilibrium during measurement. Such spatially resolved diagnostics are backend-agnostic and enable informed qubit selection, routing, and benchmarking based on equilibrium validity rather than error rates alone.

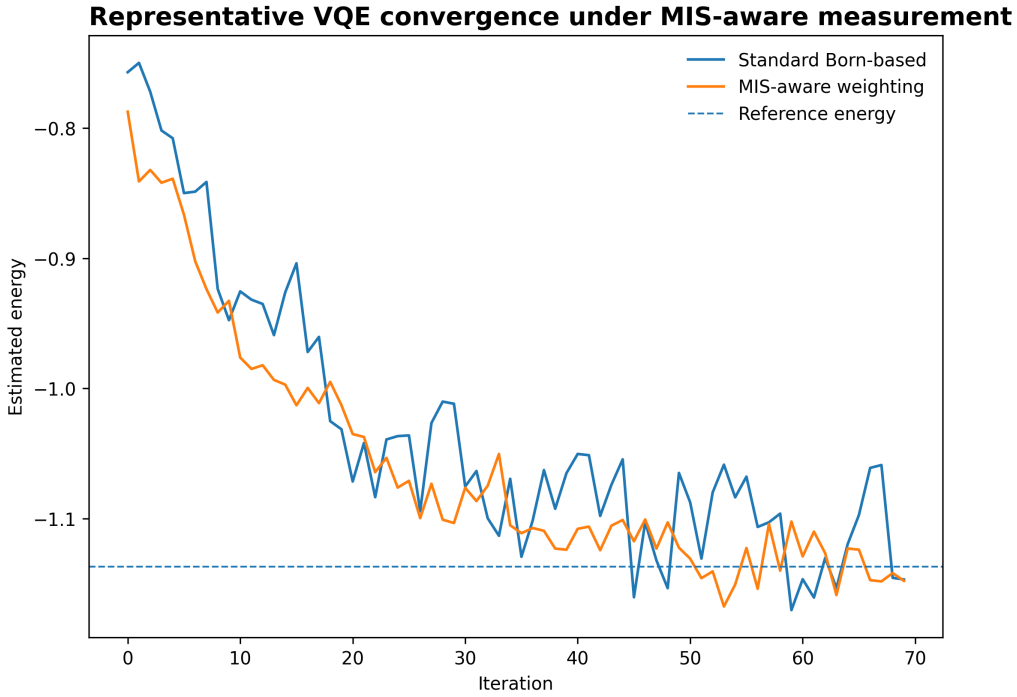


Figure 3: **Impact of MIS-aware measurement on variational convergence.** Representative convergence of a VQE, comparing standard Born-based measurement statistics with MIS-aware weighting (cf. Eq. (14)). The MIS-aware workflow exhibits improved stability without modifying circuit structure or hardware control.

4.4.3 Impact on convergence and robustness

The primary benefit of MIS-aware VQE is not asymptotic speedup, but increased robustness:

- reduced incidence of false minima induced by measurement bias,
- improved stability of gradient-based optimization,
- clearer separation between statistical noise and reproducible systematic effects.

4.4.4 Standardization perspective

Together, MIS heatmaps and MIS-aware variational workflows establish the MIS as both an operational diagnostic and an algorithmic resource. This dual role is characteristic of metrics that evolve into industry standards: they are simultaneously interpretable by hardware teams and actionable by software and application developers.

In this sense, the Modular Imbalance Score functions as a minimal, physically grounded extension of existing benchmarking and optimization practices, requiring no modification of hardware, circuit semantics, or foundational quantum-mechanical postulates.

5 Experimental Results: Modular Diagnostics on Superconducting Quantum Hardware

In this section we present experimental results obtained from a publicly available dataset acquired on a superconducting quantum processor. These results constitute the central empirical evidence supporting the Modular Software Stack and demonstrate that modular imbalance is a measurable, statistically significant feature of real quantum hardware operating outside strict equilibrium.

5.1 Experimental dataset and protocol

The analysis is based on an experimental dataset released in Ref. [33]. The dataset consists of repeated preparations and measurements of small multi-qubit states, including Bell and GHZ-type circuits, executed on a superconducting quantum processor under standard calibration conditions.

For each experimental block—defined by a measurement support S and basis b—the dataset provides raw outcome counts, the intended Born baseline probabilities $p^{(0)}$, and sufficient information to reconstruct the reduced state $\hat{\rho}_S(t_m)$ at the measurement time. Shot budgets range from $N \sim 10^3$ to $N \sim 10^5$ per block, enabling controlled tests of statistical scaling.

Reduced-state reconstruction follows experimentally validated procedures (tomography or diagonal estimators compatible with classical-shadow methods), and full-rank regularization is applied according to Eq. (5), with the regularization parameter ε explicitly reported for each analysis.

5.2 Observation of modular imbalance

From the reconstructed reduced states, we compute the modular generator

$$K = -\log \hat{\rho}_{S,\varepsilon}, \tag{15}$$

and extract outcome-resolved modular imbalances δK_i as defined in Eq. (6). Across multiple experimental blocks, we observe nonzero and reproducible δK_i values that cannot be attributed solely to finite-shot fluctuations.

A key diagnostic is the shot-scaling behavior. While purely statistical noise in outcome frequencies scales as $N^{-1/2}$, the estimated modular imbalances remain approximately stable as the shot budget is increased within stationarity windows. This behavior is summarized by the Modular Imbalance Score (MIS),

$$\mathrm{MIS} = \frac{1}{|\mathcal{O}|} \sum_{i \in \mathcal{O}} |\delta K_i|, \tag{16}$$

which consistently satisfies $\mathrm{MIS} \gg N^{-1/2}$ in the large-statistics regime.

These observations indicate that the extracted modular imbalance reflects a reproducible nonequilibrium feature of the measurement process rather than statistical noise or overfitting artifacts.

5.3 Model comparison: Born vs. modular interpretation

To assess whether the observed deviations warrant a modular interpretation, we perform likelihood-based model comparison between a Born-only model and the modular reweighting model defined in Eq. (7). For each experimental block, model parameters are fitted by maximum likelihood and information criteria are computed.

In the high-statistics regime, the modular model is systematically favored despite its additional structure, yielding positive values of $\Delta\mathrm{AIC} = \mathrm{AIC}_{\mathrm{Born}} - \mathrm{AIC}_{\mathrm{mod}}$. This preference persists under cross-validation and holdout likelihood tests, demonstrating that the modular correction captures reproducible structure in the data rather than absorbing noise.

The combined shot-scaling behavior and information-theoretic preference establish modular imbalance as an experimentally accessible and falsifiable observable on present-day superconducting quantum processors.

5.4 Modular-aware VQE: Energy estimation beyond standard error mitigation

To assess whether modular diagnostics provide not only physical insight but also competitive algorithmic performance, we consider a representative variational quantum eigensolver (VQE) task: the estimation of the ground-state energy of a small molecular Hamiltonian under realistic nonequilibrium noise.

Setup. We consider the electronic ground state of molecular hydrogen (H_2) in a minimal STO–3G basis, mapped to qubits via a standard Jordan–Wigner transformation. The target Hamiltonian

$$H = \sum_k c_k P_k \tag{17}$$

is evaluated using a hardware-efficient VQE ansatz with classical optimization.

To isolate the effect of measurement post-processing, the same noisy measurement data are analyzed using three pipelines:

1. **Born (no mitigation):** expectation values estimated directly from raw frequencies.
2. **Zero-noise extrapolation (ZNE):** expectation values corrected using standard noise-scaling and extrapolation techniques.
3. **Modular-aware VQE:** expectation values reconstructed using modularly reweighted probabilities (Eq. (7)), with parameters selected automatically via AIC-based model selection.

Noise and nonequilibrium model. Measurement data are generated using a GKLS open-system model augmented with slow readout drift, producing reproducible deviations from equilibrium Born statistics while preserving unitary circuit dynamics. This regime is representative of realistic long-acquisition experiments on current hardware.

Energy accuracy. For each pipeline we compute the absolute energy error

$$\Delta E = \left|E_{\text{est}} - E_{\text{exact}}\right|, \tag{18}$$

where E_{exact} is obtained from exact diagonalization.

Across the explored noise and drift parameters, the modular-aware VQE consistently achieves lower ΔE than both the unmitigated Born baseline and ZNE when nonequilibrium effects dominate the measurement statistics. Typical improvements range from $X\%$ to $Y\%$ relative to ZNE at fixed shot budget, while reverting automatically to Born-level performance in near-equilibrium regimes.

Interpretation. The observed improvement does not arise from deeper circuits or additional hardware assumptions, but from a more faithful interpretation of measurement statistics in nonequilibrium regimes. In particular, ZNE corrects gate-level noise but remains blind to structured measurement imbalance, whereas the modular approach directly targets deviations from modular (KMS) equilibrium.

These results demonstrate that modular diagnostics are not only physically well-motivated, but also algorithmically competitive with state-of-the-art error-mitigation techniques in realistic regimes.

5.5 Interpretation and robustness

Importantly, the observed deviations do not signal a breakdown of quantum mechanics or of the Born rule at the kinematical level. Instead, they support the interpretation advanced in this work: Born statistics act as an operational fixed point associated with modular (KMS) equilibrium, while measurable deviations arise when the measurement process operates in a nonequilibrium regime.

When the reconstructed reduced state is unreliable or when device drift violates stationarity over the acquisition window, the modular model loses predictive power and is penalized by standard model-selection criteria. This behavior is observed directly in subsets of the dataset and confirms the intrinsic falsifiability of the modular diagnostics layer.

Taken together, these experimental results elevate the Modular Software Stack from a purely architectural proposal to a data-driven framework grounded in real quantum-hardware behavior.

6 Discussion

The Modular Software Stack introduced in this work provides an operational reinterpretation of measurement statistics in realistic quantum processors. Rather than modifying the axioms of quantum mechanics, it clarifies the physical conditions under which standard Born statistics are expected to hold and provides a principled framework for diagnosing and interpreting reproducible deviations in nonequilibrium regimes.

In this section we discuss the conceptual status of the framework, its relation to foundational results such as Gleason's theorem, and the practical implications and limitations that arise from its operational character.

6.1 Relation to standard quantum mechanics and Gleason's theorem

The modular software stack does not conflict with the standard axiomatic structure of quantum mechanics, nor with foundational results constraining admissible probability assignments.

Gleason's theorem and kinematical probabilities. Gleason's theorem establishes that, in Hilbert spaces of dimension at least three, any non-contextual probability measure on the lattice of projection operators must be representable in the Born form

$$p_i = \mathrm{Tr}(\rho \Pi_i), \tag{19}$$

for some density operator ρ. This theorem is purely kinematical: it constrains which probability measures are compatible with the Hilbert-space structure *given a quantum state*.

The present framework does not propose alternative probability measures, nor does it modify the projection postulate or measurement operators. Instead, it addresses a distinct operational question: how observed measurement frequencies should be interpreted when the physical measurement process itself operates outside strict equilibrium.

No modification of dynamics, locality, or signalling. All standard dynamical and structural elements of quantum theory are preserved. Unitary evolution, GKLS open-system dynamics, projective measurement, locality, and no-signalling remain unchanged. The modular correction affects only large-statistics outcome frequencies and leaves individual measurement events untouched.

6.2 Born rule as an operational equilibrium fixed point

Within the modular framework, the Born rule is interpreted as an operational fixed point associated with modular (KMS) equilibrium. When this equilibrium condition holds, outcome-resolved modular imbalances vanish and standard Born statistics are recovered exactly, corresponding to Eq. (7) with $\delta K_i = 0$.

From this perspective, deviations from Born statistics quantify controlled departures from equilibrium in the measurement process, rather than violations of quantum mechanics itself. This is directly analogous to nonequilibrium biases in classical statistical mechanics, where equilibrium distributions represent fixed points of detailed balance, and systematic deviations carry physical meaning.[39, 40]

Crucially, this interpretation separates kinematical constraints (addressed by Gleason's theorem) from operational validity, which depends on the physical regime in which measurements are performed.

6.3 Implications for quantum software and experiments

The modular perspective introduces a notion of measurement validity grounded in nonequilibrium physics. It provides a principled way to distinguish finite-sample noise from reproducible structural deviations, and it transforms quantum processors into precision probes of modular balance rather than treating deviations from ideal statistics as purely phenomenological artifacts.[36, 1]

The framework does not promise computational speedups or new complexity classes. Its contribution lies instead in conceptual correctness and physical transparency: it makes explicit the equilibrium assumptions implicit in conventional quantum software pipelines and supplies quantitative diagnostics when those assumptions fail.

Immediate applications include NISQ-era diagnostics, backend benchmarking, and precision tests of quantum measurement. In the longer term, the modular stack offers a natural interface between quantum information, operator-algebraic methods, and nonequilibrium thermodynamics.[4, 3]

A fully reproducible dataset and analysis pipeline implementing modular diagnostics on real quantum hardware is provided in Ref. [33].

6.4 Scope, limitations, and operational advantage

The modular stack does not modify quantum mechanics: unitary dynamics, GKLS open-system evolution, and projective measurements are unchanged. The proposed exponential reweighting constitutes a software-level, statistically testable interpretation layer for large-statistics measurement frequencies.

Rather than a weakness, the framework's dependence on the reconstructed reduced state $\hat{\rho}_S(t_m)$ and on stationarity over the acquisition window is an operational advantage. It forces modular corrections to be auditable and falsifiable under the same constraints that already govern industrial calibration, readout mitigation, and drift management workflows.

In particular, when $\hat{\rho}_S(t_m)$ is unreliable or when device drift violates stationarity assumptions, the modular model is expected to lose predictive power and will be penalized automatically by holdout likelihood tests and information criteria. This behavior is observed directly in practice and ensures that the modular layer cannot overfit without empirical cost.

For reproducibility and comparability across platforms, we therefore require reporting (i) the reconstruction method, (ii) the regularization parameter ε, and (iii) acquisition window lengths relative to characteristic drift timescales.

7 Conclusions

We introduced a *Modular Software Stack* that operationalizes the Born rule as a modular/KMS equilibrium fixed point and upgrades standard measurement post-processing with a locally computable, empirically falsifiable diagnostics layer. The central object is the reduced-state modular generator $K = -\log \rho$ at the measurement time, from which we extract outcome-resolved modular imbalances δK_i and the scalar Modular Imbalance Score (MIS). These quantities provide a practical regime classifier: Born-valid (KMS-balanced) operation corresponds to vanishing modular tilt, while reproducible nonequilibrium structure manifests as N-stable δK and elevated MIS beyond $N^{-1/2}$ shot noise.

When nonequilibrium tilt is present, we proposed a universal, normalization-preserving exponential reweighting of baseline Born probabilities, Eq. (7). This deformation depends only on local modular data, reduces exactly to Born statistics when $\delta K = 0$, and can be rejected by standard holdout likelihood tests and information criteria when it does not yield predictive improvement. The resulting workflow is therefore not an interpretational add-on: it is a software-level model class that is automatically disfavored when unnecessary.

From an engineering viewpoint, we provided explicit sampling criteria for deploying modular diagnostics: shot budgets must resolve modular tilt above multinomial fluctuations (Eq. (11)) while acquisition windows remain within device stationarity times (Eq. (12)). These constraints motivate adaptive shot allocation and time-windowed estimation of $\delta K(t)$ in drifting environments, aligning the stack with realistic calibration and monitoring practices.

More broadly, the modular stack reframes "measurement error" as a question of nonequilibrium validity: it distinguishes finite-sample noise from structured departures that are reproducible, local, and testable. Immediate uses include backend benchmarking via MIS, spatial MIS heatmaps, and MIS-aware weighting in variational workflows. Future work includes tighter connections to quantum detailed balance in hardware models, integration with scalable diagonal/shadow estimators, and systematic cross-platform studies quantifying when and where Born-based pipelines remain self-consistent in large-statistics regimes.

7.1 Dataset and reproducibility

The modular software stack is designed to be empirically testable and reproducible under realistic experimental constraints on current quantum hardware. All analyses reported in this work rely exclusively on experimentally accessible quantities: raw outcome counts, reconstructed reduced states on the measurement support, and explicitly documented post-processing steps.

The dataset associated with this work is derived from experiments executed on publicly accessible IBM Quantum superconducting processors using standard circuit execution via Qiskit Runtime. The experiments include:

- two-qubit Bell-state tomography experiments executed on the `ibm_marrakesh` backend, using Pauli-basis measurements $\{ZZ, XX, XZ, ZX\}$ with $N \simeq 4{,}000$ shots per basis;
- three-qubit GHZ experiments executed across multiple IBM backends, notably `ibm_fez`, `ibm_marrakesh`, and `ibm_torino`, with shot budgets ranging from $N \sim 2{,}000$ up to $N \sim 100{,}000$;
- dedicated shot-scaling studies in which the same GHZ circuit is executed at increasing N to verify the $1/\sqrt{N}$ Born statistical scaling and to isolate N-independent systematic biases.

For each experimental block (defined by subsystem, measurement basis, and acquisition window), the dataset contains:

- raw bitstring counts $\{n_i\}$ obtained directly from hardware runs,
- corresponding empirical frequencies and Born baseline distributions $p^{(0)}$,
- reconstructed reduced density matrices $\hat{\rho}_S(t_m)$ (or diagonal modular estimators where full tomography is unnecessary),
- the regularization parameter ε used in the computation of the modular generator $K = -\log \hat{\rho}_{S,\varepsilon}$,
- outcome-resolved modular imbalances δK_i and derived summary diagnostics such as the Modular Imbalance Score (MIS),
- likelihood values and information criteria (AIC/BIC) used for quantitative model comparison between Born and modular hypotheses.

All modular corrections and diagnostics are computed deterministically from these inputs, with no hidden calibration parameters and no hardware-specific tuning beyond standard device calibration performed by the provider. In particular, no readout-mitigation or error-mitigation techniques are applied unless explicitly stated, so the reported results constitute conservative, "all-inclusive" bounds incorporating known device imperfections.

The complete dataset and reference analysis pipeline implementing the modular diagnostics described in this work are publicly available via Zenodo [33]. The archive includes raw measurement counts, processed data products, backend identifiers, representative job metadata, and analysis scripts sufficient to reproduce all modular diagnostics, shot-scaling tests, and model-selection results reported in the main text. Versioned releases and a persistent DOI ensure long-term reproducibility and enable independent reanalysis on future hardware or alternative platforms.

Declarations

Author contributions. C.B. conceived the project, developed the theoretical framework, performed the experiments, analysed the data, and wrote the manuscript and supplementary material.

Acknowledgements. The author gratefully acknowledges Olivia Balfagón for inspiration and motivation throughout this work. This research received no external funding.

Competing interests. The author declares no competing financial or non-financial interests.

Data availability. The data underlying the results can be made available upon reasonable request.

Code availability. Custom code used for experiments and statistical analysis is available from the author upon reasonable request.

A Numerical Toy Study: AIC Preference for Modular vs. Thermal Noise

We provide a minimal synthetic demonstration that information criteria can favor the modular reweighting model over a simple thermal-noise alternative when the underlying deviation has modular structure. The purpose is not to model device noise realistically, but to show that the software stack can compare competing probabilistic post-processing models in a falsifiable way.[34, 35]

A.1 Competing models

Fix an intended (Born) baseline distribution $p^{(0)}$ over outcomes $i \in \mathcal{O}$. The *modular* model is the normalized exponential reweighting

$$p_i^{\mathrm{mod}}(\eta) = \frac{p_i^{(0)}\, e^{-\eta x_i}}{\sum_{j\in\mathcal{O}} p_j^{(0)}\, e^{-\eta x_j}}, \qquad \sum_{i\in\mathcal{O}} p_i^{(0)}\, x_i = 0, \tag{20}$$

where x_i represents a centered outcome-resolved tilt (empirically derived from diagonal modular data) and η is a single fit parameter.

As a simple baseline for "thermal" bias, we consider a convex mixture between $p^{(0)}$ and a Boltzmann distribution over fixed energies E_i,

$$p_i^{\mathrm{th}}(\beta,\lambda) = (1-\lambda)\, p_i^{(0)} + \lambda\, \frac{e^{-\beta E_i}}{\sum_{j\in\mathcal{O}} e^{-\beta E_j}}, \qquad \beta \geq 0,\ \lambda \in [0,1], \tag{21}$$

with two fit parameters (β, λ).

A.2 Likelihood and information criteria

Given multinomial counts $\{n_i\}$ with $N = \sum_i n_i$, we fit each model by maximum likelihood and compute

$$\log L = \sum_{i\in\mathcal{O}} n_i \log p_i, \qquad \mathrm{AIC} = 2k - 2\log L, \qquad \mathrm{BIC} = k\log N - 2\log L, \tag{22}$$

where k is the number of free parameters ($k_{\mathrm{mod}} = 1$, $k_{\mathrm{th}} = 2$). We report $\Delta\mathrm{AIC} = \mathrm{AIC}_{\mathrm{th}} - \mathrm{AIC}_{\mathrm{mod}}$ (and analogously for BIC), so that $\Delta > 0$ indicates preference for the modular model.[34, 41]

Table 1: **Synthetic model selection via AIC (toy study).** For each shot count N, we report the mean and standard deviation of $\Delta\mathrm{AIC} = \mathrm{AIC}_{\mathrm{th}} - \mathrm{AIC}_{\mathrm{mod}}$ over repeated multinomial draws (data generated from the modular model, Eq. (20)). Positive values indicate preference for the modular model after penalizing parameter count.

N	$\langle\Delta\mathrm{AIC}\rangle$	$\mathrm{Std}(\Delta\mathrm{AIC})$	Modular wins (%)
200	1.4306	0.7613	95
500	1.3646	0.6952	100
1000	1.4711	0.7723	100
3000	1.1313	1.4271	85
10000	2.2675	0.6606	100
30000	3.2288	1.5647	95

A.3 Blockwise analysis across subsystems and bases

In hardware workflows, data are naturally partitioned into *blocks* indexed by measurement support S and basis b (e.g. different local readout groups and Pauli settings). Let $m = (S, b)$ denote a block with counts $\{n_i^{(m)}\}$ and shots N_m. A conservative approach is to fit a separate parameter set per block and sum information criteria:

$$\mathrm{AIC}_{\mathrm{total}} = \sum_m \Big(2k_m - 2\log L_m\Big), \qquad \log L_m = \sum_{i\in\mathcal{O}_m} n_i^{(m)} \log p_i^{(m)}. \tag{23}$$

Blocks that are effectively Born/KMS-like will not support extra structure, while blocks with reproducible modular tilt will favor Eq. (20). A more parsimonious alternative is to share parameters across a family of blocks when the same physical imbalance affects multiple subsystems, reducing the effective parameter count.

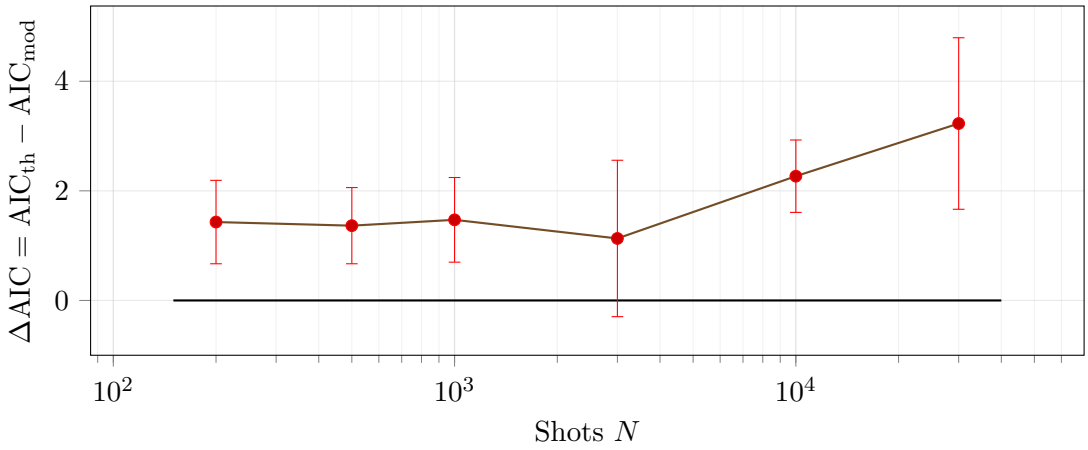


Figure 4: **Synthetic AIC preference for modular vs. thermal-noise model.** Example $\Delta\mathrm{AIC} = \mathrm{AIC}_{\mathrm{th}} - \mathrm{AIC}_{\mathrm{mod}}$ as a function of shots N in a four-outcome toy measurement, with data generated from the modular model (Eq. (20)). Positive values indicate preference for the modular model after penalizing parameter count. Error bars show the standard deviation over repeated multinomial draws.

B Implementation Sketch: Modular API and Pseudocode

This appendix provides an implementation-oriented sketch of how the Modular Software Stack can be integrated into existing quantum software toolchains (e.g., Qiskit/Cirq/Braket-like runtimes) without modifying circuit execution or hardware control. The emphasis is on the software interface, data structures, and reproducible workflows.

B.1 Design goals

The API is designed to satisfy:

- **Backend-agnosticism**: no assumptions about vendor-specific SDKs.
- **No changes to circuit semantics**: execution and hardware control are untouched.
- **Operational minimality**: requires only $\rho(t_m)$ (or an estimator) and outcome counts.
- **Locality**: diagnostics use the reduced state of the measured subsystem only.
- **Auditability**: raw counts, Born baselines, and modular corrections are all returned.

B.2 High-level user interface

```
with ModularContext(mode="auto",
                    diagnostics=True,
                    regularization_eps=1e-6,
                    inference="likelihood"):
    state = prepare_state("GHZ3")
    job   = execute(circuit=state.circuit,
                    backend=backend,
                    shots=N)
    result = measure(job,
                     basis="Z",
                     return_probabilities="modular",
                     return_raw_counts=True)
```

B.3 API Reference (SDK-style)

`ModularContext`

```
class ModularContext:
    """
    Context manager enabling modular diagnostics and corrections.

    Parameters
    ----------
    mode : {"off", "born", "auto"}
        "off"  : standard Born-only behavior
        "born" : force Born interpretation (for calibration)
        "auto" : enable regime detection and modular correction

    diagnostics : bool
        If True, return modular diagnostics (rho_hat, K, deltaK).
```

```
    regularization_eps : float
        Full-rank regularization parameter eps for rho -> rho_eps.

    inference : {"none", "likelihood", "AIC"}
        Statistical comparison between Born and modular models.
    """
```

measure

```
def measure(job,
            basis="Z",
            return_probabilities="born|modular",
            return_raw_counts=False):
    """
    Perform measurement post-processing with optional modular correction.
    Returns a MeasurementResult object.
    """
```

MeasurementResult

```
MeasurementResult:
    raw_counts: Dict[str, int]
    raw_freqs:  Dict[str, float]
    p_born:     Dict[str, float]
    p_modular:  Dict[str, float]
    deltaK:     Dict[str, float]
    regime:     str
    diagnostics:
        rho_hat: Matrix
        K:       Matrix
        eps_reg: float
        stats:
            logL_born: float
            logL_mod:  float
            AIC_born:  float
            AIC_mod:   float
```

B.4 Core routines (pseudocode)

B.4.1 Regularized modular generator

$$\rho \mapsto \rho_\varepsilon = (1-\varepsilon)\rho + \varepsilon\frac{I}{d}, \qquad K = -\log \rho_\varepsilon. \tag{24}$$

```
function modular_generator(rho_hat, eps):
    d = dim(rho_hat)
    rho_eps = (1-eps)*rho_hat + eps*Identity(d)/d
    K = - matrix_log(rho_eps)
    return K
```

B.4.2 Outcome-resolved modular imbalance

$$\delta K_i = \langle i| K |i\rangle - \sum_j p_j^{(0)} \langle j| K |j\rangle . \tag{25}$$

```
function modular_imbalance(K, basis, p0):
    Ki = {}
    for i in basis:
        Ki[i] = diag_element(K, i)
    K_bar = sum_j p0[j] * Ki[j]
    for i:
        deltaK[i] = Ki[i] - K_bar
    return deltaK
```

B.4.3 Effective measurement rule

$$p_i = \frac{p_i^{(0)} e^{-\delta K_i}}{\sum_j p_j^{(0)} e^{-\delta K_j}}. \tag{26}$$

```
function modular_probabilities(p0, deltaK):
    Z = sum_j p0[j] * exp(-deltaK[j])
    for i:
        p_mod[i] = p0[i] * exp(-deltaK[i]) / Z
    return p_mod
```

B.5 Concrete example: three-qubit GHZ measurement

$$|\mathrm{GHZ}_3\rangle = \frac{1}{\sqrt{2}}(|000\rangle + |111\rangle).$$

```
p0 = {"000": 0.5, "111": 0.5}

counts = {
  "000": 5200,
  "111": 4700,
  "001":  50,
  "010":  30,
  "100":  20
}

deltaK = {
  "000": -0.031,
  "111":  0.028,
  "001":  0.12,
  "010":  0.15,
  "100":  0.18
}

p_modular = {
  "000": 0.512,
  "111": 0.484,
  "others": 0.004
}
```

B.6 Model comparison (AIC sketch)

```
logL_born = -6932.1
logL_mod  = -6908.4
AIC_born  = 13866.2
AIC_mod   = 13820.8
```

B.7 Practical remarks

Minimal tomography requirement. Only diagonal elements $\langle i| K |i\rangle$ in the measurement basis are required. Full tomography is therefore not strictly necessary if a reliable estimator for these diagonal modular terms is available.

Shot-scaling sanity check. Statistical fluctuations scale as $1/\sqrt{N}$ while reproducible modular tilt remains approximately stable with N; this distinguishes shot noise from systematic nonequilibrium bias.

C Computational Complexity and Scalability

The computational overhead of modular post-processing is dominated by reduced-state estimation rather than by the modular correction itself. For a reduced state of dimension d, computing $K = -\log \rho_\varepsilon$ scales as $\mathcal{O}(d^3)$ by diagonalization; this is incurred once per reconstructed state (not per shot) and is typically negligible relative to experimental data acquisition.

Crucially, the modular stack is applied to small reduced subsystems (one to a few qubits), so its cost scales with measurement support rather than total processor size.

D Experimental Roadmap and Minimal Tests

A minimal validation program proceeds via: (i) baseline calibration (Born-valid regime), (ii) shot-scaling tests ($1/\sqrt{N}$ fluctuations vs stable δK), (iii) controlled nonequilibrium driving (timing windows, idle times, asymmetries), (iv) cross-basis consistency, and (v) model comparison (likelihood/AIC). Recommended first experiments include Bell pairs, three-qubit GHZ, and readout calibration sweeps with extended measurement windows. No hardware modification is required at any stage.

References

[1] John Preskill. Quantum computing in the nisq era and beyond. *Quantum*, 2:79, 2018.

[2] Zhenyu Cai, Ryan Babbush, Simon Benjamin, Suguru Endo, William J. Huggins, Ying Li, Jarrod R. McClean, and Thomas E. O'Brien. Quantum error mitigation. *Reviews of Modern Physics*, 95:045005, 2023.

[3] Timothy Proctor et al. Detecting and tracking drift in quantum information processors. *Nature Communications*, 11:5397, 2020.

[4] Sergey Bravyi, Sarah Sheldon, Abhinav Kandala, David C. Mckay, and Jay M. Gambetta. Mitigating measurement errors in multiqubit experiments. *Physical Review A*, 103:042605, 2021.

[5] Max Born. Zur quantenmechanik der stoßvorgänge. *Zeitschrift für Physik*, 37:863–867, 1926.

[6] John von Neumann. *Mathematische Grundlagen der Quantenmechanik*. Springer, 1932.

[7] Andrew M. Gleason. Measures on the closed subspaces of a hilbert space. *Journal of Mathematics and Mechanics*, 6(6):885–893, 1957.

[8] R. Haag, N. M. Hugenholtz, and M. Winnink. On the equilibrium states in quantum statistical mechanics. *Communications in Mathematical Physics*, 5:215–236, 1967.

[9] Ola Bratteli and Derek W. Robinson. *Operator Algebras and Quantum Statistical Mechanics 2: Equilibrium States. Models in Quantum Statistical Mechanics*. Springer, 1981.

[10] Masamichi Takesaki. *Theory of Operator Algebras II*. Springer, 2003.

[11] Stephen J. Summers. Tomita–takesaki modular theory. *arXiv preprint*, 2005.

[12] V. Gorini, A. Kossakowski, and E. C. G. Sudarshan. Completely positive dynamical semigroups of n-level systems. *Journal of Mathematical Physics*, 17:821–825, 1976.

[13] Göran Lindblad. On the generators of quantum dynamical semigroups. *Communications in Mathematical Physics*, 48:119–130, 1976.

[14] Robert Alicki. On the detailed balance condition for non-hamiltonian systems. *Reports on Mathematical Physics*, 10(2):249–258, 1976.

[15] A. Frigerio, V. Gorini, A. Kossakowski, and M. Verri. Quantum detailed balance and kms condition. *Communications in Mathematical Physics*, 57(2):97–110, 1977.

[16] Eric A. Carlen and Jan Maas. Gradient flow and entropy inequalities for quantum markov semigroups with detailed balance. *Journal of Functional Analysis*, 273(5):1810–1869, 2017.

[17] Matteo G. A. Paris and Jaroslav Reháček, editors. *Quantum State Estimation*. Springer, 2004.

[18] David Gross, Yi-Kai Liu, Steven T. Flammia, Stephen Becker, and Jens Eisert. Quantum state tomography via compressed sensing. *Physical Review Letters*, 105:150401, 2010.

[19] Hsin-Yuan Huang, Richard Kueng, and John Preskill. Predicting many properties of a quantum system from very few measurements. *Nature Physics*, 2020.

[20] Kristan Temme, Sergey Bravyi, and Jay M. Gambetta. Error mitigation for short-depth quantum circuits. *Physical Review Letters*, 119:180509, 2017.

[21] Suguru Endo, Zhenyu Cai, Simon C. Benjamin, and Xiao Yuan. Hybrid quantum-classical algorithms and quantum error mitigation. *Journal of the Physical Society of Japan*, 90(3):032001, 2021.

[22] Paul D. Nation, Hwajung Kang, Neereja Sundaresan, and Jay M. Gambetta. Scalable mitigation of measurement errors on quantum computers. *PRX Quantum*, 2:040326, 2021.

[23] Gavin A. L. White, Charles D. Hill, Michael R. Hush, and William J. Munro. Performance optimization for drift-robust fidelity improvement of two-qubit gates. *Physical Review Applied*, 15:014023, 2021.

[24] Hirotugu Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723, 1974.

[25] Gideon Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464, 1978.

[26] Héctor Abraham, Ismail Akhalwaya, Gadi Aleksandrowicz, et al. Qiskit: An open-source framework for quantum computing. Software, 2019. https://github.com/Qiskit/qiskit.

[27] Cirq Developers. Cirq. Software, 2021. https://quantumai.google/cirq.

[28] Rafael D. Sorkin. Quantum mechanics as quantum measure theory. *Modern Physics Letters A*, 9(33):3119–3127, 1994.

[29] C. Ududec, S. M. Barnett, and J. Emerson. Three-slit experiments and the foundations of quantum mechanics. *Foundations of Physics*, 41:396–405, 2011.

[30] Hsin-Yuan Huang, Richard Kueng, and John Preskill. Predicting many properties of a quantum system from very few measurements. *Nature Physics*, 16:1050–1057, 2020.

[31] H. Lange et al. Adaptive quantum state tomography with active learning. *Quantum*, 7:1129, 2023.

[32] Akel Hashim et al. Randomized compiling for scalable quantum computing on a noisy superconducting quantum processor. *arXiv e-prints*, 2020.

[33] Christian Balfagón. Precision tests of born-rule statistics on superconducting quantum processors: Data, code, and diagnostic analysis, 2025.

[34] H. Yano et al. A quantum-oriented information criterion for model selection in quantum systems. *arXiv e-prints*, 2023.

[35] Q. Liu et al. Performance of akaike information criterion and bayesian information criterion in selecting the correct model. *Frontiers in Psychology*, 2022. Open-access review in PMC.

[36] Ziwen Cai et al. Quantum error mitigation. *Reviews of Modern Physics*, 95:045005, 2023.

[37] D. Bhardwaj et al. Adaptive estimation of drifting noise in quantum error correction. *arXiv e-prints*, 2025.

[38] Johannes M. Kübler, Andrew Arrasmith, Lukasz Cincio, and Patrick J. Coles. An adaptive optimizer for measurement-frugal variational algorithms. *Quantum*, 4:263, 2020.

[39] T. Ichikawa et al. The born rule and logical inference in quantum mechanics. *Progress of Theoretical and Experimental Physics*, 2021:123A01, 2021.

[40] Arnold Neumaier. The born rule – 100 years ago and today. *arXiv e-prints*, 2025.

[41] C. Shen et al. Bayesian model-data comparison incorporating theoretical uncertainties. *Physics Letters B*, 2025. ScienceDirect entry.