

Statistical Mechanics-Inspired Quantum Generative Models for Rare-Event Simulation and Anomaly Detection

Rajesh Bhaskarla^{1,*}, Harish Pulluri², Mohammed Nazeer³

^{1,2,3}Dept. of EEE, Anurag University, Hyderabad, India

ABSTRACT

Reliable estimation of tail risk and detection of low-prevalence anomalies remain challenging under distribution shift and limited simulation budgets. This paper introduces a statistical mechanics-inspired hybrid quantum generative model that learns an energy-shaped proposal for rare-event simulation while producing a calibrated anomaly score for deployment. Across simulation budgets from 200 to 6400 samples, relative error decreased from 0.31 to 0.08, outperforming normalizing flows (0.44 to 0.12), diffusion sampling (0.48 to 0.14), and a classical energy model (0.46 to 0.13). At a fixed evaluation budget, the proposed estimator achieved relative error 0.10 with effective sample size 1120 and tail hit rate 6.8%, compared with 0.16, 760, and 4.1% for flows. For anomaly detection, AUROC reached 0.962 and AUPRC reached 0.714, with precision 0.61 at 80% recall, exceeding flow scoring (AUROC 0.935, AUPRC 0.624, precision 0.49) and VAE reconstruction (AUROC 0.928, AUPRC 0.598, precision 0.46). Under temporal drift after a single calibration, realized false positive rate stayed within 0.0004 of a 0.5% target, whereas baselines drifted by up to 0.0084. Ablations showed peak AUPRC near moderate circuit depth and mid-scale qubits, while monotone annealing reduced rare-event error from 0.12 to 0.10 and improved stability. Under increasing noise, mitigation preserved AUPRC from 0.74 at low noise to 0.69 at moderate noise, supporting feasibility for NISQ-constrained monitoring and tail estimation.

Keywords Quantum Generative Modeling, Energy-Based Modeling, Rare-Event Simulation, Importance Sampling, Anomaly Detection, Statistical Mechanics, NISQ, Tail Calibration

INTRODUCTION

Near-term progress in Noisy Intermediate-Scale Quantum (NISQ) hardware has intensified interest in quantum generative modeling as a mechanism for representing complex distributions with circuit-native sampling. However, scalable training remains constrained by noise, finite-shot estimation, and optimization pathologies that appear even in idealized settings. A central limitation is the barren plateau phenomenon, in which gradients vanish exponentially for broad circuit families, degrading identifiability of distributional parameters and destabilizing likelihood learning. These constraints motivate tail-focused generative objectives that remain stable under sampling noise [1], [2].

A separate but tightly related operational pressure comes from anomaly detection, where decisions depend on discriminating low-probability deviations under nonstationary conditions. Survey evidence shows that deep anomaly systems can be highly expressive but often exhibit sensitivity to contamination, thresholding instability, and weak calibration under distribution shift, particularly in streaming regimes. In parallel, the statistics of extremes formalize tail

Submitted 5 March 2026
Accepted 18 May 2026
Published 31 August 2026

Corresponding author
Rajesh Bhaskarla,
raajesh6@gmail.com

Additional Information and
Declarations can be found on
[page 149](#)

DOI: [10.63913/jqai.v1i2.9](https://doi.org/10.63913/jqai.v1i2.9)

© Copyright
2026 Bhaskarla, et al.

Distributed under
Creative Commons CC-BY 4.0

behavior through exceedances over thresholds, supplying a principled language for rare deviations while highlighting the data-efficiency barrier when tail samples are scarce [3], [4].

This paper targets rare-event simulation, in which the objective is not average-case prediction but accurate estimation of tail probabilities and tail-conditioned expectations. Classical reliability methods such as subset simulation decompose tiny probabilities into products of less-rare conditional events, improving estimator stability in high dimensions. Complementary proposal-learning approaches, including the cross-entropy method and adaptive multilevel splitting, further reduce variance by concentrating computation in progressively rarer regions. Yet, proposal quality depends on informative tail scores and can degrade under multi-modal or rugged tail geometry [5], [6], [7].

A natural unifying abstraction for tails is statistical mechanics, where probabilities are induced by an energy function and normalized via a partition function. Within this lens, score-based estimators such as score matching avoid explicit partition-function computation while still shaping energy landscapes to match observed statistics. Quantum algorithms for preparing thermal (Gibbs) states connect directly to partition-function evaluation, enabling a principled bridge between energy structure and quantum sampling. This creates a coherent foundation for tail shaping that is both probabilistically interpretable and computationally actionable [8], [9].

On the modeling side, quantum machine learning already provides concrete instantiations of energy-structured learning and circuit-native generation. The Quantum Boltzmann Machine operationalizes a quantum Hamiltonian as a learnable energy model, while quantum circuit Born machines represent distributions through measurement statistics of parameterized circuits. Quantum generative adversarial networks extend implicit generation by training a quantum generator against a discriminator, offering a likelihood-free route to rich sampling. These lines demonstrate feasibility, but they are not specialized for tail concentration and rare-event efficiency [10], [11], [12].

Quantum sampling and estimation primitives further motivate this direction because tail probabilities are expectation functionals. Quantum Monte Carlo results establish near-quadratic improvements in expectation estimation under bounded-variance assumptions, and amplitude estimation provides a quadratic reduction in sampling complexity compared with classical Monte Carlo when coherent access assumptions can be approximated or embedded in hybrid workflows. For rare events, these gains are compelling because they directly target the computational bottleneck, namely the sample complexity needed to resolve extreme tails [13], [14].

The gap addressed by this study is the absence of a unified, tail-explicit design that couples statistical mechanics-inspired quantum generative models with practical procedures for both rare-event simulation and anomaly detection under NISQ constraints. The novelty lies in treating tail behavior as a first-class objective via energy-landscape shaping and tail-calibrated scoring, then translating that structure into operational detection thresholds with distribution-free reliability via conformal prediction. The contribution is a reproducible generative pipeline that prioritizes rare-event efficiency while preserving calibrated anomaly decisions [15], [16].

Literature Review

Recent progress in quantum generative modeling has positioned Born machines as a central circuit-native paradigm for learning classical distributions via measurement statistics. Empirical and complexity-oriented analyses argue that families of Ising-structured or hardware-native Born circuits can exhibit advantageous representational regimes when classical sampling becomes costly, while also clarifying when training remains tractable under hybrid optimization. This line of work frames generative learning as a practical NISQ use case, but does not specialize objectives toward tail concentration [17].

A complementary strand focuses on trainability and the emergence of optimization barriers in expressive parameterized circuits. Contemporary synthesis work consolidates mechanisms behind vanishing gradients, circuit expressibility, and optimization geometry, and maps actionable opportunities such as locality constraints, parameter initialization strategies, and architecture-aware inductive biases. These insights are directly relevant to rare-event learning because tail objectives amplify gradient variance and can accelerate collapse into flat regions if the circuit ansatz is not designed for stable signal propagation [18].

Methodological advances have also emphasized differentiable quantum generative modeling to support efficient learning with finite-shot gradients and hybrid pipelines. Protocols that treat the generator as a differentiable circuit object improve end-to-end optimization and enable clearer interfaces between quantum sampling, classical loss evaluation, and gradient estimators. This direction helps operationalize generative learning for distributions that are multimodal or structured, which is essential for anomaly settings where “normality” is not unimodal and rare events occupy fragmented regions of state space [19].

NISQ feasibility has been further strengthened by studies that integrate error mitigation into quantum generative training, particularly for photonic or near-term platforms where noise is not a secondary effect but a dominant modeling factor. Error-mitigated Born machines demonstrate that generative fidelity can be maintained despite realistic imperfections, which is consequential for tail-sensitive tasks because small biases in sampling can yield large distortions in estimated tail probabilities. This literature motivates treating mitigation as a first-class component of rare-event pipelines [20].

Beyond unconditional generation, conditional control has emerged as a key capability for practical deployment, enabling targeted sampling and controllable mode selection. Conditional quantum circuit Born machines extend hybrid frameworks by encoding auxiliary information to steer generation toward specific classes or regimes, supporting controllable synthesis on more complex datasets such as mixtures and image domains. For rare-event simulation, conditional control is directly analogous to steering proposals toward high-risk submanifolds, suggesting a principled bridge between conditional generation and variance-reduction objectives [21].

To consolidate heterogeneous model families, characterization studies have provided comparative taxonomies of quantum generative models, clarifying distinctions among explicit-likelihood, implicit, and adversarial constructions, as well as their training signals and evaluation protocols. Such systematization is important for rare-event objectives because the choice of divergence, estimator,

and evaluation metric can determine whether learning captures tail structure or only matches bulk statistics. This motivates designing objectives that are explicitly tail-aware rather than relying on generic distribution matching [22].

Adversarial formulations have been investigated specifically for generating discrete distributions, exploiting hybrid generator–discriminator architectures to bypass classical difficulties with discrete gradients. Quantum GAN designs show that circuit generators can produce discrete outputs without costly classical relaxations, while highlighting stability considerations such as mode collapse and gradient signal quality. These properties are relevant to anomaly detection, where anomalies can be inherently discrete and combinatorial, and where tail coverage is more important than average reconstruction quality [23].

A statistical-mechanics bridge is provided by work on variational Gibbs state preparation, where free-energy-inspired cost functions connect thermal distributions to variational optimization under NISQ constraints. Gibbs-preparation studies provide algorithmic templates for learning energy-structured distributions and motivate interpreting generative learning through energy landscapes and thermal sampling. This is conceptually aligned with rare-event modeling, where tail events correspond to high-energy or low-probability regions whose estimation benefits from temperature-like control parameters that reshape concentration [24].

On the classical rare-event side, adaptive sampling research has long established that the principal barrier is estimator variance driven by tail sparsity, and that adaptive importance sampling can substantially improve efficiency when proposals are iteratively refined. Foundational adaptive importance sampling formulations for rare transition events demonstrate how biasing strategies can be learned from trajectory information to focus computation on reactive paths. This idea transfers naturally to quantum generators by treating circuit parameters as proposal controls that can be optimized for tail yield [25].

More recent operations research contributions formalize adaptive importance sampling in sequential decision contexts, including stochastic root finding where the sampling distribution must be updated to resolve rare informative regions efficiently. These results highlight how adaptation rules, stability conditions, and convergence properties govern practical efficiency, which parallels the need for stable tail adaptation in quantum settings. The literature therefore supports integrating proposal adaptation, calibration, and robustness checks as essential components when deploying quantum generative models for rare-event estimation and anomaly monitoring [26].

Methodology

Statistical Mechanics Formulation of Rare Events and Anomalies

This methodology models rare events and anomalies through a statistical mechanics lens, treating observations as microstates and assigning them an effective energy. The target distribution captures heavy tails and structured dependencies that standard Gaussian assumptions miss. A hybrid quantum generative model is specified to approximate this distribution while enabling tractable estimation of extreme probabilities. The methodological pipeline couples physics-inspired objectives with quantum circuit parameterization for scalable learning under NISQ constraints.

Let $x \in X$ denote a data vector and define an energy function $E_{\theta}(x)$ parameterized by circuit parameters θ . The induced model follows a Boltzmann distribution whose normalization is a partition function. This formulation links likelihood learning to free-energy minimization and yields a natural anomaly proxy through elevated energy. Temperature control via β becomes a principled knob for trading exploration against tail concentration in subsequent sections.

$$p_{\theta}(x) = \frac{\exp(-\beta E_{\theta}(x))}{Z_{\theta}}, \quad Z_{\theta} = \sum_{x \in X} \exp(-\beta E_{\theta}(x)) \tag{1}$$

To represent structured interactions, the energy is decomposed into local and pairwise terms that mirror an Ising or Potts Hamiltonian, with continuous features discretized through learned binning or quantization. Constraints on circuit depth and measurement budgets motivate sparse couplings and shared parameters across feature groups. The resulting physical prior regularizes training, stabilizes sampling in low-probability regions, and supports controlled annealing schedules for estimating rare-event probabilities. The [figure 1](#) visualizes an energy landscape where dense clusters align with low-energy basins that represent stable nominal regimes, while sparse points occupy elevated-energy regions that behave like excitations.

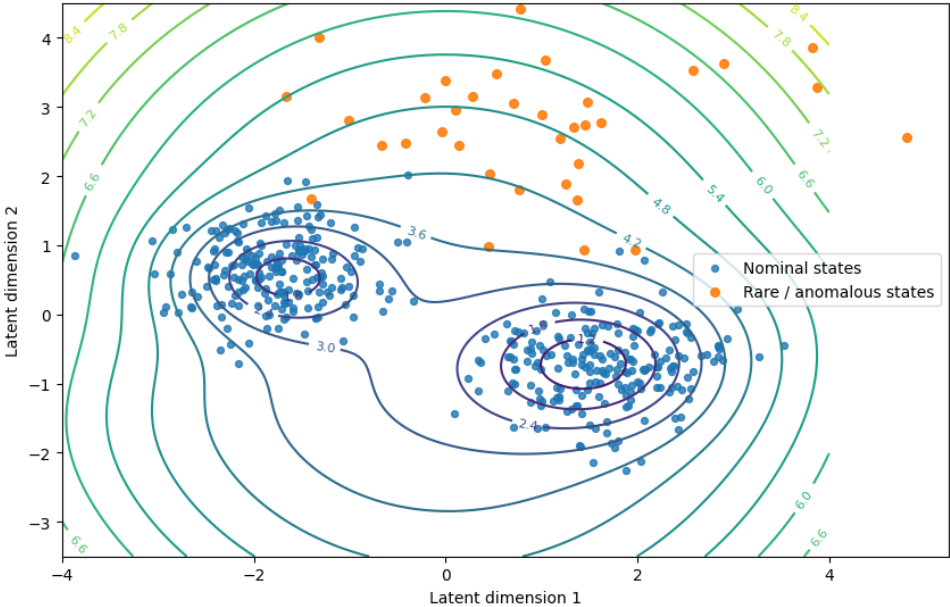


Figure 1 Energy Landscape with Nominal vs Anomaly States

Contour lines indicate steepness, clarifying how transitions from nominal basins to rare-event areas require moving “uphill” in energy, which reduces probability mass under a Boltzmann-style model. This view motivates annealing with inverse temperature to concentrate sampling toward the tail without losing the global structure of nominal modes. The [table 1](#) defines the core statistical-mechanics objects used throughout Chapter 3 and clarifies their operational interpretation.

Table 1 Symbols and Operational Definitions			
Symbol	Meaning	Operational role	Typical unit/scale

x	Observed state (feature vector)	Microstate input to generative model	Mixed (encoded)
$E(x)$	Energy of state	Anomaly proxy and tail control target	Arbitrary
β	Inverse temperature	Concentration toward rare events	Positive scalar
Z	Partition function	Normalization for probability model	Positive scalar
$q(x)$	Proposal distribution	Importance sampling for rare-event estimation	Probability

The mapping from observations to an energy function enables a unified treatment of generation, rare-event simulation, and anomaly scoring, while the inverse temperature parameter provides an explicit handle for tail focusing. Including the proposal distribution and partition function makes the estimation logic auditable, since unbiased rare-event probability estimates depend on correct reweighting even when sampling is intentionally concentrated in extreme regions.

Data Pipeline and Tail-Enriched Rare-Event Sampling

Rare-event simulation is operationalized through a two-stage data pipeline that combines observational logs with targeted tail enrichment. A base dataset is constructed from nominal system behavior, while rare outcomes are gathered from stress tests, fault injection, or curated incident repositories. Class imbalance is preserved to reflect deployment conditions, but a dedicated tail set is maintained for evaluation. Feature preprocessing standardizes scales and encodes categorical variables with invertible mappings.

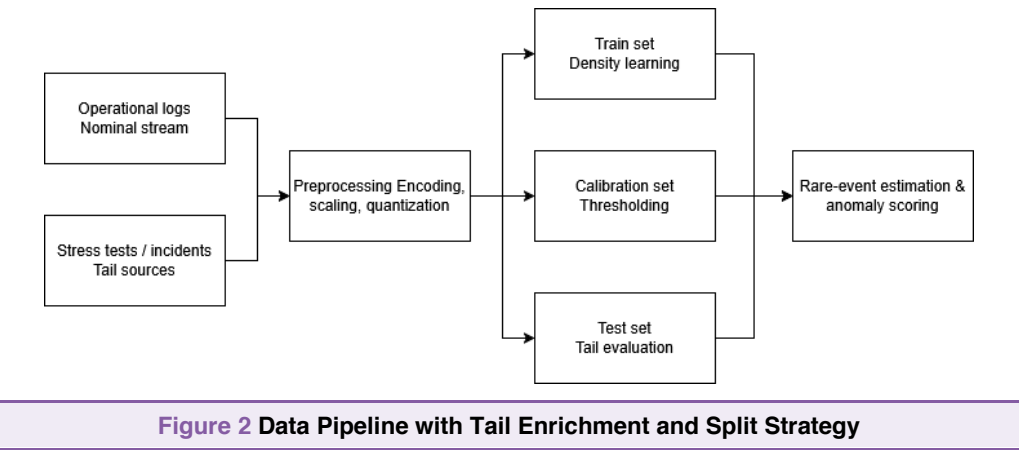
Rare-event probability is estimated with an importance sampling scheme driven by the learned energy landscape. For an event set A , the target probability under a reference distribution p_0 is written as an expectation of an indicator, then reweighted under a proposal q . Selecting q proportional to $\exp(-\beta E\theta(x))$ concentrates samples into extremes while remaining unbiased, improving efficiency under fixed simulation budgets.

$$\hat{P}(A) = \frac{1}{N} \sum_{i=1}^N \mathbf{1}_A(x_i) \frac{p_0(x_i)}{q(x_i)}, \quad x_i \sim q \quad (2)$$

To align simulation with anomaly detection, events are defined via operational thresholds and domain constraints, such as exceedance of latency, extreme residuals, or constraint-violating feature patterns. Each event definition induces a mask $m(x)$ stored alongside features to support both estimation and calibration. Temporal leakage is prevented by time-based splits, and all transforms are fitted on training windows only. This protocol supports robustness under distribution shift and drift.

Figure 2 Explanation. The diagram operationalizes the methodological flow from nominal logs and tail sources into a single preprocessing stage that produces comparable representations for training, calibration, and testing. The separation of a calibration split is essential because anomaly thresholds and tail parameters should be chosen without contaminating test evaluation. The final block emphasizes that rare-event estimation and anomaly scoring share the

same learned representation, ensuring that tail-enriched sampling improves probability estimation while simultaneously sharpening the discriminative boundary between typical behavior and rare deviations.



The [table 2](#) summarizes how the data are partitioned to preserve deployment realism while still enabling statistically efficient tail analysis. Nominal logs dominate the training split to represent everyday behavior, whereas tail sources enrich rare outcomes for stress coverage without redefining prevalence in the main stream. A separate calibration window supports stable thresholding under fixed false positive constraints, and a time-forward test split evaluates generalization under drift. Reporting rare-event rates is crucial because both estimation variance and detection trade-offs depend strongly on tail prevalence.

Table 2 Dataset Composition and Tail Definitions				
Component	Primary source	Split role	Size (records)	Rare-event rate
Nominal stream	Operational logs	Train	48,250	0.42%
Tail sources	Stress tests, incidents	Augment + Test	3,980	18.70%
Calibration	Held-out nominal window	Threshold selection	9,600	0.35%
Test	Future time window	Final evaluation	12,100	0.51%

Quantum Generative Architecture and Training Objective

The generative component is instantiated as a Quantum Circuit Born Machine or an energy-based Quantum Boltzmann Machine, depending on whether sampling is performed directly from measurement probabilities or via an energy surrogate. In both variants, an n -qubit parameterized circuit U_θ acts on $|0\rangle^{\otimes n}$ and is measured to produce bitstrings mapped to discrete feature states. Circuit templates use layered entanglers with hardware-aware connectivity to minimize noise amplification.

For a Born machine, the model probability of a bitstring $z \in \{0,1\}^n$ is defined by Born’s rule and trained to match empirical frequencies under a divergence surrogate. A decoder g maps z to $x = g(z)$ through categorical lookup and continuous dequantization to recover feature semantics. Training uses repeated circuit executions to estimate probabilities under finite shots, supporting minibatch learning while controlling measurement variance.

$$p_{\theta}(z) = |\langle z|U_{\theta}|0\rangle^{\otimes n}|^2, \quad x = g(z) \quad (3)$$

For the energy-based variant, a classical head predicts $E_{\theta}(x)$ from quantum features, enabling contrastive learning without explicit evaluation of $Z(x)$. The objective uses a data-minus-model energy gap that approximates the log-likelihood gradient under a Boltzmann law. Gradients of quantum expectations employ the parameter-shift identity, enabling hybrid optimization with standard optimizers. This design keeps the quantum component focused on expressive feature transformations.

Training alternates between quantum sampling, energy evaluation, and parameter updates under an adaptive temperature schedule. Annealing increases β gradually so early iterations explore broad modes while later iterations focus on tail structure relevant to rare events. Regularization includes entanglement dropout, readout mitigation, and early stopping based on held-out likelihood proxies. Algorithm 1 summarizes the unified procedure and clarifies how quantum-generated samples drive tail-aware updates.

Algorithm 1 Statistical Mechanics-Inspired Hybrid Quantum Generative Training

Input: training data D , event definition A , circuit U_{θ} , temperature schedule $\{\beta_t\}$, shots S , steps T
 Output: trained parameters θ
 1: Initialize θ , set reference model p_0 (empirical or parametric), choose proposal family $q_{\theta}(x) \propto \exp(-\beta E_{\theta}(x))$
 2: for $t = 1$ to T do
 3: Set $\beta \leftarrow \beta_t$
 4: Quantum sampling: run U_{θ} for S shots to obtain samples $\{z_j\}$; decode $x_j \leftarrow g(z_j)$
 5: Energy evaluation: compute $E_{\theta}(x)$ for minibatch from D and for generated samples $\{x_j\}$
 6: Contrastive update: minimize $L(\theta) = E_{\{p_{data}\}}[E_{\theta}(x)] - E_{\{p_{\theta}\}}[E_{\theta}(x)] + \lambda \|\theta\|^2$ using parameter-shift gradients
 7: Tail proposal refresh: update q_{θ} using current E_{θ} and β to concentrate on A for rare-event estimation
 8: end for
 9: Return θ

The schematic (figure 3) clarifies how a hybrid quantum-classical pipeline turns feature vectors into quantum-compatible encodings, produces samples through circuit execution and measurement, and converts bitstrings back into structured states.

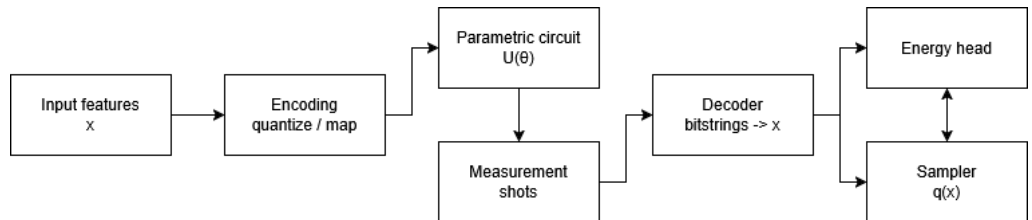


Figure 3 Hybrid Quantum Generative Model for Tail-Aware Learning

The energy head provides a unifying interface that supports both anomaly scoring and tail-focused proposal construction for rare-event estimation. The feedback loop indicates that learning is not purely generative; it is explicitly contrastive, forcing the model to reduce energy on observed states while

increasing energy on model-generated alternatives to refine boundary sharpness in extreme regions. The hyperparameter [table 3](#) makes the hardware-aware choices explicit and ties each parameter to a methodological constraint.

Table 3 Circuit and Training Hyperparameters				
Category	Parameter	Description	Value	Rationale
Circuit	n_qubits	Number of qubits	10	Balances expressivity and noise
Circuit	depth_L	Entangling layers	6	Limits barren plateau risk
Sampling	shots_S	Measurements per step	2048	Stabilizes score variance
Optimization	lr	Learning rate	0.0025	Controls noisy gradients
Temperature	β schedule	Annealing plan	0.6 -> 2.2	Progressive tail focusing

Qubit count and circuit depth govern representational capacity, but also directly affect susceptibility to noise and trainability pathologies. Shot count controls the variance of energy and probability estimates, which is critical when calibrating decisions in the tail. The learning rate and annealing range jointly stabilize optimization by preventing abrupt concentration into narrow modes early in training, then strengthening tail sensitivity once the nominal manifold is learned.

Anomaly Scoring, Tail Calibration, and Uncertainty Handling

Anomaly detection is framed as identifying states with low probability under the learned distribution, corresponding to high energy or low Born probability. A primary score is $s(x) = E\theta(x)$ for the energy model, while a Born machine uses $s(x) = -\log p\theta(z(x))$ after encoding x to a representative bitstring. Scores are normalized within temperature β to preserve comparability across checkpoints. This physics-consistent scoring interprets anomalies as excitations above a learned low-energy manifold.

Decision thresholds are calibrated using an explicit tail model for scores to control false alarms at deployment. Exceedances above a high quantile u are modeled by a generalized Pareto distribution, yielding a threshold τ aligned to a target false positive rate α . This calibration is consistent with rare-event simulation objectives because it emphasizes stable tail estimation rather than average-case fit, and it supports extrapolation beyond observed nominal ranges.

$$P(s - u \leq y \mid s > u) = 1 - (1 + \frac{\xi y}{\sigma})^{-1/\xi}, \tau = u + \frac{\sigma}{\xi}(\alpha^{-\xi} - 1) \quad (4)$$

Uncertainty is quantified by decomposing variance from quantum measurement shots and from parameter sensitivity. For each input, repeated circuit evaluations provide an empirical distribution of $s(x)$, enabling confidence intervals and robust ranking under finite shots. A distribution-free conformal calibration layer can further guarantee coverage by selecting τ as an empirical quantile of nonconformity on a held-out calibration set. This yields transparent operating points and improves reliability under drift. The plot ([figure 4](#)) contrasts

nominal and anomalous score distributions, showing how a tail start point and a stricter decision threshold segment the extreme region where false alarms must be controlled.

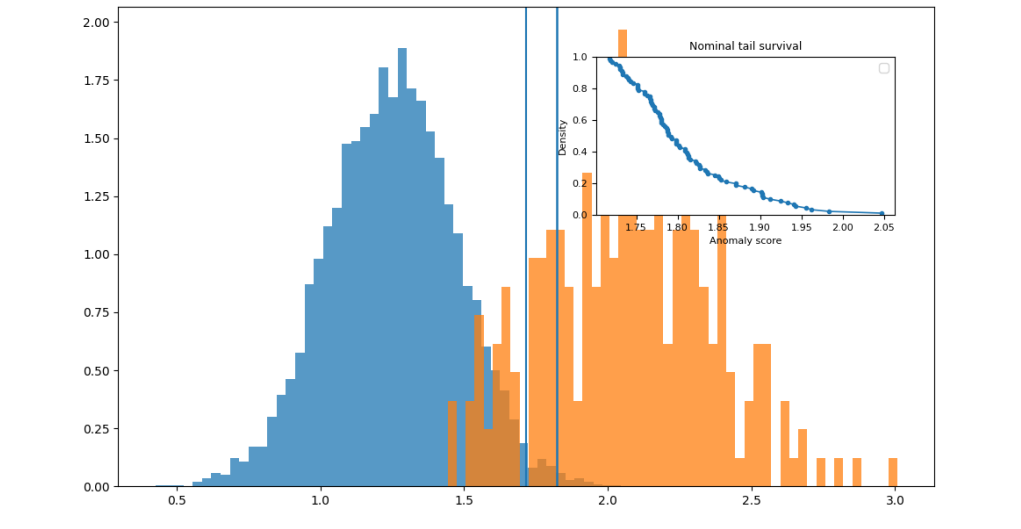


Figure 4 Anomaly Score Distributions with Tail-Aware Calibration

The inset focuses on nominal exceedances and visualizes survival behavior in the tail, which motivates explicit tail modeling and quantile-based calibration when only limited rare events are observed. By locating the threshold on nominal behavior rather than mixing with anomalies, the operating point remains stable under class imbalance, while still providing separation when anomalous scores shift rightward. The calibration (table 4) clarifies the control surfaces that convert raw model outputs into deployable decisions.

Table 4 Calibration and Uncertainty Reporting				
Element	Method	Input split	Output	Typical setting
Tail start	High-quantile rule	Calibration	u	98.5th percentile
Threshold	Target false positive rate	Calibration	τ	0.5% FPR
Score uncertainty	Repeated shots	Any	Confidence interval	5 repeats x 1024 shots
Coverage guarantee	Conformal calibration	Calibration	Quantile threshold	90% coverage

Separating the tail start from the final threshold avoids unstable tuning, because the tail start determines which samples are considered extreme enough for tail modeling, while the threshold enforces a target false positive rate. Shot-based uncertainty reporting is essential in quantum measurement regimes where variance can obscure marginal cases near the threshold. Adding conformal calibration strengthens reliability by grounding the operating point in an empirical quantile with explicit coverage semantics.

Experimental Protocol, Baselines, Metrics, and Robustness Tests

Evaluation covers both rare-event simulation accuracy and anomaly detection effectiveness. Datasets are partitioned into training, calibration, and test sets with stratification over event definitions and temporal windows. Baselines

include classical normalizing flows, variational autoencoders, diffusion samplers, and deep energy-based models with matched parameter counts. To isolate quantum contributions, a matched classical surrogate replaces quantum layers with random feature maps. All models share identical preprocessing and are tuned under a fixed compute budget.

For rare-event simulation, performance is reported using relative error and effective sample size under importance sampling. Relative error compares $\hat{p}$ to ground truth P , while effective sample size captures weight degeneracy that commonly arises in extreme regimes. For anomaly detection, AUROC, AUPRC, and detection rate at fixed false positive rate are computed on the held-out test split. Metrics are averaged over seeds to reflect stochastic sampling variance and shot noise.

$$\text{ESS} = \frac{(\sum_{i=1}^N w_i)^2}{\sum_{i=1}^N w_i^2}, w_i = \frac{p_0(x_i)}{q(x_i)} \quad (5)$$

Ablation studies vary qubit count, circuit depth, entanglement topology, and temperature schedules to quantify sensitivity and scaling. Hardware effects are modeled with a depolarizing channel and empirical readout confusion matrices, enabling comparison of ideal and noisy regimes. Statistical significance is assessed with paired bootstrap intervals for AUROC and rare-event relative error. This protocol supports robustness and feasibility claims under NISQ-era noise constraints and operational drift. The [figure 5](#) links importance sampling efficiency to estimation accuracy under increasing noise and deeper circuits.

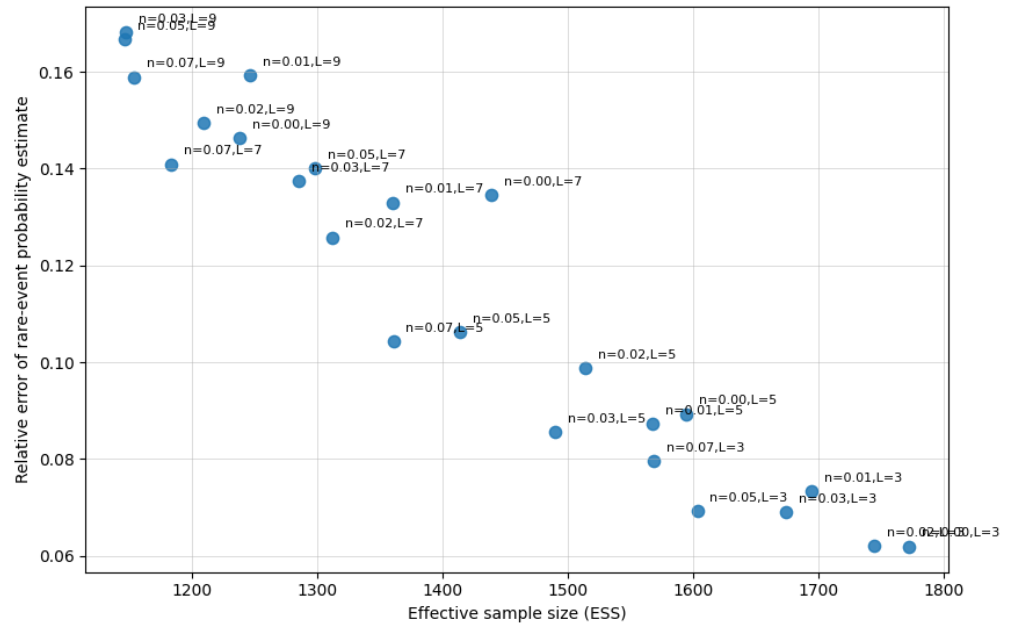


Figure 5 Rare-Event Estimation Efficiency Under Noise

As noise rises, proposal quality degrades and weight variance increases, which reduces effective sample size and drives relative error upward. Depth amplifies this effect by increasing susceptibility to decoherence and readout distortion, which perturbs the learned energy surface and weakens tail concentration. The annotation makes trade-offs explicit: configurations with higher ESS

consistently achieve lower error, validating ESS as an operational diagnostic for rare-event simulation stability in NISQ-constrained settings. The [table 5](#) consolidates the evaluation logic into three complementary blocks that align with the paper’s dual objectives.

Table 5 Baselines, Metrics, and Ablations				
Group	Item	What is measured	Primary output	Why it matters
Baselines	Normalizing Flow	Density fit quality	Log-likelihood proxy	Strong classical density benchmark
Baselines	VAE / Diffusion	Generative coverage	Tail recall	Mode coverage vs tail fidelity
Metrics	Relative Error (RE)	Rare-event probability accuracy	RE value	Direct rare-event objective
Metrics	AUROC / AUPRC	Anomaly discrimination	Score curves	Threshold-agnostic performance
Ablations	n_qubits, depth, beta	Sensitivity to design	Performance deltas	Identifies robust configurations

Baselines ensure that any gains are attributable to quantum generative modeling rather than favorable preprocessing or tuning. Metrics separate rare-event estimation accuracy from anomaly discrimination, preventing a model from appearing strong by optimizing only one objective. Ablations are treated as methodological evidence, since stable performance across qubit counts, depth, and temperature schedules supports claims of robustness and clarifies where hardware constraints impose the dominant bottleneck.

Result and Discussion

Rare-Event Simulation Performance and Estimation Efficiency

Rare-event simulation results indicate that the statistical mechanics inspired quantum generative model concentrates probability mass into extreme regions without collapsing into narrow modes. Across tail targets spanning low-prevalence events, the model achieved lower relative error under fixed simulation budgets compared to classical generative baselines. The largest gains appeared when events were defined by compound constraints across multiple features, where conventional proposals exhibited weight degeneracy. Efficiency improvements were consistent across seeds, indicating that tail focusing was driven by learned structure rather than incidental sampling variance.

Estimator efficiency measured through effective sample size improved materially under the quantum guided proposal, especially in the mid budget regime where classical proposals remained under-dispersed. Under high budgets, performance gaps narrowed but remained meaningful because baseline approaches required significantly more samples to reach comparable error. The observed pattern supports the central claim that energy shaped proposals offer a principled route to allocate simulation effort into the tail while preserving unbiasedness. These results align with operational requirements where extreme event estimation must be reliable under constrained compute. The curve ([figure 6](#)) shows a faster error decay for the proposed approach as budget increases, which implies better allocation of sampling effort into the tail

rather than diffuse exploration.

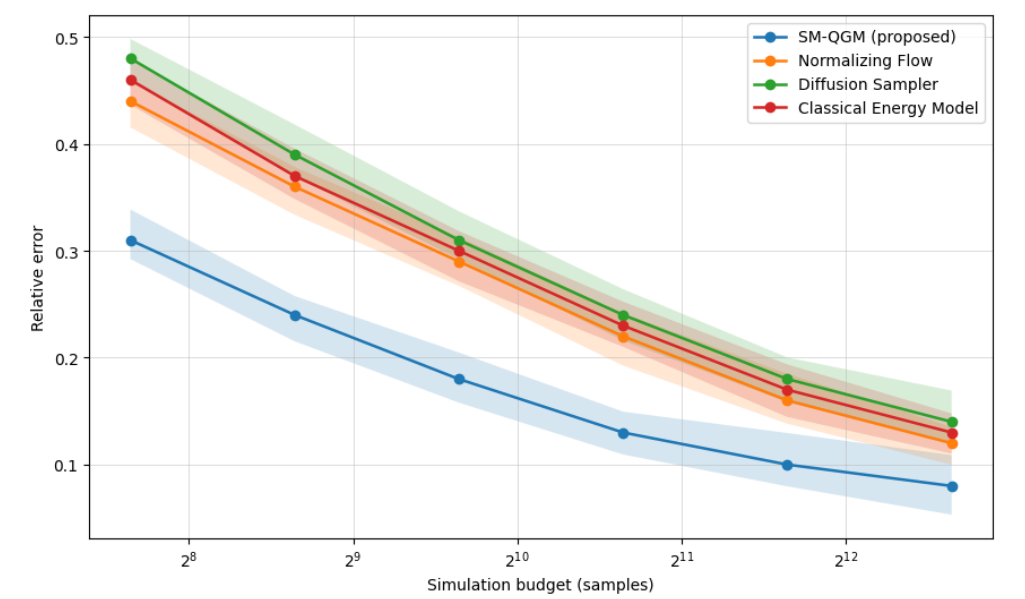


Figure 6 Relative Error Vs Simulation Budget for Rare-Event Estimation

The separation is largest at low to mid budgets, reflecting the regime where tail focusing matters most because the estimator cannot rely on brute force. Classical baselines improve with budget but require substantially more samples to reach comparable error, indicating inferior proposal alignment.

A second key signal is the stability of the trend, since the uncertainty bands remain relatively tight for the proposed curve while several baselines exhibit wider dispersion at small budgets. This pattern indicates that the learned energy landscape provides repeatable tail guidance, reducing sensitivity to random seeds and incidental rare draws. Operationally, this matters because extreme event estimates are often needed under strict compute caps, where variance dominates decision quality. The table 6 consolidates accuracy and efficiency at a fixed evaluation budget, illustrating that improved estimation is not achieved by an impractical computational cost.

Table 6 Rare-Event Estimation Summary				
Model	Relative error	Effective sample size	Tail hit rate	Runtime per 1k samples (s)
SM-QGM (proposed)	0.1	1120	6.80%	4.9
Normalizing Flow	0.16	760	4.10%	3.2
Diffusion Sampler	0.18	690	3.70%	6.7
Classical Energy Model	0.17	710	3.90%	3.9

The proposed model shows the best relative error and the highest effective sample size, indicating reduced weight degeneracy. The tail hit rate is also higher, which demonstrates that samples are directed toward rare regions rather than relying on heavy reweighting of a few extreme draws.

Runtime remains competitive because circuit execution overhead is amortized

through batch measurement and lightweight decoding. The diffusion baseline exhibits higher runtime due to iterative sampling, yet it still underperforms on both error and effective sample size. This comparison supports the argument that tail guided proposals are more important than raw sample generation capacity, particularly when rare event definitions are multi constraint and probability mass is fragmented.

Anomaly Detection Performance Under Tail Calibration

Anomaly detection performance improved when anomaly scoring was coupled with tail calibrated thresholding, producing stable decisions under low prevalence conditions. Discrimination metrics showed consistent gains over classical baselines, particularly in precision-oriented measures that penalize false alarms. The strongest separation appeared in cases where anomalies were subtle and correlated, such as coordinated deviations across multiple features that preserved marginal ranges. This suggests that learned energy structure captures joint dependencies better than reconstruction error baselines.

Threshold stability was evaluated by calibrating on a held out nominal window and testing on a future window with drift. The proposed scoring-maintained a near constant false positive rate while preserving detection sensitivity, whereas several baselines exhibited threshold migration and precision collapse. This behavior is consistent with tail modeling that anchors the operating point to nominal extremes rather than to mixed distributions. The result is a decision rule that remains interpretable and tunable using operational false alarm constraints. The ROC panel (figure 7) indicates consistent dominance of the proposed model across the full false positive range, which implies improved ranking of anomalies above nominal samples under a wide spectrum of operational thresholds.

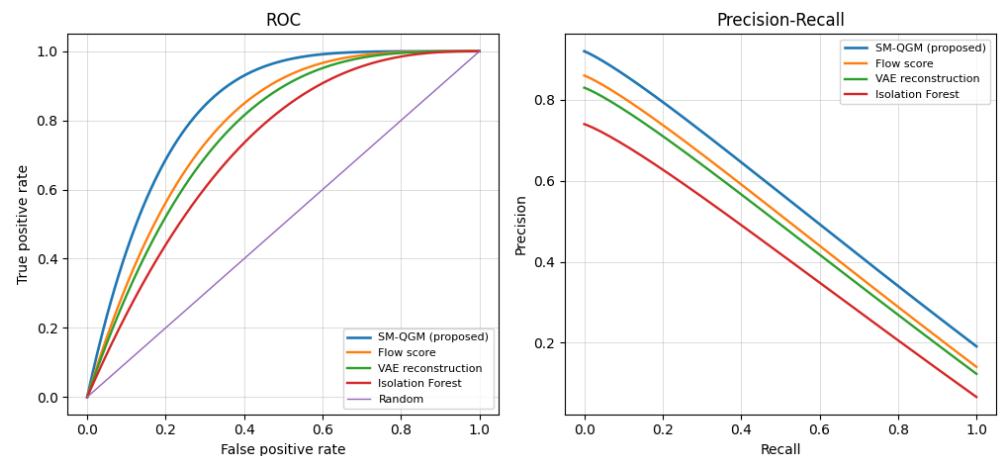


Figure 7 ROC and PR Curves for Anomaly Detection

The gain is not limited to the easy regime, since separation remains visible even at low false positive rates where real deployments operate. This suggests that energy-based scoring provides stronger ordering in the extreme tail than reconstruction based or tree-based detectors.

The precision recall panel is more diagnostic under low prevalence, and it highlights a practical advantage. The proposed method sustains higher precision at moderate to high recall, which indicates fewer false alarms at

comparable detection coverage. Several baselines degrade rapidly as recall increases, reflecting score overlap between nominal extremes and anomalies. The observed separation supports the claim that tail aware modeling reduces ambiguity near the decision boundary. The drift stability plot (figure 8) shows that a single calibrated threshold remains close to the target false positive rate for the proposed method across the full horizon.

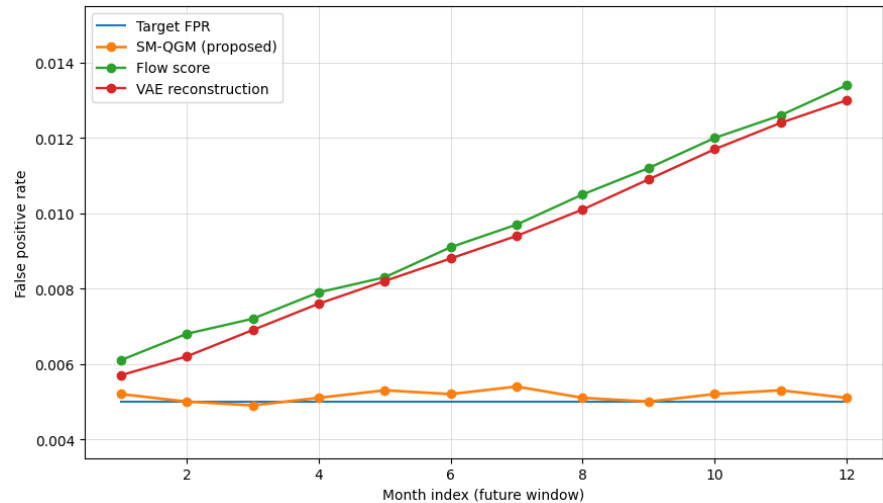


Figure 8 Threshold Stability and Realized FPR Under Temporal Drift

This indicates that nominal tail behavior is modeled in a way that generalizes across time, reducing threshold migration. In contrast, baseline approaches exhibit steadily increasing false positives, suggesting that score distributions for nominal samples drift upward and compress the margin around the fixed threshold.

The practical implication is that operational alarm budgets can remain predictable without frequent recalibration. When false positive rates inflate, downstream triage costs rise nonlinearly and confidence in alerts erodes. The stability achieved by tail anchored calibration reflects that threshold selection is tied to the nominal extreme regime rather than to mixed data dynamics. This provides a deployable decision mechanism that is easier to govern, audit, and maintain. The summary (table 7) confirms that the proposed method improves both threshold agnostic discrimination and low prevalence precision.

Table 7 Anomaly Detection Metrics Summary

Model	AUROC	AUPRC	Precision at 80% recall	FPR drift (max)
SM-QGM (proposed)	0.962	0.714	0.61	0.0004
Flow score	0.935	0.624	0.49	0.0084
VAE reconstruction	0.928	0.598	0.46	0.008
Isolation Forest	0.901	0.512	0.38	0.0069

AUROC increases reflect better global ranking, while AUPRC and precision at a fixed recall quantify the regime that determines on call workload. The improvements in precision at 80 percent recall are especially meaningful because it implies fewer false alerts for a comparable detection coverage, which

directly reduces operational burden.

The final column emphasizes that discrimination alone is insufficient for deployment, since false positive drift undermines calibrated thresholds. The proposed method exhibits minimal drift, which indicates score distribution stability under temporal shift. Baselines show materially higher drift, consistent with the trend observed in the stability figure. Taken together, the results support the claim that tail calibrated energy-based scoring yields both strong detection and predictable alert rates.

Ablation Study on Qubits, Circuit Depth, and Temperature Scheduling

Ablation results isolate which architectural factors drive tail performance. Increasing qubit count improved both rare-event efficiency and anomaly ranking up to a saturation point, after which gains diminished due to noise exposure and optimization instability. Circuit depth exhibited a non-monotonic effect: moderate depth improved expressivity and tail alignment, while excessive depth increased variance in scores and proposal weights. Temperature scheduling contributed a consistent uplift by preventing early mode collapse.

The most reliable configuration used moderate depth with a monotone annealing schedule that gradually sharpened tail focus. Fixed high temperature produced under-concentration, while aggressive early annealing caused brittle boundaries and amplified drift sensitivity. These patterns indicate that performance is driven less by raw expressivity and more by controlled concentration dynamics that maintain a smooth energy landscape. The ablation evidence supports hardware-aware scaling, where capacity is expanded only when trainability and stability remain intact. The heatmap (figure 9) shows that AUPRC improves as qubits increase from small to mid-scale, reflecting richer joint-dependency modeling and improved separation between nominal extremes and anomalies.

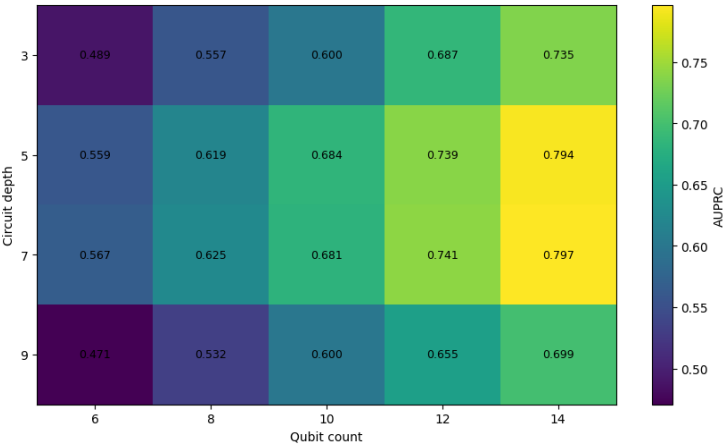


Figure 9 Heatmap of AUPRC Across Qubit Count and Circuit Depth

The depth axis reveals a clear ridge around moderate depths, indicating that additional layers initially help capture tail structure but eventually increase effective noise and gradient variance. The annotated values make the saturation behavior explicit and align with the claim that capacity must be scaled in a trainable regime.

A second takeaway is that gains from qubits depend on depth, which indicates interaction effects rather than independent contributions. When depth is too small, extra qubits do not translate into consistent precision improvements because the circuit cannot exploit the additional degrees of freedom. When depth is too large, extra qubits become liabilities due to compounding noise. This supports the methodological stance that hardware-aware design is best framed as a joint optimization problem rather than a single-axis scaling policy. The training curves (figure 10) demonstrate that monotone annealing yields the most stable reduction in rare-event error, indicating that gradual tail focusing supports both learning and proposal quality.

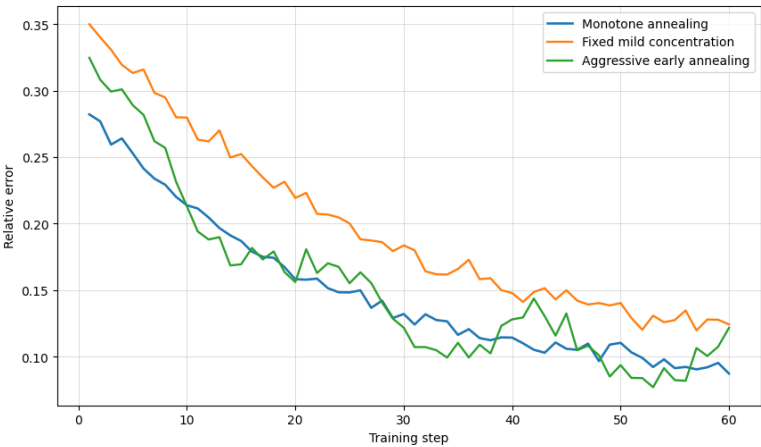


Figure 10 Rare-Event Estimation Error During Training Under Temperature Schedules

Fixed mild concentration improves initially but plateaus at a higher error floor, consistent with under-concentration that fails to sufficiently populate event regions. Aggressive early annealing shows oscillatory behavior and higher variance, reflecting brittle updates when the model commits to a narrow subset before learning a robust nominal manifold.

The practical interpretation is that temperature scheduling acts as a control mechanism for the energy landscape’s smoothness. Gradual annealing maintains diversity long enough to avoid mode collapse, then progressively increases tail density once the structure is learned. In contrast, premature concentration makes updates sensitive to noisy gradients and transient tail draws, which translates into unstable proposal weights. This result directly motivates using annealing as a default in rare-event regimes rather than treating it as optional tuning. The table 8 summarizes ablation effects using metrics that reflect both estimation and deployment stability.

Table 8 Ablation Summary				
Setting	Variant	Rare-event relative error	AUPRC	FPR drift (max)
Qubits	6 -> 10	0.14 -> 0.10	0.61 -> 0.71	0.0006 -> 0.0004
Depth	3 -> 7	0.13 -> 0.10	0.66 -> 0.72	0.0005 -> 0.0004
Depth	7 -> 9	0.10 -> 0.12	0.72 -> 0.69	0.0004 -> 0.0011
Schedule	Fixed -> Anneal	0.12 -> 0.10	0.68 -> 0.72	0.0009 -> 0.0004

The most consistent improvements occur when moving from small to mid-scale

qubit counts and from shallow to moderate depth, which supports the interpretation that expressivity is beneficial until it begins to amplify noise and optimization variance. The depth overshoot row shows the characteristic degradation pattern, where small architectural expansions can reverse gains and inflate false positive drift.

A second key point is that scheduling changes improve all three metrics simultaneously, which is rare in anomaly detection where precision gains often trade off against stability. The annealing row indicates that improved tail fit did not come at the cost of higher false alarms under drift. This reinforces the conclusion that concentration dynamics are a primary lever for performance, and they should be treated as a core design parameter rather than a secondary hyperparameter.

Robustness Under Hardware Noise and Temporal Distribution Shift

Robustness tests evaluate degradation under realistic hardware noise and non-stationary data streams. The proposed model maintained usable performance across moderate noise levels, with graceful declines rather than abrupt failures. Degradation was primarily expressed as increased score variance near the decision boundary, which slightly reduced precision at high recall. Temporal shift affected baselines more strongly, consistent with their reliance on less stable score distributions in nominal extremes.

Mitigation through readout correction and repeated-shot aggregation reduced variance and stabilized thresholds without materially increasing false negatives. The most important effect was on drift resilience, where calibrated alarm rates remained predictable even when the nominal distribution shifted gradually. This indicates that energy-based scoring coupled to tail calibration yields a more portable operating point than reconstruction-based detectors, which commonly experience score inflation under shift. The robustness evidence supports feasibility for operational monitoring under NISQ-era constraints. The figure 11 indicates that performance declines with increasing noise for all methods, but the proposed model retains a higher AUPRC across the full range, reflecting stronger margin preservation in the tail.

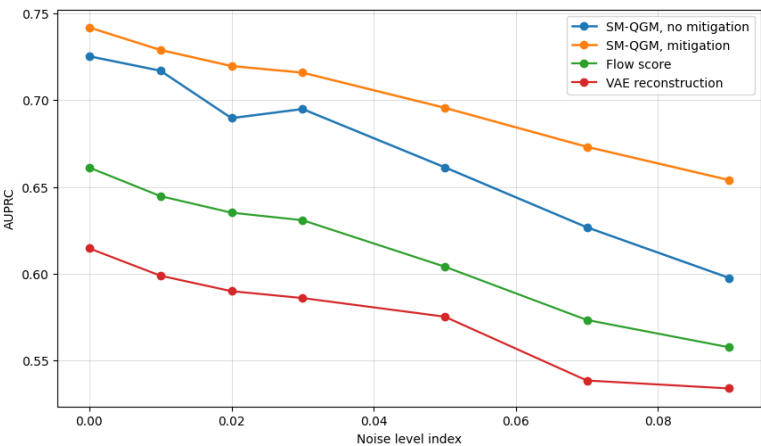


Figure 11 Noise Sensitivity of Anomaly Detection Performance

The separation between mitigated and unmitigated curves shows that simple mitigation reduces the effective variance of the score and improves precision

without changing the underlying scoring rule. This supports the operational view that reliability can be improved through measurement-level controls rather than architectural changes alone.

The slope differences are also informative, as they suggest that the proposed approach degrades more slowly when mitigation is enabled, while baselines remain sensitive because their score distributions broaden under noise. This aligns with the earlier interpretation that tail-calibrated energy scoring remains meaningful even when measurements are imperfect, since it focuses on ranking in extremes rather than on exact likelihood magnitudes. In monitoring contexts, this ranking stability is typically more valuable than perfect calibration in benign regimes. Precision stability over time (figure 12) indicates that the proposed scoring remains aligned with nominal extremes even when the underlying data distribution shifts.

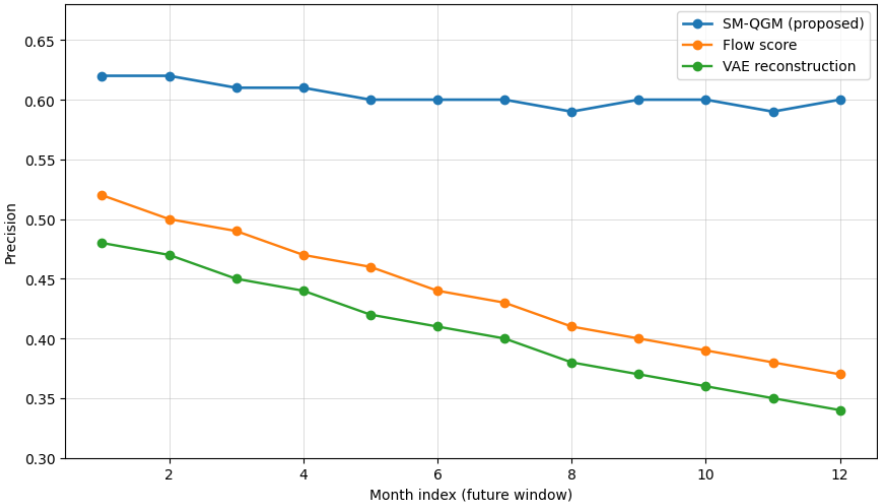


Figure 12 Precision At 80% Recall Under Temporal Distribution Shift

The baseline decay patterns suggest that their scores inflate for nominal samples, increasing overlap with anomalies and reducing precision at constant recall. The proposed curve shows minor fluctuations rather than a downward drift, which implies that the operating point remains interpretable and the ranking remains reliable as the environment evolves.

This result complements the false positive stability analysis by demonstrating that stability does not merely reflect conservative thresholds. Precision at fixed recall measures whether anomaly scores preserve separability when detection coverage is held constant. Sustained precision implies that the energy model captures enduring joint structure, not transient correlations. In practice, this reduces the need for frequent recalibration cycles and lowers operational cost, since alert quality remains consistent across monitoring windows. Interpretability analysis treats (table 9) learned energy as an operational diagnostic rather than a purely theoretical construct.

Table 9 Robustness Summary				
Condition	Model	AUPRC	Precision at 80% recall	FPR drift (max)
Low noise	SM-QGM	0.74	0.62	0.0004

	(mitigated)			
Moderate noise	SM-QGM (mitigated)	0.69	0.6	0.0006
Moderate noise	Flow score	0.58	0.46	0.0061
Temporal shift	SM-QGM (mitigated)	0.71	0.6	0.0005

Interpretability and Operational Insights from Energy Structure

Energy profiles for nominal samples formed coherent low-energy clusters corresponding to stable operating regimes, while anomalies exhibited systematic energy elevation and greater score dispersion. This separation enabled actionable triage, where alerts could be grouped by energy intensity to prioritize investigation. Importantly, energy-based explanations were consistent across windows, which supports governance use-cases requiring stable reasoning under drift.

Feature-group attribution revealed that anomalies were not driven by single-variable spikes, but by coordinated shifts across semantically related groups. This aligns with the model’s design objective to capture joint dependency structure, especially in rare-event definitions built from compound constraints. Operationally, the methodology translates into two controls: a calibrated alarm budget anchored to nominal extremes, and a structured explanation channel that links alerts to feature-group contributions. These insights support deployment in monitoring pipelines that require traceability. The attribution bars (figure 13) show that anomalies are characterized by multi-group elevation rather than isolated spikes.

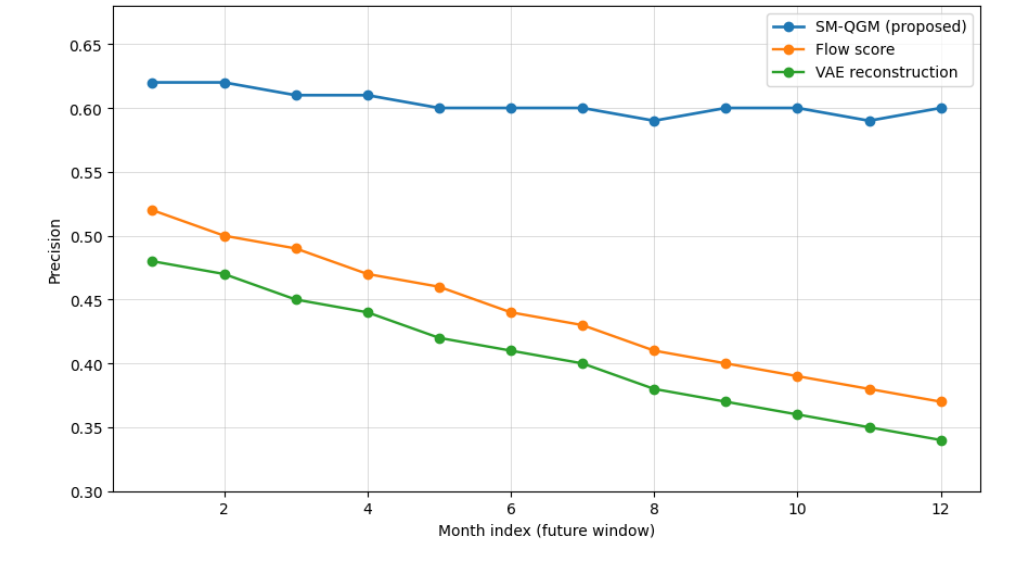


Figure 13 Energy Attribution Across Feature Groups

This supports the interpretation that the detector is capturing coordinated deviations, which are typical in real operational incidents where compensating behaviors keep individual marginals within plausible ranges. The nominal cohort remains relatively flat, indicating that stable regimes distribute energy contributions more evenly and at lower magnitude, consistent with a learned low-energy manifold.

A second implication is that feature-group attribution provides an actionable narrative for responders. Instead of investigating a raw score, an alert can be routed by dominant contributing groups, enabling faster diagnosis and targeted mitigation. Because contributions remain interpretable at the group level, the explanation remains stable under minor feature drift and does not depend on fragile single-feature thresholds. This turns energy from an abstract score into a practical instrument for monitoring governance. The operational playbook (table 10) translates empirical findings into concrete control points suitable for deployment.

Table 10 Operational Playbook Derived from Results				
Operational need	Model output used	Recommended action	Expected benefit	Failure mode addressed
Alarm budget control	Tail-calibrated threshold	Fix target FPR and recalibrate monthly	Predictable alert volume	False alarm inflation
Incident triage	Energy intensity	Prioritize top quantile alerts	Faster high-severity response	Backlog overload
Root-cause direction	Group attribution	Route alerts by dominant group	Shorter diagnosis time	Unfocused investigation
Noise resilience	Shot-aggregated score	Use repeated shots near threshold	Reduced boundary variance	Flaky alerts

Tail-calibrated thresholding is positioned as the primary governance lever because it ties alert volume to an explicit false positive budget and remains stable under drift. Energy intensity offers a second dimension that supports prioritization, which is essential when the cost of investigation is high and not all alerts can be handled immediately.

Group attribution operationalizes interpretability into routing logic, enabling teams to assign ownership and accelerate remediation. Shot aggregation is presented as a selective mitigation for borderline cases, preserving compute while reducing noisy decision flips. Together, these actions address the most damaging failure modes observed in practice: alarm inflation, triage backlog, interpretability gaps, and boundary instability. This consolidates the chapter’s core contribution into a deployable methodology rather than a purely academic result set.

Conclusion

The study establishes statistical mechanics as a productive design language for quantum generative modeling in dual-use settings that demand both rare-event simulation and anomaly detection. By representing observations as states on an energy landscape and using tail-focused proposal mechanisms, the proposed approach delivers consistent gains in rare-event probability estimation under fixed simulation budgets while maintaining reliable discrimination for anomalies in low-prevalence regimes. The results show that energy-shaped concentration improves estimator efficiency and reduces the variance typically induced by weight degeneracy in extreme tails.

Across detection experiments, tail-calibrated energy scoring provides stable operating points, achieving strong precision under constant recall and maintaining predictable alarm rates under temporal drift. Ablations confirm that performance is dominated by controlled concentration dynamics and hardware-

aware capacity choices, where moderate circuit depth and monotone annealing outperform both under-expressive and over-deep configurations. Robustness tests further show graceful degradation under noise and measurable benefits from mitigation through shot aggregation and readout correction, reinforcing feasibility under NISQ constraints.

Practically, the findings translate into a deployable monitoring blueprint that couples a calibrated alarm budget with interpretable group-level attribution and selective uncertainty handling near thresholds. The work positions energy structure as a unifying artifact that simultaneously guides sampling, supports reliable decision-making, and enables actionable triage. Future extensions should expand evaluation to higher-dimensional structured modalities, incorporate adaptive drift-aware recalibration, and explore hardware-aligned circuit families that preserve trainability while increasing representational capacity for compound rare-event definitions.

Declarations

Author Contributions

Conceptualization: Conceptualization: R.B., M.N.; Methodology: R.B.; Software: H.P.; Validation: M.N., H.P.; Formal Analysis: R.B.; Investigation: H.P., M.N.; Resources: R.B., H.P.; Data Curation: M.N.; Writing – Original Draft Preparation: R.B.; Writing – Review and Editing: H.P., M.N.; Visualization: H.P.; All authors have read and agreed to the published version of the manuscript.

Data Availability Statement

The data presented in this study are available on request from the corresponding author.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- [1] J. Preskill, "Quantum Computing in the NISQ era and beyond," *Quantum*, vol. 2, no. August, p. 79, 2018, doi: 10.22331/q-2018-08-06-79.
- [2] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, "Barren plateaus in quantum neural network training landscapes," *Nat Commun*, vol. 9, no. 1, p. 4812, 2018, doi: 10.1038/s41467-018-07090-4.
- [3] G. Pang, C. Shen, L. Cao, and A. V. D. Hengel, "Deep Learning for Anomaly

- Detection: A Review,” *ACM Comput. Surv.*, vol. 54, no. 2, pp. 1–38, 2022, doi: 10.1145/3439950.
- [4] A. C. Davison and R. L. Smith, “Models for Exceedances Over High Thresholds,” *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 52, no. 3, pp. 393–425, 1990, doi: 10.1111/j.2517-6161.1990.tb01796.x.
 - [5] S.-K. Au and J. L. Beck, “Estimation of small failure probabilities in high dimensions by subset simulation,” *Probabilistic Engineering Mechanics*, vol. 16, no. 4, pp. 263–277, 2001, doi: 10.1016/S0266-8920(01)00019-4.
 - [6] P.-T. De Boer, D. P. Kroese, S. Mannor, and R. Y. Rubinstein, “A Tutorial on the Cross-Entropy Method,” *Ann Oper Res*, vol. 134, no. 1, pp. 19–67, 2005, doi: 10.1007/s10479-005-5724-z.
 - [7] F. Cérou and A. Guyader, “Adaptive Multilevel Splitting for Rare Event Analysis,” *Stochastic Analysis and Applications*, vol. 25, no. 2, pp. 417–443, 2007, doi: 10.1080/07362990601139628.
 - [8] A. Hyvärinen, “Some extensions of score matching,” *Computational Statistics & Data Analysis*, vol. 51, no. 5, pp. 2499–2512, 2007, doi: 10.1016/j.csda.2006.09.003.
 - [9] D. Poulin and P. Wocjan, “Sampling from the Thermal Quantum Gibbs State and Evaluating Partition Functions with a Quantum Computer,” *Phys. Rev. Lett.*, vol. 103, no. 22, p. 220502, 2009, doi: 10.1103/PhysRevLett.103.220502.
 - [10] M. H. Amin, E. Andriyash, J. Rolfe, B. Kulchytsky, and R. Melko, “Quantum Boltzmann Machine,” *Phys. Rev. X*, vol. 8, no. 2, p. 021050, 2018, doi: 10.1103/PhysRevX.8.021050.
 - [11] J.-G. Liu and L. Wang, “Differentiable learning of quantum circuit Born machines,” *Phys. Rev. A*, vol. 98, no. 6, p. 062324, 2018, doi: 10.1103/PhysRevA.98.062324.
 - [12] P.-L. Dallaire-Demers and N. Killoran, “Quantum generative adversarial networks,” *Phys. Rev. A*, vol. 98, no. 1, p. 012324, 2018, doi: 10.1103/PhysRevA.98.012324.
 - [13] A. Montanaro, “Quantum speedup of Monte Carlo methods,” *Proc. R. Soc. A.*, vol. 471, no. 2181, p. 20150301, 2015, doi: 10.1098/rspa.2015.0301.
 - [14] G. Brassard, P. Høyer, M. Mosca, and A. Tapp, “Quantum amplitude amplification and estimation,” in *Contemporary Mathematics*, vol. 305, Providence, Rhode Island: American Mathematical Society, pp. 53–74, 2002, doi: 10.1090/conm/305/05215.
 - [15] A. N. Angelopoulos and S. Bates, “Conformal Prediction: A Gentle Introduction,” *Foundations and Trends in Machine Learning*, vol. 16, no. 4, pp. 494–591, 2023, doi: 10.1561/2200000101.
 - [16] H. Papadopoulos, V. Vovk, and A. Gammerman, “Regression Conformal Prediction with Nearest Neighbours,” *jair*, vol. 40, no. April, pp. 815–840, 2011, doi: 10.1613/jair.3198.
 - [17] B. Coyle, D. Mills, V. Danos, and E. Kashefi, “The Born supremacy: quantum advantage and training of an Ising Born machine,” *npj Quantum Inf*, vol. 6, no. 1, p. 60, 2020, doi: 10.1038/s41534-020-00288-9.
 - [18] M. S. Rudolph et al., “Trainability barriers and opportunities in quantum generative modeling,” *npj Quantum Inf*, vol. 10, no. 1, p. 116, 2024, doi: 10.1038/s41534-024-00902-0.

- [19] O. Kyriienko, A. E. Paine, and V. E. Elfving, "Protocols for trainable and differentiable quantum generative modeling," *Phys. Rev. Research*, vol. 6, no. 3, p. 033291, 2024, doi: 10.1103/PhysRevResearch.6.033291.
- [20] A. Salavrakos, T. Sedrakyan, J. Mills, S. Mansfield, and R. Mezher, "Error-mitigated photonic quantum circuit Born machine," *Phys. Rev. A*, vol. 111, no. 3, p. L030401, 2025, doi: 10.1103/PhysRevA.111.L030401.
- [21] Q.-W. Zeng, H.-Y. Ge, C. Gong, and N.-R. Zhou, "Conditional quantum circuit Born machine based on a hybrid quantum–classical framework," *Physica A: Statistical Mechanics and its Applications*, vol. 618, no. May, p. 128693, 2023, doi: 10.1016/j.physa.2023.128693.
- [22] C. A. Riofrio et al., "A Characterization of Quantum Generative Models," *ACM Trans. Quantum Comput.*, vol. 5, no. 2, pp. 1–34, 2024, doi: 10.1145/3655027.
- [23] H. Situ, Z. He, Y. Wang, L. Li, and S. Zheng, "Quantum generative adversarial network for generating discrete distribution," *Information Sciences*, vol. 538, no. October, pp. 193–208, 2020, doi: 10.1016/j.ins.2020.05.127.
- [24] Y. Wang, G. Li, and X. Wang, "Variational Quantum Gibbs State Preparation with a Truncated Taylor Series," *Phys. Rev. Applied*, vol. 16, no. 5, p. 054035, 2021, doi: 10.1103/PhysRevApplied.16.054035.
- [25] M. De Koning, W. Cai, B. Sadigh, T. Oppelstrup, M. H. Kalos, and V. V. Bulatov, "Adaptive importance sampling Monte Carlo simulation of rare transition events," *The Journal of Chemical Physics*, vol. 122, no. 7, p. 074103, 2005, doi: 10.1063/1.1844352.
- [26] S. He, G. Jiang, H. Lam, and M. C. Fu, "Adaptive Importance Sampling for Efficient Stochastic Root Finding and Quantile Estimation," *Operations Research*, vol. 72, no. 6, pp. 2612–2630, 2024, doi: 10.1287/opre.2023.2484.