Stress-Information Duality in QUBO Formulations for Power Grid Sensor Placement

R.J. Mathews
mail.rjmathews@gmail.com
Chattanooga, TN

***Abstract*—Sensor placement for power grid stress monitoring is commonly formulated as a QUBO with distance-based dispersion objectives. We show this produces catastrophically poor placements on radial networks — worse than 99% of random alternatives — because electrical distance anti-correlates with mutual information in tree topologies (Laplacian proxy r = -0.37; DC cascade ensemble r = -0.45). Maximizing sensor dispersion maximizes information redundancy.**

The root cause is a stress-information duality specific to radial networks: bus-level stress centrality is a sufficient proxy for information centrality. In the IEEE 33-bus tree, junction nodes are simultaneously stress attractors and informationally privileged under the tree Markov property. Any QUBO formulation preserving stress-weight dominance — coverage, low-alpha dispersion, or explicit MI — achieves 100% hub detection accuracy. The dispersion-dominant formulation ($\alpha \geq 1.0$) fails because its distance term drowns the stress signal. An alpha sensitivity sweep across all 237,336 feasible placements (k=5) maps a sharp phase transition at $\alpha \approx 0.8$ where hub accuracy collapses discontinuously from 100% to 20%. A comparative experiment on the IEEE 57-bus meshed transmission system confirms the structural prediction: no phase transition occurs (100% hub accuracy at all alphas), because the meshed topology's redundant paths break the tree Markov property that causes the duality. A DC power flow cascade ensemble (2,000 scenarios) confirms that the MI-distance anti-correlation holds under realistic stress propagation (empirical r = -0.45 vs proxy r = -0.37), and the dispersion-optimal placement captures the least information of all tested formulations under both proxy and empirical covariance.

These findings emerge from a formulation-to-validation pipeline that compares four QUBO objectives across classical solvers (brute-force, SA, MILP, greedy) and evaluates each placement against cascade detection metrics. The methodology — formulate, solve optimally, validate against domain metrics, diagnose divergence — is generalizable to any engineering QUBO application. As a quantum-readiness demonstration, we execute QAOA on a 20-qubit subproblem, recovering the exact optimum at all depths (p=1-4) with multi-start COBYLA (9/10 success rate), confirming that the corrected QUBOs are executable on near-term quantum hardware.

Keywords: QUBO formulation, sensor placement, power system monitoring,
stress-information duality, radial networks, validation methodology, QAOA

I. Introduction

A. Motivation: QUBO Formulation Quality Dominates Solver Quality

Combinatorial optimization problems in power systems — sensor placement, topology reconfiguration, contingency screening — are natural candidates for QUBO encoding and quantum optimization. But formulating an engineering problem as a QUBO requires choosing an objective function, and the wrong choice can render the mathematical optimum operationally useless. This paper demonstrates that QUBO formulation quality dominates solver quality: a perfectly solved wrong formulation produces worse placements than random guessing, while several different correct formulations all succeed regardless of solver choice.

Power grid stress monitoring via the Jacobian commutator Λ_G[1] requires optimal sensor placement — a combinatorial problem (n-choose-k) that scales exponentially and is naturally QUBO-encodable. The standard approach encodes stress-weighted dispersion as the objective, using electrical distance as a proxy for information independence — an assumption inherited from the facility-location literature that fails for structured networks. We show that in radial networks, this assumption produces placements worse than 99% of random alternatives, and we map the failure mode to a sharp phase transition in the dispersion weight parameter alpha. The root cause is structural: in tree-topology networks, electrical distance anti-correlates with mutual information, so maximizing sensor dispersion maximizes information redundancy.

The quantum aspect of this work (QAOA circuit design and simulator validation) confirms that these QUBOs are executable on quantum hardware, but the formulation finding applies equally to classical QUBO solvers — simulated annealing, D-Wave quantum annealing, and mixed-integer linear programming all produce the same operationally poor placement when given the same dispersion-dominant objective.

This paper asks: when QUBO-optimal sensor placement fails operationally, is the failure in the solver (quantum or classical) or in the formulation? And can the failure mode be diagnosed from the network's information structure?

B. The Λ_GDiagnostic

The power flow Jacobian J encodes the sensitivity of bus voltages and angles to power injections [9]. Under time-varying loading conditions, J evolves, and its time derivative Jdot captures the rate of change of this sensitivity structure. The commutator diagnostic Λ_G= ||[J, Jdot]||_F, introduced and validated in [1, 2], measures the degree to which the Jacobian's sensitivity structure fails to commute with its own time evolution. Nonzero $\Lambda_G$indicates that the system's linearized dynamics are rotating in state space — a signature of localized stress concentration that eigenvalue-based indicators (e.g., minimum singular value of J) do not capture.

The commutator admits a bus-level decomposition: $\Lambda_G^{(i)}$quantifies the contribution of bus i to the total system stress. In cascade simulations on IEEE test systems [1], $\Lambda_G^{(i)}$consistently identifies the network's high-centrality junction buses as stress attractors — notably, stress concentrates at topological hubs rather than at the buses where contingency events originate. The 2003 Northeast blackout exhibited precisely this pattern: cascade propagation concentrated at transmission corridor intersections, not at the initial trip locations [10].

This bus-level decomposition motivates sensor placement: instrumenting the k buses with highest expected $\Lambda_G^{(i)}$provides disproportionate observability of cascade stress. However, choosing k locations from n candidates is a combinatorial optimization problem — C(n, k) grows exponentially in k — and the optimal placement depends on the objective function used to define "best coverage." This paper formulates the problem as a QUBO and demonstrates that objective function choice has a larger effect on placement quality than solver choice. Full classical validation of $\Lambda_G$on IEEE 14-bus, 33-bus, and 118-bus systems is provided in [1].

C. Prior Work and Gap

Quantum computing applications in power systems fall into three categories, each at a different maturity level.

Quantum phase estimation (QPE) for eigenanalysis. Eskandari et al. [4] demonstrated QPE [14] for power system small-signal stability analysis, computing eigenvalues of the state matrix. However, power flow Jacobians are generally non-Hermitian, and Hermitianization (embedding into a larger Hermitian matrix) doubles the qubit count and introduces spectral artifacts. The commutator-based diagnostic $\Lambda_G$avoids this barrier because ||[J, Jdot]||_F is a Frobenius norm of Hermitian-structured matrices, but extracting this norm quantumly requires trace estimation rather than QPE (Section V).

Linear system solvers (HHL) for power flow. The HHL algorithm [13] solves linear systems in $\mathcal{O}(\text{polylog}(n))$ time, suggesting application to the linearized power flow equations J * delta_V = delta_P. However, the quantum speedup applies to producing the solution as a quantum state; extracting classical bus voltages requires O(n) measurements, negating the advantage for full-state problems such as voltage stability indices (L-index). For the sensor placement problem in this paper, power flow is a classical preprocessing step — the quantum component addresses the combinatorial optimization, not the power flow itself.

Combinatorial optimization via quantum annealing and QAOA. This is the most mature quantum application in power systems. D-Wave quantum annealers have been applied to unit commitment, optimal power flow subproblems, and network reconfiguration, where the natural QUBO structure of binary decision variables maps directly to Ising Hamiltonians. QAOA [3] provides a gate-model alternative with theoretical performance guarantees for certain problem classes [11]. Our work falls in this category: sensor placement is a constrained binary optimization problem (select k buses from n) that maps to a QUBO with n binary variables.

Gap this paper fills. The sensor placement literature — both classical PMU placement [20, 21] and quantum optimization for power systems [22, 23] — focuses on solver performance (approximation ratios, convergence speed, noise robustness) and takes the objective formulation as given. The question of how QUBO objective choice affects the operational quality of solutions has received limited attention. In particular, no prior work diagnoses topology-dependent failure modes of standard distance-based objectives or provides end-to-end validation connecting QUBO-optimal solutions to domain-specific cascade detection metrics. The formulation gap we discover (Section VII-D) and its structural explanation via stress-information duality (Section VIII-D) are, to our knowledge, novel contributions to both the quantum optimization and power systems sensor placement literature.

D. Contributions

1. **End-to-end QUBO pipeline with validation loop.** We formulate Λ_G-optimal sensor placement as four QUBO variants (stress-dispersion, coverage, minimax, information-theoretic), verified by brute-force enumeration of all 237,336 feasible solutions (33-bus, k=5). The validation loop connects QUBO-optimal solutions to operational cascade detection metrics, providing a reusable methodology for any engineering QUBO application. Full reproducible pipeline available as open source.

2. **Formulation gap discovery: stress-information duality in radial networks.** The dispersion-dominant QUBO produces placements worse than 99% of random alternatives (29% coverage, 20% hub accuracy). The root cause is structural: in the radial (tree) topology, bus-level stress centrality is a sufficient proxy for information centrality — junction nodes are simultaneously stress attractors and informationally privileged under the tree Markov property. Any formulation preserving stress-weight dominance produces good placements (coverage, greedy, low-alpha, MI-QUBO all achieve 100% hub accuracy); only $\alpha \geq 1.0$ fails, because the dispersion term drowns the stress signal. An alpha sensitivity sweep maps a sharp phase transition at $\alpha \approx 0.8$. MI anti-correlates with distance (proxy r = -0.37; confirmed at r = -0.45 under a DC cascade ensemble). A comparative experiment on the IEEE 57-bus meshed system (22 cycles) confirms the topology-dependence prediction: no phase transition occurs (100% hub accuracy at all alphas), and the MI-distance anti-correlation weakens (r = -0.29 vs -0.37).

3. **QAOA simulator validation with landscape analysis.** On a 20-qubit reduced subproblem, QAOA recovers the exact optimum at all depths (p=1-4) with multi-start COBYLA

(9/10 success rate). Parameter landscape analysis (40x40 grid) confirms multiple local minima; non-monotonic single-seed performance is an optimizer artifact. Classical baselines (exact, greedy 97.2%, SA 100%, MILP 100%) on the full 33-bus instance establish rigorous comparison.

4. **Corrected Hadamard test assessment and fault-tolerant roadmap.** We provide an honest reassessment of quantum commutator estimation with a worked numerical example quantifying the Frobenius norm gap (single-state measurement captures 2.8% of trace), and present qubit/depth/error-rate requirements for real-time monitoring at scale.

II. Problem Formulation

A. The Sensor Placement Problem

Given: - Power network graph G = (V, E) with n = |V| buses - Bus-level stress functional $\Lambda_G^{(i)}$(computed classically from simulation/historical data) - Sensor budget k « n - Electrical distance matrix D (impedance-weighted shortest paths)

Find: binary vector $x \in \{0,1\}^n$ with $\sum_i x_i = k$ that maximizes a coverage objective.

Test network: IEEE 33-bus radial distribution system [6]. This network is a tree graph (33 buses, 32 branches, no loops). The radial topology has qualitative implications for sensor placement: stress propagates along single paths from the substation, so sensors at opposite feeder endpoints observe correlated trunk-line stress despite being maximally separated. This topology-dependent behavior — established in the TPWRS paper [1] — predicts that dispersion-based objectives may be mismatched to radial networks (validated in Section VII-D). A comparative experiment on the IEEE 57-bus meshed transmission system (78 edges, 22 cycles) confirms that meshed networks do not exhibit the same formulation gap (Section VII-E).

We formulate and test four candidate objective functions:

Objective 1: Weighted Maximum Coverage

```
max  sum_i  w_i * z_i
s.t. z_i <= sum_{j: d(i,j) <= r} x_j   (bus i
↪  covered if sensor within radius r)
     sum_i x_i = k                         (budget
      ↪  constraint)
     x_i, z_i in {0,1}
```

Where weights w_i = topological centrality (betweenness or degree) of bus i.

- Rationale: covers high-connectivity hubs where stress concentrates (per TPWRS cascade results showing buses 45, 26, 13, 77 as consistent stress attractors)

Objective 2: Stress-Weighted Coverage

```
max  sum_i  Lambda_G^(i)_max * z_i
```

Where $(\Lambda_G^{(i)})_{\max}$is the maximum observed bus-level stress across a library of simulated contingencies.

- Rationale: directly targets buses with highest diagnostic value - Limitation: requires pre-computed contingency library (classical preprocessing)

Objective 3: Minimax Observability Gap

```
min  max_i  (1 - z_i) * w_i
```

Minimize the worst-case unobserved stress.

- Rationale: defensive strategy ensuring no high-value bus is unmonitored - Maps naturally to a different QUBO structure (useful for comparing QAOA performance across problem Hamiltonians)

Objective 4: Information-Theoretic (Mutual Information)

The information-theoretic objective maximizes the mutual information [16] between sensor observations and the latent stress field, penalizing redundancy between sensors:

```
max  sum_i H_i x_i  -  beta * sum_{i<j} MI(i,j) x_i
↪  x_j
s.t. sum_i x_i = k
```

where H_i = (1/2) log(2*pi*e * Sigma_ii) is the marginal entropy at bus i, and $\mathrm{MI}(i,j) = -\frac{1}{2}\log\bigl(1-\rho_{ij}^2\bigr)$ is the Gaussian mutual information between buses i and j, with rho_ij the correlation coefficient from the covariance matrix Sigma.

The covariance Sigma is estimated from the graph Laplacian pseudoinverse L^+scaled by stress weights: $\Sigma = W^{1/2}\,L^+\,W^{1/2}$, where $W = \mathrm{diag}(w)$. $L^+$is the covariance of a random walk on the graph, not of empirical $\Lambda_G$values. The structural justification is the tree Markov property [17]: buses separated by a junction are conditionally independent given the junction's observation. A DC cascade ensemble (2,000 scenarios) validates the proxy's qualitative predictions (Section VII-D): the MI-distance anti-correlation is stronger under cascade stress (r = -0.45) than the proxy predicts (r = -0.37), and placement rankings are preserved.

The total information captured by sensor set S is measured by the conditional entropy reduction (mutual information between S and the full field):

```
I(S; Lambda_G) = (1/2) log [ det(Sigma_{full}) /
↪  det(Sigma_{full|S}) ]
```

where $\Sigma_{\text{full}|S} = \Sigma_{NN} - \Sigma_{NS}\Sigma_{SS}^{-1}\Sigma_{SN}$ is the Schur complement (N = non-sensor buses, S = sensor buses). This quantity is submodular in S [15], meaning the greedy heuristic on the MI objective directly achieves at least (1 - 1/e) ≈ 63.2% of the optimal MI. (Our greedy heuristic optimizes the dispersion objective, not MI directly, so the Krause guarantee does not formally apply — but the 98.5% ratio observed in Section VII-D is consistent with submodularity.)

- Rationale: directly maximizes information gain, penalizes redundancy instead of rewarding distance. In radial networks, MI anti-correlates with electrical distance (proxy r = -0.37, confirmed at r = -0.45 under DC cascade ensemble), so this objective produces qualitatively different placements than dispersion. - Key advantage: the QUBO form is identical to dispersion (quadratic in x_i) but with MI replacing distance. No auxiliary variables needed — stays at n qubits.

B. QUBO Reformulation

Each objective above is a constrained binary optimization problem with the cardinality constraint sum_i x_i = k. To formulate as a Quadratic Unconstrained Binary Optimization (QUBO) problem, we absorb the constraint into the objective via a penalty term:

```
H_QUBO = -sum_i sum_j Q_{ij} x_i x_j + lambda *
↪  (sum_i x_i - k)^2
```

The penalty term expands as:

```
lambda * (sum_i x_i - k)^2 = lambda * [sum_i x_i +
↪  2*sum_{i<j} x_i x_j - 2k*sum_i x_i + k^2]
```

where we use the binary identity $x_i^2 = x_i$. For the stress-weighted dispersion objective (Objectives 2-3), the Q matrix has entries:

```
Q_{ii} = -w_i - lambda(1 - 2k)     (diagonal:
↪  stress weight + penalty)
Q_{ij} = -alpha * d_{ij} - 2*lambda  (off-diagonal:
↪  dispersion + penalty)
```

For the MI-QUBO (Objective 4), the off-diagonal dispersion term alpha * d_ij is replaced by the MI penalty beta * MI(i,j), but the QUBO structure is identical — both are quadratic in x_i with no auxiliary variables, requiring exactly n qubits.

For the coverage objective (Objective 1), the coverage indicator z_i requires auxiliary binary variables, doubling the problem to 2n qubits. This structural difference has practical implications for quantum hardware (Section III).

Q matrix properties (33-bus, stress-dispersion, k=5, alpha=1.0, lambda=10): The Q matrix is 33x33 symmetric with all 528 off-diagonal entries nonzero (dense, due to all-pairs distance terms). Diagonal entries range from -91.0 to -90.0, encoding stress weights and penalty contributions. Off-diagonal entries range from 19.0 to 19.99, encoding dispersion distances and penalty coupling. The Q matrix and its Ising representation are archived in the supplementary repository for reproducibility.

Penalty coefficient selection. The penalty lambda = 10.0 was validated by brute-force enumeration of all 237,336 feasible solutions (k=5 from n=33): the QUBO minimizer exactly matches the original constrained objective maximizer. Insufficient penalty (lambda < 5) produces infeasible QUBO minima; excessive penalty (lambda > 50) flattens the objective landscape, degrading optimizer convergence. At lambda = 10.0, QAOA feasibility rates range from 0.5% (p=1) to 7.5% (p=4) on the 20-bus subproblem, indicating that post-selection for the cardinality constraint remains essential. These low feasibility rates are a known limitation of penalty-based constraint encoding; the production pathway is constrained QAOA with Hamming-weight-preserving mixers (see Section III-D).

C. Ising Hamiltonian Mapping

For execution on gate-model quantum hardware (QAOA) or quantum annealers (D-Wave), the QUBO is mapped to an Ising Hamiltonian via the substitution x_i = (1 - Z_i)/2, where Z_i is the Pauli-Z operator on qubit i:

```
H_Ising = sum_i h_i Z_i + sum_{i<j} J_{ij} Z_i Z_j
↪  + const
```

The local fields h_i and couplings J_ij are linear functions of the Q matrix entries:

```
h_i = -(1/2) * Q_{ii} - (1/4) * sum_{j != i} Q_{ij}
J_{ij} = (1/4) * Q_{ij}
```

The constant offset does not affect the optimization but is tracked for energy comparison with classical solutions.

The graph structure of J_ij determines hardware requirements. For the stress-dispersion QUBO, all 528 pairwise couplings are nonzero (complete graph), because the dispersion term involves all-pairs electrical distances. This dense connectivity has implications for hardware embedding: D-Wave's Pegasus topology requires chain embedding for all-to-all coupling, introducing overhead proportional to chain length. Trapped-ion hardware (IonQ Aria/Forte [12]) provides native all-to-all connectivity via shared motional modes, requiring no embedding overhead — a structural advantage for dense QUBOs of this type. The MI-QUBO has comparable density (all pairs with nonzero MI), while the coverage QUBO is sparser (only pairs within the detection radius interact), potentially favoring superconducting architectures.

D. Classical Baselines

We implement four classical solvers for the 33-bus stress-dispersion QUBO (k=5, alpha=1.0) to establish rigorous baselines for QAOA comparison.

1. **Exact solution** (brute force): 237,336 feasible solutions enumerated in 0.83s. Optimal sensors: buses 16, 17, 21, 24, 32. Objective: 9.180 (stress: 1.864, dispersion: 7.316). 2. **Greedy heuristic** (iterative marginal gain): 97.2% approximation ratio, <0.01s. Selects 5, 16, 17, 21, 32 — notably includes bus 5 (highest centrality hub) instead of bus 24. This difference proves critical in the validation loop (Sec. VII-D). 3. **Simulated annealing** (10^5 iterations, T_0=10, swap moves): 100% approximation ratio (finds exact optimum), 1.78s. 4. **MILP via PuLP/CBC** (open-source, McCormick linearization of quadratic terms): 100% approximation ratio, 88.5s. Significantly slower than brute force at this scale due to linearization overhead.

Method	Objective	Approx. Ratio	Time (s)	Sensors
Exact (brute force)	9.180	1.000	0.83	16, 17, 21, 24, 32
Greedy	8.924	0.972	<0.01	5, 16, 17, 21, 32
Simulated annealing	9.180	1.000	1.78	16, 17, 21, 24, 32
MILP (CBC)	9.180	1.000	88.5	16, 17, 21, 24, 32

III. QAOA Circuit Design and Compilation

A. QAOA Overview

The Quantum Approximate Optimization Algorithm (QAOA) [3] constructs a parameterized quantum circuit that alternates between a cost unitary U_C(gamma) = exp(-i *gamma* H_C) encoding the Ising Hamiltonian from Section II-C, and a mixer unitary U_M(beta) = exp(-i *beta* H_M) with H_M = sum_i X_i that drives transitions between computational basis states. Starting from the equal superposition state $|+\rangle^{\otimes n}$, the circuit applies p layers of (U_C, U_M) pairs, producing the variational state:

```
|gamma, beta> = U_M(beta_p) U_C(gamma_p) ...
↪  U_M(beta_1) U_C(gamma_1) |+>^n
```

The 2p parameters (gamma_1,...,gamma_p, beta_1,...,beta_p) are optimized by a classical outer loop (COBYLA in this work) to minimize the expected Ising energy <gamma,beta|H_C|gamma,beta>. The depth parameter p controls the circuit's expressibility: at p = 1, QAOA produces a restricted family of states; as p grows, the circuit can approximate the ground state arbitrarily well [11]. In practice, increasing p improves theoretical expressibility but introduces more local minima in the parameter landscape (Section VII-A).

B. Circuit Compilation

The cost unitary U_C(gamma) decomposes into pairwise ZZ interactions and single-qubit Z rotations. Each ZZ_ij interaction exp(-i *gamma* J_ij * Z_i Z_j) is implemented as a CNOT-Rz(2*gamma*J_ij)-CNOT sequence (3 gates per coupling).

Single-qubit terms exp(-i *gamma* h_i * Z_i) require one Rz gate each. The mixer U_M(beta) requires one Rx(2*beta) gate per qubit.

For the 20-bus reduced subproblem (190 nonzero J_ij couplings), the compiled circuit resources per QAOA layer are:

QAOA Depth	CNOT Gates	Rz Gates	Rx Gates	Circuit Depth	Total Gates
p=1	380	210	20	115	630
p=2	760	420	40	177	1,240
p=3	1,140	630	60	239	1,850
p=4	1,520	840	80	301	2,460

For the full 33-bus problem, the dense QUBO produces 528 ZZ terms per layer (1,056 CNOTs per layer, 2,112 at p=2), placing it at the coherence boundary of current trapped-ion hardware.

Circuit depth scales linearly with p: depth = 62p + 53 (measured from compiled Qiskit circuits). On trapped-ion hardware with native all-to-all connectivity, no SWAP routing is required, and some MS (Molmer-Sorensen) gates can execute in parallel depending on the addressing scheme, reducing effective depth by approximately 2-4x. For p=2 on the full 33-bus problem, the estimated native depth after IonQ transpilation is 200-350.

C. Parameter Optimization

The classical outer loop uses COBYLA (Constrained Optimization BY Linear Approximation) with maxiter=200 and initial step size rhobeg=0.5. Across all depths, the optimizer converged in 30-66 function evaluations: p=1 required 30 evaluations (43s wall-clock on the Aer statevector simulator), while p=4 required 66 evaluations (90s). Each function evaluation involves simulating the full QAOA circuit and sampling 16,384 shots.

For hardware execution, we recommend a parameter transfer strategy: optimize parameters on the noiseless simulator, then execute a single circuit evaluation on hardware with the optimized parameters. This minimizes the shot budget (one circuit vs. hundreds of optimization iterations) and avoids noisy gradient estimation. Barren plateau effects were not observed at p <= 4 for 20 qubits, but 33-qubit hardware runs should monitor gradient magnitudes carefully at p >= 3.

D. Constraint Handling and Noise Considerations

Feasibility rates and the penalty-based approach. All QAOA results in this paper use penalty-based constraint encoding: the cardinality constraint sum_i x_i = k is absorbed into the cost function with penalty coefficient lambda=10 (Section II-B). This approach is simple to implement but produces low feasibility rates — only 0.5% of shots at p=1 and 7.5% at p=4 satisfy the constraint (Table VII-1). The remaining 92.5-99.5% of shots are discarded via post-selection, representing wasted quantum resources.

Constrained QAOA with Hamming-weight-preserving mixers. The production pathway for cardinality-constrained QUBO problems is constrained QAOA [24], which replaces the standard X-mixer with a Hamming-weight-preserving mixer that restricts evolution to the feasible subspace x : sum_i x_i = k. The ring mixer H_ring = sum_i (X_i X_i+1 + Y_i Y_i+1) preserves Hamming weight by construction: each term swaps a pair of adjacent qubits (|01> <-> |10>), leaving the total number of 1s unchanged. When initialized in any feasible state (e.g., the Dicke state $|D_n^k\rangle$, an equal superposition of all weight-k bitstrings), constrained QAOA produces 100% feasible shots by construction — eliminating the post-selection overhead entirely.

The trade-off is circuit complexity: the ring mixer requires O(n) two-qubit gates per layer (versus O(n) single-qubit Rx gates for the standard mixer), and the Dicke state initialization requires O(nk) gates [25]. For our 20-bus, k=3 problem, this adds 60 two-qubit gates for initialization and 40 per mixer layer — modest compared to the 380 CNOT gates per cost layer. The feasibility guarantee makes this the preferred approach for hardware execution, where every shot is expensive.

We use penalty-based encoding in this work because (a) it matches the standard QAOA formulation used in prior benchmarks [19], enabling direct comparison, and (b) the noiseless simulator setting makes post-selection computationally free. For hardware deployment, constrained QAOA is not optional — it is required. At 0.5% feasibility (p=1), a 16,384-shot circuit produces only 80 valid bitstrings. On hardware with $0.01-$0.10 per shot, the post-selection overhead alone makes penalty-based QAOA economically unviable. Constrained QAOA with Hamming-weight-preserving mixers [24] eliminates this overhead entirely and should be considered the default for any cardinality-constrained QUBO on quantum hardware.

Noise mitigation for hardware execution. All QAOA results use noiseless statevector simulation (Qiskit Aer). For planned hardware execution, two additional noise mitigation strategies are applicable:

1. **Measurement error mitigation.** Confusion matrix calibration corrects for readout errors, which are typically 1-3% per qubit on trapped-ion hardware. For 20 qubits, the expected readout fidelity per bitstring is 0.98^{20} 67%.
2. **Zero-noise extrapolation (ZNE).** Gate noise can be partially mitigated by running circuits at multiple noise levels and extrapolating to zero noise. This is applicable if the hardware supports noise amplification via gate folding.

IV. Head-to-Head Comparison and Scaling Analysis

A. Simulator and Classical Baselines

Table IV-1 consolidates all solver results on a common scale. Classical methods solve the full 33-bus problem (k=5); QAOA is validated on the 20-bus reduced subproblem (k=3).

Method	System	Approx. Ratio	Time (s)	Qubits	Platform
Exact (brute force)	33-bus	1.000	0.83	—	Classical
Greedy heuristic	33-bus	0.972	<0.01	—	Classical
Simulated annealing	33-bus	1.000	1.78	—	Classical
MILP (CBC)	33-bus	1.000	88.5	—	Classical
QAOA p=1 (sim)	20-bus	1.000	43	20	Aer statevector
QAOA p=2 (sim)	20-bus	0.952	45	20	Aer statevector
QAOA p=3 (sim)	20-bus	0.952	76	20	Aer statevector
QAOA p=4 (sim)	20-bus	1.000	90	20	Aer statevector

The full 33-bus QUBO requires 33 qubits with 528 ZZ couplings per layer. Statevector simulation requires 2^{33} = 8.6

billion amplitudes (128 GB), exceeding local compute. Matrix product state (MPS) simulation was attempted but proved prohibitively slow due to the dense all-to-all coupling structure. The 20-bus reduced subproblem (top-20 buses by stress weight, 190 couplings) validates QAOA circuit correctness and convergence behavior. Full 33-bus execution requires quantum hardware.

B. Hardware Execution (Pending)

IonQ trapped-ion hardware access is under evaluation (direct API, Amazon Braket, or Azure Quantum). The QAOA circuit for 33-bus at p=2 requires 2,112 CNOT gates (1,056 ZZ terms), which is at the coherence boundary of current IonQ Forte specifications [18]. Parameter transfer (optimize on simulator, single execution on hardware) is the recommended strategy to minimize shot budget.

D-Wave quantum annealing is deferred to a subsequent study. The 33-variable dense QUBO embeds on D-Wave Advantage (5000+ qubits, Pegasus graph) but requires significant chain overhead due to all-to-all connectivity. This comparison remains planned future work.

C. Scaling Analysis

At 33 buses, classical methods win decisively — brute force solves the problem in under one second. The relevant question is where the classical-quantum crossover occurs.

Scaling crossover table. Table IV-2 shows concrete scaling for exact enumeration, simulated annealing, and QAOA (p=2, dense QUBO) as a function of network size. Brute force throughput is calibrated from the measured 237,336 evaluations in 0.83s (286,000 evaluations/sec). SA time is extrapolated linearly from the measured 1.78s at n=33. QAOA gate counts reflect the dense stress-dispersion QUBO (n(n-1)/2 ZZ couplings).

n (buses)	k	C(n,k)	Brute Force	SA (est.)	QAOA p=2 Gates
33	5	237,336	0.8 s	1.8 s	3,234
57	5	4.2 M	14.6 s	3.1 s	9,690
69	10	340 B	13.8 days	3.7 s	14,214
118	10	97.5 T	11 years	6.4 s	41,654
300	10	1.4×10^{18}	$1.6 * 10^5$ yr	16 s	269,700
1,000	15	6.9×10^{32}	intractable	54 s	2,999,000

The crossover is clear: exact enumeration becomes intractable at n 70-100 (depending on k), while SA remains fast but provides no optimality guarantees. QAOA circuit size scales as $\mathcal{O}(p\,n^2)$ for the dense QUBO, reaching 270,000 two-qubit gates at n=300. At current trapped-ion gate fidelities (99.5%), a 270,000-gate circuit has negligible success probability without error correction. However, two factors shift the crossover favorably:

- *Sparse QUBOs.* The coverage QUBO has O(n * d_avg) couplings with d_avg 3-6, giving O(p * n) depth — polynomial, not quadratic. At n=300, this reduces the gate count by 50x. - *Error correction.* Projected logical gate fidelities of 99.99%+ (Tier 2 hardware, Section VI) make circuits of 10^4–10^5 logical gates feasible.

We do not claim quantum advantage at any tested scale. Instead, we validate functional QAOA execution on noiseless simulators for a power systems optimization problem, establish the formulation challenges (objective alignment, QUBO density), and identify the hardware specifications required for advantage at operationally relevant scales (n > 300, where exact methods become intractable).

V. Commutator Estimation: Reassessment

Sections V and VI address the quantum computational pathway for $\Lambda_G$monitoring itself (as opposed to sensor placement optimization). These sections are independent of the formulation gap analysis in Sections VII-VIII and may be read separately.

A. Correcting Prior Claims

An earlier version of [1] claimed that the Hadamard test enables "transformational" quantum speedup for $\Lambda_G$computation. This section provides a corrected analysis, identifying three specific errors in the original claim.

First, the Hadamard test estimates <psi|[A,B]|psi> for a single state |psi>, but Λ_G= ||[J,Jdot]||_F requires the full Frobenius matrix norm — a trace over all basis states. These are fundamentally different quantities: a single-state expectation value versus a complete-basis trace.

Second, the original analysis omitted the cost of block-encoding a time-varying sparse Jacobian at each timestep. Block encoding has cost $\mathcal{O}(d\,\mathrm{polylog}(n))$ per encoding, where d is the matrix sparsity (typically 3-6 for power grids). Re-encoding at the 10-60 Hz monitoring refresh rate dominates any downstream speedup.

Third, power grid Jacobians have $\mathcal{O}(n\,d)$ nonzero entries with d 3-6. Sparse classical commutator computation is $\mathcal{O}(n\,d^2)$, which is effectively linear in n. The quantum advantage window is therefore much narrower than the naive comparison of $\mathcal{O}(n^3)$ dense classical versus $\mathcal{O}(\mathrm{polylog}(n))$ quantum suggests.

B. What the Hadamard Test Actually Provides

Given unitaries $U_A = e^{iAt}$ and $U_B = e^{iBt}$, the Hadamard test estimates:

```
Re(<psi| U_A^dag U_B^dag U_A U_B |psi>) = 1 - t^2 *
↪  <psi|[A,B]^dag [A,B]|psi> + O(t^4)
```

This yields $\langle\psi|[A,B]^\dagger[A,B]|\psi\rangle$ for a single state $|\psi\rangle$. To recover the full Frobenius norm $\|[A,B]\|_F^2 = \mathrm{Tr}\big([A,B]^\dagger[A,B]\big)$, which classically requires explicit commutator evaluation. complete basis (O(n) Hadamard tests, no speedup), or use randomized trace estimation (Section V-C). The following worked example quantifies the gap between single-state measurement and the full trace.

Worked example (2-bus system, 4x4 Jacobian). Consider a minimal 2-bus system with Jacobian J and its time derivative Jdot, each 4x4 (2 buses x 2 state variables). Let:

```
J = [[2, -1, 0, 0],      Jdot = [[0.1, 0, 0, 0],
     [-1, 3, -1, 0],             [0, 0.2, 0, 0],
     [0, -1, 2, -1],             [0, 0, 0.3, 0],
     [0, 0, -1, 1]]              [0, 0, 0, -0.1]]
```

J is symmetric tridiagonal (a simplified Jacobian structure for a 2-bus radial system). Jdot is diagonal (non-uniform load ramp rates at each state variable).

Computing C = [J, Jdot] = J*Jdot - Jdot*J explicitly (since Jdot is diagonal, (J*Jdot)_ij = J_ij Jdot_jj and (Jdot*J)_ij = *Jdot_ii* J_ij, giving C_ij = J_ij * (Jdot_jj - Jdot_ii)):

```
C = [[ 0,    -0.1,   0,     0   ],
     [ 0.1,   0,    -0.1,   0   ],
     [ 0,     0.1,   0,     0.4 ],
     [ 0,     0,    -0.4,   0   ]]
```

C is antisymmetric (as expected for the commutator of symmetric matrices). The true $\Lambda_G$is:

```
||C||_F^2 = sum_{ij} |C_ij|^2 = 2(0.01 + 0.01 +
↪  0.16) = 0.36
Lambda_G = ||C||_F = 0.60
```

The stress is concentrated in the (3,4)/(4,3) entries (0.4 and -0.4), corresponding to the bus pair with the largest Jdot contrast (0.3 vs -0.1). Now consider the Hadamard test with each basis state:

```
<e_1|C^dag C|e_1> = 0.0100   (2.8% of trace)
<e_2|C^dag C|e_2> = 0.0200   (5.6% of trace)
<e_3|C^dag C|e_3> = 0.1700  (47.2% of trace)
<e_4|C^dag C|e_4> = 0.1600  (44.4% of trace)
Sum = 0.3600 = ||C||_F^2      OK
```

A single Hadamard test with |e_1> yields sqrt(0.01) = 0.10, while the true Λ_G= 0.60 — an underestimate by 6x. The stress is concentrated in the (3,4) subspace, but the Hadamard test with |e_1> sees only the (1,*) column. Even within this n=4 system, a single state captures only 2.8% of the trace. The distribution is highly non-uniform: |e_3> and |e_4> together capture 91.6%, while |e_1> and |e_2> capture only 8.4%. In larger systems, the mismatch worsens: for n=33, a single state captures 3% on average; for n=10^5 (continental), 0.001%. This is the fundamental measurement gap that randomized trace estimation partially bridges.

C. Randomized Trace Estimation

The Hutchinson estimator [5] provides a partial solution by approximating the trace of an implicit positive semidefinite matrix M as:

```
Tr(M) ~= (1/m) * sum_{j=1}^{m} z_j^T M z_j
```

where z_j are random vectors with $\mathbb{E}[zz^\top] = I$ (Rademacher or Gaussian). Applied to $M = C^\dagger C$ with $C = [A, B]$: $M = C^\dagger C$ with $C = [A, B]$:

```
||[A,B]||_F^2 = Tr(C^dag C) ~= (1/m) *
↪  sum_{j=1}^{m} <z_j| C^dag C |z_j>
```

In the quantum setting, |z_j> are Haar-random quantum states. Since E_Haar[|psi><psi|] = I/n (the maximally mixed state scaled by 1/n), we have:

```
E_{|psi> ~ Haar} [<psi|C^dag C|psi>] = Tr(C^dag C)
↪  / n
```

Therefore the quantum Hutchinson estimator requires the factor of n:

```
||[A,B]||_F^2 ~= (n/m) * sum_{j=1}^{m} <psi_j|C^dag
↪  C|psi_j>
```

The factor of n arises from the normalization of Haar-random quantum states (which have Each expectation $\langle\psi_j|C^\dagger C|\psi_j\rangle$ is estimated via the Hadamard test. Total cost: is estimated via Hadamard test. Total cost:

```
O(m * shots_per_state * encoding_cost)
```

Where $m = \mathcal{O}(1/\epsilon^2)$ for ϵ-relative error [**?**].

The complexity comparison between classical and quantum approaches is:

Step	Classical (sparse)	Quantum (Hadamard + trace est.)
Compute $[J, \dot{J}]$	$\mathcal{O}(n\,d^2)$	Not computed explicitly
Frobenius norm	$\mathcal{O}(n\,d)$	$\mathcal{O}(m\,C_{\text{encode}} \cdot \text{shots})$
Total per timestep	$\mathcal{O}(n\,d^2)$	$\mathcal{O}(m\,d\,\text{polylog}(n) \cdot \text{shots})$
Re-encoding at 10 Hz	—	O(10 * C_encode) amortization

For d = 5 (typical grid sparsity), classical computation is O(25n) per timestep. Quantum requires O(m d polylog(n) * shots) with m 100 for 1% accuracy and shots 1000, giving approximately 500,000 * $\log(n)^c$ operations. The crossover occurs at n 10^5 to 10^6 depending on constant factors — continental scale, not regional. The Hadamard approach becomes relevant only for grids larger than 50,000 buses, and only after fault-tolerant quantum computers with low-overhead block encoding exist. For regional grids (< 10,000 buses), classical sparse computation wins decisively.

D. Bus-Pair Monitoring: A Near-Term Narrowing

A more modest formulation avoids the full trace estimation by monitoring specific diagonal elements of $\left[J, \dot{J}\right]^\dagger \left[J, \dot{J}\right]$ corresponding to high-risk buses identified by QAOA sensor placement:

```
Lambda_G^(i) ~= sqrt( <e_i| [J,Jdot]^dag [J,Jdot]
↪  |e_i> )
```

This requires ONE Hadamard test per monitored bus per timestep. If QAOA identifies k = 10 critical buses, that's 10 Hadamard circuits per timestep — potentially viable on near-term hardware IF encoding can be made efficient.

Whether the power flow Jacobian's sparsity can be exploited for efficient block encoding remains an open question. Sparse-access block encoding [7] achieves O(d *polylog(n)); if d = 5, this is O(5* polylog(n)), potentially fast enough for bus-pair monitoring at moderate n.

This connects the two quantum components into a pipeline: QAOA (Sections III-IV) identifies where to monitor; Hadamard tests (if encoding can be made efficient) measure the stress at those locations.

E. Quantum-Amenable vs. Classical Pipeline Components

For completeness, we identify which components of the $\Lambda_G$monitoring pipeline benefit from quantum computation and which do not:

- **Classical only:** AC power flow solution (nonlinear, iterative — no known quantum speedup), Jacobian extraction (byproduct of power flow), shock metric S computation (ratio of two Frobenius norms), emergency intervention targeting (minimum voltage identification). - **Quantum-amenable:** Sensor placement optimization (QAOA, Sections III-IV — near-term for n > 300 buses) and commutator norm estimation for continuous monitoring (Hadamard test with trace estimation, this section — fault-tolerant era only, n > 50,000 buses).

VI. Fault-Tolerant Roadmap

We identify three capability tiers for quantum-enhanced power grid monitoring, each with distinct hardware requirements:

Tier	Capability	Logical Qubits	Gate Fidelity	Timeline (est.)
1 (NISQ, now)	QAOA sensor placement (33-bus)	33	99.5%	Available
2 (Early FT)	QAOA sensor placement (1000-bus) + bus-pair monitoring	1000	99.99%	2028-2032
3 (Full FT)	Real-time commutator norm estimation (continental)	10^4	99.999%	2032+

A. Resource Estimates

Tier 2 (the first operationally useful tier): QAOA for 1000-bus sensor placement: - 1000 logical qubits - O(p *n* d) = O(3 *1000* 5) = 15,000 entangling gates per layer - At 99.99% logical gate fidelity: circuit success probability after p=2 layers = $0.9999^{30000} \approx 0.05 \rightarrow$ need error correction OR error mitigation - Surface code overhead: 1000-3000 physical qubits per logical qubit $\rightarrow 10^6$ to $3*10^6$ physical qubits - This is beyond current hardware but within projected 2028-2032 roadmaps

Tier 3 (continental-scale monitoring): Block-encoded Jacobian for 100,000-bus grid: - Block encoding: $\mathcal{O}(d\,\mathrm{polylog}(n))$ T-gates per encoding; for $d \approx 5$ and $n \sim 10^5$ a rough order-of-magnitude is 10^2–10^3 T-gates per encoding. - Hadamard test: O(encoding) per measurement - Randomized trace estimation: m = 100 random states - Total per timestep: 100 * 1000 = 10^5 T-gates - At projected MHz logical clock rates: 0.1 s per timestep $\rightarrow$ 10 Hz feasible - Qubit count: O(polylog(100000)) $\approx$ O(50-100) logical qubits for the Hadamard circuit itself, but encoding ancillas may push to O(1000)

These estimates are speculative: T-gate counts depend on encoding scheme specifics that have not been optimized for power flow Jacobians.

B. Hardware Roadmap Alignment

Provider	Announced Roadmap	Relevant Tier
IonQ	Forte (2024): 36 algorithmic qubits	Tier 1
IonQ	2028 target: 1000 logical qubits	Tier 2
IBM	Flamingo (2025): 1000+ qubits	Tier 1-2 (with error mitigation)
IBM	Blue Jay (2033): 100,000 qubits	Tier 3
Google	Willow (2024): 105 qubits, below-threshold EC	Tier 1 $\rightarrow$ 2 bridge
D-Wave	Advantage2 (2024): 1200+ qubits	Tier 1 (annealing only)

Hardware roadmap data should be updated at time of submission to reflect current manufacturer announcements.

C. The Geometric Algebra Connection

The commutator $\left[J, \dot{J}\right] = 2\left(J \wedge \dot{J}\right)$ is the bivector component of the geometric product in the Clifford algebra over $\mathbb{R}^{n\times n}$ [**?**]. Quantum computers natively represent spinor transformations — the exponential map of bivectors. This suggests a deeper connection: the Jacobian frame evolution that $\Lambda_G$measures may be representable as a quantum spin system whose dynamics can be directly simulated rather than classically computed and then quantum-encoded. Exploring this "physics-native" encoding — where the grid's non-commutative dynamics are mapped to a quantum Hamiltonian without intermediate classical Jacobian computation — is a longer-term research direction that could eliminate the encoding bottleneck entirely. This connection is speculative but points toward a genuine research program linking geometric algebra, power system dynamics, and quantum simulation.

VII. Results

A. QAOA Simulator Validation (20-Bus Reduced Problem)

The full 33-bus problem requires 2^{33} statevector amplitudes (128 GB), exceeding noiseless simulator capacity. We validate QAOA on a 20-bus subproblem comprising the top-20 buses by stress weight, with k=3 sensors. The reduced problem preserves the electrical distance structure and stress-weight distribution of the full network. The exact optimum on this subproblem is sensors 17, 21, 32 with objective 3.980.

Metric	p=1	p=2	p=3	p=4
Approx. ratio	**1.000**	0.952	0.952	**1.000**
Original objective	3.980	3.789	3.789	3.980
Sensors (bus IDs)	17,21,32	17,24,32	17,24,32	17,21,32
Feasibility rate	0.46%	4.03%	4.44%	7.50%
Feasible bitstrings	71	410	449	390
Circuit depth	115	177	239	301
CNOT gates	380	760	1,140	1,520
Optimizer evals (COBYLA)	30	32	53	66
Wall-clock time	43.2 s	45.3 s	76.2 s	90.2 s
Shots (final eval)	16,384	16,384	16,384	16,384

Key observations:

1. **Non-monotonic performance is an optimizer artifact, not an expressibility limit.** With a single COBYLA initialization, p=1 and p=4 recover the exact optimum (ratio 1.000) while p=2 and p=3 converge to a suboptimal placement (ratio 0.952). To disambiguate optimizer trapping from circuit expressibility, we ran 10 COBYLA restarts with random initializations at each depth:

Depth	Restarts finding optimum	Mean ratio	Worst ratio
p=2	9/10	0.984	0.843
p=3	9/10	0.965	0.650

Both depths can express the optimal solution; the original single-seed failures were initialization-dependent. This confirms that COBYLA with random restarts is sufficient for this problem scale, while highlighting that production QAOA deployments should use multi-start optimization.

2. **Parameter landscape analysis (p=1).** The full (gamma, beta) cost landscape on a 40x40 grid (Fig. 1 — heatmap with COBYLA convergence point, data in `results/landscape_p1.npz`) reveals multiple local minima. The grid-search global minimum lies at gamma=0.322 (0.10pi), beta=0.967 (0.31pi), energy=70.1 — while the original COBYLA converged to gamma=1.431 (0.46pi), beta=0.315 (0.10pi), energy=777.6. Despite converging to a higher-energy point, COBYLA's solution still yields the correct optimal bitstring upon post-selection, because the feasible subspace contains the optimum at both parameter settings. This landscape structure — where post-selection rescues suboptimal parameters — may not persist at larger scales or with hardware noise.

Missing figure: quantum_sensor_placement/figures/fig2_landscape_p1.pdf

Fig. 1. QAOA p=1 parameter landscape (20-bus, k=3). Heatmap shows expected Ising energy over a 40x40 (gamma/pi, beta/pi) grid. Red star: grid-search global minimum (E=70). White circle: COBYLA convergence point (E=778). The multiple basins confirm that single-seed COBYLA can miss the global minimum, motivating multi-start optimization.

3. **Feasibility rate increases with depth but remains low.** Post-selection feasibility rises from 0.46% (p=1) to

7.50% (p=4), indicating that deeper circuits better concentrate amplitude on the k=3 feasibility subspace. At 16,384 shots, even p=1 produces 71 feasible bitstrings — sufficient for reliable solution extraction on a noiseless simulator. On hardware, where each shot is expensive, these rates imply a 13-200x shot overhead. Constrained QAOA with Hamming-weight-preserving mixers [24] eliminates this overhead entirely by restricting evolution to the feasible subspace (Section III-D).

4. **All depths produce feasible solutions.** No QAOA depth failed to find at least one feasible solution, validating the penalty-based constraint encoding with coefficient lambda=10.

5. **Circuit resource scaling.** CNOT count scales as 380p (from 190 ZZ couplings per layer), and circuit depth scales as 62p + 53. On trapped-ion hardware with all-to-all connectivity, these circuits are executable without routing overhead.

B. Full 33-Bus Problem: Classical Baseline Comparison

For the full IEEE 33-bus problem (k=5), we compare four classical methods. The brute-force exact solution enumerates all C(33,5) = 237,336 feasible placements.

Method	Objective	Approx. Ratio	Time (s)	Sensors
Exact (brute force)	9.1796	1.000	0.83	16, 17, 21, 24, 32
Simulated annealing	9.1796	1.000	1.78	16, 17, 21, 24, 32
MILP (CBC solver)	9.1796	1.000	88.48	16, 17, 21, 24, 32
Greedy (marginal gain)	8.9235	0.972	<0.01	5, 16, 17, 21, 32

Key observations:

1. **SA and MILP find the exact optimum.** Simulated annealing (10,000 swap iterations) and MILP with McCormick linearization (CBC solver) both recover the brute-force optimum, confirming that the QUBO landscape is navigable by classical metaheuristics at this scale.

2. **Greedy achieves 97.2% approximation ratio.** The greedy heuristic differs on one bus: it selects bus 5 (the highest-centrality hub, w=1.000) instead of bus 24 (a mid-feeder bus, w=0.286). This substitution reduces the dispersion component from 7.316 to 7.060 while maintaining identical stress coverage. The greedy placement is suboptimal under the QUBO objective but, as shown in Section VII-D, superior under cascade detection metrics.

3. **MILP is unexpectedly slow.** The McCormick linearization of the quadratic dispersion terms introduces $\mathcal{O}(n^2)$ auxiliary variables, making CBC solve time (88.5 s) two orders of magnitude slower than SA (1.78 s). For production use, SA or direct QUBO solvers are preferred at this scale.

C. QUBO Verification

The QUBO formulation is verified by comparing the QUBO minimizer with the original constrained objective across all 237,336 feasible solutions:

- **Consistency check:** The QUBO-optimal and original-objective-optimal solutions are identical: sensors 16, 17, 21, 24, 32 with objective 9.1796. - **Objective decomposition:** Stress component = 1.864, dispersion component = 7.316. The dispersion term dominates the objective by a factor of 3.9:1. - **Q matrix properties:** 33x33 symmetric, diagonal range [-91, -90], off-diagonal range [19, 19.99]. The Ising representation has 528 nonzero J_ij couplings (fully connected graph — no zero couplings).

D. Validation: Does the Placement Actually Improve Λ_GMonitoring?

This section closes the loop between QUBO-optimal sensor placement and the original $\Lambda_G$cascade detection problem. We evaluate each placement using proxy metrics derived from the electrical distance matrix (detection radius = 3.0).

Disambiguating QAOA failure from subproblem failure. The QAOA sensors are selected from a 20-bus subset while classical sensors are from the full 33-bus set. To isolate QAOA algorithmic quality from subproblem definition effects, we include the 20-bus exact optimum evaluated on the same 33-bus validation metrics:

Placement	Coverage	Avg Distance	Hub Accuracy	Detected Hubs
Coverage QUBO (r=3)	**96.6%**	1.209	**100%**	2,4,5,6,7
Greedy	73.8%	**1.009**	**100%**	2,4,5,6,7
alpha=0.3 optimal	76.5%	0.576	**100%**	2,4,5,6,7
Exact optimal (alpha=1)	29.0%	4.557	20%	2
20-bus exact (k=3)	17.1%	6.063	0%	
QAOA p=1 = 20-bus exact	17.1%	6.063	0%	
QAOA p=4 = 20-bus exact	17.1%	6.063	0%	

QAOA p=1 and p=4 produce sensors **identical** to the 20-bus exact optimum 17, 21, 32. The 0% hub accuracy is therefore a subproblem-definition issue (the dispersion objective pushes the 20-bus optimum away from hubs), not a QAOA algorithmic failure. All five stress hubs 2, 4, 5, 6, 7 are present in the 20-bus subset — the optimizer simply does not select them because dispersion rewards peripheral placement.

> **Finding 1: The algorithm succeeded; the formulation failed.** QAOA recovers the exact mathematical optimum of the QUBO (verified against brute-force enumeration). The operational failure — 0% hub accuracy, 17% coverage — is entirely attributable to the dispersion-dominant objective function. This distinction between *solver failure *(algorithm does not find the optimum) and* formulation failure* (the optimum itself is operationally poor) is central to this paper's contribution. Classical solvers (SA, MILP) exhibit the same formulation failure: they also find the exact optimum, which is also operationally poor. The fix is not a better solver — it is a better formulation.

Random placement distribution (100 seeds). To test whether the QUBO-optimal's poor cascade detection is statistically significant, we evaluate 100 random k=5 placements:

Metric	Median	IQR	% beating QUBO-optimal
Weighted coverage	78.9%	[72.3%, 83.2%]	**99%** (99/100)
Avg detection distance	1.132	[0.807, 1.603]	**98%** (98/100)
Hub accuracy	100%	—	**95%** (95/100)

The QUBO-optimal placement at alpha=1.0 is worse than 99% of random placements on coverage and 95% on hub accuracy. This is not a statistical fluke — the dispersion objective systematically produces the worst possible placement for cascade detection.

Alpha sensitivity and phase transition (Fig. 2). We sweep alpha over [0, 2] and evaluate each optimal placement through validation metrics (brute-force over all 237,336 placements at each alpha value):

A **sharp phase transition** occurs between alpha=0.7 and alpha=1.0: bus 5 (the highest-centrality hub, w=1.000) is

alpha	Stress Frac.	Sensors	Coverage	Hub Acc.
0.0	100%	2,3,4,5,6	60.8%	100%
0.1	91.6%	2,4,5,6,17	64.9%	100%
0.3	64.5%	2,5,17,21,32	76.5%	100%
0.5	47.5%	5,17,21,24,32	76.5%	100%
0.7	39.3%	5,17,21,24,32	76.5%	100%
1.0	**20.3%**	**16,17,21,24,32**	**29.0%**	**20%**
2.0	11.3%	16,17,21,24,32	29.0%	20%

Missing figure: quantum_sensor_placement/figures/fig1_alpha_phase_transition.pdf

Fig. 2. Alpha phase transition in the stress-dispersion QUBO. Hub detection accuracy (red, left axis) and weighted coverage (blue, right axis) vs dispersion weight alpha. A discontinuous collapse occurs at $\alpha \approx 0.8$: hub accuracy drops from 100% to 20% as bus 5 (highest-centrality hub) is replaced by bus 16 (peripheral). Green annotations show the stress fraction of the objective at each alpha. The gray band marks the phase transition region.

replaced by bus 16 (a peripheral feeder bus, w=0.286). Hub accuracy collapses from 100% to 20%, and coverage drops from 76.5% to 29.0%. The transition is discontinuous — there is no intermediate placement.

Coverage-based QUBO resolves the gap. The coverage objective (Objective 1), which maximizes the sum of stress weights within detection radius of any sensor, finds sensors 2, 8, 15, 19, 29 at radius=3.0 — achieving 96.6% weighted coverage and 100% hub detection. This confirms that the formulation gap is in the objective, not the optimization method.

Information-theoretic diagnosis: stress centrality IS information centrality. To understand why the dispersion-dominant formulation fails, we construct a proxy covariance matrix from the graph Laplacian pseudoinverse: $\Sigma = W^{1/2}\, L^{+}\, W^{1/2}$, where W = diag(w) and L is the graph Laplacian (see Section II-A, Objective 4). This encodes the tree Markov property of the radial network [17]. We compute pairwise mutual information $\mathrm{MI}(i,j) = -\frac{1}{2}\log\big(1-\rho_{ij}^2\big)$ and conditional entropy reduction I(S; Λ_G) = (1/2) log[det(Sigma) / det(Sigma|S)] for each placement.

Caveat and partial validation: proxy covariance vs cascade stress. The Laplacian pseudoinverse $L^{+}$is the covariance of a random walk on the graph, not of empirical $\Lambda_G$values. To test whether the proxy's structural predictions survive a more realistic stress model, we run a DC power flow cascade ensemble: 2,000 random contingency scenarios (load perturbations drawn uniformly from 0.5-2.0x base case) with simplified cascade propagation (overloaded lines tripped, flow redistributed, up to 3 cascade rounds). From the resulting 2,000 x 33 empirical stress matrix, we compute empirical bus-bus MI and compare with the proxy.

Metric	L^{+}Proxy	Cascade Empirical
MI-distance correlation r	-0.37	**-0.45**
Top-5 stress hub overlap with proxy	—	3/5
Proxy MI vs empirical MI (Pearson r)	—	0.60 ($p < 10^{-52}$)
Proxy MI vs empirical MI (Spearman rho)	—	0.28 ($p < 10^{-10}$)
Dispersion-optimal has lowest $I(S;\Lambda_G)$	Yes	**Yes**

The Pearson-Spearman divergence (r = 0.60 vs rho = 0.28) reflects a known property of Gaussian MI: the proxy and empirical MI share a strong linear relationship (Pearson), but the rank ordering of individual bus pairs is noisier (Spearman). This is expected because the proxy captures the graph-structural component of MI (which pairs are connected, through how many intermediaries) while the cascade ensemble adds load-dependent variance that reshuffles fine-grained pair rankings. The critical observation is that placement-level conclusions are preserved: the same placements rank best and worst under both covariance models, because placement quality depends on aggregate information capture, not individual pair rankings.

The cascade validation confirms the proxy's structural predictions:

1. *The MI-distance anti-correlation is stronger, not weaker, under cascade stress.* The empirical r = -0.45 exceeds the proxy r = -0.37 in magnitude, meaning the proxy is conservative: the actual stress field has even more anti-correlation between distance and information than L^{+}predicts. The proxy underestimates the severity of the formulation gap.

2. *The placement ranking is preserved.* Under empirical cascade covariance, the dispersion-optimal placement (alpha=1.0) captures the least mutual information (I = 9.35 nats) of all tested placements, while coverage captures the most (I = 18.23 nats). The ordering matches the proxy analysis exactly.

3. *The topological argument is proxy-independent.* The tree Markov property — that junction nodes d-separate downstream branches — is a property of the graph, not of the covariance model. The structural prediction (junction privilege $\rightarrow$ stress-weight dominance $\rightarrow$ formulation gap on trees) follows from topology alone.

4. *The topology comparison provides out-of-sample confirmation.* The prediction "meshed networks should not exhibit the phase transition" was derived from the proxy analysis and confirmed on the IEEE 57-bus system (Section VII-E).

The cascade ensemble is a simplified DC model, not a full TPWRS transient simulation with nonlinear dynamics. Full validation against TPWRS cascade data (Future Work item 9) would test whether the specific MI values hold under AC power flow with realistic contingency sequences. However, the cascade results here establish that the proxy's qualitative predictions — anti-correlation direction, placement ranking, and formulation gap — are robust to the choice of stress model.

Placement	Sensors	Coverage	Hub Acc.	I_proxy	I_cascade
Disp. alpha=0.3	2,5,17,21,32	76.5%	**100%**	**11.50**	**15.54**
MI-QUBO (beta=2.0)	7,17,21,24,32	76.6%	**100%**	11.00	12.12
MI-QUBO (beta=1.0)	8,17,21,24,32	66.3%	80%	10.98	—
Coverage (r=3)	2,8,15,19,29	**96.6%**	**100%**	10.90	**18.23**
Greedy	5,16,17,21,32	73.8%	**100%**	10.82	12.06
Exact (alpha=1.0)	16,17,21,24,32	29.0%	20%	10.63	9.35

I_proxy = I(S; Λ_G) under L^{+}covariance (nats). I_cascade = I(S; Λ_G) under empirical covariance from a 2,000-scenario DC cascade ensemble. Both columns show the same pattern: dispersion-optimal captures the least information, while stress-preserving formulations capture the most. The cascade column amplifies the gap — the dispersion- optimal's 9.35 nats is nearly 2x worse than coverage's 18.23 nats — because the empirical MI-distance anti-correlation (r = -0.45) is stronger than the proxy predicts (r = -0.37).

The data reveals something more interesting than "MI-QUBO is right, dispersion is wrong":

1. **The alpha=0.3 dispersion QUBO captures MORE mutual information than the MI-QUBO.** At alpha=0.3 (I_proxy=11.50 nats), the stress term dominates the objective, selecting hub-proximate sensors that are informationally diverse.

The MI-QUBO at beta=2.0 (I_proxy=11.00 nats) is beaten by a dispersion formulation with the right alpha. This is not accidental — it is structural.

2. **The real pattern: every stress-preserving objective succeeds.** Coverage (100% hub accuracy), greedy (100%), alpha=0.3 (100%), MI-QUBO beta=2.0 (100%) — despite being completely different formulations, they all produce operationally effective placements. The only formulation that fails is $\alpha \geq 1.0$ (20% hub accuracy). What these successful formulations share is that stress weighting dominates the objective.

3. **Stress centrality IS information centrality in radial networks.** High-centrality buses (bus 5: betweenness = 1.000) are simultaneously stress attractors — where $\Lambda_G$cascade signatures concentrate per the TPWRS finding [1] — and informationally privileged junction nodes in the tree Markov structure, where a single observation provides information about all downstream branches [17]. The stress weights encode the information structure of the network, not just operational priority.

4. **The alpha phase transition is the signal-to-noise boundary.** The dispersion term adds noise (rewarding distance, which anti-correlates with information gain at r = -0.37 proxy / -0.45 cascade). The stress term carries signal (rewarding centrality, which correlates with information gain). At $\alpha \approx 0.8$, the noise exceeds the signal: bus 5 (the highest-centrality hub, w=1.000) is replaced by bus 16 (peripheral, w=0.286), and hub accuracy collapses discontinuously.

5. **Dispersion-optimal has the LOWEST I(S; Λ_G) under both proxy and cascade.** The dispersion-exact placement captures 10.63 nats (proxy) / 9.35 nats (cascade) — less than every other formulation. Under cascade covariance, the gap widens: coverage captures 18.23 nats vs dispersion's 9.35. Maximizing dispersion in a radial network actively minimizes information gain about the stress field.

6. **Greedy achieves 98.5% of MI-optimal, consistent with submodularity.** The greedy heuristic (I=10.82 nats) vs the MI-QUBO at beta=1.0 (I=10.98 nats) gives a ratio of 0.985. This is consistent with the (1-1/e) submodularity bound from Krause et al. [15], though the bound applies to greedy on the MI objective directly, while our greedy optimizes the dispersion objective. The 98.5% ratio reflects the stress- information correlation: greedy's marginal-gain criterion selects high-stress buses, which are also high-information buses in the radial topology.

Topology-dependence prediction (confirmed in Section VII-E). This analysis predicts that the formulation gap is specific to radial (tree) networks where the tree Markov property creates the stress-information duality. In meshed transmission networks, multiple redundant paths break this Markov structure, reducing the MI-distance anti-correlation. The dispersion objective should be adequate for meshed topologies where distance better approximates information independence. Section VII-E validates this prediction on the IEEE 57-bus meshed system: no phase transition occurs (100% hub accuracy at all alphas), and the MI-distance correlation weakens from r = -0.37 to r = -0.29. This connects to the TPWRS paper [1], which established that radial and meshed networks exhibit qualitatively different stress propagation.

Generalizability: validation-loop methodology. The sequence — (1) formulate QUBO, (2) solve optimally, (3) validate against domain metrics, (4) discover that mathematical optimality diverges from operational effectiveness — applies to any quantum optimization for engineering: logistics routing, scheduling, network design. The lesson is that QUBO formulation quality matters more than solver quality, and domain validation must precede expensive quantum hardware execution.

E. Topology-Dependence Validation: IEEE 57-Bus Meshed Network

To test the prediction that the formulation gap is specific to radial (tree) networks, we run the identical pipeline on the IEEE 57-bus meshed transmission system (57 buses, 78 branches, 22 independent cycles). Stress weights are computed from impedance-weighted betweenness centrality (same method as 33-bus). The 57-bus alpha sweep uses simulated annealing (100,000 iterations per alpha, C(57,5) = 4,187,106 too large for brute-force).

Missing figure: quantum_sensor_placement/figures/fig3_topology_comparison.pdf

Fig. 3. Topology comparison: alpha phase transition on radial vs meshed networks. Left: IEEE 33-bus (tree, 0 cycles) — hub accuracy collapses from 100% to 20% at $\alpha \approx 0.8$. Right: IEEE 57-bus (meshed, 22 cycles) — hub accuracy remains 100% at all alphas. The formulation gap is topology-dependent, confirming the structural prediction from the tree Markov property.

Metric	33-bus (radial)	57-bus (meshed)
Topology	Tree (0 cycles)	Meshed (22 cycles)
Edges	32	78
MI-distance correlation r (proxy)	**-0.37**	-0.29
Hub acc. at alpha=0.0	100%	100%
Hub acc. at alpha=0.7	100%	100%
Hub acc. at alpha=1.0	**20%**	**100%**
Hub acc. at alpha=2.0	20%	100%
Coverage at alpha=1.0	29.0%	100%
Phase transition present	**Yes** ($\alpha \approx 0.8$)	**No**

The results confirm the topology-dependence prediction on all three dimensions:

1. **MI-distance anti-correlation weakens on meshed networks.** The 57-bus meshed network shows r = -0.29, weaker than the 33-bus radial r = -0.37. In meshed networks, alternative paths reduce the anti-correlation between distance and information — nearby buses are still correlated, but distant buses also share information through redundant paths. The anti-correlation does not vanish entirely because physical proximity still matters, but it is insufficient to cause the formulation gap.

2. **The alpha phase transition disappears on meshed networks.** On the 57-bus meshed system, hub accuracy remains at 100% for ALL tested alpha values (0.0 to 2.0). Even at alpha=2.0 (maximally dispersion-dominant), the optimizer selects sensors that detect all five stress hubs. The dispersion term is not antagonistic to hub detection on meshed networks because spreading sensors apart does not push them away from hubs — the meshed topology distributes hub-like connectivity across more of the network.

3. **Coverage remains at 100% across all alphas.** The 57-bus placements achieve perfect weighted coverage regardless of alpha, compared to the 33-bus collapse from 76.5% to 29.0%. This reflects the denser connectivity of the meshed network:

with 78 edges vs 32, each sensor covers more buses within the detection radius.

The structural explanation: In the 33-bus tree, all shortest paths between peripheral buses pass through the same small set of junction nodes (buses 2-7). The dispersion objective pushes sensors to opposite extremities, away from these hubs. In the 57-bus meshed network, 22 independent cycles create multiple shortest paths between any two buses. High-centrality nodes are distributed more uniformly, and the dispersion objective cannot systematically avoid them — spreading sensors apart in a meshed network naturally places some near hubs.

Caveat: The 57-bus alpha sweep uses simulated annealing rather than brute-force enumeration. While SA may not find the true optimum at each alpha, the consistent 100% hub accuracy across all 8 alpha values and multiple SA seeds makes it highly unlikely that the true optimum would show lower hub accuracy. The qualitative conclusion — no phase transition on meshed networks — is robust.

VIII. Discussion

A. What We Demonstrated

1. **Stress-information duality in radial networks, confirmed by topology comparison.** Bus-level stress centrality is a sufficient proxy for information centrality. All stress-preserving QUBO formulations (coverage, greedy, low-alpha dispersion, MI-QUBO) produce operationally effective placements with 100% hub accuracy. Only the dispersion-dominant formulation ($\alpha \geq 1.0$) fails, because the dispersion term drowns the stress signal that carries the information structure. The alpha=0.3 dispersion QUBO captures MORE mutual information (I_proxy = 11.50 nats) than the explicitly designed MI-QUBO (I_proxy = 11.00 nats) — because at low alpha, the stress weights dominate, and stress weights encode the tree's information structure. The alpha phase transition at 0.8 marks the boundary where dispersion noise exceeds the stress-information signal. The IEEE 57-bus meshed network shows no phase transition (100% hub accuracy at all alphas), confirming that the failure mode is topology-specific.

2. **End-to-end validation-loop methodology.** From QUBO formulation through solution to domain-metric validation, the pipeline discovers that mathematical optimality diverges from operational effectiveness — and diagnoses why. This methodology is reusable for any engineering QUBO application.

3. **QAOA quantum-readiness.** On a 20-qubit reduced problem, QAOA recovers the exact optimum at all depths (p=1-4) with multi-start COBYLA (9/10 success). Parameter landscape analysis (40x40 grid) confirms multiple local minima. Classical baselines (exact, SA, MILP, greedy) on the full 33-bus instance establish rigorous comparison.

B. Limitations

1. **No quantum advantage at any tested scale.** Classical simulated annealing solves the 33-bus problem to optimality in 1.78 seconds. Even at 20 qubits, QAOA on a noiseless simulator takes 43-90 seconds per depth. The value proposition for quantum methods lies at larger scales (300+ buses) where exact enumeration becomes intractable.

2. **No hardware execution.** All QAOA results use noiseless statevector simulation. Hardware noise, gate errors, and decoherence will degrade approximation ratios. Trapped-ion hardware execution (IonQ Aria, 25 qubits; Forte, 36 qubits) is the planned next step but is not included in this work.

3. **No proof that QAOA scales better than classical heuristics.** The theoretical QAOA advantage for constrained optimization is an active research question. Our 20-qubit results validate correctness, not scaling supremacy. Empirical scaling studies at 50-100+ qubits are needed to establish practical crossover points.

4. **Alternative QUBOs not yet tested with QAOA.** The coverage QUBO, MI-QUBO, and low-alpha dispersion QUBO all resolve the formulation gap classically, but have not been validated through QAOA. The coverage QUBO's auxiliary-variable structure (2n qubits) makes it more expensive on quantum hardware. The MI-QUBO and low-alpha dispersion have the same n-qubit structure and are directly compatible with the existing QAOA pipeline, but their landscape properties (local minima, optimizer convergence) may differ from the alpha=1.0 case.

C. When Quantum Becomes Relevant

Scale	Exact Classical	Heuristic (SA)	QAOA (p=2, dense)	Winner
33 bus, k=5	0.8 s	1.8 s	3,234 gates (sim)	Classical
118 bus, k=10	11 years	6 s	41,654 gates	Classical (SA)
300 bus, k=10	10^5 years	16 s	269,700 gates	Classical (SA)*
1,000 bus, k=15	intractable	54 s	3.0 M gates	Uncertain

SA finds high-quality but not provably optimal solutions. At n=300+, no classical method provides optimality certificates. QAOA with error correction provides a principled alternative once hardware reaches Tier 2 capability (Section VI).

D. Stress-Information Duality: Why the Formulation Gap Is Structural

The most actionable finding of this work is that QUBO formulation quality matters more than solver quality at the current scale. But the deeper insight is WHY the formulation gap exists and what it reveals about the relationship between stress centrality and information centrality in radial networks.

The key observation is not that "MI-QUBO is right and dispersion is wrong." The alpha=0.3 dispersion QUBO captures MORE mutual information (I_proxy = 11.50 nats) than the explicitly designed MI-QUBO (I_proxy = 11.00 nats at beta=2.0). Instead, the data shows that **any objective preserving stress-weight dominance produces good placements**, while **only dispersion-dominant objectives fail**. This points to a structural property, not an objective-design insight.

Stress centrality = information centrality in tree networks. In the IEEE 33-bus radial system, the stress weights w_i are derived from betweenness centrality — a topological measure of how many shortest paths pass through bus i. In a tree graph, these same junction nodes are informationally privileged: the tree Markov property [17] implies that a junction node d-separates all downstream buses, making it a sufficient statistic for the conditional distribution of downstream stress. The

TPWRS paper [1] independently established that cascade stress concentrates at these same high-centrality buses.

This triple coincidence — topological centrality, information privilege, and stress concentration — is not coincidental. It is a structural property of tree-topology networks where all paths converge through junction nodes. The stress weights encode the network's information structure because they reflect the same topological feature (path convergence) that determines conditional independence.

The dispersion term is noise, not signal. The dispersion objective rewards distance between sensors. But in the radial network, MI anti-correlates with electrical distance (r = -0.37 proxy, -0.45 cascade): nearby buses share MORE information, because they connect through fewer trunk-line segments. The dispersion term pushes sensors toward peripheral, informationally redundant positions. When alpha is small (stress dominates), this noise is overwhelmed. When alpha reaches 0.8, the noise exceeds the signal, and the optimizer replaces hub-proximate sensors with peripheral ones — collapsing hub accuracy discontinuously from 100% to 20%.

Quantified evidence:

1. **MI-distance anti-correlation (r = -0.37 proxy, r = -0.45 empirical).** Computed from the Laplacian covariance proxy across all 528 bus pairs. A DC power flow cascade ensemble (2,000 scenarios, Section VII-D) confirms the anti-correlation is stronger under cascade stress (r = -0.45) than the proxy predicts (r = -0.37), meaning the proxy is conservative. Proxy and empirical MI are correlated at r = 0.60 ($p < 10^{-52}$). Full TPWRS transient simulation would test nonlinear propagation effects.

2. **Consistent success of stress-preserving formulations.** Coverage (100% hub accuracy), greedy (100%), alpha=0.3 (100%), alpha=0.5 (100%), alpha=0.7 (100%), MI-QUBO beta=2.0 (100%) — six different objectives all succeed because they all select hub-proximate sensors. MI-QUBO at beta=1.0 is an intermediate case (80% hub accuracy): the MI redundancy penalty is large enough to shift one sensor but insufficient to fully redirect the placement toward hubs. This confirms that the critical factor is stress-weight dominance, not information-theoretic sophistication — even an explicit MI objective can underperform when its penalty term partially reproduces the dispersion failure mode.

3. **Discontinuous phase transition at $\alpha \approx$ 0.8.** The critical substitution (bus 5 $\rightarrow$ bus 16) occurs in a single discrete step. There is no gradual degradation. This discontinuity reflects the discrete nature of the combinatorial optimum: a small change in alpha causes a qualitative jump in the optimal solution.

4. **99/100 random placements beat the dispersion-optimal.** The dispersion-optimal is not borderline — it is an extreme outlier on the wrong side of virtually every random alternative, confirming that the formulation gap is severe, not marginal.

Topology-dependence confirmation (Section VII-E). The stress-information duality depends on the tree Markov property, which holds only for tree (radial) networks. We tested this prediction on the IEEE 57-bus meshed transmission system (78 edges, 22 independent cycles). The results confirm all three predictions: - The MI-distance anti-correlation weakens: r = -0.29 (57-bus) vs r = -0.37 (33-bus) - The alpha phase transition disappears: 100% hub accuracy at ALL tested alphas (0 to 2.0) - The dispersion-dominant formulation produces perfect placements on the meshed network

The 57-bus results are not merely "consistent with" the prediction — they are a clean confirmation. The 33-bus phase transition (80% hub accuracy drop at alpha=0.7->1.0) vanishes entirely on the meshed network (0% drop). This connects to the TPWRS paper [1], which established that radial and meshed networks exhibit qualitatively different $\Lambda_G$behavior. The QUBO formulation gap is the optimization-side manifestation of that diagnostic-side finding: the same tree topology that creates distinctive cascade propagation also creates a distinctive QUBO failure mode.

E. Recommendations for the Quantum Power Systems Community

1. **Validate QUBO formulations against domain metrics before quantum execution.** Our formulation gap finding demonstrates that mathematical optimality can diverge from operational effectiveness. Close-the-loop validation is not optional.

2. Combinatorial optimization is the most credible near-term quantum application in power systems. Sensor placement, topology reconfiguration, and contingency screening are naturally QUBO-encodable and scale combinatorially.

3. **Stop claiming QPE speedups for non-Hermitian Jacobians** without addressing the Hermitianization cost. The commutator-based diagnostic $\Lambda_G$has a natural quantum pathway, but the advantage threshold is 10^4–10^5 buses.

4. **The encoding bottleneck deserves more attention.** Getting classical grid data (impedance matrices, Jacobians) into quantum states is the binding constraint for quantum monitoring — not the algorithmic speedup itself.

IX. Conclusion

We presented a formulation-to-validation pipeline for QUBO-based sensor placement on the IEEE 33-bus radial distribution system, discovering and diagnosing a fundamental formulation gap in distance-based QUBO objectives.

Central result: stress-information duality. In radial networks, bus-level stress centrality is a sufficient proxy for information centrality. The tree Markov property makes junction nodes simultaneously stress attractors and informationally privileged, so any QUBO formulation preserving stress-weight dominance produces operationally effective placements. The standard dispersion-dominant formulation fails because its distance term — which anti-correlates with information gain (r = -0.37 proxy, -0.45 cascade) — drowns the stress signal at $\alpha \geq$ 0.8. The formulation gap is a structural property of tree-topology networks: a comparative experiment on the IEEE 57-bus meshed system shows no phase transition (100% hub accuracy at all alphas), confirming the prediction.

Methodology contribution. The validation loop — formulate, solve optimally, validate against domain metrics, diagnose divergence — is the paper's most transferable contribution. The key lesson is Finding 1: when the QUBO-optimal placement fails operationally, the failure lies in the formulation, not the solver (quantum or classical). SA, MILP, and QAOA all find the same mathematical optimum; all produce the same operationally poor placement. The fix is a better objective function, not a better algorithm. This diagnostic methodology generalizes to any engineering QUBO application.

QAOA quantum-readiness. As a supporting validation, QAOA on a 20-qubit subproblem recovers the exact optimum at all tested depths (p=1-4) with multi-start COBYLA (9/10 success rate), confirming that the corrected QUBOs are executable on near-term quantum hardware. Production deployment should adopt constrained QAOA with Hamming-weight- preserving mixers [24] to eliminate the post-selection overhead observed with penalty- based encoding (0.5-7.5% feasibility rates). Classical baselines on the full 33-bus instance establish that quantum advantage is not claimed at this scale; the contribution is a validated formulation pipeline ready for hardware execution at scales where classical methods become intractable.

X. Future Work

1. **Extended topology comparison (IEEE 118-bus, 300-bus).** The 57-bus meshed comparison (Section VII-E) confirms the topology-dependence prediction. Extending to the IEEE 118-bus system (186 branches, 70 cycles) and larger synthetic networks would map the relationship between cycle count and formulation gap magnitude. Of particular interest: partially meshed distribution systems (radial backbone with tie switches), which are the most operationally relevant topology class.

2. **Coverage QUBO through QAOA.** The coverage objective resolves the formulation gap classically, but its 2n-qubit auxiliary-variable structure has not been tested with QAOA. Evaluate whether the coverage QUBO's different Ising structure (sparser couplings due to the radius threshold) produces a more favorable QAOA landscape.

3. **MI-QUBO through QAOA.** The information-theoretic QUBO (Objective 4) has been validated classically (Section VII-D) but not yet executed through QAOA. Its n-qubit structure (no auxiliary variables) and comparable coupling density make it directly compatible with the existing QAOA pipeline. Comparing QAOA performance on the MI-QUBO vs the dispersion-QUBO would test whether landscape properties (local minima, barren plateaus) differ across objective functions encoding different physical priors.

4. **Hardware execution.** Execute the 20-qubit QAOA circuits on IonQ Aria (25 qubits) and the full 33-qubit problem on IonQ Forte (36 qubits). Compare hardware approximation ratios with noiseless simulator results to quantify noise impact.

5. **Scaling studies.** Extend to IEEE 69-bus, 118-bus, and 300-bus test systems. At 118+ buses, classical exact enumeration is intractable (C(118,10) > 10^{13}), providing the regime where quantum methods may offer practical advantage.

6. **Integration with $\Lambda_G$simulation.** Replace proxy validation metrics with full TPWRS cascade simulation, using time-series $\Lambda_G$values at sensor locations to measure actual detection latency and localization accuracy.

7. **Warm-starting and multi-start QAOA.** Use classical greedy or SA solutions as QAOA warm-start states [8]. Our COBYLA restart analysis shows that multi-start optimization (9/10 success rate) is essential for production QAOA at p >= 2; warm-starting may reduce the required restart count.

8. **Hub-aware constraints.** Add QUBO constraints requiring at least one sensor within electrical distance r of each identified stress hub, encoding domain knowledge directly into the optimization formulation.

9. **Empirical MI from cascade simulations.** The current MI analysis uses the graph Laplacian pseudoinverse as a covariance proxy. Replace with empirical covariance estimated from TPWRS cascade simulation ensembles, where the time-series $\Lambda_G^{(i)}$values at each bus define the joint distribution directly. This would validate whether the Laplacian-based MI structure faithfully represents the actual stress-field correlation structure.

XI. References

[1] Mathews, R.J. (2026). "Predictive Grid Stress Diagnostics via Non-Commutativity of Power Flow Jacobians." Submitted to IEEE Trans. Power Systems. [Preprint and supplementary data available upon request.] [2] Mathews, R.J. (2025). U.S. Provisional Patent 63/903,809. [3] Farhi, E., Goldstone, J., & Gutmann, S. (2014). "A Quantum Approximate Optimization Algorithm." arXiv:1411.4028. [4] Eskandari, H., et al. (2023). "Quantum computing for power system eigenanalysis." Nature Scientific Reports. [5] Avron, H. & Toledo, S. (2011). "Randomized algorithms for estimating the trace of an implicit symmetric positive semi-definite matrix." J. ACM, 58(2), 1-34. [6] Baran, M.E. & Wu, F.F. (1989). IEEE Trans. Power Delivery, 4(2), 1401-1407. [7] Camps, D., et al. (2022). "Explicit quantum circuits for block encodings of sparse matrices." arXiv:2203.10236. [8] Egger, D.J., et al. (2021). "Warm-starting quantum optimization." Quantum, 5, 479. [9] Kundur, P. (1994). Power System Stability and Control. McGraw-Hill. [10] U.S.-Canada Power System Outage Task Force (2004). Final Report on the August 14, 2003 Blackout. [11] Zhou, L., et al. (2020). "Quantum approximate optimization algorithm: Performance, mechanism, and implementation on near-term devices." PRX, 10(2). [12] Wright, K., et al. (2019). "Benchmarking an 11-qubit quantum computer." Nature Communications, 10(1), 5464. [IonQ trapped-ion architecture] [13] Harrow, A.W., Hassidim, A., & Lloyd, S. (2009). "Quantum algorithm for linear systems of equations." Phys. Rev. Lett., 103(15), 150502. [HHL algorithm] [14] Kitaev, A.Y. (1995). "Quantum measurements and the Abelian Stabilizer Problem." arXiv:quant-ph/9511026. [QPE] [15] Krause, A., Singh, A., & Guestrin, C. (2008). "Near-Optimal Sensor Placements in Gaussian Processes: Theory, Efficient Algorithms and Empirical Studies." Journal of Machine Learning Research, 9, 235-284. [16] Shannon, C.E. (1948). "A Mathematical Theory of Communication." Bell System Technical Journal, 27(3), 379-423. [17] Cover, T.M. & Thomas, J.A. (2006). Elements of Information Theory. 2nd ed. Wiley-Interscience. [Gaussian MI, conditional entropy, tree Markov property] [18] IonQ (2024). "IonQ Forte System Specifications." Technical documentation. [19] Pelofske, E., Bärtschi, A., & Eidenbenz, S. (2024). "Quantum annealing vs. QAOA: 127 qubit higher-order Ising problems on NISQ computers." Scientific Reports, 14.

[20] Xu, B. & Abur, A. (2004). "Observability analysis and measurement placement for systems with PMUs." IEEE PES Power Systems Conference and Exposition, 2, 943-946. [21] Manousakis, N.M., Korres, G.N., & Georgilakis, P.S. (2012). "Taxonomy of PMU placement methodologies." IEEE Trans. Power Systems, 27(2), 1070-1077. [22] Ajagekar, A. & You, F. (2019). "Quantum computing for energy systems optimization: Challenges and opportunities." Energy, 179, 76-89. [23] Feng, B., et al. (2023). "Quantum computing for power system optimization: A survey." Electric Power Systems Research, 218, 109196. [24] Hadfield, S., et al. (2019). "From the quantum approximate optimization algorithm to a quantum alternating operator ansatz." Algorithms, 12(2), 34. [25] Bärtschi, A. & Eidenbenz, S. (2019). "Deterministic preparation of Dicke states." Fundamentals of Computation Theory (FCT), LNCS 11651, 126-139.

XII. Appendices

A. Full Q Matrix for 33-Bus Stress-Dispersion QUBO

The Q matrix is 33x33 symmetric with: - Diagonal entries Q_ii in range [-91.00, -89.96], encoding stress weights w_i and the penalty contribution lambda(1 - 2k) for k=5, lambda=10. - Off-diagonal entries Q_ij in range [19.00, 19.99], encoding dispersion alpha*d_ij and the penalty 2*lambda. - Full numerical values archived in `results/qubo_matrix_33bus.npz` for reproducibility.

B. QAOA Circuit Structure

The custom QAOA circuit for n qubits and depth p consists of: 1. Initial layer: n Hadamard gates (equal superposition) 2. Cost unitary (per layer): For each nonzero J_ij coupling, a CNOT-Rz(2*gamma*J_ij)-CNOT sequence, plus single-qubit Rz(2*gamma*h_i) for each local field. Total per layer: 2*|E| CNOT gates + |E| + n Rz gates, where |E| is the number of nonzero couplings. 3. Mixer unitary (per layer): n Rx(2*beta) gates. 4. Measurement in computational basis.

For the 20-bus reduced problem: |E|=190 couplings, giving 380 CNOTs + 210 Rz + 20 Rx per layer. Circuit depth = 62p + 53 (accounting for parallelizable single-qubit gates).

C. IEEE 33-Bus Network Data

Standard Baran & Wu (1989) test system: 33 buses, 32 branches, radial tree topology. Branch impedances (R + jX per unit) define the electrical distance matrix via impedance-magnitude-weighted shortest paths. Network data and distance matrix archived in the supplementary code repository.

D. Stress Weight Proxy Construction

In the absence of full TPWRS $\Lambda_G$simulation data, proxy stress weights are constructed as:

w_i = normalize(betweenness_centrality(i) + 0.3 * is_leaf(i))

where betweenness centrality is computed on the impedance-weighted graph and is_leaf(i) is 1 if bus i has degree 1 (feeder endpoint vulnerability bonus). Weights are normalized to [0, 1]. The highest-weight bus is bus 5 (w=1.000, primary branching hub). When TPWRS simulation data becomes available, these proxy weights should be replaced with empirical $\Lambda_G$stress rankings from cascade simulations.