

# Madmartigan: A 96-Active-Qubit Structured-Output Benchmark on NISQ Hardware

Reference-Specific Control Separation Without Quantum Error Correction or  
Post-Selection

Frank Angelo Drew

May 13, 2026

## Abstract

This paper consolidates the technical findings from the scale-up of the original 16-qubit Madmartigan benchmark into a 96-active-qubit, six-tile benchmark package executed on IBM superconducting hardware. The central objective was not merely to increase qubit count, but to determine whether the original 16-qubit Madmartigan structural output could be replicated across multiple physical tile paths while preserving reference-specific alignment under realistic NISQ constraints: no quantum error correction, no post-selection, limited shot count, real backend calibration, topology constraints, and transpilation overhead.

The final 96-qubit package used six simultaneous 16-qubit Madmartigan tiles on IBM Marrakesh. The strongest layout, designated GLOBALPACK T6 rank-2, preserved the original four validated tile paths and added two additional scale-up tiles. This circuit ran with 96 active qubits, 156 measured qubits, seed 1337, 4096 shots per run, no QEC, and no post-selection. It transpiled to depth 495 with 1185 CZ gates, 2546 SX gates, 2489 RZ gates, 156 measurements, and 66 barriers.

Across repeated hardware runs, the raw T6 rank-2 Madmartigan circuit produced elevated XEB and HOG values across the core tile set, with five of six tiles repeatedly maintaining positive Madmartigan-reference alignment and several tiles producing XEB values above 1.0. The sixth tile remained weaker but generally positive, which is meaningful given the topology and calibration burden associated with scale-up.

The benchmark package was then hardened using a control ladder. A same-layout generic RCS control showed positive own-reference structure while repeatedly collapsing to negative or near-baseline XEB and baseline HOG against the Madmartigan seed-1337 reference. A phase-scrambled Madmartigan-adjacent control was then run with a heavier circuit burden than raw T6 rank-2; it preserved its own scrambled reference structure but failed to reproduce the Madmartigan reference band. Finally, a partial-entanglement ablation control revealed an unexpected QSCE-derived high-attractor mode: the ablated circuit produced very high own-reference XEB/HOG while retaining only weak-to-moderate overlap with the original Madmartigan reference. This showed that the QSCE/Madmartigan design family contains more than one structured-output regime: the original Madmartigan command-preservation mode and a derivative high-attractor mode.

Subsequent attempts to scale cleanly beyond 96 active qubits to 128 and 144 active qubits revealed a topology-packing wall rather than a Madmartigan signal failure. Fixed-six, fixed-four T8, and fixed-four T9 scans across Marrakesh, Fez, and Kingston failed to find enough mutually disjoint 16-qubit chain tiles under the original clean-tile constraints. The next valid large-scale experiment was therefore reframed as a 144-active-qubit backend-pressure test: preserve the proven 96-qubit T6 rank-2 core exactly, then add three best-effort stress tiles to load the remaining backend and test whether the original core survives added execution pressure.

The result of the work to date is a mature 96-active-qubit scientific evidence package: raw repeatability, same-layout generic RCS control separation, architecture-adjacent phase-scrambled control separation, and partial-entanglement ablation discovery. The package materially strengthens the claim that Madmartigan is not merely a tile-quality artifact, generic random-circuit behavior, phase-insensitive hardware bias, or shallow calibration effect. It supports the view that the Madmartigan program preserves a reference-specific structured-output band under NISQ hardware constraints and that QSCE contains a broader attractor design space.

# 1 Background and Original Benchmark Motivation

The original Madmartigan benchmark began as a 16-qubit circuit designed to test whether a QSCE-derived program could produce structured, reference-aligned output on real NISQ hardware. Unlike ordinary random circuit sampling, the Madmartigan circuit was not intended merely to produce a hard-to-simulate random distribution. Its purpose was to test whether an engineered phase/correlation program could steer hardware output toward a stable, semantically meaningful probability structure.

The conceptual basis is the QSCE premise that state preparation, phase structure, entanglement, and measurement can be treated as a command-bearing substrate. In this framing, the measured output distribution is not merely noise or random sampling; it is a collapse surface through which a quantum program can produce structured classical information. The Madmartigan circuit became a benchmark instance of this idea.

The original 16-qubit result was important because it established an initial reference distribution. The seed-1337 Madmartigan ideal top states formed the benchmark target used throughout the later scale-up work. The same 16-qubit ideal distribution remained the reference used for per-tile scoring in the larger tiled experiments.

This is important because the 96-qubit scale-up was not a new unrelated circuit. It was a replicated, tiled extension of the original 16-qubit Madmartigan program. Each tile ran the same 16-qubit Madmartigan logic, preserving the original seed and reference distribution, while the larger circuit tested whether multiple simultaneous physical embeddings could preserve the same output band under real backend pressure.

## 2 Core Research Question

The central research question was:

Can the original 16-qubit Madmartigan structural output be preserved across a multi-tile, 96-active-qubit execution on real superconducting hardware without QEC, without post-selection, and using only 4096 shots?

A secondary question followed:

If the raw benchmark succeeds, can controls demonstrate that the observed structure is not explained by tile path, generic random circuit behavior, phase-insensitive hardware bias, or simply running a deep/high-CZ circuit on the same backend?

A third question emerged during ablation testing:

If the exact Madmartigan command pathways are disrupted but the broader QSCE/Madmartigan skeleton remains, does the system collapse into noise, or does it reveal another structured-output regime?

The final answer was layered:

1. The 96-qubit raw Madmartigan circuit preserved meaningful reference-specific structure.
2. Same-layout generic RCS did not reproduce the Madmartigan band.
3. Phase-scrambled Madmartigan-adjacent control did not reproduce the Madmartigan band.
4. Partial-entanglement ablation produced a new high-attractor QSCE-derived mode, but not the original Madmartigan command band.

5. Attempts to scale cleanly to 128 and 144 active qubits using clean independent 16-qubit chain tiles were limited by topology packing, not by demonstrated signal collapse.

### 3 The 16-Qubit Madmartigan Tile Program

Each Madmartigan tile is a 16-qubit circuit composed of structured phases that mimic the broader QSCE pathway grammar:

1. Global QSCE anchoring
2. EBA-style correlation core
3. QMCA-style collapse spine
4. SQCA-style calibration lattice
5. QPSA-style impulse/fanout routing
6. BSCM-style phase mesh and clamp
7. Interleaved random single-qubit and nearest-neighbor entangling layers

Each tile is evaluated against the seed-1337 Madmartigan ideal distribution. The top ideal states for seed 1337 include the following reference bitstrings:

```
|1101010011010000>
|0101010000010000>
|1101010011010010>
|0111000000101111>
|0101100011111011>
|1101010011010101>
|0101100000010000>
|1101100010101111>
```

These ideal states are not treated merely as arbitrary bitstrings. They define the reference surface against which hardware output is scored using XEB, HOG, entropy, IPR, and top-8 overlap.

## 4 Metrics Used

### 4.1 Cross-Entropy Benchmarking Fidelity Proxy: $F_{\text{XEB}}$

For each tile, the projected 16-bit hardware counts are scored against a 16-qubit ideal probability distribution. The metric used is:

$$F_{\text{XEB}} = \sum_{\mathbf{x}} p_{\text{exp}}(\mathbf{x}) * 2^{-n} * p_{\text{ideal}}(\mathbf{x}) - 1$$

Where:

- $p_{\text{exp}}(\mathbf{x})$  is the observed hardware probability for bitstring  $\mathbf{x}$ .
- $p_{\text{ideal}}(\mathbf{x})$  is the ideal simulated probability for bitstring  $\mathbf{x}$ .
- $n = 16$  for each tile.

Higher  $F_{\text{XEB}}$  means hardware samples are landing preferentially in regions that the ideal circuit predicts to be high probability.

The critical point is that XEB is reference-dependent. A circuit can have high own-reference XEB while low Madmartigan-reference XEB if it preserves a structured distribution different from the original Madmartigan distribution.

## 4.2 Heavy Output Generation: HOG

HOG measures how much observed probability mass lands on bitstrings whose ideal probabilities exceed the ideal median. Values above 0.5 indicate heavy-output preference relative to the reference distribution.

As with XEB, HOG is reference-dependent. A control circuit may show high HOG against its own reference but baseline HOG against the Madmartigan reference.

## 4.3 Shannon Entropy

Shannon entropy measures the spread of the observed tile distribution. For 16-bit tile outputs, the maximum possible entropy is 16 bits, but with 4096 shots, observed entropy is also constrained by sampling sparsity.

The T6 rank-2 raw and controls generally occupied an entropy window around 11.7–11.9 bits per 16-bit tile. This is important because it shows that the outputs were neither single-state collapses nor fully flat over all  $2^{16}$  bitstrings. They remained high-dimensional, NISQ-limited structured distributions.

## 4.4 Inverse Participation Ratio: IPR

IPR estimates the effective number of populated states. Across the T6 rank-2 work, IPR values typically clustered in the low-to-mid thousands for 4096-shot tile projections. Similar entropy/IPR windows across raw and control circuits indicate similar distributional spread, but not necessarily the same reference identity.

## 4.5 Top-8 Overlap

Top-8 overlap counts whether the hardware’s top eight observed bitstrings intersect the ideal top eight bitstrings. At 4096 shots over a 65,536-state tile space, exact top-8 overlap is a stringent and noisy measure. It is useful as supporting evidence but should not be treated as the sole performance indicator.

# 5 Scaling Strategy: From 16 Qubits to Multiple Tiles

The scale-up strategy was to preserve the 16-qubit Madmartigan core as a tile and replicate it across physical qubit paths on IBM’s 156-qubit Heron r2-class backend. The main challenge was not simply selecting any 96 qubits. The challenge was finding 16-qubit physical paths that were sufficiently connected and sufficiently calibrated to support the local Madmartigan entanglement structure.

The final 96-qubit benchmark used six tiles:

Tile 0: [7, 8, 9, 10, 11, 18, 31, 32, 33, 34, 35, 19, 15, 14, 13, 12]  
 Tile 1: [147, 148, 149, 150, 151, 138, 131, 132, 133, 134, 135, 139, 155, 154, 153, 152]  
 Tile 2: [23, 22, 21, 36, 41, 42, 43, 44, 45, 46, 47, 57, 67, 66, 65, 56]  
 Tile 3: [92, 91, 98, 111, 110, 109, 118, 129, 128, 127, 126, 125, 108, 107, 106, 97]  
 Tile 4: [53, 54, 55, 59, 75, 74, 73, 72, 71, 70, 69, 78, 89, 88, 87, 86]  
 Tile 5: [85, 84, 83, 96, 103, 102, 101, 116, 121, 122, 123, 136, 143, 142, 141, 140]

Tiles 0–3 were the original proven fixed four. Tiles 4–5 were selected by the GLOBALPACK T6 rank-2 scan. This became the flagship 96-qubit benchmark substrate.

## 6 Calibration-Aware 64-Active-Qubit Replication

Before the final 96-active-qubit T6 rank-2 package was established, the first technically clean scale-up milestone was the calibration-aware 64-active-qubit Madmartigan replication on IBM Marrakesh. This experiment was important because it transformed the original 16-qubit Madmartigan benchmark from a single physical-region result into a spatially replicated, multi-tile hardware result.

The objective was not yet to maximize qubit count. The objective was to answer a more foundational question:

Can the exact original seed-1337 Madmartigan structural-output signature reproduce across multiple physical regions of the same superconducting backend when topology and calibration are respected?

The final 64-active-qubit configuration used:

Backend: IBM Marrakesh  
Tiles: 4  
Tile size: 16 qubits  
Active qubits: 64  
Measured qubits: 156  
Seed: 1337  
Shots: 4096  
QEC: none  
Post-selection: none  
Combined transpiled depth: 510  
CZ gates: 791  
SX gates: 1696  
RZ gates: 1661

The selected physical tile paths were:

Tile 0:

[7, 8, 9, 10, 11, 18, 31, 32, 33, 34, 35, 19, 15, 14, 13, 12]

Tile 1:

[147, 148, 149, 150, 151, 138, 131, 132, 133, 134, 135, 139, 155, 154, 153, 152]

Tile 2:

[23, 22, 21, 36, 41, 42, 43, 44, 45, 46, 47, 57, 67, 66, 65, 56]

Tile 3:

[92, 91, 98, 111, 110, 109, 118, 129, 128, 127, 126, 125, 108, 107, 106, 97]

These four tiles were not selected arbitrarily. They emerged from a calibration-aware search that excluded high-readout-error qubits and bad CZ edges, then ranked candidate tile paths by actual Madmartigan transpilation burden. The result was a physically grounded tile map that avoided the routing explosion seen in earlier naïve 9-tile attempts.

This was a critical engineering correction. Earlier monolithic and naïve tiled approaches produced output spaces or routed circuits that were not meaningfully comparable to the original 16-qubit benchmark. The 64-qubit calibration-aware layout preserved the original metric regime: each 16-qubit tile was projected and

scored against the corresponding seed-1337 ideal Madmartigan distribution.

A representative first hardware run produced:

```
Tile 0: XEB = 2.092090 | HOG = 0.734375 | Entropy = 11.8655/16 | IPR = 3583.34
Tile 1: XEB = 1.289995 | HOG = 0.691650 | Entropy = 11.8818/16 | IPR = 3644.05
Tile 2: XEB = 0.954248 | HOG = 0.639648 | Entropy = 11.9083/16 | IPR = 3738.24
Tile 3: XEB = 1.477451 | HOG = 0.684814 | Entropy = 11.8910/16 | IPR = 3679.21
```

The important observation is that all four tiles returned positive XEB and elevated HOG against the Madmartigan reference. The result did not merely show that a single 16-qubit region could preserve the Madmartigan signature. It showed that the same circuit, with the same seed and reference distribution, could reproduce across four separate calibration-aware physical regions in one simultaneous hardware execution.

The 64-qubit result was then repeated across five hardware runs. Across the full repeatability package:

```
Runs: 5
Tiles per run: 4
Total tile-level Madmartigan results: 20
Active qubits per run: 64
Shots per run: 4096
QEC: none
Post-selection: none
```

The repeated run band was:

```
XEB:      approximately 0.891 to 2.892
HOG:      approximately 0.639 to 0.763
Entropy:  approximately 11.830 to 11.908 bits
IPR:      approximately 3442 to 3745
```

All 20 tile-level XEB values were positive. All 20 tile-level HOG values were above 0.5. The entropy and IPR values remained clustered in the same broad NISQ structural-output regime rather than collapsing into a trivial low-entropy state or diffusing into uninformative baseline behavior.

This established the first clean multi-region Madmartigan replication result. The proper interpretation is:

The exact seed-1337 Madmartigan program preserved a reproducible structural-output band across four calibration-aware physical tile regions on IBM Marrakesh at 64 active qubits, with no QEC, no post-selection, and only 4096 shots.

This result became the foundation for later scale-up. The 96-qubit T6 result did not emerge from an untested one-off single-tile benchmark. It emerged after a 64-active-qubit calibration-aware replication package had already demonstrated spatial reproducibility and repeatability.

## 7 Generic RCS Control on the 64-Active-Qubit Tile Map

The 64-active-qubit replication result required a depth-comparable control. Without a control, the elevated Madmartigan XEB/HOG band could still be challenged as a function of good physical tiles, backend calibration, shot count, or generic random-circuit behavior.

The most important control at this stage was a generic RCS circuit run on the same four selected physical tiles, using the same backend, same 4096-shot count, no QEC, and no post-selection.

The generic RCS control had:

Backend: IBM Marrakesh  
Tiles: 4  
Active qubits: 64  
Measured qubits: 156  
Shots: 4096  
QEC: none  
Post-selection: none  
Depth: 581  
CZ gates: 759  
RZ gates: 2169  
SX gates: 1531

The corresponding Madmartigan circuit had:

Depth: 510  
CZ gates: 791  
RZ gates: 1661  
SX gates: 1696

This made the generic RCS control a meaningful depth- and CZ-comparable negative comparator. It was not a shallow dead circuit. It carried a similar entangling burden and used the same selected hardware regions.

The control was evaluated using dual-reference scoring:

ownXEB / ownHOG:

The generic RCS hardware output scored against its own ideal reference.

madRefXEB / madRefHOG:

The same generic RCS hardware output scored against the Madmartigan ideal reference.

This distinction is essential. A valid control does not need to be dead. In fact, the stronger control is one that preserves its own structure while failing to reproduce the Madmartigan-specific structure.

That is what happened.

Across repeated generic RCS control runs, the control showed positive own-reference behavior but collapsed against the Madmartigan reference. The Madmartigan-reference control band was:

madRefXEB: mostly negative or near zero, approximately -0.183 to +0.037  
madRefHOG: approximately 0.468 to 0.529

By contrast, the repeated Madmartigan band on the same tile map was:

Madmartigan XEB: approximately 0.891 to 2.892  
Madmartigan HOG: approximately 0.639 to 0.763

This separation is the critical result. It shows that the selected physical tiles were capable of preserving circuit structure, but generic RCS structure did not align with the Madmartigan reference distribution.

The correct interpretation is:

The 64-active-qubit generic RCS control demonstrated that the Madmartigan structural band is not explained by good qubits, backend calibration, shot count, depth/CZ burden, or generic



random-circuit behavior alone. The same physical tile map could preserve a generic circuit’s own reference structure while failing to reproduce the Madmartigan reference band.

This was the first major control separation in the scale-up campaign. It upgraded the 64-qubit result from raw repeatability to raw repeatability plus depth-comparable generic control separation.

IBM Marrakesh Physical Map: XEB Separation, Madmartigan minus Generic RCS Mad-Reference

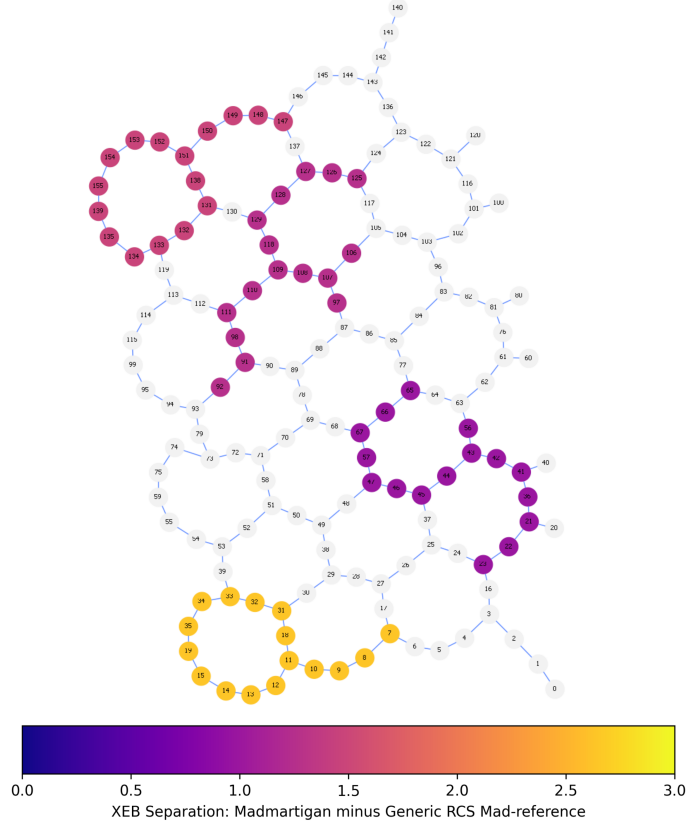


Figure 1: IBM Marrakesh physical tile map for the 64-active-qubit calibration-aware Madmartigan replication. Each qubit in a selected tile is colored by that tile’s XEB separation value, defined as raw Madmartigan-reference XEB minus same-layout generic RCS Madmartigan-reference XEB. Warmer colors indicate stronger positive reference-specific separation in favor of Madmartigan. Because XEB is a tile-level metric, each qubit in a tile is colored by the corresponding tile-level XEB separation value.

## 8 Multi-Seed Robustness on the 64-Active-Qubit Tile Map

After the 64-active-qubit seed-1337 result repeated and separated from a generic RCS control, the next question was whether the result was unique to seed 1337.

The multi-seed campaign was designed to answer:

Was seed 1337 a lucky random instance, or does the Madmartigan architecture preserve a structural-output band across multiple nearby seed instances?

The experiment used the same four calibration-aware physical tile paths:

Tile 0:

[7, 8, 9, 10, 11, 18, 31, 32, 33, 34, 35, 19, 15, 14, 13, 12]

Tile 1:

[147, 148, 149, 150, 151, 138, 131, 132, 133, 134, 135, 139, 155, 154, 153, 152]

Tile 2:

[23, 22, 21, 36, 41, 42, 43, 44, 45, 46, 47, 57, 67, 66, 65, 56]

Tile 3:

[92, 91, 98, 111, 110, 109, 118, 129, 128, 127, 126, 125, 108, 107, 106, 97]

The hardware and analysis conditions were:

Backend: IBM Marrakesh

Active qubits: 64

Tile size: 16 qubits

Shots: 4096

QEC: none

Post-selection: none

Circuit family: raw Madmartigan

Metric reference: each seed scored against its own 16-qubit ideal Madmartigan  
↪ distribution

The last point is important. Seed 1338 was not scored against seed 1337's ideal distribution. Seed 1339 was not scored against seed 1337's ideal distribution. Each seed was scored against its own ideal statevector reference. That makes the seed-robustness comparison fair.

The first seed screen tested:

Seed 1338

Seed 1339

Seed 1340

Seed 1341

The first-pass transpilation profiles were comparable:

Seed 1338: depth 506 | CZ 752

Seed 1339: depth 518 | CZ 752

Seed 1340: depth 509 | CZ 752

Seed 1341: depth 515 | CZ 752

The result split into two classes:

Strong/repeatable seed family:

1337, 1338, 1339

Weak-positive comparator seeds:

1340, 1341

This is the correct and scientifically useful outcome. It does not imply that every seed is equally strong. It shows seed-sensitive robustness rather than total seed invariance.

Seed 1338 produced a strong 4/4 tile first-pass result:

Tile 0: XEB = 1.691461 | HOG = 0.713135 | Entropy = 11.8609/16 | IPR = 3565.07  
Tile 1: XEB = 2.085808 | HOG = 0.733154 | Entropy = 11.8637/16 | IPR = 3565.07  
Tile 2: XEB = 1.303074 | HOG = 0.671875 | Entropy = 11.8986/16 | IPR = 3703.58  
Tile 3: XEB = 1.377725 | HOG = 0.687744 | Entropy = 11.8979/16 | IPR = 3705.22

Seed 1339 also produced a strong 4/4 tile first-pass result:

Tile 0: XEB = 1.120154 | HOG = 0.741455 | Entropy = 11.8780/16 | IPR = 3637.73  
Tile 1: XEB = 1.687559 | HOG = 0.791504 | Entropy = 11.8989/16 | IPR = 3708.49  
Tile 2: XEB = 1.452686 | HOG = 0.727051 | Entropy = 11.8894/16 | IPR = 3671.16  
Tile 3: XEB = 1.348320 | HOG = 0.745605 | Entropy = 11.8864/16 | IPR = 3651.98

Seeds 1340 and 1341 remained positive but weaker. Seed 1340, for example, produced:

Tile 0: XEB = 0.371807 | HOG = 0.569824  
Tile 1: XEB = 0.553700 | HOG = 0.605957  
Tile 2: XEB = 0.227333 | HOG = 0.554688  
Tile 3: XEB = 0.328323 | HOG = 0.576660

Seed 1341 showed the same general weak-positive behavior, with XEB values around 0.259–0.532 and HOG values around 0.567–0.602.

The strong seeds were then repeated. Seeds 1338 and 1339 were each run across five hardware executions, using the same four physical tile paths.

For seed 1338, the five-run repeatability package produced:

20/20 tile-level XEB values > 1.0  
20/20 tile-level HOG values > 0.67  
XEB range: approximately 1.302 to 2.207  
HOG range: approximately 0.672 to 0.741  
Entropy range: approximately 11.854 to 11.903  
IPR range: approximately 3541 to 3725

For seed 1339, the repeatability package produced:

20/20 tile-level results in the strong structural band  
XEB range: approximately 1.120 to 2.325  
HOG range: approximately 0.712 to 0.792  
Entropy range: approximately 11.874 to 11.908  
IPR range: approximately 3606 to 3738

The campaign-level result was therefore:

Seed 1337:

5 runs x 4 tiles = 20 tile results

Seed 1338:

5 runs x 4 tiles = 20 tile results

Seed 1339:

5 runs x 4 tiles = 20 tile results

Strong-seed total:

60 tile-level strong-seed Madmartigan results

Strong-seed hardware runs:

15

Active qubits per run:

64

Shots per run:

4096

QEC:

none

Post-selection:

none

Same selected physical tiles:

yes

The multi-seed campaign materially strengthens the benchmark. It shifts the narrative from:

One tuned seed produced an anomalous result.

to:

A family of Madmartigan seed instances preserved elevated XEB/HOG structural-output bands across the same calibrated physical tile map, while nearby seeds 1340 and 1341 remained positive but weaker.

That is a more credible claim than seed universality. It shows that the architecture is not a one-seed artifact, while also preserving the honest observation that seed choice affects structural-output strength.

The correct interpretation is:

Madmartigan is seed-robust across a strong adjacent seed family, but not seed-independent in the trivial sense. Seeds 1337, 1338, and 1339 formed a strong repeatable family; seeds 1340 and 1341 served as weak-positive comparator instances.

This result should be treated as the bridge between the 64-qubit calibration-aware replication and the later 96-qubit T6 scale-up. It demonstrates that the structural-output behavior was not locked to a single random instance before the program was expanded to six simultaneous tiles.

## 9 Seed-1339 Generic RCS Control

The seed-1339 result was especially important because it provided an alternate strong Madmartigan seed beyond the original seed 1337. To determine whether the seed-1339 result was also circuit-specific, a seed-matched generic RCS control was executed on the same four calibration-aware physical tiles.

The control used:

Backend: IBM Marrakesh

Selected tiles: 4

Active qubits: 64  
Measured qubits: 156  
Shots per run: 4096  
Runs: 5  
QEC: none  
Post-selection: none  
Reference Madmartigan seed: 1339  
Generic RCS seed: 1339

The transpiled control profile was:

Depth: 638  
Size: 4906  
CZ gates: 759  
SX gates: 1649  
RZ gates: 2342  
Measurements: 156  
Barriers: 40

This control had a substantial entangling burden and a full-chip measurement profile while preserving the same four-tile analysis structure. It was not a shallow or trivial comparator.

The control used two references:

Generic RCS own reference:

The ideal distribution of the generic RCS control itself.

Seed-1339 Madmartigan reference:

The ideal distribution of the seed-1339 Madmartigan circuit.

This allowed the experiment to distinguish between two questions:

1. Does the generic RCS circuit preserve its own structure?
2. Does the generic RCS circuit reproduce the seed-1339 Madmartigan structural band?

The answer to the first question was yes. The generic RCS control produced positive own-reference metrics.

The answer to the second question was no.

Across five runs and twenty tile-level control results, the generic RCS control remained near baseline against the seed-1339 Madmartigan reference:

mad1339RefXEB: approximately -0.102 to +0.100

mad1339RefHOG: approximately 0.475 to 0.507

A representative set of run-level results illustrates the behavior.

Run 1:

Tile 0: ownXEB = 0.982068 | ownHOG = 0.675537 | mad1339RefXEB = -0.008476 | mad1339RefHOG  
↪ = 0.502930

Tile 1: ownXEB = 0.796436 | ownHOG = 0.639404 | mad1339RefXEB = -0.060635 | mad1339RefHOG  
↪ = 0.490723

Tile 2: ownXEB = 0.436362 | ownHOG = 0.594971 | mad1339RefXEB = 0.073642 | mad1339RefHOG  
↪ = 0.497559

Tile 3: ownXEB = 0.181834 | ownHOG = 0.542236 | mad1339RefXEB = -0.008808 | mad1339RefHOG  
↪ = 0.502197

Run 4 showed even cleaner negative-reference behavior:

Tile 0: ownXEB = 0.967758 | ownHOG = 0.662109 | mad1339RefXEB = -0.076174 | mad1339RefHOG  
↪ = 0.493652

Tile 1: ownXEB = 0.833863 | ownHOG = 0.645264 | mad1339RefXEB = -0.078770 | mad1339RefHOG  
↪ = 0.483643

Tile 2: ownXEB = 0.349608 | ownHOG = 0.594971 | mad1339RefXEB = -0.102146 | mad1339RefHOG  
↪ = 0.474854

Tile 3: ownXEB = 0.265415 | ownHOG = 0.552490 | mad1339RefXEB = 0.037639 | mad1339RefHOG  
↪ = 0.495361

The control result is important because it generalizes the control-separation claim beyond the original seed-1337 benchmark. The seed-1337 result had already separated from generic RCS. The seed-1339 control showed that an alternate strong Madmartigan seed also separates from generic RCS when scored against its own Madmartigan reference.

The correct interpretation is:

Control separation is not limited to the original seed-1337 benchmark. A strong alternate seed also separates cleanly from a generic RCS control when tested against its own Madmartigan reference distribution.

This weakens two major objections:

1. The result was caused by one lucky seed.
2. The result was caused by good physical tiles rather than circuit-specific structure.

The seed-1339 generic RCS control used the same backend, same selected physical tiles, same active-qubit count, same shot count, same no-QEC/no-post-selection conditions, and comparable CZ burden. Yet it did not reproduce the seed-1339 Madmartigan band.

The concise result is:

Seed-1339 Madmartigan remained structurally elevated; seed-1339 generic RCS stayed baseline against the seed-1339 Madmartigan reference.

This should be included because it extends the control ladder from single-seed validation into multi-seed structural specificity.

## 10 First Viable Scale-6 Stress Layout at Depth 831

Before the optimized GLOBALPACK T6 rank-2 layout was selected, the first successful Scale-6 Madmartigan expansion was achieved using a deeper 96-active-qubit configuration. This result should be preserved in the benchmark narrative because it demonstrated depth resilience before the lower-depth optimized T6 layout was found.

The first viable Scale-6 layout preserved the original four fixed calibration-aware tiles and added two expansion tiles:

Tile 0:

[7, 8, 9, 10, 11, 18, 31, 32, 33, 34, 35, 19, 15, 14, 13, 12]

Tile 1:

[147, 148, 149, 150, 151, 138, 131, 132, 133, 134, 135, 139, 155, 154, 153, 152]

Tile 2:

[23, 22, 21, 36, 41, 42, 43, 44, 45, 46, 47, 57, 67, 66, 65, 56]

Tile 3:

[92, 91, 98, 111, 110, 109, 118, 129, 128, 127, 126, 125, 108, 107, 106, 97]

Tile 4:

[86, 87, 88, 89, 78, 69, 70, 71, 72, 73, 74, 75, 59, 55, 54, 53]

Tile 5:

[85, 84, 83, 96, 103, 102, 101, 116, 121, 122, 123, 136, 143, 142, 141, 140]

The configuration was:

Backend: IBM Marrakesh

Tiles: 6

Active qubits: 96

Measured qubits: 156

Seed: 1337

Shots: 4096

QEC: none

Post-selection: none

Full 156-qubit statevector claim: none

Per-tile reference: 16-qubit seed-1337 Madmartigan ideal distribution

The transpiled profile was:

Depth: 831

CZ: 1212

SX: 2595

RZ: 2493

This configuration is important because it tested the Madmartigan fixed-core structure under a much harsher circuit burden than the later optimized 495-depth T6 rank-2 layout. The question was not whether the circuit could survive in the best optimized embedding. The question was whether the original four-tile structural core could remain reference-aligned under a deep, high-CZ, 96-active-qubit NISQ execution.

The answer was yes.

A representative first hardware run produced:

Tile 0: XEB = 2.997705 | HOG = 0.778564 | Entropy = 11.8267/16 | IPR = 3430.92

Tile 1: XEB = 1.169695 | HOG = 0.675049 | Entropy = 11.8882/16 | IPR = 3667.95

Tile 2: XEB = 1.368898 | HOG = 0.676025 | Entropy = 11.8860/16 | IPR = 3659.95

Tile 3: XEB = 1.057950 | HOG = 0.661621 | Entropy = 11.9148/16 | IPR = 3765.08

Tile 4: XEB = 0.218251 | HOG = 0.577881 | Entropy = 11.8962/16 | IPR = 3695.42

Tile 5: XEB = 0.272104 | HOG = 0.560547 | Entropy = 11.8838/16 | IPR = 3658.35

The most important feature is the separation between the preserved fixed core and the expansion tiles. Tiles

0–3, corresponding to the original validated Madmartigan core, remained strongly positive against the seed-1337 Madmartigan reference. Tile 0 produced an especially high XEB of 2.997705 with HOG of 0.778564. Tiles 1–3 all remained above XEB 1.0 with HOG values between roughly 0.66 and 0.68.

Tiles 4 and 5 were weaker but still positive. They did not reproduce the strength of the original four-tile core, but they also did not destroy the core signal or collapse into negative reference alignment. Their weak-positive behavior is important because the purpose of the first Scale-6 layout was to apply additional backend and transpilation pressure while testing whether the original Madmartigan structure could survive.

The 831-depth layout was then repeated across additional runs. Across runs 1–5, the original four fixed tiles remained structurally dominant:

Tile 0 mean XEB  $\approx$  2.380

Tile 1 mean XEB  $\approx$  1.352

Tile 2 mean XEB  $\approx$  1.170

Tile 3 mean XEB  $\approx$  1.059

Tile 4 mean XEB  $\approx$  0.296

Tile 5 mean XEB  $\approx$  0.199

Fixed-core mean XEB  $\approx$  1.490

Expansion-tile mean XEB  $\approx$  0.248

All six tiles remained positive across all five runs. The original four fixed tiles remained the dominant structural channels. The expansion tiles remained persistently weak-positive rather than strong.

This result demonstrates depth resilience of the original four-tile Madmartigan core under 96-active-qubit / 1212-CZ scale pressure. That is a serious result because an uncorrected NISQ circuit at depth 831 with more than 1200 CZ gates would ordinarily be expected to suffer major structural degradation. Instead, the original reference-aligned tile core repeatedly remained above baseline.

The correct interpretation is:

The first viable Scale-6 layout demonstrated that the original four-tile Madmartigan core could survive deep 96-active-qubit execution pressure. The later 495-depth GLOBALPACK T6 rank-2 layout should therefore be understood as an optimized successor, not the only successful scale-up.

This establishes the developmental sequence:

Original 16-qubit Madmartigan benchmark

-> calibration-aware four-tile / 64-active-qubit replication

-> multi-seed 64-active-qubit robustness

-> first viable 96-active-qubit Scale-6 stress layout at depth 831

-> optimized GLOBALPACK T6 rank-2 layout at depth 495

-> controlled 96-active-qubit evidence package

The 831-depth run is therefore a major part of the benchmark narrative. It shows that Madmartigan’s structural preservation was not fragile, not dependent only on the later optimized rank-2 layout, and not limited to shallow or lightly burdened executions.



IBM Marrakesh Physical Map: 96q Scale-6 Depth-831 XEB Separation, Madmartigan minus Generic RCS Mad-Reference

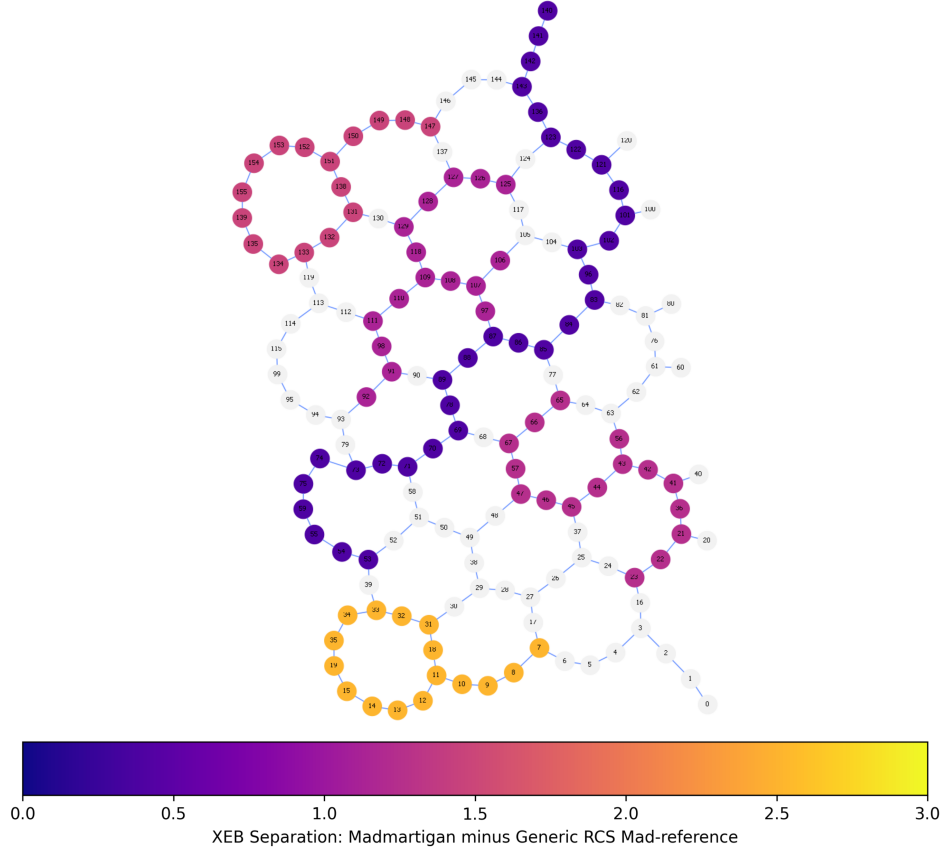


Figure 2: IBM Marrakesh physical tile map for the first viable 96-active-qubit Scale-6 stress layout at depth 831. Each qubit in a selected tile is colored by that tile’s XEB separation value, defined as raw Madmartigan-reference XEB minus same-layout generic RCS Madmartigan-reference XEB. The map shows that the original four-tile core remained strongly separated from control even under the deeper 831-depth execution burden. Because XEB is a tile-level metric, each qubit in a tile is colored by the corresponding tile-level XEB separation value.

## 11 Raw T6 Rank-2 Benchmark Configuration

The GLOBALPACK T6 rank-2 raw benchmark was configured as follows:

Backend: IBM Marrakesh  
Total measured qubits: 156  
Active qubits: 96  
Tiles: 6  
Tile size: 16  
Seed: 1337  
Shots: 4096  
QEC: none  
Post-selection: none

Full 156-qubit statevector claim: none

Per-tile ideal reference: 16-qubit seed-1337 Madmartigan distribution

The transpiled profile was:

Depth: 495

Size: 6376

SX: 2546

RZ: 2489

CZ: 1185

Measure: 156

Barrier: 66

This matters because the benchmark was not shallow, not simulated-only, and not post-selected. It was a real 96-active-qubit circuit on IBM hardware with more than one thousand CZ gates.

## 12 Raw T6 Rank-2 Hardware Results

The raw T6 rank-2 results showed persistent reference-specific structure across repeated runs. The most important pattern was that Tiles 0–4 repeatedly produced positive, often elevated, XEB and HOG against the original Madmartigan seed-1337 reference. Tile 5 was weaker but generally remained positive.

A representative Run 5 showed:

Tile 0: XEB = 1.990524 | HOG = 0.727539 | Entropy = 11.8663/16 | IPR = 3595.63

Tile 1: XEB = 1.417889 | HOG = 0.687988 | Entropy = 11.8909/16 | IPR = 3680.83

Tile 2: XEB = 1.111165 | HOG = 0.675781 | Entropy = 11.8830/16 | IPR = 3651.98

Tile 3: XEB = 1.270121 | HOG = 0.669678 | Entropy = 11.8967/16 | IPR = 3698.68

Tile 4: XEB = 1.358467 | HOG = 0.682373 | Entropy = 11.8952/16 | IPR = 3700.31

Tile 5: XEB = 0.265650 | HOG = 0.580566 | Entropy = 11.8933/16 | IPR = 3688.92

The strongest interpretation is not that every tile performed equally. The correct interpretation is that a 96-active-qubit uncorrected hardware execution preserved a multi-tile Madmartigan reference band, with the original core and most of the expanded tile set remaining positive and structured under real hardware pressure.

The result is especially meaningful because the circuit used only 4096 shots. With a 16-bit tile space of 65,536 possible outputs, 4096 shots produce a sparse sample. Sustained XEB/HOG elevation under sparse sampling suggests nontrivial reference alignment rather than simple overcounting of a few repeated states.

## 13 Two-Layout Interpretation of the 96-Qubit Scale-6 Result

The 96-active-qubit Scale-6 campaign should not be understood as a single layout result. It produced two distinct but complementary 6-tile findings.

The first was the 831-depth stress layout:

Layout A: First viable Scale-6 stress layout

Depth: 831

CZ: 1212

Purpose: harsh depth/routing stress test

Primary result: original four fixed tiles remained structurally elevated

The second was the optimized GLOBALPACK T6 rank-2 layout:

Layout B: GLOBALPACK T6 rank-2

Depth: 495

CZ: 1185

Purpose: optimized lower-depth six-tile benchmark substrate

Primary result: structural band distributed across five of six tiles repeatably

The 831-depth layout showed that the original four-tile Madmartigan core could survive deep execution pressure. The GLOBALPACK T6 rank-2 layout then improved the physical embedding, reduced the transpiled depth from 831 to 495, and brought the fifth tile into the stronger structural band.

This is the correct developmental interpretation:

The 831-depth result established depth-resilient core preservation. The 495-depth GLOBALPACK T6 rank-2 result established the stronger controlled evidence package.

The optimized GLOBALPACK T6 rank-2 result showed the following aggregate behavior across runs 1–5:

Tile 0 mean XEB  $\approx 1.943$

Tile 1 mean XEB  $\approx 1.444$

Tile 2 mean XEB  $\approx 1.009$

Tile 3 mean XEB  $\approx 1.266$

Tile 4 mean XEB  $\approx 1.294$

Tile 5 mean XEB  $\approx 0.362$

Tile 0 mean HOG  $\approx 0.721$

Tile 1 mean HOG  $\approx 0.696$

Tile 2 mean HOG  $\approx 0.666$

Tile 3 mean HOG  $\approx 0.669$

Tile 4 mean HOG  $\approx 0.678$

Tile 5 mean HOG  $\approx 0.586$

The aggregate interpretation was:

5/6 tiles repeatedly formed a strong or near-strong structural band.

6/6 tiles remained positive across all runs.

Tile 5 remained the weakest channel but did not collapse negative.

Mean all-tile XEB  $\approx 1.220$ .

Mean all-tile HOG  $\approx 0.669$ .

This is why GLOBALPACK T6 rank-2 became the flagship 96-active-qubit evidence package. It did not merely preserve the original four fixed tiles; it extended meaningful reference-aligned structure across five of the six simultaneous tile channels.

The comparison can be summarized as follows:

831-depth Scale-6 layout:

Strength:

Shows original four-tile core survives harsh depth pressure.

Weakness:

Expansion tiles remain weak-positive.

Best interpretation:

Depth-resilient core preservation.

495-depth GLOBALPACK T6 rank-2:

Strength:

Distributes the structural band across five of six tile channels.

Weakness:

Tile 5 remains weak.

Best interpretation:

Optimized 96-active-qubit evidence package.

The two-layout result is stronger than either layout alone. The 831-depth run prevents the benchmark from looking dependent on a carefully optimized low-depth mapping. The 495-depth run shows that optimization improves structural distribution without abandoning the original fixed core.

The final interpretation should therefore be:

The Scale-6 campaign produced both stress resilience and optimized replication. The first viable 831-depth layout demonstrated that the original four-tile core could survive deep 96-active-qubit pressure. The later 495-depth GLOBALPACK T6 rank-2 layout converted that survival into the strongest controlled 96-active-qubit evidence package, with five of six tile channels repeatedly preserving meaningful Madmartigan-reference structure.

## 14 Interpretation of Raw T6 Rank-2

The raw T6 rank-2 result should be interpreted as a controlled scale-up of the 16-qubit Madmartigan benchmark into a six-tile 96-active-qubit configuration.

The key claims supported by the raw run are:

1. The seed-1337 Madmartigan tile distribution can be replicated across multiple physical tile paths.
2. The signal survives real hardware transpilation and execution at depth 495 and 1185 CZ gates.
3. The original four fixed tiles remain structurally elevated under simultaneous multi-tile pressure.
4. Additional tiles can be added, though with varying quality due to topology/calibration constraints.
5. The observed output remains in a high-dimensional NISQ distributional regime, not a trivial low-entropy collapse.

However, raw runs alone are not sufficient for a mature benchmark claim. Controls were needed to test whether the signal could be explained by tile quality, generic RCS behavior, phase-insensitive structure, or hardware bias.

## 15 Control Ladder Overview

The completed 96-qubit evidence package contains three major control classes:

1. Same-layout generic RCS control

2. Phase-scrambled Madmartigan-adjacent control
3. Partial-entanglement ablation control

Each control uses dual-reference scoring where applicable:

Own-reference scoring:

Hardware output compared to the control circuit's own ideal reference.

Madmartigan-reference scoring:

The same hardware output compared to the original seed-1337 Madmartigan reference.

This dual-reference structure is critical. It prevents the controls from being dismissed as dead circuits. A useful control may preserve its own structure while failing to reproduce the Madmartigan-specific structure.

## 16 Generic RCS Control

### 16.1 Configuration

The generic RCS control used the same six physical tile paths as the raw T6 rank-2 Madmartigan run, the same backend, the same 96 active qubits, the same 156 measured qubits, the same 4096 shots, no QEC, and no post-selection.

The control body used ten alternating random single-qubit and nearest-neighbor entangling layers, with deterministic RZ padding on layers 1, 3, 5, and 7.

The transpiled profile was:

```
Depth: 581
Size: 6615
RZ: 3249
SX: 2151
CZ: 1059
Measure: 156
Barrier: 60
```

### 16.2 Results

The generic RCS control was not dead. It showed positive own-reference structure across all tiles and runs. However, when the same hardware outputs were scored against the Madmartigan seed-1337 reference, XEB went negative or near baseline and HOG remained near 0.5.

Representative Run 5:

```
Tile 0: ownXEB = 1.692178 | ownHOG = 0.705322 | mad1337RefXEB = -0.099475 | mad1337RefHOG
↪ = 0.479492
Tile 1: ownXEB = 1.347122 | ownHOG = 0.691162 | mad1337RefXEB = -0.140105 | mad1337RefHOG
↪ = 0.483887
Tile 2: ownXEB = 0.667600 | ownHOG = 0.641602 | mad1337RefXEB = -0.062471 | mad1337RefHOG
↪ = 0.489990
Tile 3: ownXEB = 0.684727 | ownHOG = 0.604004 | mad1337RefXEB = -0.037480 | mad1337RefHOG
↪ = 0.482910
```

Tile 4: ownXEB = 1.251121 | ownHOG = 0.674316 | mad1337RefXEB = -0.099479 | mad1337RefHOG  
 $\hookrightarrow$  = 0.476562  
 Tile 5: ownXEB = 0.783225 | ownHOG = 0.639648 | mad1337RefXEB = -0.094468 | mad1337RefHOG  
 $\hookrightarrow$  = 0.481445

### 16.3 Interpretation

The generic RCS control answers the question:

Does same-layout generic random circuit behavior reproduce the Madmartigan reference band?

The answer is no.

The control proved that the same physical tile paths and backend can preserve a generic RCS circuit’s own structure, but that structure does not align with the seed-1337 Madmartigan reference. This rules out a major class of trivial explanations: the Madmartigan result is not merely caused by tile quality, backend bias, or any random circuit on the same physical paths.

IBM Marrakesh Physical Map: 96q T6 Rank-2 XEB Separation, Madmartigan minus Generic RCS Mad-Reference

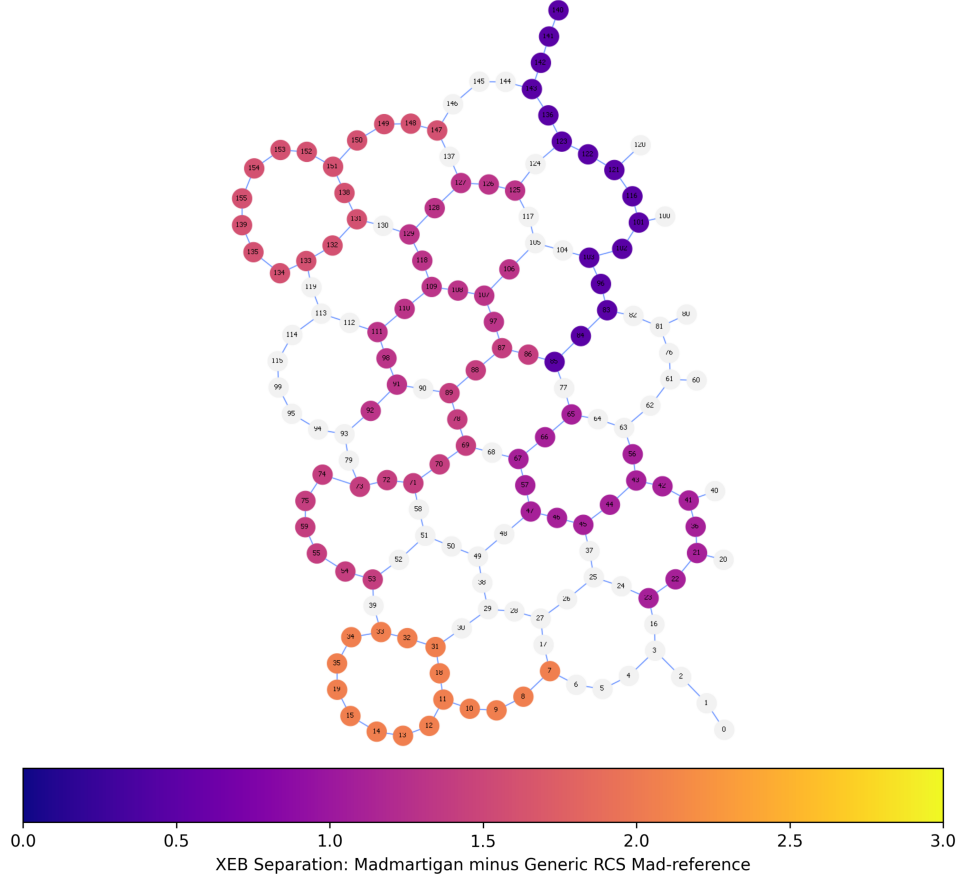


Figure 3: IBM Marrakesh physical tile map for the 96-active-qubit GLOBALPACK T6 rank-2 benchmark. Each qubit in a selected tile is colored by that tile’s XEB separation value, defined as raw Madmartigan-reference XEB minus same-layout generic RCS Madmartigan-reference XEB. Warmer colors indicate stronger positive reference-specific separation in favor of Madmartigan. Because XEB is a tile-level metric, each qubit in a tile is colored by the corresponding tile-level XEB separation value

## 17 Phase-Scrambled Control

### 17.1 Configuration

The phase-scrambled control was designed as an architecture-adjacent control. Unlike generic RCS, it remained closer to the Madmartigan family but scrambled the phase-command structure. It used the same six T6 rank-2 tile paths, the same backend, 96 active qubits, 156 measured qubits, 4096 shots, no QEC, and no post-selection.

The transpiled profile was:

Depth: 661  
Size: 6754  
SX: 2787  
RZ: 2551  
CZ: 1260  
Measure: 156  
Barrier: 66

This control was heavier than the raw T6 rank-2 benchmark in both depth and CZ count. That is important because it prevents the easy criticism that the control failed simply because it was too shallow or too lightly loaded.

### 17.2 Results

The phase-scrambled control preserved measurable own-reference structure, but it did not reproduce the Madmartigan reference band.

Representative Run 1:

Tile 0: ownXEB = 1.387368 | ownHOG = 0.710449 | mad1337RefXEB = 0.002663 | mad1337RefHOG  
↪ = 0.508301  
Tile 1: ownXEB = 0.924559 | ownHOG = 0.645996 | mad1337RefXEB  $\approx$  0.000011 | mad1337RefHOG  
↪ near baseline  
Tile 3: ownXEB = 0.406329 | ownHOG = 0.568359 | mad1337RefXEB = -0.056523 | mad1337RefHOG  
↪ = 0.491699  
Tile 4: ownXEB = 0.558252 | ownHOG = 0.638916 | mad1337RefXEB = -0.000456 | mad1337RefHOG  
↪ = 0.511963  
Tile 5: ownXEB = 0.803081 | ownHOG = 0.642334 | mad1337RefXEB = -0.091253 | mad1337RefHOG  
↪ = 0.472168

Across repeated runs, Madmartigan-reference XEB stayed near zero or modestly positive/negative, far below raw Madmartigan levels, while Madmartigan-reference HOG remained near baseline.

### 17.3 Interpretation

The phase-scrambled control answers the question:

Does a Madmartigan-adjacent circuit with comparable or heavier depth/CZ burden reproduce the Madmartigan band if the phase-command structure is scrambled?

The answer is no.

This is stronger than generic RCS because it tests an architecture-adjacent perturbation rather than an unrelated random-circuit body. The result supports the claim that the raw Madmartigan signal depends on the specific phase/correlation program, not merely on depth, CZ burden, or physical tile layout.

## 18 Partial-Entanglement Ablation Control

### 18.1 Initial Motivation

The first entanglement ablation was too destructive. It removed the entangling backbone so completely that the circuit transpiled to a shallow CZ-free profile. That was useful as a destructive ablation, but not a serious architecture-adjacent control.

A stronger partial-entanglement ablation was then built. It preserved broad Madmartigan phase sections and substantial entangling burden while breaking the exact EBA chain, QMCA collapse spine, SQCA reciprocal lattice, QPSA fanout targets, and BSCM parity/clamp alignment.

### 18.2 Configuration

The partial-entanglement ablation used the same six T6 rank-2 tile paths, 96 active qubits, 156 measured qubits, 4096 shots, no QEC, and no post-selection.

The transpiled profile was:

```
Depth: 634
Size: 7650
SX: 3620
RZ: 2101
CZ: 1773
Measure: 156
Barrier: 66
```

This control was significantly heavier than raw T6 rank-2 in CZ count.

### 18.3 Ideal Reference Difference

The partial-entanglement ablation’s own ideal top probabilities were higher than the original Madmartigan top probabilities. The partial-ablation top states were around 0.00156–0.00218, while the Madmartigan seed-1337 top states were around 0.00080–0.00134.

This matters because XEB rewards samples that land on high-probability ideal states. A more peaked own-reference distribution can generate higher ownXEB if preserved well by hardware.

### 18.4 Results

The partial-entanglement ablation produced extremely high own-reference XEB/HOG. Representative Run 3:

```
Tile 0: ownXEB = 6.940017 | ownHOG = 0.880859 | mad1337RefXEB = 0.209806 | mad1337RefHOG
↪ = 0.561768 | Entropy = 11.7257/16 | IPR = 3127.74
Tile 1: ownXEB = 4.372260 | ownHOG = 0.842285 | mad1337RefXEB = 0.242883 | mad1337RefHOG
↪ = 0.570801 | Entropy = 11.8019/16 | IPR = 3366.22
```



Tile 2: ownXEB = 4.897565 | ownHOG = 0.828857 | mad1337RefXEB = 0.243306 | mad1337RefHOG  
 $\hookrightarrow$  = 0.579590 | Entropy = 11.8064/16 | IPR = 3383.87  
Tile 3: ownXEB = 0.893227 | ownHOG = 0.652832 | mad1337RefXEB = 0.097842 | mad1337RefHOG  
 $\hookrightarrow$  = 0.533447 | Entropy = 11.8508/16 | IPR = 3538.00  
Tile 4: ownXEB = 3.324858 | ownHOG = 0.803223 | mad1337RefXEB = 0.387527 | mad1337RefHOG  
 $\hookrightarrow$  = 0.587646 | Entropy = 11.8449/16 | IPR = 3518.71  
Tile 5: ownXEB = 1.916966 | ownHOG = 0.726562 | mad1337RefXEB = 0.192001 | mad1337RefHOG  
 $\hookrightarrow$  = 0.535645 | Entropy = 11.8813/16 | IPR = 3642.47

## 18.5 Interpretation

This was not a clean negative control in the same sense as generic RCS or phase-scrambled control. It did not collapse entirely to baseline against the Madmartigan reference. Instead, it produced a new structured-output regime.

The central discovery is:

Breaking the exact Madmartigan entanglement pathways does not destroy all QSCE-family structure. It produces a derivative, high-attractor operating mode that is strongly preserved against its own reference while only weakly-to-moderately overlapping the original Madmartigan reference.

This reveals two distinguishable QSCE-related regimes:

Raw Madmartigan:

Reference-specific command preservation.

Partial-entanglement ablation:

High-gain QSCE-derived attractor mode with different reference identity.

The ablation’s high ownXEB/HOG does not mean it “beats Madmartigan at being Madmartigan.” It means it preserves its own more concentrated ablated reference distribution extremely well. The lower Madmartigan-reference XEB/HOG proves that it is not the same attractor identity as the original Madmartigan band.

This is a critical distinction: entropy/IPR show that the ablation lives in a similar broad high-dimensional attractor regime, while cross-reference XEB/HOG show that the attractor identity is different.

## 19 Comparative Summary of Raw and Control Results

### 19.1 Raw T6 Rank-2 Madmartigan

- 96 active qubits
- Depth 495
- 1185 CZ gates
- 4096 shots
- No QEC
- No post-selection
- Preserved positive Madmartigan-reference structure across the tile set
- Strongest evidence for reference-specific command-band preservation

## 19.2 Generic RCS Control

- Same six physical tile paths
- 96 active qubits
- Depth 581
- 1059 CZ gates
- Positive own-reference structure
- Negative/near-baseline Madmartigan-reference XEB
- Baseline-like Madmartigan-reference HOG
- Rules out generic same-layout RCS as the explanation

## 19.3 Phase-Scrambled Control

- Same six physical tile paths
- 96 active qubits
- Depth 661
- 1260 CZ gates
- Heavier than raw T6 rank-2
- Positive own-reference structure
- Near-baseline Madmartigan-reference behavior
- Rules out depth/CZ burden and phase-insensitive architecture-adjacent explanations

## 19.4 Partial-Entanglement Ablation

- Same six physical tile paths
- 96 active qubits
- Depth 634
- 1773 CZ gates
- Very high own-reference XEB/HOG
- Weak-to-moderate Madmartigan-reference overlap
- Reveals a QSCE-derived high-attractor mode rather than a clean negative control

## 20 Why the Evidence Package Is Scientifically Strong at 96 Active Qubits

The 96-qubit package is strong because it is not merely a raw result. It includes:

1. Raw positive benchmark
2. Repeatability

3. Same-layout generic control
4. Architecture-adjacent phase control
5. Entanglement ablation / attractor discovery
6. Consistent no-QEC/no-post-selection methodology
7. Sparse 4096-shot execution
8. Real backend execution
9. Per-tile ideal reference metrics
10. Clear distinction between own-reference and Madmartigan-reference scoring

A reviewer can still ask for independent reproduction, backend replication, and formal statistical significance testing. But as an internal or preliminary external evidence package, this is now mature. It demonstrates not only that the raw signal exists, but that it behaves differently from controls in a reference-dependent way.

## 21 Attempted Scaling Beyond 96 Active Qubits

After the 96-qubit evidence package was established, the next objective was to scale to 144 active qubits using nine 16-qubit tiles.

The first approach was to preserve the full T6 rank-2 six-tile core and search for three additional clean disjoint 16-qubit tiles. This failed across Marrakesh, Fez, and Kingston. The scans found no viable T7, T8, or T9 clean extensions when the entire fixed-six core was preserved.

The second approach was to preserve only the original four validated tiles and search for five additional clean 16-qubit tiles to reach T9 / 144 active qubits. This also failed across Marrakesh, Fez, and Kingston.

The third approach was to preserve the original four and search for four additional clean 16-qubit tiles to reach T8 / 128 active qubits. This also failed across Marrakesh, Fez, and Kingston.

The repeated failure pattern indicates a topology-packing wall, not a Madmartigan signal failure. The Heron r2 156-qubit heavy-hex topology does not appear to support enough mutually disjoint clean 16-qubit chain tiles under the preserved-core constraints.

This matters because the conclusion is not:

Madmartigan failed at 128 or 144 active qubits.

The more accurate conclusion is:

The clean independent 16-qubit-chain replication model saturated at T6 / 96 active qubits under the current preserved-core constraints and available backend topology.

## 22 Reframing the 144-Qubit Experiment as Backend Pressure

Because clean T8/T9 packing failed, the 144-qubit objective was reframed correctly.

Instead of requiring nine clean independent 16-qubit tiles, the new experiment preserves the proven 96-qubit T6 rank-2 core exactly and adds three best-effort stress tiles from the remaining backend qubits.

The purpose becomes:

Test whether the original proven T6 core survives while the rest of the backend is loaded with additional Madmartigan-style stress tiles.

This is not a clean independent 9-tile benchmark. It is a valid 144-active-qubit backend-pressure experiment.

The primary win condition is:

1. Original core Tiles 0–3 remain elevated.
2. T6 extra core Tiles 4–5 remain meaningfully positive if possible.
3. Stress Tiles 6–8 are allowed to degrade.
4. The circuit executes at 144 active qubits without QEC or post-selection.

This reframing is scientifically honest and strategically useful. It preserves the controlled 96-qubit evidence package and tests its resilience under additional full-backend load.

## 23 Research Conclusions

### 23.1 Main Result

The original 16-qubit Madmartigan benchmark successfully scaled into a controlled 96-active-qubit evidence package on real IBM hardware.

The raw T6 rank-2 benchmark showed that the seed-1337 Madmartigan reference band could be preserved across six simultaneous 16-qubit tiles, with five of six tiles repeatedly producing positive and often elevated Madmartigan-reference XEB/HOG.

### 23.2 Control Result

Same-layout generic RCS did not reproduce the Madmartigan reference band. Phase-scrambled Madmartigan-adjacent control did not reproduce the Madmartigan reference band. These controls strongly support the interpretation that the raw Madmartigan result is not explained by generic random-circuit behavior, tile quality alone, or phase-insensitive circuit burden.

### 23.3 Ablation Discovery

The partial-entanglement ablation revealed a second QSCE-derived operating regime. It generated extremely high own-reference XEB/HOG while retaining only weak-to-moderate Madmartigan-reference overlap. This suggests the QSCE/Madmartigan design family contains a tunable attractor manifold: one mode optimized for command/reference preservation and another potentially useful as a high-gain activation kernel.

### 23.4 Scale-Up Boundary

Attempts to cleanly scale beyond 96 active qubits using independent 16-qubit chain tiles ran into a topology-packing wall across Marrakesh, Fez, and Kingston. This does not invalidate the 96-qubit result. It shows that the next stage of scaling requires either backend-pressure tests, flexible non-chain tile geometries, smaller tile sizes, or topology-native full-chip circuit design.

## 24 Significance

The significance of this work is that it moves Madmartigan from a single 16-qubit benchmark into a controlled, multi-tile, 96-active-qubit research package. The evidence now includes raw structural preservation, repeatability, same-layout generic controls, phase-scrambled architecture-adjacent controls, and ablation-based discovery.

For QSCE, this is important because it supports two claims simultaneously:

1. The Madmartigan program can preserve a specific intended reference band under NISQ hardware conditions.
2. The broader QSCE design space can generate multiple structured-output regimes, including a high-attractor derivative mode.

For quantum communications, signaling, and cyber-hardening, the relevance is that structured, reference-specific quantum output may be usable as a command-bearing or authentication-bearing primitive. The IPCM prototype previously demonstrated a quantum-to-classical handoff pathway. The Madmartigan scale-up strengthens the substrate side of that story: not merely that a quantum result can be decoded and transmitted, but that structured output can survive multi-tile NISQ execution with controls showing reference specificity.

## 25 National Security and Dual-Use Implications

### 25.1 Strategic Relevance

The 96-active-qubit Madmartigan evidence package has direct implications for national-security research because the demonstrated behavior is not limited to abstract quantum benchmarking. The work points toward a quantum-native signaling and command-preservation primitive: a structured output band that survives real NISQ hardware execution, remains distinguishable from generic RCS behavior, and can be evaluated against a predefined reference distribution.

In defense and intelligence contexts, the relevant question is not only whether a quantum processor can solve a conventional computational problem. The more operationally urgent question is whether quantum hardware can generate, preserve, classify, and hand off structured signals in ways that remain useful under noise, imperfect calibration, limited shots, backend variability, and contested operating conditions.

The Madmartigan scale-up results bear directly on that question. The raw T6 rank-2 benchmark demonstrated reference-specific structure at 96 active qubits without QEC, without post-selection, and with only 4096 shots. The generic RCS and phase-scrambled controls showed that this structure is not trivially reproduced by same-layout random behavior or by phase-disrupted architecture-adjacent circuits. The partial-entanglement ablation then revealed a high-attractor QSCE-derived mode, suggesting the design family may contain both command-preserving and robust activation-oriented regimes.

Taken together, these results suggest a potential class of quantum-native signal engines whose outputs are not treated as ordinary random samples but as structured, reference-dependent state collapses that can be decoded, classified, and routed into classical systems.

### 25.2 Quantum Communications Implications

The most immediate communications implication is that Madmartigan-style structured-output preservation may support quantum-assisted command signaling. Unlike conventional quantum communication models

that focus primarily on key distribution, entanglement distribution, or long-distance qubit transmission, this work focuses on preserving a structured quantum output distribution that can be decoded into classical meaning.

That distinction matters. In many operational environments, the practical bottleneck is not only secure transmission. It is trustworthy signal generation, authentication, classification, and downstream action under uncertainty. A quantum system that repeatedly produces a reference-specific structured output band could function as a signal primitive whose validity is determined by whether the observed collapse pattern remains aligned with the expected reference.

In this framing, the Madmartigan circuit is not merely a benchmark. It is a candidate quantum-native signal source. Its output can be evaluated for:

reference alignment,  
heavy-output concentration,  
entropy/IPR operating regime,  
threshold crossing,  
state-band classification,  
and command-routing suitability.

This could support future communication systems in which quantum hardware contributes a nonclassical signaling layer, not by transmitting qubits end-to-end, but by generating structured state-dependent events that are difficult to spoof without knowing the circuit, seed, reference distribution, tile mapping, and expected statistical band.

The IPCM prototype already demonstrated a quantum-to-classical handoff pathway. The 96-qubit Madmartigan package strengthens the substrate behind that handoff by showing that structured quantum output can scale beyond a single 16-qubit tile into a multi-tile hardware execution while retaining reference-specific behavior.

### 25.3 Signaling in Contested Environments

Contested environments impose conditions that resemble, in practical terms, many of the constraints present in NISQ hardware: noise, interference, incomplete information, degraded channels, uncertain timing, adversarial observation, and limited trust in data provenance.

The Madmartigan evidence package is relevant because it demonstrates a signal structure that is not dependent on perfect hardware conditions. The circuit operated under real backend noise, calibration variability, sparse sampling, nontrivial transpilation depth, and no error correction. Yet the raw benchmark preserved a measurable reference-specific output band, while controls failed to recreate that band.

For contested signaling, that suggests a possible architecture in which the validity of a signal is not based solely on a static key, fixed waveform, or deterministic classical packet. Instead, validity could be tied to whether an observed quantum-generated output lies inside an expected reference band. A receiver or downstream system would not merely ask, “Did a bit arrive?” It could ask:

Did the measured state band align with the authorized reference?  
Did the entropy/IPR profile remain inside the accepted operating window?  
Did HOG/XEB-style measures indicate the correct structured source?  
Did the output cross a mission-defined activation threshold?  
Was the signal consistent with the expected quantum-to-classical command map?

This creates a potential pathway for resilient signaling in which authenticity and command validity derive from structured collapse behavior rather than from ordinary packet identity alone.

## 25.4 Cyber-Hardening Implications

Cyber-hardening is fundamentally about preserving trust, integrity, authentication, and resilience under adversarial pressure. The Madmartigan scale-up results are relevant because they suggest a way to generate command-bearing state signatures that are tied to a quantum circuit’s reference distribution.

A classical adversary can copy packets, replay messages, spoof headers, manipulate logs, or inject forged traffic. A QSCE/Madmartigan-style system would not eliminate those threats by itself, but it could add a quantum-native validation layer. The classical system could require a command event to be accompanied by a valid quantum-derived collapse signature. The signature would not merely be a static token; it would be a statistical state-band outcome tied to a specific circuit, backend configuration, seed, and expected reference behavior.

This could support cyber-hardening in several ways:

1. **Command authentication**

A command is accepted only if the quantum output falls inside the authorized reference band.

2. **Anti-spoofing**

Spoofed classical traffic would lack the correct quantum-derived state signature.

3. **Tamper indication**

Phase disruption, pathway alteration, or random replacement should move outputs away from the expected reference band, as demonstrated by the controls.

4. **Resilient fallback signaling**

If ordinary channels are degraded, a quantum-generated command trigger could still provide a high-confidence activation cue.

5. **Hardware-bound identity**

Signal validity could be bound to circuit behavior on a specific quantum backend or calibrated device class.

The control results are especially important for cyber-hardening. Generic RCS did not reproduce the Madmartigan band. Phase scrambling did not reproduce the Madmartigan band. Partial entanglement ablation produced a different attractor identity. This means the reference band is not trivially interchangeable with arbitrary structured output. That property is exactly what a cyber-hardening layer would need: sensitivity to the correct source structure and failure under incorrect or tampered structure.

## 25.5 Dual-Use Risk and Opportunity

The same features that make QSCE/Madmartigan relevant to secure signaling also make it dual-use. A quantum-native structured-output engine could support defensive authentication, resilient command signaling, and cyber-hardening. But any technology that improves command signaling, covert activation, or system resilience can also be misused if applied to unauthorized control, evasion, or adversarial infrastructure.

Therefore, the responsible framing is not that these results should be openly operationalized without constraint. The proper national-security framing is:

The work should be evaluated under controlled disclosure,

mission-specific requirements,  
appropriate export-control awareness,  
independent technical review,  
and government or trusted-lab validation.

The results are not yet a fielded system. They are a controlled hardware evidence package and an emerging utility substrate. That distinction matters. The immediate implication is not deployment; it is portfolio awareness, protected evaluation, and mission-relevant prototyping.

## 25.6 Relevance to DoD Portfolio Awareness

For a DoD or national-security portfolio, the importance of this work is that it challenges the assumption that useful quantum output must wait for fully fault-tolerant quantum computation. The Madmartigan package suggests that uncorrected NISQ hardware may already support structured signaling primitives if the objective is not universal computation but reference-specific output preservation and quantum-to-classical activation.

This is particularly relevant to portfolios concerned with:

resilient communications,  
command-and-control hardening,  
cyber defense,  
mission authentication,  
anti-spoofing,  
secure signaling,  
quantum-assisted sensing-to-command loops,  
and contested-environment autonomy.

The 96-qubit evidence package gives portfolio reviewers a more mature basis for assessment than the original 16-qubit benchmark alone. It shows scale-up, repeatability, control separation, and ablation behavior. It also identifies the current scaling boundary as a topology-packing issue rather than a demonstrated collapse of the signal itself.

## 25.7 Operational Hypothesis

The operational hypothesis emerging from this work is:

A QSCE/Madmartigan-style circuit can function as a quantum-native structured signal generator whose measured output distribution can be classified against an authorized reference band and handed off to classical systems for authentication, routing, or activation.

If validated independently, this could support a new class of quantum-assisted cyber-hardening tools where quantum hardware does not replace classical networks but supplies a high-integrity signal primitive inside them.

## 25.8 Recommended National-Security Next Steps

The appropriate next steps for national-security evaluation are:

1. Independent reproduction of the 96-qubit T6 rank-2 package.
2. Controlled rerun of raw, generic RCS, phase-scrambled, and partial-ablation circuits on a trusted bac
3. Formal definition of command-band acceptance thresholds.



4. IPCM-style quantum-to-classical handoff using the 96-qubit reference band.
5. Red-team testing against spoofing, phase scrambling, seed mismatch, and circuit perturbation.
6. Evaluation under degraded backend conditions and calibration drift.
7. Mission-specific prototype design for secure command signaling or cyber-hardening.
8. Export-control and dual-use review before broader disclosure or commercialization.

The key evaluation question should be:

Can a quantum-generated reference band be used as a reliable, hard-to-spoof command-authentication or signaling primitive in a contested environment?

The current results do not fully answer that question, but they justify asking it seriously.

## 25.9 National-Security Conclusion

The Madmartigan 16-to-96-qubit scale-up materially strengthens the case that QSCE may have national-security relevance beyond conventional quantum computing benchmarks. The work suggests that structured quantum output can be preserved, classified, and differentiated from controls on real NISQ hardware. That capability has direct implications for quantum communications, signaling, and cyber-hardening because it points toward quantum-derived command signatures that may be useful in contested environments.

The most defensible conclusion is:

The 96-active-qubit Madmartigan package should be treated as a controlled evidence base for evaluating quantum-native signaling and cyber-hardening primitives, not as a finished operational system. It warrants protected portfolio awareness, independent validation, and mission-specific prototyping under appropriate dual-use safeguards.

## 26 Data Availability and Reproducibility Package

A central requirement for evaluating the Madmartigan scale-up results is traceability from each reported metric back to the underlying circuit artifact, hardware execution record, raw counts, tile projection, and analysis output. For this reason, the experimental package is organized as a reproducibility bundle rather than as a narrative-only benchmark summary.

The data package for this work includes the following artifact classes:

### 1. Serialized quantum circuits

The executed circuits are preserved in IBM/Qiskit-compatible `.qpy` format. These files preserve the circuit structure used for hardware execution and allow an external evaluator to reload the same circuit object for inspection, transpilation comparison, or backend-specific rerun.

### 2. QASM3 circuit exports

QASM3 exports are preserved for human-readable inspection of the circuit bodies, gate structure, measurement layout, and tile-level construction. These exports provide an additional format for review independent of the serialized QPY files.

### 3. Raw hardware counts

Raw count dictionaries are preserved for each reported hardware run. These files contain the measured

bitstring counts returned from the backend before any tile-level projection or metric aggregation. They provide the direct empirical basis for the reported XEB, HOG, entropy, IPR, and top-output analyses.

#### 4. Backend and transpilation metadata

Metadata files are preserved for each run, including backend name, selected physical qubit paths, active qubit count, total measured qubit count, shot count, seed, run identifier, transpiled depth, operation counts, and no-QEC/no-post-selection status. This metadata allows each result to be traced to the exact hardware execution conditions.

#### 5. Tile maps and selected physical layouts

The physical tile maps used for the 64-active-qubit and 96-active-qubit experiments are preserved. These maps identify the physical qubits assigned to each 16-qubit Madmartigan tile and allow an evaluator to verify that raw, control, phase-scrambled, and ablation circuits were run on the intended physical tile paths.

#### 6. Per-run analysis CSV files

Analysis CSVs are preserved for each run and control class. These files contain the projected tile-level metrics used in the paper, including XEB, HOG, Shannon entropy, IPR, top-output summaries, and reference-overlap values where applicable.

#### 7. Control-circuit outputs

The reproducibility bundle includes raw and analyzed outputs for same-layout generic RCS controls, phase-scrambled controls, and partial-entanglement ablation controls. This allows the reported control separation to be audited directly rather than inferred only from summary text.

#### 8. Scanner and topology-packing outputs

The scanner outputs used to select viable tile paths and diagnose T8/T9 topology-packing limits are preserved. These include candidate tile lists, selected pack files, transpilation summaries, and failure summaries for fixed-six and fixed-four scaling attempts.

This reproducibility structure is important because the claims in this paper depend on reference-specific comparison, not merely on raw hardware execution. Each reported XEB or HOG value is meaningful only relative to a defined ideal reference distribution, a specific tile projection, and a specific physical circuit execution. Therefore, the raw bitstring counts, ideal reference distributions, selected tile paths, and analysis scripts must remain coupled as a single evidence package.

The intended audit chain for any reported result is:

```
QPY / QASM3 circuit artifact
->
backend transpilation and metadata
->
hardware job and raw counts
->
tile-level projection
->
ideal-reference comparison
->
XEB / HOG / entropy / IPR metrics
->
reported table or paper claim
```

This chain applies equally to the raw Madmartigan circuits and to all control circuits. In particular, the same-layout generic RCS controls and phase-scrambled controls are not treated as informal comparisons. They are preserved as full hardware-executed circuit families with their own raw counts, metadata, and analysis outputs.

The reproducibility package also addresses the sparse-sampling concern inherent in 4096-shot sampling over a 16-bit tile space. Because each tile has  $2^{16} = 65,536$  possible bitstrings, the empirical distribution is necessarily sparse. For that reason, representative summary statistics alone are insufficient. The raw count files and per-tile analysis CSVs are retained so that an evaluator can recompute the reported metrics, examine the empirical distribution directly, compare raw and control outputs, and calculate confidence intervals or additional statistical tests.

No full 96-qubit statevector claim is made for the 96-active-qubit experiments. The analysis is intentionally tile-projected: each 16-qubit tile is scored against its corresponding 16-qubit ideal reference distribution. This design preserves the original Madmartigan benchmark metric regime while allowing simultaneous multi-tile hardware execution. The reproducibility bundle therefore preserves both the full measured 156-bit hardware output and the projected 16-bit tile distributions used for per-tile scoring.

The current data package is intended to support three levels of review:

1. **Internal technical review**

Verification that the reported numbers in the paper match the stored analysis CSVs and raw count files.

2. **Independent rerun preparation**

Use of QPY/QASM3 files, tile maps, and metadata to reproduce the circuits on the same or comparable IBM backend subject to backend availability and calibration drift.

3. **External statistical audit**

Independent recomputation of XEB, HOG, entropy, IPR, confidence intervals, effect sizes, and raw-versus-control separation from the preserved count and analysis files.

Accordingly, the results should be understood not as isolated representative screenshots or selected run summaries, but as a traceable hardware evidence package. The paper reports the condensed scientific findings; the reproducibility bundle preserves the underlying circuit, hardware, and analysis artifacts required to inspect, rerun, or challenge those findings.

## 27 Limitations

The work remains subject to several important limitations:

1. It has not yet been independently reproduced by an external lab.
2. It remains backend- and calibration-sensitive, and additional backend-diverse runs are needed to determine how well the observed structured-output behavior generalizes across devices, calibration cycles, and hardware topologies.
3. The 96-qubit result uses per-tile 16-qubit reference scoring, not a full 96-qubit statevector reference. This is intentional and preserves the original Madmartigan tile metric regime, but it limits the claim to tile-projected reference preservation rather than full 96-qubit distributional fidelity.
4. The T8/T9 clean scale-up was blocked by topology packing under the tested independent 16-qubit chain-tile model. This does not invalidate the T6 result, but it shows that further scaling will require

backend-pressure testing, flexible non-chain tile geometries, smaller tile sizes, or topology-native circuit design.

5. The partial-ablation attractor mode does not yet have assigned operational semantics. Although it produced strong own-reference structure, additional work is needed to determine whether this mode can be used as a stable activation, signaling, or command-band primitive.
6. The statistical analysis currently establishes strong raw-versus-control separation within the available dataset, but additional external audit, independent recomputation, and pre-registered reruns would further strengthen the evidentiary posture.
7. The reproducibility package includes circuit artifacts, metadata, raw counts, and analysis outputs, but external reviewers must still validate the artifact chain, rerun selected circuits where feasible, and confirm that the metric pipeline reproduces the reported values.
8. The present work does not frame its result as universal quantum advantage, full fault tolerance, or a general replacement for quantum error correction. Instead, it identifies a narrower and potentially disruptive benchmark regime: repeatable, reference-specific structured-output preservation on uncorrected NISQ hardware, with no post-selection and strong same-layout control separation. This may represent a practical pre-fault-tolerant utility pathway, but broader claims require independent reproduction, backend-diverse validation, and external audit.

These limitations do not erase the result. They define the next scientific steps and clarify the boundary between the demonstrated benchmark claim and broader future implications.

## 28 Future Work

Several follow-on directions remain important for extending the present benchmark package.

### 28.1 Independent Reproduction

The highest-priority next step is independent reproduction of the 96-active-qubit T6 rank-2 benchmark package by an external evaluator or trusted laboratory. Such reproduction should use the provided QPY/QASM3 circuit artifacts, selected tile maps, raw-count analysis pipeline, and control-circuit definitions to verify whether the reported raw-versus-control separation persists under independent execution.

### 28.2 Backend-Diverse Validation

The present results were obtained on IBM Marrakesh. Additional executions on comparable superconducting backends and across multiple calibration cycles would help determine whether the observed structured-output preservation is backend-specific, calibration-window-specific, or more generally reproducible across Heron-class devices and other hardware topologies.

### 28.3 Topology-Native Scale-Up

Clean independent 16-qubit tile replication encountered a topology-packing wall beyond the T6 / 96-active-qubit regime. Future scale-up should therefore explore topology-native circuit layouts, flexible non-chain tile geometries, smaller tile units, and backend-pressure designs that preserve the validated core while adapting more naturally to the physical coupling graph.

## 28.4 Statistical and Reproducibility Audit

The present paper includes aggregate statistical analysis and a reproducibility package. Future work should include independent recomputation of the reported XEB, HOG, entropy, IPR, confidence intervals, and raw-versus-control effect sizes from the preserved raw counts and analysis CSVs. This would further strengthen the audit chain from circuit artifact to reported benchmark claim.

## 28.5 Operational Mapping of Structured Output Bands

The Madmartigan benchmark demonstrates reference-specific structured-output preservation, but additional work is needed to assign operational semantics to stable output bands. Future experiments should test whether high-confidence reference bands can support command authentication, threshold classification, IPCM-style quantum-to-classical handoff, or other structured signaling primitives.

## 28.6 Ablation-Mode Characterization

The partial-entanglement ablation revealed a high-attractor QSCE-derived mode distinct from the original Madmartigan command-preservation band. Future work should determine whether this attractor mode is seed-robust, backend-stable, and operationally useful as a separate activation or signaling primitive.

## 28.7 Dual-Use and Export-Control Review

Because structured quantum-output preservation may have implications for secure signaling, authentication, and cyber-hardening, future development should proceed with appropriate dual-use review, export-control awareness, and protected evaluation before broader operational disclosure or commercialization.

## 28.8 Evidence Package Consolidation

Prepare a full reproducibility bundle:

```
raw T6 rank-2 script
raw T6 rank-2 QPY/QASM3
raw counts and analysis CSVs
generic RCS control scripts/results
phase-scrambled control scripts/results
partial-entanglement ablation scripts/results
scanner outputs for fixed-six, fixed-four T8, and fixed-four T9
topology-packing failure summary
```

## 28.9 Utility Testing

For raw Madmartigan:

- Map high-confidence output bands to command labels.
- Test threshold classification stability.
- Connect to IPCM-style quantum-to-classical decode.

For partial-ablation attractor mode:

- Determine whether its high-ownXEB attractor can be assigned stable command semantics.

- Test across seeds.
- Test whether the attractor remains robust under backend-pressure loading.

## 29 Final Summary

The Madmartigan scale-up endeavor successfully transformed the original 16-qubit benchmark into a controlled 96-active-qubit evidence package. The raw T6 rank-2 circuit preserved reference-specific Madmartigan structure across six simultaneous tiles on IBM Marrakesh without QEC, without post-selection, and with only 4096 shots. Same-layout generic RCS and phase-scrambled controls failed to reproduce the Madmartigan reference band, while partial-entanglement ablation revealed a new QSCE-derived high-attractor mode rather than invalidating the original claim.

The attempt to go beyond 96 active qubits revealed a topology-packing wall for clean independent 16-qubit tiles across Marrakesh, Fez, and Kingston. This led to the correct next-stage reframing: preserve the proven 96-qubit core and pressure-test it inside a 144-active-qubit circuit using best-effort stress tiles.

The result is a strong, coherent, and scientifically defensible scale-up narrative:

The original Madmartigan benchmark is no longer just a 16-qubit anomaly. It has matured into a 96-active-qubit controlled NISQ evidence package showing reference-specific structured-output preservation, control separation, and a broader QSCE attractor design space.

## A Statistical Significance and Control-Separation Analysis

This appendix reports aggregate statistics computed from the preserved analysis CSVs in the Madmartigan reproducibility bundle. The purpose is not to replace the raw-count audit chain, but to quantify the separation between raw Madmartigan executions and same-layout controls using the tile-level metrics reported throughout the paper. The analysis focuses on  $F_{XEB}$ , HOG, Shannon entropy, and IPR at the 16-qubit tile-projection level.

### A.1 Sparse-Sampling Context

Each tile is a 16-qubit output space with  $2^{16} = 65,536$  possible bitstrings. Each hardware execution used 4096 shots, so the empirical distribution is necessarily sparse. This makes raw representative examples insufficient by themselves. The correct statistical question is whether elevated Madmartigan-reference  $F_{XEB}$  and HOG persist across repeated runs, tiles, seeds, and layouts, and whether same-layout controls collapse toward the Madmartigan-reference baseline.

The preserved data support this question because each reported hardware run is available as raw counts, metadata, and per-tile analysis CSV. The statistics below therefore treat each tile-level run as one observation in the repeated hardware evidence package.

### A.2 Aggregate Raw and Control Distributions

Table 1 summarizes the primary Madmartigan-reference  $F_{XEB}$  and HOG distributions used in the paper. Confidence intervals are 95% t-intervals over the tile-level observations.

Table 1: Primary Madmartigan-reference XEB and HOG summary statistics.

Dataset	n	XEB mean [95% CI]	XEB range	HOG mean [95% CI]	HOG range
64q seed 1337 raw	20	1.503 [1.204, 1.803]	0.891 to 2.892	0.688 [0.668, 0.708]	0.639 to 0.763
64q 1337 generic control	20	-0.085 [-0.118, -0.052]	-0.183 to 0.037	0.492 [0.484, 0.500]	0.468 to 0.529
64q seed 1338 raw	20	1.735 [1.571, 1.899]	1.302 to 2.207	0.709 [0.699, 0.719]	0.672 to 0.741
64q 1339 generic control	20	-0.014 [-0.039, 0.010]	-0.102 to 0.100	0.492 [0.488, 0.496]	0.475 to 0.507
64q seed 1339 raw	20	1.653 [1.489, 1.816]	1.120 to 2.325	0.757 [0.745, 0.768]	0.712 to 0.792
64q seed 1340 weak	4	0.370 [0.153, 0.587]	0.227 to 0.554	0.577 [0.543, 0.611]	0.555 to 0.606
64q seed 1341 weak	4	0.387 [0.195, 0.579]	0.259 to 0.532	0.582 [0.555, 0.608]	0.567 to 0.602
96q depth-831 generic control	30	-0.117 [-0.137, -0.096]	-0.225 to -0.010	0.482 [0.477, 0.487]	0.454 to 0.503
96q T6 generic control	30	-0.104 [-0.119, -0.089]	-0.187 to -0.022	0.486 [0.482, 0.489]	0.471 to 0.504
96q T6 partial-ablation Mad-ref	30	0.217 [0.188, 0.246]	0.092 to 0.388	0.563 [0.555, 0.570]	0.531 to 0.590
96q T6 phase-scrambled control	30	-0.009 [-0.026, 0.007]	-0.126 to 0.074	0.495 [0.491, 0.500]	0.472 to 0.512
96q T6 rank-2 raw	30	1.220 [1.031, 1.409]	0.266 to 2.389	0.669 [0.653, 0.686]	0.581 to 0.759
96q depth-831 raw	30	1.076 [0.792, 1.359]	0.107 to 2.998	0.648 [0.623, 0.673]	0.547 to 0.779

### A.3 Control-Separation Tests

Table 2 reports raw-versus-control separation for the main comparisons. Cohen’s  $d$  is calculated using the pooled standard deviation, and positive values indicate that the raw Madmartigan distribution is higher than the control distribution. Welch tests and Mann–Whitney tests are reported as distributional checks, not as substitutes for independent external replication.

### A.4 64-Active-Qubit Strong-Seed Family

The 64-active-qubit strong-seed family consists of seeds 1337, 1338, and 1339 on the same four calibration-aware physical tile paths. Across those three seeds, the data contain 60 tile-level Madmartigan observations.

Table 2: Raw-versus-control separation statistics.

Comparison	Metric	Raw mean	Control mean	Diff	95% boot CI	Cohen d	Welch p	MW p
64q seed-1337 raw vs. same-tile generic RCS	XEB	1.503	-0.085	1.588	[1.331, 1.875]	3.487	8.10e-10	6.80e-08
64q seed-1337 raw vs. same-tile generic RCS	HOG	0.688	0.492	0.196	[0.177, 0.216]	6.054	1.48e-16	6.79e-08
64q seed-1339 raw vs. seed-matched generic RCS	XEB	1.653	-0.014	1.667	[1.521, 1.821]	6.680	4.48e-15	6.80e-08
64q seed-1339 raw vs. seed-matched generic RCS	HOG	0.757	0.492	0.264	[0.254, 0.275]	14.615	1.06e-24	6.78e-08
96q depth-831 raw vs. same-layout generic RCS	XEB	1.076	-0.117	1.193	[0.935, 1.474]	2.215	1.74e-09	3.02e-11
96q depth-831 raw vs. same-layout generic RCS	HOG	0.648	0.482	0.166	[0.142, 0.190]	3.396	2.59e-14	3.01e-11
96q GLOBALPACK T6 raw vs. same-layout generic RCS	XEB	1.220	-0.104	1.324	[1.145, 1.500]	3.686	9.50e-15	3.02e-11
96q GLOBALPACK T6 raw vs. same-layout generic RCS	HOG	0.669	0.486	0.184	[0.167, 0.199]	5.719	1.27e-20	3.01e-11
96q GLOBALPACK T6 raw vs. phase-scrambled Mad-ref	XEB	1.220	-0.009	1.229	[1.050, 1.404]	3.420	6.24e-14	3.02e-11
96q GLOBALPACK T6 raw vs. phase-scrambled Mad-ref	HOG	0.669	0.495	0.174	[0.157, 0.189]	5.352	2.76e-20	3.01e-11
96q GLOBALPACK T6 raw vs. partial-ablation Mad-ref	XEB	1.220	0.217	1.002	[0.822, 1.180]	2.768	7.63e-12	8.99e-11
96q GLOBALPACK T6 raw vs. partial-ablation Mad-ref	HOG	0.669	0.563	0.107	[0.089, 0.123]	3.095	9.08e-15	8.97e-11
64q strong seed family vs. seed-matched generic controls	XEB	1.630	-0.050	1.680	[1.563, 1.803]	4.579	1.34e-36	3.18e-17
64q strong seed family vs. seed-matched generic controls	HOG	0.718	0.492	0.226	[0.215, 0.237]	6.725	5.95e-52	3.18e-17

The aggregate strong-seed Madmartigan-reference distribution has mean  $F_{XEB} = 1.630$  and mean HOG  $= 0.718$ . All 60/60 tile-level  $F_{XEB}$  values are positive and all 60/60 HOG values are above 0.5.

By contrast, the two seed-matched 64-active-qubit generic RCS controls for seeds 1337 and 1339 contain 40 tile-level control observations. Their aggregate Madmartigan-reference mean  $F_{XEB}$  is -0.050, and their aggregate Madmartigan-reference mean HOG is 0.492. This separation is the central statistical result at the 64-active-qubit level: the same physical tile map can preserve control-own structure, but generic RCS controls do not reproduce the Madmartigan-reference band.

## A.5 96-Active-Qubit Scale-6 Results

The first viable 96-active-qubit Scale-6 stress layout at depth 831 produced 30 tile-level observations across five runs. Its mean Madmartigan-reference  $F_{XEB}$  is 1.076, with mean HOG 0.648. All 30/30 tile-level  $F_{XEB}$  values are positive. This supports the interpretation that the original four-tile core survived a harsher 96-active-qubit circuit burden.

The optimized GLOBALPACK T6 rank-2 layout produced 30 tile-level observations across five runs. Its mean Madmartigan-reference  $F_{XEB}$  is 1.220, with mean HOG 0.669. All 30/30 tile-level  $F_{XEB}$  values are positive and all 30/30 HOG values are above 0.5.

The same-layout GLOBALPACK T6 generic RCS control has mean Madmartigan-reference  $F_{XEB} = -0.104$  and mean Madmartigan-reference HOG  $= 0.486$ . The raw-versus-generic comparison yields very large positive effect sizes for both  $F_{XEB}$  and HOG, as shown in Table 2.

## A.6 Sparse-Sampling Robustness Interpretation

The sparse-sampling objection is not answered by a single large  $F_{XEB}$  value. It is answered by repeatability and control separation. Across the preserved CSVs, elevated Madmartigan-reference  $F_{XEB}$  and HOG recur across runs, tiles, seeds, and 64/96-active-qubit layouts. Same-layout generic RCS controls preserve their own reference structure but collapse toward Madmartigan-reference baseline. Phase-scrambled controls likewise preserve some own-reference structure while failing to reproduce the raw Madmartigan band. The partial-entanglement ablation is not a clean negative control; it demonstrates a different high-attractor own-reference mode with weaker Madmartigan-reference overlap.

Therefore, the statistically defensible conclusion is not that sparse sampling is impossible, nor that the results prove a philosophical claim of deterministic steering. The defensible conclusion is narrower: the observed Madmartigan-reference elevation is repeatable, reference-specific, and strongly separated from same-layout controls in the preserved tile-level data.



## B Circuit Serialization and Reproducibility Manifest

The reproducibility archive contains the artifacts needed to audit the reported circuit construction, hardware execution, raw-count storage, and analysis pipeline. Table 3 summarizes the artifact classes detected in the archive.

Table 3: Reproducibility artifact manifest summary.

Artifact class	Files	MB
analysis_csv	84	0.36
metadata_json	109	0.23
other	6	3.69
pack_selection	20	0.11
qasm_export	33	3.61
qpy_circuit	34	11.62
raw_counts	82	82.16
scanner_output	33	7.70
selected_layout_or_tile_asset	5	0.04

The archive includes serialized QPY circuits, QASM3 exports, raw counts, metadata, selected tile assets, scanner outputs, and pack-selection summaries. This supports the intended audit path:

```
QPY / QASM3 circuit artifact
->
backend metadata and transpilation profile
->
hardware counts
->
tile-level projection
->
ideal-reference comparison
->
XEB / HOG / entropy / IPR metrics
->
reported paper claim
```

The manifest CSV accompanying this appendix lists each file path, artifact class, top-level campaign folder, file extension, and byte size.