TECHNICAL NOTE · SYNDROMICS · WISH APPLES VENTURES PRIVATE LIMITED · 18 NOVEMBER 2026

Measuring a processor's real noise from its own syndromes: first run on `ibm_kingston`

What we ran, what we found, how it compares with existing methods, and what we do not yet know.

Summary

We ran a distance-5 repetition code (9 qubits, 10 rounds, 30,000 shots) on IBM's `ibm_kingston` Heron processor through the free Open Plan. From the detection events alone, Syndromics fitted a physical noise model of the line: one scale factor per measurement, idle and two-qubit-gate component, starting from IBM's published calibration. We then asked two questions.

1. **Does the measured model decode better?** Yes. On shots held out from the fit, a matching decoder given the Syndromics model made **half as many logical errors** as one given IBM's calibration: **0.40% → 0.20%**. It was better in **6 of 6** train/test splits, including a time-ordered split.
2. **Where is the extra noise?** At the **q128–q129 end of the line**: that gate is at least 4× its calibrated error, and q128 has the shortest T2 on the line in IBM's own data. Mid-circuit measurement on every measure qubit is 1.7–5× worse than the published end-of-circuit readout.

0.40% → 0.20%	**6 / 6**	**q128–q129**	**1.7–5×**
logical error, IBM calibration vs Syndromics model, held-out real shots	train/test splits where Syndromics beat the calibration (5 random + 1 time split)	localised as the main excess error: gate ≥4× calibration; q128 T2 = 53 µs	mid-circuit measure + reset error vs IBM's published readout, measured in the same job

The honest comparison. Standard model-free correlation re-weighting from the QEC literature reaches the **same** decoding accuracy on this data (0.199% vs our 0.198%). On a small code the decoding gain alone is therefore not unique to us. What Syndromics adds is **attribution**: a physical model that names the gate or qubit responsible. That attribution can drive qubit placement and recalibration as well as decoding, and it may be reusable across circuits. The next experiment tests that reuse.

Scope. One run, one 9-qubit line, one day, a repetition code (bit-flip errors only). IBM's calibration snapshot was 5 days old at run time. These are first results, not a characterisation of the processor.

1. The experiment

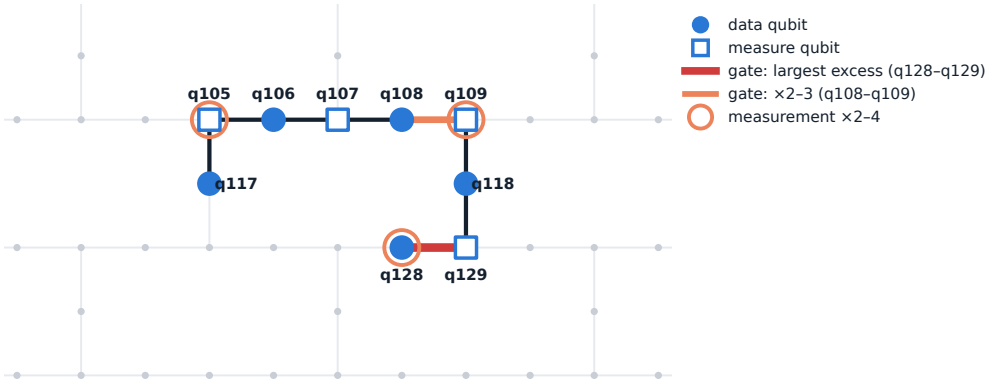


Figure 1. The 9-qubit line on `ibm_kingston` (part of the 156-qubit heavy-hex lattice). Five data qubits (filled) alternate with four measure qubits (squares). Colours mark the findings of Section 3.

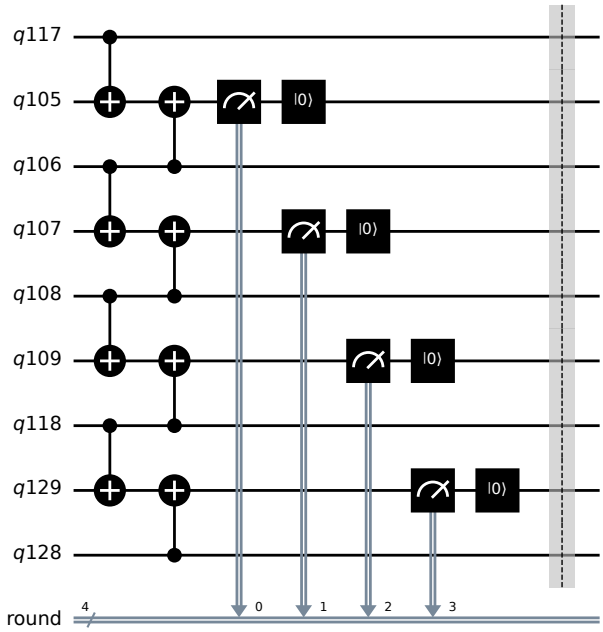


Figure 2. One round of the repetition code, written with CNOTs. Each measure qubit checks the parity of its two neighbours, is measured mid-circuit, then reset. The circuit runs 10 rounds, then measures all data qubits. On hardware Qiskit compiled it to IBM's native gates: 80 CZ, 80 SX, 45 measurements and 40 resets, with the rounds interleaved.

Run details	
Backend	`ibm_kingston` (IBM Heron r2, 156 qubits), IBM Quantum Open Plan
Job	`dbukcjvf8rhs73bmf9l0`, submitted 18 November 2026, 06:32 UTC
Qubits (line order)	117, **105**, 106, **107**, 108, **109**, 118, **129**, 128 (measure qubits in bold)
Circuit	Repetition code, distance 5, 10 rounds, 30,000 shots, logical state |0⟩
Same-job references	4 circuits × 30,000 shots on the same qubits: end-of-circuit readout of |0⟩ and |1⟩, and 6 rounds of mid-circuit measure + reset of |0⟩ and |1⟩
IBM calibration used	Snapshot dated 13 November 2026, 20:27 IST
Software	Syndromics 0.4.3; Qiskit 2.6, Qiskit Runtime SamplerV2, Stim 1.16, PyMatching 2.4

Qubit line chosen automatically by the kit, as the line with the best published errors among usable qubits. Without decoding, the logical answer was wrong in 2.2% of shots; the detection-event rate was 3.0%.

2. Method

From syndromes to a noise model

1. **Detection events.** Each measure-qubit result is compared with the previous round's; a change is a detection event (44 detectors per shot: 4 measure qubits × 11 time steps, the last from the final data measurement).
2. **Nominal model.** A Stim circuit of exactly this code with IBM's published per-qubit and per-gate errors. Its error mechanisms are grouped into physical *components*: each measurement, each qubit's idle/single-qubit noise, each two-qubit gate.
3. **Fit.** For each component, Syndromics fits one scale factor so that the model reproduces the measured detection rates and pairwise detector correlations. A sparsity penalty keeps components at their calibrated value unless the data clearly says otherwise.
4. **Output.** A corrected circuit and detector error model for the decoder, plus a list of the components that are clearly worse than calibrated (≥2× and ≥3σ).

How the decoding test avoids fooling itself

Every noise model is fitted on one half of the shots and judged on the other half, which the fit never saw. We repeat this for five random splits and one time split (fit on the first 15,000 shots, decode the last 15,000), so slow drift during the job cannot flatter the result. The decoder is the same in every case: PyMatching minimum-weight perfect matching. Only the noise model it is given changes.

Noise model given to the decoder	What it knows
Uniform	One median value per error type for the whole line: no device-specific information.
IBM calibration	IBM's published per-qubit readout, per-gate CZ error and T1/T2-derived idle error.
Correlation re-weighting	The standard model-free method: each decoding-graph edge weight estimated directly from detector pair correlations (p_{ij}). No physical model.
Syndromics	IBM's calibration, corrected component by component by the fit above.

> **A bug we found on this data, and fixed.** In some splits the first version of the fit made decoding *worse* (0.8%). The fit had added a crosstalk term between data qubits q118 and q128. In a repetition code that term becomes an edge in the decoding graph that flips the logical answer, and the decoder took it as a shortcut. Fitted crosstalk is now reported as a diagnostic but kept out of the decoder. After the fix, every start point and fit setting we tried gave 0.18–0.22%. This data set is now an automated regression test in our code.

3. Results

3.1 The measured model halves logical errors on held-out shots

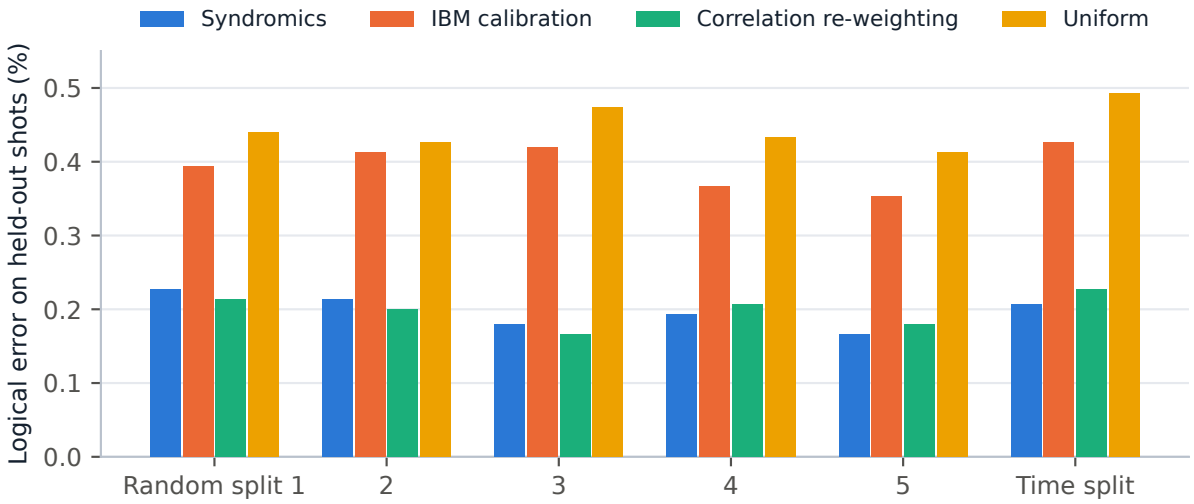


Figure 3. Logical error on the held-out half for each of the six splits. Same decoder throughout; only the noise model differs.

Noise model	Mean logical error	Range over 6 splits	vs IBM calibration
No decoding (raw logical flips)	2.22%	—	—
Uniform	0.447%	0.413–0.493%	+13%
IBM calibration	0.396%	0.353–0.427%	baseline
Correlation re-weighting (literature)	0.199%	0.167–0.227%	−50%
Syndromics	**0.198%**	0.167–0.227%	**−50%**

Each split decodes 15,000 shots, so one logical error is 0.0067 percentage points; the calibration model makes about 60 errors per split and Syndromics about 30. In the time split, Syndromics fitted on the first half scored 0.207% on the second half against 0.427% for the calibration.

Why the calibration model loses

Two effects stand out in the data (Sections 3.2 and 3.3). Mid-circuit measurement is worse than the end-of-circuit readout that calibration reports, and one end of the line has an error source the calibration does not show. Both change which error chains are likely, so the calibrated decoder picks the wrong correction more often.

3.2 Where the extra noise is

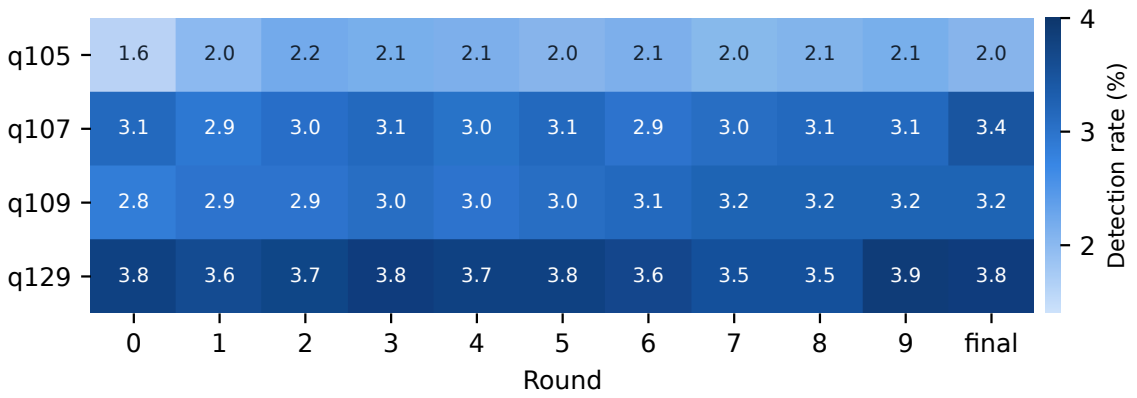


Figure 4. Detection-event rate for each measure qubit and round (raw data, no model). q129, next to q128, fires most often in every round: the excess is steady through the job, not a one-off event.

Component (fit on all 30,000 shots)	vs calibration	Supporting evidence
Two-qubit gate q129–q128	**≥4×** (4–17×)	Largest finding in every fit setting. q128 has T2 = 53 µs, the lowest on the line (others 129–285 µs); q129 has the highest detection rate (Fig. 4).
Two-qubit gate q108–q109	2–3×	q108 T2 = 129 µs, second lowest.
Measurement q105, q109, q128	2–4×	Consistent with the same-job mid-circuit references (Fig. 5).

A repetition code localises a fault well (which qubit or pair) but splits it only roughly between a qubit's measurement and its neighbouring gate; the q128–q129 ratio ranges from 4× to 17× across fit settings while its location does not change. A targeted two-qubit characterisation of q128–q129 would settle it.

3.3 Mid-circuit measurement is worse than published readout

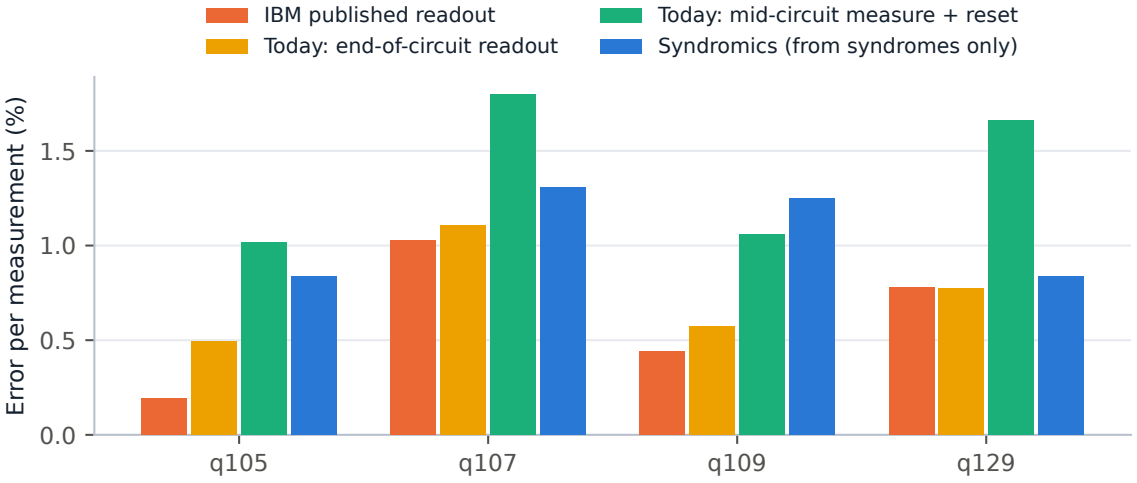


Figure 5. Error per measurement on the four measure qubits: IBM's published readout, today's end-of-circuit readout and today's mid-circuit measure + reset (both from the same job), and Syndromics' estimate from the syndromes alone.

End-of-circuit readout today is close to IBM's published values. Mid-circuit measure + reset, which is what an error-correction circuit uses every round, is 1.7–5× higher (q105: 0.20% published, 1.02% mid-circuit). A decoder weighted from published readout therefore underweights measurement errors. On q105, q107 and q109, Syndromics' estimate from the syndromes follows the mid-circuit references rather than the published values. On q129 it stays near the published value, because the fit attributes that end's excess to the q129–q128 gate instead: the measurement-versus-gate ambiguity noted above.

4. What this shows, and what it does not

Supported by this run

- On real Heron hardware, a model fitted from syndromes decodes better than the published calibration, on held-out data, consistently.
- The fit localises the excess error to a specific region (q128–q129), corroborated by raw detection rates and IBM's own T2 data.
- Mid-circuit measurement is materially worse than published readout on this line.
- The full analysis runs in seconds on a laptop from one short job.

Not yet shown

- Any advantage over correlation re-weighting in decoding accuracy: on this code they tie.
- That the result holds on other lines, other days or other processors.
- Phase errors: a repetition code sees only bit flips. Surface or qLDPC codes are needed.
- Whether the q128–q129 finding is a gate or a qubit problem (needs targeted characterisation).

Next experiments

Experiment	Question	Cost
Transfer test	Fit the model on this circuit, then decode a *different* circuit on the same qubits (more rounds, longer code). A physical model can be reused; correlation re-weighting must be re-measured for each circuit. This is where we expect Syndromics to differ.	One short job
Repeat across lines and days	Does the 2× gain hold? Does the q128–q129 excess persist, move, or vanish after recalibration?	A few minutes of QPU time per day
Chip-wide health check	Seven lines covering 115 of 156 qubits: which components have changed since calibration?	One job (already in the kit)
Surface code, then qLDPC	Extend the fit to codes that see both error types; supply measured priors to belief-propagation decoders such as Relay-BP.	Joint work with IBM

5. Reproduce it

The code is public at `github.com/wishapples/syndromics` (MIT licence), and the self-service demo kit (v0.4.3) includes the raw shots of this job. No IBM account or QPU time is needed to reproduce every number in this note:

```
pip install -r requirements.txt
cd example_ibm_kingston_2026-11-18
syndromics-ibm-demo reanalyse        # prints the tables of Sections 3.1 and 3.2 in a few seconds
```

With an IBM Quantum account, `syndromics-ibm-demo validate --backend <any Heron> --yes` runs the same experiment on a line it chooses, in one job of five circuits.

Contact
Vishal J Shah, Chief Executive Officer, Wish Apples Ventures Private Limited (Syndromics)
ceo@wishapples.com · +91 96644 74477 · syndromics.wishapples.com · github.com/wishapples/syndromics

IBM, Qiskit and Heron are trademarks of IBM. This note reports an experiment on IBM Quantum Platform using a free Open Plan account; it does not imply IBM endorsement. 18 November 2026.