DEM-Informed GNN Decoder for Exploiting Y-Biased Noise Correlations in the Distance-3 Surface Code

Matei Dumitrescu • Bachelor Thesis, Infinidata Lab • Supervised by Rihan Hai & Tim Littau

−60.3%
logical error rate vs. plain MWPM

−24.7%
vs. belief-matching

430k
parameters, trained on 5M samples

$d = 3$ rotated surface code · Y-biased circuit-level noise ($\eta = 100$) · paired holdout, 95% CIs [+56.4, +64.2] & [+16.6, +31.9]

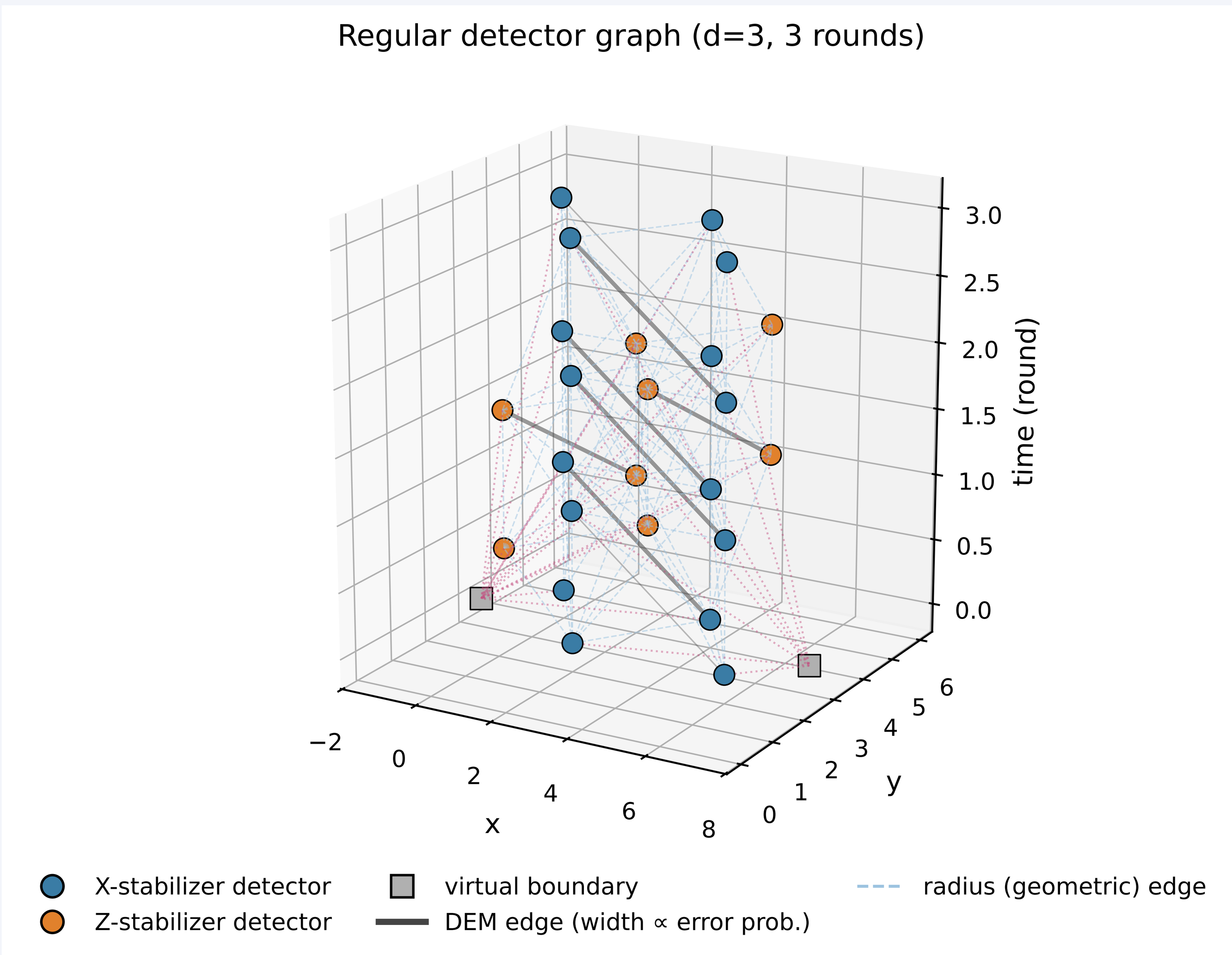
The blind spot in matching decoders

- MWPM decodes the X and Z syndromes on **two independent graphs**.
- A Y **error** (or a correlated two-qubit fault) flips both an X *and* a Z detector at once; independent matching sees **two unrelated defects** and throws the X-Z correlation away.
- **Y-biased noise** (η: $\eta = 1$ unbiased, $\eta = 100$ strongly Y-dominated) makes this correlated class the dominant error.

Our decoder

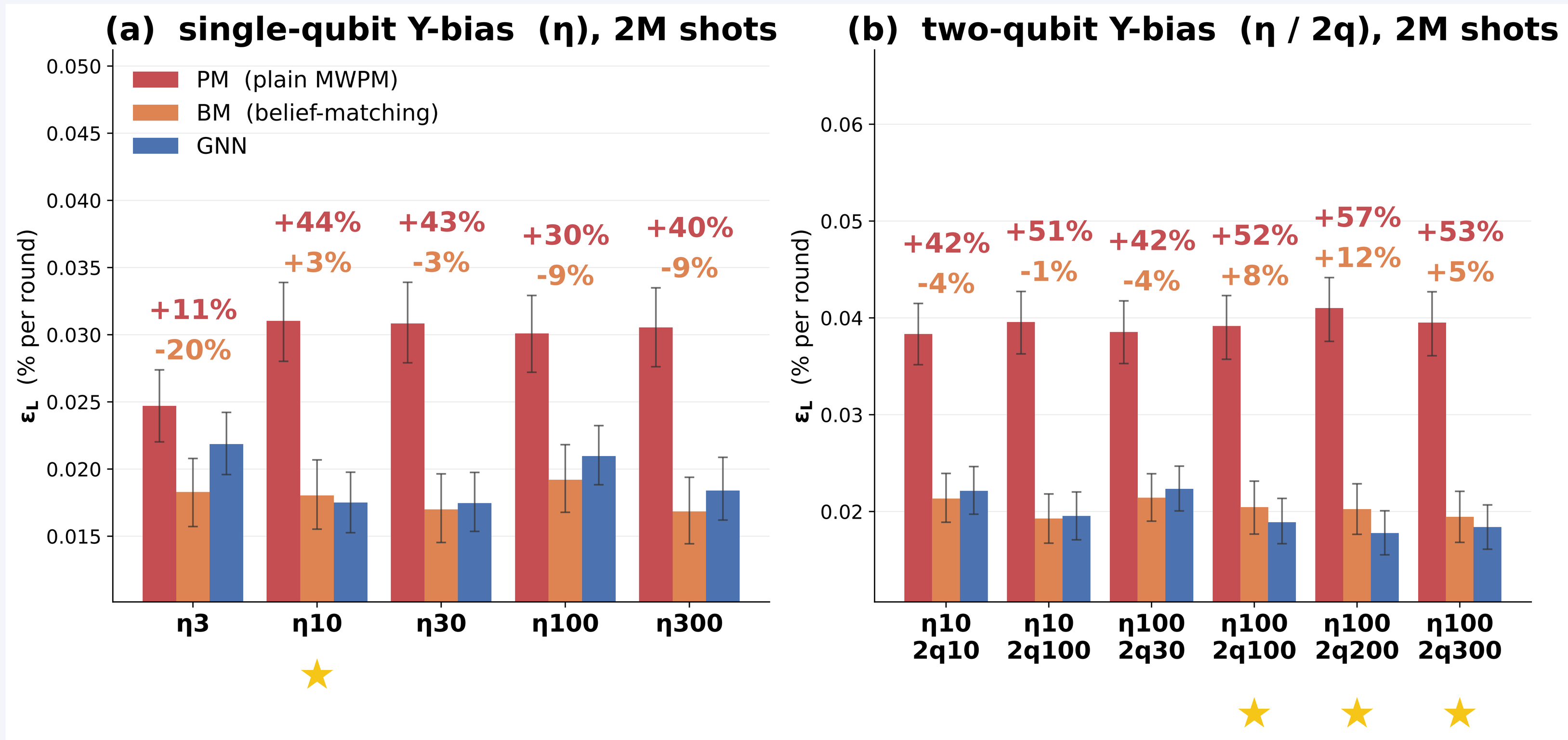
- A **DEM-informed, DEM-weighted edge-feature GNN**: nodes are detectors, edges are DEM fault mechanisms weighted by fault probability.
- **6 message-passing layers**, width 128, $\approx$430k parameters; message passing reads the **joint** X-Z **syndrome** that matching discards.
- Trained with **focal loss** on the natural error distribution plus threshold calibration; evaluated on **seed-disjoint** holdout shots.

The DEM detector graph



The decoder graph is **built directly from the detector error model (DEM)**: detector nodes in space and time, with DEM-weighted edges (width $\propto$ error probability) that span the X and Z sublattices wherever a common fault couples them.

It beats both baselines under correlated bias



Per-round logical error rate ε_L across bias settings (2M shots). Both the **GNN** and **belief-matching** cut **plain MWPM** by **30–57%**. The GNN's extra edge over the strong belief-matching baseline (**starred bars:** $+5$ **to** $+12\%$, McNemar $p < 10^{-6}$) appears only under strong *two-qubit* bias, where it keeps the joint X-Z syndrome coupled instead of decoding the two graphs separately.

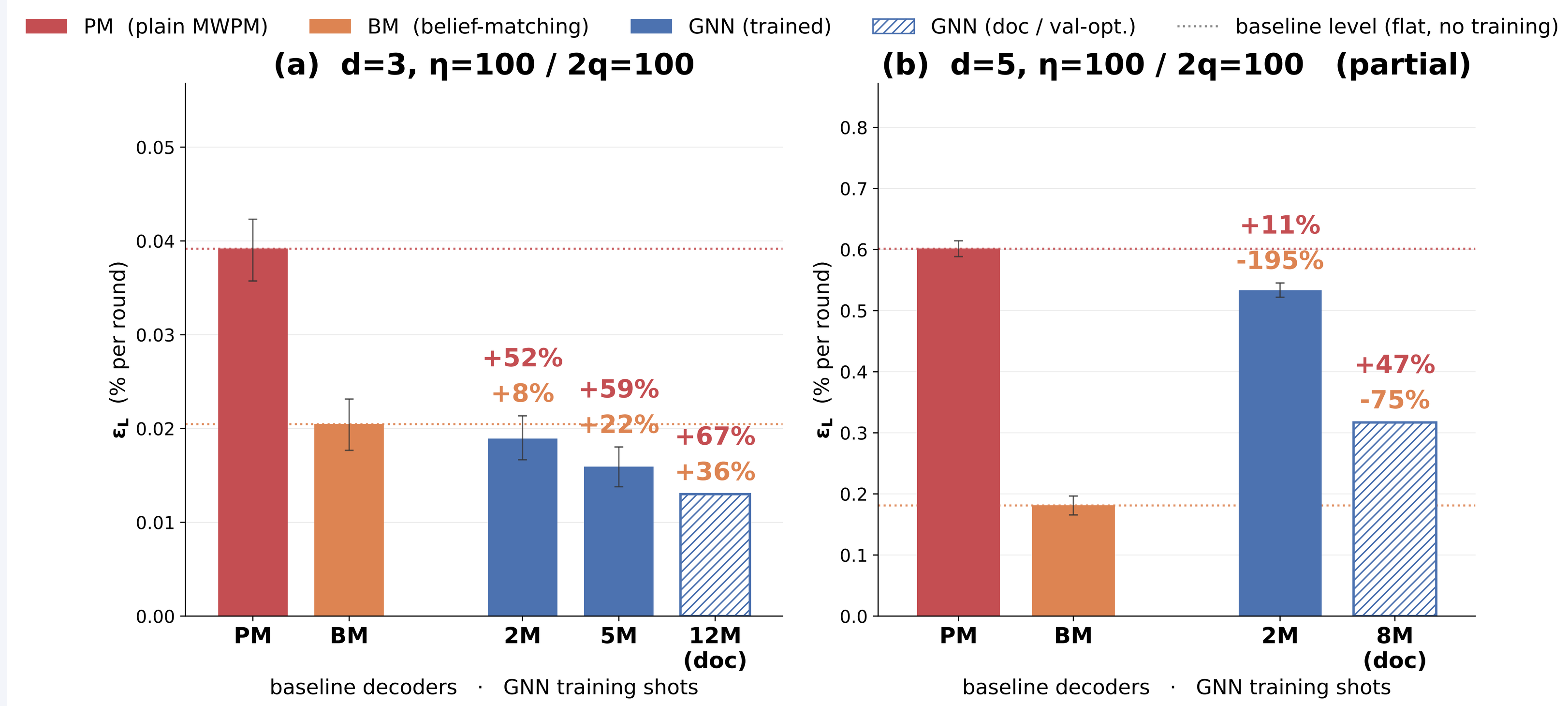
Why a learned decoder helps

- Belief-matching reweights the matching graph with belief propagation, but its last step is still **independent minimum-weight matching** on the separate X and Z graphs.
- The GNN instead keeps the **joint** X-Z **syndrome coupled** through message passing, capturing correlated faults the reweighting drops.
- Small in general, this residual **dominates under strong two-qubit bias**.

Conclusion

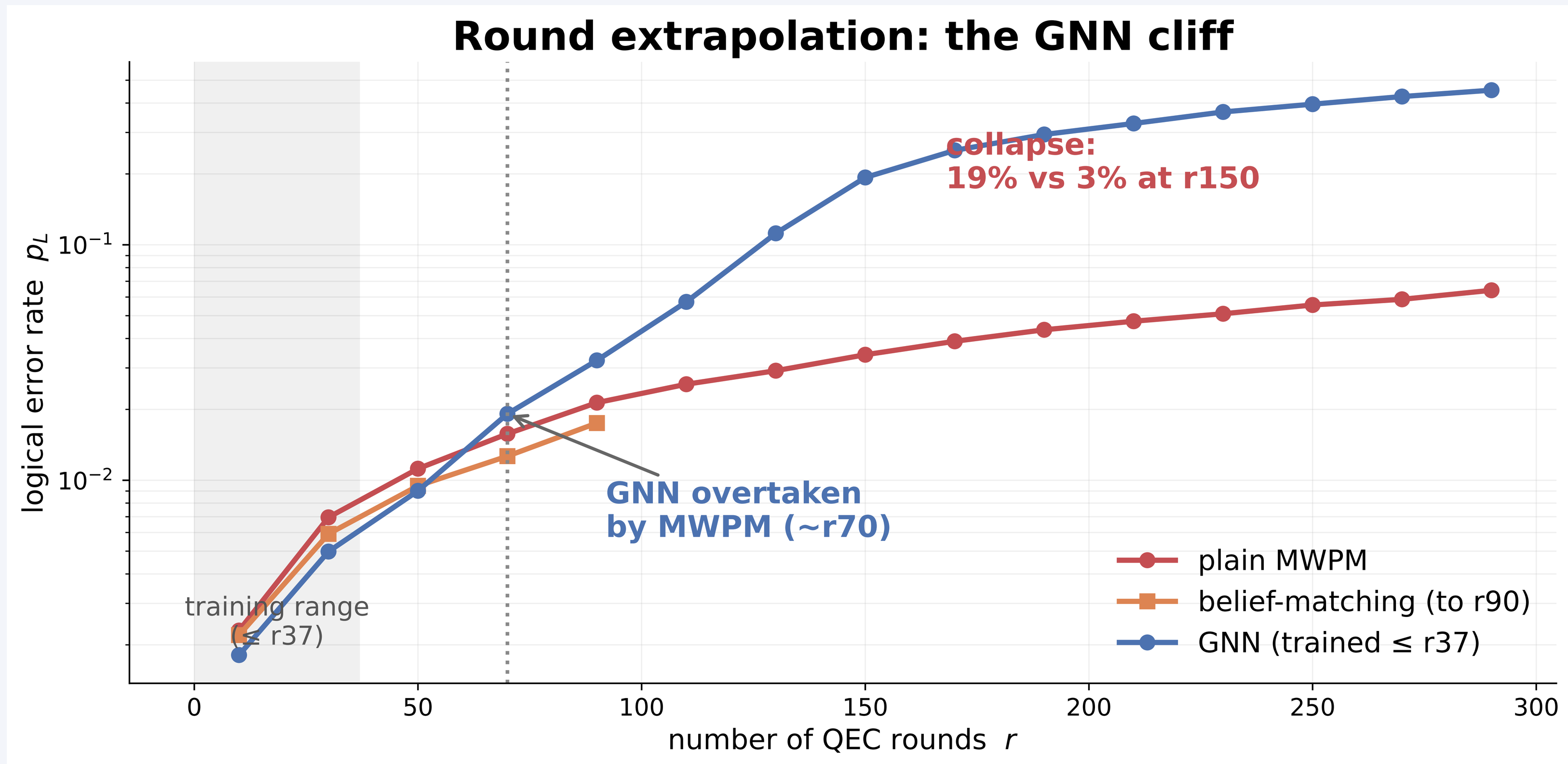
A DEM-informed GNN **recovers the** X-Z **correlation matching discards**, beating plain MWPM by up to 60% and belief-matching by $\sim$25% at $d = 3$ with 430k parameters. Under **unbiased noise the two tie**; the open gaps (larger codes, longer experiments) are **data and generalization**, not capacity.

It's data, not architecture



Same architecture, more shots: the reduction vs. **MWPM** climbs +52/59/67% (2M/5M/12M) and vs. **belief-matching** +8/22/34%, with no saturation. $d = 5$ **needs far more data**: $3\times$ the stabilizers and weight-2 correction blow up the syndrome space, so at a fixed 8M budget belief-matching overtakes, a data-budget limit rather than a capacity ceiling.

But it doesn't extrapolate in time



Per-round p_L as experiments run longer than training (≤ 37 rounds). The GNN tracks matching out to $\sim$50 rounds, then crosses **plain MWPM** near round 70 and **collapses** far beyond its training regime (19% vs. 3% at 150 rounds). Generalizing across experiment *length*, not raw accuracy, is the open problem.

Setup: Stim `rotated_memory_z`, $d = 3$, $p = 0.001$, rounds 1–17 (depth to 37) • **Baselines:** plain MWPM & belief-matching (BP + matching) • metric: per-round logical error rate ε_L