

Figure 10: Toric code - depolarizing noise model results (including SAQ without CPND and BP-OSD)

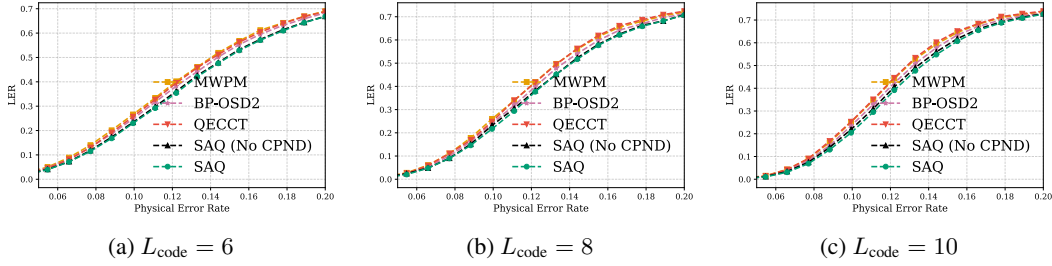


Figure 11: Toric code - independent noise model results (including SAQ without CPND and BP-OSD)

## E SUPPLEMENTARY EXPERIMENTAL RESULTS

### E.1 PERFORMANCE COMPARISON: SAQ (WITHOUT CPND) VS. QECCT AND BP-OSD

To provide a direct and fair comparison with the QECCT architecture, the following experiments evaluate the LER based on our network’s raw physical error prediction,  $\hat{e}$  (the output of the syndrome stream). We also include a full comparison against the BP-OSD decoder, as requested.

This provides an “apples-to-apples” comparison, as this raw prediction is the direct equivalent of the end-to-end QECCT model’s final output. This LER result is distinct from our full SAQ-Decoder performance, which uses  $\hat{e}$  as an informative prior for the CPND stage to construct a final, syndrome-consistent recovery operator.

Figure 10 presents the performance of toric codes under depolarizing noise. In particular, Figure 10c shows that at  $p = 0.2$ , taking SAQ ( $\text{LER} = 6.06 \times 10^{-1}$ ) as a reference, MWPM, BP-OSD, and QECCT incur LER increases of approximately 26%, 23%, and 19%, respectively, whereas SAQ (No CPND) increases the LER by only 2.8%. This indicates that SAQ (No CPND) suffers only a very small marginal degradation relative to SAQ, while maintaining the high marginal gap from the classical and neural baselines. A similar behavior is observed in Figure 11, which reports toric-code performance under independent noise.

Crucially, the performance of SAQ (No CPND) remains close to that of our full, CPND-enabled SAQ decoder, while maintaining a clear performance gap over all other baselines, including QECCT, MWPM, and BP-OSD. This trend is consistent across other code families and physical error-rate regimes, such as the rotated surface code under depolarizing noise (Figure 12) and independent noise (Figure 13). Moreover, the results show that our SAQ Transformer architecture, even when used as a standalone decoder, consistently and significantly outperforms BP-OSD across the entire range of tested physical error rates, demonstrating that the performance gains are not merely due to post-processing, but stem from a more expressive and accurate neural architecture.

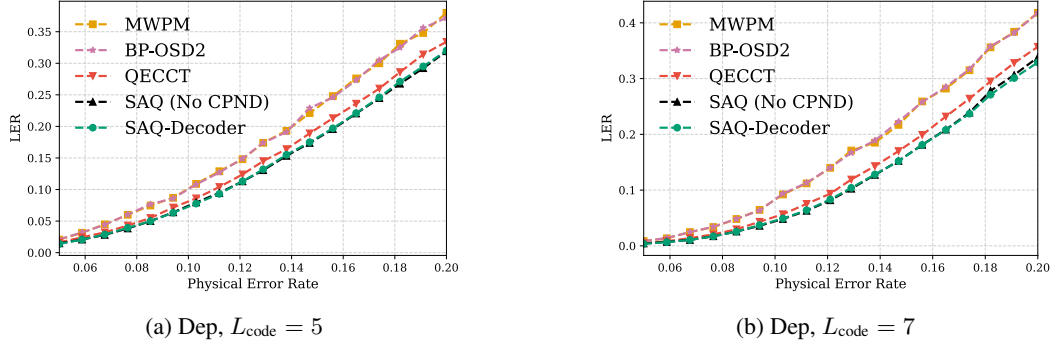


Figure 12: Rotated surface code results

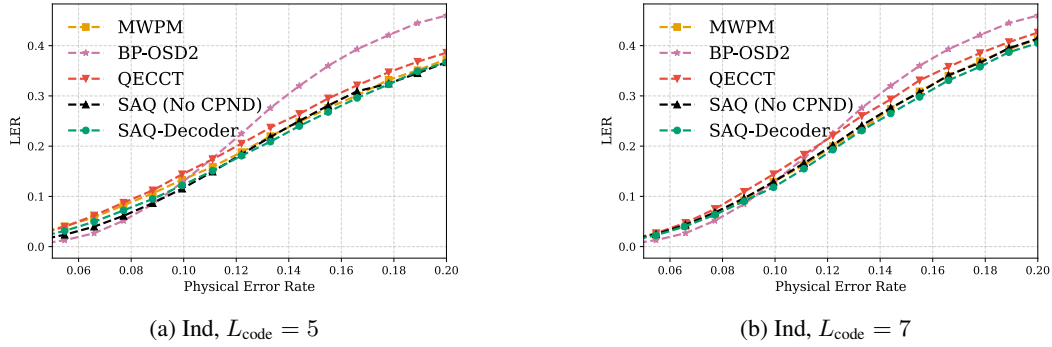


Figure 13: Rotated surface code results

## E.2 SCALABILITY ANALYSIS FOR LARGER CODE DISTANCES

To further probe the scalability of our framework, which is critical for practical fault-tolerance, we extended our analysis to a rotated surface code with  $L_{code} = 11$ . We evaluated this larger code under both independent and depolarizing noise models. The results, presented in Figure 14, demonstrate that SAQ-Decoder scales robustly. The decoder not only maintains its superior performance but also preserves the significant LER gap over both MWPM and QECCT, confirming that the efficiency and accuracy gains of our architecture generalize effectively to larger code distances. In particular, Figure 14b shows that at  $p = 0.2$ , taking SAQ ( $LER = 3.88 \times 10^{-1}$ ) as a reference, the LER of MWPM and QECCT increases by approximately 21% and 8.8%, respectively, whereas SAQ (No CPND) increases the LER by only about 3.1%. This indicates that SAQ (No CPND) incurs only a very small marginal degradation relative to SAQ, while the classical and neural baselines remain substantially worse at the same physical error rate.

## E.3 VALIDATING GENERALITY

To rigorously validate the generalizability of our SAQ-Decoder framework beyond surface codes, we conducted new experiments on two distinct code families: the color code ( $L_{code} = 3, 5$ ) and the repetition code ( $L_{code} = 3, 5$ ). Critically, both experiments were performed under realistic, multi-round circuit-level noise models. These experiments demonstrate robustness confirm our framework’s claim of generality, as it is fundamentally agnostic to the code family.

### E.3.1 COLOR CODE WITH 2-ROUND CIRCUIT NOISE

Figure 15a illustrates the LER performance on a distance 3 color code with 2 rounds of circuit noise, demonstrating robustness with high marginal gaps from the baselines. In Figure 15b SAQ-Decoder significantly outperforms all baselines across the entire range of physical error rates. At the highest analyzed rate of  $p = 2.00e - 02$ , our SAQ-Decoder achieves an LER of  $LER = 1.76 \times 10^{-1}$ . This

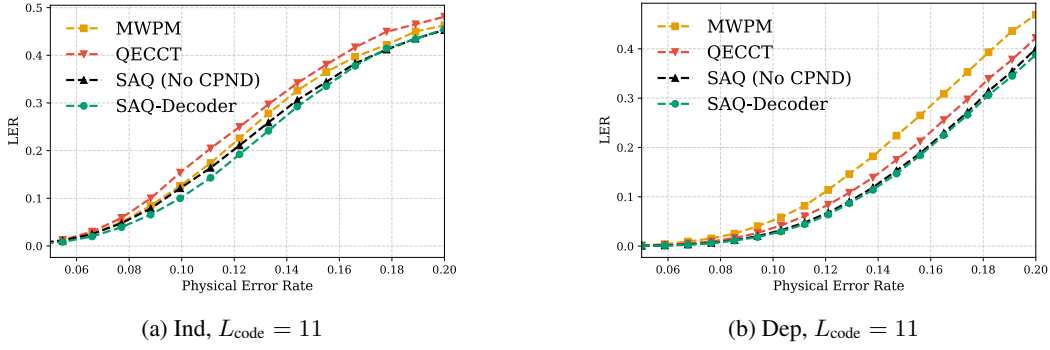
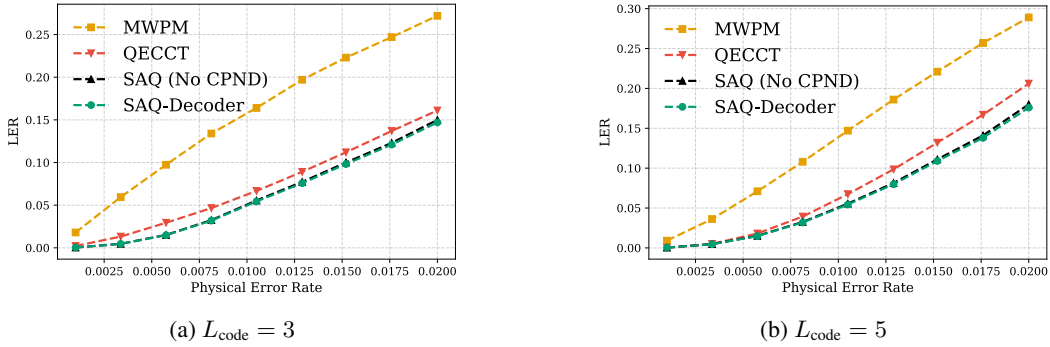
Figure 14: Rotated surface code ( $L_{code} = 11$ ) results.

Figure 15: Color code with circuit noise results.

is 17.0% lower than QECCT ( $LER = 2.06 \times 10^{-1}$ ) and 64.2% lower than the MWPM baseline ( $LER = 2.89 \times 10^{-1}$ ). We also note that the full SAQ-Decoder (with CPND) is 2.2% better than the (SAQ (No CPND)).

### E.3.2 REPETITION CODE WITH 3-ROUND CIRCUIT NOISE

Figure 16a presents the LER results for a distance 3 repetition code with 3 rounds of circuit noise, showing that the decoder remains robust with performance gaps relative to the baseline methods. Similarly, Figure 16b shows the results for a distance 5 repetition code with 3 rounds of circuit noise. While the performance of all decoders is closer on this code, SAQ-Decoder consistently maintains the lowest logical error rate. At  $p = 2.50e - 01$ , SAQ-Decoder's LER of  $4.60 \times 10^{-1}$  is 1.83% lower than QECCT ( $LER = 4.68 \times 10^{-1}$ ) and 2.61% lower than MWPM ( $LER = 4.72 \times 10^{-1}$ ).

## F THE USE OF LARGE LANGUAGE MODELS (LLMs)

LLMs were employed to assist in several aspects of this research and manuscript preparation. For the literature review, LLMs aided in the identification and sourcing of relevant related works to ensure comprehensive coverage of the field. During the research process, LLMs were consulted for research ideation, though these explorations did not yield beneficial outcomes that influenced the final work. In manuscript preparation, LLMs assisted with improving grammar, enhancing textual transitions between sections, and refining the exposition of technical concepts for better clarity and readability. For software development, LLMs provided assistance in code writing, GPU acceleration optimizations, and debugging code issues. Despite these auxiliary uses, all core research contributions, experimental design, theoretical insights, and scientific conclusions presented in this work are entirely the product of the authors' original research and analysis.

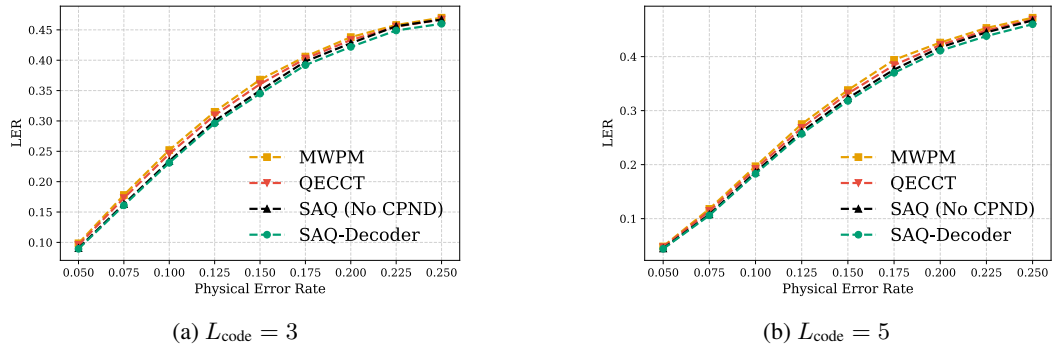


Figure 16: Repetition code with circuit noise results.