Quantum Fidelity - Lab 1

Part I: Two IBM devices, properties, and comparison

I selected two IBM Quantum devices, ibm_torino and ibm_brisbane, and gathered their qubit-level and gate-level properties. The statistics include relaxation time (T1), dephasing time (T2), readout error, single-qubit gate errors, and two-qubit gate errors. Summary metrics (min, max, mean, median) and histograms were used to assess device performance.

For ibm_torino, the mean T1 was about 1.72×10^8 μs and the mean T2 about 1.39×10^8 μs. Readout error averaged 0.0425, single-qubit gate errors averaged 0.00646, and two-qubit gate errors averaged 0.0353. For ibm_brisbane, the mean T1 was about 2.22×10^8 μs, the mean T2 was about 1.35×10^8 μs, and the mean readout error was 0.0330. Single-qubit gate errors were much lower, averaging 0.00026, and two-qubit gate errors also averaged lower at 0.0176.

The histograms show broad distributions of T1 and T2 values for both devices, with most qubits clustered between 1×10^8 and 2.5×10^8 μs. Readout error distributions highlight that Brisbane qubits are more sharply concentrated at low error values, while Torino has a longer tail with higher-error outliers.

Overall, ibm_brisbane performs better than ibm_torino. While coherence times are comparable across the two devices, Brisbane has clearly lower readout error, single-qubit error, and two-qubit error rates. Since two-qubit errors are the dominant contributor to overall circuit reliability, Brisbane is the more favorable device for running quantum circuits. I think this is expected because ibm brisbane has fewer qubits and connections when compared to torino.

Part II: GHZ scaling on ideal and two noisy simulators

GHZ circuits with $n \in \{2,3,\ldots,20\}$ qubits were studied on three backends: an ideal simulator, a noisy simulator derived from ibm_torino, and a noisy simulator derived from ibm_brisbane. Fidelity was defined as $F = 1 - d(p,q)$, where $d(p,q)$ is the total variation distance between the noisy and ideal outcome distributions. $F=1$ represents perfect agreement with the ideal distribution, and $F=0$ indicates no overlap.

The results show that the ideal simulator maintains fidelity near 1 across all qubit counts, as expected. Both noisy simulators display monotonic degradation in fidelity as n increases. For torino, fidelity drops from ~0.98 at n=2 to ~0.52 at n=20. Brisbane shows a similar trend, decreasing from ~0.97 at n=2 to ~0.45 at n=20. This demonstrates how larger GHZ circuits amplify the effects of device noise.

Resource analysis confirms the cause of this trend. The number of two-qubit (CX) gates and the circuit depth both grow roughly linearly with n, reaching over 50 CX gates and depths exceeding 150 at n=20. When fidelity is plotted directly against the number of CX gates, we see a clear negative correlation: more entangling gates lead to greater noise accumulation and lower fidelity. Brisbane circuits transpile to deeper circuits on average, which explains why its fidelities are slightly lower than Torino despite having lower gate error rates at the device level.

The relative fidelity plots, defined as F_noisy/F_ideal, further illustrate the scalability challenge. Relative fidelity falls below 0.8 by n≈8 and drops below 0.5 by n≈18. A common acceptance threshold of $F \geq 0.90$ is met only up to n≈4–5 for both devices, beyond which the circuits become too noisy to reliably represent the GHZ state.

In summary, Part II demonstrates that while small GHZ circuits can be simulated with high fidelity, scaling quickly reveals the limitations of noisy devices. The number of two-qubit gates and circuit depth drive the fidelity decay, and even devices with relatively low gate error rates cannot maintain high fidelity for GHZ states beyond ~5–6 qubits.

Part III: Fidelity at Scale (ESP vs measured)

Circuit reliability without simulation using Estimated Success Probability (ESP) was estimated, compounding per-gate success probabilities from device calibrations after transpilation: ESP = ∏(1−e_g), where e_g is the device-reported gate error for gate g on the specific qubits used. We applied this to GHZ(n) for n=2…20 on ibm_torino and ibm_brisbane, then compared ESP to the measured simulator fidelities from Part II (defined as F=1−TVD to the ideal GHZ distribution).

ESP decreases with n for both devices, tracking the overall trend seen in the measured fidelities, but it systematically overestimates performance. Quantitatively, the average (ESP − measured) gaps are 0.178 (median 0.206) for torino and 0.170 (median 0.157) for brisbane. This bias is expected because ESP multiplies gate-level success probabilities and ignores several real effects that degrade performance in practice: readout error, decoherence during idle time (T1/T2 over circuit depth), crosstalk/correlated errors, calibration drift, and routing overhead not fully captured by static gate-error numbers.

Across n, torino's ESP curve is often comparable to or slightly higher than brisbane's despite brisbane's lower per-gate error rates from Part I. This is explained by transpilation cost: brisbane's layouts produced more CX gates and larger depths for the same GHZ(n), so the product of (1−e_g) compounds over more two-qubit operations and reduces ESP. The measured fidelities show an even steeper decline, reinforcing that two-qubit gate count and depth dominate scalability and that ESP is optimistic relative to reality.