For an n-qubit GHZ state, say we measure the all-X observable, ideally without any noise, the measurement distribution would be a uniform superposition of all states with even 1's. There are 2^(n-1) such even-Hamming-weight states. In practice, our shot count is much lower than 2^(n-1) when n is large, since we usually just have 10k shots, and it is very easy to have a gigantic number of 2^(n-1) for some large n. Therefore, every shot in practice is going to produce a unique bitstring with close-to-unit probability. Therefore, our measured bitstring must be a uniform distribution over the 'measured bitstrings'.

Then, as argued and intuitively evident in our Appendix D, readout error mitigation becomes ineffective as a statistical method.

Now, unable to mitigate readout errors at all, one can be convinced that to get a measured (noisy) bit-string with an odd Hamming weight would be to have an odd number of flips (readout errors) on any of the equally probable, even-Hamming-weight (ideal) bit-string. Assuming reasonably that for each qubit there is a symmetric 1% misclassification rate (this is a typical, and even good, measurement error rate for current-generation hardware), this amounts to compute the probability for a binomial variable X~Binomial(n, 0.01) to be odd, i.e., Pr(X~Binomial(n, 0.01) is odd).

The probability for a binomial variable X~Binomial(n, p) to be odd is:

$$\Pr\{X \text{ odd}\} = \frac{1}{2}\left[1-(1-2p)^n\right]$$

The derivation is as follows

$$\begin{aligned}
((1-p)+p)^n &= \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \\
&= \sum_{k=0}^{\lfloor n/2 \rfloor} \binom{n}{2k} p^{2k} (1-p)^{n-2k} + \sum_{k=0}^{\lfloor n/2 \rfloor} \binom{n}{2k+1} p^{2k+1} (1-p)^{n-(2k+1)} \\
&= \Pr\{X \text{ even}\} + \Pr\{X \text{ odd}\} \\
((1-p)-p)^n &= \sum_{k=0}^{n} \binom{n}{k} (-p)^k (1-p)^{n-k} \\
&= \sum_{k=0}^{\lfloor n/2 \rfloor} \binom{n}{2k} p^{2k} (1-p)^{n-2k} - \sum_{k=0}^{\lfloor n/2 \rfloor} \binom{n}{2k+1} p^{2k+1} (1-p)^{n-(2k+1)} \\
&= \Pr\{X \text{ even}\} - \Pr\{X \text{ odd}\}
\end{aligned}$$

Therefore, for p=0.01, we have

n	Probability of odd flips
1	0.0100
2	0.0198
8	0.0774
32	0.3180
128	0.4961

For example with n=32, we get about ~31% out of the, say 10k, measured bit-strings which will be evaluated to have an expectation value of -1. Hence, the total expectation value will be 0.69 *(+1) + 0.31*(-1) = 0.38 instead of a perfect 1 for the noiseless case.

This shows why the fidelity estimation of an n-qubit GHZ state by sampling its stabilizers [39,50,59,70] underestimates the fidelity much more strongly.