On The Cost of Fault-Tolerainzing Shallow Circuits.

Michael Ben-Or David Ponarovsky

November 7, 2024

Abstract

In this work we study the overall depth overhead cost required for constructing fault tolerance circuits. We focus on shallow depth circuits classes, In particular, $\mathbf{QAC}_0$, $\mathbf{QNC}_{0,f}$ and $\mathbf{QNC}_1$ and certain knowns problem candidates for demonstrating quantum advantage such as factoring [Sho97] and Instantaneous Quantum Polynomial-time [BMS17], [Pal+24]. We only give a partial answers, Yet, clues that might pave the way towards a full understanding the complexity versus fault tolerance trade-off.

1 Introduction.

The question about the feasibility of computation under noise is almost as ancient as the computer science field itself, initialized by Von Neumann [Neu56] at the time that classical computation putted in debuts. Time been pass and the followed works had pointed that not even a polynomial computation in the presence of noise is still reasonable but one can implement a fault tolerance version at a most constant times cost at the circuits depth [Pip85]. Or in asymptotic sense, classical computation in the presence of noise is as exactly hard as computation in ideal environment.

Once again, the feasibility question raised again, this time regarding quantum computing, and while an intensive work has been done, and also succeed to prove that polynomial quantum computation can be made fault tolerance, [AB99],[Got14] and even with only constant overhead at the original circuit width [Gro19], the required depth over-head is till not well understood. We stress out that in all the familiar constructions, in construct to Pippenger [Pip85], original constant-depth gates are mapped to asymptotically grow[1] depth gates.

This work address the above, We ask whether a magnitude depth overhead is an unavoidable price that one has to pay. And, in particular, whether an ideal $\mathbf{QNC}_1$ circuits can be computed in noisy-$\mathbf{QNC}_1$ circuits. We show how using the ideas presented in [Gro19] and [Pip85] gives almost immediately $\mathbf{QNC}_{0,f} \subset$ noisy-$\mathbf{QNC}_1$ and that sampling from IQP [BMS17] also can be done in logarithmic depth circuits.

2 Notations.

We denote by C_g the good qLDPC code [Din+22] [PK21] [LZ22b], and by C_{ft} the concatenation code presented at [AB99] (ft stands for fault tolerance). For a code C_y, we use Φ_y, E_y, D_y to denote the channel maps circuits into the their matched circuits compute in the code space, the encoder, and the decoder, respectively. We use Φ_U to denote the 'Bell'-state storing the gate U. We say that a state $|\psi\rangle$ is at a distance d from a quantum code C if there exists an operator U that sends $|\psi\rangle$ into C such that U is spanned on Paulis with a degree of at most d. Sometimes, when the code being used is clear from the context, we will say that a block B of qubits has absorbed at most d noise if the state encoded on B is at a distance of at most d from that code.

[1] Note, that here, classical computation is also counted in the overall depth cost

3 Notations, Definitions and Construction.

The notation used in this paper follows standard conventions for coding theory. We use n to represent the length of the code, k for the code's dimension, and ρ for its rate. The minimum distance of the code will be denoted as d, and the relative distance, i.e., d/n, as δ. In this paper, n and k will sometimes refer to the number of physical and logical bits. Codes will be denoted by a capital C followed by either a subscript or superscript. When referring to multiple codes, we will use the above parameters as functions. For example, $\rho(C_1)$ represents the rate of the code C_1. Square brackets are used to present all these parameters compactly, and we use them as follows: $C = [n, k, d]$ to declare a code with the specified length, dimension, and distance. Any theorem, lemma, or claim that states a statement that is true in the asymptotic sense refers to a family of codes. The parity check matrix of the code will be denoted as H, with the rows of H representing the parity check equations. The generator matrix of the code will be denoted as G, with the rows of G representing the basis of codewords. The syndrome of a received word will be denoted as s, which is the result of multiplying r by the transpose of H. We use $C^\perp$ to denote the dual code of C, which is defined such that any codeword of it $z \in C^\perp$ is orthogonal to any $x \in C$, meaning $z \cdot x = 0$, where the product is defined as $x \cdot z = \sum_i x_i z_i$. $C^\top$ stands for the code obtained by taking the parity check matrix of C and transposing it.

In this paper, we define the triple product $\mathbb{F}_2^n \times \mathbb{F}_2^n \times \mathbb{F}_2^n \to \mathbb{Z}$ as $|x \cdot y \cdot z| = \sum_i^n x_i y_i z_i$. Similarly, we define the binary product $|x \cdot y|$, noting that this product differs from the standard product by mapping into $\mathbb{Z}$ rather than $\mathbb{F}_2$. For $w \in \mathbb{F}_2^n$, we use the super operator $\cdot|_w$ to map an operator originally defined in an n-dimensional space to an operator that only acts on coordinates restricted to w. For example, $x|_w$ is the vector in $\mathbb{F}_2^{|w|}$ obtained by taking the values of x on coordinates where w is not zero. $|x \cdot y|_w = \sum_{i:w_i \neq 0} x_i y_i$ and $C|_w$ is the code obtained by taking the codewords of C restricted to w.

Definition 3.1. *Let C, $\tilde{C}$ be linear binary codes at the same length, We will say that $\tilde{C}$ is a Triorthogonal with respect to C if:*

1. $\tilde{C} \subset C$
2. *$|x \cdot y \cdot z|$ is even for $x, y, z \in C$ such that at least one of x, y, z belongs to $\tilde{C}$.*
3. *$|x \cdot y|$ is even for $x, y \in C$ such that at least one of x, y belongs to $\tilde{C}$.*

If a code C is Triorthogonal with respect to itself then we will say that C is a self Triorthogonal code.

For example, the empty code, that contains only the zero code word, i.e $C = \{0\}$, is a Triorthogonal with respect to any code. In fact for proving **??** taking the empty code is sufficient. For other example, the Triorthogonal codes defined in [BH12] are Triorthogonal with respect to themself.

A quantum code over n qubits is an embedding of $\mathcal{H}_2^{\otimes k}$ as a subspace of $\mathcal{H}_2^{\otimes n}$. Similar to classical codes, we will call n and k the physical and logical qubits. The embeddings of states in $\mathcal{H}_2^{\otimes k}$ are called codewords or encoded states. In addition, we will use the term "logical operator" (i.e. logical X_i) to describe an operator that acts on the code space exactly as it would act on the logical space $\mathcal{H}_2^{\otimes k}$ (in our example, turning on and off the encoded state corresponds to the ith qubit exactly as X_i acts as Pauli X on the ith qubit in $\mathcal{H}_2^{\otimes k}$). We will denote by X and Z the single X and Z Pauli operators, by X_i the application of X on the ith qubit and nothing else (identity) on the rest of the qubits. By $X^{(v)}$ for some $v \in \mathbb{F}_2^n$, we mean the operator composed by applying X on each of the qubits whose index is a non-trivial coordinate of v and identity elsewhere. In a similar fashion, we define $Z^{(v)}$. When the context is clear, we will allow ourselves to omit the brackets, i.e. Z^v. The weight of a Pauli operator is the number of coordinates on which the operator acts non-trivially. Recall that the set of Pauli $+I$ spans all the Hermitian matrices. We say that the Pauli weight of an operator is the maximal weight of a Pauli in its Pauli decomposition. For example, consider the operator $A = IXX + ZII$, the weight of A is 2. The distance of a quantum code is the minimal weight of an operator that takes one codeword to another. We use the standard bracket notation to describe quantum states and in addition, we define for a vector space $A \subset \mathbb{F}_2^n$ the notation $|A\rangle$ to represent the uniform superposition of all the vectors belonging to that space, namely:

$$|A\rangle = \frac{1}{\sqrt{|A|}} \sum_{x \in A} |x\rangle$$

We define in the same way the notation to hold for affine spaces, $|x + A\rangle$. We will use $\propto$ to denote a quantum states up to normalization factor, for example $|\psi\rangle \propto |0\rangle + |1\rangle$ means that $|\psi\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$. A CSS code is a quantum code defined by a pair of classical codes C_X and C_Z, satisfying $C_Z^\perp \subset C_X$, such that any codeword of it has the form $|x + C_Z^\perp\rangle$, where $x \in C_X$. We will use Q to refer to a CSS

code in general and use $C_X/C_Z^\perp$ to refer to the vectors associated with the X-generators or the encoded states in the computational basis. In the same way, $C_Z/C_X^\perp$ refers to the Q in the phase basis. We will say that a CSS code Q is a LDPC if C_X and C_Z are both LDPC codes. Our construction uses the classical Tanner code [Tan81], the expander codes [SS96], and Hyperproduct code (quantum expanders) [LTZ15], [TZ14], [BFS23]. We will not describe these constructions and refer the reader to those papers for further information.

4 Todo:

1. Move to encoding each qubit by logarithmic width (instead of chanks) the reason is that the gate teleportation becomes complicated when it applied over higher dimension.
2. Then showing for 2-qubit gates set that is indeed works.
3. Treating separately to noise observed in two qubits gates.

5 Fault tolerance Toffoli.

[COMMENT] In that section the $\cdot$ operation is the pair wise product (pair wise AND).

Assume that $\bar{0}, \bar{1} \in C_X$ and that they belong to two different cosets of $C_X/C_Z^\perp$. Let $x, y \in \{\bar{0}, \bar{1}\}$.

$$\begin{aligned}
&\sum_{z,z',w\in C_Z^\perp} |z\rangle\,|z'\rangle\,|w\rangle \\
&\sum_{z,z',w\in C_Z^\perp} |z\rangle\,|z'\rangle\,|w+z\cdot z'\rangle \\
&\sum_{z,z',w\in C_Z^\perp} |z+x\rangle\,|z'+y\rangle\,|w+z\cdot z'\rangle \\
&\sum_{z,z',w\in C_Z^\perp} |z+x\rangle\,|z'+y\rangle\,|x\cdot y+x\cdot z'+y\cdot z+zz'+w+z\cdot z'\rangle \\
&\sum_{z,z',w\in C_Z^\perp} |z+x\rangle\,|z'+y\rangle\,|x\cdot y+x\cdot z'+y\cdot z+w\rangle
\end{aligned} \tag{1}$$

Since $x, y \in \{\bar{0}, \bar{1}\}$ we have that $x \cdot z'$ equals to either z' or $\bar{0}$. Hence $\sum_{w\in C_Z^\perp} |\xi + x\cdot z + w\rangle = \sum_{w\in C_Z^\perp} |\xi + w\rangle$. So the idea is the following, suppose that one has to compute Toffoli at time t over the registers R_1, R_2, R_3. First, at time 0, he initialize a logical zero $|C_Z^\perp\rangle$ in each register, then he compute pairwise Toffoli R_1, R_2 into R_3. That gives the ket $\sum_{z,z',w\in C_Z^\perp} |z\cdot z' + w\rangle$, immediately afterwords encode R_3 again into a good quantum code. Denote by τ the time required for decoding R_3 back, at time $t-\tau$ start to decode R_3. Eventually at time time t compute again the transversal Toffoli, by Equation (1) we gets the desired.

By similar arguments exhibited at Claim 7.3 one can show that the errors behaves according to a Pauli noise channel. **[COMMENT]** That is not correct, since the concatenation construction assumes that all the registers initialized to physical zeros in the begging of the computation.

5.1 Another Idea, $z \cdot z'$ cann't contribute too mach.

Clearly we have that $|z\cdot z'| \leq |z|, |z'|$ therefore we have that $\mathbf{Pr}_{z,z'\in C_Z^\perp}[|z\cdot z'| \geq t] \leq \mathbf{Pr}_{z\in C_Z^\perp}[|z| \geq t]$. Now assume that the tanner code by which the code defined is bipartite graph and denote by z_+, z_- the grouping of the z's generators supported on the even and the odd vertices of the graph. By triangle inequality $|z| = |z_+ + z_-| \leq |z_+| + |z_-|$, So if $|z| > t$ then at least one of $|z_-|, |z_+|$ is greater than $t/2$. Hence via the union bound:

$$\mathbf{Pr}_{z\in C_Z^\perp}[|z|] \leq \mathbf{Pr}_{z\in C_Z^\perp}\left[\bigcup_{i\in\pm} |z_i| \geq t/2\right] \leq \sum_{i\in\pm} \mathbf{Pr}_{z\in C_Z^\perp}[|z_i| \geq t/2]$$

Since any two positive (negative) generators are disjoint we have that $|z_+|$ is a sum of the independent random variables each stands for the weight contributed by a positive vertex. Let us denote by V^+, V^-

the positive and the negative vertices and for each vertex $v \in V$ we will denote by $_v$ the bits of z restricted to v edges. So $|z_\pm| = \sum_{v\in V^\pm} |z_v|$. For simplicity assume that $|V^+| = |V^-| = n/2$ and that $\mathbf{E}_{z\in C_A\otimes C_B}[|z|] = \mu$. Then we can use concentration inequality to have:

$$\mathbf{Pr}_{z\in C_Z^\perp}[|z|] \leq \sum_{i\in\pm} \mathbf{Pr}_{z\in C_Z^\perp}\left[\sum_{v\in V^i} |z_v| \geq t/2\right] \leq 2e^{-(\mu-\frac{t}{2})n}$$

Thus if $\mu - \gamma \geq O(1)$ (from Claim 7.2) then with high probability the Toffoli is computed up to reducible error.

6 The Noise Model

7 Fault Tolerance (With Resets gates) at Linear Depth.

Claim 7.1. *There exists a value $p_{th} \in (0,1)$ such that if $p < p_{th}$, then any quantum circuit C with a depth of D and a width of W can be computed by a p-noisy circuit C', which allows for resets. The depth of C' is at most $\max\{O(D), O(\log(WD))\}$.*

7.1 Initializing Magic for Teleportation gates and encodes ancillaries.

The Protocol:

1. Initialization of zeros: The qubits are divided into blocks of size $|B|$. Each block is encoded in C_g using $D_{ft}\Phi_{ft}[E_g]\,|0^{|B|}\rangle$.
2. Initialization of Magic for Teleportation gates: The gates in the original circuit are encoded in C_g using $D_{ft}\Phi_{ft}[E_g]\,|\Phi_U\rangle$.
3. Gate teleportation: Each gate in the original circuit is replaced by a gate teleportation.
4. Error reduction: After the initialization step, at each time tick, each block runs a single round of error reduction.

Claim 7.2 (From [LZ22a])**.** *Assuming that an error $|e| \leq \gamma n$, i.e e is supported on less than γn bits, then a single correction round reduce e to an error e' such that $|e'| < \nu|e|$.*

Claim 7.3. *The gate $D_{ft}\Phi_{ft}[E_g]$ initializes states encoded in C_g subject to a $3p$-noise channel.*

Proof. Clearly, with high probability, $\Phi_{ft}[E_g]$ successfully encodes into $C_{ft}\circ C_g$, let's say with probability $1 - \frac{1}{poly(n)}$. Denote by E_i and D_i the encoder and decoder at the ith level of the concatenation construction. Consider the decoder under $\mathcal{N}$ action: $P_2D_1P_2D_2, .., P_{i-1}D_iP_i$, by the fault-tolerance construction, a logical error at the ith stage occurs with probability p^{2^i}. Therefore, by the union bound, the probability that in one of the steps the circuit absorbs an error that is not corrected is less than $p + p^2 + p^4 + .. < 2p$. Hence, any decoded qubit absorbs noise with probability less than $2p$.

Thus, overall, we can bound the probability of a single qubit being faulty by:

$$\begin{aligned}\mathbf{Pr}[\text{fault}] &= \mathbf{Pr}[\text{fault}|\Phi_{ft}[E_g]]\cdot\mathbf{Pr}[\Phi_{ft}[E_g]] + \mathbf{Pr}\left[\text{fault}|\overline{\Phi_{ft}[E_g]}\right]\cdot\mathbf{Pr}\left[\overline{\Phi_{ft}[E_g]}\right]\\ &\leq \mathbf{Pr}[\text{fault}|\Phi_{ft}[E_g]] + \mathbf{Pr}\left[\overline{\Phi_{ft}[E_g]}\right] \leq 2p + \frac{1}{poly(n)} \leq 3p\end{aligned}$$

Remark 7.1. *In our construction, we use the concatenation code to encode blocks of length $\log(n)$. Therefore, any $poly(n)$ in the above should be replaced by $\log(n)$. However, this does not affect anything since the inequality does not depend on n.*

□

Claim 7.4. *With a probability $1 - \frac{WD}{|B|}\cdot D2e^{-2|B|(\beta-p)}$, the total amount of noise absorbed in a block at any given time t, is less than γn.*

Proof. Consider the ith block, denoted by B_i. By applying Hoeffding's inequality, we have that the probability that more than $\beta|B|$ qubits are flipped at time t is less than $2e^{-2|B|(\beta-p)}$. By using the union bound over all blocks at all time locations, we can conclude that with probability $1 - \frac{WD}{|B|} \cdot D2e^{-2|B|(\beta-p)}$, the noise absorbed in a block is less than $|\beta|B$ for the entire computation.

Let X_t denote the support size of the error over B_i at time t. Using Claim 7.2, we can bound the total amount of error absorbed by a block until time t as follows:

$$X_t \leq \nu \cdot (X_{t-1} + \beta|B|) \leq \nu(\gamma + \beta)|B| \leq \gamma|B|$$

□

Claim 7.5. *The total depth of the circuit is* $O(D) + O(\log^c |B|)$.

Proof. The gate for encoding $|B|$-length blocks in C_g is a Clifford gate and can therefore be computed in $O(\log |B|)$ depth. The encoding of the magic/bell states is done by first computing them in the logical space (un-encoded qubits) and then encode them using the encoder. Hence, the fault-tolerant version of both initializing ancillaries and magic states/bell states costs $O((\log |B|) \cdot \log^c(|B| \log |B|))$ [2] depth [AB99]. Backing into C_g from C_{ft} by decoding the concatenation code takes exactly as long as the encoding, namely $O((\log |B|) \cdot \log^c(|B| \log |B|))$.

Then, using the bell measurements, any of the logical gates takes $O(1)$ depth. Since we only perform a single round of error correction, the remaining computation until the last decoding stage takes at most constant time of the original depth. Finally, we pay $O(\log |B|)$ for complete decoding. Summing all, we get:

$$\begin{aligned} &O(\log |B| \cdot \log^c(|B| \log |B|)) + O(D) + O(\log |B|) \\ =&O(D) + O(\log^c |B|) \end{aligned}$$

□

Assuming that W is polynomial in D, taking the block length to be $|B| = \log((W \cdot D)^c)$, as shown in Claim 7.4, results in a linear fault tolerance construction with a success probability of $1 - \frac{1}{\log^{c_2}(W \cdot D)}$. This means that the fault tolerance version of circuits in $\mathbf{QNC}_1$ has a logarithmic depth. Additionally, using the construction in [Aha+96] produces a polynomial fault tolerance circuit in the reversible gates setting. **[COMMENT]** We missed the fact that it requires non trivial classical computation to compute what gate should be applied after the gate teleportation (i.e $UPU^\dagger$).

References

[Neu56] J. von Neumann. "Probabilistic Logics and the Synthesis of Reliable Organisms From Unreliable Components". In: *Automata Studies*. Ed. by C. E. Shannon and J. McCarthy. Princeton: Princeton University Press, 1956, pp. 43–98. ISBN: 9781400882618. DOI: doi:10.1515/9781400882618-003. URL: https://doi.org/10.1515/9781400882618-003.

[Tan81] R. Tanner. "A recursive approach to low complexity codes". In: *IEEE Transactions on Information Theory* 27.5 (1981), pp. 533–547. DOI: 10.1109/TIT.1981.1056404.

[Pip85] Nicholas Pippenger. "On networks of noisy gates". In: *26th Annual Symposium on Foundations of Computer Science (sfcs 1985)*. 1985, pp. 30–38. DOI: 10.1109/SFCS.1985.41.

[Aha+96] D. Aharonov et al. *Limitations of Noisy Reversible Computation*. 1996. arXiv: quant-ph/9611028 [quant-ph]. URL: https://arxiv.org/abs/quant-ph/9611028.

[SS96] M. Sipser and D.A. Spielman. "Expander codes". In: *IEEE Transactions on Information Theory* 42.6 (1996), pp. 1710–1722. DOI: 10.1109/18.556667.

[2] The width of the original circuit is $|B|^2$ so the number of locations is $|B|^2 \cdot \log |B|$

[Sho97] Peter W. Shor. "Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer". In: *SIAM Journal on Computing* 26.5 (Oct. 1997), pp. 1484–1509. DOI: 10.1137/s0097539795293172. URL: https://doi.org/10.1137%2Fs0097539795293172.

[AB99] Dorit Aharonov and Michael Ben-Or. *Fault-Tolerant Quantum Computation With Constant Error Rate*. 1999. arXiv: quant-ph/9906129 [quant-ph].

[BH12] Sergey Bravyi and Jeongwan Haah. "Magic-state distillation with low overhead". In: *Physical Review A* 86.5 (2012), p. 052329.

[Got14] Daniel Gottesman. *Fault-Tolerant Quantum Computation with Constant Overhead*. 2014. arXiv: 1310.2984 [quant-ph].

[TZ14] Jean-Pierre Tillich and Gilles Zemor. "Quantum LDPC Codes With Positive Rate and Minimum Distance Proportional to the Square Root of the Blocklength". In: *IEEE Transactions on Information Theory* 60.2 (Feb. 2014), pp. 1193–1202. DOI: 10.1109/tit.2013.2292061. URL: https://doi.org/10.1109%2Ftit.2013.2292061.

[LTZ15] Anthony Leverrier, Jean-Pierre Tillich, and Gilles Zemor. "Quantum Expander Codes". In: *2015 IEEE 56th Annual Symposium on Foundations of Computer Science*. IEEE, Oct. 2015. DOI: 10.1109/focs.2015.55. URL: https://doi.org/10.1109%2Ffocs.2015.55.

[BMS17] Michael J. Bremner, Ashley Montanaro, and Dan J. Shepherd. "Achieving quantum supremacy with sparse and noisy commuting quantum computations". In: *Quantum* 1 (Apr. 2017), p. 8. ISSN: 2521-327X. DOI: 10.22331/q-2017-04-25-8. URL: http://dx.doi.org/10.22331/q-2017-04-25-8.

[Gro19] Antoine Grospellier. "Constant time decoding of quantum expander codes and application to fault-tolerant quantum computation". Theses. Sorbonne Université, Nov. 2019. URL: https://theses.hal.science/tel-03364419.

[PK21] Pavel Panteleev and Gleb Kalachev. *Asymptotically Good Quantum and Locally Testable Classical LDPC Codes*. 2021. DOI: 10.48550/ARXIV.2111.03654. URL: https://arxiv.org/abs/2111.03654.

[Din+22] Irit Dinur et al. *Good Locally Testable Codes*. 2022. DOI: 10.48550/ARXIV.2207.11929. URL: https://arxiv.org/abs/2207.11929.

[LZ22a] Anthony Leverrier and Gilles Zémor. *Decoding quantum Tanner codes*. 2022. arXiv: 2208.05537 [quant-ph]. URL: https://arxiv.org/abs/2208.05537.

[LZ22b] Anthony Leverrier and Gilles Zémor. *Quantum Tanner codes*. 2022. arXiv: 2202.13641 [quant-ph].

[BFS23] Nouédyn Baspin, Omar Fawzi, and Ala Shayeghi. *A lower bound on the overhead of quantum error correction in low dimensions*. 2023. DOI: 10.48550/ARXIV.2302.04317. URL: https://arxiv.org/abs/2302.04317.

[Pal+24] Louis Paletta et al. "Robust sparse IQP sampling in constant depth". In: *Quantum* 8 (May 2024), p. 1337. ISSN: 2521-327X. DOI: 10.22331/q-2024-05-06-1337. URL: http://dx.doi.org/10.22331/q-2024-05-06-1337.