Characterizing local noise in QAOA circuits

Jeffrey Marshall, Filip Wudarski, Stuart Hadfield and Tad Hogg

1 QuAIL, NASA Ames Research Center, Moffett Field, California 94035, United States of America
2 USRA Research Institute for Advanced Computer Science, Mountain View, California 94043, United States of America

E-mail: jmarshall@usra.edu

Keywords: QAOA, NISQ devices, quantum optimization

Abstract

Recently Xue et al (2019 arXiv:1909.02196) demonstrated numerically that QAOA performance varies as a power law in the amount of noise under certain physical noise models. In this short note, we provide a deeper analysis of the origin of this behavior. In particular, we provide an approximate closed form equation for the fidelity and expected cost in terms of the noise rate, system size, and circuit depth. As an application, we show these equations accurately model the trade off between larger circuits which attain better expected cost values, at the expense of greater degradation due to noise.

1. Introduction

We study noise in Quantum Approximate Optimization algorithm (QAOA) circuits [1]. We consider the original formulation of QAOA, with transverse field mixer, though a generalization to the Quantum Alternating Operator Ansatz exists [2]. Here, a QAOA circuit of \( N \) qubits is specified by a cost Hamiltonian diagonal in the computational basis:

\[
H_c = \sum_{i=1}^{N} h_i \sigma_i^z + \sum_{j=1}^{N} j_{ij} \sigma_i^+ \sigma_j^- + ... ,
\]

(1)

as well as 2\( d \) angles \((\gamma, \beta)\), \( d \) being the high-level depth of the circuit \( U \) (i.e., number of QAOA rounds):

\[
U(\gamma, \beta) = \prod_{k=1}^{d} e^{-i\beta_j H_{\gamma}} e^{-i\gamma_j H_{\gamma}}.
\]

(2)

The Hamiltonian \( H_{\gamma} \) with off-diagonal elements, known as the ‘mixing’ Hamiltonian, is given by \( H_{\gamma} = \sum_{i=1}^{N} \sigma_i^z \). We assume throughout the initial state is an equal superposition over all computational basis states, \(|\psi_0\rangle = |+\rangle^{\otimes N}, \text{with } |+\rangle = \frac{1}{\sqrt{2}} (|0\rangle + |1\rangle)\).

The goal of QAOA is to find approximate solutions to optimization problems phrased as equation (1), which can be achieved by finding ‘good’ angles \((\gamma, \beta)\) in order to minimize or maximize the expected cost,

\[
C(\gamma, \beta) = \langle \psi_0 | U^{\dagger} H_{\gamma} U | \psi_0 \rangle.
\]

(3)

Due to the repetitive nature of a QAOA circuit, it is amenable to a mathematical analysis studying the presence of noise, as a function of system size \( N \) and depth \( d \). In particular a noise channel which acts identically during each block can be used to extract certain scaling characteristics of performance. Following Xue et al [3], we study in more detail the behaviour of the cost expectation function equation (3), when a layer of local depolarizing noise acts after each QAOA round, as in figure 1. Xue et al demonstrated that under such models,

---

3 The ‘dots’ in equation (1) denote higher order terms in \( \sigma \), though in our simulations we only consider up to quadratic terms. In our simulations we use fully connected problems of various sizes, without local fields \((h_i = 0)\) where \( j_{ij} \in \{-1, 1\} \) randomly chosen. Optimization is done via several random trials of BFGS.
though gradients become flatter, the general landscape does not change much. This has also been observed in experiments at low depths [4, 5]. This simplification allows us to answer in more detail questions pertaining to how local noise causes deviations in the expected cost and what the relative trade-offs are with respect to system size, circuit size (depth) and the noise rate.

2. Noise model

We assume a model such that after each QAOA round a layer of local noise $E_p$ is applied to each qubit (see figure 1), of the form

$$E_p \rho = (1-p) \rho + \frac{p}{M} \sum_{j=1}^{M} K_j \rho K_j^\dagger,$$

(4)

where $p \in [0,1]^\frac{1}{2}$. With this freedom of $p$, any quantum map can be phrased as in equation (4), but we will focus on cases with $p < 1$ where there is a term proportional to the state $\rho$, which includes depolarizing and dephasing noise.

Generalizing to the case of $N$ qubits, we write $E_p^{(n)}$ for $n = 1, ..., N$ to denote the above noise channel acting on qubit $n$, and the identity channel on all other $N-1$ qubits. A single round of QAOA followed by a noise layer is therefore of the form

$$\rho_j = \prod_{n=1}^{N} E_p^{(n)}[U(\gamma, \beta) \rho_0 U(\gamma, \beta)^\dagger],$$

(5)

where $\gamma, \beta$ are the $d = 1$ angles, and $\rho_0 = |\psi_0\rangle \langle \psi_0|$. Note the noise layer can also be written $\prod_{n=1}^{N} E_p^{(n)} = \bigotimes_{n=1}^{N} E_p^{(n)}$.

We focus initially on $d = 1$. Let us define $|\psi_1\rangle = U|\psi_0\rangle$, the noiseless QAOA-1 output. For ease of notation (though this restriction is not necessary), let us assume the $K_j$ are unitary (e.g. depolarizing noise). We can interpret the map $E_p^{(n)}|\psi_1\rangle$ statistically, as in [3], as applying noise operator $K^{(n)}_j$ (defined as identity on all qubits, and $K_j$ on $n$-th qubit) with probability $\frac{p}{M}$, resulting in the quantum state $|\psi_j^{(n)}\rangle = K^{(n)}_j |\psi_1\rangle$. Adopting the notation

$$|\psi_j^{(n)}\rangle = |\psi^{(n_1, n_2, ..., n_{N-1})}\rangle = K^{(n_1)}_h K^{(n_2)}_h ... |\psi_1\rangle,$$

(6)

Equation (5) can be written as

$$\rho_j = \sum_{m=0}^{N} (1-p)^{N-m} \left( \frac{p}{M} \right)^m \sum_{\lambda \lambda} |\psi^{(n_\lambda, m_\lambda)}\rangle \langle \psi^{(n_\lambda, m_\lambda)}|,$$

(7)

where $\tilde{n}_m, \tilde{j}_m$ are length $m$ vectors, with each entry in $\tilde{n}_m$ distinct (i.e. no repeats), $m$ specifies the total number of noise operators acting; $m = 0$ is the case where no noise acts, and $m = N$ means every qubit has a noise operator applied. The second sum in equation (7) is over $M^m \times \binom{N}{m}$ unique terms, using that each index in $\tilde{j}$ can run from 1 to $M$ (repeats allowed), and that $\tilde{n}$ has elements ranging from 1 to $N$, but with the restriction each element is unique.

4 For now we assume no restriction on the $K_j$ other than that $E_p$ is a quantum map, and so to preserve the trace, $\sum_j K_j^\dagger K_j = M I$

5 For qubit-local depolarizing noise, $M = 4$, with $K_j = \sigma_\alpha$, the identity and three Pauli $x, y, z$ operators.

6 $K^{(n)}_j = I \otimes ... \otimes K_j \otimes I \otimes I$, with $K_j$ on the $n$-th qubit.
3. Approximate analytical expressions for QAOA performance under noise

Of interest is how \( \rho_1 \) compares to the ideal QAOA-1 output, \(|\psi_1\rangle\). Two key quantities are the fidelity between the two, and the difference in expected cost.

3.1. Fidelity

First, let us consider the fidelity, or overlap, between \( \rho_1 \) and \(|\psi_1\rangle\)

\[
F_1 = \langle \psi_1 | \rho_1 | \psi_1 \rangle. \tag{8}
\]

In [3], it is posited, and shown numerically for the parameters studied, that the fidelity fits to a generic polynomial \( F_1 = (1 - p)^N \). One aim of this note is to justify such an equation in the case of noise which acts locally on each qubit, though a similar analysis could be performed for \( k \)-local noise.

First we write, using equation (7) and (8),

\[
\sum_{n, n_1} \langle \psi_1 | \psi_1^{(n_1)} \rangle^2 = M^m \binom{N}{m} \langle \psi_1 | \psi_1^{(n_1)} \rangle^2, \tag{9}
\]

where the ‘overline’ represents the average (mean) value of the overlaps (i.e. over all possible \( n_{i_m}, n_{i_0} \)).

A reasonable assumption is that the average on the right depends only on the parameter \( m \), and can therefore be written as a function \( f_m = \langle \psi_1 | \psi_1^{(n_{i_0})} \rangle^2 \). With this, we can write the fidelity equation (8) as

\[
F_1 = \sum_{m=0}^{N} \binom{N}{m} (1 - p)^N m^m f_m. \tag{10}
\]

An example of this is shown in figure 2 for a typical problem instance.

It now becomes clearer where the binomial form discovered in [3] comes from. First note we have \( f_0 = 1 \) by definition, and we expect \( f_N \approx 2^{-N} \) (the overlap squared of two random states). While it is tempting to set here \( f_m = 2^{-m} \), as the scaling is in general noise dependent, we use the slightly less restrictive form (satisfying \( f_0 = 1 \))

and write instead \( f_m \approx 1 + \alpha (\kappa^{-m} - 1) \), and so, by the binomial theorem

\[
F_1 \approx 1 + \alpha \left[ 1 - p \left( \frac{\kappa - 1}{\kappa} \right)^N \right]. \tag{11}
\]

Due to the generality of the form \( f_m \), we expect that many systems of interest will follow equation (11), though we mention that in our simulations here, we only consider depolarizing noise.

Notice that this equation predicts a slightly different dependence on \( N \) compared to the assumed form of [3], though both are equivalent for sufficiently small noise rate \( p \ll 1 \). In this limit, one can relate the exponent \( \delta \) found in [3] as \( \delta = \frac{\alpha}{\kappa} \). For \( \alpha = 0.996, \kappa = 2.71 \) from figure 2, one obtains \( \delta = 0.63 \).
As demonstrated in the inset of figure 2, our formula (equation (11)) also applies to when the noise is dominant (e.g. \( p > 0.5 \)), and not just for ‘small’ \( p \). Notice that the full curve over \( p \in [0, 1] \) cannot be replicated by a formula \( (1 - p)^{15} \), which does not give the appropriate behaviour in the large \( p \) limit.

3.2. Cost

The expected cost for noisy QAOA-1 is given by

\[
C_{\text{noise}}^{(1)} = \text{Tr}[H_1 \rho_i].
\]  
(12)

For well chosen angles (\( \gamma, \beta \)) using QAOA to minimize, we expect the cost to increase under the noise, i.e. \( C_{\text{noise}}^{(1)} > \langle \psi_i | H_1 | \psi_i \rangle = C_{\text{ideal}}^{(1)} \). Reference [3] posits that the dependence can be written as

\[
C_{\text{noise}}^{(1)} = (1 - p)^p \eta^{N} C_{\text{ideal}}^{(1)},
\]  
(13)

for some \( \eta > 0 \), where we use, matching the condition in the large noise limit \( p \to 1 \), the Haar random expectation \( C_{\text{ideal}} = \frac{1}{2N} \text{Tr}[H_1] = 0 \) by equation (1). For well chosen angles, one expects \( C_{\text{ideal}}^{(1)} < 0 \).

Performing a similar analysis as above via equations (7), (12), we write

\[
\sum_{\rho_m} \langle \psi^{(\text{full})}_m | H_1 | \psi^{(\text{full})}_m \rangle = M^m \left( \frac{N}{m} \right) \langle \psi^{(\text{ideal})}_m | H_1 | \psi^{(\text{ideal})}_m \rangle.
\]  
(14)

For \( m = 0 \), the average term on the right is precisely \( C_{\text{ideal}}^{(1)} \), and for \( m = N \) it can be approximated by the Haar random expectation, which here is 0. As before, if this quantity on the right only depends (to a good enough approximation) on \( m \), we can replace it by a function \( c_m = \langle \psi^{(\text{ideal})}_m | H_1 | \psi^{(\text{ideal})}_m \rangle \). This structure suggests writing, as before, \( c_m \approx \alpha + \tilde{\alpha} \chi^{-m} \), where \( \alpha + \tilde{\alpha} \approx C_{\text{ideal}}^{(1)} \).

With this, again by the binomial theorem,

\[
C_{\text{noise}}^{(1)} \approx \alpha + (C_{\text{ideal}}^{(1)} - \alpha) \left( 1 - p \left( \frac{\chi - 1}{\chi} \right) \right)^N,
\]  
(15)

showing a slightly different form from [3]. In the small \( p \) limit however, one can extract the relation \( \eta = \frac{C_{\text{ideal}} - \alpha \chi^{-1}}{C_{\text{ideal}}} \). Using the values from figure 3, we get \( \eta = 0.28 \) for this example. Note that \( C_{\text{ideal}} \) in equation (15) typically also carries an \( N \) dependence.

As in the case of fidelity, our equation also works to good accuracy in nearly the full range \( p \in [0, 1] \), i.e. in particular where \( p > 0.5 \) is not a ‘small’ parameter, which is shown in the inset of figure 3.

3.3. QAOA-\( d \)

For QAOA with \( d \) rounds (QAOA-\( d \)) we can extend the above analysis, which sheds light onto the depth versus noise trade-off. One can write an exact expression for \( \rho_{d,b} \).

\[ \text{For other problem sets and/or noise model where } C_{\text{noise}} \text{ tends to a non-zero value under increasing noise, it is easy to modify equation (15).} \]
1. Introduction

Recent experiments of [4] have shown consistencies with our results, namely a clear reduction in performance with system-size for problems requiring SWAP gates, and similarly an observation of optimal circuit depth with respect to noise. For native problems not requiring SWAP gates (and hence presumably less noisy), another tradeoff was observed where for $N < 25$ the noise effect can be balanced by the optimal cost also decreasing in tandem.

Going forward it is worthwhile to compare more directly our theory to other noise types found in hardware, for example, two-local noise on two-qubit gates, as well as spatial correlations due to effects such as cross-talk. This could be achieved, for example, by artificially altering the noise in the circuit and studying the scaling, among other possible experiments.

Acknowledgments

We are grateful for support from NASA Ames Research Center, the AFRL Information Directorate under grant F4HBKC4162G001, the Office of the Director of National Intelligence (ODNI) and the Intelligence Advanced Research Projects Activity (IARPA), via IAA 145 483, and NASA Academic Mission Services, Contract No. NNA16BD14C. The views and conclusions contained herein are those of the authors and should not be

\[
\rho_d = \sum_{m=0}^{Nd} (1 - p)^{Nd-m} \left( \frac{p}{M} \right)^m \sum_{j_m \ell_m r_m} |\psi_{j_m\ell_m r_m}^{(i)}\rangle \langle \psi_{j_m\ell_m r_m}^{(i)}|,
\]

where the second sum has a new index, $\tilde{l}$ representing the QAOA layer (round) in which the noise operator acts. That is, the elements of the length $m$ vectors $(\tilde{l}_m, \tilde{r}_m, \tilde{j}_m)$ tell us respectively which layer, qubit and operator the noise is acting. The second sum is over $M^m$ terms. For notational transparency, the general form of the term in the second sum is:

\[
|\psi_{j_m\ell_m r_m}^{(i)}\rangle = K_{j_m}^{\tilde{l}_m} U_d ... K_{j_2}^{\tilde{l}_m} U_1 K_{j_1}^{\tilde{l}_m} U_1 |\psi_i\rangle,
\]

where $U_k$ is the ideal QAOA unitary in round $k$, and $K_{j_k}^{\tilde{l}_m} = K_{m=0}^{\tilde{l}_m} ... K_{m=m}^{\tilde{l}_m}$ are the noise operators acting in that round (with $\sum_k m_k = m$). Following the form of equations (11), (15), one would simply replace $'N'$ by $'Nd'$ for $F_d$ and $c^{(d)}$.

Of interest is the trade-off between performing a greater number of rounds, ideally obtaining a lower cost, but at the expense that there is additional noise acting to raise the cost. Clearly for large $p$ there is no benefit to going beyond $d = 1$ since the output will be close to the maximally mixed state. In figure 4 we see for a given $p$ there is an optimal choice of depth. Only for very small noise levels is it beneficial to go to large depths. Once the depolarizing probability exceeds 2% there is no reason to go beyond $d = 1$. For this example, if $p > 0.25$, the curves have inverted from the order at $p = 0$, i.e., here $d = 1$ is the optimal choice.

4. Discussion

Figure 4. Cost $C = G^{(d)}_{\text{min}}(p)$ as a function of depolarizing probability for various circuit depths $d$, for a typical $N = 6$ instance (see footnote 3). For larger values of $p > 0.3$ the curves do not cross again, and all converge to the Haar random cost of 0 as $p \to 1$. The black dash-dot line is the optimal cost (ground state), which is found to a good approximation for $d \geq 4$ in the noiseless case ($p = 0$).

Each colored dash line is from a non-linear least squares fit to a function of form $G^{(d)}_{\text{min}}$ (i.e. Equation (15) with $N \to Nd$), and matches the data (colored solid lines) very well.

We are grateful for support from NASA Ames Research Center, the AFRL Information Directorate under grant F4HBKC4162G001, the Office of the Director of National Intelligence (ODNI) and the Intelligence Advanced Research Projects Activity (IARPA), via IAA 145 483, and NASA Academic Mission Services, Contract No. NNA16BD14C. The views and conclusions contained herein are those of the authors and should not be
interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of ODNI, IARPA, AFRL, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purpose notwithstanding any copyright annotation thereon. We used QuTiP in our simulations [6], and our code is publicly available.

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary information files).

ORCID iDs

Jeffrey Marshall ▼ https://orcid.org/0000-0002-5011-1281
Filip Wudarski ▼ https://orcid.org/0000-0002-0911-8342
Stuart Hadfield ▼ https://orcid.org/0000-0002-4607-3921
Tad Hogg ▼ https://orcid.org/0000-0001-8452-399X

References

[1] Farhi E, Goldstone J and Gutmann S 2014 A quantum approximate optimization algorithm arXiv:1411.4028
[2] Hadfield S, Wang Z, O’Gorman B, Rieffel E G, Venturelli D and Biswas R 2019 From the quantum approximate optimization algorithm to a quantum alternating operator Ansatz Algorithms 12 34
[3] Xue C, Chen Z-Y, Wu Y-C and Guo G-P 2019 Effects of quantum noise on quantum approximate optimization algorithm arXiv:1909.02196
[4] Arute F et al 2019 Quantum approximate optimization of non-planar graph problems on a planar superconducting processor arXiv:2004.04197
[5] Abrams D M, Didier N, Johnson B R, da Silva M P and Ryan C A 2019 Implementation of the XY interaction family with calibration of a single pulse arXiv:1912.04424
[6] Johansson J R, Nation P D and Nori F 2012 QuTiP: An open-source python framework for the dynamics of open quantum systems Comput. Phys. Commun. 183 1760