Exploring superadditivity of coherent information of noisy quantum channels through genetic algorithms

Govind Lal Sidhardh,[1] Mir Alimuddin,[1,2] and Manik Banik[2]
[1]*School of Physics, IISER Thiruvananthapuram, Vithura, Kerala 695551, India*
[2]*Department of Theoretical Sciences, S. N. Bose National Center for Basic Sciences, Block JD, Sector III, Salt Lake, Kolkata 700106, India*

(Received 17 January 2022; revised 2 July 2022; accepted 6 July 2022; published 26 July 2022)

Machine learning techniques are increasingly being used in fundamental research to solve various challenging problems. Here we explore one such technique to address an important problem in the quantum communication scenario. While transferring quantum information through a noisy quantum channel, the utility of the channel is characterized by its quantum capacity. Quantum channels, however, display an intriguing property called superadditivity of coherent information. This makes the calculation of quantum capacity a hard computational problem involving optimization over an exponentially increasing search space. In this work, we first utilize a neural network *Ansatz* to represent quantum states, and then we apply an evolutionary optimization scheme to address this problem. We find regions in the three-parameter space of qubit Pauli channels where coherent information exhibits this superadditivity feature. We characterized the quantum codes that achieve high coherent information, finding several nontrivial quantum codes that outperform the repetition codes for some Pauli channels. For some Pauli channels, these codes display very high superadditivity of the order of 0.01, much higher than the observed values in other well-studied quantum channels. We further compared the learning performance of the neural network *Ansatz* with the raw *Ansatz* to find that in the three-shot case, the neural network *Ansatz* outperforms the raw representation in finding quantum codes of high coherent information. We also compared the learning performance of the evolutionary algorithm with a simple particle swarm optimization scheme, and we show empirical results indicating comparable performance, suggesting that the neural network *Ansatz* coupled with the evolutionary scheme is indeed a promising approach to finding nontrivial quantum codes of high coherent information.

DOI: 10.1103/PhysRevA.106.012432

I. INTRODUCTION

The primary goal of any communication protocol is the reliable transfer of information from one spacetime point to another. However, in almost all practical scenarios, the communication lines get interrupted with unavoidable noises, and hence their successful implementation crucially depends on mitigation of those undesirable noises. From a rigorous mathematical point of view, the problem of communication in the presence of noise was first addressed by Claude Shannon in his seminal 1948 paper [1]. Shannon modeled a noisy classical communication line, also called a classical channel, as a stochastic map. This map is from some random input variable used by a sender for encoding information to some output random variable received by a distant receiver. The utility of such a classical channel is then characterized by its capacity, which is given by the mutual information between input and output random variables—an entropic quantity—optimized over all possible input probability distributions. Interestingly, the additivity property of mutual information makes parallel uses of many identical copies of a channel equivalent to multiple uses of its single copy [2]. Furthermore, the devices used for storing and transferring information are classical objects whose working rules are fundamentally governed by classical physics.

Quantum information theory utilizes peculiar features of the quantum world to devise novel communication protocols that are advantageous over their classical counterpart [3], and in some cases do not have any classical analog [4–12]. While modeling the quantum communication scenario, classical input and output variables are replaced by respective input and output Hilbert spaces, and accordingly, the channel action is most generally described by a completely positive trace-preserving (CPTP) map from the operator space on input to operator space on output Hilbert space [13]. Depending on what purpose a quantum channel is used for, its merit of utility is quantified by different quantities. For instance, the classical capacity of a quantum channel quantifies how good it is to transfer classical information [14,15]. In contrast, the private capacity captures its efficacy to transfer classical information in a secure way from a sender to a receiver so that no eavesdropper can know it [16]. On the other hand, the quantum capacity of a quantum channel captures its utility to transfer quantum information [16–18]. While the quantum capacity is defined through an entropic quantity—coherent information of its multiple uses in the asymptotic limit—obtaining a single-letter expression for this quantity is extremely difficult in general. Unlike the classical capacity of a classical channel, the quantum capacity of a quantum channel can exhibit a striking superadditive

phenomenon,[1] where two different channels each with zero quantum capacity can be used together to send quantum information with nonzero rate [22–24]. Later, it was shown that there exists a quantum channel whose N-parallel use cannot send quantum information, whereas its $(N+1)$-parallel use becomes useful for transferring quantum information, with N taking arbitrary integer values [25]. The superadditivity phenomenon makes the study of quantum communication more interesting as parallel use of many copies of a channel can be more beneficial than multiple uses of a single channel. This phenomenon of superadditivity has been explicitly demonstrated in the case of the one-parameter family of depolarizing channels, for which the repetition codes are known to display superadditivity at three channel-use and five channel-use level [26–29]. More recently, superadditivity of coherent information was demonstrated for two channel-uses for the two-parameter family of dephrasure channels, again using the repetition code [30,31]. Recently, quantum codes based on w-states were shown to display superadditivity for generalized erasure channels [32]. However, for more general classes of quantum channels, such simple quantum codes may not display the superadditivity feature, and therefore one needs to enlarge the search over the code space.

In this work, we utilize a machine learning-based optimization technique, as pioneered in Ref. [26], to explore superadditivity of coherent information for a general class of qubit Pauli channels that includes a depolarizing channel as a special case. This study of Pauli channels is important because of its considerable generality and the practical utility of modeling more general quantum noises as Pauli channels using techniques like Pauli projection [33] in quantum error correction applications [34]. For the class of qubit depolarizing channels, no two-qubit state is known to display superadditivity of coherent information [30]. Does this carry over to the class of generalized Pauli channels as well? If not, in what regions of the three-parameter space can this superadditivity be proven. In this work, we attempt to address these questions by building on the ideas from Ref. [26]. Specifically, we use genetic algorithms (GAs) on the neural network *Ansatz* to map out regions of superadditivity and characterize quantum codes of high coherent information for the three-parameter family of qubit Pauli channels. In addition, we also investigate the learning performance of the scheme by comparing it with (i) cases in which a raw *Ansatz* is used instead of the neural network *Ansatz*; and (ii) other cases in which particle swarm optimization is used instead of GAs.

The remaining sections of the paper are organized as follows: In Sec. II A, a brief introduction to the problem of quantum capacities is given. A brief introduction to GAs and methods are given in Secs. II B and II C, respectively. The main findings in our work have been presented in Sec. III. Lastly, we present our conclusions in Sec. IV. A detailed analysis of numerical learning performance is given in Appendix A.

[1]Importantly, classical capacity and private capacity of quantum channels also exhibit superadditive behavior [19–21].

II. PRELIMINARIES

In this section, we first review the idea of superadditivity of coherent information of a quantum channel, and then we will briefly discuss some concepts in machine learning that will be relevant to our purpose.

A. Superadditivity of coherent information and quantum capacity

Before moving to the quantum case, let us recall that a noisy classical channel is a stochastic map $\mathcal{N}: P(X) \to P(Y)$, where $P(X)$ and $P(Y)$, respectively, denote probability distributions on the input and output random variables X and Y. Shannon's noisy channel coding theorem [1] specifies the capacity of such a channel as

$$\mathcal{C}(\mathcal{N}) = \max_{P(X)} H(X:Y), \tag{1}$$

where $H(X:Y) := H(X) + H(Y) - H(XY)$ is the mutual information between the input and output distributions, with $H(X) := -\sum_{x\in X} p(x)\log_2 p(x)$ being the Shannon entropy, and the optimization is over the input probability distributions $P(X) \equiv \{p(x) \mid p(x) \geqslant 0 \text{ and } \sum_{x\in X} p(x) = 1\}$.

In a quantum scenario, a communication process is more generally described by a channel (CPTP map) $\Lambda: \mathcal{L}(\mathcal{H}_{\text{in}}) \to \mathcal{L}(\mathcal{H}_{\text{out}})$, where $\mathcal{L}(\mathcal{Z})$ is the operator space acting on the Hilbert space $\mathcal{Z}$, and $\mathcal{H}_{\text{in}}, \mathcal{H}_{\text{out}}$ are the respective Hilbert spaces for the input and output systems of the channel. Complete positivity of Λ assures positivity of the map when it is applied on the part of a composite system, i.e., $(\Lambda_A \otimes \mathbb{I}_R)\rho_{AR} \geqslant 0 \,\forall\, \rho_{AR} \in \mathcal{P}(\mathcal{H}_A \otimes \mathcal{H}_R)$, where $\mathcal{P}(\mathcal{Z}) \subset \mathcal{L}(\mathcal{Z})$ denotes the set of positive operators acting on $\mathcal{Z}$, and $\mathbb{I}_z$ denotes the identity map on $\mathcal{L}(\mathcal{Z})$. Furthermore, the trace-preserving condition implies $\text{Tr}[\Lambda_z(\rho)] = 1 \,\forall\, \rho \in \mathcal{D}(\mathcal{Z}) \subset \mathcal{P}(\mathcal{Z})$, where $\mathcal{D}(\mathcal{Z})$ denotes the set of density operators acting on $\mathcal{Z}$. The quantum capacity $Q(\Lambda)$ of a quantum channel Λ is given by the following regularized expression [16–18]:

$$Q(\Lambda) := \lim_{n\to\infty} \frac{1}{n} Q^{(1)}(\Lambda^{\otimes n}), \tag{2}$$

where

$$Q^{(1)}(\Lambda^{\otimes n}) := \max_{\rho_S} \mathcal{I}(\rho_S, \Lambda^{\otimes n}) \tag{3}$$

is the n-shot channel capacity, where n identical copies of Λ are used in parallel. The maximization in Eq. (3) is over all valid quantum states $\rho_S \in \mathcal{D}(\mathcal{H}_{\text{in}}^{\otimes n})$. In general, ρ_S's are mixed states, although in calculations it is often convenient to introduce a purified state $|\Psi_{SR}\rangle \in \mathbb{C}^{2n}$ such that $\text{Tr}_R(|\Psi_{SR}\rangle\langle\Psi_{SR}|) = \rho_S$. With this purified state, the expression for the quantity $\mathcal{I}$, called coherent information of $\Lambda^{\otimes n}$ for the input ρ_S, reads

$$\begin{aligned}\mathcal{I}(\rho_S, \Lambda^{\otimes n}) &\equiv \mathcal{I}(|\psi_{SR}\rangle, \Lambda^{\otimes n}) \\ &= S(\Lambda^{\otimes n}(\rho_S)) - S(\Lambda^{\otimes n} \otimes \mathbb{I}_R(|\psi_{SR}\rangle\langle\psi_{SR}|)).\end{aligned} \tag{4}$$

Here $S(\rho) = -\text{Tr}\,(\rho\log_2(\rho))$ is the Von Neumann entropy [3], and $\mathbb{I}_R$ is the identity channel on the purifying environment R. Coherent information of a channel Λ is then defined as $\mathcal{I}(\Lambda) \equiv \max_{|\psi_{SR}\rangle} \mathcal{I}(|\psi_{SR}\rangle, \Lambda)$.

Note that the expression of capacity in Eq. (1) for a classical channel follows from a similar regularized formula by invoking the additivity of mutual information, i.e., $\mathcal{C}(\mathcal{N}^{\otimes n}) = n\mathcal{C}(\mathcal{N})$ [2]. This greatly simplifies the optimization problem over an unbounded number channel use to a tractable one in the single-letter space. One might hope that such a simplification might exist for the quantum scenario as well; however, that is not the case. There are a few classes of channels that exhibit weak additivity, like the degradable channel, for which the quantum capacity reduces to coherent information [16], but this is not true in general. In fact, one can find channels that have strictly superadditive coherent information, i.e., $\mathcal{I}(\mathcal{N}^{\otimes n}) > n\mathcal{I}(\mathcal{N})$ [22–26,35]. This is quite interesting, as parallel use of multicopy channels with properly constructed quantum codes can yield more transmission rates than using the channel multiple times. Unfortunately, this appears with a downside; the optimization needs to be carried out over an unbounded number of channel uses, rendering the problem intractable. Yet all is not lost, as it is known that $\frac{1}{n}Q^{(1)}(\Lambda^{\otimes n})$ is lower bound to the quantum capacity [16–18], i.e.,

$$Q(\Lambda) \geqslant \frac{1}{n}Q^{(1)}(\Lambda^{\otimes n}) \geqslant \frac{1}{n}\mathcal{I}(|\psi_{SR}\rangle, \Lambda^{\otimes n}) \tag{5}$$

for any purified quantum state $|\psi_{SR}\rangle$. Therefore, even though Eq. (2) cannot be solved exactly in many cases, by finding quantum codes with high coherent information, Eq. (5) can be used to get better lower bounds to the quantum capacity of the channel. However, finding such high coherent codes becomes difficult as n increases because the space becomes exponentially large and more complex. Therefore, instead of relying on raw representations and systematic search, one has to switch to some form of variational *Ansatz* and metaheuristic search mechanism.

In this work, we restrict our study in qubit channels, i.e., channels whose input and output systems are qubits, $\Lambda : \mathcal{L}(\mathbb{C}^2) \to \mathcal{L}(\mathbb{C}^2)$. We are interested in a three-parameter family of such channels called the Pauli channel, and they are defined as

$$\begin{aligned}
&\Lambda(\rho) = p_0\sigma_0\rho\sigma_0 + p_1\sigma_1\rho\sigma_1 + p_2\sigma_2\rho\sigma_2 + p_3\sigma_3\rho\sigma_3,\\
&\sigma_0 := \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \sigma_1 := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},\\
&\sigma_2 := \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}, \quad \sigma_3 := \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix};\\
&\text{with } p_i \geqslant 0\,\forall\, i, \ \sum_{i=0}^{3} p_i = 1, \ \text{and } \rho \in \mathcal{D}(\mathbb{C}^2).
\end{aligned} \tag{6}$$

When a quantum state ρ is sent through this channel, physically, the effect can be thought of as if the Pauli σ_i gate is applied on the system's state with probability p_i. We are interested in mapping out the regions in the parameter space where the 2-shot quantum capacity $[Q^{(1)}(\Lambda^{\otimes 2})]$ is provably superadditive. Starting with a neural network *Ansatz* to represent the quantum state $|\psi_{SR}\rangle$ and by following a simple evolutionary strategy for optimization, we found regions in three-parameter space for which coherent information is superadditive. Before dwelling on the details of the result, we quickly review some of the fundamentals of genetic algorithms.

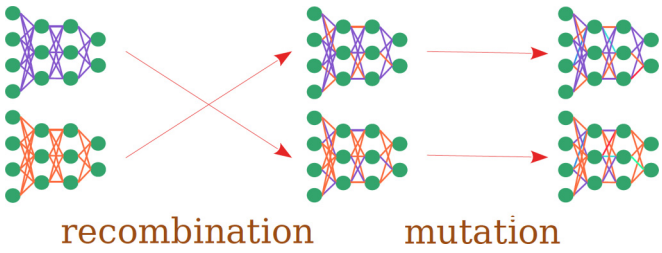


FIG. 1. Evolution of neural networks: An illustration showing how neural network states evolve to create new generations.

B. Genetic algorithms

At their core, machine learning techniques attempt to find patterns hidden in data and to exploit these learned patterns in control and decision-making. With the advent of deep learning, complemented by the ever-increasing computational power, machine learning approaches have solved problems that were once thought to be impossible. One of the most striking demonstrations of the power of modern machine learning/AI occurred in 2016 when Google's *AlphaGo* defeated the world champion in the game of *Go*. Apart from this, deep learning techniques have shown exceptional performance in problems like computer vision, natural language processing, medical diagnosis, and more. The success of modern machine learning in such diverse fields has prompted researchers to apply these techniques to fundamental research. Not surprisingly, many physicists have applied machine learning tools in their niche as well [36–40]. Interestingly, there have also been attempts toward developing methods for machine-assisted discovery of physical principles from experimental data [41].

In this work, we use neural networks (NNs) to represent quantum states, and these NNs are trained using genetic algorithms (GAs). GAs are a class of metaheuristic search/optimization algorithms inspired by the process of evolution by natural selection. All GAs start with a population of candidate solutions that are evolved over multiple generations to create better (fitter) solutions to the optimization problem. To simulate the various processes in natural selection, there are several biologically inspired operations involved in a GA. The typical GA begins with a randomly initialized set of candidate solutions. The candidate solutions are thought of as individuals of a population. Each individual is assigned a *fitness* that measures the quality of the individual. The population is left to evolve to produce fitter individuals and hopefully to generate a good enough solution to the optimization problem. Typically, the evolution process involves the following three fundamental steps (see Fig. 1):

(i) *Crossover/recombination*: Several individuals of the population are selected and are crossed over by an appropriately chosen scheme to produce new individuals. Individuals with high fitness are more likely to be crossed over. The intuition is that crossing over can bring together useful features of fit individuals to produce fitter individuals.

(ii) *Mutation*: The new individuals produced through crossover are subjected to mutation. Mutation introduces more variations into the population.

(iii) *Selection*: The fitter individuals are promoted to the next generation, while individuals with less fitness are

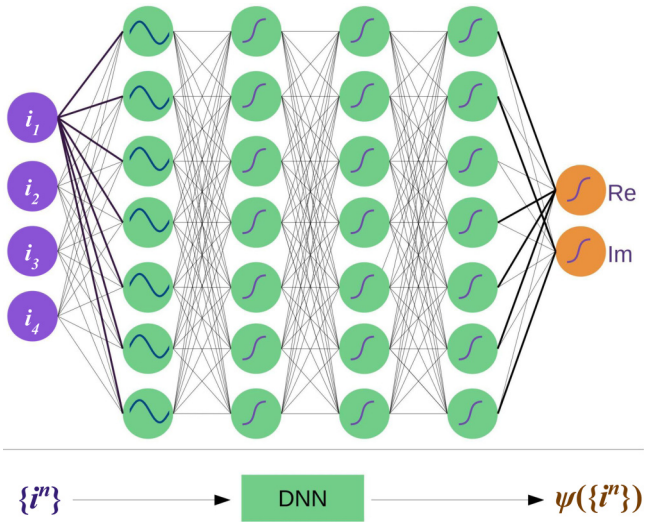


FIG. 2. The neural network *Ansatz*: A simple illustration of the neural network *Ansatz* used in this study. The given NN represents a 4-qubit quantum state. The inputs given to the NN are binary strings of length four and represent the corresponding computational basis states. The two outputs from the NN represent the real and imaginary parts of the amplitude corresponding to the input basis state. The first layer has a cosine activation, while all subsequent layers are given a tanh activation function.

removed from the population with high probability. GAs are derivative-free optimization methods, making them particularly useful for optimizing discontinuous objective functions and landscapes filled with local minima. Since GAs are good at avoiding local optima in complex optimization landscapes, they are often the tool of choice for global optimization problems. That said, GAs have their limitations. Since information like the gradient is not considered in the optimization process, GAs can be resource-hungry, requiring large population sizes and many fitness function evaluations. This can result in a substantial computational overhead if the fitness function evaluation is computationally expensive. In such cases, one often has to resort to parallel computing or computationally efficient approximations of the fitness function.

C. Methods

We represented quantum states using a NN *Ansatz* [42,43] as shown in Fig. 2. More specifically, this is a feed forward (FF) neural network, which we chose following its superior performance in benchmarking problems compared to other NN architectures like restricted Boltzmann machines, as reported in Ref. [26]. If there are n-qubits in the problem, a pure quantum state $|\Psi\rangle$ can be expanded in the computational basis as

$$|\Psi\rangle = \sum_{\{i^n\}} \frac{1}{C} \psi(\{i^n\}) |\{i^n\}\rangle, \tag{7}$$

where $\psi(\{i^n\})$ is the complex amplitude of the computational basis state $|\{i^n\}\rangle$ identified by the bit string $\{i^n\}$, and C is the normalization factor. In the NN *Ansatz*, the state $|\Psi\rangle$ is represented using a NN with n-input neurons and two output neurons. When fed in with a bit string $\{i^n\}$, the NN outputs two values corresponding to the real and imaginary parts of $\psi(\{i^n\})$. The weights of the NNs are the variational parameters in this representation that may be varied using an appropriate optimization technique. There are no formal rules about the choice of activation functions; however, a cosine activation for the first hidden layer and tanh activation for all subsequent hidden layers are known to work very well in practice [26,42], and they are employed in this work. We implemented the fully connected deep neural network architecture using the Sequential API of the TENSORFLOW package [44].

An immediate question that follows is whether using such a NN *Ansatz* is needed in the first place. Why not use a raw encoding (*Ansatz*), where the individual represents the coefficients of the state vector in a chosen basis, directly without going through the pain of introducing more parameters in the form of a neural network? This question has already been answered in Ref. [26], where they provided empirical evidence that showed the superior performance of the NN *Ansatz* compared to raw encodings. For completeness, we have also analyzed the performance of the NN *Ansatz* and found that for the two-shot problem, using a NN *Ansatz* might be overkill, but in the three-shot case, the NN *Ansatz* significantly outperforms the raw encoding. See Appendix for more details.

To fully define the optimization problem, we need a loss function (the function to be optimized). We formulate the problem as a minimization problem, and hence the loss function is simply the negative of the coherent information. Initially, we tried several gradient-based optimizers implemented in TENSORFLOW, including the standard "Adam" optimizer [45]. These techniques worked well in 2-qubit space, but as the space became large, all gradient-based algorithms almost always converged to some local minima with vanishing coherent information. This is expected since in the noisy regime, regions of positive, coherent information are likely to cover only a small fraction of the Hilbert space volume, and all product states ($|\Psi_{SR}\rangle = |\psi_S\rangle \otimes |\phi_R\rangle$) form local minima with vanishing coherent information. One trick to get around this problem is to add bias into the search space by penalizing product states. However, our analysis suggested that this is not a very robust approach.

When the optimization landscape is filled with local optima, a better strategy is to use some kind of gradient free-optimization methods like particle swarm optimization (PSO) or GAs. Reference [26] studied both PSO and a simple GA in benchmark problems and found that both methods show comparable performance. In this work, we choose to use a simple genetic algorithm as presented in Ref. [46]. The algorithm comes prebuilt with the DEAP (Distributed Evolutionary Algorithms in Python) package [47] as *deap.algorithms.eaSimple*. In addition to being an easy-to-use package, DEAP implementations are also easily parallelizable. For completion, we have also made a detailed comparison of the performance of the GA scheme we employed to a simple PSO variant, concluding that for the problem at hand, both schemes are comparable in terms of performance. We have also observed some interesting nuances in the behavior of these algorithms with variations in depth of the NN and other metaparameters. Detailed discussions can be found in Appendix.

Individuals (candidate solutions) in the population are represented as a list consisting of weight matrices of the neural

network. We employed a recombination scheme such that each of these weight matrices independently undergoes a two-dimensional (2D) crossover mechanism [48], where randomly chosen submatrices of each weight matrix are swapped with the corresponding submatrix of the other parent individual. In some sense, these weight matrices are similar to chromosomes in that recombination (crossover) happens only between weights of the corresponding layers of two individuals. We also added a Gaussian mutation of predetermined mean and standard deviation to the new individuals produced after crossover. Figure 1 gives a simplified illustration of the crossover and mutation process described above. We now have all the machinery necessary to find quantum codes with high coherent information for the Pauli channel.

III. RESULTS AND DISCUSSION

A. One-shot capacity of the Pauli channel

To prove the superadditivity of coherent information for a channel, one must know its one-shot quantum capacity $[Q^{(1)}(\Lambda)]$ exactly. For the depolarizing channel, which is a special case of the class of Pauli channels, it is known that the maximally entangled state maximizes the single-shot coherent information [49]. It turns out that this result also carries over to the class of Pauli channels as shown in Refs. [50,51]. Thus for a Pauli channel $\Lambda_{\vec{p}}$ as in Eq. (6), the one-shot quantum capacity is given by

$$Q^{(1)}(\Lambda_{\vec{p}}) = 1 - H(\vec{p}), \tag{8}$$

where $\vec{p} = (p_0, p_1, p_2, p_3)$ is the probability vector, and $H(\,)$ is the Shannon entropy of the probability distribution.

The *quantum code* that maximizes the single-shot coherent information is the maximally entangled state given by

$$|\Psi_{SR}\rangle = \frac{1}{\sqrt{2}}(|0\rangle_R \otimes |0\rangle_S + |1\rangle_R \otimes |1\rangle_S). \tag{9}$$

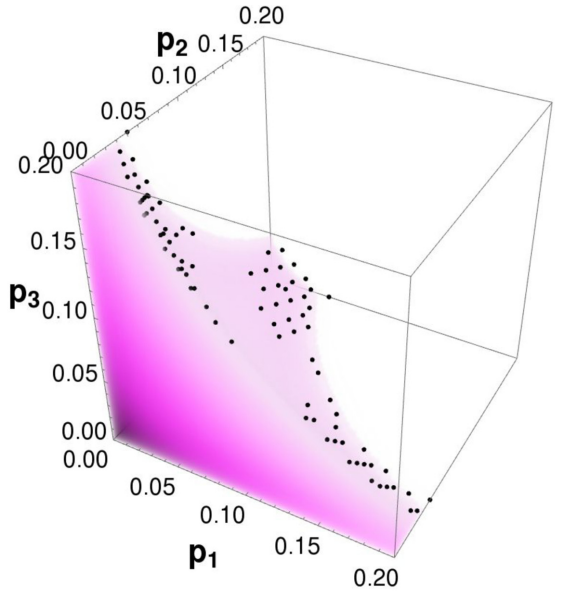
Figure 3 shows coherent information of the maximally entangled state for Pauli channels with parameters in the range $0 < p_i < 0.2$, $i \in \{1, 2, 3\}$. As seen from the figure, coherent information is high and positive in the low-noise regime. As the channel gets noisier, the single-shot coherent information quickly reduces to zero. We have explored this three-parameter space for the two channel-use and three channel-use case to find quantum codes that display superadditivity of coherent information.

B. Results of the GA-based optimization

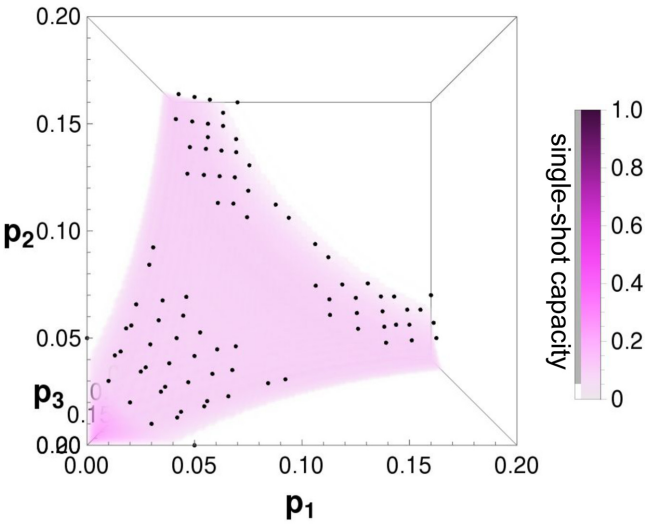
First we will consider the problem of finding good quantum codes with high two-shot coherent information. After analyzing results, presenting regions with superadditivity and the quantum codes with high two-shot coherent information, we move to the case of three-shot coherent information.

1. Optimizing two-shot coherent information

Quantum states were represented using a fully connected neural network with a 4-neuron input layer, 4 hidden layers with a width of 4 neurons each, and a 2-neuron output layer. The first hidden layer has a cosine activation function, and others have a "tanh" activation function. All optimizations



(a) Population size:50



(b) Population size:300

FIG. 3. Result of GA optimization: The single-shot capacity of the class of Pauli channels is shown in the density plot, color-coded as detailed on the color bar. Each point in the $p_1p_2p_3$ space represents a Pauli channel. We studied the superadditivity of points in a 3D grid of width 0.01. The black spots in the graph represent Pauli channels that display superadditivity of coherent information among these points.

were done using a simple evolutionary algorithm (described in [46]) as implemented in the DEAP package [47]. We set the crossover probability to 0.5, mutation probability to 0.2, and a population consisting of a total of 300 individuals evolved for 300 generations. Mutations were modeled as a simple Gaussian addition with a mean of 0.5 and a standard deviation of 0.25. Each attribute was allowed to mutate with a probability of 0.5. The selection operation was based on a tournament selection scheme with three individuals participating in each tournament. The averaged out learning curves of the

algorithm with different metaparameter settings can be seen in Appendix.

Optimization results were very promising with $\frac{1}{2}\mathcal{I}(|\Psi\rangle, \Lambda^{\otimes 2})$ almost always reaching at least as high as the one-shot quantum capacity, where $|\Psi\rangle \in (\mathbb{C}^4)^{\otimes 2} \equiv \mathbb{C}^2_{S_1} \otimes \mathbb{C}^2_{S_2} \otimes \mathbb{C}^2_{R_1} \otimes \mathbb{C}^2_{R_2}$. This is a good indication that the optimization algorithm is covering our desired region of the 4-qubit space and not getting stuck in most local optima. Interestingly, in many points in the search space, the algorithm found quantum codes $|\Psi\rangle \in (\mathbb{C}^4)^{\otimes 2}$ with coherent information strictly greater than one-shot capacity, i.e.,

$$\tfrac{1}{2}\mathcal{I}(|\Psi\rangle, \Lambda^{\otimes 2}(p_1, p_2, p_3)) > Q^{(1)}(\Lambda(p_1, p_2, p_3)), \tag{10}$$

hence establishing superadditivity of coherent information for those channels.

The set of Pauli channels for which the algorithm found superadditive quantum codes is shown as black spots in Fig. 3, along with the channel's one-shot capacity. As can be seen from the figure, there appears to be an apparent symmetry in the regions with superadditivity, with respect to the permutations of the parameters p_1, p_2, and p_3. As expected, we could not find any superadditivity for two channel-use along the main diagonal; the diagonal represents the family of depolarizing channels. To cross-check this observation, we ran a more refined optimization for the depolarizing channel for the two channel-use case, with a population of 500 individuals that evolved for 1000 generations; even then, the algorithm could not find any superadditive quantum codes. Therefore, this observation suggests that, most likely, two channel-uses cannot result in superadditivity for the family of depolarizing channels.

The next immediate step is to find the quantum codes that showed superadditivity in various regions; for all we know, the quantum states learned may vary from point-to-point displaying a nontrivial dependence on the three parameters. But surprisingly, after looking at the numerical value of the states and accounting for equivalences up to local unitaries in the purifying qubits, we found that the algorithm found only three distinct quantum codes that display superadditivity of coherent information. Moreover, the regions of superadditivity (shown in Fig. 4) separate into three distinct regions, depending on which quantum code

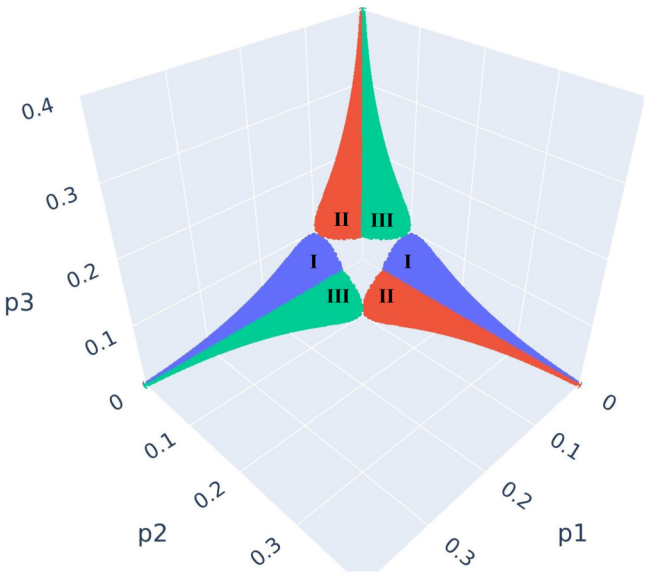


FIG. 4. Regions of two-shot superadditivity obtained from the three quantum codes: The blue region (marked with "I") shows the highest two-shot coherent information when code $|\Psi(\mathrm{I})\rangle$ is used. Similarly, the red region (marked "II") corresponds to code $|\Psi(\mathrm{II})\rangle$ and the green region (marked "III") corresponds to code $|\Psi(\mathrm{III})\rangle$. p_1, p_2, p_3 are, respectively, the probabilities of three Pauli errors.

gives maximal separation between one-shot and two-shot quantum capacity. The three (unnormalized) quantum codes the algorithm learned are given by

$$\begin{aligned}
|\Psi(\mathrm{I})\rangle &\equiv |zz\rangle|\bar{z}z\rangle + |zz\rangle|\bar{z}\bar{z}\rangle + |\bar{z}\bar{z}\rangle|zz\rangle + |\bar{z}\bar{z}\rangle|z\bar{z}\rangle;\\
|\Psi(\mathrm{II})\rangle &\equiv |xx\rangle|\bar{x}x\rangle + |xx\rangle|\bar{x}\bar{x}\rangle + |\bar{x}\bar{x}\rangle|xx\rangle + |\bar{x}\bar{x}\rangle|x\bar{x}\rangle;\\
|\Psi(\mathrm{III})\rangle &\equiv |yy\rangle|\bar{y}y\rangle + |yy\rangle|\bar{y}\bar{y}\rangle + |\bar{y}\bar{y}\rangle|yy\rangle + |\bar{y}\bar{y}\rangle|y\bar{y}\rangle;
\end{aligned} \tag{11}$$

where $|a\rangle$ ($|\bar{a}\rangle$) is the $+1$ (-1) eigenstate of the Pauli σ_a operator, $a \in \{x, y, z\}$. Note that $|\Psi(\mathrm{I})\rangle$ is invariant under the ZZ operator on the first two qubits. Similarly, the codes $|\Psi(\mathrm{II})\rangle$ and $|\Psi(\mathrm{III})\rangle$ are invariant under the action of XX and YY, respectively. The expression for two-shot quantum capacity when $|\Psi(\mathrm{I})\rangle$ is passed through the Pauli channel $\Lambda_{\vec{p}} \equiv \Lambda(p_1, p_2, p_3)$, with $\vec{p} = (p_1, p_2, p_3)$ and $p_0 = 1 - p_1 - p_2 - p_3$, is given by

$$\begin{aligned}
Q^{(2)}(|\Psi(\mathrm{I})\rangle, \Lambda_{\vec{p}}) = {} & p_1p_2\log_2(2p_1p_2) + p_0p_3\log_2(2p_0p_3) + \tfrac{1}{2}\left(p_1^2 + p_2^2\right)\log_2\left(p_1^2 + p_2^2\right) + \tfrac{1}{2}\left(p_0^2 + p_3^2\right)\log_2\left(p_0^2 + p_3^2\right)\\
& - (p_1 + p_2)(p_0 + p_3)\log_2((p_1 + p_2)(p_0 + p_3)) + (p_0p_2 + p_1p_3)\log_2(p_0p_2 + p_1p_3)\\
& + (p_0p_1 + p_2p_3)\log_2(p_0p_1 + p_2p_3) - \tfrac{1}{2}(p_0 + p_3)^2\log_2\left(\tfrac{1}{2}\left((p_1 + p_2)^2 + (p_0 + p_3)^2\right)\right)\\
& - \tfrac{1}{2}(p_1 + p_2)^2\log_2\left(\tfrac{1}{2}\left((p_1 + p_2)^2 + (p_0 + p_3)^2\right)\right).
\end{aligned} \tag{12}$$

Given Eq. (12), the symmetry of the quantum codes and that of the Pauli channel can be used to show that

$$\begin{aligned}
Q^{(2)}(|\Psi(\mathrm{II})\rangle, \Lambda(p_1, p_2, p_3)) &= Q^{(2)}(|\Psi(I)\rangle, \Lambda(p_3, p_2, p_1)),\\
Q^{(2)}(|\Psi(\mathrm{III})\rangle, \Lambda(p_1, p_2, p_3)) &= Q^{(2)}(|\Psi(I)\rangle, \Lambda(p_1, p_3, p_2)).
\end{aligned}$$

The regions in the three-parameter space of p_1, p_2, and p_3 that show superadditivity using these codes are depicted in Fig. 4. It should be pointed out that there are overlaps between regions that show superadditivity, for instance $|\Psi(\mathrm{I})\rangle$ also satisfies Eq. (10) in certain regions in Region:II and Region:III closer to the p_2p_1-plane, but these values are less than what one can achieve from $|\Psi(\mathrm{II})\rangle$ and $|\Psi(\mathrm{III})\rangle$, respectively.

Figure 4 gives the general structure of regions where superadditivity has been shown, but it tells very little about

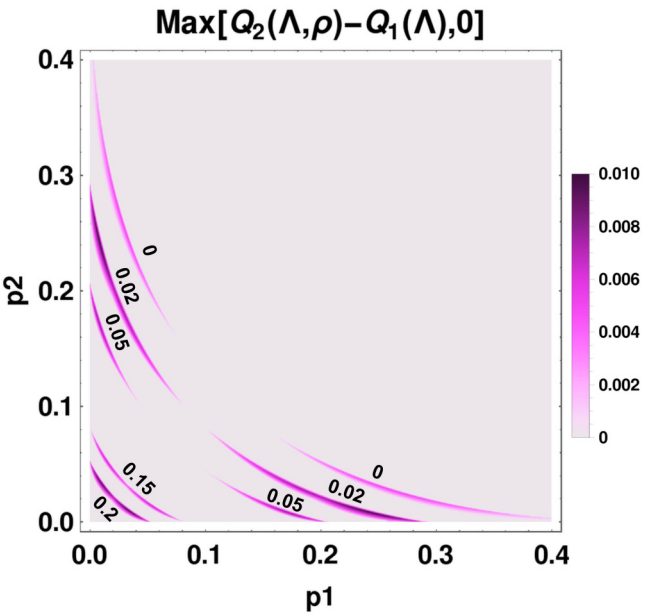


FIG. 5. 2D density plot of superadditivity for fixed p_3: $\max[Q^{(2)}(\Psi, \Lambda_{\vec{p}}) - Q^{(1)}(\Lambda_{\vec{p}}),0]$ is plotted for fixed values of $p_3 \in \{0, 0.02, 0.05, 0.15, 0.2\}$, where $Q^{(2)}(\Psi, \Lambda_{\vec{p}}) \equiv \max[Q^{(2)}(|\Psi(\mathrm{I})\rangle, \Lambda_{\vec{p}}), Q^{(2)}(|\Psi(\mathrm{II})\rangle, \Lambda_{\vec{p}}), Q^{(2)}(|\Psi(\mathrm{III})\rangle, \Lambda_{\vec{p}})]$. All colored regions correspond to regions of superadditivity, with the amount of superadditivity shown by the "color bar" to the right of the plot.

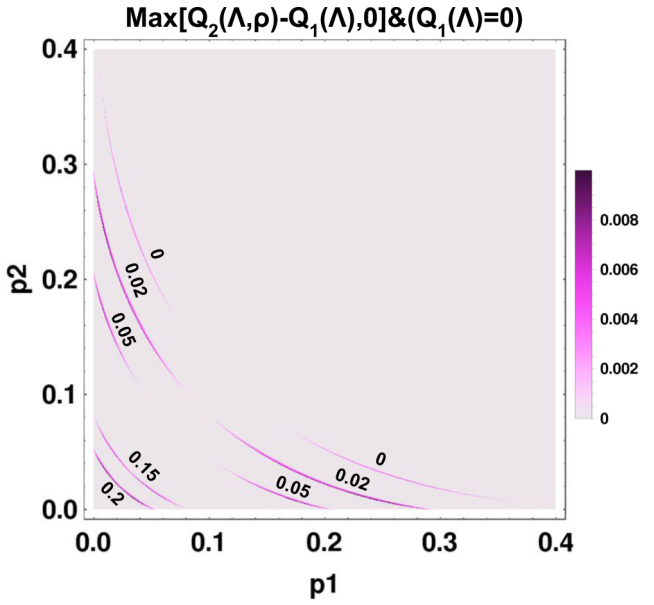


FIG. 6. Pauli channels with $Q^{(1)}(\Lambda_{\vec{p}}) = 0$ but $Q^{(2)}(\Lambda_{\vec{p}}, \rho) > 0$: $\max[Q^{(2)}(\lambda_{\vec{p}}, \rho) - Q^{(1)}(\Lambda_{\vec{p}}),0]$ is plotted for fixed values of $p_3 \in \{0, 0.02, 0.05, 0.15, 0.2\}$, whenever $Q^{(1)}(\Lambda_{\vec{p}}) = 0$. All colored regions correspond to regions of superadditivity, with the amount of superadditivity shown by the "color bar" to the right of the plot.

the three-dimensional structure of the regions that display superadditivity and the magnitude of superadditivity. To get around this problem, two-dimensional density plots were created, as shown in Fig. 5, for different fixed values of p_3 ($\in \{0, 0.02, 0.05, 0.15, 0.2\}$). The colored regions in the plot represent regions of superadditivity, with the magnitude of superadditivity encoded in the color as detailed in the color bar. Superadditivity of two-shot coherent information has been shown for several two-parameter families of quantum channels like the dephrasure channel and the generalized amplitude damping channel [26]. Each stripe in Fig. 5 effectively describes another two-parameter family with parameters p_1 and p_2, which shows superadditivity of two-shot coherent information. Figure 4 coupled with Fig. 5 gives a better description of the regions where superadditivity can be shown and the magnitude of this superadditivity. From an experimentalist's perspective, it might be useful to find Pauli channels that exhibit the maximal gap between one-shot quantum capacity and two-shot quantum capacity while using the codes Eq. (11). Numerically, we calculated the maximal gap to be $\max_{\vec{p}}[Q^{(2)}(\Lambda_{\vec{p}}, |\Psi(I)\rangle) - Q^{(1)}(\Lambda_{\vec{p}})] \approx 0.010\,234\,2$, achieved using the Pauli channel with $\vec{p} = (p_1, p_2, p_3) = (0.225\,688, 0.008\,011\,96, 0.026\,304\,1)$ (and other permutations of p_1, p_2, and p_3). This is, to the best of our knowledge, much higher than the gap obtained in other well-known channels such as the generalized erasure channel [52], the dephrasure channel, the generalized amplitude damping channel, and the depolarizing channel [26], making such Pauli channels good candidates for experimental studies.

It is also interesting to find out subclasses of Pauli channels for which $Q^{(1)}(\Lambda_{\vec{p}}) = 0$ but $Q^{(2)}(\Psi, \Lambda_{\vec{p}}) \neq 0$. These are the channels where, on single-use of the channel, no quantum information can be communicated, but on multiple parallel uses, communication is viable. Pauli channels that show this property are shown in Fig. 6. This region is identical to that of Fig. 5, except that it is "thinner" than that in Fig. 5.

2. Optimizing three-shot coherent information

Up to now, we have restricted our discussions to the two-shot capacity. It can be seen that the depolarizing channel (corresponding to the main diagonal in Fig. 5) does not show superadditivity under two channel uses using the quantum codes shown in Eq. (11). In fact, no quantum codes are known that achieve superadditivity for the two-shot case for the depolarizing channel. However, it is known that by using the repetition code, the depolarizing channel shows superadditivity under three channel-use (and higher uses as well) [26]. It is, therefore, interesting to study the three-shot capacity of the Pauli channels. Using GA-based search with a Schmidt *Ansatz* [26] for representing quantum states, we searched through the three-parameter space of Pauli channels for the superadditivity phenomenon, finding the regions marked in Fig. 7. We found three distinct quantum codes that displayed the highest coherent information in our analysis. These are as follows:

$$\begin{aligned} |\Phi(\mathrm{I})\rangle &\equiv |zzz\rangle|z\rangle + |\bar{z}\bar{z}\bar{z}\rangle|\bar{z}\rangle; \\ |\Phi(\mathrm{II})\rangle &\equiv |xxx\rangle|x\rangle + |\bar{x}\bar{x}\bar{x}\rangle|\bar{x}\rangle; \\ |\Phi(\mathrm{III})\rangle &\equiv |yyy\rangle|y\rangle + |\bar{y}\bar{y}\bar{y}\rangle|\bar{y}\rangle, \end{aligned} \tag{13}$$

where the ket symbols are defined as earlier. These are essentially the famous repetition codes that have been used to show superadditivity for the depolarizing channels. The last qubit

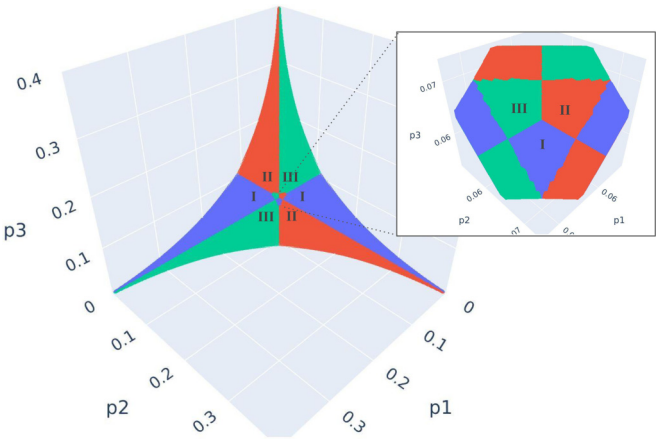


FIG. 7. Regions of three-shot superadditivity obtained from the three quantum codes: The blue region (marked with "I") shows the highest three-shot coherent information when code $|\Phi(\mathrm{I})\rangle$ is used. Similarly, the red region (marked "II") corresponds to code $|\Phi(\mathrm{II})\rangle$ and the green region (marked "III") corresponds to code $|\Phi(\mathrm{III})\rangle$. p_1, p_2, p_3 are, respectively, the probabilities of three Pauli errors.

in Eqs. (13) is the purifying qubit. The regions in the three-parameter space that show superadditivity using these codes are shown in Fig. 7. The regions look similar to that in Fig. 4 for the two-shot case, but we emphasize that although there is an overlap between the two regions in Figs. 7 and 4, neither of the regions fully contains the other. It is clear from Fig. 7 that there are Pauli channels of the form $\vec{p} = (p_1, p_2, p_3) = ((1+\epsilon_1)p, (1+\epsilon_2)p, (1+\epsilon_3)p)$ with sufficiently small ϵ_i that does not show superadditivity under two channel uses (using the quantum codes we described), but it does show superadditivity under three uses. The depolarizing channel is, of course, a special case of these channels. It is also interesting to note that there is a flip in the good quantum codes near the center. This again reveals the nontrivial relation between noise and "goodness" of quantum codes. The exact reason for this behavior is an interesting subject for future study.

We further calculated the expression for three-shot quantum capacity when $|\Phi(I)\rangle$ is passed through the Pauli channel $\Lambda_{\vec{p}} \equiv \Lambda(p_1, p_2, p_3)$, with $\vec{p} = (p_1, p_2, p_3)$ and $p_0 = 1 - p_1 - p_2 - p_3$. It can be written as

$$
\begin{aligned}
&Q^{(3)}(|\Phi(I)\rangle, \Lambda_{\vec{p}}) \\
&= \tfrac{1}{3}p_2\left(3p_1^2 + p_2^2\right)\log_2\left(p_2^3 + 3p_1^2p_2\right) + \tfrac{1}{3}p_1\left(p_1^2 + 3p_2^2\right)\log_2\left(p_1^3 + 3p_2^2p_1\right) - (p_1+p_2)(p_0+p_3)\log_2\left(\tfrac{1}{2}(p_1+p_2)(p_0+p_3)\right) \\
&\quad + \left[p_0\left(p_1^2+p_2^2\right) + 2p_1p_2p_3\right]\log_2\left(p_0\left(p_1^2+p_2^2\right) + 2p_1p_2p_3\right) + \left[2p_0p_1p_2 + \left(p_1^2+p_2^2\right)p_3\right]\log_2\left(2p_0p_1p_2 + \left(p_1^2+p_2^2\right)p_3\right) \\
&\quad + \tfrac{1}{3}p_3\left(3p_0^2+p_3^2\right)\log_2\left(p_3^3+3p_0^2p_3\right) + \tfrac{1}{3}p_0\left(p_0^2+3p_3^2\right)\log_2\left(p_0^3+3p_3^2p_0\right) - \tfrac{1}{3}[1-3(p_1+p_2)(p_0+p_3)] \\
&\quad \times \log_2\left(\tfrac{1}{2}[1-3(p_1+p_2)(p_0+p_3)]\right) + \left[2p_0p_2p_3 + p_1\left(p_0^2+p_3^2\right)\right]\log_2\left(2p_0p_2p_3 + p_1\left(p_0^2+p_3^2\right)\right) \\
&\quad + \left[2p_0p_1p_3 + p_2\left(p_0^2+p_3^2\right)\right]\log_2\left(2p_0p_1p_3 + p_2\left(p_0^2+p_3^2\right)\right). \qquad (14)
\end{aligned}
$$

Like in the previous case, given Eq. (14), the symmetry of the quantum codes and that of the Pauli channels can be used to show that

$$
\begin{aligned}
Q^{(3)}(|\Phi(\mathrm{II})\rangle, \Lambda(p_1,p_2,p_3)) &= Q^{(3)}(|\Phi(I)\rangle, \Lambda(p_3,p_2,p_1)), \\
Q^{(3)}(|\Phi(\mathrm{III})\rangle, \Lambda(p_1,p_2,p_3)) &= Q^{(3)}(|\Phi(I)\rangle, \Lambda(p_1,p_3,p_2)).
\end{aligned}
$$

Here as well, we estimated numerically the maximal gap while using the quantum codes in the set of Eq. (14) to be $\max_{\vec{p}}[Q^{(3)}(\Lambda_{\vec{p}}, |\Phi(I)\rangle) - Q^{(1)}(\Lambda_{\vec{p}})] \approx 0.012\,740\,6$, achieved using the Pauli channel with $\vec{p} = (p_1, p_2, p_3) = (0.007\,306\,49, 0.240\,303, 0.022\,323\,4)$ (and other permutations of p_1, p_2, and p_3). Interestingly, the optimal gap in this case is much higher than what was observed in the case of two channel uses discussed earlier.

The Schmidt *Ansatz* is somewhat restrictive since the algorithm does not cover the entire 6-qubit space. Therefore, it is possible that with the NN *Ansatz* as in Eq. (7), the ML algorithm may find other new and nontrivial codes that outperform the repetition codes. Next, we explore this possibility. Using an NN *Ansatz* with 6 input neurons, 6 hidden layers, and a width of 6 units, we ran the same GA and, to our surprise, found several nontrivial codes that outperformed the repetition codes. These codes are as follows:

$$
\begin{aligned}
|\chi(\mathrm{I})\rangle \equiv{}& (|zzz\rangle + |\bar{z}\bar{z}\bar{z}\rangle)|zzz\rangle + (|z\bar{z}z\rangle - |\bar{z}z\bar{z}\rangle)|zz\bar{z}\rangle \\
&+ (|zzz\rangle - |\bar{z}\bar{z}\bar{z}\rangle)|z\bar{z}z\rangle + (|z\bar{z}z\rangle + |\bar{z}z\bar{z}\rangle)|z\bar{z}\bar{z}\rangle; \\
|\chi(\mathrm{II})\rangle \equiv{}& (|xxx\rangle + |\bar{x}\bar{x}\bar{x}\rangle)|xxx\rangle + (|x\bar{x}x\rangle - |\bar{x}x\bar{x}\rangle)|xx\bar{x}\rangle \\
&+ (|xxx\rangle - |\bar{x}\bar{x}\bar{x}\rangle)|x\bar{x}x\rangle + (|x\bar{x}x\rangle + |\bar{x}x\bar{x}\rangle)|x\bar{x}\bar{x}\rangle; \\
|\chi(\mathrm{III})\rangle \equiv{}& (|yyy\rangle + |\bar{y}\bar{y}\bar{y}\rangle)|yyy\rangle + (|y\bar{y}y\rangle - |\bar{y}y\bar{y}\rangle)|yy\bar{y}\rangle \\
&+ (|yyy\rangle - |\bar{y}\bar{y}\bar{y}\rangle)|y\bar{y}y\rangle + (|y\bar{y}y\rangle + |\bar{y}y\bar{y}\rangle)|y\bar{y}\bar{y}\rangle. \qquad (15)
\end{aligned}
$$

The regions in the three-parameter space where these codes exhibit superadditivity are shown in Fig. 8. As seen from the figure, the regions are very similar to those seen in Fig. 4. This is interesting since Eqs. (15) and (11) (by padding it for the three-shot case by using dummy qubits) are not equivalent. There are also points where Eq. (15) outperforms the repetition codes (for instance, at $p_1 = 0.0$, $p_2 = 0.002$, $p_3 = 0.382$). Other such points can be seen in Fig. 9, where we have shown points in the three-parameter space where either of the two sets—Eq. (13) or Eq. (15)—display superadditivity. The points are color-coded according to the best performing codes among the two sets. We have further calculated the expression

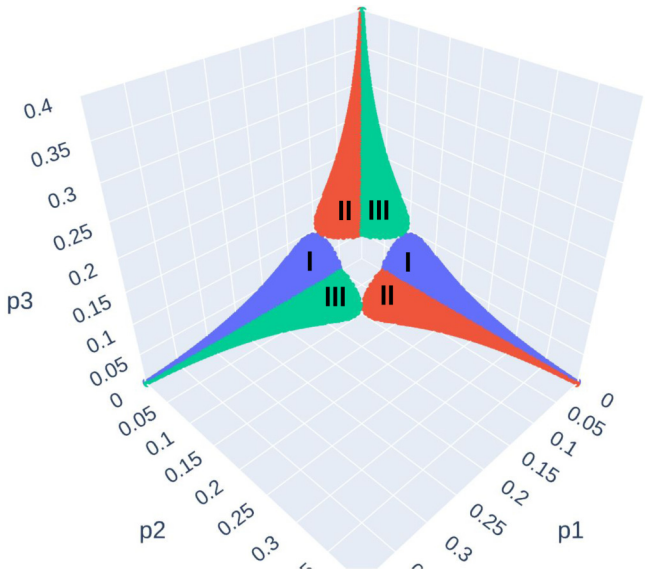


FIG. 8. Regions of three-shot superadditivity obtained from the quantum codes in Eq. (15): The blue region (marked with "I") shows the highest three-shot coherent information when code $|\chi(\mathrm{I})\rangle$ is used. Similarly, the red region (marked "II") corresponds to code $|\chi(\mathrm{II})\rangle$ and the green region (marked "III") corresponds to code $|\chi(\mathrm{III})\rangle$. p_1, p_2, p_3 are, respectively, the probabilities of three Pauli errors.

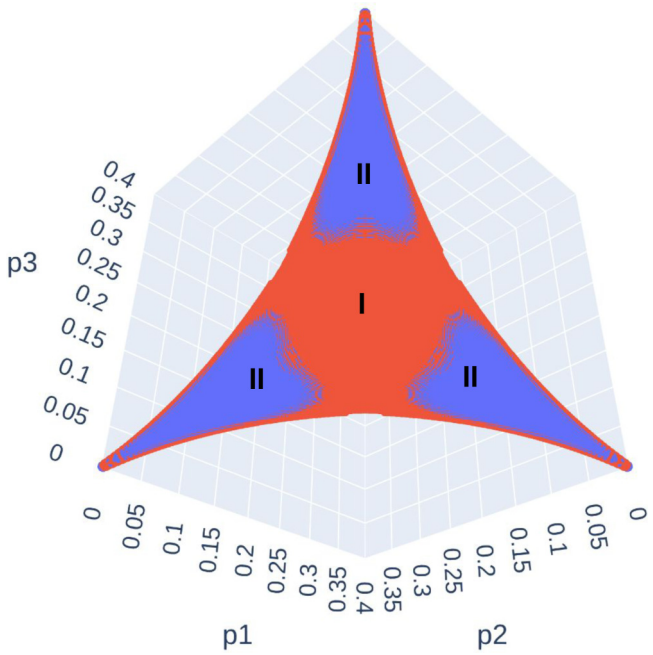


FIG. 9. Regions of three-shot superadditivity obtained from the quantum codes in Eqs. (15) and (13): The red region (marked "I") corresponds to points where Eq. (13) shows maximum superadditivity, while the blue region (marked "II") corresponds to points where Eq. (15) outperforms Eq. (13).

for the three-shot capacity of these codes,

$$
\begin{aligned}
Q^{(3)}(|\chi(I)\rangle, \Lambda_{\vec{p}}) = \frac{1}{3}\Big[& 2p_1^2p_2\log_2\left(2p_1^2p_2\right) + 2p_1p_2^2\log_2\left(2p_1p_2^2\right) + 2p_3^2p_0\log_2\left(2p_3^2p_0\right) + 2p_3p_0^2\log_2\left(2p_3p_0^2\right) \\
& + 2p_1p_2p_3\log_2(2p_1p_2p_3) + 2p_1p_2p_0\log_2(2p_1p_2p_0) + 2p_1p_3p_0\log_2(2p_1p_3p_0) + 2p_2p_3p_0\log_2(2p_2p_3p_0) \\
& + p_1\left(p_1^2+p_2^2\right)\log_2\left(p_1\left(p_1^2+p_2^2\right)\right) + p_1\left(p_3^2+p_0^2\right)\log_2\left(p_1\left(p_3^2+p_0^2\right)\right) + p_2\left(p_3^2+p_0^2\right)\log_2\left(p_2\left(p_3^2+p_0^2\right)\right) \\
& + p_3\left(p_3^2+p_0^2\right)\log_2\left(p_3\left(p_3^2+p_0^2\right)\right) + p_2\left(p_1^2+p_2^2\right)\log_2\left(p_2\left(p_1^2+p_2^2\right)\right) + p_3\left(p_1^2+p_2^2\right)\log_2\left(p_3\left(p_1^2+p_2^2\right)\right) \\
& + p_0\left(p_3^2+p_0^2\right)\log_2\left(p_0\left(p_3^2+p_0^2\right)\right) + \left(p_1^2+p_2^2\right)p_0\log_2\left(\left(p_1^2+p_2^2\right)p_0\right) \\
& + 2p_1(p_1p_3+p_2p_0)\log_2(p_1(p_1p_3+p_2p_0)) + 2p_1(p_1p_0+p_2p_3)\log_2(p_1(p_1p_0+p_2p_3)) \\
& + 2p_2(p_1p_3+p_2p_0)\log_2(p_2(p_1p_3+p_2p_0)) + 2p_2(p_1p_0+p_2p_3)\log_2(p_2(p_1p_0+p_2p_3)) \\
& + 2p_3(p_1p_3+p_2p_0)\log_2(p_3(p_1p_3+p_2p_0)) + 2p_3(p_1p_0+p_2p_3)\log_2(p_3(p_1p_0+p_2p_3)) \\
& + 2p_0(p_1p_3+p_2p_0)\log_2(p_0(p_1p_3+p_2p_0)) + 2p_0(p_1p_0+p_2p_3)\log_2(p_0(p_1p_0+p_2p_3)) - 2(p_1+p_2) \\
& \times (p_0+p_3)\log_2\left(\frac{1}{2}(p_1+p_2)(p_0+p_3)\right) - \left((p_1+p_2)^2+(p_0+p_3)^2\right)\log_2\left(\frac{(p_1+p_2)^2+(p_0+p_3)^2}{4}\right)\Big].
\end{aligned}
\tag{16}
$$

As earlier, the symmetry of the quantum codes and that of the Pauli channels can be used to show that

$$
\begin{aligned}
Q^{(3)}(|\chi(\mathrm{II})\rangle, \Lambda(p_1,p_2,p_3)) &= Q^{(3)}(|\chi(\mathrm{I})\rangle, \Lambda(p_3,p_2,p_1)), \\
Q^{(3)}(|\chi(\mathrm{III})\rangle, \Lambda(p_1,p_2,p_3)) &= Q^{(3)}(|\chi(\mathrm{I})\rangle, \Lambda(p_1,p_3,p_2)).
\end{aligned}
$$

Numerically, the maximal gap attained by these quantum codes is approximately 0.006 815 35 when $\vec{p} = (p_1, p_2, p_3) =$ (0.008 246 09, 0.220 845, 0.027 740 4) (and other permutations of p_1, p_2, and p_3).

IV. CONCLUSION

We showed that a neural network representation of quantum states coupled with a simple evolutionary strategy could find quantum codes with high coherent information for the class of Pauli channels. We considered the three-parameter

family of Pauli channels and found quantum channels that display superadditivity of two-shot coherent information. We further characterized the quantum codes that the algorithm learned and surprisingly found that the algorithm only learned three distinct quantum codes. With the exact form of the quantum codes at hand, we calculated the exact form of the two-shot coherent information with these codes and further mapped regions of superadditivity in the three-parameter space, thus obtaining the structure as shown in Fig. 4. Our analysis showed that certain Pauli channels show a large gap between single-shot quantum capacity and two-shot quantum capacity; this high gap of around 0.01 is much larger than what is seen in other well-studied classes of quantum channels, and therefore we think these Pauli channels will be of use in experimental studies of quantum capacities.

For better structural visualization, we investigated the superadditivity property under the constant noise of one of the parameters (p_3), effectively obtaining two-parameter families, plotted in Fig. 5, that display superadditivity of two-shot coherent information. Furthermore, we find Pauli channels that cannot transmit quantum information in single use, but parallel use of their two copies can communicate quantum information reliably with non-zero rate, a prominent example of "$0 + 0 > 0$." The channels having the above feature are plotted in Fig. 6.

We also applied the same approach to the three channel-use case of Pauli channels. With a Schmidt *Ansatz* the algorithm found three distinct quantum codes, which are variants of the famous repetition code. The superadditive region obtained by using these codes is shown in Fig. 7. Our inspection is quite consistent with the studies of the depolarizing channels, which shows superadditivity under three channel uses but not under two uses of the channel; in fact, this seems to remain true for other one-parameter quantum channels that are "close," in terms of the noise structure, to the depolarizing channel, as can be seen from Fig. 4.

The true power of machine learning approaches lies in finding nontrivial solutions to complex problems. Recognizing the possibility of such nontrivial quantum codes with high superadditivity, we used the NN *Ansatz* in Eq. (7) to search through the 6 qubit Hilbert space, eventually discovering the set Eq. (15). In fact, there are Pauli channels where this code outperforms the repetition codes.

The results shed light on the behavior and properties of the class of Pauli channels; this would be of much use in the context of near-term quantum computing, considering the usefulness of Pauli channels as a practical model for real quantum noise. Moreover, as discussed before, the high superadditivity displayed by some of the Pauli channels makes them an ideal candidate for experimental studies. The success of the GA-based optimization technique with neural networks in finding new and nontrivial quantum codes is yet another instance of the power of machine learning techniques in aiding fundamental research and, more importantly, quantum information research.

Note that these calculations become prohibitive as the number of channel uses increases; therefore, an interesting problem would be finding efficiently computable fitness measures that approximate coherent information. All our calculations involved basic techniques from the respective fields, and their success suggests that it is worthwhile to implement advanced GA-based optimization techniques and neural network representations to find more efficient and computationally feasible schemes. It would also be interesting to use these techniques to find generalizations of the codes in Eqs. (11) and (15) to higher channel uses, eventually finding better bounds to the quantum capacity of these channels.

ACKNOWLEDGMENTS

G.L.S. is thankful for the DST-INSPIRE fellowship. M.A. and M.B. acknowledge funding from the National Mission in Interdisciplinary Cyber-Physical systems from the Department of Science and Technology through the I-HUB Quantum Technology Foundation (Grant No. I-HUB/PDF/2021-22/008). M.B. acknowledges support through the research grant of INSPIRE Faculty fellowship from the Department of Science and Technology, Government of India and the startup research grant from SERB, Department of Science and Technology (Grant No. SRG/2021/000267). The authors gratefully acknowledge the use of the *Padmanabha* computational cluster made available through the center for High Performance Computation at IISER-TVM.

APPENDIX: ANALYSIS OF LEARNING PERFORMANCE

In this Appendix, we present the numerical performance of the optimization scheme used, and we show empirical evidence that suggests that the NN *Ansatz* coupled with the presented GA algorithm is a promising approach to find quantum codes with high coherent information. In this Appendix, we ask three main questions: (i) Is a NN *Ansatz* needed in the first place to find good quantum codes? (ii) How does the GA's learning performance change as the NN's size is varied? (iii) How does the employed GA perform in comparison to the well-studied case of PSO?

As we will soon elaborate on, we found the following through simulations: (i) A NN *Ansatz* is not necessary for finding good quantum codes in the two-shot case. On the other hand, in the three-shot case, only the algorithms trained with a NN *Ansatz* could find good quantum codes that showed superadditivity. This suggests that *a priori*, a NN *Ansatz* is a good starting point for exploring the superadditivity of quantum channels. (ii) Given fixed parameters, the GA showed improved learning performance as the depth of the NN increased. We also found a positive correlation between learning performance and population size. This contrasts with a simple PSO variant, which performed worse as the depth of the NN increased. (iii) By comparing the performance of the GA and the PSO variant in test cases, we found that their performances are comparable. In particular, the employed PSO algorithm always performed superior to the GA when a RAW *Ansatz* was used. RAW here refers to parametrizing a quantum code explicitly by its coefficients in the computational basis. In sharp contrast, GA outperformed the PSO algorithm when the NN *Ansatz* was used.

Before elaborating on these points, we briefly introduce the PSO algorithm used in this paper.

1. Particle swarm optimization

Particle swarm optimization (PSO) is yet another derivative-free metaheuristic search method that is used extensively in global optimization problems. A variant of PSO for the quantum codes search was used extensively in Ref. [26].

PSO works in the following mechanism: Initially, a set of candidate solutions (here called particles) are randomly generated in the optimization landscape. Each particle will be indexed by "i" and equipped with a position vector (x_i) encoding the candidate solution and a velocity vector (v_i) encoding the particle's direction of motion in the optimization landscape. Each particle also "remembers" the best position it has occupied in the landscape and the best position achieved (considering all the particles) since the start of the optimization. We also set an upper and lower bound to the velocity components that can be achieved by each of the particles.

A snippet of the PSO scheme is given in the following. The update equation is the same as given in Ref. [26]:

N_GEN #Number of generations

N_Particles #Number of particles

α #Intertia Parameter

β #Self-Interaction Parameter

γ #Mutual-Interaction Parameter

x_i^B #Best position in particle-i's history

x^G #Best position found in the entire history

$v_{\min}$, $v_{\max}$ #Min and Max of velocity.

Initialize all particles randomly

Repeat N_GEN times:

  For each particle-i:

    For each position/velocity component-k:

      $\beta_r \leftarrow \beta * \text{random}(0, 1)$

      $\gamma_r \leftarrow \gamma * \text{random}(0, 1)$

      $v_i(k) \leftarrow \alpha v_i(k) + \beta_r[x_i^B(k) - x_i(k)] + \gamma_r[x^G(k) - x_i(k)]$

      if $v_i(k) > v_{\max}$:

        $v_i(k) \leftarrow v_{\max}$

      if $v_i(k) < v_{\min}$:

        $v_i(k) \leftarrow v_{\min}$

        $x_i(k) \leftarrow x_i(k) + v_i(k)$

      end loop

      if $\text{Fitness}(\text{x}_\text{i}) > \text{Fitness}(\text{x}_\text{i}^\text{B})$:

        $x_i^B \leftarrow x_i$

  end loop

  $x^G \leftarrow \arg\max_i[\text{Fitness}(x_i)]$

end loop

return $\text{Fitness}(x^G)$

The PSO algorithm described above was implemented in Python using the DEAP package.[2]

Now we move to the analysis of the learning performance.

2. Learning performance

a. Two uses of the Pauli channel

We first study the problem of finding optimal quantum codes with high 2-shot coherent information. Purified quantum codes then belong to a 4-qubit Hilbert space. We first investigate if employing a NN *Ansatz* is helpful for this problem. For this, we compare the learning performance with the GA and PSO with a RAW *Ansatz* and NN *Ansätzes* of varying depth. As mentioned before, in the RAW *Ansatz*, we treat the coefficients of the state vector (in the computational basis) as the optimization parameters. Throughout this section, we fix the quantum channel to be a Pauli channel with parameters $p1 = 0.003$, $p2 = 0.285$, and $p3 = 0.008$. This point belongs to the region (in Fig. 4) where we found quantum codes that display superadditivity of coherent information.

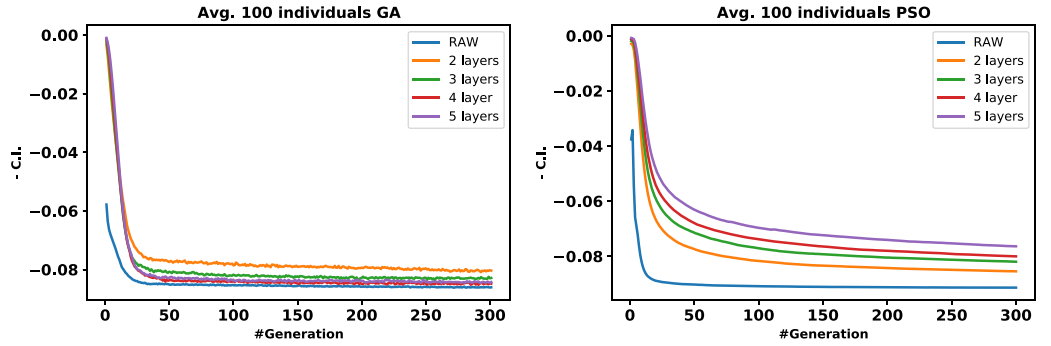
The genetic algorithm was executed with 100 individuals, which evolved for 300 generations. As described in the main text, a "2D crossover" type recombination strategy was used with a crossover probability of 0.5. A Gaussian mutation of mean 0.5 and standard deviation of 0.25 was used to mutate individuals. The mutation scheme is such that each individual has a chance of 0.2 for being picked for mutation. Once picked, each of its attributes (encodings) has a chance of 0.5 for undergoing a Gaussian addition.

The PSO algorithm was executed with 100 particles, which evolved for 300 generations. We also set $\alpha = \beta = \gamma = 0.5$. The maximum and minimum velocity of each component was set to 1 and -1, respectively.

We compared the learning performance of the GA and PSO with the RAW *Ansatz* and deep NN *Ansätzes* of depths 5, 4, 3, and 2 layers. As described in the main text, these are fully connected NNs of a width of four units each. They are chosen to have a cosine activation function in the first layer and "tanh" activation function in all subsequent layers. The learning performance of these algorithms with varying NN depths is shown in Fig. 10. As inferred from the figure, the population's average fitness consistently improves as the optimization proceeds in time. The first observation is that the RAW *Ansatz* performs better than an NN *Ansatz* in both cases. There is also a significant difference between the performance of the GA and PSO. As inferred from Fig. 10, the performance of the GA increases as the NN depth increases. This is in sharp contrast to the case of PSO, where the performance is observed to degrade as the NN's depth increases.

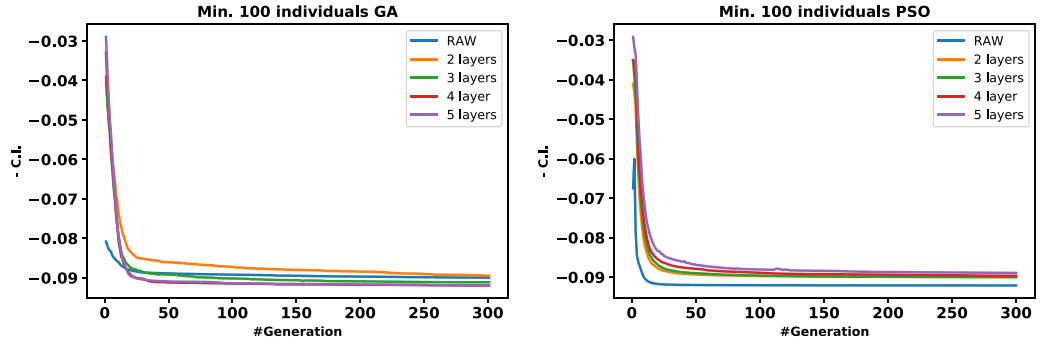
We also note that with a RAW *Ansatz*, PSO performed better than the GA, while with the NN *Ansatz*, the GA performed superior to PSO. This is a trend we observed throughout the analysis, as shall be mentioned in the coming sections. This does not, of course, mean that GAs are superior to PSO when equipped with an NN *Ansatz*. For one, the employed PSO variant and the GA are the simplest among a whole class of advanced PSO and GA variants. Moreover, even for the

[2]See Ref. [53] for an implementation of the above algorithm.



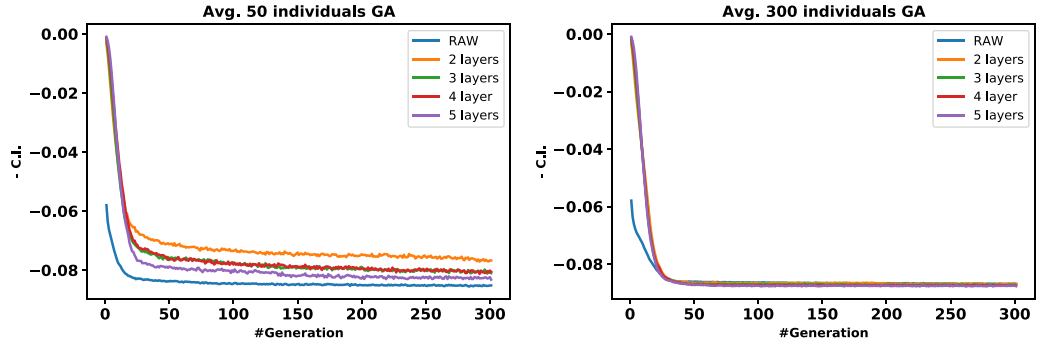
(a) GA (b) PSO

FIG. 10. Learning curves for GA/PSO with NNs of increasing depth. A population of 100 individuals is evolved in both cases. RAW represent the case when the NN *Ansatz* is not used and the coefficients of the wave function are treated as the optimization variables. Each learning curve shows the average fitness of the population averaged over 100 learning trajectories. (a) Going down the y-axis near the asymptote, the lines are in the order "2 layers," "3 layers," "5 layers," "4 layers," and "RAW." (b) Top curve corresponds to "5 layers" followed by "4 layers," "3 layers," "2 layers," and finally "RAW."



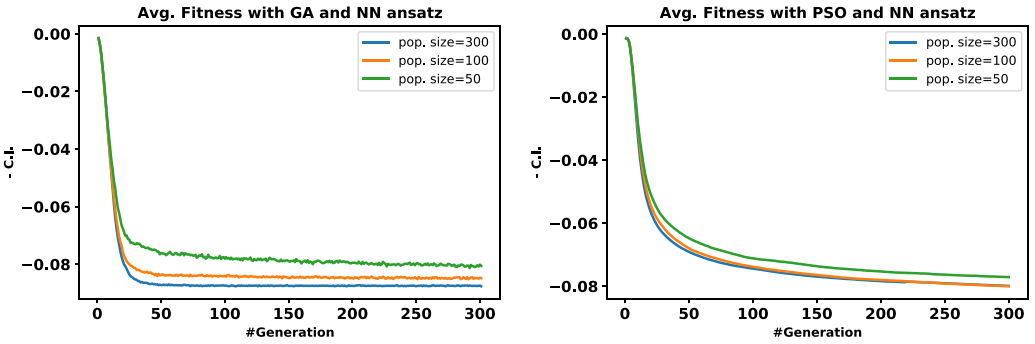
(a) GA (b) PSO

FIG. 11. Learning curves for GA/PSO with NNs of increasing depth. A total of 100 individuals are evolved in each case. RAW represents the case when the NN *Ansatz* is not used and the coefficients of the wave function are treated as the optimization variables. Each learning curve shows the best solution of a generation averaged over 100 learning trajectories. (a) Going down the y-axis near the asymptote, the lines are in the order "2 layers," "RAW," "3 layers," "4 layers," and "5 layers." (b) Top curve corresponds to "5 layers" followed by "4 layers," "3 layers," "2 layers," and finally "RAW."



(a) Population size:50 (b) Population size:300

FIG. 12. Learning curves for GA with NNs of increasing depth. Each learning curve shows the average fitness of the population averaged over 100 learning trajectories. (a) The initial population size is set to 50 individuals. The top curve corresponds to "2 layers," followed by "3 layers," "4 layers," "5 layers," and finally "RAW" at the bottom. (b) The initial population size is set to 300 individuals. The high population size resulted in an insensitivity to NN depth and hence the learning curves are overlapped and indistinguishable.



(a) GA (b) PSO

FIG. 13. Learning curves for GA and PSO with increasing population size. A NN with a depth of 4 layers is used as the optimization *Ansatz*. Each learning curve shows the average fitness of the population averaged over 100 learning trajectories. (a) A population size of 300 (bottom curve) shows optimal performance, followed by a population size of 100 (middle) and then by a size of 50 (top). (b) Going down the y-axis near the asymptote, the lines are in the order "pop. size = 50," "pop. size = 100," "pop. size = 300."

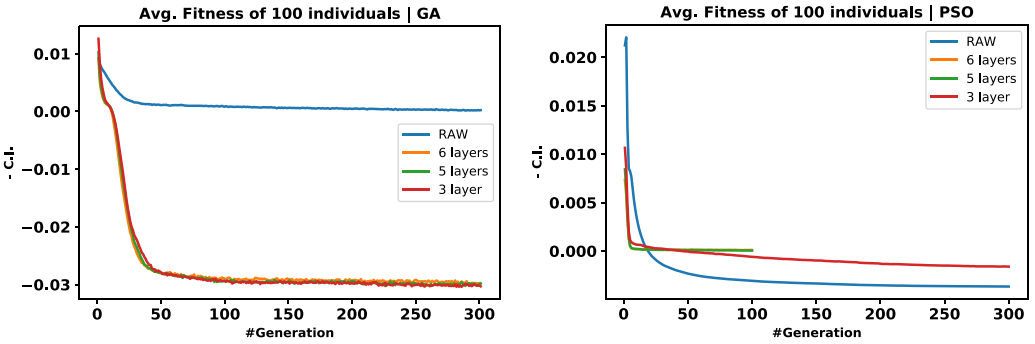
algorithms considered, the evidence provided is not decisive since the performance may change as the metaparameters are changed. So, at best, all we can conclude is that in general, PSO and GA are comparable for the optimization task considered. We will return to this point later.

In the previous discussion, we looked at the population's average fitness as a measure of learning performance. Perhaps a more useful metric is the best solution found at each generation. The learning curves with this new metric are shown in Fig. 11. As inferred from the figure, the main observations from the previous discussions carry over to this case as well.

It is also interesting to see how the population size affects learning performance. For this, we redid the previous calculations with a population size of 50 and 300 and found that in all these cases, the main observations made earlier carry over without fail. The corresponding learning curves for the GA are shown in Fig. 12. The learning curves for the PSO algorithm are relatively insensitive to different population sizes and are hence not shown. It is clear by comparing Fig. 10 to Fig. 12 that as the population size increases, the sensitivity of the learning performance to the depth of the NN decreases to the point of saturation at a population of 300. This shows a tradeoff between population size and depth of the NN, as intuitively expected.

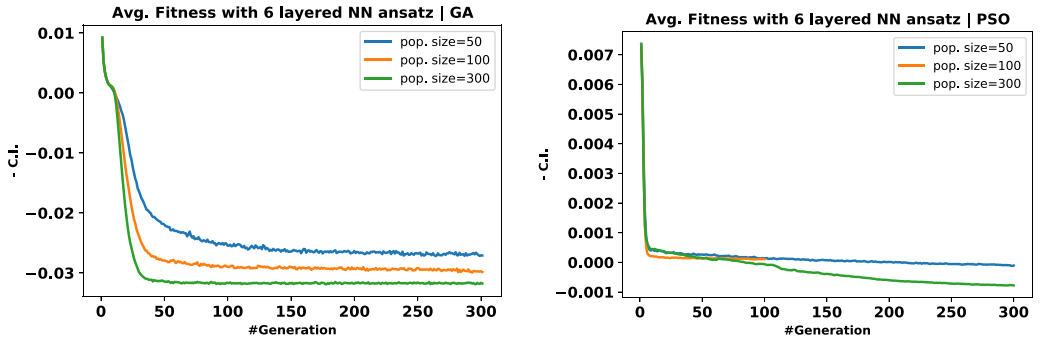
It would also be illuminating to see the dependence of the learning performance on population size. This is shown in Fig. 13, where the population dependence of GA/PSO was studied with a NN *Ansatz* of depth 4. As intuitively expected, we observed a positive correlation between learning performance and population size. The trend is observed irrespective of the algorithm (GA/PSO), *Ansatz* (RAW/NN), or learning metric (Avg/Min) used. Therefore, only the learning performance of GA and PSO with a 4-layered NN *Ansatz* and average fitness metric is shown in Fig. 13. We also point out an observed relative insensitivity of the employed PSO scheme to population size.

The preceding discussion focused on two-shot coherent information. A recurring observation we made is that in this case, the kind of NN *Ansatz* we used is not of much use when compared with the RAW *Ansatz*. This is likely due to the small size of the 4-qubit Hilbert space. We also saw that in this case, the GA scheme employed improves performance as the depth of the NN is increased or as the population size is increased. We also compared the GA with a simple PSO



(a) GA (b) PSO

FIG. 14. Learning curves for GA/PSO with NNs of increasing depth. RAW represent the case when the NN *Ansatz* is not used and the coefficients of the wave function are treated as the optimization variables. Each learning curve shows the average fitness of the population averaged over 100 learning trajectories. (a) The top curve corresponds to RAW *Ansatz*. Learning curves corresponding to 6 layers, 5 layers, and 3 layers are largely indistinguishable. (b) The top curve corresponds to "6 layers," followed by "5 layers," "3 layers," and finally the RAW *Ansatz*.



(a) GA (b) PSO

FIG. 15. Learning curves for GA and PSO with increasing population size. A NN with a depth of 6 layers is used as the optimization *Ansatz*. Each learning curve shows the average fitness of the population averaged over 100 learning trajectories. (a) The top curve corresponds to a population size of 50, followed by a population size of 100 and finally a population size of 300 individuals. (b) The top curves, corresponding to a population size of 50 and 100, are largely indistinguishable near the asymptote; these are followed by a green curve corresponding to a population size of 300.

variant and showed that the performance of both algorithms is comparable. Interestingly, we saw a degradation of PSO's performance with increased depth of the NN. Similar to the GA, the performance of PSO was seen to improve with increased population size; still, the sensitivity to the population size is weaker compared to the GA.

b. Three uses of the Pauli channel

Now we look at the 3-shot problem. The quantum codes now belong to a 6-qubit Hilbert space. We will consider the Pauli channel with parameters $p_1 = 0.006$, $p_2 = 0.022$, and $p_3 = 0.247$ as a testbed to study the performance of the various schemes. This is the Pauli channel for which we found the quantum codes with the highest 3-shot superadditivity. We start with the following important question: Can a NN *Ansatz* outperform RAW optimization in this 6-qubit problem? To answer this, we first studied the learning dynamics of GA/PSO with NNs of varying depth and compared it with RAW optimization. The metaparameters of GA and PSO are the same as used in the previous section. The resulting learning curves are shown in Fig. 14, where RAW optimization is compared with three NNs with depths 6, 5, and 3. The width of each NN is set to 6 units, and the activation functions are chosen as earlier.

From Fig. 14, it is clear that the NN *Ansatz* performs superior to the RAW optimization in the three-shot problem. We also note that while the employed GA scheme coupled with a NN *Ansatz* finds quantum codes of high coherent information, the PSO variant fails at this. In the case of RAW optimization, however, PSO performs slightly better than the GA; yet no code of high coherent information was found. We also observe that the learning curves for the GA are largely insensitive to the depth of the NN. The learning curves behave similarly with a "Best Solution" metric and are not shown here to avoid redundancy.

We now look at the effect of population size on learning performance. For this, we fix a NN *Ansatz* of depth 6 and analyze the learning performance with populations of sizes 300, 100, and 50. The results are shown in Fig. 15. As intuitively expected, performance improvement is observed as the population size is increased. As earlier, the employed PSO scheme is relatively insensitive to population size.

The study revealed that a NN *Ansatz* outperforms RAW optimization for finding quantum codes with high 3-shot coherent information. This displays the utility of machine learning approaches to finding good quantum codes in optimization problems, consistent with the results obtained in Ref. [26].

Throughout the preceding discussion, we also observed that with the RAW *Ansatz*, the employed PSO variant outperforms the GA. In contrast, we saw that when a NN *Ansatz* is used, the GA was seen to outperform the PSO scheme significantly. We also noticed that the performance of GA improves with the depth of the NN *Ansatz*, while PSO displayed degradation with depth. However, these observations should be treated with caution as this does not imply that GAs outperform PSO schemes in all cases of interest. For one, we used the simplest variants of GAs and PSO schemes; things might

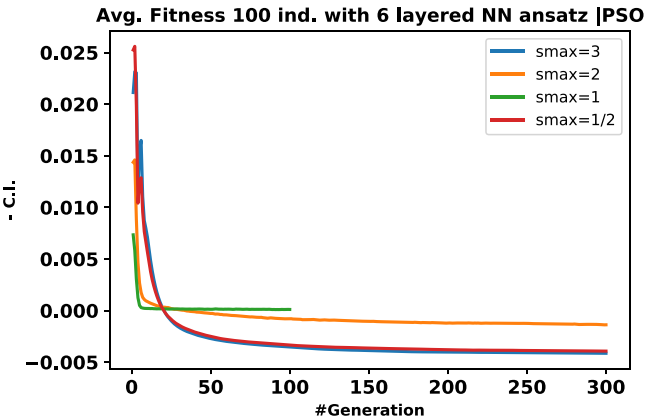


FIG. 16. Learning curves for PSO with different max. velocities. A 6-layered NN is used as the optimization *Ansatz*. Each learning curve shows the average fitness of the population, averaged over 100 learning trajectories. The top curve corresponds to smax $= 1$, followed by the curve with a maximum speed of 2. The curves corresponding to smax $= 3$ and smax $= 1/2$ are largely indistinguishable from each other and show maximum performance.

change significantly when more advanced methods are used. Moreover, we have not explored the learning performance for all possible values of the associated metaparameters. There might be different metaparameter settings where PSO will perform comparable to, or even better than, the GA with the NN *Ansatz*. To illustrate the dependence on metaparameters, we have evaluated the learning curves of PSO with varying max velocity. This is shown in Fig. 16. As observed from the figure, the learning performance depends (rather nontrivially) on the maximum allowed speed for each velocity component.

Compiling all these observations suggests that a GA coupled with a NN *Ansatz* significantly outperforms RAW optimization and is, therefore, a good optimization scheme for finding quantum codes of high coherent information. We also saw that, in the chosen metaparameter settings, the GA is more robust for use with an NN *Ansatz* compared to the RAW *Ansatz*. We also pointed out that this should not be taken for the superior performance of GA schemes to PSO schemes. At best, we may conclude that for the problem at hand, the two methods are comparable.

[1] C. E. Shannon, A mathematical theory of communication, Bell Syst. Tech. J. **27**, 379 (1948).

[2] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley, New York, 1991)

[3] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2000).

[4] C. H. Bennett and S. J. Wiesner, Communication via One- and Two-particle Operators on Einstein-Podolsky-Rosen States, Phys. Rev. Lett. **69**, 2881 (1992).

[5] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters, Teleporting an Unknown Quantum State Via Dual Classical and Einstein-Podolsky-Rosen Channels, Phys. Rev. Lett. **70**, 1895 (1993).

[6] C. H. Bennett and G. Brassard, Quantum cryptography: Public key distribution and coin tossing, Theor. Comput. Sci. **560**, 7 (2014).

[7] S. S. Bhattacharya, A. G. Maity, T. Guha, G. Chiribella, and M. Banik, Random-receiver quantum communication, PRX Quantum **2**, 020350 (2021).

[8] G. Chiribella, M. Banik, S. S. Bhattacharya, T. Guha, M. Alimuddin, A. Roy, S. Saha, S. Agrawal, and G. Kar, Indefinite causal order enables perfect quantum communication with zero capacity channels, New J. Phys. **23**, 033039 (2021).

[9] T. Guha, M. Alimuddin, S. Rout, A. Mukherjee, S. S. Bhattacharya, and M. Banik, Quantum advantage for shared randomness generation, Quantum **5**, 569 (2021).

[10] S. Saha, S. S. Bhattacharya, T. Guha, S. Halder, and M. Banik, Advantage of quantum theory over nonclassical models of communication, Ann. Phys. **532**, 2000334 (2020).

[11] M. Vaisakh, R. K. Patra, M. Janpandit, S. Sen, M. Banik, and A. Chaturvedi, Mutually unbiased balanced functions and generalized random access codes, Phys. Rev. A **104**, 012420 (2021).

[12] S. Agrawal, R. Tarafder, G. Smith, A. Roy, and M. Banik, Better transmission with lower capacity: Lossy compression over quantum channels, arXiv:2105.06617.

[13] M. M. Wilde, From classical to quantum shannon theory, arXiv:1106.1445.

[14] B. Schumacher and M. D. Westmoreland, Sending classical information via noisy quantum channels, Phys. Rev. A **56**, 131 (1997).

[15] A.S. Holevo, The capacity of the quantum channel with general signal states, IEEE Trans. Inf. Theor. **44**, 269 (1998).

[16] I. Devetak, The private classical capacity and quantum capacity of a quantum channel, IEEE Trans. Inf. Theor. **51**, 44 (2005).

[17] S. Lloyd, Capacity of the noisy quantum channel, Phys. Rev. A **55**, 1613 (1997).

[18] P. W. Shor, *The Quantum Channel Capacity and Coherent Information* (Lecture Notes, MSRI Workshop on Quantum Computation, 2002).

[19] M. B. Hastings, Superadditivity of communication capacity using entangled inputs, Nat. Phys. **5**, 255 (2009).

[20] K. Li, A. Winter, X. Zou, and G. Guo, Private Capacity of Quantum Channels is Not Additive, Phys. Rev. Lett. **103**, 120501 (2009).

[21] G. Smith and J. A. Smolin, Extensive Nonadditivity of Privacy, Phys. Rev. Lett. **103**, 120503 (2009).

[22] G. Smith and J. Yard, Quantum communication with zero-capacity channels, Science **321**, 1812 (2008).

[23] J. Oppenheim, For quantum information, two wrongs can make a right, Science **321**, 1783 (2008).

[24] G. Smith, J. A. Smolin, and J. Yard, Quantum communication with gaussian channels of zero quantum capacity, Nat. Photon. **5**, 624 (2011).

[25] T. Cubitt, D. Elkouss, W. Matthews, M. Ozols, D. Pérez-García, and S. Strelchuk, Unbounded number of channel uses may be required to detect quantum capacity, Nat. Commun. **6**, 6739 (2015).

[26] J. Bausch and F. Leditzky, Quantum codes from neural networks, New J. Phys. **22**, 023005 (2020).

[27] D. P. DiVincenzo, P. W. Shor, and J. A. Smolin, Quantum-channel capacity of very noisy channels, Phys. Rev. A **57**, 830 (1998).

[28] G. Smith and J. A. Smolin, Degenerate Quantum Codes for Pauli Channels, Phys. Rev. Lett. **98**, 030501 (2007).

[29] J. Fern and K. B. Whaley, Lower bounds on the nonzero capacity of pauli channels, Phys. Rev. A **78**, 062335 (2008).

[30] F. Leditzky, D. Leung, and G. Smith, Dephrasure Channel and Superadditivity of Coherent Information, Phys. Rev. Lett. **121**, 160501 (2018).

[31] S. Yu, Y. Meng, R. B. Patel, Y.-T. Wang, Z.-J. Ke, W. Liu, Z.-P. Li, Y.-Z. Yang, W.-H. Zhang, J.-S. Tang, C.-F. Li, and G.-C. Guo, Experimental Observation of Coherent-Information Superadditivity in a Dephrasure Channel, Phys. Rev. Lett. **125**, 060502 (2020).

[32] S. N. Filippov, Multiletter codes to boost superadditivity of coherent information in quantum communication lines with polarization dependent losses, Phys. Rev. A **105**, 062606 (2022).

[33] J. J. Wallman and J. Emerson, Noise tailoring for scalable quantum computation via randomized compiling, Phys. Rev. A **94**, 052325 (2016).

[34] B. M. Terhal, Quantum error correction for quantum memories, Rev. Mod. Phys. **87**, 307 (2015).

[35] A. Tehrani and R. Pereira, The coherent information on the manifold of positive definite density matrices, J. Math. Phys. **62**, 042201 (2021).

[36] M. Schuld, I. Sinayskiy, and F. Petruccione, An introduction to quantum machine learning, Contemp. Phys. **56**, 172 (2015).

[37] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, Quantum machine learning, Nature (London) **549**, 195 (2017).

[38] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborová, Machine learning and the physical sciences, Rev. Mod. Phys. **91**, 045002 (2019).

[39] L. Lamata, Quantum machine learning and quantum biomimetics: A perspective, Mach. Learn.: Sci. Technol. **1**, 033002 (2020).

[40] J. Carrasquilla, Machine learning for quantum matter, Adv. Phys. X **5**, 1797528 (2020).

[41] R. Iten, T. Metger, H. Wilming, L. del Rio, and R. Renner, Discovering Physical Concepts with Neural Networks, Phys. Rev. Lett. **124**, 010508 (2020).

[42] Z. Cai and J. Liu, Approximating quantum many-body wave functions using artificial neural networks, Phys. Rev. B **97**, 035116 (2018).

[43] H. Saito, Solving the bose–hubbard model with machine learning, J. Phys. Soc. Jpn. **86**, 093001 (2017).

[44] M. Abadi *et al.*, TensorFlow: Large-scale machine learning on heterogeneous systems (2015), software available from tensorflow.org.

[45] D. P Kingma and J. Ba, Adam: A method for stochastic optimization, arXiv:1412.6980.

[46] T. Bäck, D. B. Fogel, and Z. Michalewicz, *Evolutionary Computation 1: Basic Algorithms and Operators* (CRC, Boca Raton, FL, 2018).

[47] F-A Fortin, F-M De Rainville, M-A Gardner, M. Parizeau, and C. Gagné, Deap: Evolutionary algorithms made easy, J. Mach. Learn. Res. **13**, 2171 (2012).

[48] J. P. Cohoon and W. D. Paris, Genetic placement, IEEE Trans. Comput.-Aided Des. Integr. Circ. Syst. **6**, 956 (1987).

[49] M. M. Wilde, *Quantum Information Theory* (Cambridge University Press, Cambridge, 2013).

[50] J. Bausch and F. Leditzky, Error thresholds for arbitrary pauli noise, SIAM J. Comput. **50**, 1410 (2021).

[51] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters, Mixed-state entanglement and quantum error correction, Phys. Rev. A **54**, 3824 (1996).

[52] S. N. Filippov, Capacity of trace decreasing quantum operations and superadditivity of coherent information for a generalized erasure channel, J. Phys. A **54**, 255301 (2021).

[53] https://deap.readthedocs.io/en/master/examples/pso_basic.html.