

PAPER • OPEN ACCESS

## A Monte Carlo Tree Search approach to QAOA: finding a needle in the haystack

To cite this article: Andoni Agirre *et al* 2025 *New J. Phys.* **27** 043014

View the [article online](#) for updates and enhancements.

You may also like




- [Spin-dynamics of pairs of submerged-4f-shell atoms in cylindrically symmetric harmonic traps: connecting multi-channel models to effective analytical approaches](#)  
Hui Li, Eite Tiesinga, Ming Li et al.
- [The composite particle duality: a new class of topological quantum matter](#)  
Gerard Valentí-Rojas, Joel Priestley and Patrik Öhberg
- [Determining absolute neutrino mass using quantum technologies](#)  
A A S Amad, F F Deppisch, M Fleck et al.



## PAPER

## A Monte Carlo Tree Search approach to QAOA: finding a needle in the haystack

## OPEN ACCESS

RECEIVED  
17 January 2025REVISED  
17 March 2025ACCEPTED FOR PUBLICATION  
31 March 2025PUBLISHED  
11 April 2025Original Content from  
this work may be used  
under the terms of the  
[Creative Commons  
Attribution 4.0 licence](https://creativecommons.org/licenses/by/4.0/).Any further distribution  
of this work must  
maintain attribution to  
the author(s) and the title  
of the work, journal  
citation and DOI.Andoni Agirre<sup>1,2,3,\*</sup> , Evert van Nieuwenburg<sup>4</sup>  and Matteo M Wauters<sup>2,5,6,\*</sup> <sup>1</sup> Donostia International Physics Center (DIPC), Manuel Lardizabal Pasealekua 4, 20018 Donostia, Basque Country, Spain<sup>2</sup> Center for Quantum Devices, Niels Bohr Institute, Copenhagen University, Universitetsparken 5, 2100 Copenhagen, Denmark<sup>3</sup> Department of PMAS: Physics, Chemistry and Technology, University of the Basque Country (UPV/EHU), Manuel Lardizabal Pasealekua 3, 20018 Donostia, Basque Country, Spain<sup>4</sup>  $\langle aQa \rangle$  at Lorentz Institute and Leiden Institute of Advanced Computer Science, Leiden University, PO Box 9506, 2300 RA Leiden, The Netherlands<sup>5</sup> CNR-INO Pitaevskii BEC Center and Department of Physics, University of Trento, Via Sommarive 14, I-38123 Povo, Trento, Italy<sup>6</sup> INFN-TIFPA, Trento Institute for Fundamental Physics and Applications, Via Sommarive 14, I-38123 Povo, Trento, Italy

\* Authors to whom any correspondence should be addressed.

E-mail: [andoni.agirre@dipc.org](mailto:andoni.agirre@dipc.org) and [matteo.wauters@unitn.it](mailto:matteo.wauters@unitn.it)**Keywords:** variational quantum algorithms, quantum optimization, artificial intelligence, NISQ algorithms**Abstract**

The search for quantum algorithms to tackle classical combinatorial optimization problems has long been one of the most attractive yet challenging research topics in quantum computing. In this context, variational quantum algorithms (VQAs) are a promising family of hybrid quantum–classical methods tailored to cope with the limited capability of near-term quantum hardware. However, their effectiveness is hampered by the complexity of the classical parameter optimization which is prone to getting stuck either in local minima or in flat regions of the cost-function landscape. The clever design of efficient optimization methods is therefore of fundamental importance for fully leveraging the potential of VQAs. In this work, we approach parameter optimization as a sequential decision-making problem and tackle it with an adaptation of Monte Carlo Tree Search, a powerful artificial intelligence technique designed for efficiently exploring complex decision graphs. We show that leveraging regular parameter patterns deeply affects the decision-tree structure and allows for a flexible and noise-resilient optimization strategy suitable for near-term quantum devices. Our results shed further light on the interplay between artificial intelligence and quantum information and provide a valuable addition to the toolkit of variational quantum circuits.

**1. Introduction**

Overcoming the limitations of classical computing for problems hard to solve on classical machines, typically those requiring exponentially large resources, has been one of the main drives behind the development of quantum computation since the first pioneering works on quantum algorithms. Indeed, some celebrated examples, such as Grover's algorithm for unstructured search [1], its generalization to more complex problems such as amplitude estimation [2] and minimum finding [3], and Shor's algorithm for factorization [4] display significant speedup over classical methods, motivating further research in the quest for quantum advantage. However, the realization of fault-tolerant quantum computers, a prerequisite for fully exploiting this potential, remains a formidable challenge. In the interim, leveraging available technology necessitates exploring alternative strategies, such as variational quantum algorithms [5–8] (VQAs), to harness the computational power of current quantum devices.

VQAs employ a hybrid approach, combining classical optimization loops with (shallow) parameterized quantum circuits, offering a practical means of quantum computation with existing hardware. In the last years, they have been at the center of intense research activities and their proposals range from quantum optimization [6, 9] and ground state preparation [10–12] to quantum compiling [13, 14] and quantum

machine learning [15–17]. Designing efficient methods capable of dealing with hardware noise, for the classical optimization of the variational parameters, is crucial for their success. In this work, we propose a framework inspired by artificial intelligence techniques displaying remarkable resilience against noise.

One of the most used algorithms in the VQA family is the quantum approximate optimization algorithm [11, 18–24] (QAOA), which aims at finding an (approximate) solution to a classical combinatorial optimization problem mapped onto a cost Hamiltonian  $H_C$ , diagonal in the computational basis. To find its ground state, which minimizes the energy and thus also the classical cost function, QAOA starts from the ground state of an auxiliary mixing Hamiltonian  $H_M$ . Typically, one chooses  $H_M = \sum_i \sigma_i^x$ , as its ground state  $|+\rangle^{\otimes n}$  is easy to prepare and has a uniform weight on each state in the computational basis. Then, QAOA prepares the variational state

$$|\gamma, \beta\rangle = \prod_{i=1}^P e^{-i\beta_i H_M} e^{-i\gamma_i H_C} |+\rangle^{\otimes n}, \quad (1)$$

where  $P$  sets the circuit depth and  $(\gamma, \beta)$  is the array of variational parameters. Their optimal value is found by minimizing  $F_P(\gamma, \beta) = \langle \gamma, \beta | H_C | \gamma, \beta \rangle$ . QAOA has been applied to a wide range of problems, both classical and quantum, and many approaches have been proposed to improve its original formulation [23]. Examples include several heuristics to improve the parameter optimization [21, 25–28], reinforcement-learning guided exploration of the energy landscape [29, 30], and addition of further unitary operations to make the ansatz more expressive [31–34].

Despite the flexibility and popularity of VQAs, one of their main weaknesses appears in the classical optimization loop [35], which may suffer from hardware and measurement noise or the presence of flat regions in the cost function, known as barren plateaus [36]. Indeed, both effects either slow down the exploration of parameter space or prevent it from escaping local minima, in particular when the classical optimization algorithm relies on the gradient of the cost function. To tame these issues, gradient-free approaches have been suggested as appealing alternatives for the practical implementation of VQAs. However, even local minimization methods based on finite differences (also known as pseudo-gradient methods) suffer from the exponential scaling of resources when facing barren plateaus [37]. True gradient-free global optimizers are promising candidates to overcome such difficulties. Among these, reinforcement-learning [38] inspired approaches emerged as intriguing choices, both for the parameter optimization [29, 30, 39] and circuit design [32, 40].

In this context, Monte Carlo Tree Search [41] (MCTS) appears as a compelling gradient-free and robust optimization method for VQAs in the presence of noise. Being the backbone of many recent achievements in machine learning of complex combinatorial games, most notably chess and Go [42], MCTS has gained substantial fame and has been widely used for many different tasks across many different fields [43, 44]. As such, MCTS has already been successfully applied in the context of quantum annealing schedule optimization [39, 45], control problems [46] and circuit design [40]. Its problem-agnostic nature, the ease with which it can be integrated into deep learning frameworks, and, fundamentally, the different approach to the optimization task compared to more traditional algorithms, make it a promising candidate for enhancing the performance of VQAs with complex parameter landscapes, such as QAOA.

In this work, we investigate the potential of MCTS in identifying optimal strategies for QAOA applied to instances of the optimization versions of 3-satisfiability (3-SAT) and MaxCut, both NP-hard classical optimization problems [47]. We show that a suitably modified version of MCTS offers a compelling practical route to design robust optimal QAOA schedules, with exceptional resilience against noise in the cost function. The crucial modification to MCTS aims at exploiting the parameter concentration and regularity inherent in many optimal QAOA schedules [27, 48–50]. This version, which we denote as iterative search-space restricted MCTS (SSR-MCTS), progressively builds the solutions at each depth  $P$  from the previous optima, inspired by iterative methods presented in [11, 21]. On the opposite, we find that a vanilla version of MCTS spectacularly fails in navigating the QAOA energy landscape already at a small circuit depth. This happens because of the particular fractal-like structure of the leaf nodes, the endpoints of the decision tree, where good strategies for QAOA are uniformly and sparsely distributed among bad ones. Iterative SSR-MCTS drastically simplifies the leaf-node structure of the QAOA decision tree, allowing MCTS to reliably find optimal schedules for different choices of the system size, circuit depth, or problem instance. The combination of noise resilience, gradient-free optimization and good scalability in the quantum circuit volume makes our results well-suited for applications in current quantum hardware.

Our work showcases the emergence of optimization complexity in VQAs from the unusual perspective of sequential decision processes. Understanding the connection with other complexity estimators for the energy landscapes of QAOA can further unravel the interplay between artificial intelligence methods and quantum optimization, clarifying when the former provides valuable resources for NISQ devices.

The rest of the paper is organized as follows. In section 2 we recap the general structure of MCTS algorithms (section 2.1), describe the MCTS approach for QAOA (section 2.2) and explain the SSR variant and its iterative version (section 2.3). In sections 3 and 4 we present our results on instances of 3-SAT and unweighted MaxCut on 3-regular graphs, respectively. We draw our conclusion and present future outlooks in section 5. Further results and analysis are presented in the appendices: in appendix A we briefly discuss a combination of MCTS and gradient descent to construct a global optimization algorithm, in appendix B we look at an example MCTS decision tree for QAOA, in appendix C we test the performance of various iterative SSR-MCTS variants on noisy cost landscapes, and in appendix D we study the resource cost of our algorithm.

## 2. MCTS-guided QAOA optimization

In this section, we first give a brief introduction to the main concepts behind MCTS and its core elements. Then, we frame QAOA as a decision process and investigate the peculiar structure that makes it unsuitable for standard MCTS exploration. With this motivation, we describe in detail the modification needed to simplify the tree structure and optimize the policy for tackling QAOA. A reader already familiar with MCTS can skip directly to section 2.2.

### 2.1. Monte Carlo Tree Search

#### 2.1.1. General concepts

A decision tree is a structure that represents the possible paths originating from a sequential set of discrete action choices. Each branching point (*node*) of the tree represents a set of decisions taken insofar, while the *children* of that node are the possible choices stemming from it. We denote the starting point of the tree as the *root* node and the terminal states as *leaf* nodes. Each leaf node has an associated reward depending on the success of that particular strategy. As decision trees grow exponentially with their depth, clever algorithms are necessary to navigate such structures.

MCTS explores a given tree domain via random samples guided by some specified selection and playout policies, to find a path down the decision tree that maximizes a reward associated with the task at hand [41, 43]. Starting from the root node, it incrementally builds an asymmetric replica of the parts of the tree (*branches*) it deems most promising. The algorithm repeatedly runs a four-stage cycle, learning about the tree by growing the part stored in memory. At the same time, it keeps track and continuously updates the cumulative score  $w_i$  of each stored node  $i$  and the number of times it has been visited by the algorithm  $n_i$ . The cycles stop when the allotted computational budget is exceeded, at which point the algorithm makes a decision and suggests a strategy to follow. This usually corresponds either to the child node with the maximum average score  $w_i/n_i$ , or the one with the highest visit counts  $n_i$ . The larger the number of cycles, the more complete the reconstruction of the tree will be, leading to more accurate statistics of the nodes and improving the quality of the final choice. If the number of cycles is unbounded, MCTS is eventually equivalent to a brute force search and the policy converges to the optimal strategy [44].

#### 2.1.2. The MCTS cycle

The four stages of an MCTS cycle are (i) *selection*, (ii) *expansion*, (iii) *rollout*, and (iv) *backpropagation*.

- i. In the *selection stage*, starting from the root, the algorithm travels the part of the tree stored in memory by selecting for each node its child that maximizes a specified selection policy. The most common choice is the Upper Confidence Bounds applied to Trees (UCT) policy [41, 43], which selects node  $a^*$  among the children of node  $i$ ,  $A(i)$ , such that

$$a^* = \arg \max_{a \in A(i)} \left\{ \frac{w_a}{n_a} + C \sqrt{\frac{2 \ln n_i}{n_a}} \right\}, \quad (2)$$

where  $n_i$  is the number of times the agent has visited the parent node  $i$ . The first term in (2) rewards the exploitation of well-scoring paths, while the second one encourages the exploration of less-known branches. The constant  $C$  balances the two terms. The selection stage runs until a node with a child not yet added to the stored tree is selected.

- ii. In the *expansion stage*, one unexpanded child of the node reached in the selection stage is expanded, i.e. added to the tree in memory with a score and a visit count of zero, and the algorithm travels to the newly expanded node.
- iii. In the *rollout stage*, random moves are played from the newly expanded node, without expanding them, until a leaf node of the original tree is reached. The reward associated with that leaf node is evaluated in this stage.

- iv. In the *backpropagation stage*, the scores and visits are updated with the result from the rollout. The algorithm travels back to the root node increasing the visit counts of the nodes along the path by one and appropriately adding the reward obtained in the rollout to their cumulative scores.

The first two stages define the tree policy. They can be readily modified by changing the strategy of selection used in (i), and the way or the number of children that are expanded in (ii) [39, 51, 52]. The remaining two stages, where the leaf nodes are sampled and the expected rewards of different branches are updated, define the default policy of the algorithm. This policy can be modified by changing how the rollouts are carried out, such as by making the playouts follow some task-specific heuristics [52–54]. Overall, MCTS is independent of the specific problem it is facing, and can often be successfully used with very little knowledge of the corresponding action space. In cases where the standard version fails, the loose definition of its policies makes the algorithm highly flexible when developing problem-specific modifications [43, 44].

## 2.2. MCTS for quantum optimization

Using MCTS for QAOA parameter optimization requires formulating the task as a sequential decision problem. A straightforward way to achieve this is to build a decision tree of depth  $2P$  and branching factor  $b$ . At each depth  $i \in [1, 2P]$  the algorithm must choose a value for the parameter  $\theta_i$ , where  $\theta = \{\gamma_1, \beta_1, \dots, \gamma_P, \beta_P\}$ , out of  $b$  discrete options,  $\theta_i \in \{\theta_i^{(1)}, \dots, \theta_i^{(b)}\}$ , and the range of the parameters depends on the spectrum and the symmetries of the Hamiltonians generating the parameterized unitary operators. As we consider Hamiltonians with an integer spectrum, the variational parameters belong to the compact interval  $[0, 2\pi)$ . Moreover, since the cost function is real, it satisfies  $F_P(-\gamma, -\beta) = F_P(\gamma, \beta)$ , meaning that, if the different  $\theta_i^{(k)}$  are evenly spaced, for every path  $\theta$  there is an equivalent one where  $\theta_i^{(k)} \rightarrow 2\pi - \theta_i^{(k)}$ . Hence, we can restrict the domain of the first parameter to  $[0, \pi]$ , effectively halving the size of the tree to a total of  $\lceil \frac{1}{2} b^{2P} \rceil$  leaf nodes, without changing the discretization step.

This approach transforms a continuous optimization problem into a discrete one; however, we do not expect it to affect QAOA performances significantly. Indeed, the algorithm's efficiency is limited by the complexity of the energy landscape and the structure of the decision tree, not by the impossibility of exactly pinpointing the target minimum because it does not fall on the discrete grid. Yet, pathological cases might exist where extremely thin minima [22, 55] and wide flat regions require more clever discretization schemes than the regular ones.

Alternative formulations of the decision tree can also be devised, where the leaf nodes correspond to coefficients of a discrete basis for continuous functions with a compact domain. These can parameterize both smooth annealing schedules [39] or QAOA strategies [21]. Modulo minor differences depending on the specific parameterization of the schedule, the tree size will grow exponentially in the number of variational parameters. This is not a limitation for MCTS, as, in many cases, it is known to handle exponentially growing domains exceptionally well [43].

As we rely on the variational principle for the optimization, the reward collected at each leaf node must be related to the energy associated with the corresponding choice of parameters  $F_P(\gamma, \beta)$ . In this work, we chose a nonlinear mapping  $r(\theta) = e^{-\nu F_P(\theta)}$  for the rewards, with  $\nu$  chosen such that the reward starts rising when energies at or below the typical ranges are considered, and the algorithm can properly gauge the goodness of the parameters in the context of the problem at hand<sup>7</sup>.

In MCTS language, optimizing the QAOA parameters is a single-player game, meaning that no opposing agent is trying to reach a different goal. This has some consequences for the optimal strategy choice: in a two-player game, a meaningful strategy is based on the assumption that the opponent is trying to win as well, and high-scoring paths based on counterproductive choices of the other agent are usually not worth exploring [44]. This is not true in a single-player game, where the goal is maximizing the score considering all eligible strategies. In view of this, we implement some changes to MCTS proposed in [57]. We first modify the criterion with which MCTS decides the final move to play, by choosing the child that leads to the path of the random simulation that obtained the highest reward. Second, we expand the highest-scoring path of each turn all the way to the leaf node in between turns, thus ensuring that MCTS memorizes such good paths in their entirety. The search still follows the UCT selection policy in (2) for the selection stage, which can under or overestimate the potential of exploration of certain regions of the search domain. But with these changes, we have the certainty that the final parameters MCTS chooses will be at least as good as those in the best memorized sequence. In the following, we will denote this modified variant as MCTS-SP (single-player).

<sup>7</sup> To ensure the theoretical bounds on the regret of the playing agent [56], the UCT selection policy with  $C = \sqrt{2}$  requires the rewards to lie in  $[0, 1]$ , although this is not strictly necessary for MCTS to perform well, and should be studied on a case-by-case basis [52].

In the following sections, however, we will still consider the standard version of MCTS. First, it is interesting to see how it compares to the single-player variant and, second, we expect the single-player modifications to no longer be beneficial in some particularly relevant cases, such as when considering a noisy cost landscape (see appendix C). In this noisy situation, the task loses its deterministic nature, and the memorized parameter sequences will likely contain inaccurate expected rewards, making the standard final choice criteria more appropriate, and rendering path memorization irrelevant.

### 2.3. Iterative search-space restricted MCTS

MCTS has been proven to work well for optimizing annealing schedules [39], although gradient-descent algorithms perform slightly better in idealized settings (noiseless and smooth cost functions), thanks also to the continuous parameter spaces [45]<sup>8</sup>. However, in our study, we found that standard MCTS is unsuited for QAOA as soon as  $P > 1$ . The modified single-player version obtains improved results but quickly fails to learn good strategies too, even in shallow circuits, as we will discuss later on in section 3. Therefore, MCTS needs to be adapted further to perform well in QAOA parameter optimization, leveraging on the structure of its energy landscape.

In this section, we describe how to modify the standard MCTS algorithm to successfully tackle QAOA. The core idea is not to modify the MCTS algorithm *per se*, but to progressively modify the space of eligible strategies. We call this method iterative search-space restriction (SSR). It consists of constraining the action space of the decision tree nodes as the circuit depth  $P$  increases, according to the knowledge of previous solutions for  $P - 1$ . Indeed, QAOA optima often display regular structures and patterns in the optimal parameters [11, 21, 30, 49], look ahead to figure 4(b) for an example. This property stems from the presence of optimal discretized quantum annealing schedules, which are typically robust with respect to small parameter changes in the Hamiltonian, system size, and number of time steps. These regular patterns appear as a monotonic increase (decrease) of the  $\gamma_i$  ( $\beta_i$ ) with the discrete-time index  $i$ , approaching a continuous function as  $P \rightarrow \infty$ . Hence, when increasing the number of layers, the new optimal parameters typically resemble a finer interpolation of the optimal schedule for the previous circuit depth.

For problem types/instances that exhibit these properties, we can implement an SSR-MCTS step in the following way: suppose that, for a depth  $P$ , the optimal parameters  $\theta^* = \{\gamma_1^*, \beta_1^*, \dots, \gamma_P^*, \beta_P^*\}$  are known. We then let MCTS search for the optimal  $P + 1$  parameters  $\tilde{\theta}^* = \{\tilde{\gamma}_1^*, \tilde{\beta}_1^*, \dots, \tilde{\gamma}_{P+1}^*, \tilde{\beta}_{P+1}^*\}$  in a mesh of  $b$  discrete values per parameter within

$$\begin{aligned}\tilde{\gamma}_i^* &\in [\gamma_{i-1}^* (1 - \delta_{P+1}), \gamma_i^* (1 + \delta_{P+1})], \\ \tilde{\beta}_i^* &\in [\beta_{i-1}^* (1 - \delta_{P+1}), \beta_i^* (1 + \delta_{P+1})],\end{aligned}\quad (3)$$

where  $i \in [0, P + 1]$ . The values for  $\beta_0^*$ ,  $\gamma_0^*$ ,  $\beta_{P+1}^*$  and  $\gamma_{P+1}^*$  can generally be heuristically chosen values that limit the range of the new parameters, and  $\delta_{P+1}$  is a *softening* constant that specifies how much MCTS is allowed to search beyond the bounds of the restricted regions of parameter space for depth  $P + 1$ . An example of a  $P = 3$  to  $P = 4$  search-space restricted search, with  $\delta_4 = 0$ , is depicted in figure 1.

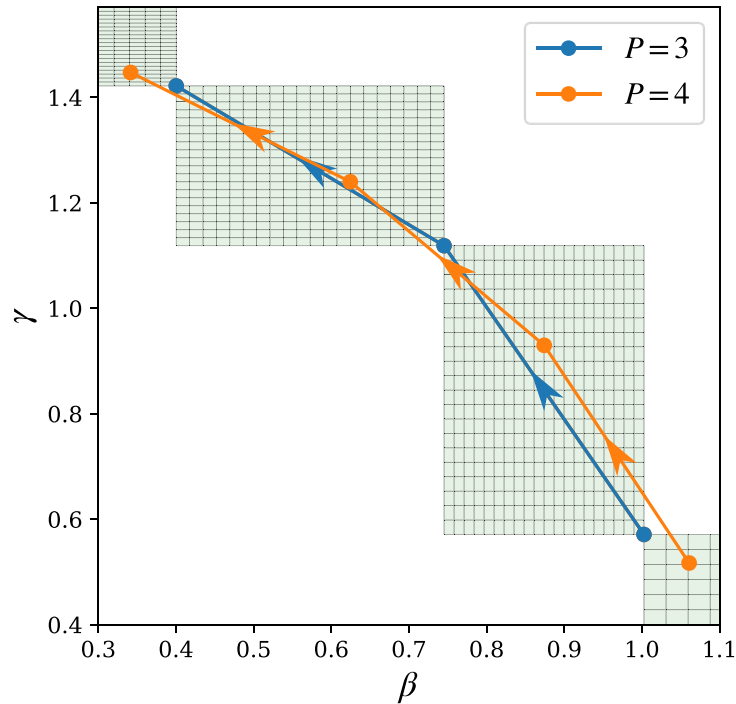
This method can be employed to look for the optimal QAOA angles one layer deeper when we know the optima for the previous depth, and its iterative application starting from  $P = 1$  allows for a systematic restriction of the action space by gradually increasing the number of parameters, without any prior knowledge of the optimal schedule. In this way, the parameters suggested by SSR-MCTS at depth  $P$  are used to determine how the  $P + 1$  search is restricted. The different  $\delta_p$  for each step can be adjusted to relax the restriction and help the algorithm correct its potential inaccuracies in previous circuit depths<sup>9</sup>. In this iterative case, the boundaries of the first and last parameters ( $i = 0$  and  $i = P + 1$ ) do not need to be specified and can generally be unrestricted, i.e. set by the bounds of the original  $P = 1$  action space<sup>10</sup>. This procedure is independent of the particular Hamiltonian instance  $H_C$  one is facing: as the number of QAOA layers increases, it progressively generates a restricted action space tailored to the problem at hand, learning the structure of the parameter space from previous iterations.

As we will see in subsequent sections, SSR-MCTS makes the energy landscape seen by the MCTS agent easier to navigate, as regions of the parameter space that are uninteresting or even hindering the optimization

<sup>8</sup> We expect the real potential of MCTS as an optimizer to lie in robust handling of noisy or complex cost-landscapes where gradient-descent optimizers are prone to get stuck in local minima.

<sup>9</sup> To make the method fully adaptive, one might introduce automatic ways to control the softening constants during the restriction procedure, which within the present implementation have to be set manually.

<sup>10</sup> In cases where their optimal values vary slowly with  $P$  and high circuit depths are considered, the restricted regions of the first and last parameters would shrink much slower relative to the others, potentially causing them to lose the required resolution to select a value close to the optimal one. This could be easily circumvented by increasing (decreasing) the first (last) boundary values for the edge parameters to lie at most a fixed distance away from the previous optima, e.g.  $\gamma_{P+1}^*$  would become  $\gamma_{P+1}^* - \varepsilon(\gamma_{P+1}^* - \gamma_P^*)$ , for some  $\varepsilon < 1$ .



**Figure 1.** Example of a search space restriction step (with  $\delta_4 = 0$  and  $b = 21$ ) when the depth is increased from  $P = 3$  to  $P = 4$ . Each point in the plot represents a pair of parameters  $(\gamma_i, \beta_i)$  and the arrows indicate the direction of increasing  $i$  for the parameter pairs. The SSR-MCTS algorithm only searches for the  $P = 4$  optimal parameters (orange) within the grid points of the highlighted green rectangles, which are generated from the  $P = 3$  optima (blue). In this example with  $\delta_4 = 0$ , all of the space outside of the green rectangles is not seen by SSR-MCTS.

task are gradually discarded. The complexity reduction of both the decision tree and the associated energy landscape is particularly evident when  $P$  is large. Indeed, the regularity regions become smaller as  $P$  increases, while the branching factor remains unchanged, making the discretization of the parameters finer at each step. Pictorially, it can be seen as zooming in the region around the target minimum with greater resolution at each iteration. For this reason, and despite this progressive simplification of the decision tree, SSR does not affect its exponential scaling with  $P$ ; in fact, after the first run, the restriction generally breaks inversion symmetry, effectively doubling the number of leaf nodes, with our formulation of the problem.

To keep the overall computational cost under control, we will consider MCTS with a number of function evaluations growing linearly with the circuit depth, leading to an overall quadratic scaling for iterative SSR-MCTS starting from an unrestricted  $P = 1$  search.

### 3. Results—3-SAT

#### 3.1. The 3-SAT problem

As a first benchmark for the MCTS-guided quantum optimization, we consider instances of boolean 3-SAT problems. 3-SAT is one of the most well-known NP-complete problems [58] and a standard benchmark for quantum optimization algorithms [39, 59]. Given a set of  $n$  classical binary variables  $x_i \in \{0, 1\}$ , it consists in finding whether a particular assignment  $\mathbf{x}^*$  exists such that a boolean expression consisting of a logical conjunction of  $m$  clauses

$$\varphi(\mathbf{x}) = \bigwedge_{a=1}^m C_a(\mathbf{x}) \quad (4)$$

is satisfied. In 3-SAT each clause is a disjunction of three literals  $C_a = l_{a_1} \vee l_{a_2} \vee l_{a_3}$ , with  $l_{a_j}$  being  $x_i$  or its negation  $\neg x_i$ , where  $i = a_j$  is the index of the  $j$ th bit entering clause  $C_a$ . The hardness of the satisfiability problem depends on the density of clauses  $\alpha \equiv m/n$ . It is in the vicinity of  $\alpha_c \sim 4.3$ , where 3-SAT has a well-known phase transition from satisfiable to unsatisfiable instances [60], that a backbone of strongly constrained variables emerges and one finds the hardest problem instances, with which existing 3-SAT solvers struggle the most [61].

A given instance  $\varphi$  can be encoded into a quantum Hamiltonian as

$$H_C = \sum_{a=1}^m \frac{1}{8} (\mathbb{1} + g_{a_1} \sigma_{a_1}^z) (\mathbb{1} + g_{a_2} \sigma_{a_2}^z) (\mathbb{1} + g_{a_3} \sigma_{a_3}^z), \quad (5)$$

where the sum runs over all clause indices,  $\sigma_{a_j}^z$  represents the Pauli operator corresponding to the  $j$ th bit in clause  $a$ , and with  $g_{a_j}$  taking the value  $+1$  if the corresponding variable enters the problem statement (4) as the negation  $\neg x_{a_j}$ , or  $-1$  if it enters as  $x_{a_j}$  instead. Equation (5) generates long-range interactions with up to three-body terms, which require careful embedding to be encoded in present-day quantum hardware [62]. Typically, they can be mapped with polynomial cost onto effective Ising Hamiltonians with equivalent low-lying states [63].

The expectation value of the cost Hamiltonian equation (5) gives the average number of unsatisfied clauses in the state under consideration. Finding the ground state of said Hamiltonian is thus equivalent to solving the NP-hard optimization version of the corresponding 3-SAT problem, also known as MAX-3-SAT [58], where the (harder) task is finding the bitstring that violates the minimum number of clauses. Clearly, this also solves 3-SAT, as the problem is satisfiable if and only if the ground state energy is zero. The phase transition of MAX-3-SAT happens for similar clause densities to  $\alpha_c$ , although it affects the hardness of the problem differently<sup>11</sup>. Even though we are technically solving MAX-3-SAT, in the following we will refer to the problem instances as ‘3-SAT instances’, as it is customary for the sake of convenience.

In this work, we work with a set of 3-SAT instances with  $\alpha = 3$ , each with a unique bit-string satisfying equation (4). Despite this density of clauses being lower than the threshold for the complexity phase transition in the thermodynamic limit [60], having a non-degenerate ground state makes the optimization hardness comparable to random problem instances in the critical regions [39, 64]. From a quantum annealing perspective, this is reflected in the quantum Hamiltonian having an energy gap that decreases exponentially with the qubit number, making annealing unfeasible for large systems. 3-SAT is known to be a formidable problem for QAOA, requiring large circuit depths to reach the ground state with high fidelity ( $P \gtrsim 45$  for 0.95 fidelity on 6 qubits [65]). However, as long as the projection over the ground state is large enough, an approximate optimization is sufficient to sample it with a finite probability. As deep variational circuits are unfeasible for current hardware, we will limit the number of layers to  $P \leq 10$ . As a mixer Hamiltonian for QAOA we use a transverse field

$$H_M = \sum_i \sigma_i^x, \quad (6)$$

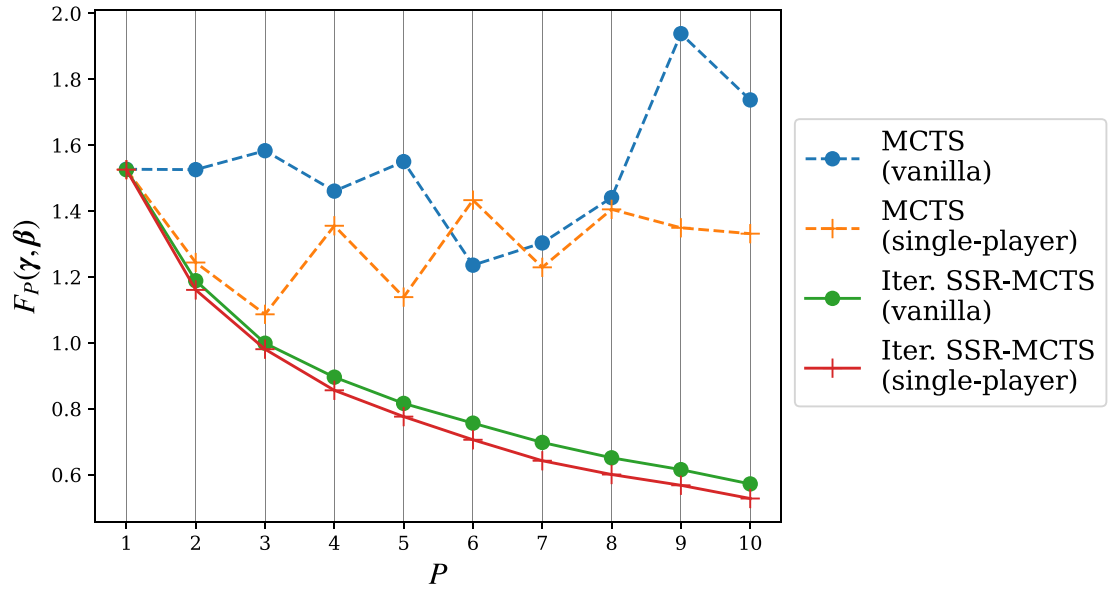
which we adopt throughout this paper. The code was developed using the open-source QuTip library [66, 67] and is available in [68]. The data supporting our findings are available in [69].

### 3.2. Unrestricted MCTS vs SSR-MCTS

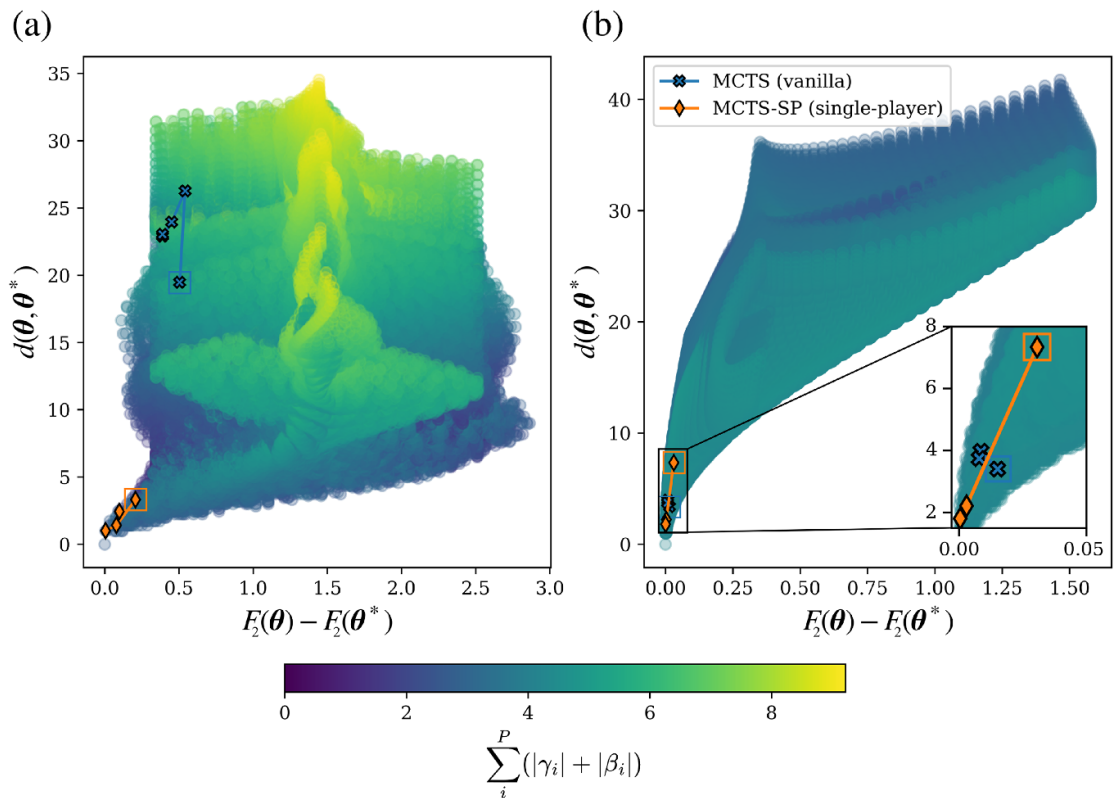
Let us now test MCTS’s ability to select good QAOA strategies for 3-SAT problems. Based on the discussion presented in section 2, we have four optimizer variants: vanilla MCTS, MCTS-SP, iterative SSR-MCTS, and iterative SSR-MCTS-SP. In figure 2, we report their performance, measured with the variational energy at the end of the optimization, as a function of the number of QAOA layers. The data refer to an example 3-SAT instance with  $n = 7$  qubits. The most striking feature is that vanilla MCTS fails spectacularly as soon as  $P > 1$ . The single-player variant mildly improves the performance, but again fails already for a modest circuit depth  $P = 4$ . On the other hand, iterative SSR completely changes the algorithm’s effectiveness, making the ground state approximation more and more accurate as  $P$  increases, and the single-player modifications now further improve the overall performance of iterative SSR-MCTS. This behavior appears consistently across all tested instances and problem sizes.

The dramatic effect of SSR already at  $P = 2$  is due to its impact on the leaf-node structure. To appreciate it, in figure 3 we compare the energy distribution among the leaf nodes for a depth  $P = 2$  when MCTS has the unrestricted  $[0, 2\pi)^{2P}$  space to search in (a), with SSR generated from the  $P = 1$  optima in (b). Each leaf node can be identified by a tuple composed by its energy difference to the optimal choice  $F_P(\boldsymbol{\theta}) - F_P(\boldsymbol{\theta}^*)$ , the corresponding tree distance  $d(\boldsymbol{\theta}, \boldsymbol{\theta}^*)$  (a measure of the deviation from the path of the optimal leaf, defined

<sup>11</sup> The hardness of MAX-3-SAT also depends on the clause density  $\alpha$ , and one expects a phase transition to occur for some  $\alpha \leq \alpha_c$ , as MAX-3-SAT subsumes the complexity of 3-SAT. As  $\alpha$  increases past  $\alpha_c$  and the problem instances become more and more over-constrained, the hardness of MAX-3-SAT, unlike that of 3-SAT, never decreases. Such high clause densities make it easy to find that problem instances are unsatisfiable, but very hard to find their optimal bitstrings. We have gone from an *easy-hard-easy* type of transition for 3-SAT to an *easy-hard* one for MAX-3-SAT [61].



**Figure 2.** Comparison of the performance of different versions of MCTS on a  $n = 7$  example 3-SAT instance. Energies of the final states are shown as a function of circuit depth for vanilla MCTS (blue), MCTS with single-player modifications (orange), iterative SSR-MCTS (green), and iterative SSR-MCTS with single-player modifications (red). The number of MCTS cycles increases linearly with  $P$ .



**Figure 3.** Distribution of all leaf nodes in the tree representing all possible combinations of parameters MCTS can choose for a circuit depth of  $P = 2$  and a branching factor  $b = 30$ , in an example 3-SAT instance. (a) Leaf nodes from the full parameter space (up to inversion symmetry), (b) leaf nodes within the restricted parameter space defined according to the  $P = 1$  optimal angles. Each leaf node is represented according to its difference in energy and tree distance (defined in appendix B.1) from the optimal node  $(F_2(\theta) - F_2(\theta^*), d(\theta, \theta^*))$ , and the color of each leaf node represents the sum of the corresponding QAOA parameters (run time). The nodes corresponding to the best possible sets of parameters are the ones on  $(0,0)$ , the bottom-left corners of the figures. For each of the two cases, we also show the five leaf nodes selected by both standard MCTS (blue crosses) and MCTS-SP (orange diamonds) after five runs, each with different numbers of MCTS cycles per turn. The symbols in the squares correspond to the results with the fewest number of cycles, and the lines connect the leaf nodes in order of increasing cycle numbers in the set  $\{200, 600, 1000, 1800, 3500\}$ .

in appendix B.1), and the effective run time  $\tau = \sum_i (|\gamma_i| + |\beta_i|)$ . Figure 3 shows the relation between these quantities.

Comparing the two panels, the effect of SSR emerges strikingly, as it deeply modifies the correlation between energy and distance from the optimal strategy. In panel (a) there are many leaf nodes with high energy but a small distance from the absolute minimum, meaning many low-energy nodes can be effectively hidden among siblings with undesirable energies. This makes gauging the quality of different moves very challenging for MCTS. Furthermore, in the case of the standard variant, even when the algorithm eventually finds a path leading to some optimal leaf node, its contribution to the exploration of the tree in future cycles and to the final chosen move is almost inconsequential: the optimal leaf node was most likely sampled very late in the search to modify the nodes' average scores (and subsequent search paths) significantly. Indeed, this feature is exactly one of the main factors that make MCTS struggle in single-agent contexts [43], suggesting that its performance might benefit from tailored policy modifications<sup>12</sup>.

If we look at SSR in panel (b), instead, a large energy difference also corresponds to a node far from the optimum. This means that good paths down the tree are more concentrated in particular branches and can be reached via similar parameter choices in the early turns of the game, making MCTS able to discover and learn strategies to navigate the cost landscape adeptly. The energy interval spanned by the possible QAOA schedules is also reduced, notice the difference between the scales on the x-axis. SSR gets rid of all terminal nodes corresponding to excessively large run times and highly excited states, thus focusing on the parameter region where fast-running and low-energy minima reside. More details on the leaf node structure are given in appendix B.

The difficulties of vanilla MCTS persist if one tries to improve its performance by increasing the number of cycles. An example is provided by the set of blue crosses in both panels of figure 3. Starting from 200 MCTS cycles per turn (boxed cross) we increase this number to  $\{600, 1000, 1800, 3500\}$ . The links connect MCTS results with a growing number of cycles. When MCTS has to search through the whole parameter landscape, panel (a), it typically gets stuck on leaf nodes far away from the optimal one. On the contrary, it consistently finds well-scoring nodes near the optimal one for the SSR case.

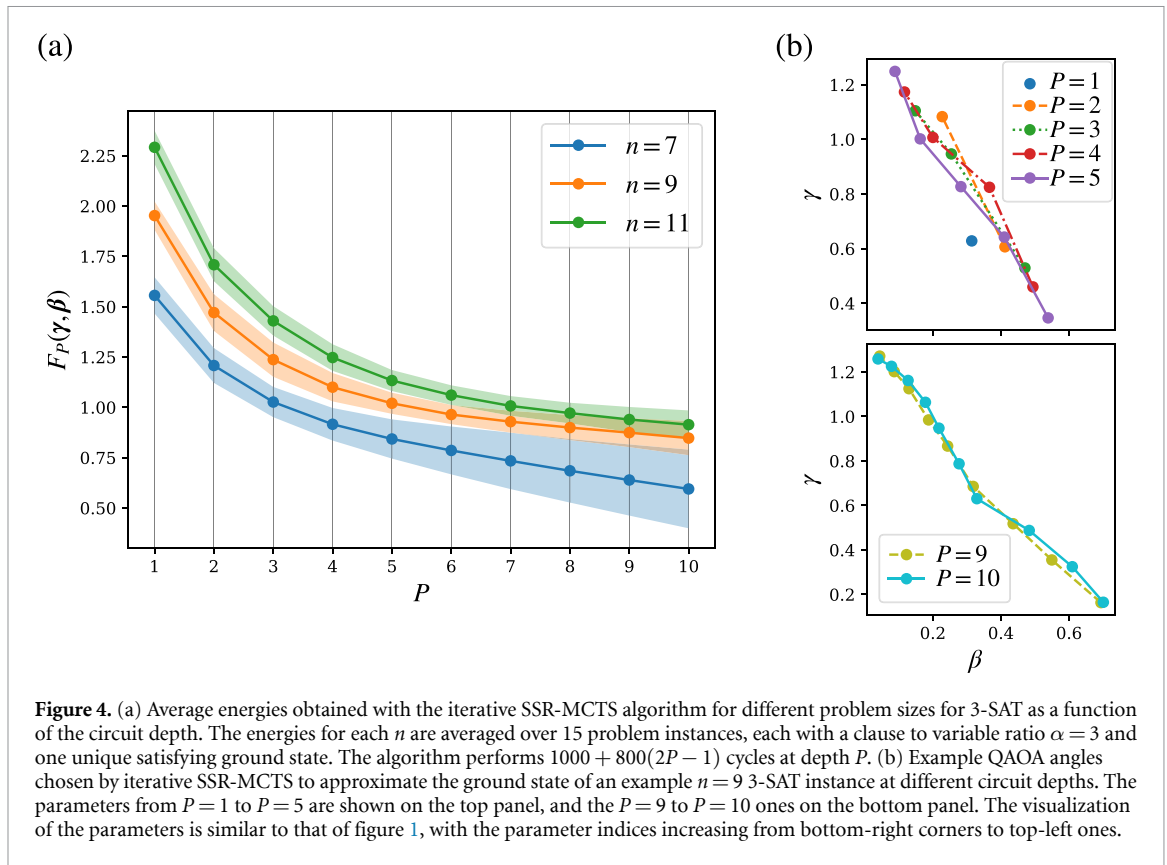
Single-player modifications (orange diamonds) can partially cure this problem, or enhance the performance of SSR-MCTS. In the problem represented in figure 3, MCTS-SP approaches the optimal leaf node as the number of cycles is increased both with and without SSR. This modified version, however, relies on finding the optimal leaf nodes within the rollout stages, which we expect to become exponentially harder in the non-SSR case, as the number of leaf nodes and local minima increases with circuit depth. This makes this single-player modified version only usable in practice to find the minima for low-depth QAOA unless iterative SSR is employed to find regular parameters; indeed, its performance plummets with increasing depth, and it fails already for  $P = 4$  (see figure 2) unless we allow the number of cycles to grow exponentially in  $P$ .

### 3.3. 3-SAT with iterative SSR-MCTS

Now that we had a glimpse into the importance of search-space restriction for transforming the QAOA decision tree into one that MCTS can efficiently find optimal strategies in, let us delve into more details of the iterative SSR-MCTS protocol we use. In each run, the algorithm performs 1000 initial cycles plus 800 more per turn, excluding the final one, while selecting the parameters. The last parameter to choose does not require further stochastic rollouts since the best move can be selected straightforwardly. Hence, even though the search space scales exponentially with  $P$ , the number evaluations of  $F_P(\gamma, \beta)$  only increases linearly with growing circuit depth,  $n_{\text{fev}} = 1000 + 800(2P - 1)$ , as two new turns are added at each subsequent  $P$ . See appendix D for an analysis of the performance dependence on the number of training cycles. We set  $C = \sqrt{2}$  in the selection policy and  $\nu = 1/2$  for the exponential mapping between energy and reward discussed in 2.2. At each depth  $P$ , the range of the parameters is bounded according to the parameters proposed for the previous depth, and a branching factor of  $b = 30$  is used to build the tree within the restricted space. The boundaries are softened with the values  $\delta = \{0, 0, 0.1, 0.05, 0.04, 0.03, 0.02, 0.01, 0.01\}$ , where  $\delta_P$  indicates how much the search is allowed to search outside of the regular regions at depth  $P$ , according to (3). We fix  $\delta_1 = 0$  since we are already considering the entire space at the shallowest depth, and  $\delta_2 = 0$  because we expect the algorithm to find the  $P = 1$  optimal parameters in every case. We set a nonlinear decrease of  $\delta_P$  with the circuit depth when  $P \geq 2$ .

Figure 4(a) reports the results for QAOA with iterative SSR-MCTS for different 3-SAT instances of varying problem sizes  $n = 7, 9, 11$  and  $m = 3n$  clauses. Each data point is the average over 15 problem instances, while the shaded areas indicate their standard deviation. The algorithm performs reliably as  $n$

<sup>12</sup> There are also cases in which the standard algorithm performs well in single-player domains, look at [41, 70–72], but unrestricted QAOA parameter optimization, as a hard deterministic single-player game, happens to not be one of those games.



**Figure 4.** (a) Average energies obtained with the iterative SSR-MCTS algorithm for different problem sizes for 3-SAT as a function of the circuit depth. The energies for each  $n$  are averaged over 15 problem instances, each with a clause to variable ratio  $\alpha = 3$  and one unique satisfying ground state. The algorithm performs  $1000 + 800(2P - 1)$  cycles at depth  $P$ . (b) Example QAOA angles chosen by iterative SSR-MCTS to approximate the ground state of an example  $n = 9$  3-SAT instance at different circuit depths. The parameters from  $P = 1$  to  $P = 5$  are shown on the top panel, and the  $P = 9$  to  $P = 10$  ones on the bottom panel. The visualization of the parameters is similar to that of figure 1, with the parameter indices increasing from bottom-right corners to top-left ones.

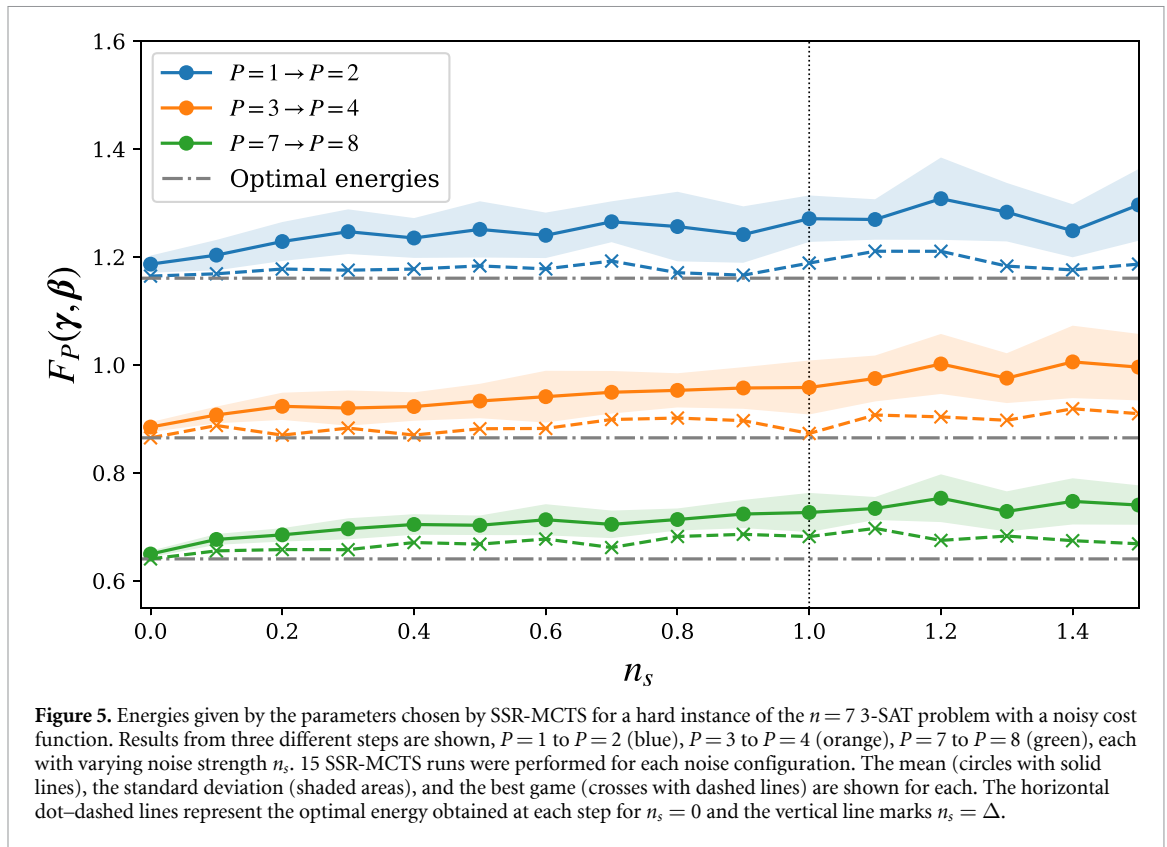
increases, as shown by the monotonic decrease of the cost function  $F_P(\gamma, \beta)$  as a function of the depth  $P$ . The energy remains high, centered around the lowest excitation, because of the known reachability deficit for QAOA applied to satisfiability problems [65].

The effect of SSR on the QAOA schedule can be appreciated from figure 4(b), where we show an example set of parameters chosen by iterative SSR-MCTS for different depths on a 3-SAT instance with  $n = 7$ . Each dataset is a sequence  $(\beta_i, \gamma_i)$  with the first pair corresponding to the point closer to the lower-right corner of the plot. As  $P$  increases, the optimal parameters clearly approach a regular behavior, linked to optimal digitized-quantum-annealing schedules and known to be associated with optimal minima strategies robust with respect to small parameter and size variations [11, 21, 30]. While iterative SSR is the analog of other iterative optimization methods of QAOA parameters [11, 21] in the context of MCTS, it is remarkable that it completely changes standard MCTS performance already at  $P = 2$ . While the regularity at large circuit depth is expected, its appearance is noteworthy when the search space is only moderately restricted.

### 3.4. Noisy rewards

A key aspect of any optimization strategy is its capacity to handle noisy energy landscapes. This is particularly relevant in practical implementations of VQAs, where expectation values are derived from measurements with stochastic outcomes. Additional qubit errors in the quantum circuit further complicate the search for optimal parameters. In this work, we focus solely on measurement noise and defer a more comprehensive analysis to future research. As monitoring uncertainty can be mitigated by increasing the number of shots, in the following, we explore the robustness of SSR-MCTS under varying levels of measurement noise.

We model the measurement uncertainty by introducing a random Gaussian variable to the rewards during the rollout phase of each MCTS cycle. Consequently, the algorithm must select variational parameters based on noisy data accumulated in the statistics of the decision tree. Specifically, we use random variables  $\mathcal{N}(0, n_s)$  with zero mean and a standard deviation determined by the noise level  $n_s$ . This addition modifies the reward function as  $F_P(\theta) \rightarrow F_P(\theta) + \mathcal{N}(0, n_s)$ . For evaluation purposes, the final assessment does not include this random noise. MCTS generally handles noisy rewards well though its effectiveness can vary based on factors like noise intensity, problem structure, and algorithm implementation [44]. The relevant energy scale for comparing noise strength is the typical energy gap  $\Delta$  between eigenvalues. As long as  $n_s \ll \Delta$ , MCTS should effectively distinguish between leaf nodes leading to states close to different eigenstates, as the node scores are already subjected to uncertainty due to stochastic sampling. However, if the noise strength approaches the gap  $n_s \sim \Delta$ , it can, in principle, significantly affect the estimation of optimal actions, making



it difficult to differentiate between strategies. In our 3-SAT formulation (see equation (5)), the energy units correspond to the number of violated clauses, with the minimum relevant energy gap being  $\Delta = 1$ .

To test the protocol's robustness against noise, we focus on the hardest 3-SAT instance (smallest energy gap in a QA setting) with  $n = 7$  and repeat 15 SSR-MCTS runs, without the single-player modifications (see appendix C for a comparison with other variants and COBYLA [73]), for 3 different jumps on  $P$  between 1 and 8, and  $n_s \in [0, 1.5]$ . Remarkably, MCTS has no issues finding the optimal parameters as it is able to average out the noise over different cycles even for large noise strengths  $n_s \gtrsim \Delta$ , as we show in figure 5. As the noise strength is increased, the average performance gets mildly worse (full circles with solid lines) but it is often able to find very good parameters in one of the runs. Indeed, the best results out of the 15 repetitions (crosses with dashed lines) drift away even more slowly than the average from the noise-free QAOA energy, as  $n_s$  increases. This remarkable robustness makes iterative SSR-MCTS well-suited for applications in noisy quantum devices, as it does not rely on gradient estimations and the uncertainty on the reward estimate can be easily taken into account when designing how the leaf nodes are sampled.

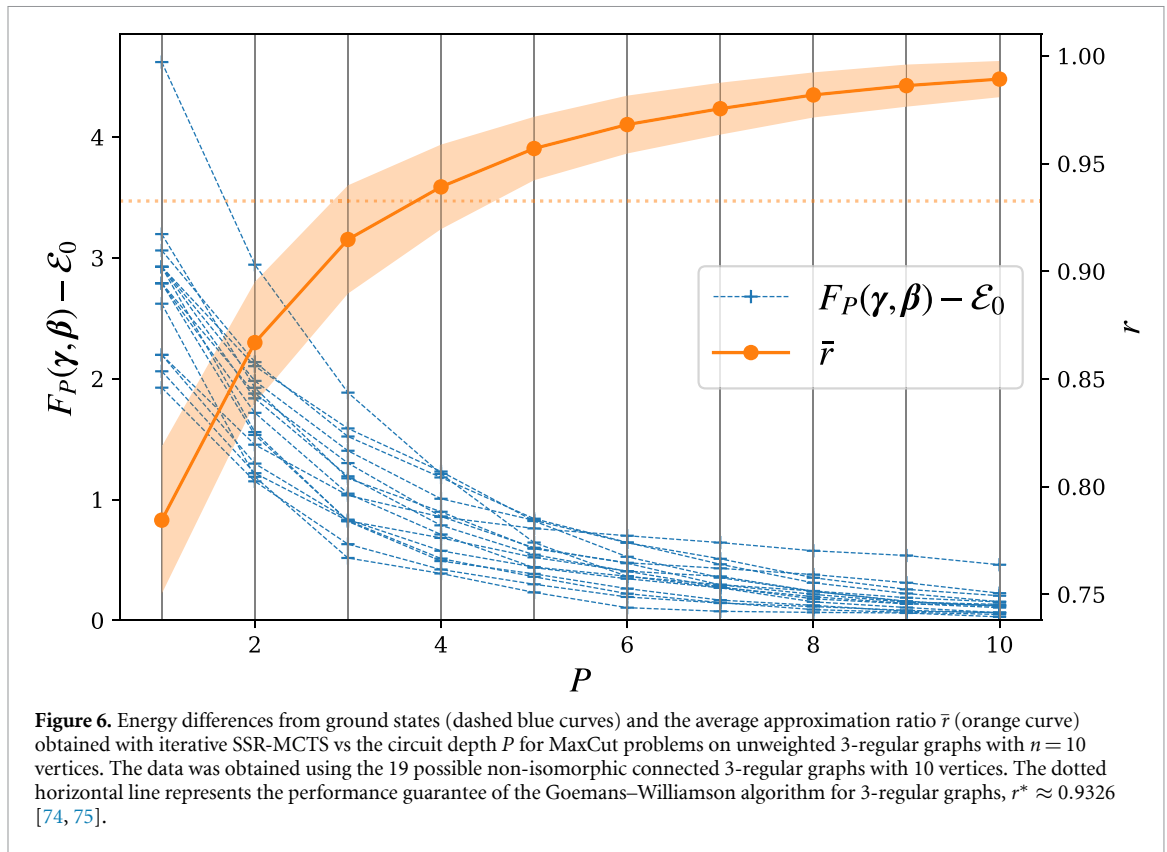
#### 4. Results—unweighted MaxCut

To confirm that our results are not unique to 3-SAT, we repeat the performance analysis on unweighted MaxCut instances. MaxCut is another prototypical NP-hard optimization problem, extensively studied in QAOA contexts [18, 21, 76–78]. It consists in classifying the vertices  $V$  of a graph  $G(V, E)$  into two sets such that the number of edges in  $E$  joining vertices between the different sets is maximal. The problem Hamiltonian for MaxCut can be expressed as an antiferromagnetic Ising model on  $G$ ,

$$H_C = \sum_{\langle i, j \rangle \in E} \frac{1}{2} \left( \mathbb{1} + \sigma_i^z \sigma_j^z \right), \quad (7)$$

with each ferromagnetic edge carrying an energetic penalty of 1 unit. The size of the cut for a configuration  $|\sigma\rangle$  with energy  $\mathcal{E} = \langle \sigma | H_C | \sigma \rangle$  on a graph with  $N_E$  edges will be given in this representation by  $C(\mathcal{E}) = N_E - \mathcal{E}$ . The ground state energy  $\mathcal{E}_0$  thus gives the maximum cut  $C_{\max} \equiv C(\mathcal{E}_0)$ . A common figure of merit for MaxCut is the approximation ratio  $r$ , which for an energy  $\mathcal{E}$  is defined as

$$r(\mathcal{E}) = \frac{C(\mathcal{E})}{C_{\max}}, \quad (8)$$



such that  $r = 1$  if the optimal cut is found. In figure 6 we show the results of iterative SSR-MCTS for 19 MaxCut problem instances on  $n = 10$  unweighted 3-regular graphs. We plot both the energy for each individual instance (blue crosses) and the average approximation ratio over all instances (orange circles). The latter is compared to the performance guarantee of the Goemans–Williamson (GW) algorithm [74, 75], which is a standard classical reference for MaxCut problems. The convergence to the ground state is smooth with increasing values of  $P$ , similar to what we observed for 3-SAT problems.

The accuracy of iterative SSR-MCTS is qualitatively similar to the results obtained with warm-start gradient descent approaches [21]. At  $P = 4$ , both optimization strategies of QAOA schedules beat, on average, the GW algorithm. At larger values of  $P$ , when  $E \rightarrow 0$ , the discretization of the parameter space marginally limits the performance of our method, preventing the algorithm from reaching the exact ground state ( $r = 1$ ) even if reachable by the variational QAOA ansatz. For instance, at  $n = 10$  and  $P = 10$ , we find  $1 - r \simeq 0.01$ , one order of magnitude larger than [21]. A larger branching factor in the decision tree would result in a finer resolution of the parameters, which mitigates this limitation at very small residual energies, at the cost of making the tree exploration more resource-expensive. Furthermore, improved search-space restriction schemes for the edge parameters, such as searching a limited distance from the previous optima as suggested in section 2.3 or making the discretization non-uniform (finer closer to the previous optima), could be another step in fully optimizing the performance of iterative SSR-MCTS for deep circuits. In general, however, we can not expect MCTS to outperform gradient-based algorithms in ideal simulations where the cost-function is noiseless *and* the energy landscape have only a few smooth minima, such as the regular MaxCut instances tested here. In experiments, instead, it can be advantageous because of its resilience against unavoidable noise. Furthermore, with single-shot measurements, sampling the ground state with high probability does not require perfect fidelity with the final variational state.

## 5. Conclusions and outlook

In this paper, we analyzed the performance of MCTS as a gradient-free optimizer for QAOA variational parameters. Despite its success in designing optimal schedules for quantum annealing [39], vanilla MCTS fails when faced with the complex energy landscape of QAOA. This complexity emerges as fractal-like decision trees, characterized by the close proximity of low- and high-energy leaf nodes, where it is impossible to discriminate between good and bad branches in polynomial time reliably. We overcame these difficulties by restricting the search space of QAOA parameters based on the knowledge of optimal parameters obtained

for shallower circuits and leveraging the well-known presence of regular schedules in QAOA [11, 21, 27, 30, 49, 50]. By testing it on 3-SAT and MaxCut instances, we showed that iterative SSR-MCTS provides a systematic way to optimize QAOA parameters within the MCTS framework. Preliminary numerical results indicate its performance is comparable with other heuristic optimization methods while offering at the same time great resilience against noise, suggesting it effectively accommodates both the potentiality and limitations of NISQ applications.

Moreover, the flexibility of MCTS suggests further possible development. First, further modifications to the tree policy can be envisioned to improve the performance and tailor it to specific VQA applications. For instance, the algorithm would likely benefit from changes to the selection criterion itself [51], such as using a different kind of policy entirely [79, 80] or modifying the UCT to favor the exploration of nodes with a large reward variance [52, 81]. MCTS can also be efficiently integrated with neural networks [39, 42] to speed up the nodes' score evaluations and improve the transferability between different problem instances or sizes. This will likely make the algorithm more robust at the price of a greater computational cost in the training phase. Finally, we expect MCTS to be paired well with many modifications of QAOA as well as other VQAs. Being problem agnostic and relying only on a well-defined reward (cost) function, which is a key ingredient of VQAs, iterative SSR-MCTS represents a reliable workhorse to add to the toolkit of hybrid quantum–classical algorithms.

Optimizing VQA parameters is generally a challenging task due to features such as barren plateaus [36, 37] and the proliferation of local minima [28], which hinder efficient exploration of the energy landscape. Warm initialization of variational parameters has been shown to mitigate these difficulties [82–84], and iterative SSR-MCTS essentially follows a similar philosophy. Despite this common ground, standard variational optimization and MCTS approach the problem from fundamentally different perspectives: The former treats QAOA as a function optimization problem over a high-dimensional manifold, while the latter frames it as a sequential decision-making process.

This distinction motivates further studies to investigate how the structural properties of the decision tree in MCTS relate to the geometry of the energy landscape itself. In particular, could the observed 'fractalization' of the decision tree be indicative of the roughness or flatness of the energy landscape [55]? Understanding such connections could offer deeper insights into the intrinsic difficulty of variational optimization, as well as reveal new ways to exploit decision-tree structures to navigate complex landscapes more effectively. Indeed, one might argue that the hardness of QAOA optimization is an inherent property of the algorithm rather than a consequence of the optimization method used. However, reinforcement learning approaches have been shown to uncover regular patterns in QAOA parameters that remain elusive to traditional warm-started optimizations [30]. Given that MCTS can be viewed as a crude form of reinforcement learning, it is possible that it could outperform direct energy landscape minimization in certain cases where conventional optimizers struggle the most.

Furthermore, alternative strategies such as Bayesian optimization [85] and surrogate methods [86, 87] have recently gained traction as powerful and robust tools for the classical loops of VQAs [88–90]. A detailed comparison of their respective strengths and limitations, as well as an investigation of possible fruitful combinations of these methods, would be instrumental in developing a more comprehensive toolkit for hybrid quantum–classical algorithms, enabling the selection of the most effective optimization method for a given problem.

Another crucial aspect in the development of quantum algorithms for NISQ devices is their resilience against errors and decoherence. These, indeed, typically degrade the performance of VQAs when the circuit depth increases, even if the optimization loop is entirely done on classical machines [91]. For more realistic applications, it is therefore important to understand whether iterative SSR-MCTS can cope with hardware noise and still suggest meaningful strategies.

Addressing these issues will advance our understanding of the mutual enhancement between quantum computing and artificial intelligence, and foster the development of quantum technologies.

## Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: <https://doi.org/10.5281/zenodo.13970332>.

## Acknowledgments

We warmly thank M Burrello and G Giedke for useful discussions and advice. AA acknowledges funding by the Department of Education of the Basque Government through the Project PIBA\_2023\_1\_0021 (TENINT), and that this work has been produced with the support of a 2023 Leonardo Grant for Researchers

in Physics, BBVA Foundation. The BBVA Foundation is not responsible for the opinions, comments and contents included in the project and/or the results derived therefrom, which are the total and absolute responsibility of the authors. MW has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101080086 NeQST. This project has been supported by the Provincia Autonoma di Trento and Q@TN, the joint lab between the University of Trento, FBK—Fondazione Bruno Kessler, INFN—National Institute for Nuclear Physics, and CNR—National Research Council. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the granting authority can be held responsible for them. This work was supported by the Dutch National Growth Fund (NGF), as part of the Quantum Delta NL programme.

## Appendix A. MCTS and gradient-descent: on possible hybrid optimizer variants

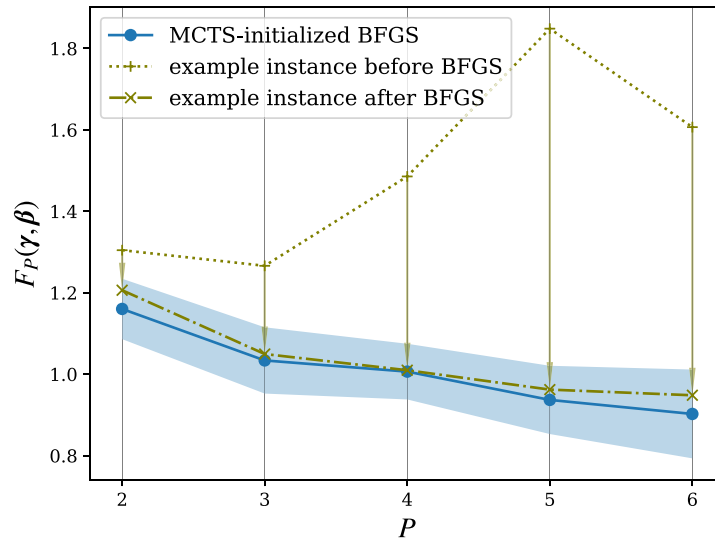
MCTS is an extremely versatile algorithm. The very loose definition of its selection and playout policies make it very convenient to modify to suit the problem at hand or develop new alternative algorithms [43, 44]. In this appendix, we showcase two of the many promising avenues opened by this approach for developing MCTS-based optimizers to find optimal QAOA strategies. The core idea is to use MCTS to survey at a coarser resolution the QAOA energy landscape while a gradient-based local minimization, for instance the BFGS algorithm [92], is tasked with finding the local minima. We look at two possible implementations.

In the first, as proposed in [39], MCTS simply suggests initial parameters for a subsequent gradient-descent search, even when the schedule chosen by MCTS gives rise to a high energy. Indeed, providing ways to efficiently initialize local (gradient-based) optimization is very valuable in the context of the complex cost landscapes of QAOA, as it is well-established that randomly initialized QAOA tends to converge to undesirable local minima [27]. We present a benchmark of this approach in figure A1, which shows the results of MCTS-initialized gradient-descent searches using the standard BFGS [92] algorithm on 10 different instances of  $n = 7$  3-SAT. For each instance, we consider the best result out of 10 runs of MCTS-initialized QAOA. The MCTS-initialization improves the results considerably from standard MCTS, showing that the parameters suggested by unmodified MCTS can, in fact, sometimes be on the slope leading to a good minimum, and not necessarily stuck in a local minimum like the high energy seems to suggest. This approach relies exactly on what makes MCTS fail in the first place, i.e. the ‘fractalization’ of the leaf nodes. While it is hard for MCTS to identify an optimal strategy, it could potentially identify something that is close enough in the parameter space to easily find it with a local minimization with continuous parameters initialized at that point. The combined resources for the MCTS initialization and the subsequent BFGS gradient descent, however, make this approach's computational cost scaling unfavorable for large parameter spaces.

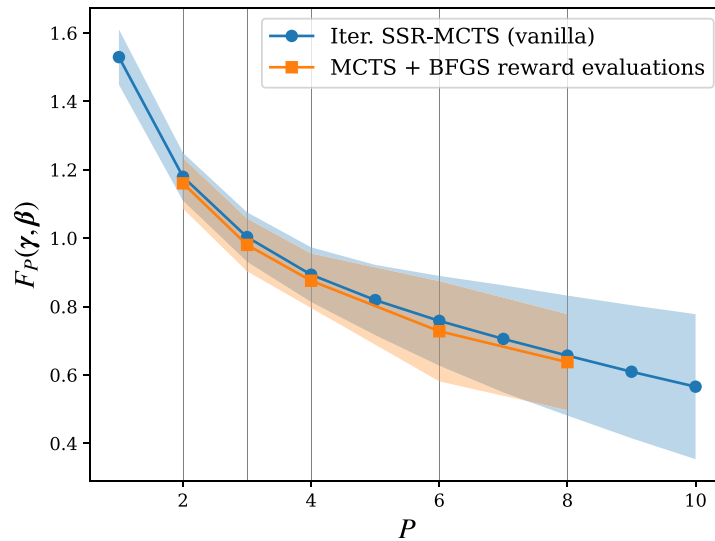
The second alternative hybrid algorithm can be devised by modifying the rollout stage of MCTS. In this version of MCTS, the reward of a leaf node corresponding to a set of QAOA angles, is instead extracted from the energy given by the parameters obtained via a BFGS search starting from the parameters of the original leaf node. Consequently, this version of MCTS sees the cost landscape as a collection of ‘flat’ surfaces, each with a cost corresponding to its particular minimum, and extending to the whole basin of attraction of said minimum. This is somewhat reminiscent of the basin-hopping algorithm for multi-variable optimization [93], with the difference that the hopping dynamics occur naturally with MCTS.

We show the results obtained with this second algorithm in figure A2, comparing it to the iterative SSR-MCTS for 10  $n = 7$  problem instances. We find MCTS does not struggle within such landscapes for the depths under consideration, at least with the problem size we are considering. The good performance obtained by MCTS in these types of basin-based landscapes can be attributed to the fact that MCTS tackles the exploration of the energy landscape in a different way than most local optimizers. Indeed, the latter start from a random point in the parameter space and then update all variational parameters (or a batch of them) by looking at the surrounding landscape. MCTS, on the other hand, sees the sequence  $(\gamma, \beta)$  as a particular strategy on a decision tree, where at each layer the next move is chosen to maximize the probable future reward and is less prone to get stuck in flat regions, as it samples other areas of the landscape and focuses on the more promising regions. This variant also especially benefits from the single-player modifications outlined in section 2.2. Concretely, sampling the lowest energy plateaus just once already implies that the optimal set of parameters will be found by the algorithm.

However, running MCTS in this manner greatly increases the number of required function evaluations for the same number of MCTS cycles, as each rollout requires a gradient-descent run. Nevertheless, the simplified landscape also seems to reduce the number of cycles required for MCTS to choose good strategies, although this does not make up for the vast increase in function evaluations given by the gradient-descent



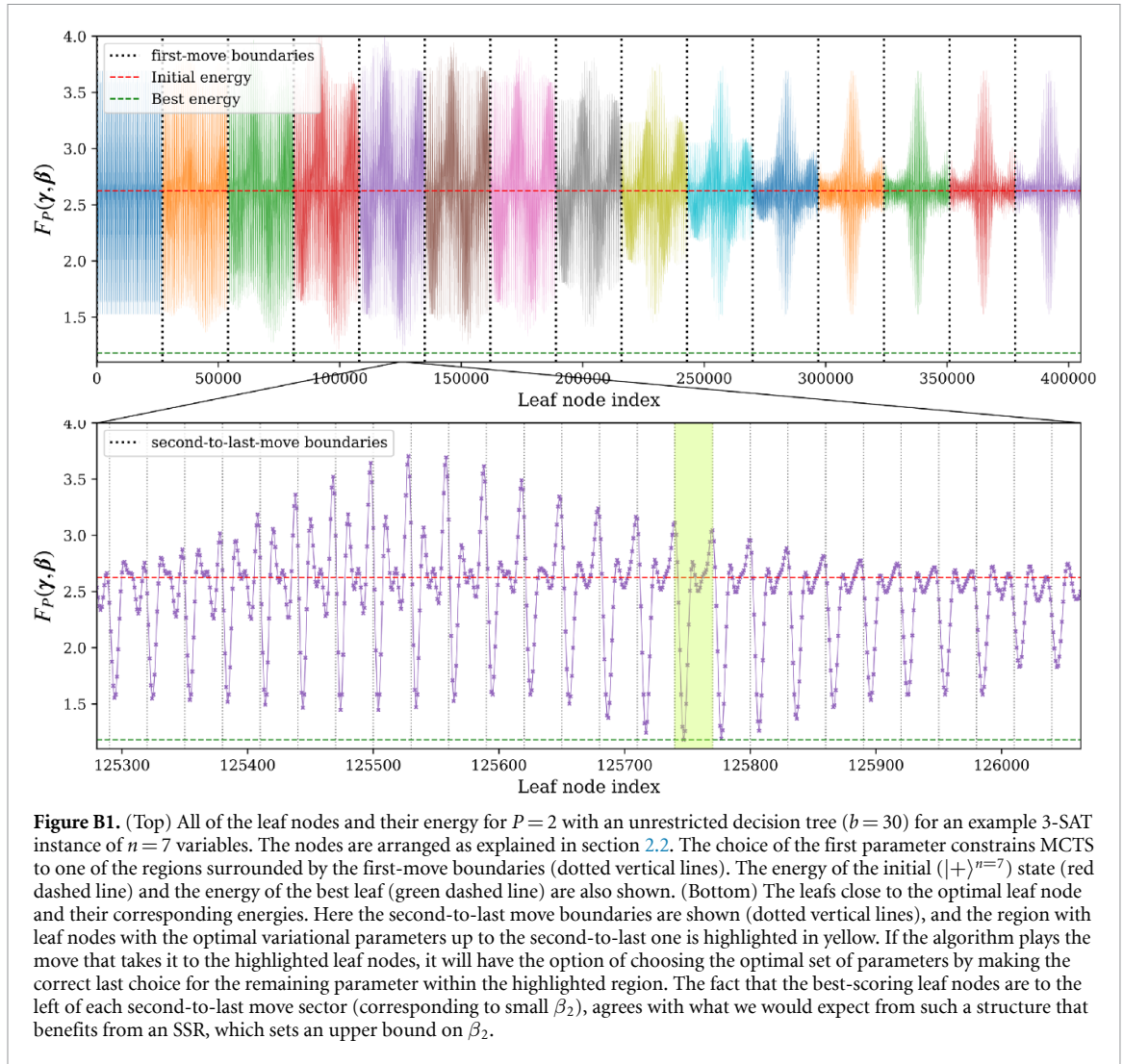
**Figure A1.** Average energies obtained with MCTS-initialized BFGS searches on 10  $n = 7$  3-SAT instances (blue). 10 games of standard (non-SSR) MCTS were run per instance, and the parameters suggested by MCTS were used as initial parameters of a BFGS search. Only the best result out of the ten games of each instance contributes to the average. For one example instance, the initial energies given by the MCTS search are shown (green '+' markers), suggesting that parameters chosen by MCTS with large residual energies can still be on the slope to a minimum with much smaller energy (green 'x' markers) in many cases.



**Figure A2.** Energies of MCTS with BFGS reward evaluations (orange) and standard MCTS with iterative SSR (blue) averaged over 10  $n = 7$  3-SAT instances. For this BFGS variant, a gradient descent search was run at each leaf node reached during the rollout stages of MCTS. In this way, the algorithm sees a landscape made up of different flat plateaus with the values of all the minima, which correspond to the regions from where the gradient points to that particular minimum. MCTS ran for just 100 cycles per turn for each instance.

runs. With 100 MCTS cycles per turn, this algorithm is able to match the performance of iterative SSR-MCTS for  $n = 7$  3-SAT, as seen in figure A2. This algorithm is, in a sense, similar in concept to the one used in [32], where MCTS is used to propose Hamiltonian sequences for creating generalized QAOA ansätze, and the goodness of a sequence is gauged by the energy of the final state obtained with the parameters outputted by a gradient-based method for the given sequence; except, in our case, the sequence is fixed (we use the standard QAOA one) and we employ this method which utilizes gradient-descent runs during the rollout stages to optimize the parameters for that fixed sequence.

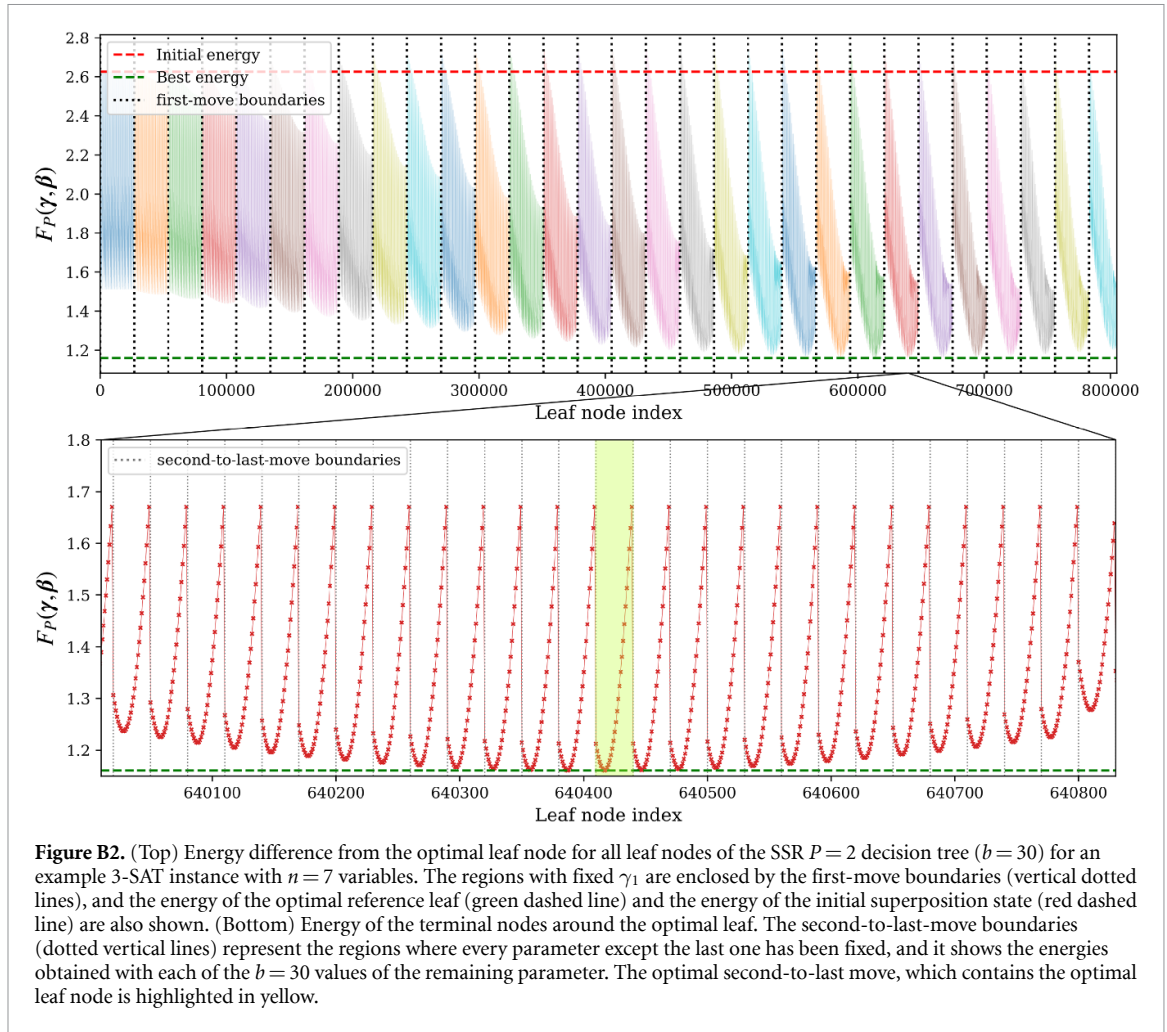
Both of these hybrid proposed methods have the advantage that the QAOA parameters they suggest are no longer discretized and that they are not bound by any problem-dependent heuristic considerations. However, the MCTS algorithm loses its main advantages, such as its gradient-free nature and remarkable resistance against noise.



## Appendix B. Why standard MCTS struggles: structure of the decision tree with and without SSR

In this appendix, we extend the analysis of the  $P = 2$  cost landscape that was discussed in section 3.2. As we found out, SSR is a requirement for standard MCTS to perform well in the task of QAOA parameter optimization. We expect this vast difference in performance between the two versions to manifest as clearly distinct decision tree structures. We now look at the MCTS trees directly for the case of the example  $n = 7$  3-SAT instance that we considered in figure 3 in order to study how the action spaces themselves differ in the unrestricted and SSR cases.

Figure B1 shows the energy difference from the optimal set of parameters for all of the  $4.05 \times 10^5$  available leaf nodes (possible combinations of the variational parameters) for the unrestricted case. In figure B2 we represent the same feature for the  $\delta_1 = 0$  SSR, with the restriction implemented according to the previous depth's optimal parameters. Notice that there are twice as many ( $8.1 \times 10^5$ ) possible outcomes in the SSR case since inversion symmetry can no longer be straightforwardly exploited. For both cases, we show the energies of all leaf nodes and represent the corresponding first-move boundaries, which enclose the regions that MCTS must choose between in the first turn, i.e. sets of leaf nodes sharing the same value of the first parameter ( $\gamma_1$  in our case). In addition, we also present a zoomed-in version of both figures, showing the leafs close to the terminal node corresponding to the best approximation of the ground state within the



variational manifold. In these zoomed-in figures, we also show the second-to-last move boundaries, the regions where all but the last parameter ( $\beta_2$  in this case) have been fixed<sup>13</sup>.

Visually, the two cases are vastly different. The unrestricted case shows a very complex and unpredictable distribution of leaf nodes, where good strategies are likely hard to find. The best nodes are hidden in regions where very badly scoring ones also exist, and the averaging of the UCT strategy makes standard MCTS blind to these features in favor of paths that lead to mediocre leaves more consistently. The improvements due to SSR can be argued qualitatively already from the structure of the energies in figure B1. There is one key observation: to restrict this  $P = 2$  search space, we make use of the optimal parameters ( $\tilde{\gamma}_1^*, \tilde{\beta}_1^*$ ) for the same instance at  $P = 1$ . As such,  $\tilde{\gamma}_1^*$  sets an upper bound on  $\gamma_1$ , which means that the entire right part of the leaves (those with  $\gamma_1 > \tilde{\gamma}_1^*$ ) in figure B1 will be discarded. Indeed, we see that these turn out to all be undesirable leaf nodes. Furthermore,  $\tilde{\beta}_1^*$  sets a lower bound for  $\beta_1$ , which in turn, means that the left part within each first-move sector is filtered out, and, again, we see that it is the parts with high  $\beta_1$  that house the best minima. This qualitative argument can be extended all the way to the choice of the last parameter.

In the SSR case of figure B2, in contrast, visual inspection suffices to see that the leaf node structure becomes much easier for MCTS to navigate: the first moves' quality is straightforward to estimate, with the behavior of the leaves in each sector being consistent and more regular, compared to figure B1. In particular, there are no hidden good minima among leaf nodes scoring worse on average. Notice also that the parameter restriction already preselects a region of the variational manifold where the energy is always smaller than the

<sup>13</sup> The different second-to-last boundaries in figures B1(bottom) and B2(bottom), correspond to cutaway sections with constant  $\gamma_2$  of the remaining 2D cost landscape when all but the last two parameters have been specified. This gives rise to an almost self-similar-looking structure, which slowly varies between different sectors.

initial one, marked by the dashed red line. As we saw, standard MCTS can learn to very efficiently navigate such spaces.

As a closing remark, it is important to note that a high difference in the index of two leaf nodes of these figures need not necessarily indicate that the parameters of the two leaf nodes are far away from each other in Euclidean distance in the space of the variational parameters. For this purpose, representing the leaf nodes as we did in figure 3 is more appropriate. Below we give more details on the definition of the tree distance employed in section 3.2.

### B.1. Definition of tree distance

We end this appendix with the formal definition of the tree distance used to generate figure 3. For a depth  $i$ , there are  $b$  possible values to choose for the corresponding variational parameter  $\theta_i$ ,

$$\theta_i \in \left\{ \theta_i^{(1)}, \dots, \theta_i^{(b)} \right\}. \quad (\text{B.1})$$

A leaf node is defined by a set of  $2P$  such choices, the indices of which we represent with  $c_i \in [1, b]$ ,

$$\boldsymbol{\theta} = \left\{ \theta_1^{(c_1)}, \theta_2^{(c_2)}, \dots, \theta_{2P}^{(c_{2P})} \right\}. \quad (\text{B.2})$$

We define the tree distance  $d$  between two leaf nodes given by configurations  $\boldsymbol{\theta}, \boldsymbol{\theta}'$  as the euclidean distance on the indices  $c_i, c'_i$  of their defining parameters,

$$d(\boldsymbol{\theta}, \boldsymbol{\theta}')^2 = \sum_i^{2P} (c_i - c'_i)^2. \quad (\text{B.3})$$

This is related to the (non-periodic<sup>14</sup>) distance of the two points in  $2P$  dimensional parameter space.

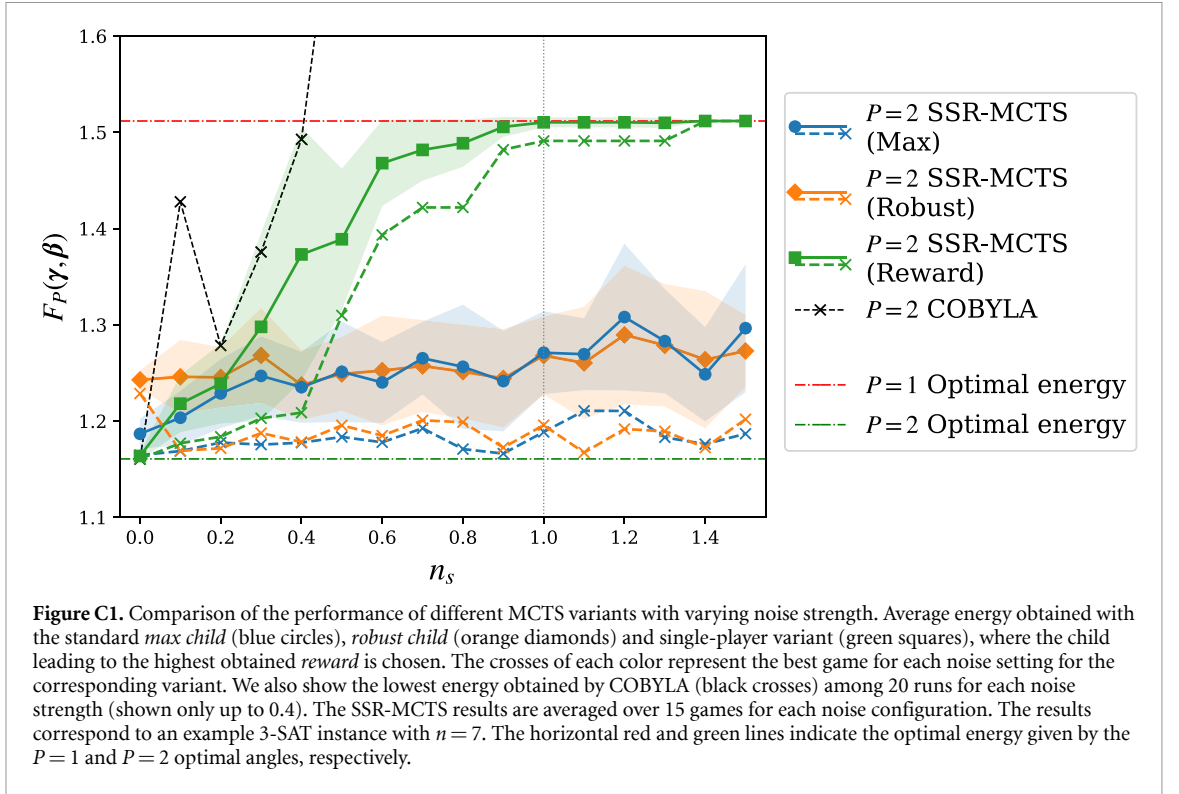
## Appendix C. MCTS variants in the midst of noise

In the main text, we utilized standard SSR-MCTS (i.e. without the single-player modifications) to optimize noisy cost functions (section 3.4). In this appendix, we briefly look at how other versions of our algorithm fare within noisy landscapes. Apart from the standard MCTS algorithm (the one which makes the final decision by choosing the child with maximum  $w_i/n_i$ , also known as *max child*), we also test the standard algorithm with another widespread final move selection criterion known as *robust child* [44, 94, 95], which chooses the child with the highest visit count  $n_i$ , as well as the single-player variant introduced in section 2.2. We also compare the different MCTS approaches with COBYLA [73], a gradient-free optimizer widely used in VQAs and as a standard baseline comparison to other optimizers [90, 96, 97], despite not being particularly optimized for noisy landscapes [90, 98]. We do not provide an explicit comparison with BFGS, a standard gradient-based benchmark algorithm for noiseless optimization, as we found it unable to converge even at the smallest noise setting that we considered.

Figure C1 shows how each of the three tested variants of MCTS and COBYLA perform in finding the optimal parameters for search-space restricted  $P = 2$  QAOA on a single instance of 3-SAT with  $n = 7$  qubits and increasing noise strength. For the MCTS variants, we report both the average performance with the associated variance (continuous lines with solid marks and shaded areas) and the best result obtained out of 15 repetitions (dashed lines with crosses). For COBYLA we only report the best result out of 20 runs (dotted black lines with crosses). In the noiseless case, the single-player variant performs best out of the MCTS-based variants as expected, as it is able to reliably store paths that lead to well-scoring sets of parameters. It also matches the quality of the COBYLA optimization, even though the latter explores a continuous energy landscape. However, their performance quickly falls off as the noise strength is increased past  $0.2\Delta$ : the memorized paths of MCTS-SP are no longer a reliable criterion for choosing moves, and COBYLA becomes extremely unstable. Indeed, The single-player variant of MCTS is designed for deterministic action spaces, a quality that is lost as soon as we add noise or errors into the reward evaluation. Interestingly, for strong noise, the single-player variant converges to the energy of the  $P = 1$  optimal parameters. We do not report data for COBYLA beyond  $n_s = 0.4$  as some points fall far outside the shown  $y$ -axis range.

The other two variants perform similarly, with a decisive improvement over the single-player variant and COBYLA at strong noise. In the  $n_s = 0$  case, *robust child* seems to consistently choose worse schedules than

<sup>14</sup> The periodic version of the tree distance can also be considered, although slightly harder to implement in the SSR case. For our analysis, both the periodic and non-periodic versions gave qualitatively similar results.



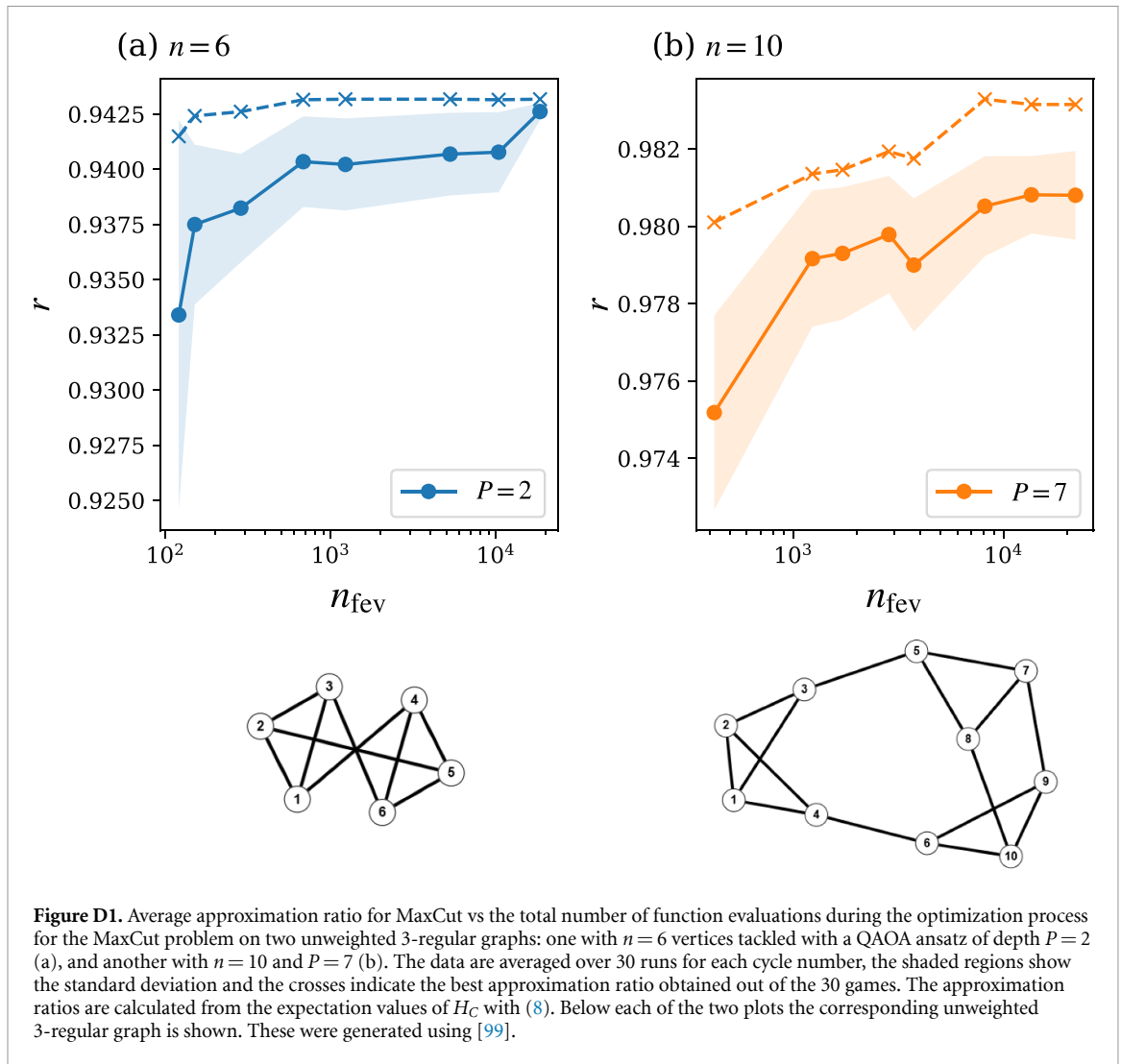
the other variants. In principle, *robust child* does not directly take the scores of the nodes into consideration for the final move selection. This is a potential advantage in noisy environments as the nodes' scores are affected by noise directly, which might suggest it might be more stable than *max child*. However, the scores also do affect how the nodes are chosen in the selection stage, which, in turn, influences the visit counts. The worse performance of robust-child in the noiseless case might also indicate that the algorithm could benefit, at least in this example case, from alternative energy-reward mappings or different  $C$  parameter values, to increase the reward difference between different leaf nodes to make the search greedier and less horizontal, and visit counts across sibling-nodes more varied.

## Appendix D. Resource cost analysis

When designing optimization strategies for VQAs, the number of required queries to the quantum hardware is an important figure of merit for assessing their performance. This also limits the circuit depths that are realistically achievable, as a larger number of variational parameters also results in a larger required number of function evaluations.

As an *anytime* algorithm [44], the number of times the cost function is evaluated  $n_{\text{fev}}$  can be very easily controlled with MCTS. The number of calls matches the number of MCTS cycles performed during the game, times the number of shots needed for the desired accuracy, and they can be stopped at any point to make the algorithm suggest a strategy. The minimum number of cycles that we can run without leaving out any options are  $2Pb$  (the depth of the tree times the branching factor), where the selected move at each turn will be the child node which gave the highest score after running one random simulation from each of the  $b$  children. Evidently, the results from such MCTS runs are bound to be very inaccurate. We also know that, as the number of cycles grows to infinity, the algorithm eventually finds the optimal leaf node [44]. In general, the training of machine-learning and AI methods is resource expensive and we expect MCTS to be no exception. In the main text, we used a heuristically chosen scaling  $n_{\text{fev}} = 1000 + 800(2P - 1)$ . Here we elucidate the performance dependence on  $n_{\text{fev}}$ , which can be used to balance between the algorithm's cost and its accuracy. Importantly, the  $n_{\text{fev}}$  cycles do not need to be evenly distributed among the  $2P$  choices that constitute the final strategy. Indeed, we find that spending more time at the beginning of the process is beneficial, as the first moves are critical for designing a successful schedule. Instead, optimizing the parameters 'deeper' in the ansatz requires less exploration of the remaining part of the tree.

Figure D1 shows the performance of iterative SSR-MCTS with varying number of total MCTS cycles and system sizes. The results in figure D1(a) are obtained with  $\delta = 0$  on a graph with  $n = 6$  vertices (shown below the main plot), while the softening is  $\delta = 0.02$  in figure D1(b), where we run the algorithm with  $P = 7$  layers



on a graph with  $n = 10$  (also shown below the plot). The branching factor is set to  $b = 30$  in both cases. The sizes of the trees in the two cases thus amount to  $8.1 \times 10^5$  and  $2.2 \times 10^{10}$  leaf nodes, respectively, out of which the algorithm samples only  $n_{\text{fev}}$  (not necessarily unique) nodes. The two panels display an improving performance as  $n_{\text{fev}}$  increases, both in the average and in the best-found approximation ratio. However, the range in which  $r$  varies is rather limited, below 2%, and it would be significantly overshadowed by the presence of noise. Hence, a few hundred cycles are enough to reach good approximations of the target state, as the search-space restriction already limits the variational ansatz in a region close to the global minimum. In particular, notice that in figure D1(b) SSR-MCTS already reaches  $r \geq 0.95$  for the minimum number of cycles considered  $n_{\text{fev}} = 420$ . Our results suggest that iterative SSR-MCTS is competitive against other state-of-the-art algorithms, such as the Bayesian Optimization in [88]. The performance can be compared on the same Max-Cut instance ( $n = 10$  and  $P = 7$ ), where SSR-MCTS obtains a better approximation ratio. A comparison of the respective resource costs, however, is less straightforward as our algorithm leverages the knowledge of optimal schedules on shallower circuits and, therefore, it depends on the details of the implementation of the iterative procedure.

A resource-savvy implementation will invest more time in the optimization of low-depth ansätze and progressively reduce the number of cycles per layer as  $P$  increases up to the desired accuracy or until the energy no longer decreases. In cases where the number of function evaluations is an important limiting factor of a particular experimental implementation of QAOA, the branching factor could also be reduced to make the algorithm converge with fewer calls at the cost of the reachability of the SSR-MCTS algorithm. These coarser results could also be used as good initial parameters for gradient-based runs (see appendix A). Additionally, integrating MCTS with neural networks, as done in [39] for annealing schedules, will likely greatly enhance our strategy. Indeed, the neural networks allow for adapting existing trained MCTS agents on new problem instances or circuit depths without having to learn the decision tree structure from scratch.

A separate discussion concerns the number of shots needed for any function evaluation. To estimate the reward with an accuracy  $\epsilon$ , typically one needs  $N_{\text{shots}} = O(1/\epsilon^2)$  single-shot measurements. Here, SSR-MCTS has two nice features that contribute towards reducing this overhead. First, the robustness against measurement noise suggests that rough estimates of the reward are sufficient for obtaining good approximations in the ground states, in particular in the early training stage of the MCTS agent. Second, The prefactor in the proportionality relation between  $N_{\text{shots}}$  and  $\epsilon^{-2}$  will likely decrease as  $P$  increases. Indeed, SSR already severely restricts the energy window spanned by the QAOA ansatz, as clearly shown by figure 3. Their energy variance is reduced accordingly. Finally, one can train the MCTS algorithm directly with single-shot measurements and update the value of the tree nodes using the energy of the sampled classical configuration. This might greatly reduce the resource requirements but further investigations are needed to gauge its efficacy and possible advantage over the approach presented here.

## ORCID iDs

Andoni Agirre  <https://orcid.org/0009-0002-8368-4438>

Evert van Nieuwenburg  <https://orcid.org/0000-0003-0323-0031>

Matteo M Wauters  <https://orcid.org/0000-0003-3261-9425>

## References

- [1] Grover L K 1996 A fast quantum mechanical algorithm for database search (arXiv:quant-ph/9605043)
- [2] Brassard G, Høyer P, Mosca M and Tapp A 2002 Quantum amplitude amplification and estimation *Quantum Computation and Information* vol 305 (*Contemporary Mathematics*) (American Mathematical Society) pp 53–74
- [3] Durr C and Hoyer P 1999 A quantum algorithm for finding the minimum (arXiv:quant-ph/9607014)
- [4] Shor P 1994 Algorithms for quantum computation: discrete logarithms and factoring *Proc. 35th Annual Symp. on Foundations of Computer Science* pp 124–34
- [5] McClean J R, Romero J, Babbush R and Aspuru-Guzik A 2016 *New J. Phys.* **18** 023023
- [6] Moll N et al 2018 *Quantum Sci. Technol.* **3** 030503
- [7] Cerezo M et al 2021 *Nat. Rev. Phys.* **3** 625–44
- [8] Bharti K et al 2022 *Rev. Mod. Phys.* **94** 015004
- [9] McClean J R, Harrigan M P, Mohseni M, Rubin N C, Jiang Z, Boixo S, Smelyanskiy V N, Babbush R and Neven H 2021 *PRX Quantum* **2** 030312
- [10] Peruzzo A, McClean J, Shadbolt P, Yung M-H, Zhou X-Q, Love P J, Aspuru-Guzik A and O’Brien J L 2014 *Nat. Commun.* **5** 4213
- [11] Mbeng G B, Fazio R and Santoro G 2019 Quantum annealing: a journey through digitalization, control, and hybrid quantum variational schemes (arXiv:1906.08948)
- [12] Cerezo M, Sharma K, Arrasmith A and Coles P J 2022 *npj Quantum Inf.* **8** 113
- [13] Khatri S, LaRose R, Poremba A, Cincio L, Sornborger A T and Coles P J 2019 *Quantum* **3** 140
- [14] Sharma K, Khatri S, Cerezo M and Coles P J 2020 *New J. Phys.* **22** 043006
- [15] Biamonte J, Wittek P, Pancotti N, Rebentrost P, Wiebe N and Lloyd S 2017 *Nature* **549** 195–202
- [16] Farhi E and Neven H 2018 Classification with quantum neural networks on near term processors (arXiv:1802.06002)
- [17] Havlíček V, Córcoles A D, Temme K, Harrow A W, Kandala A, Chow J M and Gambetta J M 2019 *Nature* **567** 209–12
- [18] Farhi E, Goldstone J and Gutmann S 2014 arXiv:1411.4028
- [19] Lloyd S 2018 Quantum approximate optimization is computationally universal (arXiv:1812.11075)
- [20] Streif M and Leib M 2019 Comparison of QAOA with quantum and simulated annealing (arXiv:1901.01903)
- [21] Zhou L, Wang S T, Choi S, Pichler H and Lukin M D 2020 *Phys. Rev. X* **10** 021067
- [22] Wauters M M, Mbeng G B and Santoro G E 2020 *Phys. Rev. A* **102** 062404
- [23] Blekos K, Brand D, Ceschini A, Chou C-H, Li R-H, Pandya K and Summer A 2024 A review on quantum approximate optimization algorithm and its variants *Phys. Rep.* **1068** 1–66
- [24] Santra G C, Jendrzewski F, Hauke P and Egger D J 2024 *Phys. Rev. A* **109** 012413
- [25] Wilson M, Stromswold S, Wudarski F, Hadfield S, Tubman N M and Rieffel E 2019 Optimizing quantum heuristics with meta-learning (arXiv:1908.03185)
- [26] Nakanishi K M, Fujii K and Todo S 2020 *Phys. Rev. Res.* **2** 043158
- [27] Sack S H and Serbyn M 2021 *Quantum* **5** 491
- [28] Lumia L, Torta P, Mbeng G B, Santoro G E, Ercolessi E, Burrello M and Wauters M M 2022 *PRX Quantum* **3** 020320
- [29] Yao J, Bukov M and Lin L 2020 Policy gradient based quantum approximate optimization algorithm *Proc. 1st Mathematical and Scientific Machine Learning Conf. (Proc. Machine Learning Research* vol 107) ed J Lu and R Ward (PMLR) pp 605–34 (available at: <https://proceedings.mlr.press/v107/yao20a.html>)
- [30] Wauters M M, Panizon E, Mbeng G B and Santoro G E 2020 *Phys. Rev. Res.* **2** 033446
- [31] Hadfield S, Wang Z, O’Gorman B, Rieffel E G, Venturelli D and Biswas R 2019 *Algorithms* **12** 34
- [32] Yao J, Lin L and Bukov M 2021 *Phys. Rev. X* **11** 031070
- [33] Zhu L, Tang H L, Barron G S, Calderon-Vargas F A, Mayhall N J, Barnes E and Economou S E 2022 *Phys. Rev. Res.* **4** 033029
- [34] Chandaran P, Hegade N N, Paul K, Albarrán-Arriagada F, Solano E, del Campo A and Chen X 2022 *Phys. Rev. Res.* **4** 013141
- [35] Bittel L and Kliesch M 2021 *Phys. Rev. Lett.* **127** 120502
- [36] McClean J R, Boixo S, Smelyanskiy V N, Babbush R and Neven H 2018 *Nat. Commun.* **9** 4812
- [37] Arrasmith A, Cerezo M, Czarnik P, Cincio L and Coles P J 2021 *Quantum* **5** 558
- [38] Sutton R S and Barto A G 2018 *Reinforcement Learning, An Introduction* 2nd edn (The MIT Press)
- [39] Chen Y-Q, Chen Y, Lee C-K, Zhang S and Hsieh C-Y 2022 Optimizing quantum annealing schedules with Monte Carlo Tree Search enhanced with neural networks *Nat. Mach. Intell.* **4** 269–78

- [40] Yao J, Li H, Bukov M, Lin L and Ying L 2022 Monte Carlo Tree Search based hybrid optimization of variational quantum circuits *Proc. Mathematical and Scientific Machine Learning (Proc. Machine Learning Research vol 190)* (PMLR) pp 49–64 (available at: <https://proceedings.mlr.press/v190/yao22a.html>)
- [41] Kocsis L and Szepesvári C 2006 Bandit based Monte-Carlo planning *Machine Learning: ECML 2006* (Springer) pp 282–93
- [42] Silver D et al 2017 Mastering chess and shogi by self-play with a general reinforcement learning algorithm (arXiv:1712.01815)
- [43] Świechowski M, Godlewski K, Sawicki B and Mańdziuk J 2023 *Artif. Intell. Rev.* **56** 2497–562
- [44] Browne C B, Powley E, Whitehouse D, Lucas S M, Cowling P I, Rohlfshagen P, Tavener S, Perez D, Samothrakis S and Colton S 2012 *IEEE Trans. Comput. Intell. AI Games* **4** 1–43
- [45] Wauters M M and van Nieuwenburg E 2022 *Nat. Mach. Intell.* **4** 810–3
- [46] Dalgaard M, Motzoi F, Sørensen J J and Sherson J 2020 *npj Quantum Inf.* **6** 6
- [47] Karp R M 1972 *Reducibility Among Combinatorial Problems* (Springer) pp 85–103
- [48] Brandao F G S L, Broughton M, Farhi E, Gutmann S and Neven H 2018 For fixed control parameters the quantum approximate optimization algorithm’s objective function value concentrates for typical instances (arXiv:1812.04170)
- [49] Akshay V, Rabinovich D, Campos E and Biamonte J 2021 *Phys. Rev. A* **104** L010401
- [50] Streif M and Leib M 2020 *Quantum Sci. Technol.* **5** 034008
- [51] Coquelin P A and Munos R 2007 Bandit algorithms for Tree Search (arXiv:cs/0703062)
- [52] Schadd M P D, Winands M H M, van den Herik H J, Chaslot G M J B and Uiterwijk J W H M 2008 Single-player Monte-Carlo Tree Search *Computers and Games* (Springer) pp 1–12
- [53] Winands M H M and Björnsson Y 2011  $\alpha\beta$ -based play-outs in Monte-Carlo Tree Search 2011 *IEEE Conf. on Computational Intelligence and Games (CIG’11)* pp 110–7
- [54] Cazenave T and Jouandea N 2007 On the Parallelization of UCT *Computer Games Workshop* (available at: <https://hal.science/hal-02310186>)
- [55] Stechly M, Gao L, Yogendran B, Fontana E and Rudolph M 2024 Connecting the Hamiltonian structure to the QAOA energy and Fourier landscape structure (arXiv:2305.13594)
- [56] Auer P, Cesa-Bianchi N and Fischer P 2002 *Mach. Learn.* **47** 235–56
- [57] Björnsson Y and Finnsson H 2009 *IEEE Trans. Comput. Intell. AI Games* **1** 4–15
- [58] Garey M R and Johnson D S 1979 *Computers and Intractability: A Guide to the Theory of NP-Completeness* 1st edn (W. H. Freeman & Co.)
- [59] Battaglia D A, Santoro G E and Tosatti E 2005 *Phys. Rev. E* **71** 066707
- [60] Mézard M, Parisi G and Zecchina R 2002 *Science* **297** 812–5
- [61] Zhang W 2001 Phase transitions and backbones of 3-SAT and maximum 3-SAT *Principles and Practice of Constraint Programming—CP 2001* (Springer) pp 153–67
- [62] Bastidas V M, Haug T, Gravel C, Kwek L C, Munro W J and Nemoto K 2020 arXiv:2009.00823 [quant-ph]
- [63] Lucas A 2014 *Front. Phys.* **2** 5
- [64] Znidarič M 2005 *Phys. Rev. A* **71** 062305
- [65] Akshay V, Philathong H, Morales M E S and Biamonte J D 2020 *Phys. Rev. Lett.* **124** 090504
- [66] Johansson J, Nation P and Nori F 2012 *Comput. Phys. Commun.* **183** 1760–72
- [67] Johansson J, Nation P and Nori F 2013 *Comput. Phys. Commun.* **184** 1234–40
- [68] Agirre A 2022 quantumMCTS (Code repository) (available at: <https://github.com/agiandoni/quantumMCTS>)
- [69] Agirre A, van Nieuwenburg E and Wauters M M 2024 Dataset for “A Monte Carlo Tree Search approach to QAOA: finding a needle in the haystack” *Zenodo* <https://doi.org/10.5281/zenodo.13970332>
- [70] Bjarnason R, Fern A and Tadeipalli P 2009 *Proc. Int. Conf. on Automated Planning and Scheduling* vol 19 pp 26–33
- [71] Kiarostami M S, Daneshvaramoli M, Khalaj Monfared S, Visuri A, Karisani H, Hosio S, Khashehchi H, Futuhi E, Rahmati D and Gorgin S 2021 On using Monte-Carlo Tree Search to solve puzzles *Proc. 2021 7th Int. Conf. on Computer Technology Applications (ICCTA’21)* (Association for Computing Machinery) pp 18–26
- [72] Godlewski K and Sawicki B 2020 MCTS based agents for multistage single-player card game 2020 *IEEE 21st Int. Conf. on Computational Problems of Electrical Engineering (CPEE)* (IEEE)
- [73] Powell M J D 1994 *A Direct Search Optimization Method That Models the Objective and Constraint Functions by Linear Interpolation* (Springer) pp 51–67
- [74] Goemans M X and Williamson D P 1995 *J. ACM* **42** 1115–45
- [75] Halperin E, Livnat D and Zwick U 2004 *J. Algorithms* **53** 169–85
- [76] Wurtz J and Lykov D 2021 *Phys. Rev. A* **104** 052419
- [77] Herrman R, Treffert L, Ostrowski J, Lotshaw P C, Humble T S and Siopsis G 2021 *Quantum Inf. Process.* **20** 289
- [78] Wang Z, Hadfield S, Jiang Z and Rieffel E G 2018 *Phys. Rev. A* **97** 022304
- [79] Chaslot G, De Jong S, Saito J T and Uiterwijk J 2006 Monte-Carlo Tree Search in production management problems *Proc. 18th BeNeLux Conf. on Artificial Intelligence* vol 9198
- [80] Cazenave T 2009 Nested Monte-Carlo search *Proc. 21st Int. Joint Conf. on Artificial Intelligence* pp 456–61
- [81] Schadd M P D 2011 Selective search in games of different complexity *PhD Thesis* Maastricht University (available at: <https://api.semanticscholar.org/CorpusID:153142732>)
- [82] Egger D J, Mareček J and Woerner S 2021 *Quantum* **5** 479
- [83] Tate R, Moondra J, Gard B, Mohler G and Gupta S 2023 *Quantum* **7** 1121
- [84] Truger F, Barzen J, Bechtold M, Beisel M, Leymann F, Mandl A and Yussupov V 2024 *ACM Comput. Surv.* **56** 1–31
- [85] Frazier P I 2018 A tutorial on Bayesian optimization (arXiv:1807.02811)
- [86] Forrester A, Sobester A and Keane A 2008 *Engineering Design via Surrogate Modelling: A Practical Guide* (Wiley)
- [87] Kübler J M, Arrasmith A, Cincio L and Coles P J 2020 *Quantum* **4** 263
- [88] Tibaldi S, Vodola D, Tignone E and Ercolessi E 2023 *IEEE Trans. Quantum Eng.* **4** 1–11
- [89] Shaffer R, Kocia L and Sarovar M 2023 *Phys. Rev. A* **107** 032415
- [90] Cheng L, Chen Y-Q, Zhang S-X and Zhang S 2024 *Commun. Phys.* **7** 83
- [91] Pelofske E, Bärtschi A, Golden J and Eidenbenz S 2023 High-round QAOA for MAX  $k$ -SAT on trapped ion NISQ devices 2023 *IEEE Int. Conf. on Quantum Computing and Engineering (QCE)* vol 01 pp 506–17
- [92] Fletcher R 1987 *Practical Methods of Optimization* (Wiley)
- [93] Wales D J and Doye J P K 1997 *J. Phys. Chem. A* **101** 5111–6

- [94] Chaslot G, Bakkes S, Szita I and Spronck P 2021 *Proc. AAAI Conf. on Artificial Intelligence and Interactive Digital Entertainment* vol 4 pp 216–7 (available at: <https://ojs.aaai.org/index.php/AIIDE/article/view/18700>)
- [95] Chaslot G, Winands M, Herik H, Uiterwijk J and Bouzy B 2008 *New Math. Nat. Comput.* **04** 343–57
- [96] Lavrijsen W, Tudor A, Müller J, Iancu C and de Jong W 2020 Classical optimizers for noisy intermediate-scale quantum devices *2020 IEEE Int. Conf. on Quantum Computing and Engineering (QCE)* pp 267–77
- [97] Egger D J, Capecchi C, Pokharel B, Barkoutsos P K, Fischer L E, Guidoni L and Tavernelli I 2023 *Phys. Rev. Res.* **5** 033159
- [98] Pellow-Jarman A, McFarthing S, Sinayskiy I, Park D K, Pillay A and Petruccione F 2024 *Sci. Rep.* **14** 16011
- [99] Graphonline 2015 Graph visualization tool (available at: <https://github.com/UnickSoft/graphonline>) (Accessed 23 September 2024)