

Scaling advantage with quantum-enhanced memetic tabu search for LABS

Alejandro Gomez Cadavid^{1,2}, Pranav Chandarana^{1,2}, Sebastián V. Romero^{1,2},
Jan Trautmann,¹ Enrique Solano^{1,*}, Taylor Lee Patti,³ and Narendra N. Hegade^{1,†}

¹Kipu Quantum GmbH, Greifswalderstrasse 212, 10405 Berlin, Germany

²Department of Physical Chemistry, University of the Basque Country EHU, Apartado 644, 48080 Bilbao, Spain

³NVIDIA, Santa Clara, California 95051, USA

(Dated: November 7, 2025)

We introduce quantum-enhanced memetic tabu search (QE-MTS), a non-variational hybrid algorithm that achieves state-of-the-art scaling for the low-autocorrelation binary sequence (LABS) problem. By seeding the classical MTS with high-quality initial states from digitized counterdiabatic quantum optimization (DCQO), our method suppresses the empirical time-to-solution scaling to $O(1.24^N)$ for sequence length $N \in [27, 37]$. This scaling surpasses the best-known classical heuristic $O(1.34^N)$ and improves upon the $O(1.46^N)$ of the quantum approximate optimization algorithm, achieving superior performance with a $6\times$ reduction in circuit depth. A two-stage bootstrap analysis confirms the scaling advantage and projects a crossover point at $N \gtrsim 47$, beyond which QE-MTS outperforms its classical counterpart. These results provide evidence that quantum enhancement can directly improve the scaling of classical optimization algorithms for the paradigmatic LABS problem.

I. INTRODUCTION

Finding low-autocorrelation binary sequences (LABS) is a long-standing and computationally demanding problem at the interface of combinatorial optimization and signal processing [1, 2]. For a given length, the goal is to find a binary sequence that minimizes the sum of squared off-peak autocorrelations. This objective directly translates into practical performance gains in technologies such as radar, where low sidelobes are crucial for distinguishing targets from background noise. A defining property that makes LABS a stringent benchmark is that there exists exactly one non-trivial instance per sequence length. This removes instance-selection bias and requires algorithms to confront the same increasingly rugged energy landscape.

The problem's difficulty is reflected in the fact that optimal sequences have been verified only up to $N = 66$ [3], underscoring its rapidly growing computational complexity. LABS can be formulated as the ground-state search of a long-range, four-local spin-glass Hamiltonian, representing a higher-order unconstrained binary optimization (HUBO) problem. This formulation has established LABS as a canonical testbed for benchmarking optimization algorithms across classical and quantum domains. A broad range of methods have been explored for LABS, including mixed-integer solvers such as CPLEX [4] and Gurobi [5], heuristic meta-optimizers such as the memetic tabu search (MTS) [6], and quantum algorithms such as the quantum approximate optimization algorithm (QAOA) [7] and recent approaches based on Pauli correlation encoding [8].

These approaches scale exponentially in computational effort, typically measured by the number of objective function evaluations. Among classical heuristics, the memetic tabu search (MTS) currently provides the best known performance, exhibiting a scaling of $O(1.34^N)$ and reliably reaching optimal

solutions up to $N \leq 64$ [9]. Despite this success, the exponential growth in cost ultimately limits the tractable sequence length, motivating the search for alternative paradigms. Quantum optimization methods have been investigated as potential routes to overcome this scaling barrier. The QAOA with twelve layers achieves a scaling of $O(1.46^N)$ for $28 \leq N \leq 40$, which can be reduced to $O(1.21^N)$ when combined with quantum minimum finding [7]. The bias-field digitized counterdiabatic quantum optimization (BF-DCQO) algorithm attains comparable scaling to twelve-layer QAOA while requiring up to six times fewer entangling gates for instances up to $N = 30$ [10–14]. However, the practical reach of these quantum algorithms remains constrained by hardware limitations. Experimental realizations have so far demonstrated QAOA with a single layer up to $N = 18$ on trapped-ion hardware [7] and BF-DCQO up to $N = 20$ on superconducting devices [10]. These restrictions highlight the need for hybrid quantum–classical schemes that can leverage the strengths of both domains to improve scaling within currently available computational resources.

In this work, we introduce a hybrid non-variational quantum-classical approach, the quantum-enhanced memetic tabu search (QE-MTS), to address the LABS problem. QE-MTS integrates DCQO [15] with MTS. The quantum stage generates low-energy candidate sequences that seed the initial MTS population, providing statistically biased starting points that guide the subsequent local search. The framework represents a concrete instance of hybrid sequential quantum computing [16], and quantum-enhanced optimization [17]. This division of roles aligns with present hardware capabilities: quantum processors efficiently sample structured bitstrings, while classical algorithms perform the combinatorial refinement through recombination, mutation, and memory-guided search. To rigorously assess algorithm performance, we conduct a comprehensive scaling analysis by benchmarking the time-to-solution as a function of system size. Our results demonstrate that QE-MTS exhibits significantly more favorable scaling behavior than standalone MTS for $N \leq 37$, with the performance gap widening as system size increases.

* enr.solano@gmail.com

† narendra.hegade@kipu-quantum.com

The remainder of this paper is organized as follows. Section II defines the LABS objective. Section III details the hybrid solver. Section IV presents empirical results for $27 \leq N \leq 37$ together with the scaling analysis. Section V concludes and discusses future extensions. Additional implementation details are provided in the appendices.

II. LOW-AUTOCORRELATION BINARY SEQUENCES

The LABS problem aims to find binary sequences $\mathbf{s} = (s_1, \dots, s_N) \in \{\pm 1\}^N$ that minimize the objective

$$E(\mathbf{s}) = \sum_{k=1}^{N-1} C_k^2, \quad \text{with} \quad C_k = \sum_{i=1}^{N-k} s_i s_{i+k}, \quad (1)$$

where C_k denotes the k -th autocorrelation coefficient. The cost function penalizes large off-peak autocorrelations and favors sequences with sharp correlation profiles, which are valuable in applications such as radar and communication signal design. Expanding Eq. (1) and mapping $s_i \mapsto \sigma_i^z$ yields a long-range Ising Hamiltonian representation that reads as

$$H_f = 2 \sum_{i=1}^{N-2} \sigma_i^z \sum_{k=1}^{\lfloor (N-i)/2 \rfloor} \sigma_{i+k}^z + 4 \sum_{i=1}^{N-3} \sigma_i^z \sum_{t=1}^{\lfloor (N-i-1)/2 \rfloor} \sum_{k=t+1}^{N-i-t} \sigma_{i+t}^z \sigma_{i+k}^z \sigma_{i+k+t}^z, \quad (2)$$

which includes both two- and four-body interactions, characteristic of higher-order spin-glass models [18]. The number of two- and four-body interaction terms increases quadratically and cubically with the system size, respectively

$$n_{\text{two}}(N) = \begin{cases} \frac{N}{2} \left(\frac{N}{2} - 1 \right), & \text{if } N \text{ even,} \\ \left(\frac{N-1}{2} \right)^2, & \text{if } N \text{ odd,} \end{cases} \quad (3)$$

$$n_{\text{four}}(N) = \begin{cases} \frac{N}{12} \left(\frac{N}{2} - 1 \right) (2N - 5), & \text{if } N \text{ even,} \\ \frac{1}{24} (N - 3)(N - 1)(2N - 1), & \text{if } N \text{ odd.} \end{cases} \quad (4)$$

These expressions provide an estimate of the computational resources required for a given instance. A quantum algorithm requiring G_{2q} and G_{4q} two-qubit gates to represent the two- and four-body interactions present in Eq. (2) would involve $O(G_{2q}N^2 + G_{4q}N^3)$ entangling operations. Given present coherence times and connectivity constraints, current quantum processors can reliably simulate systems of only up to approximately $N \lesssim 20$ [10].

The LABS problem exhibits both global bitflip and bit reversal symmetries

$$(s_1, \dots, s_N) \equiv (-s_1, \dots, -s_N) \equiv (s_N, \dots, s_1) \equiv (-s_N, \dots, -s_1),$$

which reduce the effective search space from 2^N to 2^{N-2} distinct configurations. The flip symmetry can be incorporated

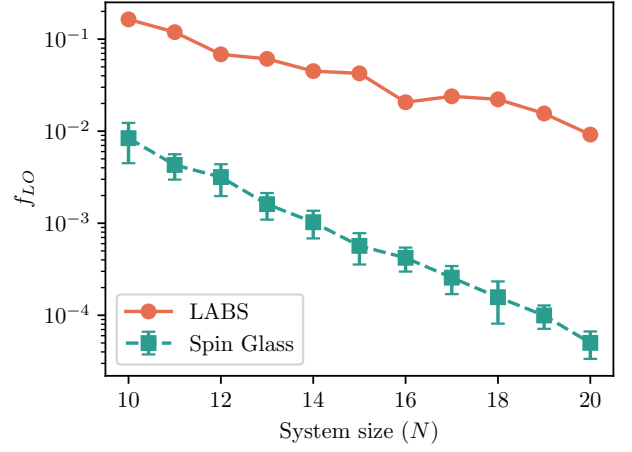


FIG. 1. Density of 1-flip local minima f_{LO} for different N . For the spin-glass case, we report the results over 10 independent instances. This quantifies the number of sequences in which no single flip reduces the energy, compared to the total number of configurations. Additionally, it supports why LABS is a more challenging problem than a common spin-glass benchmark.

at the Hamiltonian level by fixing one spin, whereas the reverse symmetry is more difficult to encode. Although these symmetries can be exploited by both classical and quantum optimizers, they offer at most a constant-factor reduction in computational effort. Since this work focuses on scaling exponents, we do not apply symmetry-based instance reduction.

A distinctive feature of LABS is its high spectral degeneracy. The number of distinct energy levels grows at most polynomially as $O(N^3)$, while the configuration space increases exponentially as $O(2^{N-2})$. This degeneracy arises from the squared autocorrelation terms in Eq. (2), which restrict energy values to multiples of four, with an upper bound of $N(N-1)(2N-1)/6$, see Appendix A. The uniform anti-ferromagnetic couplings in the Hamiltonian induce frustration, further enhancing degeneracy.

These features produce a rugged energy landscape densely populated with local minima. We quantify this ruggedness through the density of single-flip local minima

$$f_{LO}(N) = \frac{|\{s \in \{\pm 1\}^N : E(s) \leq E(s^{(i)}) \forall i\}|}{2^N}, \quad (5)$$

where $s^{(i)}$ is the configuration via flipping the i -th spin of s . This corresponds to the standard notion of one-spin-flip-stable (or metastable) states in mean-field spin glasses [19, 20], which provides a baseline for assessing ruggedness via the density of single-flip local minima of a given Hamiltonian. As shown in Fig. 1, LABS exhibits a higher density of local minima than randomly generated Sherrington–Kirkpatrick (SK) spin-glass instances [21] with Gaussian couplings $J_{ij} \sim \mathcal{N}(0, 1/N)$ and zero fields. This abundance of local traps creates problems for local search algorithms, where single-spin-flip dynamics converge to such metastable energies, and motivates our hybrid quantum-classical approach, combining non-local exploration via quantum sampling with classical refinement through metaheuristic search.

III. QUANTUM-ENHANCED ALGORITHM

A. Digitized counterdiabatic quantum optimization

Adiabatic quantum optimization aims to prepare the ground state of the problem Hamiltonian H_f . An initial state is evolved within the time window $t \in [0, T]$ under the path $H_{\text{ad}}(\lambda) = (1 - \lambda)H_i + \lambda H_f$. Here, $\lambda(t)$ toggles adiabatically the system, starting from an initial Hamiltonian H_i — whose corresponding ground state can be easily prepared— towards the problem Hamiltonian H_f , whose ground state encodes the solution of the optimization problem [23]. Relying on the adiabatic theorem, the ground state of H_f is theoretically reached in the adiabatic limit $\dot{\lambda}(t) \rightarrow 0$. In the following, we use $H_i = \sum_i h_i^x \sigma_i^x$ as initial Hamiltonian, with h_j^x the transverse fields acting on the j -th spin. In particular, we set $h_j^x = -1$ so that the ground state of H_i is $|\psi(0)\rangle = |+\rangle^{\otimes N}$ with $\sigma^x |\pm\rangle = \pm |\pm\rangle$.

To overcome the intrinsically slow adiabatic evolution, it is possible to introduce an auxiliary counterdiabatic (CD) driving contribution to accelerate the process while suppressing diabatic transitions [24, 25]. Several implementations have been proposed, with one approximating the adiabatic gauge potential by a nested-commutator series expansion [26], such that the system evolves under $H(\lambda) = H_{\text{ad}}(\lambda) + \lambda A_\lambda^{(l)}$ with

$$A_\lambda^{(l)} = i \sum_{k=1}^l \alpha_k(\lambda) O_{2k-1}(\lambda), \quad (6)$$

where $O_0(\lambda) = \partial_\lambda H_{\text{ad}}$ and $O_k(\lambda) = [H_{\text{ad}}, O_{k-1}(\lambda)]$. In the limit $l \rightarrow \infty$, this expansion converges to the exact gauge potential A_λ , which suppresses every possible transition. The coefficients α_k are obtained by minimizing the action $S_l = \text{Tr}[G_l^2]$ with $G_l = \partial_\lambda H_{\text{ad}} - i[H_{\text{ad}}, A_\lambda^{(l)}]$ (see Appendix B). In this work, we consider the first-order nested-commutator term ($l = 1$), which takes the form

$$\begin{aligned} -iO_1 = & 4 \sum_{i=1}^{N-2} \sum_{k=1}^{\lfloor \frac{N-i}{2} \rfloor} \left[h_i^x \sigma_i^y \sigma_{i+k}^z + h_{i+k}^x \sigma_{i+k}^y \sigma_i^z \right] \\ & + 8 \sum_{i=1}^{N-3} \sum_{t=1}^{\lfloor \frac{N-i-1}{2} \rfloor} \sum_{k=t+1}^{N-i-t} \sum_{p \in \{i, i+t, i+k, i+k+t\}} h_p^x \sigma_p^y \prod_{\substack{q \in \{i, i+t, i+k, i+k+t\} \\ q \neq p}} \sigma_q^z. \end{aligned} \quad (7)$$

For a fast evolution, the adiabatic contribution $H_{\text{ad}}(\lambda)$ becomes negligible compared to $H_{\text{cd}}(\lambda)$ and therefore, can be omitted. This is known as the impulse regime and has been shown to be useful while designing quantum algorithms for optimization problems [15, 27, 28], as it reduces the number of resources needed without compromising performance. Although the implementation of CD terms remains a challenge on current analog quantum platforms, their digitization has been leveraged in digital quantum computers [29]. The resulting time-evolution operator can be represented as a quantum circuit via digitization and the first-order Trotter-Suzuki decomposition, i.e., $U(T, 0) =$

$\prod_{k=1}^{n_{\text{trot}}} \exp[\Delta t \alpha_1(k\Delta t) \dot{\lambda}(k\Delta t) O_1(k\Delta t)]$, where n_{trot} is the number of Trotter steps and $\Delta t = T/n_{\text{trot}}$. In this work we take the limit $n_{\text{trot}} \rightarrow \infty$ to get the continuous time-evolution operator under $H(\lambda)$ in the impulse regime. After each circuit execution, all qubits are measured n_{shots} times, returning binary bitstring $b_0 \dots b_{N-1}$ with $b_i \in \{0, 1\}$. Each measurement translates into an energy sample of H_f . We build the DCQO circuits using an efficient decomposition into two-qubit gates, see Appendix B; and leverage CUDA-Q [30] kernels to execute them using GPUs. Particularly, the simulation is done in the range $27 \leq N \leq 37$ for $n_{\text{shots}} = 10^5$ using an *Amazon EC2 P6-b200.48xlarge* instance, see Appendix E.

B. Quantum-enhanced memetic tabu search

Memetic tabu search (MTS) is a population-based metaheuristic that combines global exploration with local intensification (via tabu search) [31]. The population maintains candidate sequences. Recombination and mutation explore new regions and then, tabu search performs a focused improvement of each offspring while preventing cycling through a short-term memory. In our quantum-enhanced MTS (QE-MTS) algorithm, DCQO supplies bitstrings that are used as the initial population, while the MTS baseline starts from a random one. Specifically, DCQO is run for a finite number of shots and the lowest bitstring is replicated K times to form the initial population, see Algorithm 1. For the MTS baseline, all K individuals are drawn randomly, which is the standard approach [6]. In both MTS and QE-MTS, we use $K = 100$, $p_{\text{comb}} = 0.9$, $p_{\text{mut}} = 1/N$ and tournament size 2. We run the MTS with a single thread to avoid over-estimation of the number of objective evaluations, and we stop the program as soon as the certified optimum is found. The time-to-solution (TTS) is measured the number of objective evaluations until the known optimal is for the first time. While this definition of TTS does not have units of time, it is possible to recover the exact runtimes from $\text{TTS} \times \tau$, where τ is the time to make one function evaluation.

IV. RESULTS

In this section, we run QE-MTS and perform a statistical analysis in order to predict the typical TTS of our approach. For each N and each algorithmic variant (MTS and QE-MTS) we run multiple independent replicates. A replicate is a distinct initial parent pool: for QE-MTS it consists of a single quantum-enhanced initial population, whereas for MTS it is an independently sampled random pool. Within each replicate we run 100 seeds and record the TTS.

A. Scaling analysis

Let $Y_{N,m,r,s}$ denote the TTS for sequence length N , method m , replicate r , and seed s . In our study, $27 \leq N \leq 37$,

Algorithm 1: Quantum-enhanced memetic tabu search

Input: Sequence length N ; population size K or initial population from a quantum sub-routine P_Q ; recombination probability p_{comb} ; maximum number of generations G_{max} ; mutation rate p_{mut} ; optional target energy E_{target}

Output: Best sequence s^* found

```

1 Initialize;
2 Use  $P_Q$  as the initial population. Set  $s^*$  to the best energy from the  $P_Q$  individuals.;
3 while  $E(s^*) > E_{\text{target}}$  and  $G \leq G_{\text{max}}$  do
4   if  $\text{rand}(0, 1) < p_{\text{comb}}$  then
5     Select two parents  $p_1, p_2$  (tournament);
6      $c \leftarrow \text{COMBINE}(p_1, p_2)$ ;           // See Appendix F
7   else
8      $c \leftarrow$  a random individual from the population;
9    $\text{MUTATE}(c, p_{\text{mut}})$ ;                     // See Appendix F
10   $c \leftarrow \text{TABUSEARCH}(c)$ ;              // See Appendix F
11  if  $E(c) < E(s^*)$  then
12     $s^* \leftarrow c$ 
13  Replace a uniformly random individual in the population by  $c$ ;
14   $G \leftarrow G + 1$ ;
15 return  $s^*$ 

```

$m \in \{\text{MTS}, \text{QE-MTS}\}$, $r \leq 100$ and $s \leq 100$. Define the per-replicate summary as $\tilde{Y}_{N,m,r} = \text{median}_s\{Y_{N,m,r,s}\}$, which yields a distribution of medians $\{\tilde{Y}_{N,m}\}_r$ for each pair (N, m) . We pick the medians to represent the TTS at a certain N since they are stable against outliers and the question we aim to answer is: how long is one typically expected to wait until the optimal solution is found. Moreover, we characterize this distribution of medians via the quantiles $Q_p(N, m) = \text{quantile}_r(\{\tilde{Y}_{N,m}\}_r; p)$, with $p \in \{0.10, 0.50, 0.90\}$. In particular, $Q_{0.50}(N, m)$ is the median of medians; $Q_{0.10}$ reflects a typical best-case (low TTS); and $Q_{0.90}$ includes typical worst-cases (high TTS). To quantify the scaling of Q_p with respect to N , we fit the log-linear model

$$\log Q_p(N, m) = \alpha_{m,p} + \beta_{m,p} N, \quad (8)$$

which assumes $Q_p(N, m) \sim (\kappa_{m,p})^N$, where $\kappa_{m,p} = \exp(\beta_{m,p})$ is the exponential scaling base.

To take into account the uncertainty coming from both replicates and seeds, we use a two-stage bootstrap. For each draw $b = 1, \dots, B$ with $B = 5000$: (i) resample replicates with replacement within each (N, m) ; (ii) for each selected replicate, resample seeds with replacement and recompute $\tilde{Y}_{N,m,r}^{(b)}$; (iii) get $Q_p^{(b)}(N, m)$ across resampled replicates; (iv) fit Eq. (8) to obtain $\beta_{m,p}^{(b)}$ and the coefficients of determination $R_{m,p}^{2,(b)}$. This bootstrapping mirrors how new data would be generated, including both between- and within-replicate variability, i.e. the variability coming from having different r and different s , respectively. We report the results with 95% confidence intervals for κ and R^2 in Table I, where the median scaling base for QE-MTS is $\kappa \approx 1.24$ (95% CI [1.23, 1.25]) versus $\kappa \approx 1.37$

TABLE I. Two-stage bootstrap 95% CIs for $\kappa = \exp(\beta)$ and R^2 ($B = 5000$ draws).

Method	Summary	κ CI	R^2 CI
QE-MTS	$Q_{0.50}$	[1.23, 1.25]	[0.86, 0.89]
QE-MTS	$Q_{0.10}$	[1.25, 1.27]	[0.83, 0.91]
QE-MTS	$Q_{0.90}$	[1.24, 1.25]	[0.90, 0.92]
MTS	$Q_{0.50}$	[1.36, 1.37]	[0.85, 0.87]
MTS	$Q_{0.10}$	[1.34, 1.36]	[0.85, 0.88]
MTS	$Q_{0.90}$	[1.37, 1.39]	[0.85, 0.87]

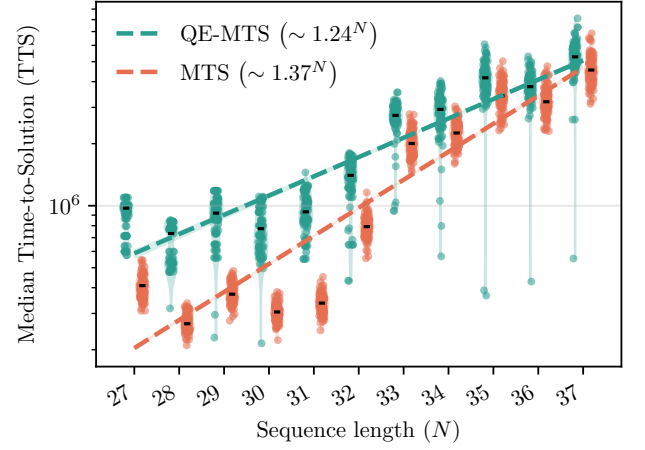


FIG. 2. Per- N distributions of per-replicate medians $\tilde{Y}_{N,m,r}$ for QE-MTS (teal) and MTS (orange) on a log-scaled TTS axis. Each dot is a replicate median; dashed curves are log-linear fits to $Q_{0.50}(N, m)$ (median-of-medians), with fitted bases $\sim 1.24^N$ (QE-MTS) and $\sim 1.37^N$ (MTS). Lower positions indicate fewer function evaluations (faster).

(95% CI [1.36, 1.37]) for MTS, indicating a shallower growth of typical TTS under QE-MTS. This suggests that scaling of MTS improves when it is quantumly enhanced. Besides, the scaling of the lower and higher quantiles, $Q_{0.10}$ and $Q_{0.90}$, also lie in a similar range of κ , indicating that the typical TTS is not expected to spread drastically as the system size grows. Nevertheless, quantiles lower than $Q_{0.10}$ may be affected by outliers, resulting in lower scaling but worse R^2 values. We also note that the obtained exponent bases may slightly change depending on the chosen range of N . For instance, including $N > 37$ changes the scaling of MTS from $O(1.37^N)$ to $O(1.34^N)$. This small change indicates that the chosen range ($27 \leq N \leq 37$) provides a reasonable reference.

In Fig. 2, we compare the typical TTS across different sequence lengths by plotting the distribution of per-replicate medians for each method. At smaller N , the MTS bulk data is notably lower than QE-MTS, indicating a smaller intercept, i.e., faster typical runs. Nevertheless, the trend of QE-MTS is a slower increase, implying a scaling advantage with respect to MTS. Moreover, we observe that in the region of $29 \leq N \leq 37$ some of the QE-MTS medians lay at lower values with respect all their MTS relatives, which is specially remarkable for $33 \leq N \leq 37$. In fact, these outliers start to be captured by $Q_{0.04}$, where κ drops down to 1.17 at the

cost of a lower quality fit with $R^2 \approx 0.5$. In the studied sequence length range, QE-MTS remains slightly above MTS in typical TTS, even though the gap visually narrows. In Appendix C, we quantify this gap by computing the log-ratio $\log_{10}(\text{TTS}_{\text{QE-MTS}}) - \log_{10}(\text{TTS}_{\text{MTS}})$ across the system size; and in Appendix D we present a simple alternative that sets the next steps to further lower the TTS such that the enhancement is two-fold, in the scaling as well as the intercept from Eq. (8).

B. Estimated crossover

To turn the qualitative pattern from Fig. 2 into a decision rule, we ask at what N the scaling advantage becomes relevant, i.e., when QE-MTS is expected provide lower typical TTS values. For this, we define a conservative crossover N_{\times} as the sequence length where the upper region of the QE-MTS data falls below the lower region of the MTS one. Concretely, from the fits (see Eq. (8)) of $Q_{0.95}(N, \text{QE-MTS})$ and $Q_{0.05}(N, \text{MTS})$. In each bootstrap draw b , N_{\times} is computed

$$N_{\times}^{(b)} = \frac{\alpha_{\text{MTS}, 0.05}^{(b)} - \alpha_{\text{QE-MTS}, 0.95}^{(b)}}{\beta_{\text{QE-MTS}, 0.95}^{(b)} - \beta_{\text{MTS}, 0.05}^{(b)}}. \quad (9)$$

The expected crossover is then estimated to occur at

$$\text{median } N_{\times} = 46.598, \quad 95\% \text{ CI } [44.937, 48.863]. \quad (10)$$

Therefore, under this conservative bulk criterion, QE-MTS becomes the safer choice at $N \gtrsim 47$. Note that this estimation does not include the cost of drawing the samples from the quantum computer, since we are interested solely in the effect of the initial quantum population. Nevertheless, in a practical scenario, it is important to include the time units of TTS, i.e., re-formulate it as $\text{TTS} = T_Q + T_C$. Here, $T_C = \tau_C \times \#$ objective-evaluations, with τ_C the effective wall-clock time to evaluate once the objective function; and $T_Q = \tau_Q \times n_{\text{shots}}$, with τ_Q the effective wall-clock time to do one shot. While this shifts the crossover point in practice, the scaling improvement of QE-MTS remains.

V. CONCLUSIONS

We have presented a hybrid non-variational quantum-classical optimization algorithm, the quantum-enhanced memetic tabu search (QE-MTS), for the low-autocorrelation binary sequence problem. By initializing a classical memetic tabu search with samples generated through digitized counterdiabatic quantum optimization, QE-MTS exhibits an empirical scaling advantage over the best-known classical heuristic. In the range $27 \leq N \leq 37$, QE-MTS lowers the typical scaling from $O(1.37^N)$ to $O(1.24^N)$. The quantum stage achieves approximately a sixfold reduction in circuit depth compared to the QAOA, providing a more practical route toward realizing quantum advantage on early fault-tolerant hardware.

These findings demonstrate that combining problem-specific quantum samplers with high-performance classical

metaheuristics can deliver measurable improvements in scaling for combinatorial optimization. More broadly, QE-MTS illustrates a class of quantum-enhanced optimization strategies in which quantum subroutines augment mature classical solvers, offering a realistic pathway toward achieving computational advantage in near-term quantum systems for real-scale and industrially relevant optimization problems.

Future work includes exploring adaptive DCQO variants to enhance the diversity of the initial population to decrease the typical time-to-solution while maintaining the observed scaling advantage. Finally, while our study evaluates a single quantum-classical pipeline, it does not discard the usage of other classical or quantum heuristics as the source of seeds.

ACKNOWLEDGMENTS

We thank Bhargava Balaganchi and Anton Simen for valuable discussions. We thank Michael Falkenthal and Sebastian Wagner for their support with GPU access.

Appendix A: Spectrum size of the LABS problem

To estimate the size of the spectrum, we start by showing that the objective function (Eq. (1)) is upper bounded by $O(N^3)$,

$$\begin{aligned} 0 \leq E(s) &= \sum_{k=1}^{N-1} \left(\sum_{i=1}^{N-k} s_i s_{i+k} \right)^2 \\ &\leq \sum_{k=1}^{N-1} \left(\sum_{i=1}^{N-k} 1 \right)^2 \\ &= \frac{N(N-1)(2N-1)}{6} \\ &= O(N^3). \end{aligned} \quad (A1)$$

Additionally, two distinct energy levels have a minimal separation of $\Delta E = E(s) - E(s') = 4$, which follows from $s_i s_{i+k} \pm s'_i s'_{i+k} = 0, \pm 2$ and

$$\Delta E = \sum_{k=1}^{N-1} \left(\sum_{i=1}^{N-k} (s_i s_{i+k} + s'_i s'_{i+k}) \right) \left(\sum_{i=1}^{N-k} (s_i s_{i+k} - s'_i s'_{i+k}) \right) \quad (A2)$$

This means that the energy levels are distributed in steps of 4 while being upper bounded $O(N^3)$. Therefore, the spectrum size also scales as $O(N^3)$ in the worst case.

Appendix B: Efficient construction of DCQO circuits

In this section, we describe how to build the DCQO circuits, which requires the computation of the first-order counterdiabatic coefficient, as well as the transpilation into quantum gates. The DCQO circuit is given by $U(T, 0) = \prod_{k=1}^{n_{\text{tot}}} \exp[\theta(k\Delta t) O_1(k\Delta t)]$, where $\theta(t) = \Delta t \alpha(t) \lambda(t)$ and

$$\begin{aligned}
-iO_1 = & 4 \sum_{i=1}^{N-2} h_i^x \sigma_i^y \sum_{k=1}^{\lfloor \frac{N-i}{2} \rfloor} \sigma_{i+k}^z + 4 \sum_{i=1}^{N-2} \sigma_i^z \sum_{k=1}^{\lfloor \frac{N-i}{2} \rfloor} h_{i+k}^x \sigma_{i+k}^y \\
& + 8 \sum_{i=1}^{N-3} h_i^x \sigma_i^y \sum_{t=1}^{\lfloor \frac{N-i-1}{2} \rfloor} \sum_{k=t+1}^{N-i-t} \sigma_{i+t}^z \sigma_{i+k}^z \sigma_{i+k+t}^z + 8 \sum_{i=1}^{N-3} \sigma_i^z \sum_{t=1}^{\lfloor \frac{N-i-1}{2} \rfloor} \sum_{k=t+1}^{N-i-t} h_{i+t}^x \sigma_{i+t}^y \sigma_{i+k}^z \sigma_{i+k+t}^z \\
& + 8 \sum_{i=1}^{N-3} \sigma_i^z \sum_{t=1}^{\lfloor \frac{N-i-1}{2} \rfloor} \sum_{k=t+1}^{N-i-t} h_{i+k}^x \sigma_{i+t}^z \sigma_{i+k}^y \sigma_{i+k+t}^z + 8 \sum_{i=1}^{N-3} \sigma_i^z \sum_{t=1}^{\lfloor \frac{N-i-1}{2} \rfloor} \sum_{k=t+1}^{N-i-t} h_{i+k+t}^x \sigma_{i+t}^z \sigma_{i+k}^z \sigma_{i+k+t}^y.
\end{aligned} \tag{B1}$$

After trotterization, we can rewrite $U(T, 0)$ as

$$\begin{aligned}
U(T, 0) = & \prod_{k=1}^{n_{\text{trot}}} \left(\prod_{i=1}^{N-2} \prod_{k=1}^{\lfloor \frac{N-i}{2} \rfloor} \exp \left[4i\theta(k\Delta t) h_i^x \sigma_i^y \sigma_{i+k}^z \right] \exp \left[4i\theta(k\Delta t) h_{i+k}^x \sigma_i^z \sigma_{i+k}^y \right] \right) \\
& \times \left(\prod_{i=1}^{N-3} \prod_{t=1}^{\lfloor \frac{N-i-1}{2} \rfloor} \prod_{k=t+1}^{N-i-t} \exp \left[8i\theta(k\Delta t) h_i^x \sigma_i^y \sigma_{i+t}^z \sigma_{i+k}^z \sigma_{i+k+t}^z \right] \exp \left[8i\theta(k\Delta t) h_{i+t}^x \sigma_i^z \sigma_{i+t}^y \sigma_{i+k}^z \sigma_{i+k+t}^z \right] \right. \\
& \left. \exp \left[8i\theta(k\Delta t) h_{i+k}^x \sigma_i^z \sigma_{i+t}^z \sigma_{i+k}^y \sigma_{i+k+t}^z \right] \exp \left[8i\theta(k\Delta t) h_{i+k+t}^x \sigma_i^z \sigma_{i+t}^z \sigma_{i+k}^z \sigma_{i+k+t}^y \right] \right).
\end{aligned} \tag{B2}$$

Defining the generalized Pauli rotations as $R_P(x) = \exp(-ixP)$, it follows that

$$\begin{aligned}
U(0, T) = & \prod_{k=1}^{n_{\text{trot}}} \left(\prod_{i=1}^{N-2} \prod_{k=1}^{\lfloor \frac{N-i}{2} \rfloor} R_{Y_i Z_{i+k}} (4\theta(k\Delta t) h_i^x) R_{Z_i Y_{i+k}} (4\theta(k\Delta t) h_{i+k}^x) \right) \\
& \times \left(\prod_{i=1}^{N-3} \prod_{t=1}^{\lfloor \frac{N-i-1}{2} \rfloor} \prod_{k=t+1}^{N-i-t} R_{Y_i Z_{i+t} Z_{i+k} Z_{i+k+t}} (8\theta(k\Delta t) h_i^x) R_{Z_i Y_{i+t} Z_{i+k} Z_{i+k+t}} (8\theta(k\Delta t) h_{i+t}^x) \right. \\
& \left. R_{Z_i Z_{i+t} Y_{i+k} Z_{i+k+t}} (8\theta(k\Delta t) h_{i+k}^x) R_{Z_i Z_{i+t} Z_{i+k} Y_{i+k+t}} (8\theta(k\Delta t) h_{i+k+t}^x) \right).
\end{aligned} \tag{B3}$$

The two-qubit block requires two entangling R_{ZZ} gates and four single-qubit rotations to be implemented, see Fig. 3. The four-qubit block requires 10 entangling R_{ZZ} gates and 28 single-qubit rotations to be implemented. Overall, the number of entangling gates to do a single trotter step of DCQO is equivalent to doing two layers of QAOA. To put these numbers in context, for $N = 67$ QAOA($p = 12$) circuits require 1.4 million entangling gates whereas DCQO ones require 236.3 thousand.

Now, the numerical computation of the first-order counterdiabatic coefficient $\alpha(t)$ can become computationally expensive even for moderately large system sizes due to the cubic scaling of the Hamiltonian terms, as shown in Eq. (4). In the following lines, we describe how it was analytically computed to make the circuit preparation more efficient.

Recall the total Hamiltonian $H(\lambda) = H_{\text{ad}}(\lambda) + i\lambda\alpha_1(\lambda)O_1(\lambda)$ from Sec. III. The coefficient $\alpha_1(\lambda)$ is obtained as $\alpha_1(\lambda) =$

$-\Gamma_1(\lambda)/\Gamma_2(\lambda)$ with $\Gamma_k(\lambda) = \text{tr}[O_k^\dagger(\lambda)O_k(\lambda)]$. In particular, $\Gamma_1(\lambda)$ reads as

$$\Gamma_1(\lambda) = \sum_{\alpha \in \{2,4\}} c_\alpha \sum_{A \in G_\alpha} \mathcal{S}_x(A), \quad c_2 = 16, \quad c_4 = 64. \tag{B4}$$

where $\mathcal{S}_x(A) = \sum_{p \in A} (h_p^x)^2$ and $G_{2,4}$ the sets including all the two- and four-body interactions present in Eq. (2), respectively. Following the same procedure for computing $\Gamma_2(\lambda)$,

$$\begin{aligned}
\Gamma_2(\lambda) = & -256 \left[64 \lambda^2 \mathcal{I}_{44} + \sum_{A \in G_2} \left(\lambda^2 \mathcal{S}_x(A) + (1 - \lambda)^2 (\mathcal{S}_{bx}(A) + 4 \mathcal{P}_x(A) - \mathcal{S}_x(A)^2) \right) \right. \\
& \left. + 4 \sum_{A \in G_4} \left(4 \lambda^2 \mathcal{S}_x(A) + (1 - \lambda)^2 (\mathcal{S}_{bx}(A) + 4 \mathcal{P}_x(A) - \mathcal{S}_x(A)^2) \right) + 4 \lambda^2 (4 \mathcal{I}_{24} + \mathcal{I}_{22}) \right],
\end{aligned} \tag{B5}$$

where $\mathcal{S}_{bx}(A) = \sum_{p \in A} (h_p^x h_p^b)^2$, $\mathcal{P}_x(A) = \sum_{p < q \in A} (h_p^x h_q^x)^2$ and $\mathcal{I}_{\alpha\beta}(\mathbf{a}, \mathbf{b}) := \{a_p | a_p = b_q, \forall (p, q) \in [1, \alpha] \times [1, \beta]\}$, with $\mathbf{a} := (a_0, \dots, a_\alpha) \in G_\alpha$, $\mathbf{b} := (b_0, \dots, b_\beta) \in G_\beta$ and $\alpha, \beta \in \{2, 4\}$.

Appendix C: Distance between MTS and QE-MTS

In Sec. IV we observed that the QE-MTS data approaches the MTS one, implying that the gap between the data points

decreases until an eventual crossover. Here, we quantify

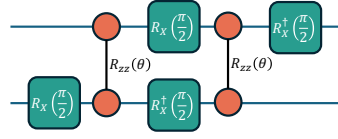


FIG. 3. Decomposition of the block of two-qubit rotations $R_{YZ}(\theta)R_{ZY}(\theta)$, requiring 2 entangling gates R_{ZZ} and 4 single-qubit gates.

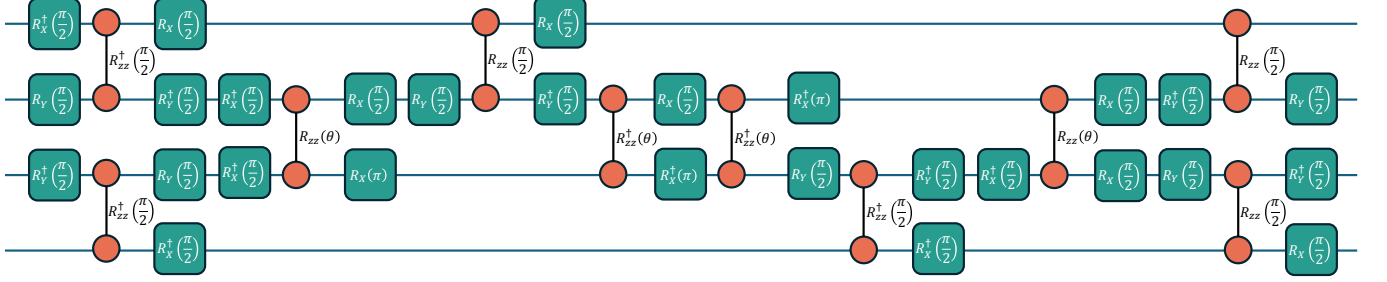


FIG. 4. Decomposition of the block of four-qubit rotations $R_{YZZZ}(\theta)R_{ZZYZ}(\theta)R_{ZZYZ}(\theta)R_{ZZYZ}(\theta)$, requiring 10 entangling gates R_{ZZ} and 28 single-qubit gates.

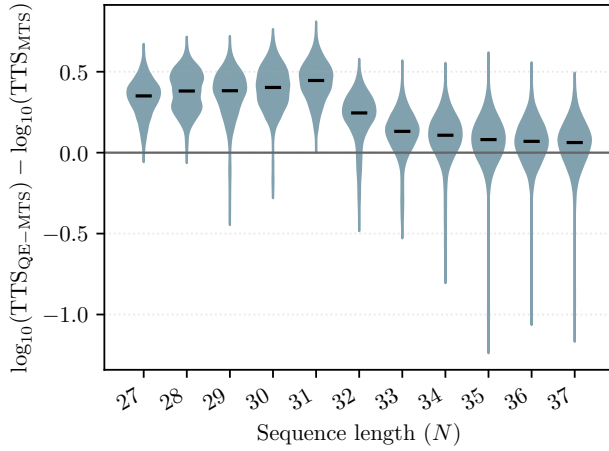


FIG. 5. Per- N distributions of $\log_{10}(\text{TTS}_{\text{QE-MTS}}) - \log_{10}(\text{TTS}_{\text{MTS}})$. Negative values indicate QE-MTS is faster (lower TTS) on a log scale. Distributions are obtained via a two-stage bootstrapping (replicates and seeds, 5000 draws per N).

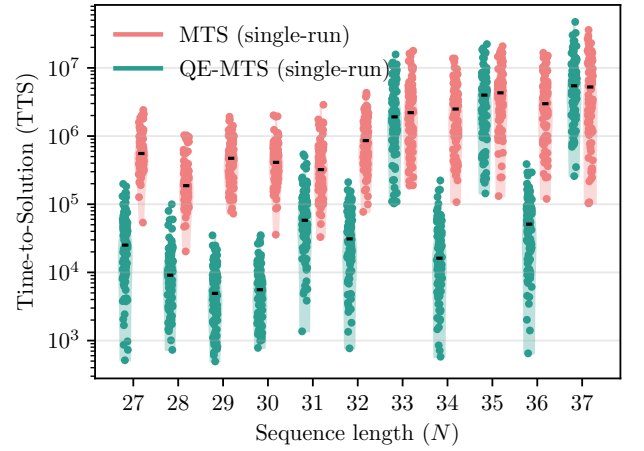


FIG. 6. Single-run comparison of time-to-solution (TTS) with sequence length N using an alternative QE-MTS, where the initial population is taken from multiple DCQO runs (teal), and MTS (orange).

this gap by computing the log-ratio $\log_{10}(\text{TTS}_{\text{QE-MTS}}) - \log_{10}(\text{TTS}_{\text{MTS}})$ across the studied system sizes. Negative values of this ratio favor QE-MTS and since this metric is logarithmic, vertical differences have a direct multiplicative meaning: a value of D implies $\text{TTS}_{\text{QE-MTS}}/\text{TTS}_{\text{MTS}} = 10^D$. In Fig. 5 we show the distribution of ratios across different sequence lengths. We observe that at small N , the distributions cluster slightly above. However, as N increases, the difference shifts systematically to negative values, indicating a growing advantage for QE-MTS. This is consistent with the smaller scaling coefficients κ in Table I and QE-MTS approaching MTS in Fig. 2. In conclusion, although plain MTS typically performs better at small sizes, the typical advantage shifts toward QE-MTS as N grows.

Appendix D: Lowering time-to-solution

One specific observation from Fig. 2 is that the TTS values for QE-MTS lie above those of MTS for system sizes in the range $27 \leq N \leq 37$. In other words, the TTS values are effectively higher in this regime, even though the slope is lower. As discussed in the main text, the crossover between QE-MTS and MTS is expected to occur around a sequence length of $N \approx 47$. This apparent discrepancy arises from the specific warm-starting procedure used to initialize MTS with quantum-enhanced solutions. Alternative initialization strategies could, in principle, lower the TTS values and bring them below the MTS reference line.

In this section, we describe one such alternative setup of QE-MTS that achieves lower TTS across multiple system sizes. This motivates future work to identify an optimal configuration where the TTS remains low while preserving the scaling advantage. In this setup, we construct the initial pop-

TABLE II. EC2 instance specifications.

Instance	GPUs	vCPUs	RAM	GPU memory	Max N
A100	8	96	1 TB	320 GB	35
H200	8	192	2 TB	1128 GB	37
B200	8	192	2 TB	1440 GB	37

ulation in Algorithm 1 using the lowest-energy bitstrings obtained from multiple independent DCQO runs. This approach makes the initialization more robust against shot noise. Figure 6 shows the performance of a single run of this QE-MTS variant compared to a typical MTS run, demonstrating that high-quality initial populations can indeed be obtained via DCQO. For most sequence lengths, the best-performing QE-MTS seeds outperform the best MTS seeds by up to two orders of magnitude. For three system sizes ($N = 33, 35, 37$), however, no enhancement was observed. Nevertheless, these isolated cases do not outweigh the consistent improvements across other system sizes and do not preclude further enhancements for larger $N > 37$.

Appendix E: GPU hardware specifications

We tested AWS *p4d.24xlarge* (A100), *p5en.48xlarge* (H200) and *P6-b200.48xlarge* (B200) instances. The B200 and H200 GPUs offer higher memory, enabling $N = 37$ in our DCQO simulations, see Fig. 7 and Table II.

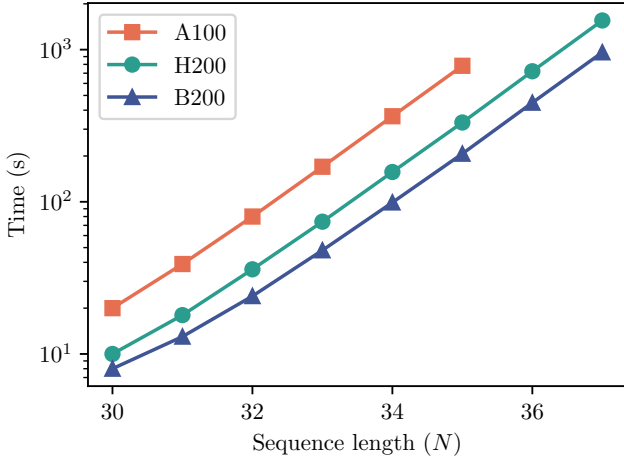


FIG. 7. Runtime comparison for NVIDIA A100, H200 and B200 GPUs for the simulation of the DCQO circuits through CUDA-Q [30].

Appendix F: Tabu search

Tabu search is a metaheuristic for combinatorial optimization that extends greedy local search with a short-term memory to avoid cycling and to encourage exploration. Recently visited states are declared tabu for a limited tenure so that the

Algorithm 2: TABUSEARCH

Input: starting sequence s_0
Output: locally improved sequence \tilde{s}

```

1  $\tilde{s} \leftarrow s_0; p \leftarrow s_0;$ 
2 Initialize  $TabuList[1:N] \leftarrow 0;$ 
3  $M \leftarrow \text{random.int}(0, N) + \lfloor N/2 \rfloor;$  //  $\in [\frac{N}{2}, \frac{3N}{2}]$ 
4 for  $t = 1$  to  $M$  do
5   Choose index  $i^*$  with minimum energy among
      $\{i \mid TabuList[i] < t\};$ 
6    $p \leftarrow \text{FLIP}(p, i^*);$ 
7    $\theta_{\min}, \theta_{\max} \leftarrow M/10, M/50$ 
      $TabuList[i^*] \leftarrow t + \text{random.int}(\theta_{\min}, \theta_{\max});$ 
8   if  $E(p) < E(\tilde{s})$  then
9      $\tilde{s} \leftarrow p$ 
10 return  $\tilde{s}$ 

```

Algorithm 3: COMBINE and MUTATE

1 Function COMBINE(p_1, p_2):
2 Choose cut point $k \in \{1, \dots, N-1\}$ uniformly;
3 **return** $p_1[1:k] \parallel p_2[k+1:N];$

4 Function MUTATE(s, p_{mut}):
5 **for** $i = 1$ **to** N **do**
6 **if** $\text{rand}(0, 1) < p_{\text{mut}}$ **then**
7 $s \leftarrow \text{FLIP}(s, i)$
8 **return** s

search can traverse non-improving regions without immediately undoing itself; an aspiration criterion lifts tabu status when a move yields a new global best [32, 33]. In our instantiation, Algorithm 2 (TABUSEARCH) iteratively inspects the one-flip neighborhood, selects the best admissible move under a bounded randomized tenure with aspiration, updates the incumbent when an improvement occurs, and runs for a randomized budget. Algorithm 3 provides the simple COMBINE (single-point crossover to seed restarts) and MUTATE (independent flips with probability p_{mut} to perturb incumbents) operators that generate starting points for the local improver.

- [1] A. Boehmer, Binary pulse compression codes, *IEEE Transactions on Information Theory* **13**, 156 (1967).
- [2] M. Schroeder, Synthesis of low-peak-factor signals and binary sequences with low autocorrelation (corresp.), *IEEE Transactions on Information Theory* **16**, 85 (1970).
- [3] T. Packebusch and S. Mertens, Low autocorrelation binary sequences, *Journal of Physics A: Mathematical and Theoretical* **49**, 165001 (2016).
- [4] I. I. Cplex, V12.10.0: User's Manual for CPLEX, International Business Machines Corporation **46**, 157 (2009).
- [5] Gurobi Optimization, LLC, *Gurobi Optimizer Reference Manual* (2024).
- [6] J. Gallardo, C. Cotta, and A. Fernández-Leiva, Finding low autocorrelation binary sequences with memetic algorithms, *Ap-*

- plied Soft Computing **9**, 1252 (2009).
- [7] R. Shaydulin, C. Li, S. Chakrabarti, M. DeCross, D. Herman, N. Kumar, J. Larson, D. Lykov, P. Minssen, Y. Sun, Y. Alexeev, J. M. Dreiling, J. P. Gaebler, T. M. Gatterman, J. A. Gerber, K. Gilmore, D. Gresh, N. Hewitt, C. V. Horst, S. Hu, J. Johansen, M. Matheny, T. Mengle, M. Mills, S. A. Moses, B. Neyenhuis, P. Siegfried, R. Yalovetzky, and M. Pistoia, Evidence of scaling advantage for the quantum approximate optimization algorithm on a classically intractable problem, *Science Advances* **10**, eadm6761 (2024).
 - [8] M. Sciorilli, G. Camilo, T. O. Maciel, A. Canabarro, L. Borges, and L. Aolita, A competitive nisq and qubit-efficient solver for the labs problem (2025), [arXiv:2506.17391 \[quant-ph\]](https://arxiv.org/abs/2506.17391).
 - [9] B. Bošković, F. Brglez, and J. Brest, Low-autocorrelation binary sequences: On improved merit factors and runtime predictions to achieve them, *Applied Soft Computing* **56**, 262 (2017).
 - [10] T. Koch, D. E. B. Neira, Y. Chen, G. Cortiana, D. J. Egger, R. Heese, N. N. Hegade, A. G. Cadavid, R. Huang, T. Itoko, T. Kleinert, P. M. Xavier, N. Mohseni, J. A. Montanez-Barrera, K. Nakano, G. Nannicini, C. O'Meara, J. Pauckert, M. Proissl, A. Ramesh, M. Schicker, N. Shimada, M. Takeori, V. Valls, D. V. Bulck, S. Woerner, and C. Zoufal, Quantum optimization benchmarking library - the intractable decathlon (2025), [arXiv:2504.03832 \[quant-ph\]](https://arxiv.org/abs/2504.03832).
 - [11] A. G. Cadavid, A. Dalal, A. Simen, E. Solano, and N. N. Hegade, Bias-field digitized counterdiabatic quantum optimization, *Phys. Rev. Res.* **7**, L022010 (2025).
 - [12] S. V. Romero, A.-M. Visuri, A. G. Cadavid, A. Simen, E. Solano, and N. N. Hegade, Bias-field digitized counterdiabatic quantum algorithm for higher-order binary optimization, *Communications Physics* **8**, 348 (2025).
 - [13] P. Chandarana, A. G. Cadavid, S. V. Romero, A. Simen, E. Solano, and N. N. Hegade, Runtime Quantum Advantage with Digital Quantum Optimization (2025), [arXiv:2505.08663 \[quant-ph\]](https://arxiv.org/abs/2505.08663).
 - [14] IBM Quantum, Isqay Quantum Optimizer - A Qiskit Function by Kipu Quantum, <https://docs.quantum.ibm.com/guides/kipu-optimization> (2025), [Online: 14/04/25].
 - [15] A. G. Cadavid, I. Montalbán, A. Dalal, E. Solano, and N. N. Hegade, Efficient digitized counterdiabatic quantum optimization algorithm within the impulse regime for portfolio optimization, *Phys. Rev. Appl.* **22**, 054037 (2024).
 - [16] P. Chandarana, S. V. Romero, A. G. Cadavid, A. Simen, E. Solano, and N. N. Hegade, Hybrid sequential quantum computing (2025), [arXiv:2510.05851 \[quant-ph\]](https://arxiv.org/abs/2510.05851).
 - [17] I. Čepaitė, N. Vaishnav, L. Zhou, and A. Montanaro, Quantum-enhanced optimization by warm starts (2025), [arXiv:2508.16309 \[quant-ph\]](https://arxiv.org/abs/2508.16309).
 - [18] E. Gardner, Spin glasses with p-spin interactions, *Nuclear Physics B* **257**, 747 (1985).
 - [19] A. J. Bray and M. A. Moore, Metastable states in spin glasses, *Journal of Physics C: Solid State Physics* **13**, L469 (1980).
 - [20] T. Aspelmeier, A. J. Bray, and M. A. Moore, Complexity of Ising Spin Glasses, *Phys. Rev. Lett.* **92**, 087203 (2004).
 - [21] D. Sherrington and S. Kirkpatrick, Solvable Model of a Spin-Glass, *Phys. Rev. Lett.* **35**, 1792 (1975).
 - [22] H. Horner, Time dependent local field distribution and metastable states in the SK-spin-glass, *The European Physical Journal B* **60**, 413–422 (2007).
 - [23] A. Lucas, Ising formulations of many NP problems, *Front. Phys.* **2**, 5 (2014).
 - [24] M. Demirplak and S. A. Rice, Adiabatic Population Transfer with Control Fields, *The Journal of Physical Chemistry A* **107**, 9937 (2003).
 - [25] M. V. Berry, Transitionless quantum driving, *Journal of Physics A: Mathematical and Theoretical* **42**, 365303 (2009).
 - [26] P. W. Claeys, M. Pandey, D. Sels, and A. Polkovnikov, Floquet-Engineering Counterdiabatic Protocols in Quantum Many-Body Systems, *Phys. Rev. Lett.* **123**, 090602 (2019).
 - [27] P. Chandarana, N. N. Hegade, I. Montalbán, E. Solano, and X. Chen, Digitized counterdiabatic quantum algorithm for protein folding, *Phys. Rev. Appl.* **20**, 014024 (2023).
 - [28] S. V. Romero, X. Chen, G. Platero, and Y. Ban, Optimizing edge-state transfer in a Su-Schrieffer-Heeger chain via hybrid analog-digital strategies, *Phys. Rev. Appl.* **21**, 034033 (2024).
 - [29] N. N. Hegade, K. Paul, Y. Ding, M. Sanz, F. Albarrán-Arriagada, E. Solano, and X. Chen, Shortcuts to Adiabaticity in Digitized Adiabatic Quantum Computing, *Phys. Rev. Appl.* **15**, 024038 (2021).
 - [30] CUDA-Q Development Team, *CUDA-Q* (2025), accessed: 2025-07-14.
 - [31] F. Glover, E. Taillard, and D. de Werra, A User's Guide to Tabu Search, *Annals of Operations Research* **41**, 3 (1993).
 - [32] F. Glover, Tabu Search—Part I, *ORSA Journal on Computing* **1**, 190 (1989).
 - [33] F. Glover, Tabu Search—Part II, *ORSA Journal on Computing* **2**, 4 (1990).