Hybrid Quantum-Classical Multi-cut Benders Approach with a Power System Application

Document Version:
Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.
Link to publication

Hybrid Quantum-Classical Multi-cut Benders Approach with a Power System Application

Nikolaos G. Paterakis, *Senior Member, IEEE*

Abstract—**Leveraging the current generation of quantum devices to solve optimization problems of practical interest necessitates the development of hybrid quantum-classical (HQC) solution approaches. In this paper, a multi-cut Benders decomposition (BD) approach that exploits multiple feasible solutions of the master problem (MP) to generate multiple valid cuts is adapted, so as to be used as an HQC solver for general mixed-integer linear programming (MILP) problems. The use of different cut selection criteria and strategies to manage the size of the MP by eliciting a subset of cuts to be added in each iteration of the BD scheme using quantum computing is discussed. The HQC optimization algorithm is applied to the Unit Commitment (UC) problem. UC is a prototypical use case of optimization applied to electrical power systems, a critical sector that may benefit from advances in quantum computing. The validity and computational viability of the proposed approach are demonstrated using the D-Wave Advantage 4.1 quantum annealer.**

Index Terms—**Benders decomposition, power systems, quantum computing, unit commitment**

I. Introduction

QUANTUM computing (QC) is an emerging technology that has the potential to help addressing hard computational problems. Although significant progress has been made in the recent years, universal error-corrected quantum computers have not been realized yet. Nonetheless, currently available quantum hardware, often called Noisy Intermediate-Scale Quantum (NISQ) [1], allows applying quantum algorithms in order to identify problems and application areas in which QC can offer an advantage with respect to classical computing. Among the many areas where a quantum speedup is expected, solving combinatorial optimization problems is one of the most prominent ones [2].

In the context of using QC for optimization, a commonly-studied class of combinatorial problems are Quadratic Unconstrained Binary Optimization (QUBO) problems [3]. A QUBO problem is represented by $\min_{\mathbf{x}} \mathbf{x}^T \mathbf{Q} \mathbf{x}$ where $\mathbf{x} \in \{0, 1\}^n$ and $\mathbf{Q} \in \mathbb{R}^{n \times n}$. QUBO problems can be transformed to Ising models ($\mathbf{x} \in \{-1, 1\}^n$) and vice versa. The limitations of NISQ hardware have triggered the development of hybrid quantum-classical (HQC) algorithms that exploit both classical and quantum resources. Popular techniques include variational approaches such as the Variational Quantum Eigensolver (VQE) [4] and the Quantum Approximate Optimization Algorithm (QAOA) [5], as well as techniques based on Grover's algorithm [6]. However, gate-based QC, which the aforementioned algorithms are designed to exploit, are still at an early development stage and, therefore, their application for solving practical-scale problems is limited [7]. On the contrary, special-purpose quantum computers, namely quantum annealers (QA), currently surpass gate-based computers both in terms of number of qubits and qubit connectivity and can be used to obtain approximate solutions of larger QUBO instances [8].

A. Decomposition-based HQC Algorithms

Although several problems can be modeled as QUBO (e.g., [9]–[11]), most optimization problems of practical interest contain both discrete and continuous variables. For instance, mixed-integer linear programming (MILP) problems are commonly used in diverse application areas, such as logistics [12], the coordination of unmanned aerial vehicles [13] and power systems [14].

In order to leverage a quantum processing unit (QPU) for solving mixed-integer problems of practical interest, decomposition-based HQC algorithms must be developed. Such HQC algorithms aim to decompose the original optimization problem into a part that can be assigned to the QPU, i.e., a QUBO problem or a problem that can be cast as a QUBO, and a part that can be solved efficiently using classical algorithms (e.g., a convex optimization problem). Recently, the development of such approaches has gained attention and various general-purpose and problem-specific techniques have been proposed.

In [15] a decomposition approach for mixed binary-continuous optimization problems with complicating constraints based on a multi-block version of the alternating direction method of multipliers (ADMM) was presented. The proposed method splits the original problem into a QUBO problem and constrained convex subproblems (SP). The QUBO problem was solved using the variational quantum eigensolver (VQE) and the quantum approximate optimization algorithm (QAOA), while the constrained convex SPs were solved with a classical solver. Despite being a general method of wide applicability, the convergence of the HQC is not guaranteed.

In [16] the Benders decomposition (BD) scheme was used as the basis for a general-purpose HQC MILP solver. BD splits the problem into a master problem (MP) that contains both binary and continuous variables and a linear programming (LP) SP. A primer on Benders decomposition is presented in Section II-B. Although BD is proven to converge to the global optimum of the original MILP problem, a major drawback of its direct application as an HQC algorithm is that it requires the discretization of the continuous variable that appears in the constraints of the MP (proxy for the SP value) at the expense of using ancillary qubits.

Apart from the aforementioned general-purpose HQC algorithms, that are based on popular decomposition techniques,

N.G. Paterakis is with the Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven, 5600MB, The Netherlands (email: n.paterakis@tue.nl).

similar ideas have been proposed for specific optimization problems (e.g., [17], [18]).

B. QC Applications in the Power & Energy Sector

Although a general-purpose decomposition-based HQC optimization approach is proposed in this paper, this study is motivated by a critical application area, namely power and energy systems. Computing has always been of paramount importance for the design and operation of power systems [19]. However, the complexity induced by the modernization of the power sector is posing computational challenges that may not be met by classical resources [20]. The energy sector is currently undergoing a transition from fossil-based to zero-carbon energy sources, triggered by global decarbonization targets. The electrical power sector is leading the energy transition through three innovation trends, namely the electrification of end-use sectors, the decentralization of energy resources and extensive digitalization. It is foreseen that electricity will become the main energy carrier due to the significant potential of integrating renewable energy resources in power systems, while the linkage with other energy carriers (e.g., transportation and gas) will be strengthened, resulting in an even more complex system [21].

In this context, QC and quantum informatics are expected to find significant applications in the energy sector. First, quantum cryptography can be leveraged in order to shield the cyber security of power systems, especially those with many distributed energy resources [22]. Also, QC has the potential to expand the current computational capabilities and facilitate the solution of complex analysis, optimization and data analytics problems in the energy domain [23]–[25], while maintaining a low energy footprint of computation [26]. Interestingly, although the prospect of benefits has been identified, technical studies that investigate the use of QC for addressing power and energy system computational problems are scarce.

In [27] the solution of the phasor measurement unit (PMU) placement problem using the D-Wave Systems 2000Q QA was investigated. The PMU placement problem was formulated as the minimum dominating set problem. It was found that for some instances the QA outperformed the classical solver CPLEX. The optimization model presented in [27] is easily converted into a QUBO, however, it does not capture the complexity of the actual problem. Similarly, in [28] fairness during the design of a heat grid was evaluated by solving a QUBO problem on a QA and the quality of the solutions was benchmarked against a graph partitioning solver without observing significant differences. In [29] simplified formulations of the facility location-allocation, unit commitment (UC) and heat exchanger network synthesis problems were solved using both the D-Wave Systems 2000Q QA and IBM Q gate-based quantum computers. It was reported that for some problem instances, contrary to the QC approach, the classical solver Gurobi failed to return an optimal solution within a given time limit. However, the problem formulations presented in [29] could either be directly recast as QUBO problems or discretization of continuous variables was performed, at the expense of using a large number of ancillary qubits in order to represent the discretized variables.

Quantum algorithms have also been applied to fundamental power system analysis. In [30] a quantum power flow algorithm that exploits the Harrow-Hassidim-Lloyd (HHL) algorithm for the solution of systems of linear equations was introduced. Although computational experiments are performed on a small-scale test system due to the limitations of NISQ technology, it was argued that an exponential quantum speedup may be attainable in the future. The HHL algorithm was also exploited in [31] as part of a workflow to assess power system security under contingencies of increasing severity. These two studies advocate that QC may find impactfull applications in risk and reliability assessment of power grids, that are currently considered intractable for complex large-scale systems.

Finally, applications of HQC algorithms that combine machine learning and quantum sampling have been applied to power system problems. For instance, in [32] a methodology to detect and classify power system faults using conditional restricted Boltzman machines and quantum generative training was proposed.

C. Contributions

The convergence properties of BD are very appealing, however, previous research proposed a straightforward implemention of the method as the basis of an HQC optimization algorithm that does not efficiently exploit NISQ resources [16]. In this paper, an alternative approach to implementing BD is investigated. Instead of discretizing the continuous variable of the MP in order to directly assign it to a QPU, a multi-cut version of BD is developed. The QPU is assigned to handle a pure binary subroutine that selects the cuts that enter the MP in order to manage its size and accelerate BD, while both the MP and the SPs are solved on a classical computer.

The contribution of this paper is threefold:

- An HQC multi-cut BD scheme that is applicable to general MILP problems without decomposable SP structure is adapted such that both classical and quantum resources can be exploited.
- A cut selection procedure that is based on two different cut selection criteria (exclusion of infeasible MP solutions, MP variable coverage) and two different cut selection strategies (minimum set cover, maximum coverage) is proposed. The QUBO reformulation of the cut selection strategies is provided and implementation details are discussed.
- The proposed HQC optimization algorithm is applied to a prototypical power system optimization problem, namely the UC problem, and the computational viability of its solution using a commercially available QA is extensively discussed.

It is to be noted that although in this paper experiments are conducted using a QA, the quantum step can also be solved on a gate-based QPU using, for instance, VQE and QAOA.

The remainder of this paper is organized as follows: in Section II the necessary theoretical background is established, while in Section III the proposed HQC multi-cut BD algorithm and the proposed cut selection procedure are detailed. Then, in Section IV the UC problem formulation that is used for

demonstration purposes is described. Numerical results are presented and discussed in Section V. Finally, conclusions are drawn in Section VI.

II. PRELIMINARIES

A. Quantum Annealing

In this section, an overview of the process of solving an optimization problem using a QA is briefly reviewed. Further details can be found in various sources, including [8] and [33]. QAs are based on the adiabatic quantum computing paradigm [34]. Quantum annealing evolves a quantum state by applying a time-dependent Hamiltonian described by (1) over a time interval $[0, T_A]$, where T_A is the annealing time:

$$H(\tau) = A(\tau)H_0 + B(\tau)H_P \tag{1}$$

The annealing path functions $A(\tau)$ and $B(\tau)$ are monotonic functions of time that satisfy the conditions $A(0) = 1$, $A(T_A) = 0$ and $B(0) = 0$, $B(T_A) = 1$ respectively. In other words, at the beggining of the annealing, the Hamiltonian is equivalent to H_0 and gradually transitions to H_P. The initial Hamiltonian H_0 for a system of V qubits is described by (2), where σ_a^x is the Pauli-x operator applied to qubit a. The initial Hamiltonian sets the qubits into an equal superposition with respect to the computational basis z (even representation of all the optimization problem solutions).

$$H_0 = -\sum_{a \in V} \sigma_a^x \tag{2}$$

The problem Hamiltonian H_P which is dominant at time T_A is described by (3) and represents the unconstrained optimization problem of interest in terms of an Ising model, whose ground state is the optimal solution. The parameters h, J are real numbers that depend on the problem to be solved (linear and quadratic biases) and σ_a^z is the Pauli-z operator applied to qubit a.

$$H_P = \sum_{a \in V} h_a \sigma_a^z + \sum_{a \in V} \sum_{b \in V, a \neq b} J_{a,b}\ \sigma_a^z \otimes \sigma_b^z \tag{3}$$

Ideally, according to the quantum adiabatic theorem, starting from the ground state of H, if the transition $A : 1 \to 0$ and $B : 0 \to 1$ is sufficiently slow, the system will remain in the ground state throughout the transition. This implies that $H(T_A)$ should also be in the ground state, i.e., it represents the optimal solution of the problem [35].

Since QAs are open systems, the final result may not be the optimal solution of the problem. For this reason, quantum annealing is applied multiple times (annealing-read cycle) in order to increase the probability of finding high quality solutions. In general, the process of solving an optimization problem using a QPU consists of three steps. First, the optimization problem must be expressed as an Ising model or, equivalently, a QUBO problem. It is possible that the logical problem formulation may impose interactions between qubits that are not directly compatible with the physical topology of the QPU. For this reason, the second step consists of finding a minor embedding of the logical problem graph such that

it is compatible with the sparse native topology of the QPU, i.e., the physical connectivity between qubits [36]. Finally, the minor-embedded problem instance is submitted to the QPU together with a set of hyperparameters, QA is performed and a set of solutions is returned.

B. Benders Decomposition

In this paper, we are concerned with MILP problems of the form (4):

$$\min_{\mathbf{x,y}} \qquad \mathbf{c}^T\mathbf{x} + \mathbf{d}^T\mathbf{y} \tag{4a}$$

$$\text{subject to} \qquad \mathbf{Ax} + \mathbf{By} \geq \mathbf{b} \tag{4b}$$

$$\mathbf{x} \in \mathbb{R}_+^n, \mathbf{y} \in \mathbb{Y}^m \tag{4c}$$

where $\mathbf{c} \in \mathbb{R}^n$, $\mathbf{d} \in \mathbb{R}^m$, $\mathbf{b} \in \mathbb{R}^q$ are vectors and $\mathbf{A} \in \mathbb{R}^{q \times n}$, $\mathbf{B} \in \mathbb{R}^{q \times m}$ are matrices of coefficients, respectively. $\mathbb{Y}$ is a set of constraints involving only the decision variables $\mathbf{y}$.

BD is a popular technique that is applied to problems with complicating decision variables such as (4) [37]. Decision variables may be considered complicating if they are involved in most of the problem constraints or render the optimization problem non-convex. If complicating variables are fixed, then the optimization problem is substantially simplified, either because it can be decomposed in SPs that can be solved independently, or because matrix $\mathbf{A}$ attains a structure which permits a straightforward solution. The central idea in BD is to decompose the original problem into an MP which contains all the integer variables and a SP in which the complicating variables $\mathbf{y}$ are fixed to tentative values (LP problem). The MP and the SP are iteratively solved. Tentative solutions $\hat{\mathbf{y}}$ are provided by solving the MP, whereas the solution of the SP yields the so-called Benders cuts, i.e., constraints that are progressively added to the MP to restrict its solution space.

Given a tentative solution of the MP ($\mathbf{y} = \hat{\mathbf{y}}$) the SP is given by (5):

$$\min_{\mathbf{x}} \qquad \mathbf{c}^T\mathbf{x} \tag{5a}$$

$$\text{subject to} \qquad \mathbf{Ax} \geq \mathbf{b} - \mathbf{B}\hat{\mathbf{y}} \tag{5b}$$

$$\mathbf{x} \in \mathbb{R}_+^n \tag{5c}$$

In practice, the dual of the SP (DSP) given by (6) is solved:

$$\max_{\mathbf{v}} \qquad V_{DSP} = (\mathbf{b} - \mathbf{B}\hat{\mathbf{y}})^T\mathbf{v} \tag{6a}$$

$$\text{subject to} \qquad \mathbf{A}^T\mathbf{v} \leq \mathbf{c} \tag{6b}$$

$$\mathbf{v} \in \mathbb{R}_+^q \tag{6c}$$

where $\mathbf{v}$ are the dual variables associated with constraints (5b). Notice that the solution space of the DSP remains the same for different tentative values $\hat{\mathbf{y}}$.

The MP is given by (7):

$$\min_{\mathbf{y},\zeta} \qquad V_{MP} = \zeta + \mathbf{d}^T\mathbf{y} \tag{7a}$$

$$\text{subject to} \qquad \mathbf{y} \in \mathbb{Y}^m \tag{7b}$$

$$\zeta \geq \zeta^{low} \tag{7c}$$

$$(\mathbf{b} - \mathbf{By})^T\mathbf{v}_i \leq \zeta, \ \forall i \in D \tag{7d}$$

$$(\mathbf{b} - \mathbf{By})^T\mathbf{u}_j \leq 0, \ \forall j \in U \tag{7e}$$

In (7) ζ is a free variable that acts as a surrogate for the value of the DSP. In order to guarantee that the MP is always bounded, an arbitrarily low bound on ζ is enforced via (7c). The set of constraints (7d), where $\mathbf{v}_i$ corresponds to an extreme point of (6), are referred to as optimality Benders cut. If the DSP is unbounded, extreme rays can be extracted instead of extreme points, giving rise to feasibility Benders cuts expressed by (7e). The sets of extreme points and extreme rays of the DSP are denoted by D and U, respectively.

In each iteration the value of the MP provides a lower bound of (4a), i.e., $LB = \hat{\zeta} + \mathbf{d}^T \hat{\mathbf{y}}$. The lower bound is monotonically increasing. A feasible solution of the DSP provides a valid upper bound $UB = (\mathbf{b} - \mathbf{B}\hat{\mathbf{y}})^T \hat{\mathbf{v}} + \mathbf{d}^T \hat{\mathbf{y}}$. The algorithm converges if $(UB - LB) < \epsilon$.

Although BD is one of the most widely applied decomposition techniques, the straightforward implementation that was presented in this section is often inefficient from a computational point of view. Several reasons, including poor feasibility and optimality cuts, initial iterations that slowly improve the bounds and slow convergence towards the end of the algorithm, have been recognized. Details about these problems, as well as a systematic review of the research that has been devoted to accelerating BD, can be found in [38].

III. METHODOLOGY

A. HQC Multi-cut Benders Decomposition

In this paper, the BD acceleration strategy that was proposed in [39] and is based on generating multiple cuts via multiple solutions (MCMS) of the MP is explored further. In particular, multiple feasibility and optimality cuts can be generated by exploiting multiple feasible (but not necessarily optimal) solutions of the MP that are available in each iteration. Although all the generated cuts may be appended to the MP in order to reduce the number of iterations, adding a large number of cuts rapidly increases the size of the MP problem and therefore, the total execution time of the algorithm. To achieve a trade-off between the increase in the size of the MP and the reduction in the number of major iterations of the decomposition algorithm, it is possible to add a subset of the cuts that are generated in each iteration. A cut selection subroutine based on pure binary problems that may be solved using QC is developed, hence the modified algorithm is an HQC optimization algorithm (HQC-MCMS). A schematic illustration of HQC-MCMS is shown in Fig. 1.

The HQC-MCMS procedure is described in Algorithm 1. Similar to the original BD algorithm, the procedure begins by setting the upper and lower bounds of the original objective function to $+\infty$ and $-\infty$, respectively (lines 1 and 2). In each iteration k, and while the difference between the upper and lower bounds remains higher than a pre-specified tolerance ϵ, a number of steps are executed. First, the MP is solved and S_k feasible solutions are obtained. Modern solvers permit the extraction of high quality feasible solutions that are found while attempting to solve a MILP problem to optimality. Note that it may be the case that $S_k < S$, where S is the number of requested solutions. If the MP is proven to be infeasible,
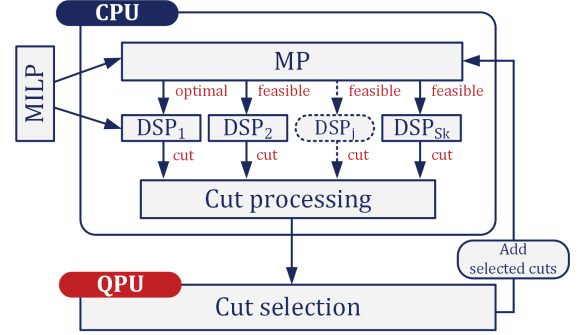

Fig. 1. Schematic representation of HQC-MCMS.

the optimization problem is also infeasible and the algorithm terminates. Otherwise, the lower bound is updated with the value V_{MP}^1 of the optimal solution (lines 5-9).

For each of the S_k available solutions the corresponding DSP is solved. Note that the DSP instances are independent and may be solved in parallel. If a DSP instance is infeasible, the algorithm terminates. If the DSP instance is unbounded, then a feasibility cut is generated and appended to the set of feasibility cuts G_k^F, whereas, if the DSP instance is optimal, then an optimality cut is generated and appended to the set of optimality cuts G_k^O (lines 10-19).

The crucial step of the HQC-MCMS algorithm is the cut selection procedure which is described by Algorithm 2 (Section III-D) and comprises two elements, namely the cut selection criterion (Section III-B) and the cut selection strategy (Section III-C). This is the step where QC can be exploited. After the execution of the cut selection subroutine, the sets of the selected feasibility and optimality cuts, $G_k^{'F}$ and $G_k^{'O}$ respectively, are added to the MP to be solved in the next iteration (lines 20-27).

Finally, the upper bound is updated using the value of the DSP corresponding to the optimal solution of the MP, V_{DSP}^1, and the iteration counter is increased. The convergence of the HQC-MCMS algorithm is discussed in Section III-E.

B. Cut Selection Criteria

1) Criterion I – cut selection based on the exclusion of infeasible solutions: In [39] cut selection was based on the observation that a subset of the generated feasibility cuts may exclude all the infeasible solutions of the MP that are identified in a given iteration. The feasibility cut that corresponds to the infeasible solution $\hat{\mathbf{y}}^i$ of the MP also excludes the infeasible solution $\hat{\mathbf{y}}^j$ if $(\mathbf{b} - \mathbf{B}\hat{\mathbf{y}}^j)^T \hat{\mathbf{u}}_i > 0$.

Let $|G_k^F|$ denote the cardinality of the set of all feasibility cuts that are generated in iteration k. Then, the $|G_k^F| \times |G_k^F|$ binary indicator matrix $\mathbf{E}$ can be constructed in order to compile information about the infeasible MP solutions that are excluded by each cut in the current iteration. Specifically, $\mathbf{E}_{ij} = 1$ if the feasibility cut i excludes the infeasible solution associated with the j-th solution of the MP, otherwise $\mathbf{E}_{ij} = 0$. Note that the diagonal elements of this matrix are always equal to one because, by definition, a feasibility cut excludes the infeasible MP solution based on which it was generated.

Algorithm 1: HQC-MCMS algorithm

Data: ϵ (tolerance), S (maximum number of MP solutions to be extracted)

1 $UB \leftarrow +\infty$;
2 $LB \leftarrow -\infty$;
3 $k \leftarrow 1$ (iteration counter) ;
4 **while** $(UB - LB) > \epsilon$ **do**
5   **Solve** MP$_k$ (7) and obtain $S_k \leq S$ solutions;
6   **if** *infeasible* **then**
7   Stop; Declare infeasibility;
8   **else**
9 $LB \leftarrow V^1_{MP}$;
10   **for** $j \leftarrow 1$ **to** S_k **do**   ▷ In parallel
11   Solve DSP$_{j,k+1}$ (6) for S_j ;
12   **if** *infeasible* **then**
13   Stop; Declare infeasibility;
14   **else if** *unbounded* **then**
15   Add cut to G^F_k ;
16   **else**
17   Add cut to G^O_k ;
18   **end**
19   **end**
20   **if** $G^F_k \neq \varnothing$ **then**
21   **Execute** Algorithm 2 on a **QPU**;
22   Add selected feasibility cuts G'^F_k to MP$_{k+1}$;
23   **end**
24   **if** $G^O_k \neq \varnothing$ **then**
25   **Execute** Algorithm 2 on a **QPU**;
26   Add selected optimality cuts G'^O_k to MP$_{k+1}$;
27   **end**
28 $UB \leftarrow \min\{UB, V^1_{DSP}\}$;
29   **end**
30 $k \leftarrow k + 1$;
31 **end**

Cut selection based on this criterion presents two potential drawbacks. First, this criterion applies only to feasibility cuts. Second, it is possible that for a given problem instance one or more cuts can exclude all the infeasible MP solutions, rendering the application of this criterion trivial.

2) Criterion II – cut selection based on MP variable coverage: It is often the case that the generated optimality and feasibility cuts are low-density. This means that the coefficient that corresponds to a decision variable of the MP in a given cut is either zero or near-zero relative to other coefficients and therefore, the contribution of a low-density cut to strengthening the MP tends to be limited [40]. For this reason, several BD acceleration strategies are based on the idea of either generating high-density Pareto optimal cuts [41] or bundles of low-density cuts [40] such that more MP decision variables are covered. In this paper, instead of focusing on the generation of cuts such that the decision variables of the MP are covered, the cuts that are generated based on multiple solutions of the MP are inspected with the purpose of

identifying a subset of feasibility (and/or optimality cuts) such that all or most of the MP decision variables are collectively covered. Note that in this paper a rather strict definition of MP variable coverage is adopted. A decision variable y_i of the MP is said to be covered in a given feasibility cut of the form $\sum_i y_i (\mathbf{B}^T \mathbf{u})_i \geq \mathbf{b}^T \mathbf{u}$, if for the i-th row of the matrix $\mathbf{B}^T \mathbf{u}$ it holds $|(\mathbf{B}^T \mathbf{u})_i| > 0$. For optimality cuts a similar definition applies if $\mathbf{u}$ is replaced by $\mathbf{v}$.

A $|G^F_k| \times m$ binary indicator matrix $\mathbf{D}^F$ can be constructed after having identified which MP decision variables are covered by a given feasibility cut in the current iteration. Specifically, $\mathbf{D}^F_{ij} = 1$ if the j-th variable of the MP is covered in cut i, otherwise $\mathbf{D}^F_{ij} = 0$. A similar matrix $\mathbf{D}^O$ ($|G^O_k| \times m$) may be constructed for optimality cuts, if cut selection is to be applied also to the set of optimality cuts. Note that some columns of $\mathbf{D}^F$ and $\mathbf{D}^O$ may be zero, that is, some decision variables may not be covered in any of the generated feasibility or optimality cuts.

Compared to Criterion I, cut selection based on the coverage of MP variables applies both to feasibility and optimality cuts and, therefore, may result in more effective MP size management. Moreover, from an implementation perspective, the construction of matrices $\mathbf{D}^F$ and $\mathbf{D}^O$ requires the evaluation of shorter expressions in comparison with $\mathbf{E}$.

C. Cut Selection Strategies

After having processed the available cuts according to one of the criteria that were presented in Section III-B, a cut selection strategy must be applied in order to identify G'^F_k and/or G'^O_k. Two such strategies based on the minimum set cover problem and the maximum coverage problem, as well as their solution using a QPU are discussed next.

1) Strategy I – minimum set cover for cut selection: The minimum set cover problem can be solved to identify a set of cuts with minimum cardinality. If cut selection is based on Criterion I, then the minimum number of feasibility cuts that exclude all the infeasible solutions in the current iteration will be identified. Similarly, if cut selection is based on Criterion II, then the minimum number of feasibility (optimality) cuts that cover all the MP decision variables that can be covered in the current iteration will be identified.

Given a set of rows matrix $\mathbf{M} \in \mathbb{R}^{|I| \times |J|}$, where I is the set of rows and J is the set of columns, the minimum set cover problem is a pure integer problem expressed by (8):

$$\min_{\boldsymbol{\chi}} \quad \sum_{i \in I} \chi_i \tag{8a}$$

$$\text{subject to} \quad \sum_{i \in I} M_{ij} \chi_i \geq 1, \ \forall j \in J \tag{8b}$$

$$\boldsymbol{\chi} \in \{0,1\}^I \tag{8c}$$

where χ_i is a binary variable that is equal to 1 if the cut i is selected and 0 otherwise. It is reiterated that the cut selection problem has to be solved in each iteration k. However, for notational simplicity, the iteration index k is dropped. Depending on the cut selection criterion, the columns of $\mathbf{M}$ represent either infeasible solutions or decision variables

of the MP (matrix $\mathbf{M}$ is accordingly replaced by matrices $\mathbf{E}$, $\mathbf{D}^F$, $\mathbf{D}^O$). Note that when Criterion II is used, there is the possibility that a number of MP decision variables cannot be covered and, therefore, (8b) should be amended by subtracting a binary slack variable from the right-hand side in order to indicate that the constraint cannot be satisfied for a particular column j. The sum of the slack variables should also be penalized in (8a). However, this can be avoided by inspecting $\mathbf{M}$ and dropping all the columns of zeros.

The minimum set cover problem is known to be NP-hard [42], i.e., it is intractable for large I and J. For this reason, various heuristics have been proposed in order to obtain approximate solutions. In this paper, the set cover problem applied to cut selection is solved using a QPU. For this reason, (8) must be recast as a QUBO problem. First, (8b) is converted to an equality constraint by adding an integer slack variable on the right hand side of the constraint and using binary expansion [43] with $\gamma_j = \lceil \log_2 \left(\sum_{i \in I} M_{ij} \right) \rceil, \ \forall j \in J$ ancillary qubits as in (9):

$$\sum_{i \in I} M_{ij}\chi_i = 1 + \sum_{\alpha=0}^{\alpha=\gamma_j-1} 2^\alpha s_{\alpha j}, \ \forall j \in J \quad (9)$$

The QUBO problem formulation is given by (10) where H is the Hamiltonian of the problem, and $\mathcal{P}_\mathcal{A}$ and $\mathcal{P}_{\mathcal{B}j}, \forall j \in J$ are positive penalties that are heuristically determined:

$$H = H_A + H_B \quad (10a)$$

where

$$H_A = \mathcal{P}_\mathcal{A} \sum_{i \in I} \chi_i \quad (10b)$$

$$H_B = \sum_{j \in J} \mathcal{P}_{\mathcal{B}j} \left[\sum_{i \in I} (M_{ij}\chi_i) - 1 - \sum_{\alpha=0}^{\alpha=\gamma_j-1} 2^\alpha s_{\alpha j} \right]^2 \quad (10c)$$

The maximum number of qubits that are required in order to represent the problem is $|I| + |J| \cdot \lceil \log_2 |I| \rceil$. If Criterion I is used then this number translates to $|G_k^F|(1 + \lceil \log_2 |G_k^F| \rceil)$. This is the case if each infeasible MP solution is excluded by every feasibility cut. If Criterion II is used, then the maximum number of qubits that are needed is $|G_k^F| + m \cdot \lceil \log_2 |G_k^F| \rceil$ if all the variables of the MP are covered by every single cut. A similar expression can be written for Criterion II applied to the set of optimality cuts.

It may be observed that for Criterion I the number of qubits that are necessary in order to represent problem (10) is independent of the size of the optimization problem (4). Instead it depends on the user-provided parameter S and is expected to be larger during the first iterations of the algorithm where mostly feasibility cuts are generated. In addition to that, inspection of the columns of matrix $\mathbf{M}$ can reveal rows (feasibility cuts) that exclude all the infeasible solutions in the current iteration and avoid triggering cut selection. The number of qubits in case Criterion II is applied depends on the number of MP variables and may be particularly large. However, for many practical applications the generated cuts are typically expected to be low-density, which implies that a

relatively large number of columns of matrix $\mathbf{M}$ are expected to be dropped because all their entries are zero. In both cases, multiple cuts may exclude the same infeasible solutions or cover the same variables. Therefore, the rows of matrix $\mathbf{M}$ can also be inspected in order to remove duplicates, keeping the row that corresponds to a higher quality MP solution.

2) Strategy II – maximum coverage for cut selection: The conservativeness of Strategy I may lead to a large number of cuts being added to the MP. To address this problem, the maximum coverage problem can be solved as an alternative cut selection strategy in order to select at most a predefined number of cuts such that, depending on the cut selection criterion that is applied, either the maximum number of infeasible solutions are excluded or the maximum number of MP decision variables are covered. To the best of the Author's knowledge, maximum coverage has not been used as a cut selection strategy in the context of BD before.

Given a matrix $\mathbf{M} \in \mathbb{R}^{|I| \times |J|}$ and a maximum number of cuts to be selected $\mathcal{M}$, the maximum coverage problem [44] is a pure integer problem that is formulated in (11):

$$\max_{\boldsymbol{\chi}, \boldsymbol{\phi}} \quad \sum_{j \in J} \phi_j \quad (11a)$$

$$\text{subject to} \quad \sum_{i \in I} \chi_i \leq \mathcal{M} \quad (11b)$$

$$\sum_{i \in I} M_{ij}\chi_i \geq \phi_j, \ \forall j \in J \quad (11c)$$

$$\boldsymbol{\chi} \in \{0,1\}^I, \boldsymbol{\phi} \in \{0,1\}^J \quad (11d)$$

The QUBO problem formulation is given by (12) where H is the Hamiltonian of the problem and $\mathcal{P}_\mathcal{A}$, $\mathcal{P}_{\mathcal{B}j}, \forall j \in J$ and $\mathcal{P}_\mathcal{C}$ are positive penalties that are heuristically determined. First, (11b) is converted into an equality constraint using integer slack variables on the left-hand side of the constraint and binary expansion introducing $\gamma = \lceil \log_2(\mathcal{M}+1) \rceil$ ancillary qubits. Note that (11b) can be replaced by an equality constraint if $\mathcal{M} \ll |I|$, thus avoiding the use of slack variables. Similarly, (11c) can be converted into an equality constraint using $\gamma_j = \lceil \log_2(\min(\mathcal{M}, (\sum_{i \in I} M_{ij})) + 1) \rceil, \ \forall j \in J$ ancillary qubits.

$$H = H_A + H_B + H_C \quad (12a)$$

where

$$H_A = -\mathcal{P}_\mathcal{A} \sum_{j \in J} \phi_j \quad (12b)$$

$$H_B = \sum_{j \in J} \mathcal{P}_{\mathcal{B}j} \left[\sum_{i \in I} (M_{ij}\chi_i) - \phi_j - \sum_{\alpha=0}^{\alpha=\gamma_j-1} 2^\alpha s_{\alpha j} \right]^2 \quad (12c)$$

$$H_C = \mathcal{P}_\mathcal{C} \left(\sum_i \chi_i - \mathcal{M} + \sum_{\alpha=0}^{\alpha=\gamma-1} 2^\alpha s_\alpha \right)^2 \quad (12d)$$

The maximum number of qubits that are required in order to represent the problem, assuming that (12d) is replaced by an equality constraint, is $|I| + |J| \cdot (1 + \lceil \log_2(\mathcal{M}+1) \rceil)$. If Criterion I is used, this number translates to $|G_k^F| + |G_k^F| \cdot$

$(1 + \lceil \log_2(\mathcal{M} + 1) \rceil)$. This is the case if each infeasible MP solution can be excluded by at least $\mathcal{M}$ cuts. If Criterion II is used, then the maximum number of qubits that are required is $|G_k^F| + m \cdot (1 + \lceil \log_2(\mathcal{M} + 1) \rceil)$ in case all the MP variables can be covered by at least $\mathcal{M}$ cuts. A similar expression can be written for Criterion II applied to the set of optimality cuts. Although these numbers may appear to be prohibitively large, in practice, the observations of Section III-C1 hold.

Algorithm 2: Cut selection procedure

Data: G_k^F, G_k^O, *cutSelectionCriterion*, *cutSelectionStrategy*, *optSelect* (boolean; whether to apply cut selection on optimality cuts), hyperparameters

1 **if** *cutSelectionCriterion = Criterion I* **then**
2   Construct $\mathbf{E}$;
3   Inspect $\mathbf{E}$;
4   **if** *cutSelectionStrategy = Strategy I* **then**
5   Solve (10) to select feasibility cuts;
6   **end**
7   **if** *cutSelectionStrategy = Strategy II* **then**
8   Solve (12) to select feasibility cuts;
9   **end**
10   Return $G_k^{'F}$;
11 **end**
12 **if** *cutSelectionCriterion = Criterion II* **then**
13   Construct $\mathbf{D}^F$;
14   Inspect $\mathbf{D}^F$;
15   **if** *cutSelectionStrategy = Strategy I* **then**
16   Solve (10) to select feasibility cuts;
17   **end**
18   **if** *cutSelectionStrategy = Strategy II* **then**
19   Solve (12) to select feasibility cuts;
20   **end**
21   **if** *optSelect* **then**
22   Construct $\mathbf{D}^O$;
23   Inspect $\mathbf{D}^O$;
24   **if** *cutSelectionStrategy = Strategy I* **then**
25   Solve (10) to select optimality cuts;
26   **end**
27   **if** *cutSelectionStrategy = Strategy II* **then**
28   Solve (12) to select optimality cuts;
29   **end**
30   **end**
31   Return $G_k^{'F}$ and $G_k^{'O}$;
32 **end**

D. Cut Selection Procedure

The cut selection problem that is invoked in lines 21 and 25 of Algorithm 1 is described by Algorithm 2. If cut selection is based on Criterion I, the rows of matrix $\mathbf{E}$ are inspected in order to identify cuts that exclude the same infeasible solutions (line 3). If such duplicate rows are found, the row corresponding to the cut associated with an MP solution with smaller objective function value is kept, while the rest of

the rows are dropped. Then, depending on the cut selection strategy, the corresponding optimization problem is solved (lines 5 and 8) and $G_k^{'F}$ is returned. If Criterion II is used, matrices $\mathbf{D}^F$ and $\mathbf{D}^O$ are inspected depending on whether cut selection is applied both to feasibility and optimality cuts (lines 14 and 23). First, their rows are inspected similarly to the rows of $\mathbf{E}$. Then, since it may not be possible to cover a number of MP variables by any of the cuts (all column entries are zero), the respective columns are dropped. Depending on the cut selection strategy, the corresponding optimization problem is solved (lines 25 and 28) using the modified matrix.

The inspection step may significantly reduce the size of the matrices before the cut selection optimization problems are solved. Additionally, all three matrices are inspected in order to identify whether a single cut can either exclude all the infeasible solutions (Criterion I) or covers all the MP variables that can be covered (Criterion II). If such a cut is found, the cut selection procedure terminates without triggering the solution of an optimization problem and the set of selected feasibility and/or optimality cuts that is returned contains only a single cut.

E. Convergence of HQC-MCMS

Despite the fact that the quantum step does not necessarily provide an optimal (or even feasible) solution of the cut selection problem, the HQC-MCMS algorithm is guaranteed to always converge to the global optimum of Problem (4). This is due to the fact that sets G_k^F and G_k^O and, therefore, any non-empty subsets $G_k^{'F}$ and $G_k^{'O}$ contain valid cuts [39]. Moreover, the MP is solved to optimality at each iteration. Then, the feasibility or optimality cut that is generated at the current iteration using the optimal value of the MP in the DSP can always be added in the next iteration to prevent $G_k^{'F}$ and $G_k^{'O}$ from being empty. Note that according to [45], it is also possible to avoid solving the MP to optimality in each iteration while guaranteeing convergence; however, the practical implications are out of the scope of this study.

IV. Use Case: The Unit Commitment Problem

The UC problem is a fundamental power system optimization problem that aims to optimaly schedule and dispatch the available generation and demand side resources [46]. Various BD schemes have been applied in order to solve different versions of this problem [47]–[51]. To demonstrate the applicability of the proposed HQC algorithm a simple version of the UC problem with typical operational and network constraints is adopted. First, in Section IV-A the complete MILP problem formulation is presented. Then, the DSP and MP formulations are derived in Sections IV-B and IV-C.

A. MILP Formulation

The UC problem formulation is described by (13a)-(13o). The notation that is used is presented in Table I. The dual variables associated with each constraint set are denoted by greek letters that are displayed in parentheses next to the corresponding equation. Note that all the dual variables associated with

TABLE I
NOTATION USED IN THE UC FORMULATION

Sets and indices

$g\,(G)$	Index (set) of generators
$t\,(T)$	Index (set) of time
$i, j\,(I)$	Indices (set) of buses
I^0	Set of reference bus
$l(L)$	Index (set) of loads

Parameters

A_{ig}^G	Generator incidence matrix; 1 if generator g is connected to bus i
A_{il}^L	Load incidence matrix; 1 if load l is connected to bus i
C_g	Energy cost of generator g (\$/MWh)
D_{lt}	Demand of load l in time t (MW)
NLC_g	No-load cost of generator g (\$/h)
SUC_g	Start-up cost of generator g (\$)
SDC_g	Shut-down cost of generator g (\$)
P_g^{max}	Maximum power output of generator g (MW)
P_g^{min}	Minimum power output of generator g (MW)
P_g^{ini}	Initial power output of generator g (MW)
RU_g	Ramp-up rate of generator g (MW/h)
RD_g	Ramp-down rate of generator g (MW/h)
X_{ij}	Reactance of line (i, j) (pu)
F_{ij}^{max}	Maximum power flow through line (i, j) (MW)
B_{ij}	(i, j) admittance matrix element (pu)
u_g^{ini}	Initial state of generator g; 1 if generator g was online before the beginning of the scheduling horizon
ζ^{low}	Lower bound of the proxy variable ζ (\$)

Decision variables

P_{gt}	Power output of generator g in time t (MW)
θ_{it}	Voltage angle of bus i in time t (rad)
ζ	Proxy for the DSP value (\$)
u_{gt}	Binary variable; 1 if generator g is online in time t
y_{gt}	Binary variable; 1 if generator g is starting-up in time t
z_{gt}	Binary variable; 1 if generator g is shutting-down in time t

the inequality constraints are non-negative, whereas the dual variables associated with equality constraints are unrestricted.

$$\min_{\boldsymbol{P},\boldsymbol{\theta},\boldsymbol{u},\boldsymbol{y},\boldsymbol{z}} \quad TC$$

$$\text{where } TC = \sum_t \sum_g (C_g P_{gt}$$
$$+ NLC_g u_{gt} + SUC_g y_{gt} + SDC_g z_{gt}) \quad (13\text{a})$$

subject to:

$$y_{gt} - z_{gt} = u_{gt} - u_{g(t-1)} \quad \forall g, t > 1 \quad (13\text{b})$$

$$y_{gt} - z_{gt} = u_{gt} - u_g^{ini} \quad \forall g, t = 1 \quad (13\text{c})$$

$$-P_{gt} \ge -P_g^{max} u_{gt} \quad \forall g, t \quad (\mu_{gt}^+) \quad (13\text{d})$$

$$P_{gt} \ge P_g^{min} u_{gt} \quad \forall g, t \quad (\mu_{gt}^-) \quad (13\text{e})$$

$$-P_{gt} + P_{g(t-1)} \ge -RU_g \quad \forall g, t > 1 \quad (\nu_{gt}^+) \quad (13\text{f})$$

$$-P_{gt} \ge -RU_g - P_g^{ini} \quad \forall g, t = 1 \quad (\nu_g^{0+}) \quad (13\text{g})$$

$$P_{gt} - P_{g(t-1)} \ge -RD_g \quad \forall g, t > 1 \quad (\nu_{gt}^-) \quad (13\text{h})$$

$$P_{gt} \ge -RD_g + P_g^{ini} \quad \forall g, t = 1 \quad (\nu_g^{0-}) \quad (13\text{i})$$

$$\frac{1}{X_{ij}}(\theta_{jt} - \theta_{it}) \ge -F_{ij}^{max} \quad \forall (i,j)|X_{ij} \neq 0, t \quad (\psi_{ijt}) \quad (13\text{j})$$

$$\theta_{i,t} = 0 \quad \forall t, i \in I^0 \quad (\lambda_t^0) \quad (13\text{k})$$

$$\sum_{j \in I} B_{ij}\theta_{jt} - \sum_{g \in G} P_{gt}A_{ig}^G = -\sum_{l \in L} D_{lt}A_{il}^L \quad \forall i, t \quad (\lambda_{it}) \quad (13\text{l})$$

$$P_{gt} \ge 0 \quad \forall g, t \quad (13\text{m})$$

$$\theta_{it} \in \mathbb{R} \quad \forall i, t \quad (13\text{n})$$

$$u_{gt}, y_{gt}, z_{gt} \in \{0, 1\} \quad \forall g, t \quad (13\text{o})$$

The objective function is expressed by (13a) and stands for the minimization of the total energy and commitment cost across the time horizon. Constraints (13b) and (13c) determine the generator commitment logic. The power output of generators is limited by (13d) and (13e). The intertermporal constraints (13f)-(13i) limit the change in the power output of generators in consecutive time periods according to their up and down ramp rates. Network constraints are modeled in terms of a DC power flow approximation. The power that flows through transmission lines is constrained by (13j), while (13k) fixes the voltage angle at the reference bus. The power balance at each bus is determined by (13l). Finally, (13m)-(13o) determine the domain of the decision variables.

B. Dual Subproblem

For a tentative solution $\hat{u}_{gt}$ of the MP, the DSP is an LP problem that is expressed by (14a)-(14k).

$$\max_{\boldsymbol{\mu}^+,\boldsymbol{\mu}^-,\boldsymbol{\nu}^+,\boldsymbol{\nu}^-,\boldsymbol{\nu}^{0-},\boldsymbol{\nu}^{0+},\boldsymbol{\psi},\boldsymbol{\lambda},\boldsymbol{\lambda}^0} \quad UC^{DSP}$$

$$\text{where } UC^{DSP} = \sum_t \sum_g \left(\mu_{gt}^- P_g^{min} - \mu_{gt}^+ P_g^{max}\right)\hat{u}_{gt}$$

$$-\sum_t \sum_g \left(\nu_{gt}^+ RU_g + \nu_{gt}^- RD_g\right)$$

$$+\sum_g \left[\nu_g^{0+}\left(-P_g^{ini} - RU_g\right) + \nu_g^{0-}\left(P_g^{ini} - RD_g\right)\right]$$

$$-\sum_t \sum_i \left[\lambda_{it}\sum_l \left(D_{lt}A_{il}^L\right) + \sum_{j|X_{ij}\neq 0}\psi_{ijt}F_{ij}^{max}\right] \quad (14\text{a})$$

subject to:

$$-\mu_{gt}^+ + \mu_{gt}^- - \nu_g^{0+} + \nu_g^{0-} + \nu_{g(t+1)}^+ - \nu_{g(t+1)}^-$$
$$-\sum_i A_{ig}^G \lambda_{it} \le C_g \quad \forall g, t = 1 \quad (14\text{b})$$

$$-\mu_{gt}^+ + \mu_{gt}^- - \nu_{gt}^+ + \nu_{gt}^-$$
$$-\sum_i A_{ig}^G \lambda_{it} \le C_g \quad \forall g, t = |T| \quad (14\text{c})$$

$$-\mu_{gt}^+ + \mu_{gt}^- - \nu_{gt}^+ + \nu_{g(t+1)}^+ + \nu_{gt}^- - \nu_{g(t+1)}^-$$
$$-\sum_i A_{ig}^G \lambda_{it} \le C_g \quad \forall g, t \in (1, |T|) \quad (14\text{d})$$

$$\lambda_t^0 + \sum_j \left(\lambda_{jt}B_{ji}\right)$$
$$+ \sum_{j|X_{ij}\neq 0}\left(\frac{\psi_{jit} - \psi_{ijt}}{X_{ij}}\right) = 0 \quad \forall i \in I^0, t \quad (14\text{e})$$

$$\sum_j \left(\lambda_{jt}B_{ji}\right)$$
$$+ \sum_{j|X_{ij}\neq 0}\left(\frac{\psi_{jit} - \psi_{ijt}}{X_{ij}}\right) = 0 \quad \forall i \in I - I^0, t \quad (14\text{f})$$

$$\mu_{gt}^+, \mu_{gt}^-, \nu_{gt}^+, \nu_{gt}^- \ge 0 \quad \forall g, t \quad (14\text{g})$$

$$\nu_g^{0-}, \nu_g^{0+} \geq 0 \quad \forall g \tag{14h}$$

$$\psi_{ijt} \geq 0 \quad \forall (i,j)|X_{ij} \neq 0, t \tag{14i}$$

$$\lambda_t^0 \in \mathbb{R} \quad \forall t \tag{14j}$$

$$\lambda_{it} \in \mathbb{R} \quad \forall i, t \tag{14k}$$

Constraints involving binary variables (e.g., generator minimum up and down time, reserve capacity constraints, etc.) can be directly added to the formulation of Section IV-A since they do not affect the DSP.

C. Master Problem

Given $G_k^{'F}$ and $G_k^{'O}$, the MP is a MILP problem that is expressed by (15a)-(15e).

$$\min_{\zeta, \boldsymbol{u}, \boldsymbol{y}, \boldsymbol{z}} \quad \zeta + \sum_t \sum_g \left(NLC_g u_{gt} + SUC_g y_{gt} + SDC_g z_{gt} \right) \tag{15a}$$

subject to:

$(13b), (13c)$

$$\left\{ \sum_t \sum_g \left(\hat{\mu}_{gt}^- P_g^{min} - \hat{\mu}_{gt}^+ P_g^{max} \right) u_{gt} \right.$$
$$- \sum_t \sum_g \left(\hat{\nu}_{gt}^+ RU_g + \hat{\nu}_{gt}^- RD_g \right)$$
$$+ \sum_g \left[\hat{\nu}_g^{0+} \left(-P_g^{ini} - RU_g \right) + \hat{\nu}_g^{0-} \left(P_g^{ini} - RD_g \right) \right]$$
$$- \sum_t \sum_i \left[\hat{\lambda}_{it} \sum_l \left(D_{lt} A_{il}^L \right) + \sum_{j|X_{ij} \neq 0} \hat{\psi}_{ijt} F_{ij}^{max} \right]$$
$$\left. \leq \zeta \right\}^{(\kappa)} \in \bigcup_{\kappa=2}^k G_\kappa^{'O} \tag{15b}$$

$$\left\{ \sum_t \sum_g \left(\hat{\mu}_{gt}^- P_g^{min} - \hat{\mu}_{gt}^+ P_g^{max} \right) u_{gt} \right.$$
$$- \sum_t \sum_g \left(\hat{\nu}_{gt}^+ RU_g + \hat{\nu}_{gt}^- RD_g \right)$$
$$+ \sum_g \left[\hat{\nu}_g^{0+} \left(-P_g^{ini} - RU_g \right) + \hat{\nu}_g^{0-} \left(P_g^{ini} - RD_g \right) \right]$$
$$- \sum_t \sum_i \left[\hat{\lambda}_{it} \sum_l \left(D_{lt} A_{il}^L \right) + \sum_{j|X_{ij} \neq 0} \hat{\psi}_{ijt} F_{ij}^{max} \right]$$
$$\left. \leq 0 \right\}^{(\kappa)} \in \bigcup_{\kappa=2}^k G_\kappa^{'F} \tag{15c}$$

$$\zeta \geq \zeta^{low} \tag{15d}$$

$$u_{gt}, y_{gt}, z_{gt} \in \{0, 1\} \quad \forall g, t \tag{15e}$$

The optimality and feasibility cuts that are appended to the MP up to the current iteration k are expressed by (15b) and (15c) respectively. Constraint (15d) is necessary in order to prevent the problem from being unbounded in the first iteration of the algorithm.

V. NUMERICAL EXPERIMENTS

A. Implementation Details

The HQC-MCMS algorithm was implemented in Python 3.9 using the Pyomo package [52]. All the classical MILP problems were solved using the Gurobi 9.5.0 solver [53] with a *MIPGap* of 0 % on a workstation with 2 Intel Xeon processors (24 cores, 3 GHz) and 128 GB of RAM. In order to find multiple solutions of the MP the solution pool functionality of Gurobi was used. The *PoolSearchMode* parameter was set to 1. This means that the solver continues the search for feasible solutions after the optimal solution has been found, however, no guarantees are provided about the quality of the additional feasible solutions.

The QUBO problem instances were solved using the D-Wave Advantage 4.1 QA that can be accessed via Amazon Braket. The D-Wave Advantage 4.1 QPU relies on the Pegasus topology and features more than 5000 qubits and 35000 couplers [54]. Embedding of the problem on the physical QPU graph was performed using the *minorminer* package using the default settings. For all problems that were submitted to the QPU the annealing time was set to 10 μs and the anneal-read cycle was repeated 1000 times. It is to be noted that the hyperparameters associated with solving problems on a QPU may impact the solution time and quality. Thus, they should be carefully tuned. However, due to the considerable utilization cost of QPUs and the diversity of problem instances that are solved, only a limited number of tuning attempts were performed in this study. The chain strength value was set to 150% of the largest interaction coefficient observed in the problem Hamiltonian. According to the recommendation in [2] for general binary integer linear problems for problems of the type (10), $\mathcal{P}_\mathcal{A} = 1$ and $\mathcal{P}_{\mathcal{B}j} = |I|$, $\forall j \in J$ and for problems of the type (12), $\mathcal{P}_\mathcal{A} = 1$, $\mathcal{P}_{\mathcal{B}j} = |I| + |J|$, $\forall j \in J$ and $\mathcal{P}_\mathcal{C} = |I| + |J|$, where I, J are the sets of rows and columns of matrix $\mathbf{M}$ respectively.

B. Input Data

To investigate the applicability of the proposed methodology on the UC problem, the test systems displayed in Fig. 2 were randomly generated. The network topologies were generated based on the methodology that was proposed in [56]. First, buses are uniformly placed within a fixed area with width and height of 1. Then, given a distance requirement (in this study $[0, 0.4]$) between neighboring buses, the set of transmission lines are selected by sampling a Poisson distribution with its parameter set to 2.67. The reactance of a transmission line is considered proportional to its length. For simplicity a factor of 1/10 is assumed for all lines. Subsequently, the type of each bus is decided. It is assumed that 50% of the buses are load buses, 20% are generator buses and 30% have both loads and generators connected. In case the number of buses is such that the aforementioned percentages do not result in integers with a sum equal to the number of buses, the remaining buses are considered to be transfer buses (i.e., they do not connect loads or generators). The capacity of each line is sampled from a uniform distribution ranging from 15% to 35% of the total generating capacity of the system. The individual loads
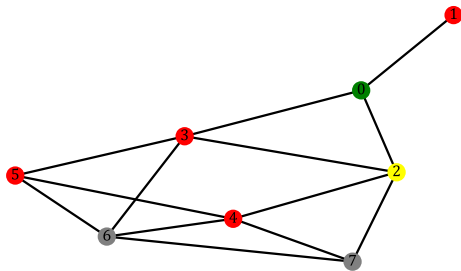
(a) 8-bus power system with 13 transmission lines and 1071 MW generating capacity
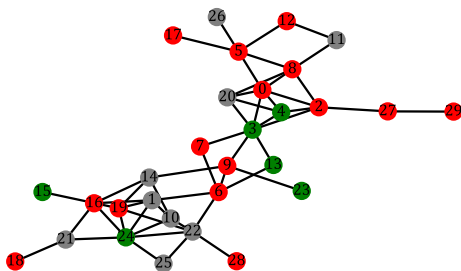
(b) 30-bus power system with 51 transmission lines and 5395 MW generating capacity

Fig. 2. The test power systems. The nodes are color-coded. Red: load bus, Green: generation bus, Gray: load and generation bus, Yellow: transfer bus. Without loss of generality, Bus 0 is defined as the reference bus.
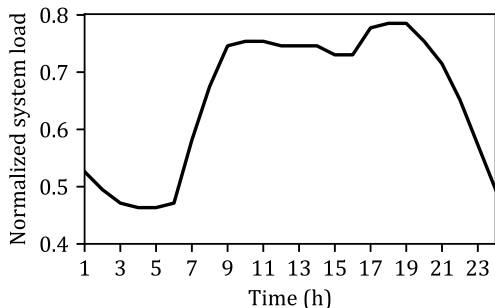

Fig. 3. Normalized system load. Extracted from [55].

are assigned a percentage of the hourly normalized system load (with respect to the maximum generating capacity of the system) that is portrayed in Fig. 3 for 24 time periods using the Dirichlet distribution. The Dirichlet distribution satisfies the requirements that each bus load fraction is positive and that the sum of the fractions is 1. Finally, generator parameters are constructed using the values presented in Table II.

C. Results and Discussion

In this section, the results of the numerical experiments are reported. For all the experiments the algortihm terminates if $\frac{UB-LB}{UB}100\% < 0.5\%$. For the 8-bus system, the whole 24-hour load profile is used (72 MP decision variables are involved in the DSP), while 10 solutions of the MP are

TABLE II
GENERATOR PARAMETERS. RANGES IMPLY SAMPLING A VALUE FROM A UNIFORM DISTRIBUTION

Parameter	Value
P_g^{max}	$[60, 600]$
P_g^{min}	$[20\% P_g^{max}, 40\% P_g^{max}]$
RU_g	$\max(P_g^{min}, [20\% P_g^{max}, 40\% P_g^{max}])$
RD_g	RU_g
SUC_g	$[5, 1600]$
SDC_g	SUC_g
C_g	$[5, 30]$
NLC_g	$[3C_g, 6C_g]$
u_g^{ini}	$\{0, 1\}$
P_g^{ini}	$P_g^{min} u_g^{ini}$

requested from the solver. For the 30-bus system 30 solutions of the MP are requested, however, considering the full 24 periods the heuristic would not result in finding a feasible minor-embedding within the timeout limit of 1000s for all the cases. For this reason, only the first 8 time periods were considered (120 MP decision variables variables involved in the DSP). For both power systems, Strategy II was applied using $\mathcal{M} = 3$ (both for feasibility and optimality cuts). It should be noted that the lowest-energy solution returned by the QPU did not always satisfy the constraints of the cut selection strategy. In these cases, the lowest-energy valid solution was kept. If none of the solutions in the returned sample set were valid, the solution with the lowest number of constraint violations was implemented.

1) Performance comparison of cut selection criteria and strategies: In order to assess the performance of the HQC-MCMS algorithm the cut selection procedure described in Algorithm 2 is executed both on a classical computer and a QA. Detailed computational characteristics for the two test systems are presented in Tables III and IV. The values presented in these tables are averaged over 5 executions of the algorithm for each case. The straightforward implementation of BD and the addition of all the available cuts to the MP are used as benchmarks of the performance of the multi-cut strategies. Moreover, a case in which the optimal solution as well as three randomly selected solutions of the MP are used to generate cuts is presented in order to establish the fact that Algorithm 2 performs a non-trivial cut selection. For ease of reference, the different combinations of cut selection strategies and cut selection criteria are denoted by C1-C12. The time for completing a quantum computation is given by $T_Q \approx T_P + \rho(T_A + T_R)$, where T_Q is the QPU access time, T_P is the quantum programming time, ρ is the number of times an anneal-read cycle is repeated, T_A is the annealing time and T_R is the time required to read a measurement. It is conventional to report ρT_A as an equivalent to the classical CPU time [27], however, T_Q is reported in Tables III and IV for the cut selection component for the sake of completeness.

In Table III it can be seen that all multi-cut strategies result in a reduction in the number of iterations with respect to the straightforward implementation of BD of up to 69.6%. Since the 8-bus power system problem instance is small, adding all the generated cuts at each iteration to the MP results

TABLE III
SOLUTION TIMES FOR THE 8-BUS SYSTEM. ALL TIMES ARE IN (S)

	BD	All cuts	Random	Classical						Quantum					
				C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
cutSelectionStrategy	-	-	-	I			II			I			II		
cutSelectionCriterion	-	-	-	I	II	II	I	II	II	I	II	II	I	II	II
optSelect	-	-	-	-	F	T	-	F	T	-	F	T	-	F	T
Iterations	23.00	7.00	17.60	13.00	7.00	7.00	13.00	9.00	9.00	13.00	7.00	7.00	13.40	10.00	9.80
Time	28.20	8.66	19.10	13.12	9.22	9.24	12.95	10.65	10.78	13.14	9.06	10.33	14.87	17.22	64.62
MP solution	1.74	2.17	3.77	1.77	1.63	1.30	1.71	1.86	1.75	1.76	1.58	1.39	1.89	2.10	1.68
DSP solution	26.46	6.49	15.33	8.36	6.50	6.61	8.34	7.15	7.13	8.37	6.58	6.63	8.54	7.55	7.47
M construction	-	-	-	2.37	0.32	0.42	2.29	0.49	0.54	2.27	0.33	0.40	2.30	0.52	0.58
Cut selection	-	-	-	0.62	0.78	0.92	0.62	1.16	1.35	0.33	0.39	0.50	0.43	0.82	0.94
Minor-emb. (feas.)	-	-	-	-	-	-	-	-	-	0.41	0.17	0.18	1.71	6.23	7.62
Minor-emb. (opt.)	-	-	-	-	-	-	-	-	-	-	-	1.23	-	-	46.34

T: True, F: False

TABLE IV
SOLUTION TIMES FOR THE 30-BUS SYSTEM. ALL TIMES ARE IN (S)

	BD	All cuts	Random	Classical						Quantum					
				C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12
cutSelectionStrategy	-	-	-	I			II			I			II		
cutSelectionCriterion	-	-	-	I	II	II	I	II	II	I	II	II	I	II	II
optSelect	-	-	-	-	F	T	-	F	T	-	F	T	-	F	T
Iterations	51.00	10.00	24.40	9.00	11.00	10.00	9.00	9.00	9.00	9.00	10.60	8.40	9.00	10.20	11.40
Time	110.50	80.56	69.65	53.22	64.73	35.01	52.86	38.11	29.16	56.31	109.92	1348.55	73.14	223.53	1594.90
MP solution	29.00	62.86	37.58	27.74	40.67	12.25	26.51	19.19	8.71	28.40	45.88	18.04	23.33	30.20	12.24
DSP solution	81.51	17.70	32.07	16.12	20.45	17.82	16.10	16.17	15.83	15.95	18.77	19.07	16.29	17.91	23.40
M construction	-	-	-	8.77	1.92	2.56	9.64	1.74	2.31	8.91	1.63	2.11	9.60	1.84	3.04
Cut selection	-	-	-	0.59	1.70	2.38	0.61	1.00	2.30	0.37	1.02	1.91	0.44	1.08	2.61
Minor-emb. (feas.)	-	-	-	-	-	-	-	-	-	2.67	42.62	53.43	23.49	172.51	360.04
Minor-emb. (opt.)	-	-	-	-	-	-	-	-	-	-	-	1253.99	-	-	1193.57

T: True, F: False

in the lowest total execution time, despite the increase in the MP solution time. The reason for this is that all cases in which cut selection is applied, additional components are executed, namely the construction of the indicator matrix $\mathbf{M}$ and the solution of the cut selection problem. For this small problem instance, the significant reduction in the DSP solution time (which in turn depends on the number of iterations) is dominant. Nonetheless, applying any of the cut selection strategies and criteria results in a comparable acceleration in comparison with the straightforward implementation of BD when classical resources are used.

For the 8-bus system, the total cut selection time is by up to 50% lower using QA for cases C7-C12 in comparison with their classical counterparts C1-C6. However, when QA is used, additional time is required in order to map the problem to the QPU topology, which can significantly influence the performance of the overall performance of the HQC-MCMS algorithm. It can be seen that when Strategy I is used for cut selection, significantly less time is required in order to find a minor embedding in comparison with Strategy II. Also, applying cut selection using D^F resulted in faster minor embedding in comparison with D^O. Nonetheless, for cases C7-C11 the HQC-MCMS algorithm performed better than

the straightforward implementation of BD, while C8 also performed better than its classical counterpart C2. It is to be noted that although, on average, the total solution time that is reported in C12 exceeds that of the straightforward implementation of BD, the performance of the minor embedding heuristic is highly variable (standard deviation of 50.31 s) and two instances were found to perform better than BD.

A similar analysis is performed for the 30-bus system based on the results that are presented in Table IV. First, it can be observed that all the multi-cut solution approaches result in a reduction in the number of iterations with respect to the straightforward BD implementation of up to 83.5%. Contrary to the 8-bus system, adding all the available cuts in each iteration to the MP results in a proliferation of the MP solution time that is sufficient to render it a less performant option than the application of any of the cut selection strategies when using a classical solver. Both for Strategies I and II the best performance was observed when Criterion II was used by applying cut selection both on feasibility and optimality cuts (C3 and C6) due to the maximum reduction in the size of the MP. However, the corresponding cases using QA (C9 and C12) were the least performant ones due to the excessive time that was required in order to map the corresponding problems
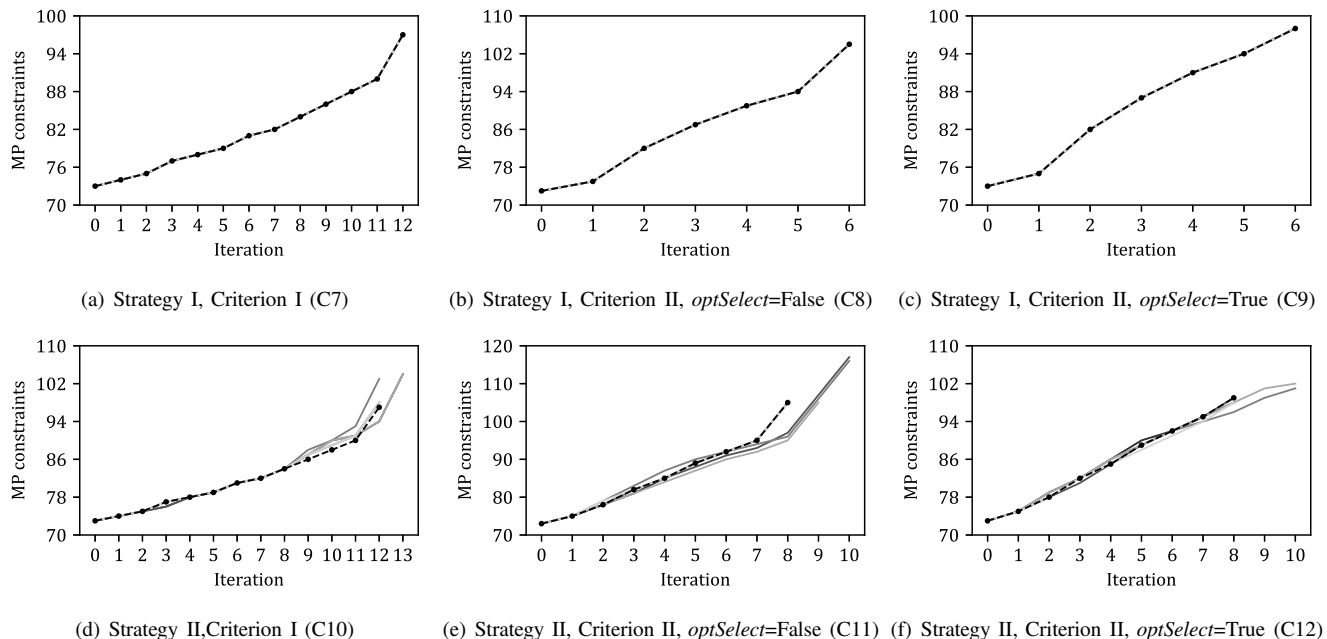
Fig. 4. Number of MP constraints in each iteration of the HQ-MCMS algorithm applied to the 8-bus system for different cut selection strategies and criteria. The trajectories corresponding to each of the 5 runs using a QPU for cut selection are represented by the gray lines. The optimal solution trajectory found using classical optimization is marked with the dashed black line.

to the QPU topology.

Similarly to the results for the 8-bus power system, except for cases C11 and C12, cut selection was faster when QA was used also for the 30-bus power system by up to 40%. The worse performance of cut selection in C11 and C12 can be attributed to the higher average number of iterations that were required for convergence to the specified tolerance in comparison with their classical counterparts (C8 and C9). Despite the significant time that is spent on minor-embedding, C7 and C10 are characterized by a lower solution time in comparison with adding all the available cuts in the MP, while C8 was slightly faster than the straightforward BD implementation.

The results on the 30-bus test system also reveal a significant difference in the time that is required for minor-embedding the logical problem graphs to the QPU topology when different cut selection strategies and cut selection criteria are applied. Minor embedding time exhibits strong dependence on the complexity of the cut selection problem that is solved using QA, as well as on the size of matrix $\mathbf{M}$ after inspection. Strategy II relies on a more complex problem in comparison with Strategy I, which explains why the total minor embedding

is more time consuming in C8 and C9 as well as C11 and C12 in comparison with C7 and C10 respectively. The size of matrix $\mathbf{M}$ is generally larger for optimality cuts in comparison with feasibility cuts. This may be attributed to the different density of feasibility and optimality cuts, which is higher for the latter. For instance, for the worst-performing instance of C9, the lowest average density of feasibility cuts was 7.92% and the highest 12.50%, whereas for the optimality cuts the lowest average density was 35.83% and the highest 55.65%.

Finally, both for the 8-bus and the 30-bus test systems, HQC-MCMS converged after a significantly lower number of iterations compared to the random cut selection benchmark. This is an indication that despite the fact that QA acts as a heuristic, the cut selection results are non-trivial.

2) Quality of the solutions obtained using QA: To evaluate the quality of the solutions that are returned by QA, the increase in the number of constraints of the MP due to the addition of feasibility and optimality cuts across iterations is displayed for the 8-bus and 30-bus test systems in Figs. 4 and 5 respectively.

For the 8-bus system it can be seen that the increase in the number of MP constraints when cut selection problem is solved using QA corresponds to that of the classical solution for cases C7-C9. This is the reason why no significant differences are observed in the MP solution time in comparison with their classical counterparts, while the same number of iterations are performed. On the contrary, differences are observed when Strategy II is employed in combination with any of the cut selection criteria (C10-C12). Although in all three cases there are instances of the quantum step which result in an optimal cut selection trajectory, the average number of iterations is increased due to the sub-optimal trajectories that are generated in some of the runs. It should be noted that

TABLE V
LARGEST QUBO PROBLEM INSTANCE SUBMITTED TO THE QPU FOR
EACH CASE (8-BUS SYSTEM)

	Decision variables	Slack variables	Interactions
C7	5	16	99
C8	7	0	21
C9	4	40	186
C10	15	18	176
C11	28	21	336
C12	48	83	763

(a) Strategy I, Criterion I (C7) (b) Strategy I, Criterion II, *optSelect*=False (C8) (c) Strategy I, Criterion II, *optSelect*=True (C9)

(d) Strategy II,Criterion I (C10) (e) Strategy II, Criterion II, *optSelect*=False (C11) (f) Strategy II, Criterion II, *optSelect*=True (C12)
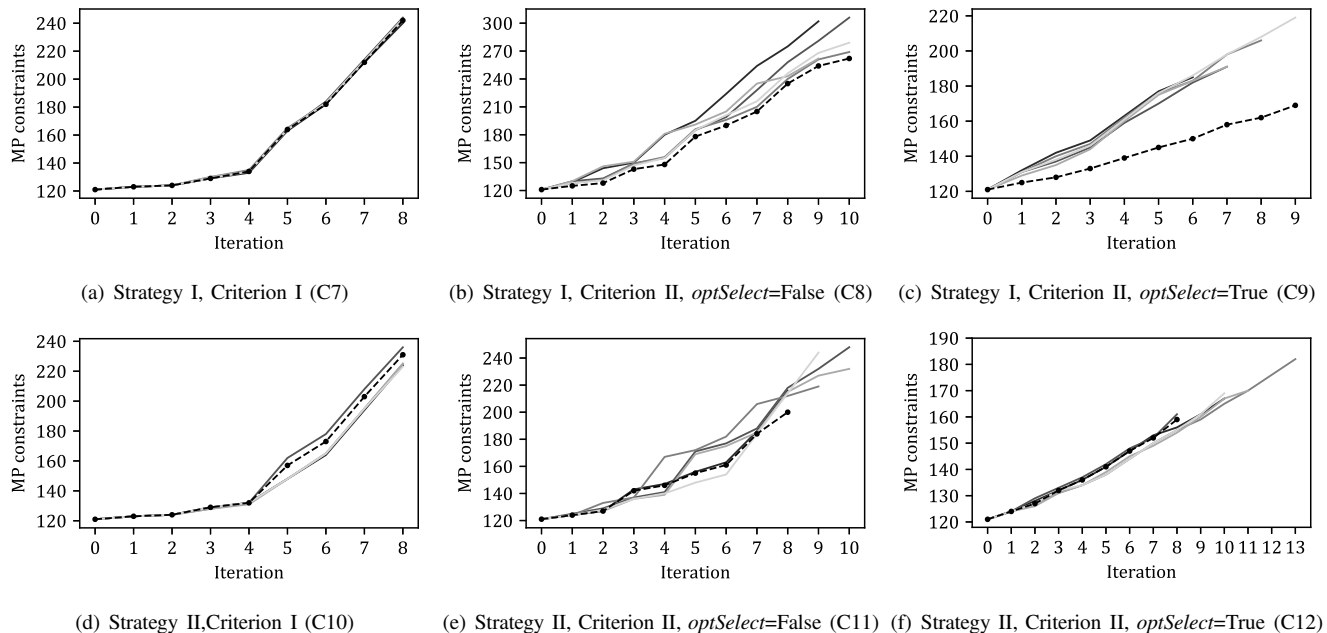
Fig. 5. Number of MP constraints in each iteration of the HQ-MCMS algorithm applied to the 30-bus system for different cut selection strategies and criteria. The trajectories corresponding to each of the 5 runs using a QPU for cut selection are represented by the gray lines. The optimal solution trajectory found using classical optimization is marked with the dashed black line.

TABLE VI
QA SOLUTION DETAILS FOR THE 30-BUS SYSTEM

| | #QPU calls | Minor embedding | | | #Infeas. solutions | Worst case infeasibility | | Largest problem instance | | |
		#Decisions with broken chains	Max. chain break fraction	Representative chain length		#Infeas. constraints	#Constraints	Decision variables	Slack variables	Interactions
C7	15	1	1.69	5	-	-	-	5	54	324
C8	39	1	1.05	7.75	-	-	-	17	78	1559
C9 (feas.)	29	3	1.72	7.25	-	-	-	18	95	1937
C9 (opt.)	27	19	5.09	34.2	2	1	43	30	287	9569
C10	15	0	-	9.33	-	-	-	35	52	494
C11	35	9	4.83	14	4	1	33	69	99	1764
C12 (feas.)	45	12	7.35	13.71	5	2	22	117	149	3307
C12 (opt.)	34	17	6.42	27.87	1	1	28	109	165	7020

#: number

the size of the problems that are submitted to the QPU is relatively small, with the largest instance solved using QA in each case reported in Table V. For this reason, the lowest-energy solutions tend to satisfy the constraints of the cut selection strategies. It is also worth mentioning that for all the problems related to this test system that were submitted to the QPU, only a single decision was made based on a sample with broken logical chains (chain break fraction of 0.76%). The comparable performance of the classical and quantum resources shows that HQC-MCMS is a viable decomposition-based HQC algorithm for small-scale optimization problems.

For the 30-bus test system, the results portrayed in Fig. 5 are indicative of the performance of QA as a heuristic for cut selection when HQC-MCMS is applied to larger-scale optimization problems. For Strategy I, although in C7 the application of QA resulted in the selection of a comparable number of cuts in each iteration with its classical counterpart, using Criterion II results in trajectories that significantly increase the size of

the MP. This is reflected on the increased MP solution time but also on the reduced average number of iterations required for convergence in C8 and C9. For Strategy II, the opposite behavior is observed. With the exception of C10 in which the results are consistent with the optimal results obtained by classical optimization, C11 and C12 are characterized by a higher number of iterations in comparison with their classical counterparts, which is an indication that a larger number of sub-optimal cut choices are made by QA. Nonetheless, despite the cut selection trajectories generated by QA in C9 and C12 departing significantly from the optimal, the cut selection remains effective in terms of MP size management for Criterion II applied both to feasibility and optimality cuts.

To provide further insight into the quality of the QA solutions for the larger 30-bus system, additional details for different cut selection strategies and criteria are presented in Table VI for all the executions of the HQC-MCMS using the QPU. The first observation is that the deterioration of the

solution quality observed in Fig. 5 is related to the size of the QUBO instances that are assigned to the QPU. Furthermore, the size of the problem impacts the qualitative characteristics of minor-embedding. Larger problem instances tend to result in more decisions being made based on samples with broken logical chains and higher maximum chain break fractions. The minor embedding of larger QUBO problems also tends to be characterized by longer qubit chains. A representative chain length is also presented, which is significantly higher when Criterion II is used, in particular when cut selection is applied to optimality cuts. For each minor embedding that is calculated if cut selection is triggered, the maximum qubit chain length is recorded and the results are averaged for each of the five runs. Their maximum across the five runs is considered as a representative chain length. Notably, QA managed to discover valid solutions in the majority of the cases, whereas in the few cases where a cut selection was made based on an infeasible solution, only a few of the problem constraints were not satisfied. Summarizing Table VI, Strategy I is associated with more favorable solution characteristics than Strategy II. The same can be said about Criterion I in comparison with Criterion II, especially when cut selection is applied to optimality cuts.

3) Practical limitations: Based on the computational experience with the UC problem, three limitations of the proposed HQC algorithm can be identified. First, although Criterion II appears to be computationally advantageous in comparison with Criterion I when classical computing resources are used, the dependence of the size of matrix $\mathbf{M}$ on the number of complicating variables imposes a limit on the size of the problem that can be embedded on the QPU. In addition to that, significant time is consumed on minor-embedding, especially when cut selection concerns optimality cuts. On the contrary, the size of matrix $\mathbf{M}$ under Criterion I depends solely on the number of MP solutions which is a user-selected parameter. Since this parameter is independent of the size of the optimization problem, Criterion I may find wider applicability to large-scale optimization problems when using a QPU to perform cut selection in comparison with Criterion II, given the limitations of NISQ hardware.

Another shortcoming of HQC-MCMS is that minor embedding has to be repeated in each iteration where the QPU is called because a different matrix $\mathbf{M}$ is available. Although performing cut selection using quantum resources may be a computationally efficient procedure itself, the impact of the time that is required in order to map the problem can adversely impact the overall performance of the algorithm. To overcome this limitation, either more efficient minor-embedding heuristics or a systematic way to exploit previously generated minor embeddings need to be developed.

The third challenge is related to finding suitable values for the hyperparameters (chain strength, annealing schedule) and weights for the quadratic penalty functions in order to guarantee that the lowest-energy solution that is found corresponds to a valid solution with high probability and reduce logical chain breakage.

VI. CONCLUSION

In this paper, a hybrid quantum-classical (HQC) multi-cut Benders decomposition (BD) strategy that exploits multiple feasible solutions of the master problem (MP) in order to generate multiple feasibility and optimality cuts was presented. Adding multiple cuts to the MP improves the convergence rate of BD. However, the increase in the size of the MP may adversely impact solution time. In order to manage the size of the MP and exploit the availability of multiple cuts, a cut selection procedure that can be assigned to a quantum computer was developed. Two different criteria and two different cut selection strategies based on pure binary problems that can be solved using quantum computing were studied. The HQC algorithm was applied to the Unit Commitment problem and computational experiments were conducted using the D-Wave Advantage 4.1 quantum annealer. Results on two test power systems showed that although it is viable for quantum resources to be used as an alternative to classical resources for cut selection for small-scale problems, current hardware limitations must be overcome and the efficiency of minor-embedding techniques should be improved before effectively applying the proposed approach to large-scale problem instances. Future research will focus on further improving the proposed HQC algorithm according to the limitations that were identified in Section V-C3 and applying it on different use cases.

REFERENCES

[1] J. Preskill, "Quantum computing in the NISQ era and beyond," *Quantum*, vol. 2, p. 79, Aug. 2018.

[2] A. Lucas, "Ising formulations of many NP problems," *Front. Phys.*, vol. 2, Feb. 2014.

[3] G. Kochenberger, J.-K. Hao, F. Glover, M. Lewis, Z. Lü, H. Wang, and Y. Wang, "The unconstrained binary quadratic programming problem: a survey," *J. Comb. Optim.*, vol. 28, pp. 58–81, July 2014.

[4] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O'Brien, "A variational eigenvalue solver on a photonic quantum processor," *Nat. Commun.*, vol. 5, 2014, Art. no. 4213.

[5] E. Farhi, J. Goldstone, and S. Gutmann, "A quantum approximate optimization algorithm," *arXiv:1411.4028*, 2014.

[6] A. Gilliam, S. Woerner, and C. Gonciulea, "Grover adaptive search for constrained polynomial binary optimization," *Quantum*, vol. 5, p. 428, Apr. 2021.

[7] G. Nannicini, "Performance of hybrid quantum-classical variational heuristics for combinatorial optimization," *Phys. Rev. E*, vol. 99, Jan. 2019, Art. no. 013304.

[8] C. C. McGeoch, "Theory versus practice in annealing-based quantum computing," *Theor. Comput. Sci.*, vol. 816, pp. 169–183, May 2020.

[9] E. Grant, T. S. Humble, and B. Stump, "Benchmarking quantum annealing controls with portfolio optimization," *Phys. Rev. Applied*, vol. 15, Jan. 2021, Art. no. 014012.

[10] S. Harwood, C. Gambella, D. Trenev, A. Simonetto, D. Bernal, and D. Greenberg, "Formulating and solving routing problems on quantum computers," *IEEE Trans. Quantum Eng.*, vol. 2, pp. 1–17, 2021, Art no. 3100118.

[11] T. Stollenwerk, E. Lobe, and M. Jung, "Flight gate assignment with a quantum annealer," in *Quantum Technology and Optimization Problems*, S. Feld and C. Linnhoff-Popien, Eds. Cham: Springer International Publishing, 2019, pp. 99–110.

[12] D. Anghinolfi, M. Paolucci, and M. Robba, "Optimal planning of door-to-door multiple materials separated waste collection," *IEEE Trans. Autom. Sci. Eng.*, vol. 13, no. 4, pp. 1448–1457, Oct. 2016.

[13] J. Yang, X. Xu, D. Yin, Z. Ma, and L. Shen, "A space mapping based 0–1 linear model for onboard conflict resolution of heterogeneous unmanned aerial vehicles," *IEEE Trans. Veh. Technol.*, vol. 68, no. 8, pp. 7455–7465, Aug. 2019.

[14] Y. Chen, F. Liu, B. Liu, W. Wei, and S. Mei, "An efficient MILP approximation for the hydro-thermal unit commitment," *IEEE Trans. Power Syst.*, vol. 31, no. 4, pp. 3318–3319, July 2016.

[15] C. Gambella and A. Simonetto, "Multiblock ADMM heuristics for mixed-binary optimization on classical and quantum computers," *IEEE Trans. Quantum Eng.*, vol. 1, pp. 1–22, 2020, Art no. 3102022.

[16] C.-Y. Chang, E. Jones, and P. Graf, "On quantum computing for mixed-integer programming," *arXiv:2010.07852*, 2020.

[17] A. Ajagekar, T. Humble, and F. You, "Quantum computing based hybrid solution strategies for large-scale discrete-continuous optimization problems," *Comput. Chem. Eng.*, vol. 132, Jan. 2020, Art no. 106630.

[18] L. Braine, D. J. Egger, J. Glick, and S. Woerner, "Quantum algorithms for mixed binary optimization applied to transaction settlement," *IEEE Trans. Quantum Eng.*, vol. 2, pp. 1–8, 2021, Art no. 3101208.

[19] J. Lin, F. Magnago, and J. M. Alemany, "Chapter 1 - optimization methods applied to power systems: Current practices and challenges," in *Classical and Recent Aspects of Power System Optimization*, A. F. Zobaa, S. H. Abdel Aleem, and A. Y. Abdelaziz, Eds. Academic Press, 2018, pp. 1–18.

[20] J. Tovar-Facio, M. Martín, and J. M. Ponce-Ortega, "Sustainable energy transition: modeling and optimization," *Curr. Opin. Chem. Eng.*, vol. 31, Mar. 2021, Art no. 100661.

[21] S. Potrč, L. Čuček, M. Martin, and Z. Kravanja, "Sustainable renewable energy supply networks optimization – The gradual transition to a renewable energy system within the European Union by 2050," *Renewable Sustainable Energy Rev.*, vol. 146, Aug. 2021, Art no 111186.

[22] Z. Tang, Y. Qin, Z. Jiang, W. O. Krawec, and P. Zhang, "Quantum-secure microgrid," *IEEE Trans. Power Syst.*, vol. 36, no. 2, pp. 1250–1263, Mar. 2021.

[23] R. Eskandarpour, K. J. Bahadur Ghosh, A. Khodaei, A. Paaso, and L. Zhang, "Quantum-enhanced grid of the future: A primer," *IEEE Access*, vol. 8, pp. 188 993–189 002, Oct. 2020.

[24] O. O. Olatunji, P. A. Adedeji, and N. Madushele, "Chapter 22 - quantum computing in renewable energy exploration: status, opportunities, and challenges," in *Design, Analysis, and Applications of Renewable Energy Systems*, ser. Advances in Nonlinear Dynamics and Chaos (ANDC), A. T. Azar and N. A. Kamal, Eds. Academic Press, 2021, pp. 549–572.

[25] A. Giani and Z. Eldredge, "Quantum computing opportunities in renewable energy," *SN Comput. Sci.*, vol. 2, 2021, Art no 393.

[26] N. Elsayed, A. S. Maida, and M. Bayoumi, "A review of quantum computer energy efficiency," in *Proc. 2019 IEEE Green Technologies Conference (GreenTech)*, Lafayette, LA, USA, Apr. 3-6, 2019, pp. 1–3.

[27] E. B. Jones, E. Kapit, C.-Y. Chang, D. Biagioni, D. Vaidhynathan, P. Graf, and W. Jones, "On the computational viability of quantum optimization for PMU placement," in *Proc. 2020 IEEE Power & Energy Society General Meeting (PESGM)*, Montreal, QC, Canada, Aug. 2-6, 2020, pp. 1–5.

[28] K. Loh, "Fairness evaluation during the conceptual design of heat grids with quantum annealers," *arXiv: 1910.09929*, 2019.

[29] A. Ajagekar and F. You, "Quantum computing for energy systems optimization: Challenges and opportunities," *Energy*, vol. 179, pp. 76–89, July 2019.

[30] F. Feng, Y. Zhou, and P. Zhang, "Quantum power flow," *IEEE Trans. Power Syst.*, vol. 36, no. 4, pp. 3810–3812, July 2021.

[31] R. Eskandarpour, P. Gokhale, A. Khodaei, F. T. Chong, A. Passo, and S. Bahramirad, "Quantum computing for enhancing grid security," *IEEE Trans. Power Syst.*, vol. 35, no. 5, pp. 4135–4137, Sep. 2020.

[32] A. Ajagekar and F. You, "Quantum computing based hybrid deep learning for fault diagnosis in electrical power systems," *Appl. Energy*, vol. 303, Dec. 2021, Art no 117628.

[33] M. W. Johnson, M. H. S. Amin, S. Gildert, T. Lanting, F. Hamze, N. Dickson, R. Harris, A. J. Berkley, J. Johansson, P. Bunyk, E. M. Chapple, C. Enderud, J. P. Hilton, K. Karimi, E. Ladizinsky, N. Ladizinsky, T. Oh, I. Perminov, C. Rich, M. C. Thom, E. Tolkacheva, C. J. S. Truncik, S. Uchaikin, J. Wang, B. Wilson, and G. Rose, "Quantum annealing with manufactured spins," *Nature*, vol. 473, p. 194–198, 2011.

[34] A. M. Childs, E. Farhi, and J. Preskill, "Robustness of adiabatic quantum computation," *Phys. Rev. A*, vol. 65, Dec. 2001, Art no 012322.

[35] E. Farhi, J. Goldstone, S. Gutmann, and M. Sipser, "Quantum computation by adiabatic evolution," *arXiv:quant-ph/0001106*, 2000.

[36] J. Cai, W. G. Macready, and A. Roy, "A practical heuristic for finding graph minors," *arXiv: 1406.2741*, 2014.

[37] A. J. Conejo, E. Castillo, R. Minguez, and R. Garcia-Bertrand, *Decomposition Techniques in Mathematical Programming*. Springer-Verlag Berlin Heidelberg, 2006.

[38] R. Rahmaniani, T. G. Crainic, M. Gendreau, and W. Rei, "The Benders decomposition algorithm: A literature review," *Eur. J. Oper. Res.*, vol. 259, no. 3, pp. 801–817, June 2017.

[39] N. B. Asl and S. A. MirHassani, "Accelerating benders decomposition: multiple cuts via multiple solutions," *J. Comb. Optim.*, vol. 37, pp. 806–826, 2019.

[40] G. K. D. Saharidis, M. Minoux, and M. Ierapetritou, "Accelerating Benders method using covering cut bundle generation," *Int. Tran. Oper. Res.*, vol. 17, pp. 221–237, Mar. 2010.

[41] L. Tang, W. Jiang, and G. K. D. Saharidis, "An improved Benders decomposition algorithm for the logistics facility location problem with capacity expansions," *Ann. Oper. Res.*, vol. 210, pp. 165–190, 2013.

[42] T. Grossman and A. Wool, "Computational experience with approximation algorithms for the set covering problem," *Eur. J. Oper. Res.*, vol. 101, no. 1, pp. 81–92, Aug. 1997.

[43] K. Tamura, T. Shirai, H. Katsura, S. Tanaka, and N. Togawa, "Performance comparison of typical binary-integer encodings in an ising machine," *IEEE Access*, vol. 9, pp. 81 032–81 039, 2021.

[44] S. Takabe, T. Maehara, and K. Hukushima, "Typical approximation performance for maximum coverage problem," *Phys. Rev. E*, vol. 97, Feb. 2018, Art no 022138.

[45] A. Geoffrion and G. Graves, "Multicommodity distribution system design by benders decomposition," *Manage. Sci.*, vol. 20, pp. 822–844, Jan. 1974.

[46] Q. P. Zheng, J. Wang, and A. L. Liu, "Stochastic optimization for unit commitment—A review," *IEEE Trans. Power Syst.*, vol. 30, no. 4, pp. 1913–1924, July 2015.

[47] Y. Wen, C. Guo, H. Pandžić, and D. S. Kirschen, "Enhanced security-constrained unit commitment with emerging utility-scale energy storage," *IEEE Trans. Power Syst.*, vol. 31, no. 1, pp. 652–662, Jan. 2016.

[48] A. Nasri, S. J. Kazempour, A. J. Conejo, and M. Ghandhari, "Network-constrained AC unit commitment under uncertainty: A Benders' decomposition approach," *IEEE Trans. Power Syst.*, vol. 31, no. 1, pp. 412–422, Jan. 2016.

[49] J. M. Alemany, F. Magnago, and D. Moitre, "Benders decomposition applied to security constrained unit commitment," *IEEE Lat. Am. Trans.*, vol. 11, no. 1, pp. 421–425, Feb. 2013.

[50] Y. Fu, Z. Li, and L. Wu, "Modeling and solution of the large-scale security-constrained unit commitment," *IEEE Trans. Power Syst.*, vol. 28, no. 4, pp. 3524–3533, 2013.

[51] Y. Fu, M. Shahidehpour, and Z. Li, "Security-constrained unit commitment with AC constraints," *IEEE Trans. Power Syst.*, vol. 20, no. 2, pp. 1001–1013, May 2005.

[52] M. L. Bynum, G. A. Hackebeil, W. E. Hart, C. D. Laird, B. L. Nicholson, J. D. Siirola, J.-P. Watson, and D. L. Woodruff, *Pyomo–optimization modeling in python*, 3rd ed. Springer Science & Business Media, 2021, vol. 67.

[53] Gurobi Optimization, LLC, *Gurobi Optimizer Reference Manual*, 2021. [Online]. Available: https://www.gurobi.com

[54] *QPU-Specific Physical Properties: Advantage_system4.1*, D-Wave User Manual 09-1262A-A, D-Wave Systems Inc., 2021.

[55] C. Ordoudis, P. Pinson, J. Morales González, and M. Zugno, "An updated version of the IEEE RTS 24-bus system for electricity market and power system operation studies," Technical University of Denmark, Tech. Rep., 2016.

[56] Z. Wang, R. J. Thomas, and A. Scaglione, "Generating random topology power grids," in *Proc. 41st Annual Hawaii International Conference on System Sciences (HICSS 2008)*, Waikoloa, HI, USA, Jan.7-10, 2008, pp. 183–183.