Defense Presentation

Hybrid Quantum-Classical Optimization for Data Center Energy System

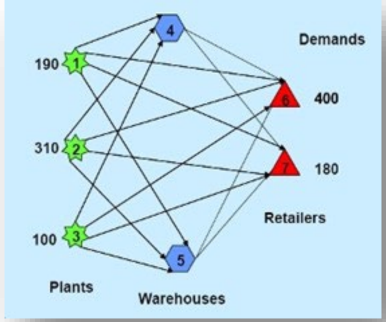
Ph.D. Candidate: Zhongqi Zhao
Date: 07/18/2025
Academic Advisor: Dr. Lei Fan and Dr. Zhu Han
Committee Member:
Dr. Miao Pan, Dr. David Mayerich, and Dr. Xiaodi Wu

Outline

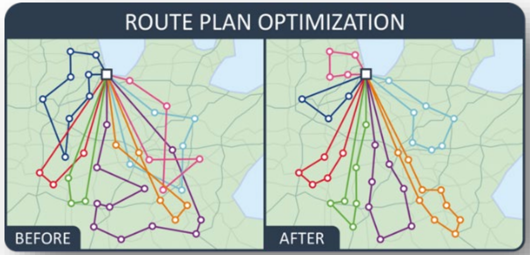
- **Introduction**
  - Mixed-integer Linear Programming & Quantum Computing
- Work 1: Hybrid Quantum Benders' Decomposition (HQC-Bend) for Mixed-integer Linear Programming and Python Package
- Work 2: Energy Management Problem in Internet Data Center Using HQC-Bend
- Work 3: Optimal Energy and LLM Training Job Scheduling for Internet Data Center Using Nonlinear HQC-Bend.
- Future Work & Conclusion

Introduction (1/7) – MILP Application

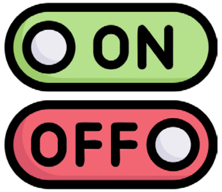
- (Mixed)-integer Linear Programming



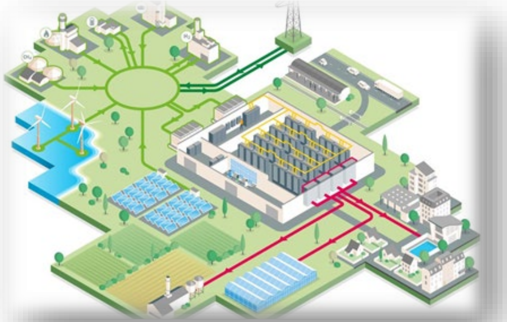
Production & Demand.



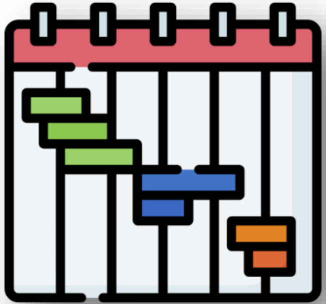
Route Planning



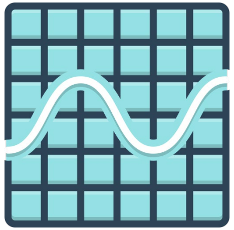
State



Energy Management

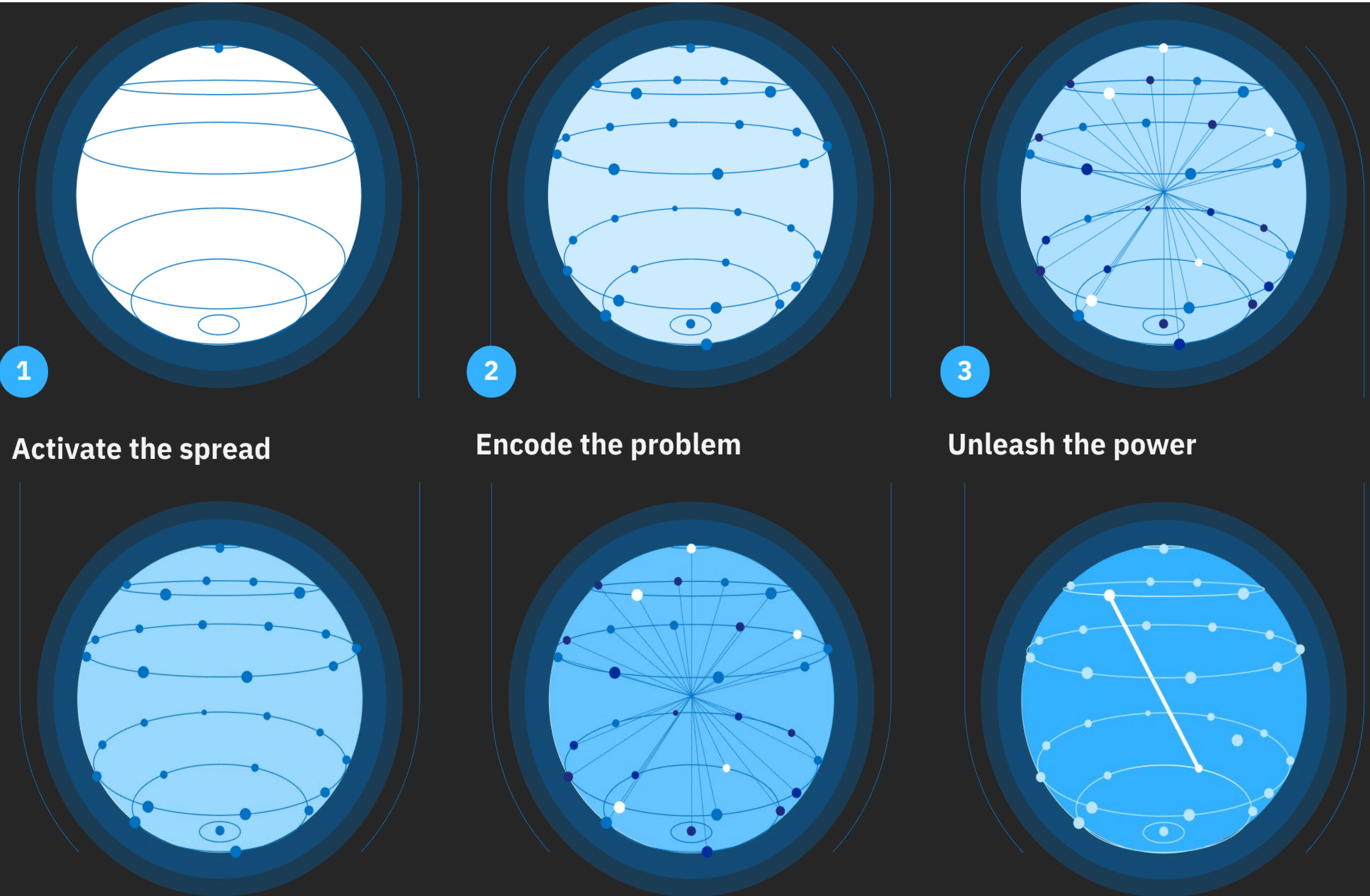


LLM Task Scheduling



Amplitude

Introduction (2/7) – Quantum Computing (QC)



Quantum computers
CAN

create vast multidimensional spaces
in very LARGE problems.

translate them back into
what we CAN use And understand

Classical computers
IS HARD

to achieve the same.

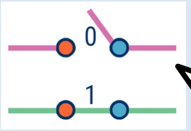
Introduction (3/7) – Quantum Computing (QC)

- What is Quantum Computing (QC):

harnesses the principles of quantum mechanics—superposition, entanglement, and interference—to process information in **parallel** and **probabilistic** ways, enabling solutions to complex problems.

- A Toy Example: "Mystery Coin" is Heads or Tails $f(x)$

Classical Computing

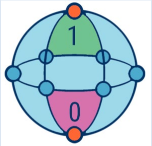


Done with transistors, which can be 0 or 1.

First $f(0)$, then $f(1)$. → $f(x)$ → Same → C; Differ → B

Constant Or Balanced?

Quantum Computing



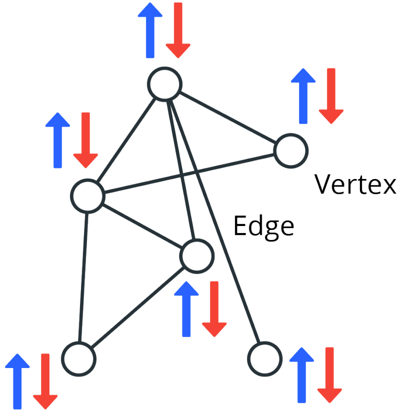
Done with superposition $|x\rangle = {}^{1}/_{\sqrt{2}}(|0\rangle + |1\rangle)$

One Query

$|x\rangle$ — H — $U_{f(x)}$ — H — $|0\rangle \rightarrow$ C; $|1\rangle \rightarrow$ B

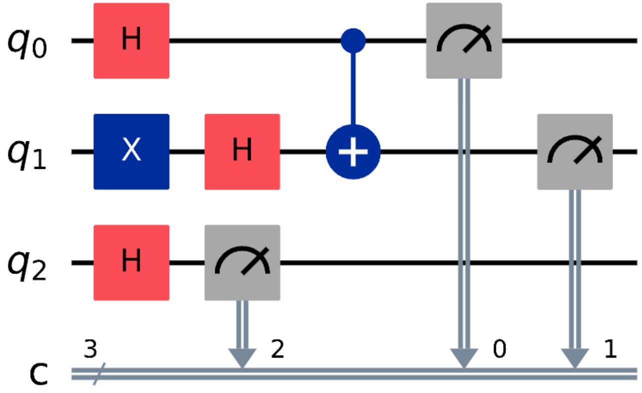
Introduction (4/7) – QC vs. Classical Computing

? ?	Advantage	Disadvantage
CPU (Central Processing Unit)	• **Versatility to Classical Algorithms**	• **Limited Parallelism -> Slower***
QPU (Quantum Processing Unit)	• **Quantum Parallelism/Tunneling**	• **Limited & Specialized Cases**



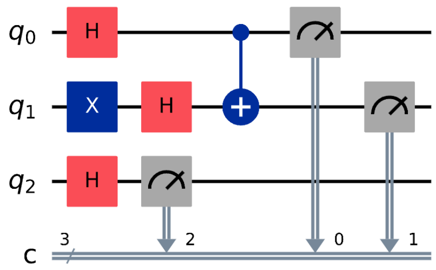
$\Longleftarrow$ Quantum Annealing (QA)

(QDC) Digital Quantum Circuit $\Longrightarrow$



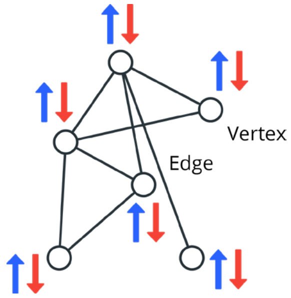
Introduction (5/7) – QA vs. DQC (MFG)

Digital Quantum Circuit



433
IBM Q
Osprey

Quantum Annealing



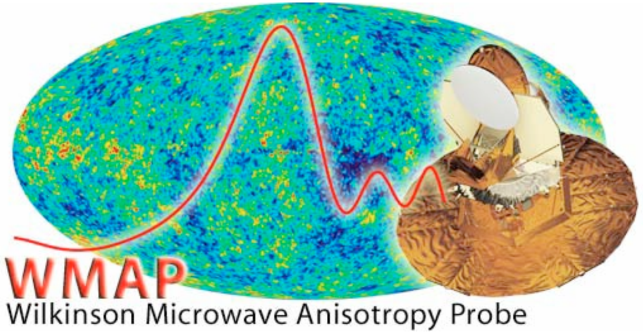
5760
D-Wave
The Quantum Computing Company™
Advanced

- Continuous/Discrete (VQA, QAOA)*

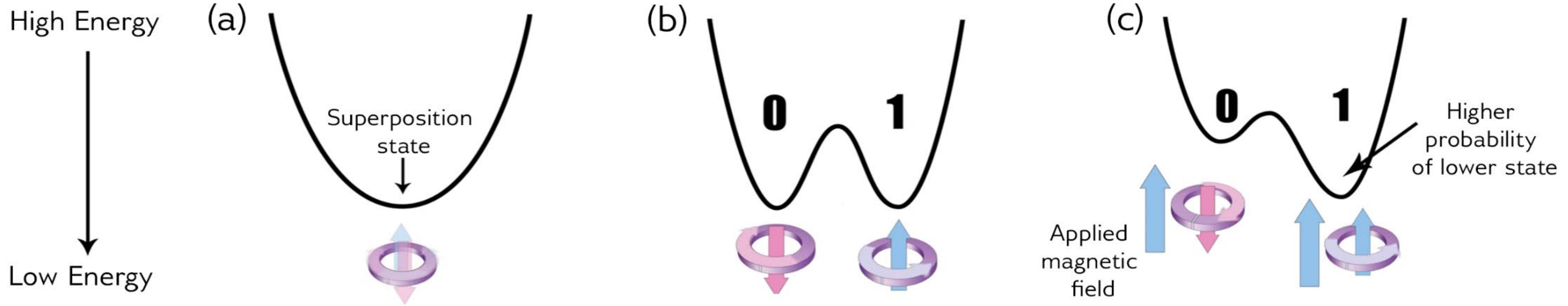
⚠ : Better performance is not Guaranteed.

- Discrete (QUBO, Ising)

⚠ : Need to translate/reformulate problem



Introduction (6/7) – QA vs. DQC (MFG)



A matter of microseconds.

5760
D-Wave The Quantum Computing Company
Advanced

- Initial Qubits
- Superposition at $|0\rangle$s and $|1\rangle$s.
- Not yet coupled.

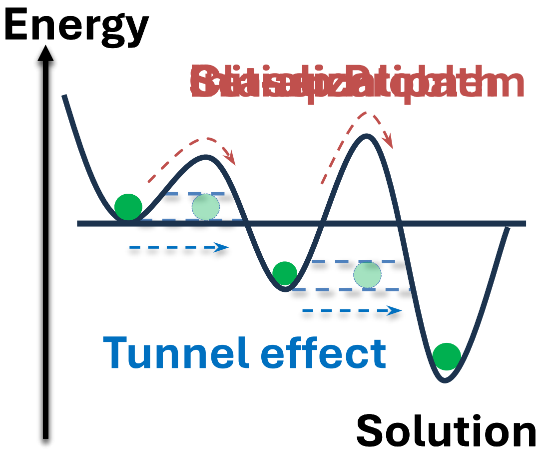
- Qubits are entangled.
- At state of many possible answers.
- Couplers & biases applied

- Inputs' energy are set.
- Lowest energy is at or closes to the optima.
- Energy → possibility

Introduction (7/7) – QA in D-WAVE

Core idea: Encoding the objective function as the eigenvalue of the final ground state of Schrodinger equation, based on Adiabatic Quantum Computing Model

$$\boldsymbol{\sigma} \in \{-1,1\}$$

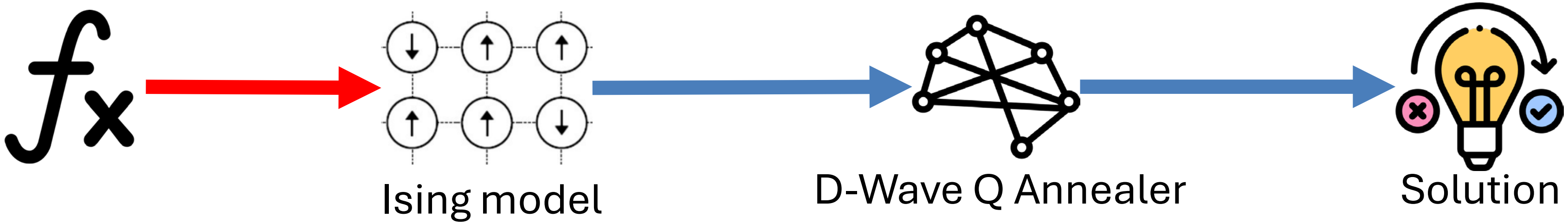


$$H_{\text{ising}} = \underbrace{-\frac{A(s)}{2}\left(\sum_i \hat{\sigma}_x^{(i)}\right)}_{\text{Initial Hamiltonian}}$$

Spins interact with applied field

$$+ \underbrace{\frac{B(s)}{2}\left(\sum_i h_i \hat{\sigma}_z^{(i)} + \sum_{i>j} J_{i,j}\hat{\sigma}_z^{(i)}\hat{\sigma}_z^{(j)}\right)}_{\text{Final Hamiltonian}}$$

Neighboring spins interact with each other



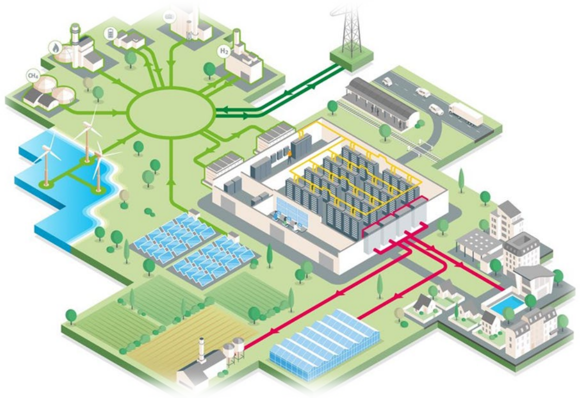
Outline

- Introduction
- **Work 1: Hybrid Quantum Benders' Decomposition (HQC-Bend) for Mixed-integer Linear Programming and Python Package**
- Work 2: Energy Management Problem in Internet Data Center Using HQC-Bend
- Work 3: Optimal Energy and LLM Training Job Scheduling for Internet Data Center Using Nonlinear HQC-Bend.
- Future Work & Conclusion

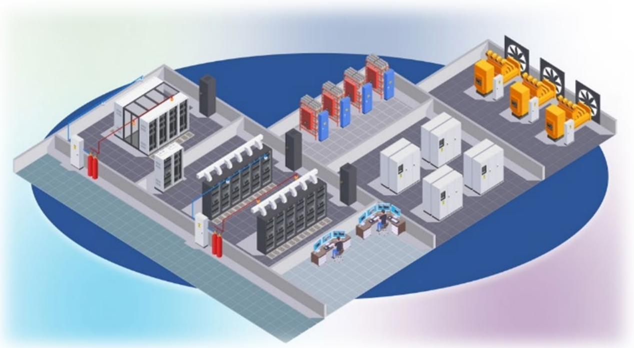
Work I: HQC-Bend for MILP & Software

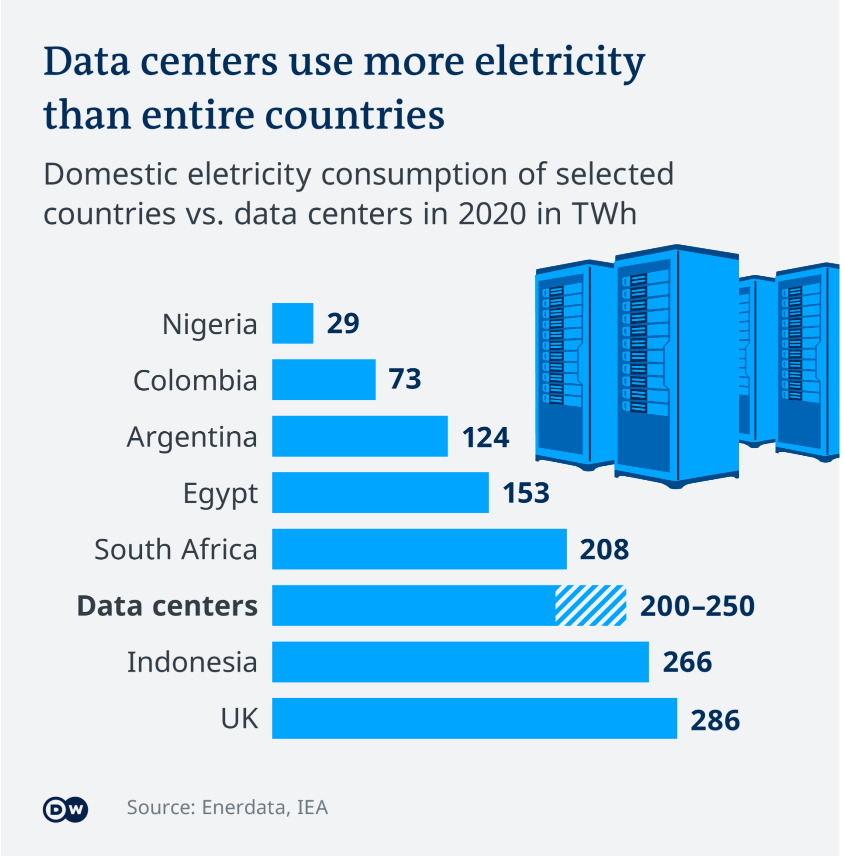
☐ Motivation: Internet Data Center (IDC) Energy Management is Vital!

Renewable energy resources is many



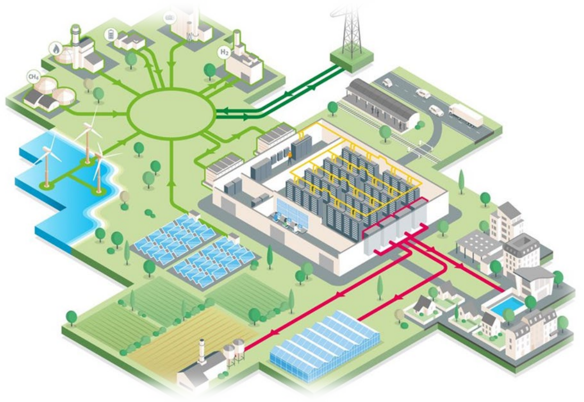
IDC server room devices are complex



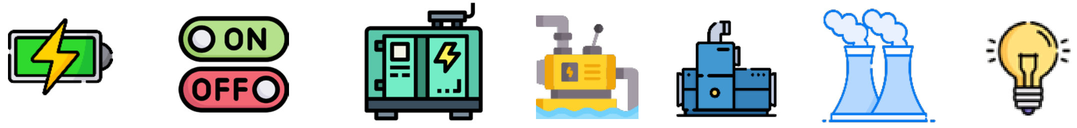


Work 1: HQC-Bend for MILP

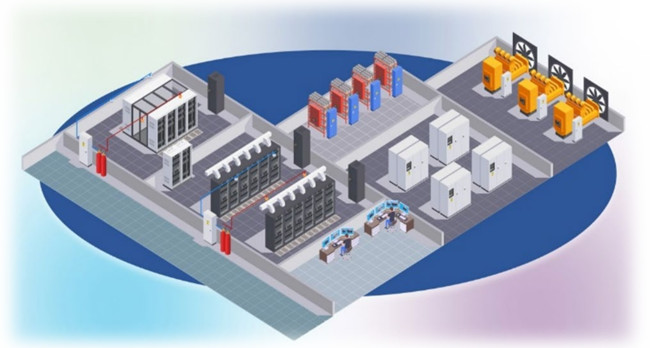
- Motivation: Internet Data Center (IDC) Energy Management is Vital!



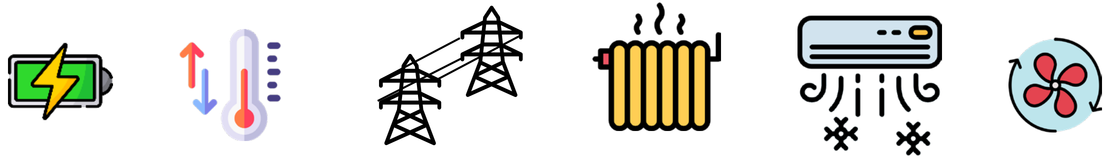
Binary variables: $x \in \mathbb{B}$



...



Continuous variables: $y \in \mathbb{R}$



...

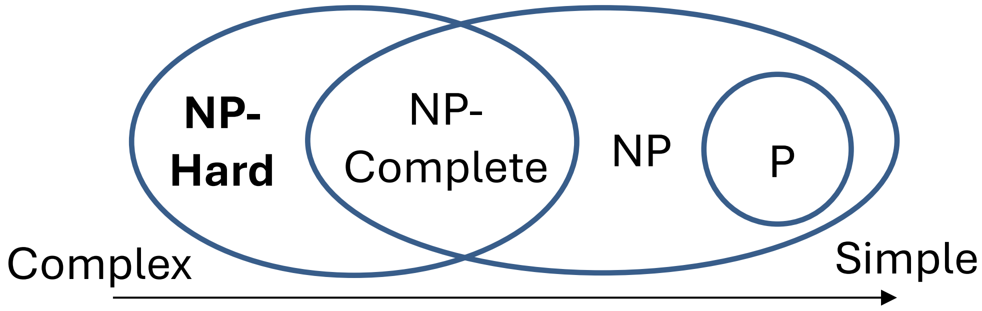
Work I: HQC-Bend for MILP

❑ (Mixed)-integer linear programming

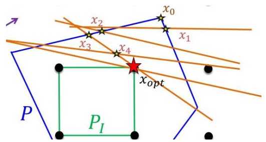
$$
\begin{aligned}
\max_{\mathbf{x},\mathbf{y}} \quad & \mathbf{c}^\intercal \mathbf{x} + \mathbf{h}^\intercal \mathbf{y} \\
\text{s.t.} \quad & \mathbf{T}\mathbf{x} \leq \mathbf{p} \\
& \mathbf{A}\mathbf{x} + \mathbf{G}\mathbf{y} \leq \mathbf{b} \\
& \mathbf{x} \in \mathbb{Z}^n, \mathbf{y} \in \mathbb{R}^m
\end{aligned}
$$

$$
\begin{aligned}
\max_{\mathbf{x},\mathbf{y}} \quad & \mathbf{c}^\intercal \mathbf{x} \\
\text{s.t.} \quad & \mathbf{A}\mathbf{x} \leq \mathbf{b} \\
& \mathbf{x} \in \mathbb{Z}^n.
\end{aligned}
$$

- MILP is **NP-Hard**,
- It **can't** be solved in polynomial time*.



Problem type	Example Problem
NP-Hard	Turing Halting Problem **(M)ILP**
NP-Complete	Graph 3-coloring
NP	Factoring
P	Graph Connectivity

- Enumerative Methods 
- Branch and Bounds 
- Cutting Planes Methods 

Work 1: HQC-Bend for MILP

Consider a Mixed-Binary LP on right-hand-side,

Classical Benders' Decomposition Algorithm is:

1) Solve the master problem (**MAP**). Obtain solution $\bar{x}$ and $\bar{\lambda}$.
2) Determine $\underline{\lambda}$ by solving the dual of the subproblem (**DSUB**).
3) If DSUB is unbounded. Add the corresponding feasibility cuts to MAP and return to Step 1. (**Feasibility Cuts**).
4) If the DSUB objective value $< \bar{\lambda}$ and finite, Add the Optimality Cuts to MAP and return to Step 1. (**Optimality Cuts**)
5) If $f(|\bar{\lambda} - \underline{\lambda}|) \leq \tau$. then we recognize the current $\bar{x}$ solves the original mixed integer program, with optimal y equal to the solution to the primal subproblem with $x = \bar{x}$.

$$\begin{aligned} \max_{\mathbf{x},\mathbf{y}} \ & \mathbf{c}^\intercal \mathbf{x} + \mathbf{h}^\intercal \mathbf{y} \\ \text{s.t. } & \mathbf{T}\mathbf{x} \leqq \mathbf{p} \\ & \mathbf{A}\mathbf{x} + \mathbf{G}\mathbf{y} \leqq \mathbf{b} \\ & \mathbf{x} \in \mathbb{Z}^n, \mathbf{y} \in \mathbb{R}^m \end{aligned}$$

$$\begin{aligned} \max_{\mathbf{x},\lambda} \ & \mathbf{c}^\intercal \mathbf{x} + \lambda \\ \text{s.t. } & \mathbf{T}\mathbf{x} \leqq \mathbf{p} \\ & (\mathbf{b} - \mathbf{A}\mathbf{x})^\intercal u^k \geq \lambda \quad \text{for } k \in K \\ & (\mathbf{b} - \mathbf{A}\mathbf{x})^\intercal r^j \geq 0 \quad \text{for } j \in J \\ & \lambda \in \mathbb{R}, \mathbf{x} \in \mathbb{B}^n \end{aligned}$$

Master Problem

Feasibility Cuts

or

Optimality Cuts

Binary solution $\mathbf{x}$

$$\begin{aligned} \min_{\mathbf{u}} \ & (\mathbf{b} - \mathbf{A}\mathbf{x})^\intercal \mathbf{u} \\ \text{s.t. } & \mathbf{G}^\intercal \mathbf{u} \geq \mathbf{h} \\ & \mathbf{u} \in \mathbb{R}^m_+ \end{aligned}$$

DSUB

Work 1: HQC-Bend for MILP

Algorithm 1 Hybrid Quantum-Classical Benders' Decomposition Algorithm

Require: Initial sets $\hat{K}$ of extreme points and $\hat{J}$ of extreme rays of Q
  $\overline{\lambda} \leftarrow +\infty$
  $\underline{\lambda} \leftarrow -\infty$
  while $|\overline{\lambda} - \underline{\lambda}| \geq \epsilon$ **do**
    $\mathbf{P} \leftarrow$ Appropriate penalties numbers or arrays
    $\mathbf{Q} \leftarrow$ Reformulate both objective and constraints and construct the QUBO formulation by using corresponding rules
    $\mathbf{x}' \leftarrow$ Solve MAP by quantum computer.
    $\overline{\lambda} \leftarrow$ Extract $\mathbf{w}$ and replace the $\overline{\lambda}$ with $\overline{\lambda}(\mathbf{w})$
    $z_{LP}(\mathbf{x}) \leftarrow$ Solve the DSUB problem
    $\underline{\lambda} \leftarrow z_{LP}(\mathbf{x})$
    if $z_{LP}(\mathbf{x}) = -\infty$ **then**
      An extreme ray j of Q is found (Feasibility Cut).
      $\hat{J} = \hat{J} \cup \{j\}$
    else if $z_{LP}(\mathbf{x}) < \overline{\lambda}$ **and** $\overline{\lambda} \neq +\infty$ **then**
      An extreme point k of Q is found (Optimality Cut).
      $\hat{K} = \hat{K} \cup \{k\}$
    end if
  end while
  return $\overline{\lambda}$, $\mathbf{x}$

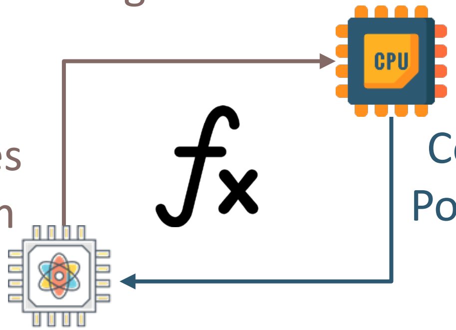
$$f(\mathbf{x}') = \mathbf{x}'^{\top}\mathbf{Q}_{\text{QUBO}}\mathbf{x}'$$

Obj. $\quad \mathbf{x}^{\top}\operatorname{diag}(\mathbf{c})\mathbf{x}$

Connection Var. $\quad + \sum_{i=-\underline{m}}^{\overline{m}_+} w_{i+\underline{m}} 2^i w_{i+\underline{m}} - \sum_{j=0}^{\overline{m}_-} w_{j+(1+\underline{m}+\overline{m}_+)} 2^j w_{j+(1+\underline{m}+\overline{m}_+)}$

Opt. Cut $\quad + \sum_{k\in K} P_k \left(\overline{\lambda}(\mathbf{w}) + (u^k)^{\top}\mathbf{A}\mathbf{x} + \sum_{l=0}^{\bar{l}^K} 2^l s^K_{kl} - b^{\top}u^k \right)^2$

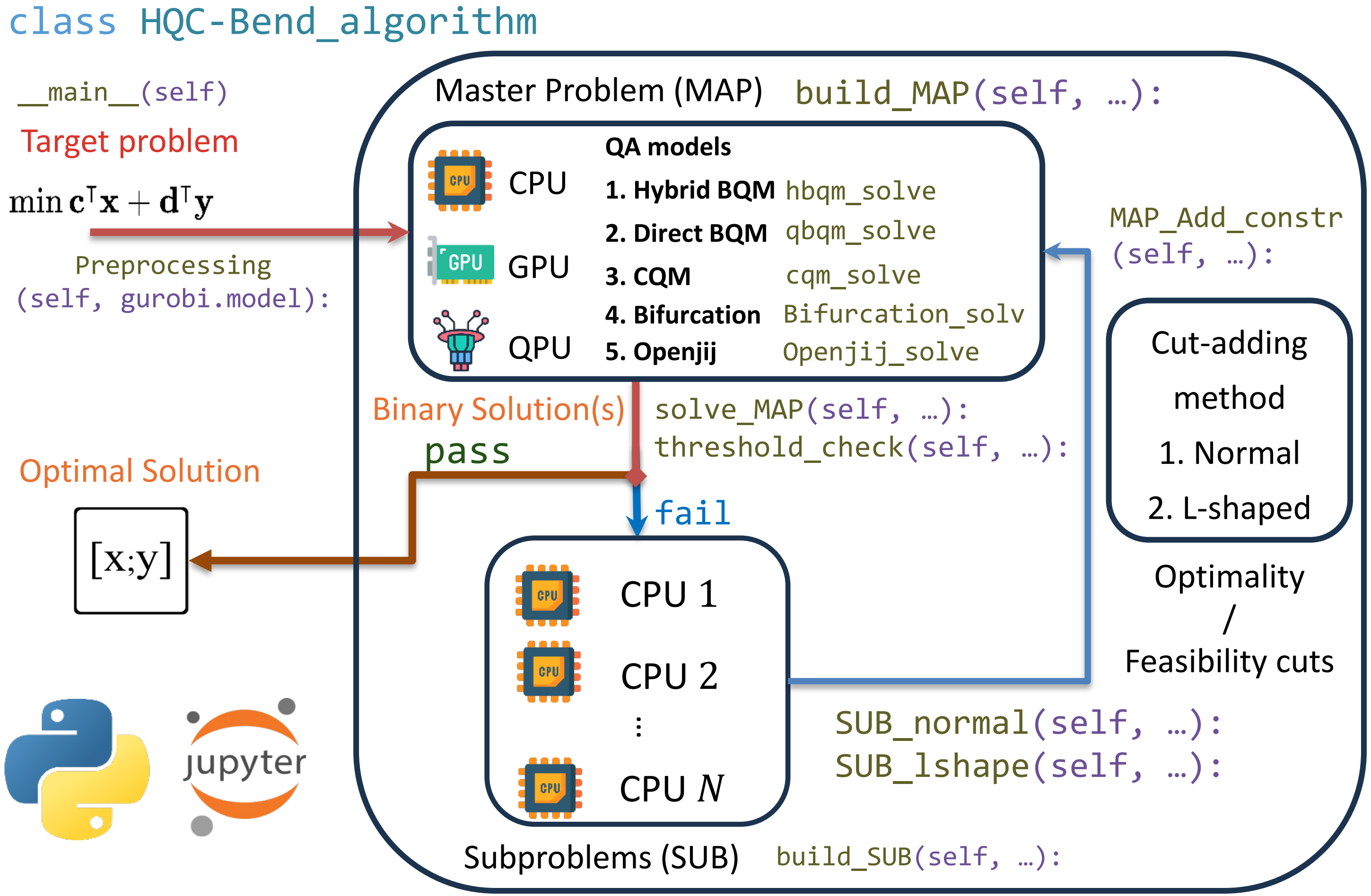
Feas. Cut $\quad + \sum_{j\in J} P_j \left((r^j)^{\top}\mathbf{A}\mathbf{x} + \sum_{l=0}^{\bar{l}^J} 2^l s^J_{kl} - \mathbf{b}^{\top}r^j \right)^2$



Work I: HQC-Bend Python Package

- Package Overview:

GitHub Page
https://github.com/djzts/HQCMCBD-API



Work I: HQC-Bend Python Package

1

1

$$
\begin{aligned}
\text{(Target Model)} \quad & \min \mathbf{c}^\top \mathbf{x} + \mathbf{d}^\top \mathbf{y} \\
\text{s.t.} \quad & \mathbf{A}\mathbf{x} + \mathbf{G}\mathbf{y} \leq \mathbf{b}, \\
& \mathbf{H}\mathbf{x} \leq \mathbf{e}, \\
& \mathbf{x} \in \mathbb{B}^n, \mathbf{y} \in \mathbb{R}^m.
\end{aligned}
$$

2

```
Preprocessing
(self, gurobi.model):
```

3 $\mathbf{c}, \mathbf{H}, \mathbf{e}$	3 $\mathbf{d}, \mathbf{A}, \mathbf{G}, \mathbf{b}$

```
build_MAP(self):
```

$$
\begin{aligned}
\text{(MAP)} \quad & \min_{\mathbf{x},\lambda,\mathbf{z}} \mathbf{c}^\top \mathbf{x} + \lambda \\
\text{s.t.} \quad & \mathbf{H}\mathbf{x} \leq \mathbf{e}, \\
& f(\lambda, \mathbf{x}) \leq 0, \ \text{(Optimality cuts)}, \\
& f(\mathbf{x}) \leq 0, \ \text{(Feasibility cuts)}, \\
& \lambda = g(\mathbf{z}), \text{(Discretization)}, \\
& \mathbf{x} \in \mathbb{B}^n, \lambda \in \mathbb{Q}, \mathbf{z} \in \mathbb{B}^l.
\end{aligned}
$$

(SUBs)

3

```
build_SUB
(self, count):
```

Preprocessing of the Target Model

the target optimization model is decomposed into multiple subproblems.

1 **Decompose Target Model**:

2 **Preprocessing**:

- Using *Gurobi* to extract **key** components:

3 **Problems Construction**:

- MAP (Master Problem):
- SUB (Subproblems):

Work I: HQC-Bend Python Package

2

CPU : Classical Gurobi solver (1-cut), Simulated Quantum Annealing Openjij Solver (1-cut)

GPU : Simulated Bifurcation (1-cut)

Simulated Bifurcation

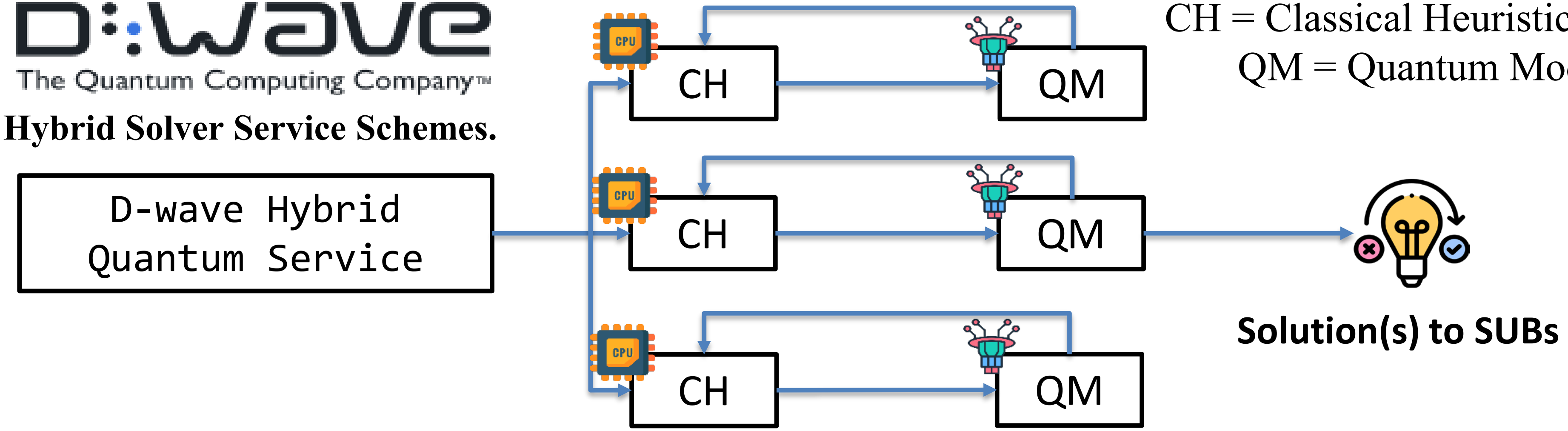
Quantum Annealing Models | ***D-wave Quantum Service*** | D-wave Hybrid Quantum Service

1. **Hybrid BQM**: `hbqm_solve(self)` 2. ***Direct BQM***: `qbqm_solve(self)` 3. **CQM**: `cqm_solve(self)`

The Quantum Computing Company™

Hybrid Solver Service Schemes.

CH = Classical Heuristic Module.
QM = Quantum Module.



Work I: HQC-Bend Python Package

3

Subproblem Solving Logic Flow

`SUB_normal:`

Normal

`SUB_lshape:`

L-shape

Optimality Cut (**OC**)

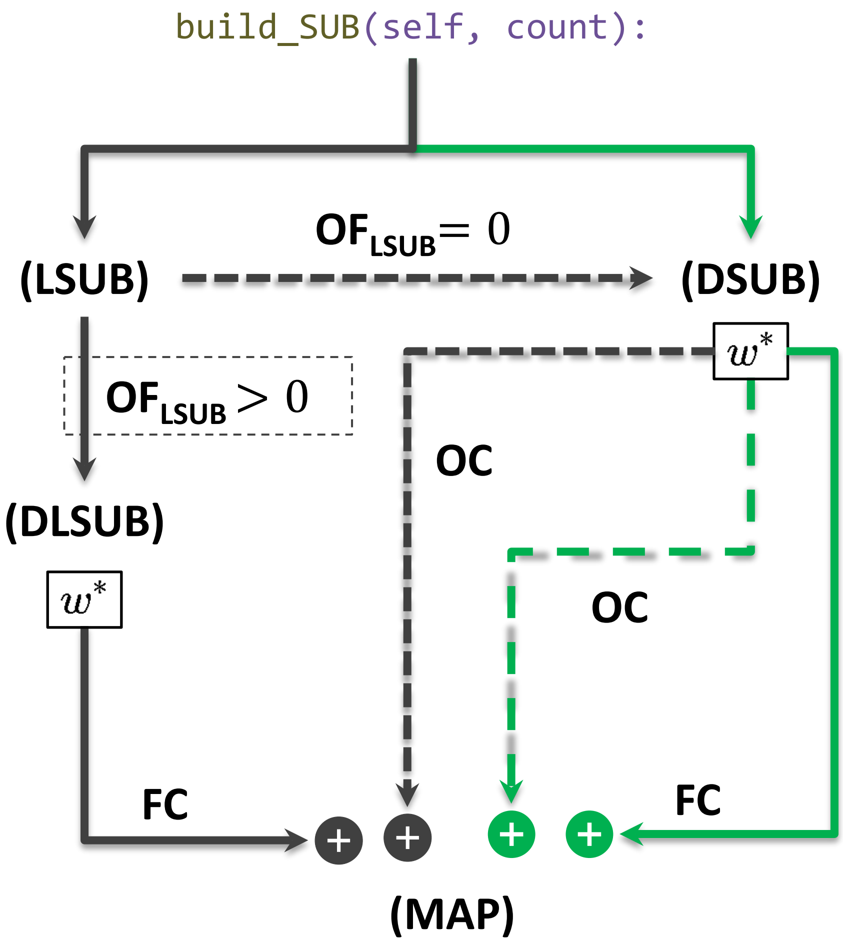
$$(\mathbf{b}-\mathbf{A}\mathbf{x})^{\mathsf{T}}\mathbf{w}^{*} \leq \lambda$$

Feasibility Cut (**FC**)

$$(\mathbf{b}-\mathbf{A}\mathbf{x})^{\mathsf{T}}\mathbf{w}^{*} \leq 0$$

Objective Function of $*$ (**OF**$_*$)

$$f_{\mathrm{obj},*}(\mathbf{x},\mathbf{y})$$



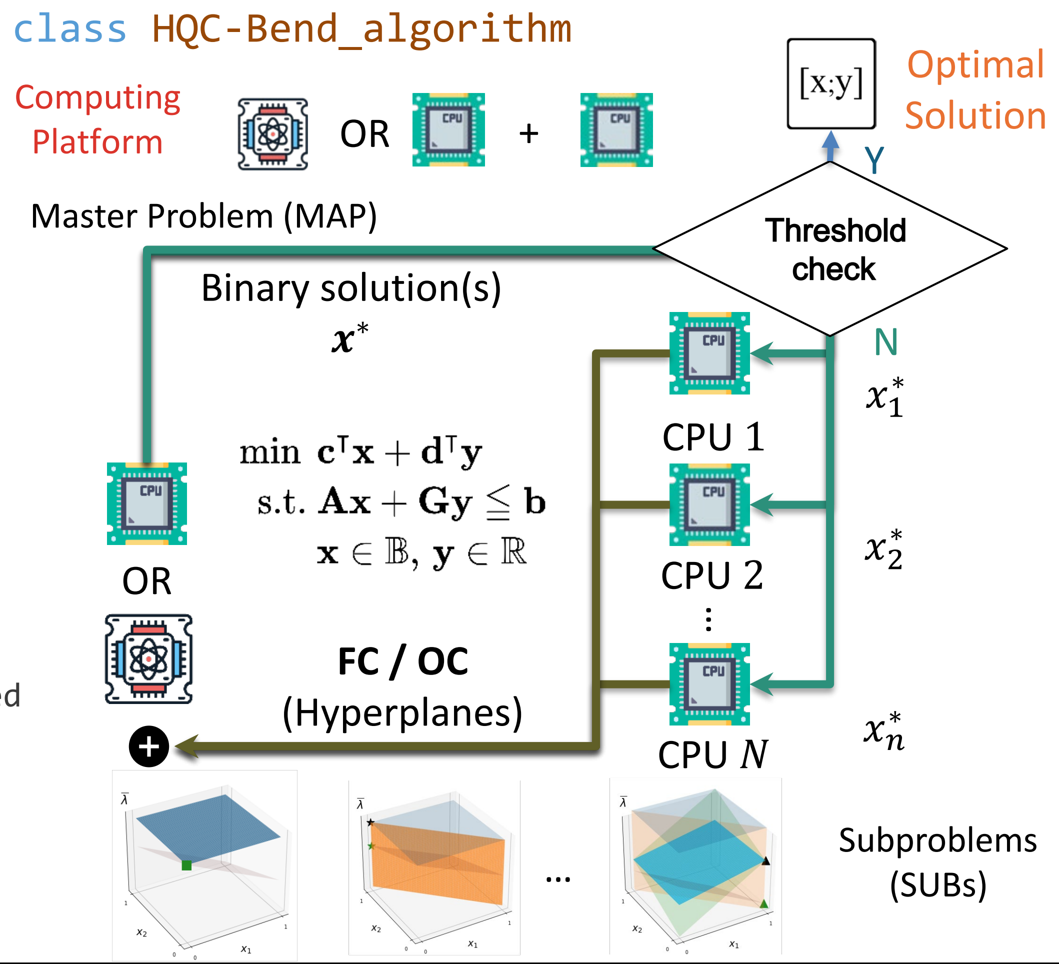
Work I: HQC-Bend Python Package

Example Python Code

```
import gurobipy as gp
from gurobipy import GRB, Model, quicksum
import numpy as np
import sys
%run HQC-Bend_notebook.ipynb
# Create a new Gurobi model
model = gp.Model(''Example'')
:
:
# Set the objective function
model.setObjective(c@x+h@y,GRB.MAXIMIZE)
# Add the constraints
model.addConstr(A@x+G@y<=b,name=''constraints'')
# Optimize the model
model.optimize()
# call the solver
Solver = HQC-Bend_algorithm(model, mode = ''default'')
Solver.run()
```

Example Text Output

```
The n-th Config file of quantum sampling is created
successfully at F:\...\Dwave_info-round-n.json.
create optimality cut 2.create optimality cut
Round n: Current Obj. value is 9.0;
lambda_upper is 17.0; lambda_lower is 11.0;
Relative gap is 54.545%; Absolute gap is 6.0.
```



Outline

- Introduction
- Work 1: Hybrid Quantum Benders' Decomposition (HQC-Bend) for Mixed-integer Linear Programming and Python Package
- **Work 2: Energy Management Problem in Internet Data Center Using HQC-Bend**
- Work 3: Optimal Energy and LLM Training Job Scheduling for Internet Data Center Using Nonlinear HQC-Bend.
- Future Work & Conclusion

Work 2: HQC-Bend in Data Center

$$\min_{\substack{u_t^{\text{dis}},\, u_t^{\text{chr}},\, p_t^{\text{dis}},\, p_t^{\text{chr}},\, x_{j,t}^{\text{chiller}},\\ x_{j,t}^{\text{tower}},\, T_{i,t}^{\text{Zone}},\, T_{i,t}^{\text{sup}},\, v_t^{\text{vent}}}} \sum_{t=0}^{T} p_t^{\text{e,g}} e_t^{\text{dc,in}}. \qquad x \in \mathbb{B}, y \in \mathbb{R}$$

The objective function: minimize the total cost of electricity imported from the grid.

$$e_t^{\text{dc,in}} = E_t^{\text{HVAC}} + E_t^{\text{Server}} + \Delta E_t^{\text{B}} - E_t^{\text{Solar}}, \ \forall t.$$

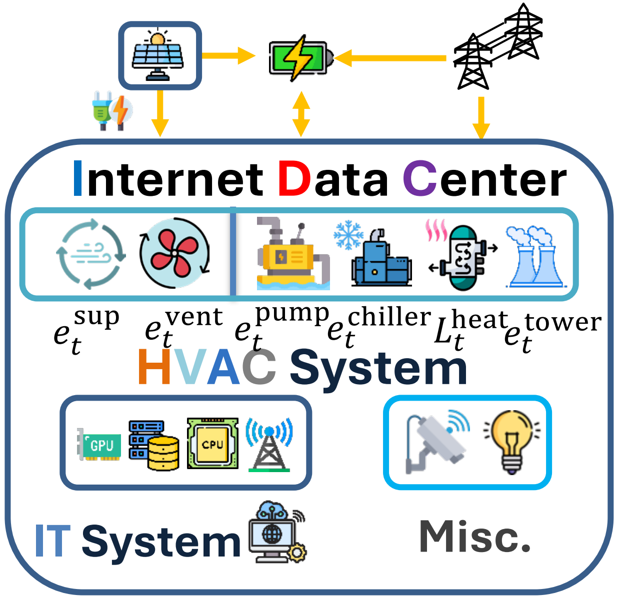
The sum of every energy sources and consumers.

$$E_t^{\text{HVAC}} = e_t^{\text{sup}} + e_t^{\text{vent}} + e_t^{\text{chiller}} + e_t^{\text{pump}} + e_t^{\text{tower}}, \ \forall t.$$

The sum of every parts' energy consumption.

$$\Delta E_t^{\text{B}} = p_t^{\text{chr}} \eta^{\text{chr}} - p_t^{\text{dis}} \cdot (\eta^{\text{dis}})^{-1}, \ \forall t.$$

Battery's (dis)charging law



$$\underline{\xi^{\text{B}}} \le E_{t+1}^{\text{B,state}} \le \overline{\xi^{\text{B}}}, \ \forall t$$

Battery status requirements at time *t*

$$E_{t+1}^{\text{B,state}} = E_t^{\text{B,state}} + \Delta E_t^{\text{B}}, \ \forall t.$$

Battery status at time *t*.

Work 2: HQC-Bend in Data Center

Detailed Formulation

$$T_{i,t}^{\text{Zone},-} \leq T_{i,t}^{\text{Zone}} \leq T_{i,t}^{\text{Zone},+}, \ \forall i, t.$$

Upper/lower bound of room temperature

$$v_t^{\text{vent}} + v_t^{\text{out}} \geq \underline{v}_t^{\text{vent}}, \ \forall t.$$

The minimum ventilation air flow speed

$$v_t^{\text{sup}} = v_t^{\text{out}} + v_t^{\text{return}}, \ \forall t.$$

The air flow speed that comes out of the AC

$$\sum_{j \in \mathbf{J}^{\text{chiller}}} x_{j,t}^{\text{chiller}} m_{j,t}^{\text{chiller}} \geq m_t^{\text{chw}}, \ \forall t.$$

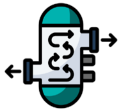
$$\sum_{j \in \mathbf{J}^{\text{tower}}} x_{j,t}^{\text{tower}} m_{j,t}^{\text{tower}} \geq m_t^{\text{conw}}, \ \forall t.$$

min capacity of chiller/condense tower water

$$T_{i,t}^{\text{sup},-} \leq T_{i,t}^{\text{sup}} \leq T_{i,t}^{\text{sup},+}, \ \forall i, t.$$

Upper/lower bound of AC temperature

$$L_t^{\text{heat}} = \left(T_t^{\text{out}} - \sum_{i \in \mathbf{I}^{\text{Zone}}} \chi_i T_{i,t}^{\text{sup}} \right) \cdot v_t^{\text{out}} c_p^{\text{air}} + \sum_{i \in \mathbf{I}^{\text{Zone}}} \chi_i \left(T_{i,t}^{\text{Zone}} - T_{i,t}^{\text{sup}} \right) \cdot v_t^{\text{return}} c_p^{\text{air}}, \ \forall t.$$



The sum of heat load in data center

$$m_t^{\text{chw}} = \frac{L_t^{\text{heat}}}{\left(T_t^{\text{chwr}} - T_t^{\text{chws}} \right) \cdot c_p^{\text{water}}}, \ \forall t.$$

$$m_t^{\text{conw}} = \frac{L_t^{\text{heat}}}{\left(T_t^{\text{conwr}} - T_t^{\text{conws}} \right) \cdot c_p^{\text{water}}}, \ \forall t.$$

The amount of chiller/condense tower water to take away the heat.

Work 2: HQC-Bend in Data Center

Detailed Formulation

$$e_t^{\text{chiller}} = \sum_{j \in \mathbf{J}^{\text{chiller}}} x_{j,t}^{\text{chiller}} \left(\beta_{0,j}^{\text{chiller}} + \beta_{1,j}^{\text{chiller}} m_{j,t}^{\text{chiller}}\right), \ \forall t.$$

$$e_t^{\text{tower}} = \sum_{j \in \mathbf{J}^{\text{tower}}} x_{j,t}^{\text{tower}} \left(\beta_{0,j}^{\text{tower}} + \beta_{1,j}^{\text{tower}} m_{j,t}^{\text{tower}}\right), \ \forall t.$$

Energy consumption of chillers & condense towers

$$e_t^{\text{pump}} = \beta_0^{\text{pump}} + \beta_1^{\text{pump}} m_t^{\text{conw}}, \ \forall t.$$

Energy consumption of pump in condense towers

$$v_t^{\text{sup}} = v_t^{\text{out}} + v_t^{\text{return}}, \ \forall t.$$

Return Air flow

$$e_t^{\text{vent}} = \beta_0^{\text{vent}} \left(v_t^{\text{vent}} - \underline{v}^{\text{vent}}\right), \ \forall t.$$

Energy consumption for ventilation

$$v_t^{\text{vent}} \geq \underline{v}^{\text{vent}}, \ \forall t.$$

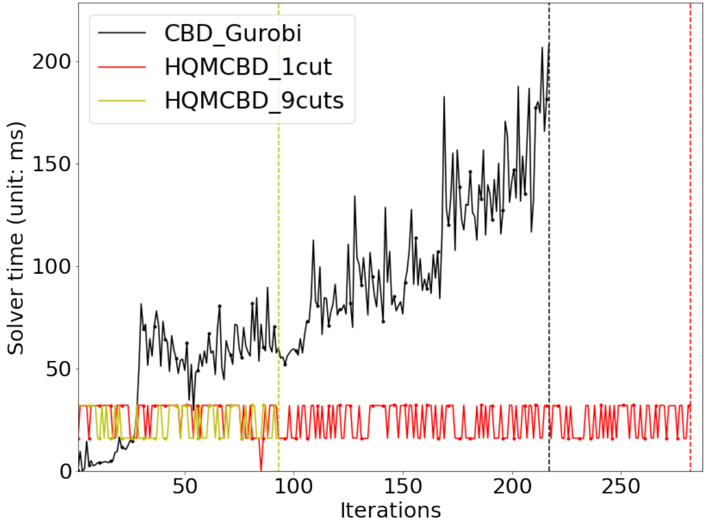
Ventilation Requirement

$$\begin{aligned} \max_{\mathbf{x},\mathbf{y}} \ & \mathbf{c}^\intercal \mathbf{x} + \mathbf{h}^\intercal \mathbf{y} \\ \text{s.t.} \ & \mathbf{T}\mathbf{x} \leq \mathbf{p} \\ & \mathbf{A}\mathbf{x} + \mathbf{G}\mathbf{y} \leq \mathbf{b} \\ & \mathbf{x} \in \mathbb{Z}^n, \mathbf{y} \in \mathbb{R}^m \end{aligned}$$

Mixed-integer linear programming (MILP)

Work 2: HQC-Bend in Data Center

- Simulation: Comparison Between Classical solver, HQC-Bend with different multi-cuts options



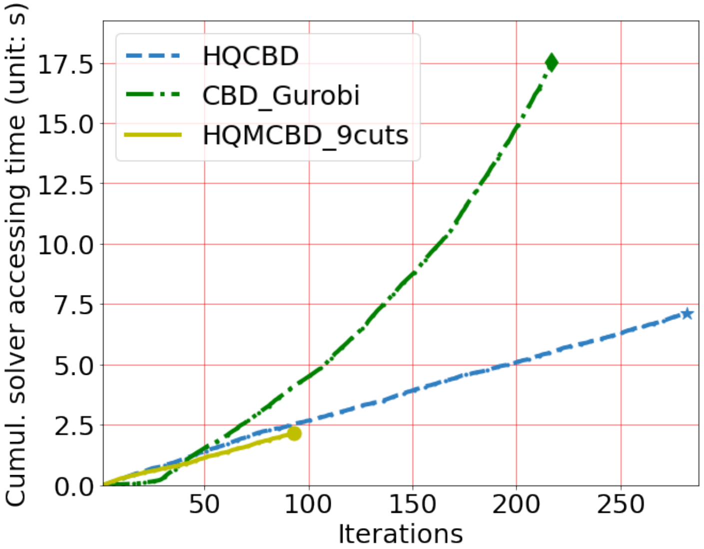
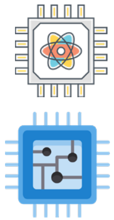


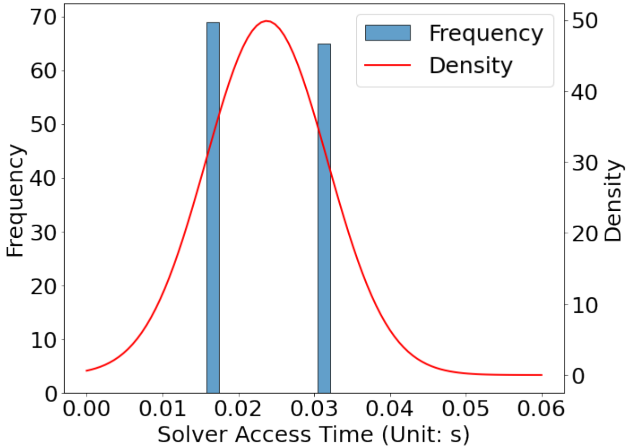
Table 1. Iteration Comparison Between **HQC-Bend** with different multi-cuts strategies

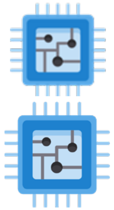
	Set-up	$\|x\|$	Iter. of CBD	Aver. iter. of HQCMBD ($N=3$)	Gain	Iter. of HQCMBD ($N=6$)			Aver. iter. of HQCMBD ($N=6$)	Gain	Aver. iter. of HQCMBD ($N=9$)	Gain
Case 1	$\{3,4,5\}$	33	117	83.67	-28%	66	74	65	68.33	-42%	56	-52%
Case 2	$\{4,2,2\}$	24	217	160	-26%	120	125	127	127.33	-41%	100	-54%

Work 2: HQC-Bend in Data Center



HQC Histogram





Classical CPU Hist.

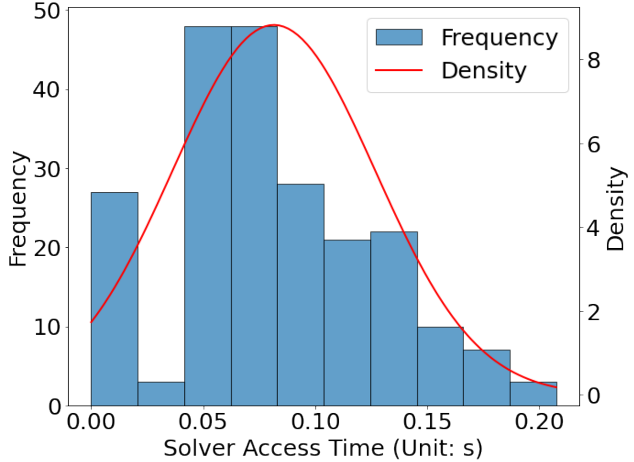


Table 2. Standard Deviation Comparison

Model \ Detail	Standard Deviation Unit: 10^{-3}	Gain
Case1 CBD	186.0	96.33%
Case1 HQCMBD	6.8	

Model \ Detail	Standard Deviation Unit: 10^{-3}	Gain
Case2 CBD	45.2	82.31%
Case2 HQCMBD	8.0	

The HQC-Bend **outperforms** the classical approach in terms of

solver ***access time***, ***Iterations***, and ***robustness***.

Outline

- Introduction
- Work 1: Hybrid Quantum Benders' Decomposition (HQC-Bend) for Mixed-integer Linear Programming and Python Package
- Work 2: Energy Management Problem in Internet Data Center Using HQC-Bend
- **Work 3: Optimal Energy Management and LLM Training Job Scheduling for Internet Data Centers Using Nonlinear HQC-Bend**.
- Future Work & Conclusion

Work 3: HQCN-Bend for IDC LLM Training

Motivation: LLM training is a core part of IDC.

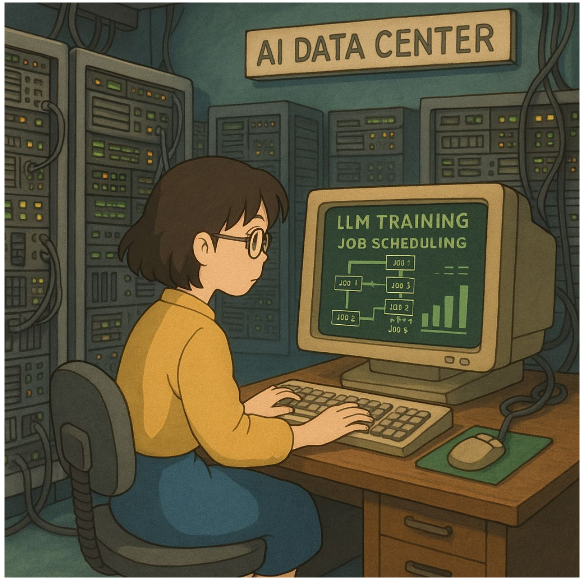
Gemini Grok GPT-3

1300MWh

US homes (130) Annually

- Is the previous work good enough?

	Work 3	**Improvement?**
IDC Number	Single	≥ 2, with data link
Power consumption	Constant	Variable to LLM tasks
LLM Training Task Scheduling	✗	✓



Work 3: HQCN-Bend for IDC LLM Training

Background: Token, Reward and Computation Resource.

Dr. Zhu Han is widely recognized as a pioneering force in the fields of wireless communications, game theory, quantum computing and network science. (GPT-4o)

```
<|im_start|>system<|im_sep|>You are a helpful assistant<
|im_end|><|im_start|>user<|im_sep|>Dr. Zhu Han is widely
recognized as a pioneering force in the fields of wirele
ss communications, game theory, quantum computing and ne
twork science. H<|im_end|><|im_start|>assistant<|im_sep|
>
```

```
200264, 17360, 200266, 3575, 553, 261, 10297, 29186, 200
265, 200264, 1428, 200266, 5822, 13, 151904, 21513, 382,
20360, 20418, 472, 261, 107046, 9578, 306, 290, 8532, 32
8, 25556, 24296, 11, 2813, 17346, 11, 48889, 34349, 326,
5402, 11222, 13, 487, 200265, 200264, 173781, 200266
```

$$T_{j,n}^{job} \approx \frac{6 \times N \times d_{\text{model},j}}{\text{n FLOPS}}$$

Performance Evaluation (GPT-like model)

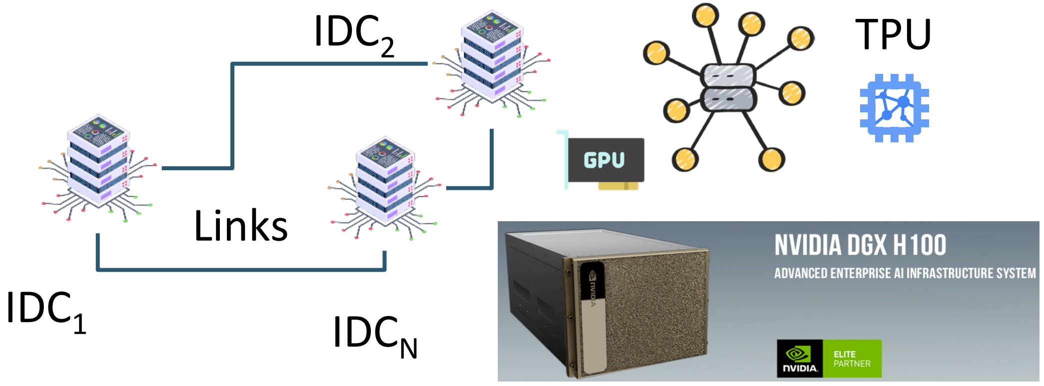
$$\text{FLOPs per token} \approx 6 \times N \times d_{\text{model}}^2$$

N is the number of transformer layers

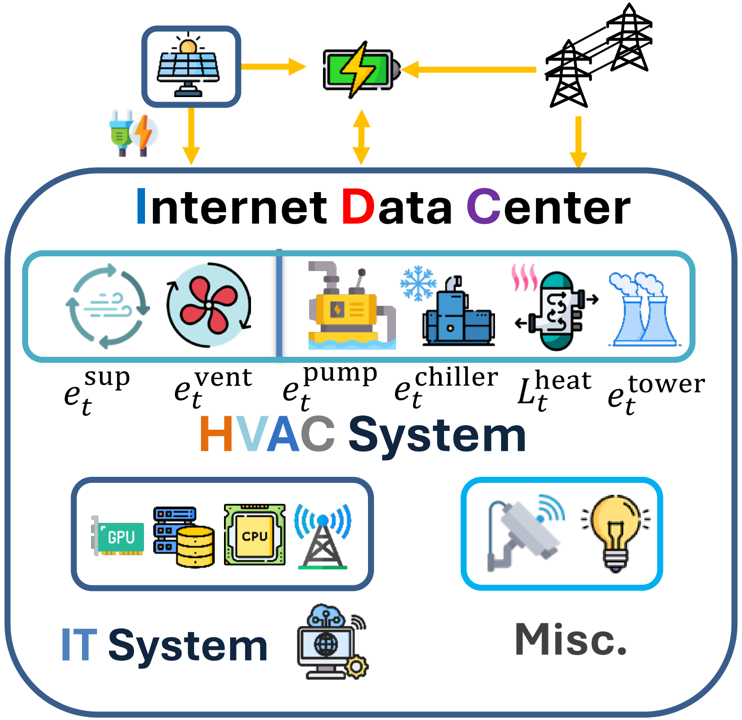
d_{model} is the hidden dimension (model width)

Work 3: HQCN-Bend for IDC LLM Training

- System Model: Multi-IDC LLM Task Scheduling and Energy Management



Param	LLM Tasks Pool			Electricity Price
IDC$_1$	LLM$_{11}$	…	LLM$_{1M}$	$c^e_{1,T}$
IDC$_2$	LLM$_{21}$	…	LLM$_{2M'}$	$c^e_{2,T}$
…	…	…	…	…
IDC$_N$	LLM$_{N1}$	…	LLM$_{NM}$	$c^e_{N,T}$



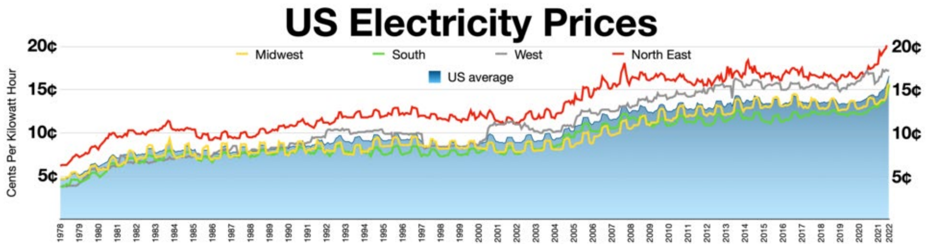
Work 3: HQCN-Bend for IDC LLM Training

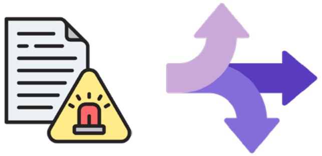
- System Model: Multi-IDC LLM Task Scheduling and Energy Management

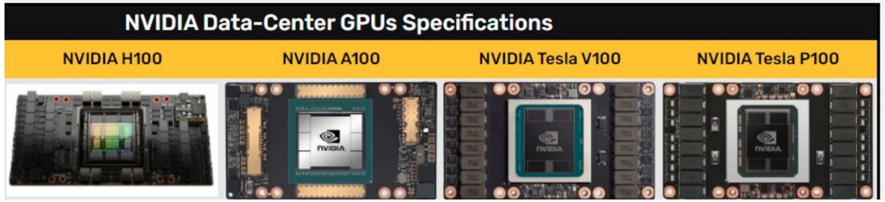
- **Challenges**:

  - LLM-Task-model-scheduling-wise:

    - Maximize the net income;
    - Local industrial electricity Price;
    - Time-sensitive LLM task completion;
    - Limited task data transmission link, which to use?
    - Computing nodes with different performance, which to use?
    - *& concerns in Work 3 in multiple IDC locations.*







Work 3: HQCN-Bend for IDC LLM Training

- System Model: Multi-IDC LLM Task Scheduling and Energy Management

- **Challenges**:

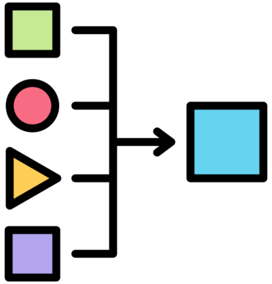
  - Algorithm-wise

  - Concerns 1: Generality of the model

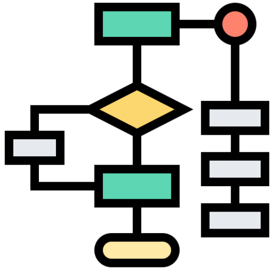
  - Concerns 2: Parameters selection

  - Concerns 3: Nonlinearity in Obj. Function

  - Concerns 4: Creating the algorithm for Mixed-integer nonlinear programming.



$$\sum_{\substack{i\in I\\ j\in J}} f(x_i\ x_j)$$



Work 3: HQCN-Bend for IDC LLM Training

- Problem Formulation: Maximize the net profit of IDCs over a period. (Binary, Continuous)

$$\max \; c^{\text{profit}} - c^{\text{loss}} - c^{\text{transfer}} - c^{\text{ebill}},$$

Objective function

$$c^{\text{profit}} = \sum_{j \in \mathcal{J}} C_j^{\text{profit}} x_j^{\text{Done}},$$

Profit decomposition

$$c^{\text{loss}} = \sum_{j \in \mathcal{J}} C_j^{\text{loss}} x_j^{\text{Abort}},$$

Monetary loss decomposition

$$c^{\text{transfer}} = \sum_{j \in \mathcal{J}} \sum_{n \in \mathcal{N}} C_{j,n}^{\text{transfer}} \cdot \boxed{x_j^{\text{TF}} \cdot x_{j,n}^{\text{job}}}.$$

Job transfer cost decomposition (*nonlinearity*)

$$c^{\text{ebill}} = \sum_{i \in \mathcal{I}} \sum_{t \in \mathcal{T}} C_{i,t}^{\text{ebill}} e_{i,t}^{\text{G}},$$

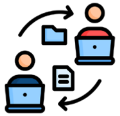
Electricity bill for IDC operation

$$x_j^{\text{Done}} + x_j^{\text{Abort}} + x_j^{\text{Hold}} = 1, \; \forall j \in \mathcal{J},$$

$$\begin{cases} x_j^{\text{Hold}} = 0, & \text{if } j \in \mathcal{J}^{\text{w}} \cup \mathcal{J}^{\text{TS}}, \\ x_j^{\text{Abort}} = 0, & \text{if } j \in \mathcal{J}^{\text{NTS}}. \end{cases}$$

x^{Done} 
x^{Abort} 
$\mathcal{J}^{\text{TS}}$ 

x^{Hold} 

x^{TF} 
$\mathcal{J}^{\text{w}}$ 

Work 3: HQCN-Bend for IDC LLM Training

1. Job scheduling (Universal)

$$\sum_{n\in\mathcal{N}} x_{j,n}^{\mathrm{job}} = x_j^{\mathrm{Done}},\ \forall j \in \mathcal{J},$$

Search all nodes to see whether the task is finished

$$u_{j,n,t}^{\mathrm{job}} \le x_{j,n}^{\mathrm{job}},\ \forall j \in \mathcal{J}, t \in \mathcal{T}, n \in \mathcal{N},$$

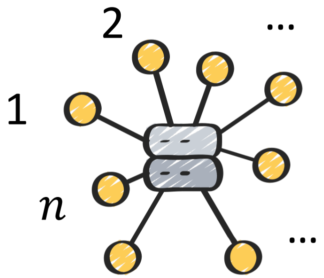
No start if the task is not assigned

$$u_{j,n,t-1}^{\mathrm{job}} - u_{j,n,t}^{\mathrm{job}} - v_{j,n,t}^{\mathrm{sd}} + v_{j,n,t}^{\mathrm{su}} = 0,\ \forall j \in \mathcal{J}, \forall t \in \mathcal{T}, n \in \mathcal{N},$$

Logical relationship between processing, start, and shutdown

$$v_{j,n,t}^{\mathrm{sd}} + v_{j,n,t}^{\mathrm{su}} \le 1,\ \forall j \in \mathcal{J}, t \in \mathcal{T}, n \in \mathcal{N}.$$

No start and finish at same time



$t = 1$

ON

OFF

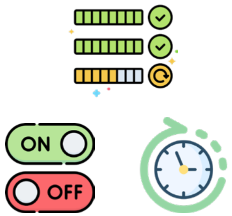
$u_{j,n,t}$	1	1	1	0	0
v	$v^{su} = 1$			$v^{sd} = 1$	

$\le x_{j,n}^{job}$

Work 3: HQCN-Bend for IDC LLM Training

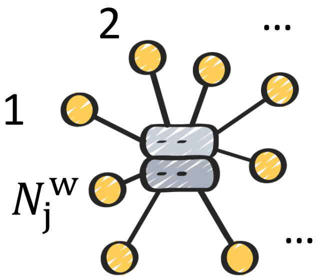
2. Current working jobs

$$x^{\text{job}}_{j,N_j^w} = x_j^{\text{Done}}, \ \forall j \in \mathcal{J}^{\text{w}},$$
$$u^{\text{job}}_{j,N_j^w,t} = x_j^{\text{Done}}, \ \forall j \in \mathcal{J}^{\text{w}}, t \in [1, T_j^w].$$



Final node state
defines task completion

Completion must occur in time



$t = 1$

$u_{j,n,t}$	1	1	1	1	0
or					
$u_{j,n,t}$	1	1	0	0	0

Work 3: HQCN-Bend for IDC LLM Training

3. Time-sensitive job (in job pool)

$$u^{\text{job}}_{j,n} = \sum_{t=1}^{T_j^{\text{TS}}} v^{\text{sd}}_{j,n,t}, \ \forall j \in \mathcal{J}^{\text{TS}}, n \in \mathcal{N},$$

$$\sum_{t=1}^{T_j^{\text{TS}}} u^{\text{job}}_{j,n,t} = T^{\text{job}}_{j,n} \cdot x^{\text{job}}_{j,n}, \ \forall j \in \mathcal{J}^{\text{TS}}, n \in \mathcal{N},$$

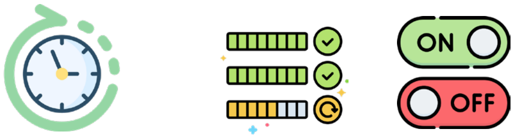
$$\sum_{\tau=1}^{T^{\text{job}}_{j,n}} v^{\text{su}}_{j,n,(t-\tau+1)} \leq u^{\text{job}}_{j,n,t}, \ \forall j \in \mathcal{J}^{\text{TS}}, n \in \mathcal{N}, t \in [1, T_j^{\text{TS}}],$$

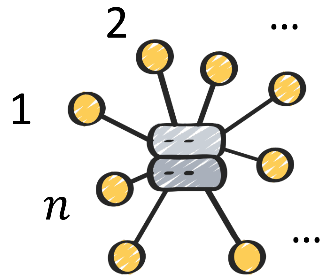
$$\sum_{n \in \mathcal{N}} \sum_{t \in [1, T_j^{\text{TS}}]} v^{\text{sd}}_{j,n,t} = x_j^{\text{Done}}, \ \forall j \in \mathcal{J}^{\text{TS}},$$

Shut down mark defines training's completion;

The task need to be training with task time;

Once the task starts. It cannot be terminated;





$t = 1$

$T^{job}_{j,n} = \max(t, T^{\text{task}}_{j,n})$

$t = T_j^{\text{TS}}$

$u_{j,n,t}$	1	1	1	0	0
v	$v^{su} = 1$			$v^{sd} = 1$	

$\leq x^{job}_{j,n}$

Work 3: HQCN-Bend for IDC LLM Training

4. Non-time-sensitive job (in job pool)

$$u^{\text{job}}_{j,n} = \sum_{t\in\mathcal{T}} v^{\text{sd}}_{j,n,t},\ \forall j\in\mathcal{J}^{\text{NTS}}, n\in\mathcal{N},$$

$$\sum_{t\in\mathcal{T}} u^{\text{job}}_{j,n,t} = T^{\text{job}}_{j,n}\cdot x^{\text{job}}_{j,n},\ \forall j\in\mathcal{J}^{\text{NTS}}, n\in\mathcal{N},$$

$$\sum_{\tau=1}^{T^{\text{job}}_{j,n}} v^{\text{su}}_{j,n,(t-\tau+1)} \le u^{\text{job}}_{j,n,t},\ \forall j\in\mathcal{J}^{\text{NTS}}, n\in\mathcal{N}, t\in\mathcal{T},$$

$$\sum_{n\in\mathcal{N}}\sum_{t\in\mathcal{T}} v^{\text{sd}}_{j,n,t} = x^{\text{Done}}_{j},\ \forall j\in\mathcal{J}^{\text{NTS}},$$

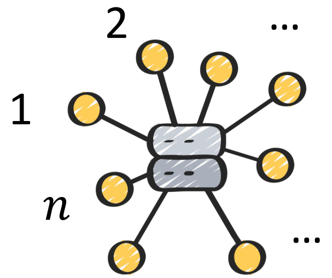
Shut down mark defines training's completion;

The task need to be training with task time;

Once the task starts. It cannot be terminated;

$t = 1$

$$T^{job}_{j,n} = \max(t, T^{task}_{j,n})$$



$u_{j,n,t}$	1	1	1	0	0
v	$v^{su} = 1$			$v^{sd} = 1$	

$\le x^{job}_{j,n}$

Work 3: HQCN-Bend for IDC LLM Training

5. Transferred job (1)

$$x_j^{\mathrm{TF}} = x_j^{\mathrm{Done}} - \sum_{\{n|\mathcal{I}^N(n)=\mathcal{I}^J(j)\}} x_{j,n}^{\mathrm{job}}, \forall j \in \mathcal{J},$$

Defines what task is transferred.

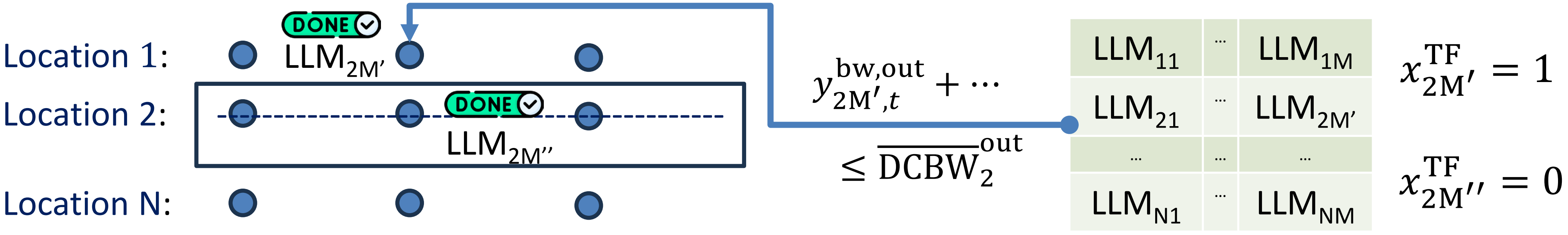
$$D_j \cdot x_j^{\mathrm{TF}} = \sum_{t=1}^{T_{\mathrm{range}}} y_{j,t}^{\mathrm{bw,out}}, \forall j \in \mathcal{J}^{\mathrm{NTS}},$$

$$D_j \cdot x_j^{\mathrm{TF}} = \sum_{t=1}^{T_j^{\mathrm{TS}}} y_{j,t}^{\mathrm{bw,out}}, \forall j \in \mathcal{J}^{\mathrm{TS}},$$

Once the task is transferred. The training data need to be upload/download to another location.

$$\sum_{\{j|\mathcal{I}^J(j)=i\}} y_{j,t}^{\mathrm{bw,out}} \le \overline{\mathrm{DCBW}}_i^{\mathrm{out}}, \forall i \in \mathcal{I}, \forall t \in \mathcal{T},$$

The uploading data size upper bound.



Work 3: HQCN-Bend for IDC LLM Training

5. Transferred job (2)

$$x^{\mathrm{TF}}_{\{j_1,\dots j_{tf}\}} = 1$$

$$y^{\mathrm{bw,out}}_{j,t} = \sum_{\{n|\mathcal{I}^N(n)\neq\mathcal{I}^J(j)\}} y^{\mathrm{bw,in}}_{j,n,t},\ \forall j\in\mathcal{J}, t\in\mathcal{T},$$

$$y^{\mathrm{bw,in}}_{j,n,t} \le \overline{y}^{\mathrm{bw,in}}_{j,n,t},\ \forall j\in\mathcal{J}, n\in\mathcal{N}, t\in\mathcal{T},$$

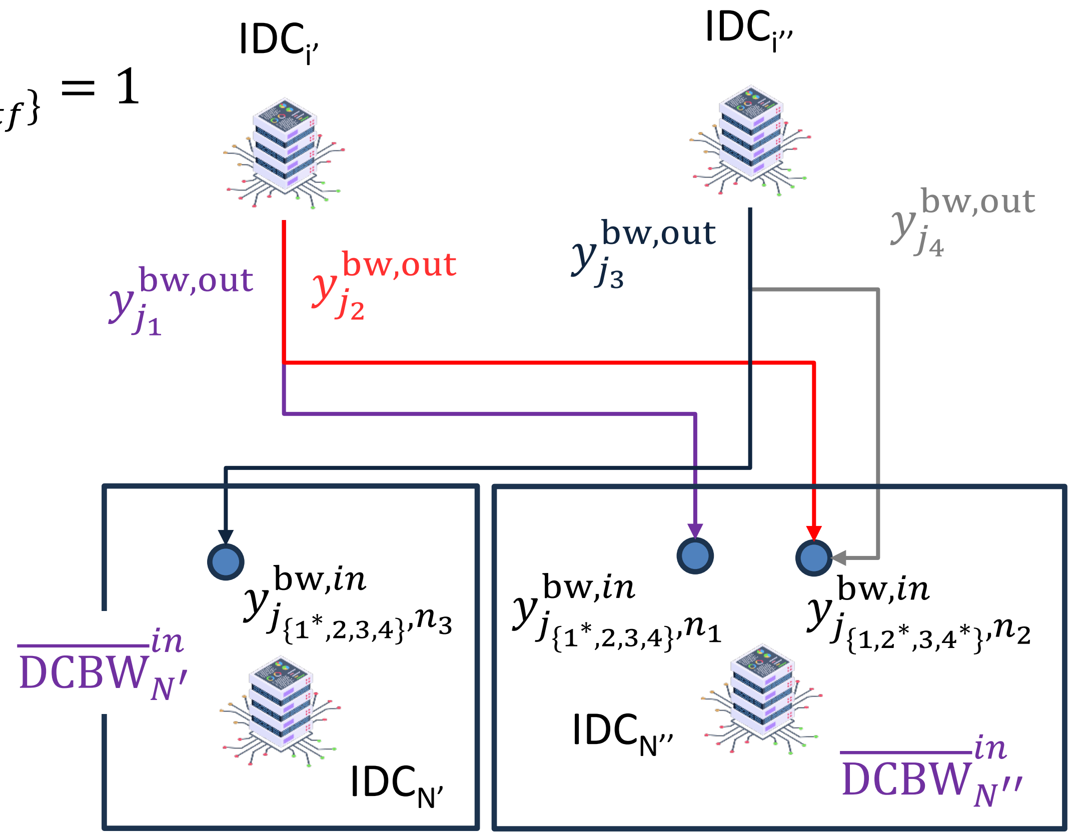
$$\overline{y}^{\mathrm{bw,in}}_{j,n,t} = \overline{y}^{\mathrm{bw,in}} x^{\mathrm{job}}_{j,n},\ \forall j\in\mathcal{J}, n\in\mathcal{N}, t\in\mathcal{T},$$

$$\sum_{\{n|\mathcal{I}^N(n)=i\}} y^{\mathrm{bw,in}}_{j,n,t} \le \overline{\mathrm{DCBW}}^{\mathrm{in}}_i,\ \forall i\in\mathcal{I},\ \forall t\in\mathcal{T},$$

Select the receiving node* to download the LLM.

The upper bound of downloading speed at node.

The upper bound of downloading speed at IDC.



$$x^{\mathrm{job}}_{j_3,n_3} = x^{\mathrm{job}}_{j_4,n_2} = x^{\mathrm{job}}_{j_1,n_1} = x^{\mathrm{job}}_{j_2,n_2} = 1$$

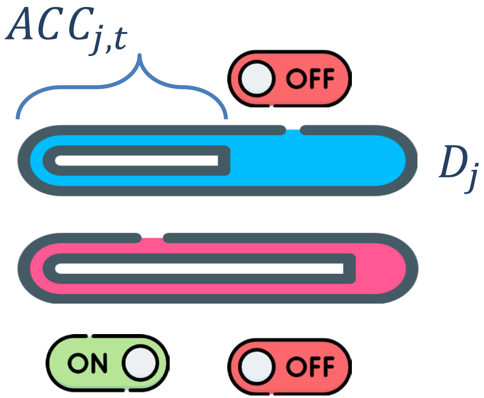
5. Transferred job (3)

$$ACC_{j,t} = \sum_{n \in \mathcal{N}} \sum_{\tau=1}^{t} y_{j,n,\tau}^{\text{bw,in}}, \ \forall j \in \mathcal{J}, t \in \mathcal{T},$$

$$ACC_{j,t} \geq D_{j,t}^{in}, \ \forall j \in \mathcal{J}, t \in \mathcal{T},$$

$$D_{j,t}^{in} = D_j \cdot v_{j,n,t+1}^{\text{su}}, \ \forall j \in \mathcal{J}, n \in \{n \mid \mathcal{I}^N(n) \neq \mathcal{I}^J(j)\}, t \in \mathcal{T}.$$

The size of the data that has been downloaded. Training can only start after it is **completely downloaded**.



	$v_{j,n,t+1}$
$ACC_{j,t} \leq D_j$	{0}
$ACC_{j,t} = D_j$	{0,1}

Job$_j$ can start now

Regulation

Ensures LLM is only transmit to **a single node**

$$D_j \cdot x_j^{\text{TF}} = \sum_{t=1}^{T_{\text{range}}} y_{j,t}^{\text{bw,out}}$$

$$y_{j,t}^{\text{bw,out}} = \sum_{\{n \mid \mathcal{I}^N(n) \neq \mathcal{I}^J(j)\}} y_{j,n,t}^{\text{bw,in}}$$

Work 3: HQCN-Bend for IDC LLM Training

6. Computing nodes Energy Modeling Overview

ON   Idle   OFF

$$e_{n,t}^{\mathrm{O,Node}} = e_{n,t}^{\mathrm{O,N,idle}} + e_{n,t}^{\mathrm{O,N,w}}, \ \forall n \in \mathcal{N}, t \in \mathcal{T},$$

$$e_{n,t}^{\mathrm{O,N,idle}} = E_{n}^{\mathrm{O,N,idle}} u_{n,t}^{\mathrm{power}}, \ \forall n \in \mathcal{N}, t \in \mathcal{T},$$

$$e_{n,t}^{\mathrm{O,N,w}} = \sum_{j \in \mathcal{J}} E_{j,n}^{\mathrm{O,N,w}} \cdot \beta_{j}^{\mathrm{TDP}} \cdot u_{j,n,t}^{\mathrm{job}}, \ \forall n \in \mathcal{N}, t \in \mathcal{T},$$

The power consumption of every node (idle, working)

$$u_{n,0}^{\mathrm{power}} = 1, \ \forall n \in \mathcal{N}^{*},$$

$$u_{j,n,t}^{\mathrm{job}} \leq u_{n,t}^{\mathrm{power}}, \ \forall j \in \mathcal{J}, \forall t \in \mathcal{T}, n \in \mathcal{N},$$

$$u_{n,t-1}^{\mathrm{power}} - u_{n,t}^{\mathrm{power}} - v_{n,t}^{\mathrm{power,sd}} + v_{n,t}^{\mathrm{power,su}} = 0, \ \forall t \in \mathcal{T}, n \in \mathcal{N}.$$

The power state of every node (idle, working)

Node Type	Power (KW)	Performance (petaFLOPS)
10 DGX1	350	17.3
10 DGX2	100	20
4 DGXA100	26	20
10 DGXA100s	15	13
1 DGXA200	14.3	72
1 DGXH100	10.2	32

TABLE I: Power and Performance Specifications of Nodes

Work 3: HQCN-Bend for IDC LLM Training

7. HVAC, Temperature, and BESS System

Those constraints are referred to [1]. However, we made several changes based our setup.

7.1 Modified Heating and Air Conditioning System

$$T_{z,t}^{\mathrm{Zone}} = T_{z,t-1}^{\mathrm{Zone}} + \sum_{z' \in adj(z)} \left(\frac{T_{z',t-1}^{\mathrm{Zone}} - T_{z,t-1}^{\mathrm{Zone}}}{C_z^{\mathrm{heat}} R_{z'z}^{\mathrm{Zone}}} \right) + \frac{\theta_{z,t}}{C_z^{\mathrm{heat}}}$$

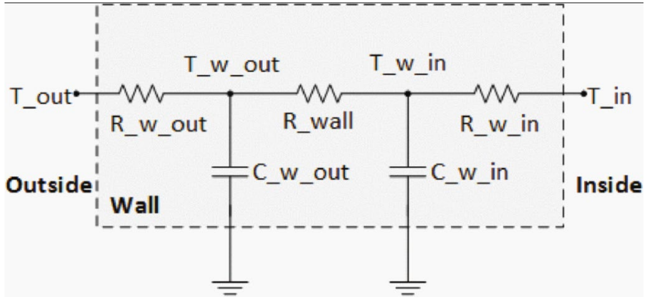
$$+ \frac{\dot{m}_{z,t}^{\mathrm{Zone}} c^{\mathrm{a,s}} \left(T_{z,t}^{\mathrm{AC}} - T_{z,t-1}^{\mathrm{Zone}} \right)}{C_z^{\mathrm{heat}}}, \forall z \in \mathcal{Z}, \forall t \in \mathcal{T},$$

$$\text{where } C_z^{\mathrm{heat}} = c^{\mathrm{a,s}} \cdot \rho^{\mathrm{air}} \cdot S_z^{\mathrm{Zone}} \cdot h_z,$$

$$\dot{m}_{z,t}^{\mathrm{Zone}} = k_z^{\mathrm{AC}} \cdot v_t^{\mathrm{AC}},$$

$$\theta_{z,t} = \xi \sum_{\{n | \mathcal{Z}(n) = z\}} E_{n,t}^{\mathrm{O,Node}}.$$

Heat from the local computing nodes.



Time Discrete Difference Room Temperature Model [2]

[1] Zhao, Zhongqi, Lei Fan, and Zhu Han. "Optimal Data Center Energy Management with Hybrid Quantum-Classical Multi-Cuts Benders' Decomposition Method." IEEE Transactions on Sustainable Energy (2023).

[2] Belić, Filip, Željko Hocenski, and Dražen Slišković. "Thermal modeling of buildings with RC method and parameter estimation." 2016 International Conference on Smart Systems and Technologies (SST). IEEE, 2016.

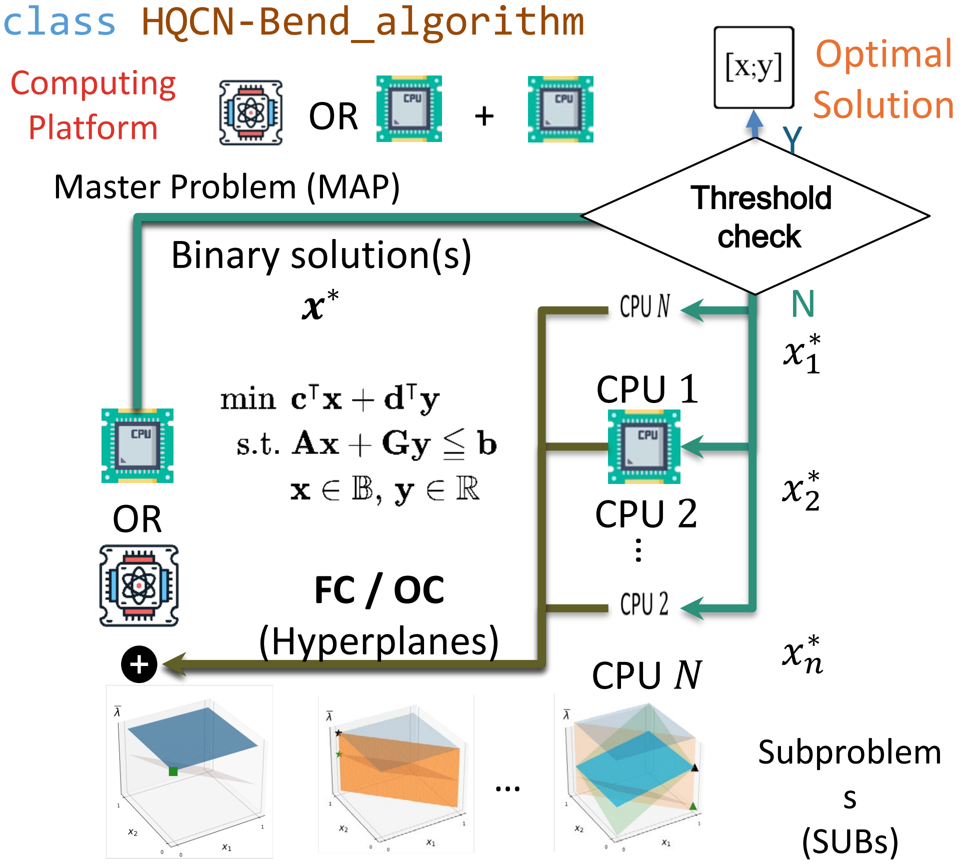
Work 3: HQCN-Bend for IDC LLM Training

- Algorithm: Hybrid Quantum-classical Nonlinear Benders' decomposition Approach

Step 1: Reformulate the objective function

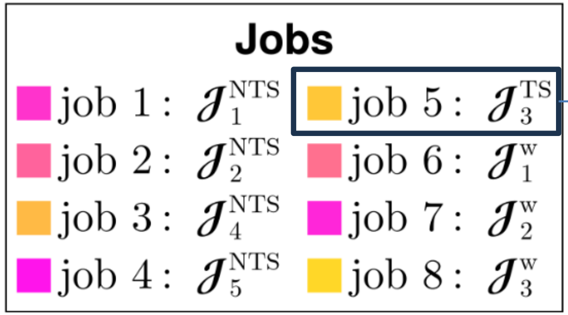
$$f(\mathbf{x}') = \mathbf{x}'^{\intercal}\mathbf{Q}_{\mathrm{QUBO}}\mathbf{x}'$$

$$\begin{aligned}
&\mathbf{x}^{\intercal}\,\mathrm{diag}(\mathbf{c})\mathbf{x} \\
&+ \sum_{i=-\underline{m}}^{\overline{m}_{+}} w_{i+\underline{m}} 2^{i} w_{i+\underline{m}} - \sum_{j=0}^{\overline{m}_{-}} w_{j+(1+\underline{m}+\overline{m}_{+})} 2^{j} w_{j+(1+\underline{m}+\overline{m}_{+})} \\
&+ \sum_{k\in K} P_k \left(\overline{\lambda}(\mathbf{w}) + (u^k)^{\intercal}\mathbf{A}\mathbf{x} + \sum_{l=0}^{\bar{l}^{K}} 2^{l} s_{kl}^{K} - b^{\intercal}u^k \right)^2 \\
&+ \sum_{j\in J} P_j \left((r^j)^{\intercal}\mathbf{A}\mathbf{x} + \sum_{l=0}^{\bar{l}^{J}} 2^{l} s_{kl}^{J} - \mathbf{b}^{\intercal} r^j \right)^2
\end{aligned}$$



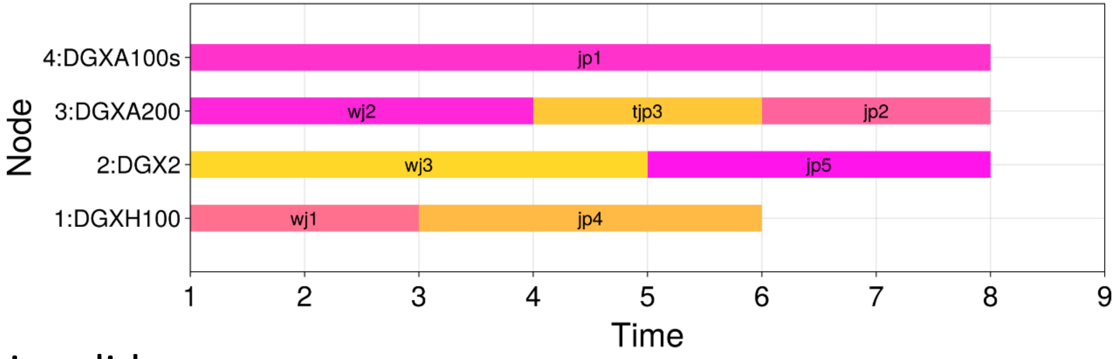
Work 3: HQCN-Bend for IDC LLM Training

Simulation Results:

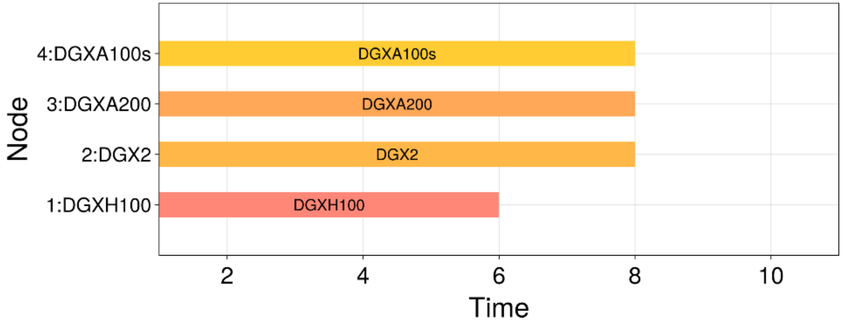


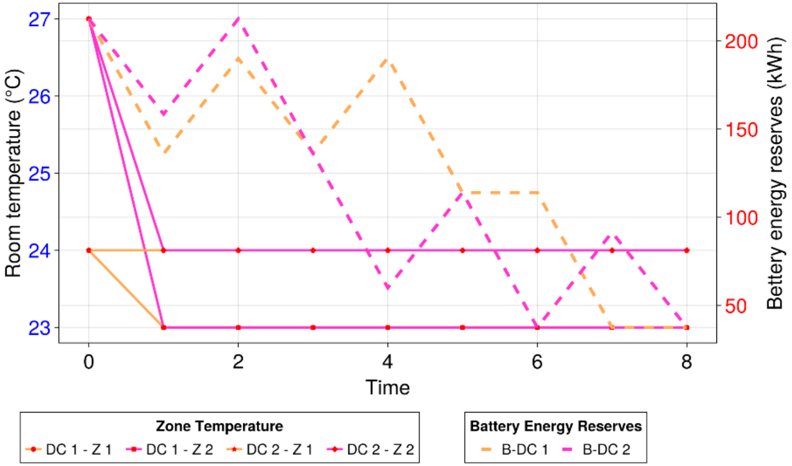
TF

- LLM Training Task scheduling



- The LLM training task / device arrangement is valid
- Zone temperature is within the bound

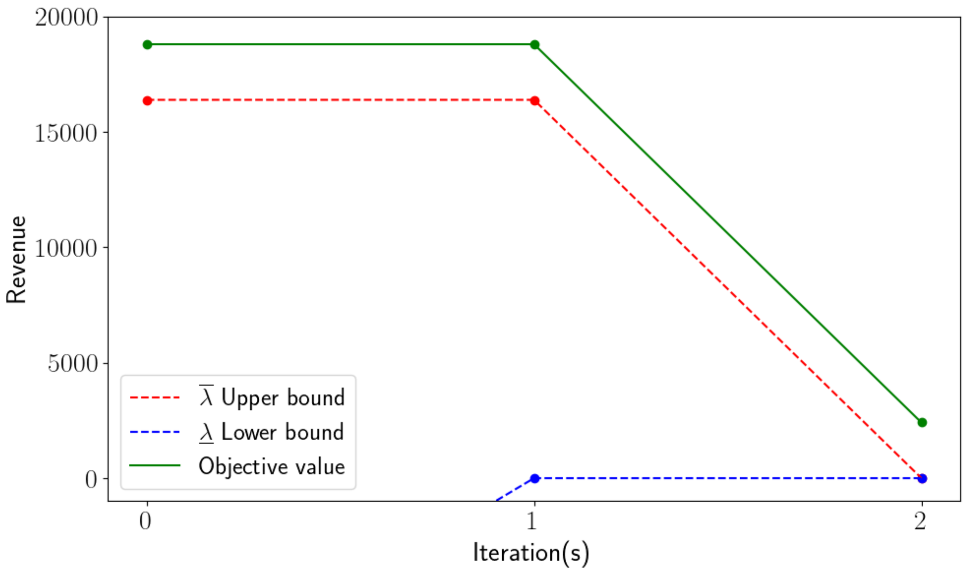




Work 3: HQCN-Bend for IDC LLM Training

Simulation Results:

- Benders Decomposition Performance and Convergence:



$\overline{\text{Solver access time}}$ / iter.

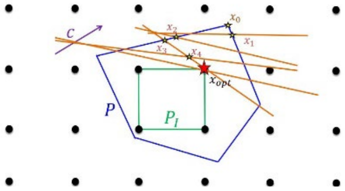
Jij	114.13s sometimes take 3,4 iterations
GPU Simulated Bifurcation	23.33s (*Max iter. = 1e4*, agent = 8192)

The GPU runs faster than the CPU based algorithm

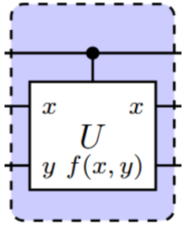
Goto, H., Tatsumura, K., & Dixon, A. R. (2019). Combinatorial optimization by simulating adiabatic bifurcations in nonlinear Hamiltonian systems. Science advances, 5(4), eaav2372.
https://www.openjij.org/

Future Work & Conclusion

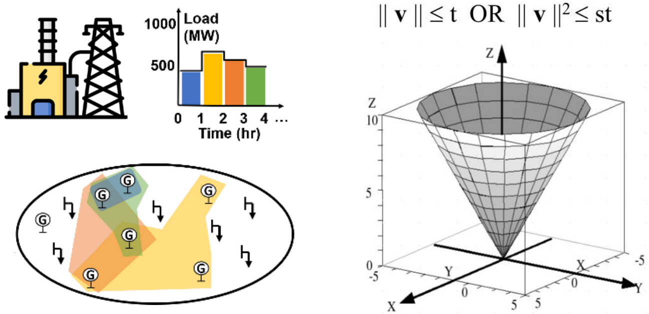
- Future Work
  - HQC-Bend package Upgrade in for MICP
    - Benders' Dual / General BD / Logic BD
    - High-order unconstraint binary optimization (HUBO) in Obj./Constraint
      - $\prod x_i \cdot x_j \cdot \ldots \cdot x_k$ to Digital Q Circuit / QUBO (QA)



Plane-cutting Method



  - Internet Data Center Model Upgrade
    - Unit commitment convex constraint plug-in (MISOCP)
    - Dynamic flexible training LLM task scheduling
    - Quantum communications

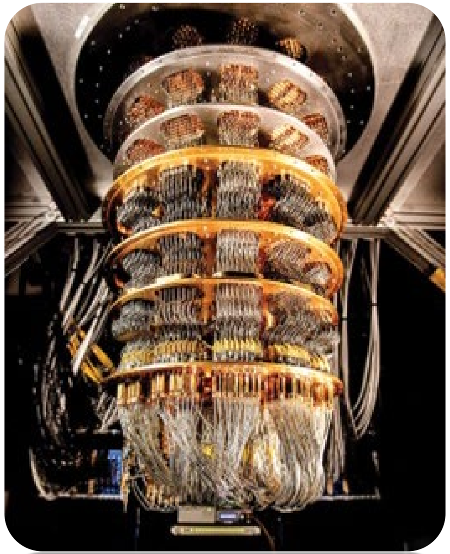


Internet Data Center Model

Future Work & Conclusion

Conclusion

- Quantum Computing (QC) provide a special way to deal with the complex MIP. By leverage Both QC and classical computation power, HQC-Bend can reduce the computation time for Mixed-integer Programming significantly.



- Work 1: HQC-Bend for MILP and Python Package
    - ✓ Reformulate the MAP of BD for the MILP problem and validate the algorithm.
    - ✓ Introduce a Python package implementing the HQC-Bend algorithm.
- Work 2: Energy Management Problem in Internet Data Center Using HQC-Bend
    - ✓ Propose a MILP model for IDC energy management.
    - ✓ the HQC-Bend approach outperforms the CBD approach in practice.
- Work 3: Optimal Energy Management and LLM Training Job Scheduling for IDC Using Nonlinear HQC-Bend
    - ✓ Propose a MINLP model for IDC LLM task scheduling & energy management
    - ✓ The HQCN-Bend method is feasible for solving certain MINLPs.

Publications

Journal

1. Zhao, Z., Fan, L., & Han, Z. (2023). Optimal Data Center Energy Management with Hybrid Quantum-Classical Multi-Cuts Benders' Decomposition Method. IEEE Transactions on Sustainable Energy.
2. Xuan, W., Zhao, Z., Fan, L., & Han, Z. (2024). Lagrangian Relaxation Based Parallelized Quantum Annealing and its Application in Network Function Virtualization. IEEE Open Journal of the Communications Society.

Conference

1. Zhao, Z., Fan, L., & Han, Z. (2022, April). Hybrid quantum benders’ decomposition for mixed-integer linear programming. In 2022 IEEE Wireless Communications and Networking Conference (WCNC) (pp. 2536-2540). IEEE.
2. Xuan, W., Zhao, Z., Fan, L., & Han, Z. (2021, October). Minimizing delay in network function visualization with quantum computing. In 2021 IEEE 18th International Conference on Mobile Ad Hoc and Smart Systems (MASS) (pp. 108-116). IEEE.
3. Zhao, Z., Fan, L., Guo, Y., Wang, Y., Han, Z., & Hanzo, L. (2024, June). QAOA-assisted benders' decomposition for mixed-integer linear programming. In ICC 2024-IEEE International Conference on Communications (pp. 1127-1132). IEEE.
4. Zhao, Z., Fan, L., Zheng, H., & Han, Z. (2023, October). Quantum Computing for Cable-Routing Problem in Solar Power Plants. In 2023 North American Power Symposium (NAPS) (pp. 1-6). IEEE.
5. Zhao, Z., Fan, L., & Han, Z. (2024, October). Optimal Energy and IT Service Emergency Schedule for Internet Data Center. In 2024 56th North American Power Symposium (NAPS) (pp. 1-6). IEEE.
6. Zhao, Z., Yao, Y., Fan, L., & Ding, F. (2024, July). Spatial-Temporal PV Hosting Capacity Estimation and Evaluation. In 2024 IEEE Power & Energy Society General Meeting (PESGM) (pp. 1-5). IEEE.

Conference/Demo

1. **Zhao, Z**, Mingze Li, Lei Fan, and Zhu Han. "HQC-Bend: A Python Package of Hybrid Quantum-Classical Multi-cuts Benders' Decomposition Algorithm." 2025 IEEE International Conference on Quantum Communications, Networking, and Computing (QCNC) / Computer Communication (INFOCOM) / International Conference on Communications (ICC). IEEE, 2025.
2. **Zhao, Z**, Lei Fan, and Zhu Han. " Optimal Data Center Energy Management and LLM Task Scheduling with Hybrid Quantum-Classical Nonlinear Benders' Decomposition Method." 2025, Ongoing.

Should I approve
Zhongqi's Defense?
Approve
Still
Approve

Thank you!

Best Professors Ever!