Quantum Annealing Inspired Algorithms for Reconstruction at High Energy Colliders

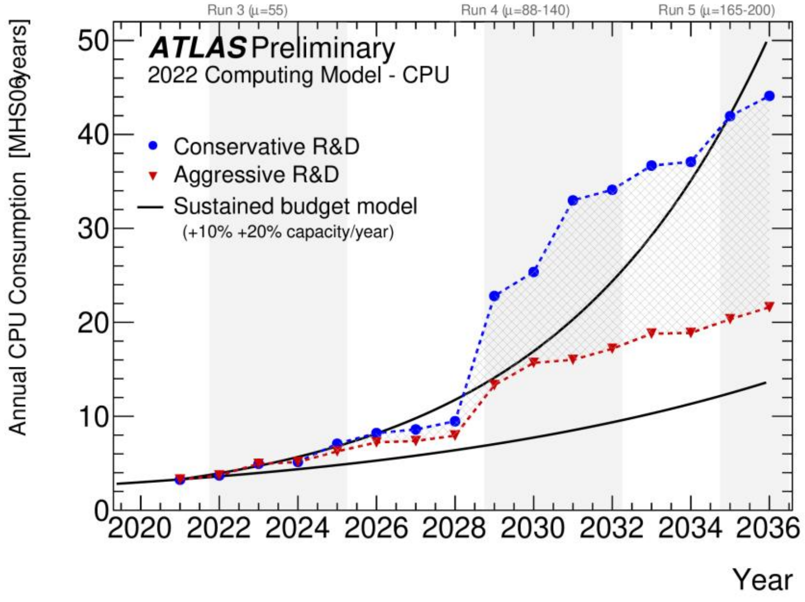
Quantum Computing & Machine Learning Workshop, August 6-8, 2024

大川(Okawa) 英希(Hideki)

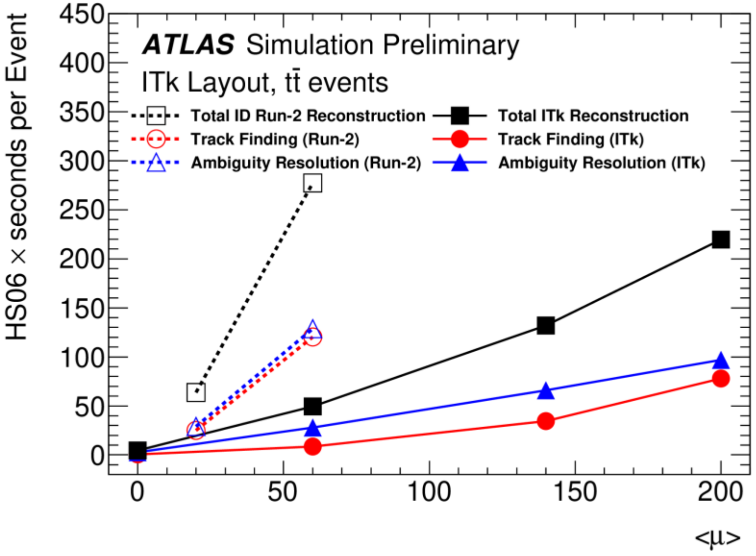
Institute of High Energy Physics, Chinese Academy of Sciences

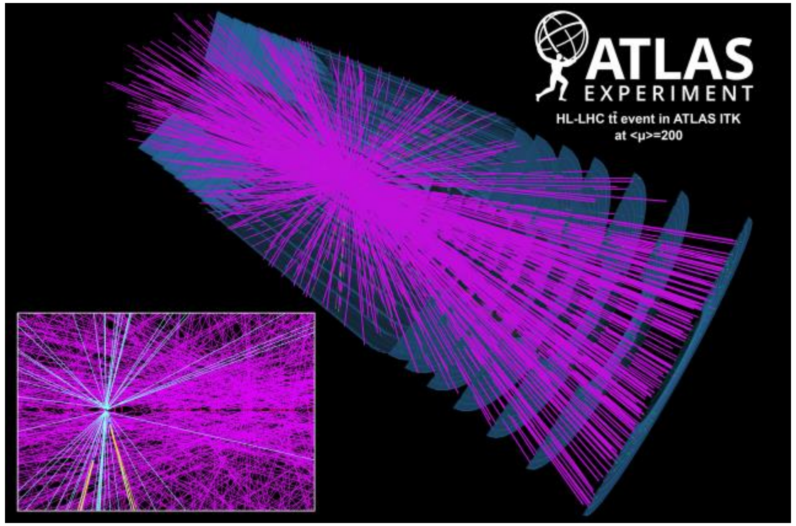
Work in collaboration with Qing-Guo Zeng, Xian-Zhe Tao, Man-Hong Yung [SUSTech/IQSE]

Reconstruction at LHC & HL-LHC



ATL-PHYS-PUB-2019-041





- At the HL-LHC, **CPU time exponentially increases with pileup**, leading to increase in annual computing cost by x10-20.
- **Tracking is the most CPU-consuming reconstruction task.**
- **Jet reconstruction is also known to be CPU-intensive.**
- GPU & ML-based approaches are actively investigated for tracking, but **quantum algorithms may also bring in innovations.**

	Run 1	Run 2	HL-LHC
μ	21	40	150-200
Tracks	~280	~600	**~7-10k**

Quantum Approaches

Quantum Gates

- Uses quantum logic gates
- General-purposed
- IBM, Google, Xanadu, IonQ, Origin Quantum (本源), QuantumCTek (国盾量子), etc.

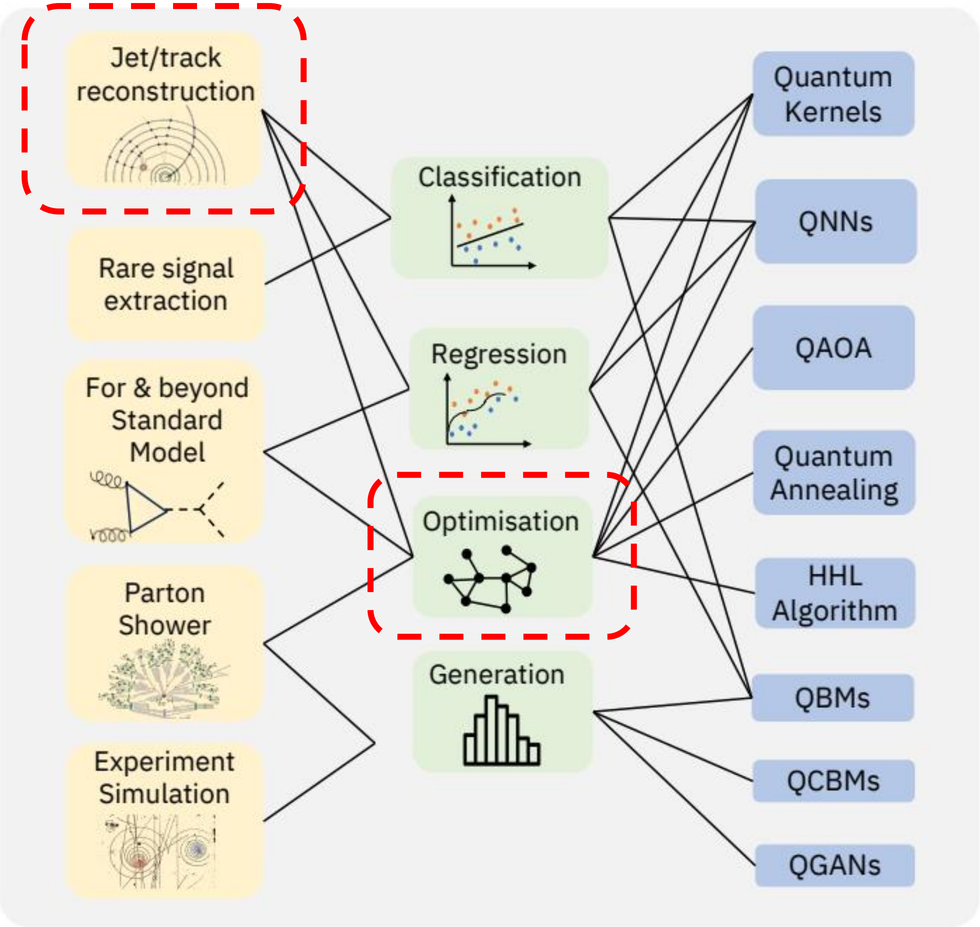
Ising machines

Quantum Annealing

- Uses adiabatic quantum evolution to seek for the ground state of a Hamiltonian
  → Only applicable to optimization problems
- Implemented in D-Wave Systems.

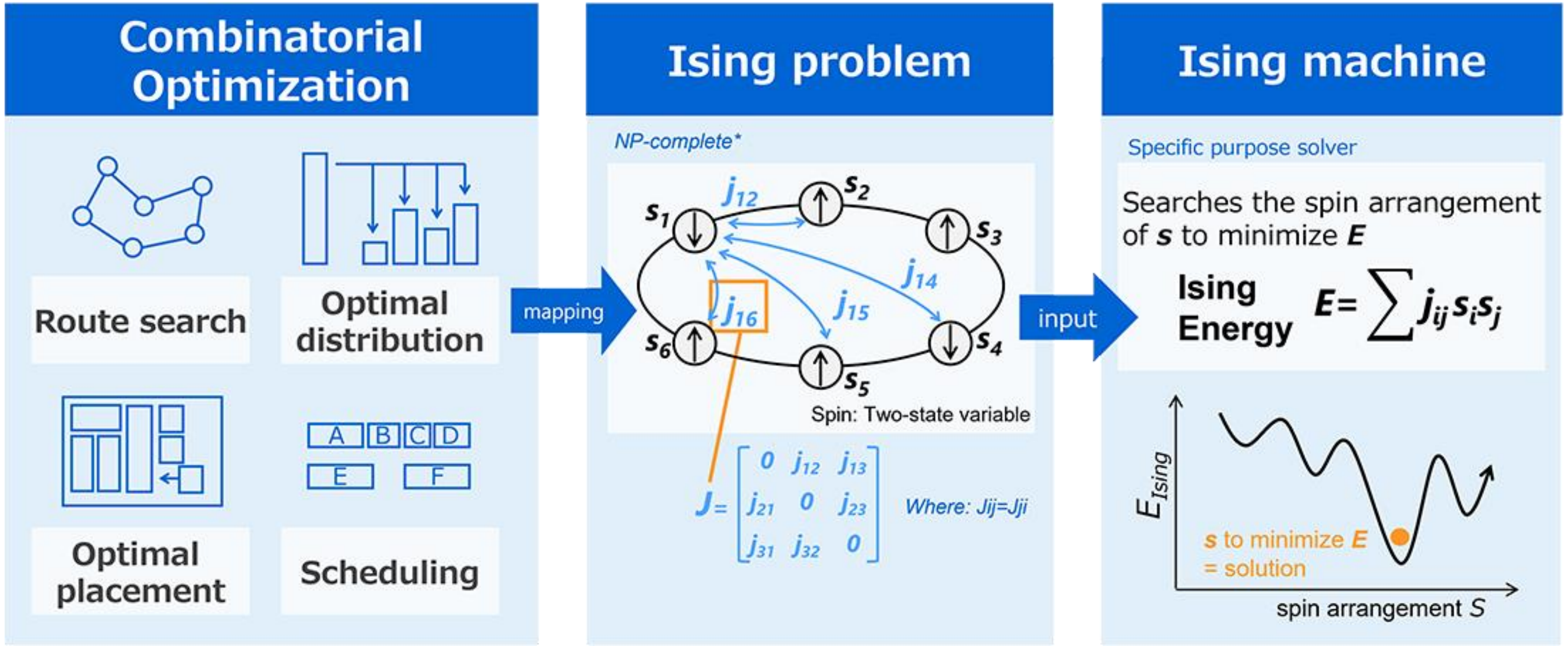
Quantum-Inspired ← Scope of this talk

- Inspired by quantum annealing.
- Simulated annealing, simulated coherent Ising machine, simulated bifurcation, etc.



QC4HEP White Paper

Combinatorial Optimization Problem



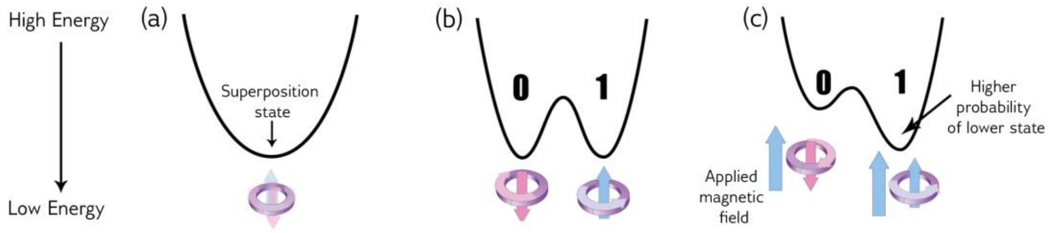
©TOSHIBA

- Combinatorial optimization problems are non-deterministic polynomial time (NP) complete problem: no efficient algorithm exists to find the solution.
- They can be mapped to Ising problems → Ising machines can provide quasi-optimal answers
- **Track & jet reconstruction can also be formulated as such problems.**

Quantum Approaches

Quantum annealing

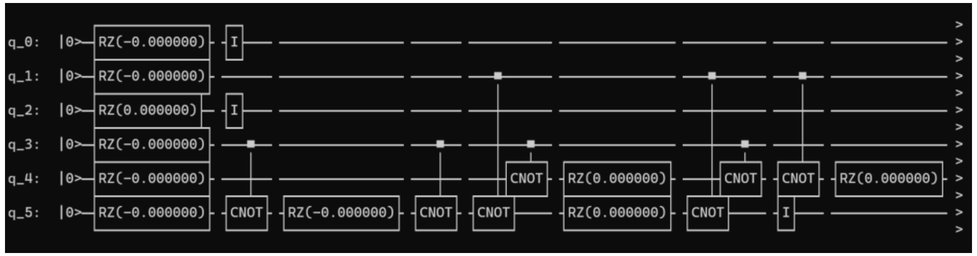
From D-wave website



- Quantum annealer looks for the global minimum of a given function with quantum tunneling.
- D-Wave currently provides 5000+ qubit service.
- Pros: High number of qubits available, although not all qubits are available for fully connected graphs (only a few hundred qubits)
- Cons: Unable to access the actual hardware from China.

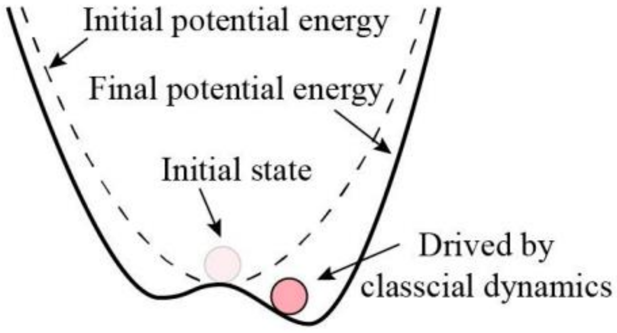
Quantum Gates

QAOA circuit implemented in Origin Quantum



- Quantum gate machines are universal, and can also solve Ising problems with variational circuits: e.g. Variational Quantum Eigensolver (VQE), Quantum Approximate Optimization Algorithm (QAOA), etc.
- Pros: Universal computing, a few platforms available in China
- Cons: Number of qubits is much less than quantum annealing

Quantum Annealing Inspired Algorithms (QAIAs)



Quantum inspired algorithm

- "Quantum-inspired" algorithms search for minimum energy through the **classical time evolution of differential equations:** simulated annealing, simulated bifurcation (SB), simulated coherent Ising machine, etc.
- **SB in particular can run in parallel unlike simulated annealing**, in which one needs to access the full set of spins & not suitable for parallel processing

Simulated Bifurcation (SB)

➤ **adiabatic Simulated Bifurcation (aSB)**

$$\dot{x}_i = \frac{\partial H_{\mathrm{SB}}}{\partial y_i} = \Delta y_i, \qquad \dot{y}_i = \frac{\partial H_{\mathrm{SB}}}{\partial x_i} = -[K x_i^2 - p(t) + \Delta] x_i + \xi_0 \sum_{j=1}^{N} J_{ij} x_j$$

➤ **ballistic Simulated Bifurcation (bSB)**

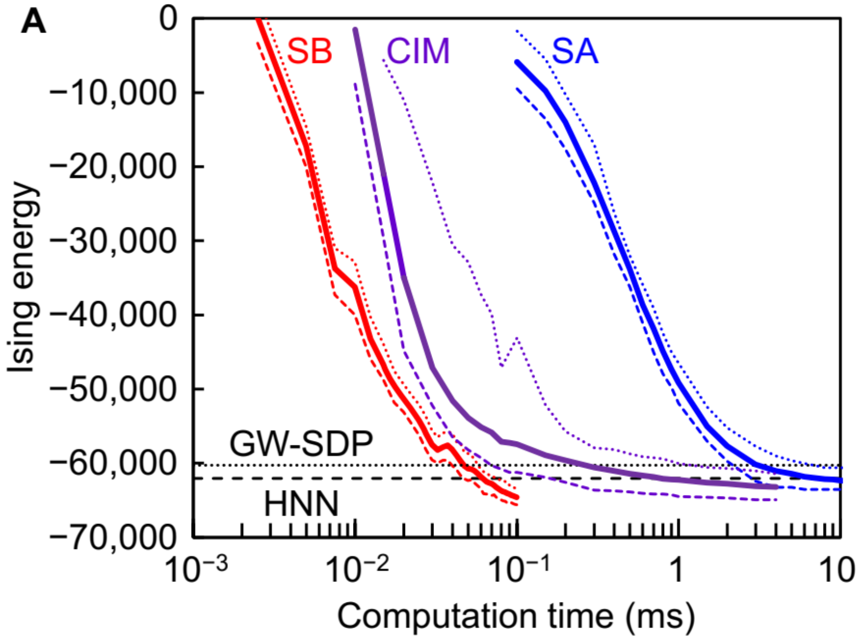
$$\dot{x}_i = \frac{\partial H_{\mathrm{SB}}}{\partial y_i} = \Delta y_i, \qquad \dot{y}_i = \frac{\partial H_{\mathrm{SB}}}{\partial x_i} = (p(t) - \Delta) x_i + \xi_0 \sum_{j=1}^{N} J_{ij} x_j$$

➤ **discrete Simulated Bifurcation (dSB)**

$$\dot{x}_i = \frac{\partial H_{\mathrm{SB}}}{\partial y_i} = \Delta y_i, \qquad \dot{y}_i = \frac{\partial H_{\mathrm{SB}}}{\partial x_i} = (p(t) - \Delta) x_i + \xi_0 \sum_{j=1}^{N} J_{ij} \mathrm{sign}(x_j)$$

M.H. Yung

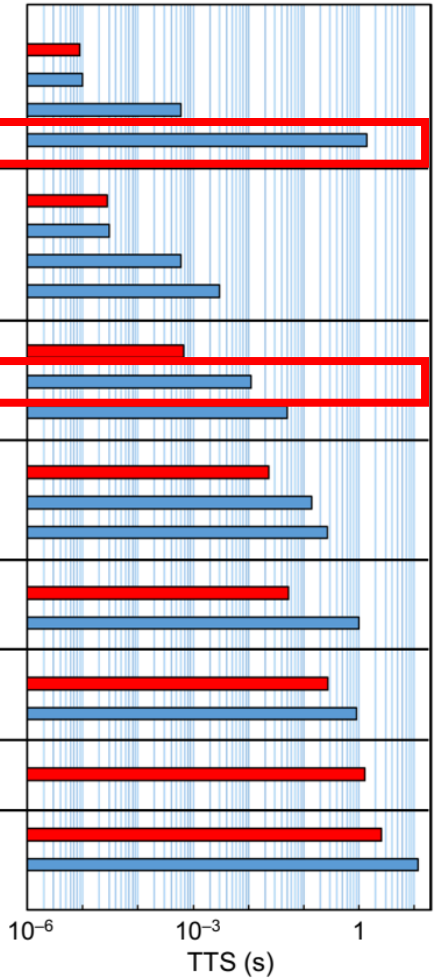
Simulated Bifurcation (SB)



- **Simulated bifurcation is known to outperform other CC algorithms as well as quantum annealing (QA) for some problems**
- Simulated Coherent Ising Machine (SimCIM) had largely degraded performance in our study, so is not presented.

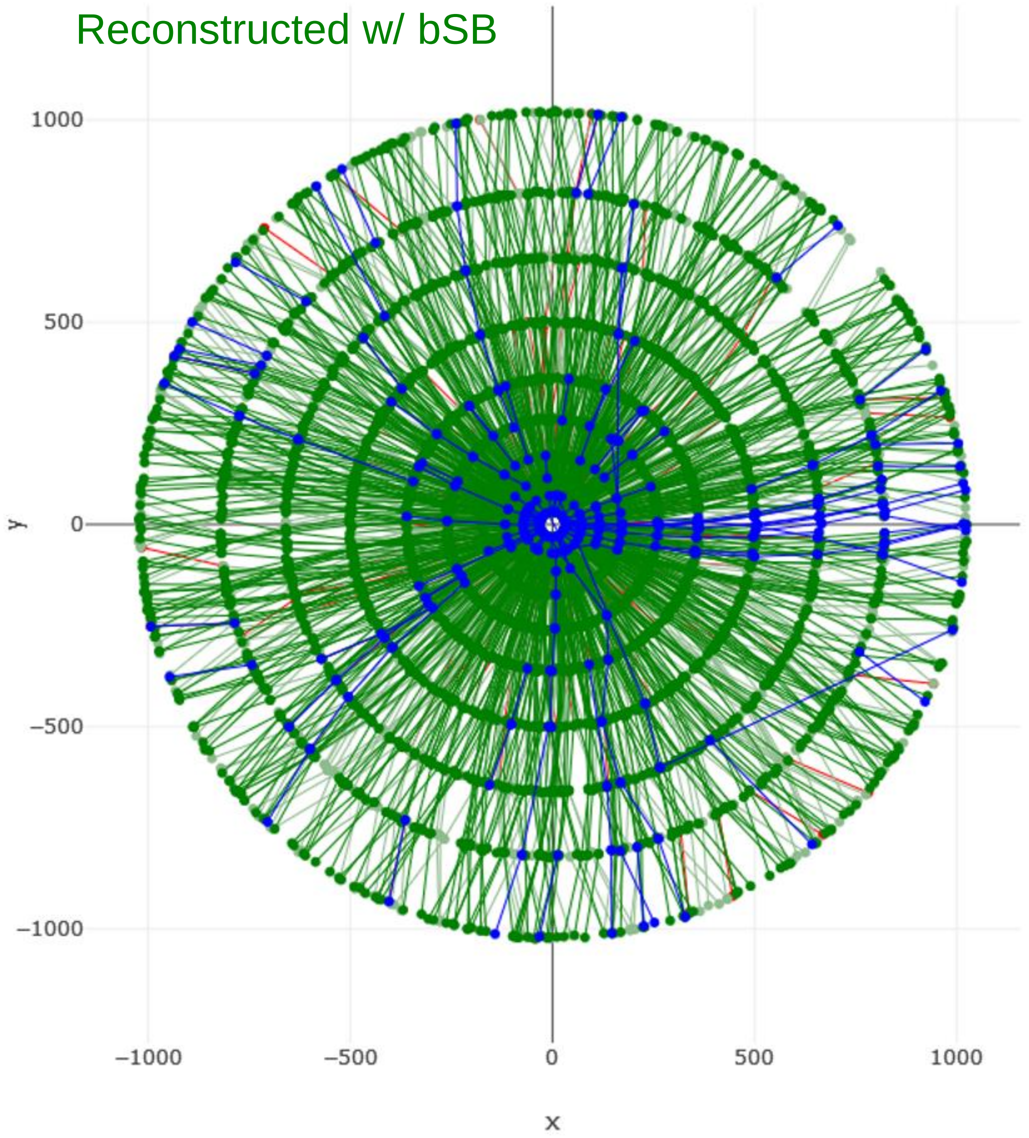
Goto et al., Sci. Adv. 2019; 5: eaav2372
Goto et al., Sci. Adv. 2021; 7: eabe7953

N	Connectivity	$J_{i,j}$	Machine	TTS
60	All-to-all	{±1}	**dSBM** RBM CIM QA	**9.2 μs** 10 μs 0.6 ms 1.4 s
100	All-to-all	{±1}	**dSBM** RBM SimCIM CIM	**29 μs** 30 μs 0.6 ms 3.0 ms
200	Sparse (Degree 3)	{0, −1}	**dSBM** QA CIM	**0.70 ms** 11 ms 51 ms
700	All-to-all	{±1}	**dSBM** SimCIM DA	**25 ms** 0.14 s 0.27 s
1024	All-to-all	{±1}	**dSBM** DA	**55 ms** 1 s
1024	All-to-all	16 bits {$-2^{15}+1,\ldots,2^{15}-1$}	**dSBM** DA	**0.29 s** 0.9 s
2000 (K_{2000})	All-to-all	{±1}	**dSBM**	**1.3 s**
2000 (G22)	Sparse (1%)	{0, −1}	**dSBM** SimCIM	**2.7 s** 12 s



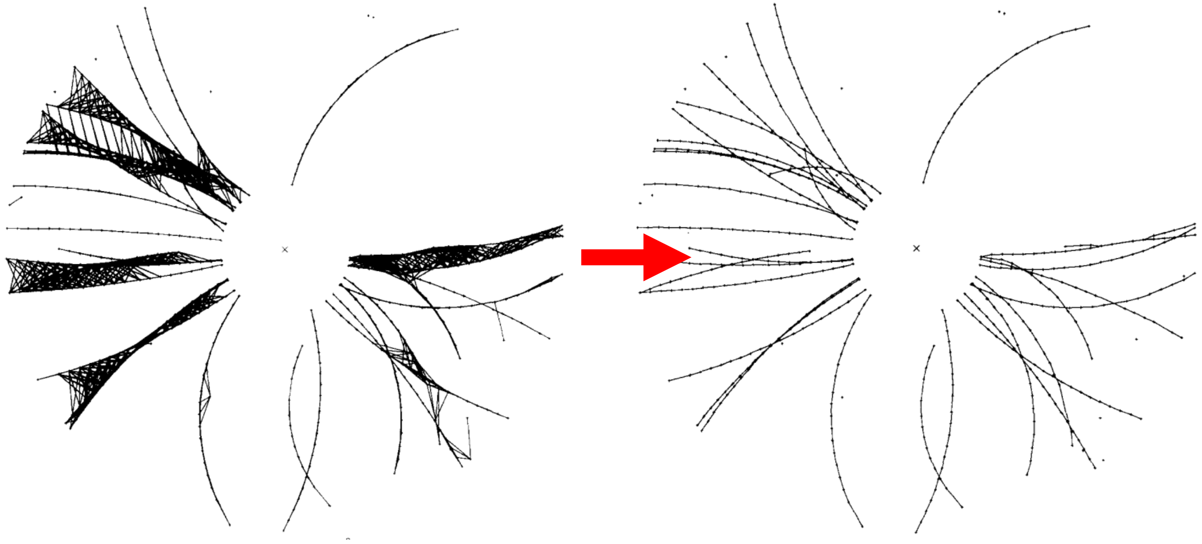
Tracking Studies

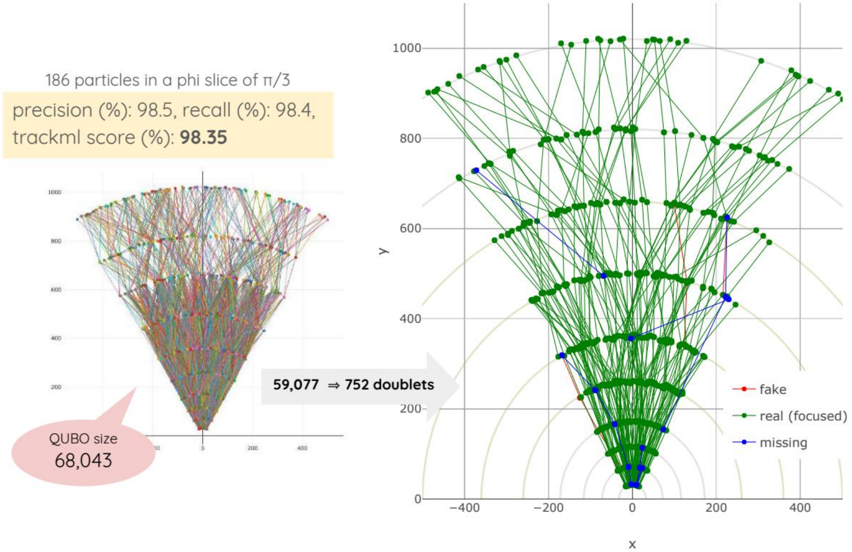
H. Okawa, Q.-G. Zeng, X.-Z. Tao, M.-H. Yung,
arXiv:2402.14718 (2024)



Tracking as Optimization Problem

- **Tracking as an optimization problem: a global approach to reconstruct tracks in one-go.** (↔iterative approach: Combined Kalman Filter)
- **Stimple-Abele & Garrido (1990):** generate all potential doublets with some cuts applied & pursue a binary classification task (i.e. solve an Ising/QUBO problem) to determine which ones should be kept.
- **Modern quantum computing versions:** quantum annealers w/ doublets (A. Zlokapa et al.) & triplet-based (F. Bapst et al.) approaches; quantum gate machines (L. Funcke et al., etc.; **H.Okawa**)





QUBO Formulation w/ Triplets

- Tracks are formed by connecting silicon detector hits: e.g. triplets (segments w/ 3 hits).
- Doublets/triplets are connected to reconstruct tracks & it can be regarded as a **quadratic unconstrained binary optimization (QUBO)** problem.

$$O(a,b,T) = \underbrace{\sum_{i=1}^{N} a_i T_i}_{\text{Quality of triplets}} + \underbrace{\sum_{i}^{N}\sum_{j<i}^{N} b_{ij} T_i T_j}_{\text{Compatibility b/w triplet pairs}}$$

Quality of triplets

Compatibility b/w triplet pairs

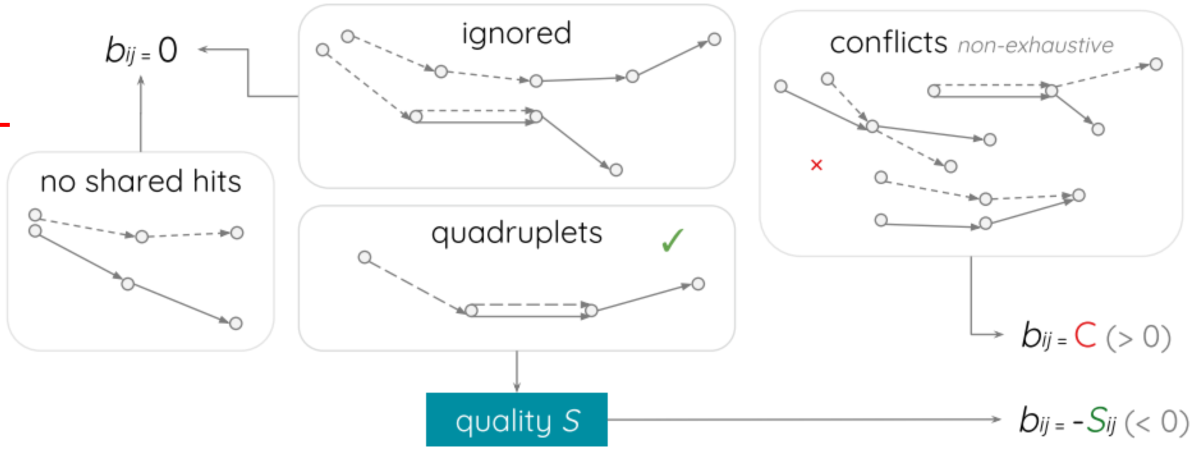
$$a_i = \alpha\left(1 - e^{\frac{|d_0|}{\gamma}}\right) + \beta\left(1 - e^{\frac{|z_0|}{\lambda}}\right),$$

b_{ij} = 0 (if no shared hit), 1 (if conflict)

= $-S_{ij}$ (if two hits are shared)

$$S_{ij} = \frac{1 - \frac{1}{2}(|\delta(q/p_{Ti}, q/p_{Tj})| + max(\delta\theta_i, \delta\theta_j))}{(1 + H_i + H_j)^2},$$

F. Bapst et al. Comp. Soft. Big Sci. 4 (2019) 1.

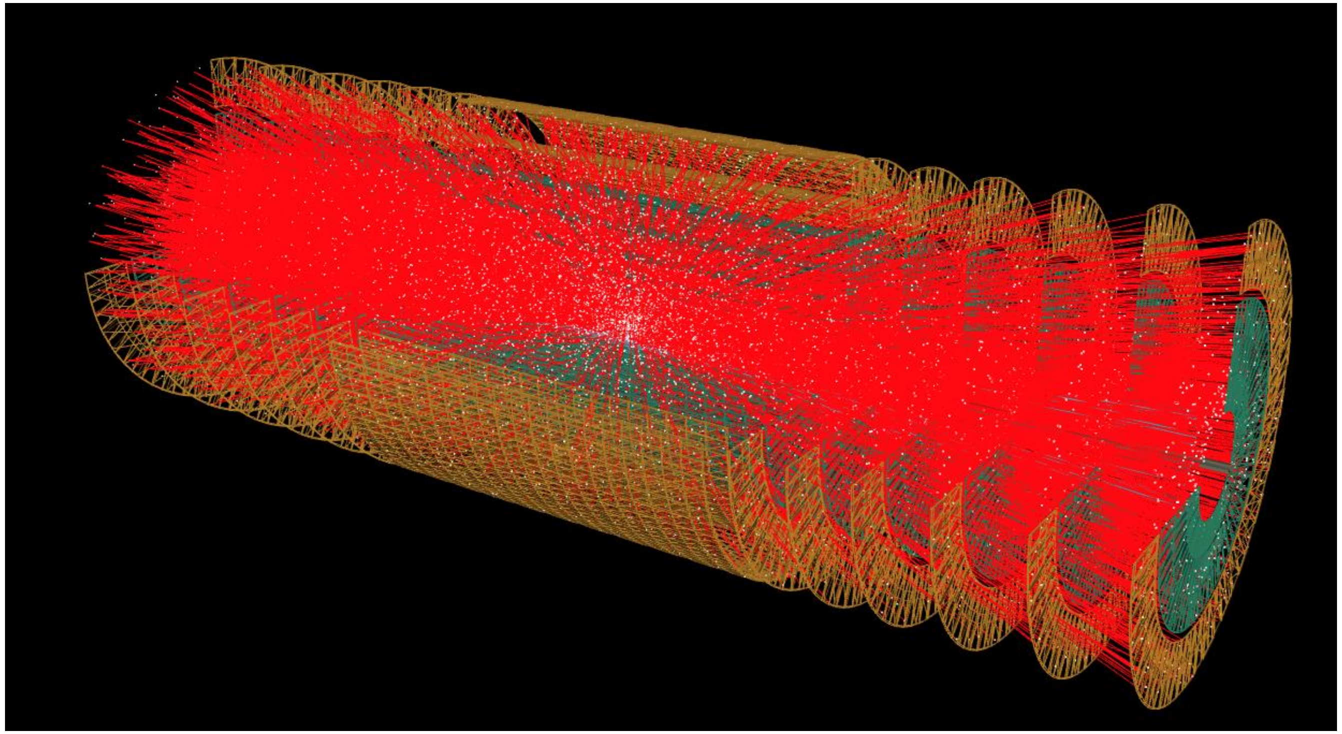


Minimizing QUBO is equivalent to searching for the ground state of the Hamiltonian.

Dataset (TrackML)

- TrackML is an open-source dataset prepared for TrackML Challenges (two competitions hosted by CERN & Kaggle).
- It is **designed w/ HL-LHC conditions (200 pileup) & run w/ fast simulation (e.g. noise, inefficiency, parametrized material effects, etc.)**
- Only tracks w/ p_T>1 GeV in the barrel are considered.
- QUBO is computed event by event using hepqpr-qallse framework.

Amrouche, S., et al., arXiv:1904.06778 (2019);
Amrouche, S., et al., Comput. Softw. Big Sci. 7(1), 1 (2023)

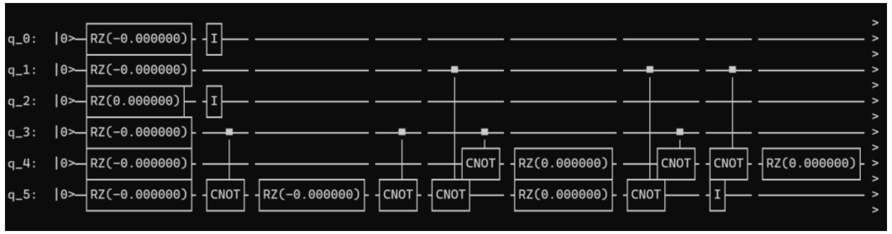


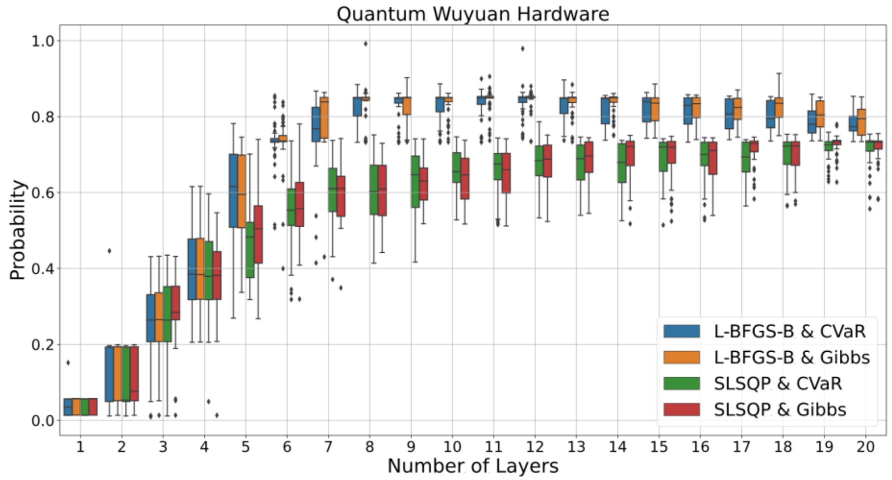
Thanks to Andreas Salzburger for suggestions and discussions!

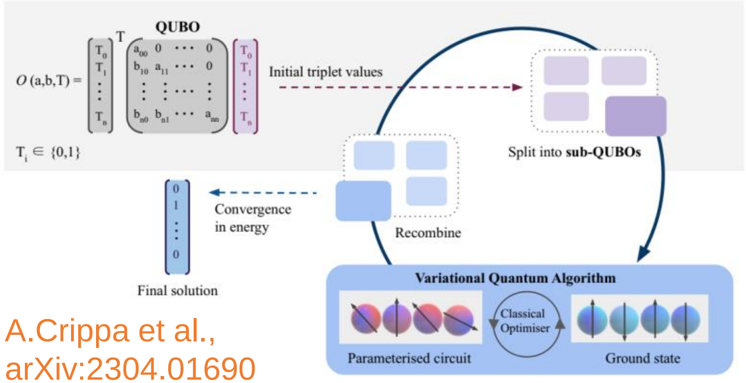
H. Okawa, Springer Communications in Computer and Information Science, 2036 (2024) 272–283, arXiv:2310.10255

Previous Study w/ Quantum Gates

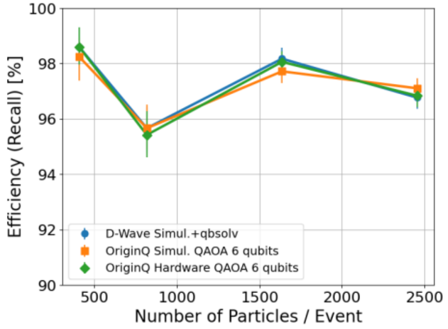
- Thorough optimization of QAOA in terms of # of layers, optimizers & loss functions.
- 6-qubit hardware (Origin Quantum Wuyuan) & simulator are used.
- Used a theoretically robust sub-QUBO method to split the problem into 6-qubit size
- Comparable performance obtained w/ the previous D-Wave studies (F. Bapst et al. Comp. Soft. Big Sci. 4 (2019) 1)

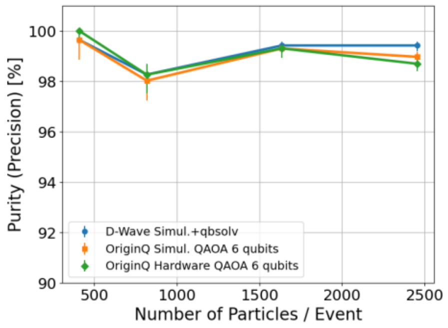




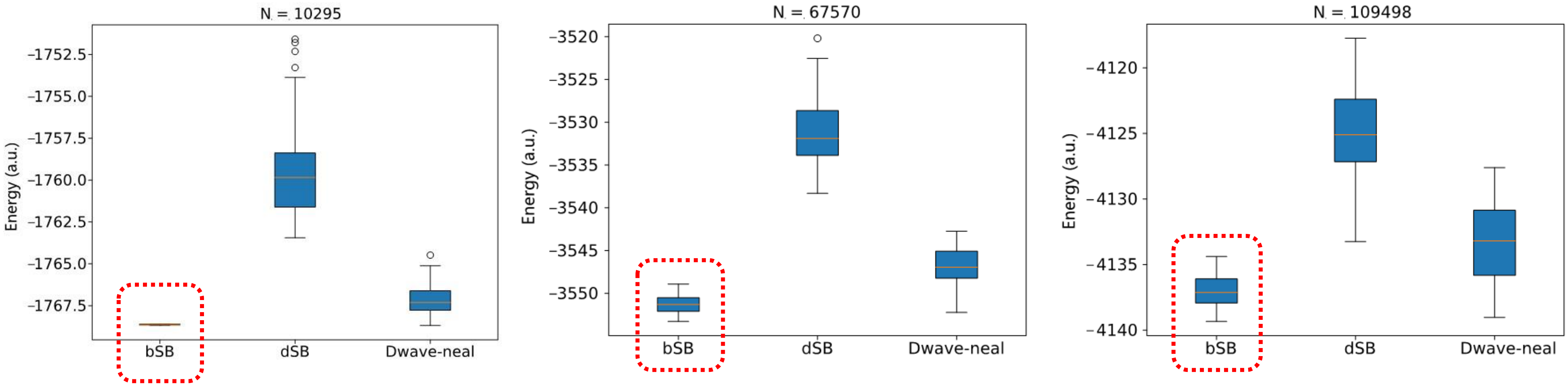


A.Crippa et al., arXiv:2304.01690



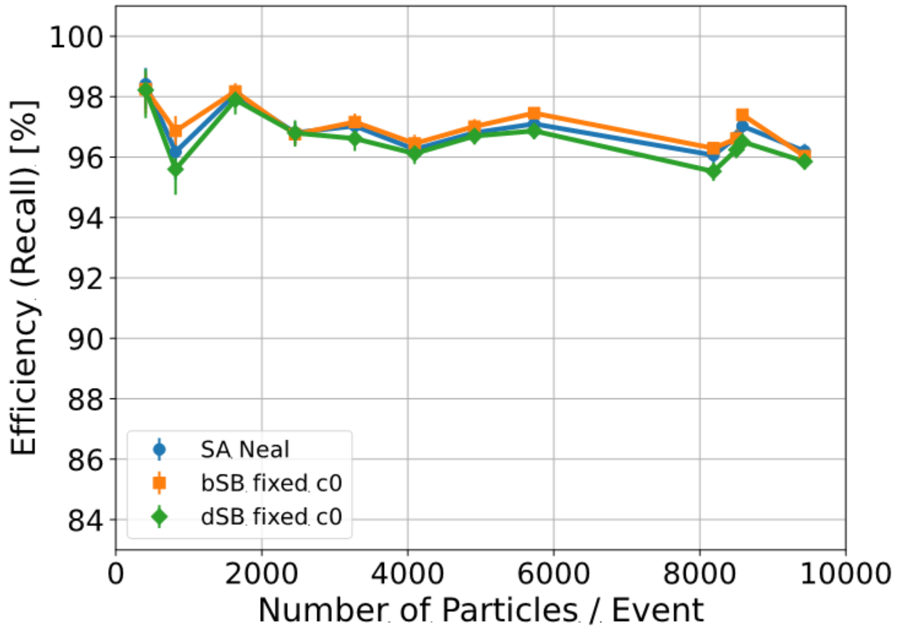


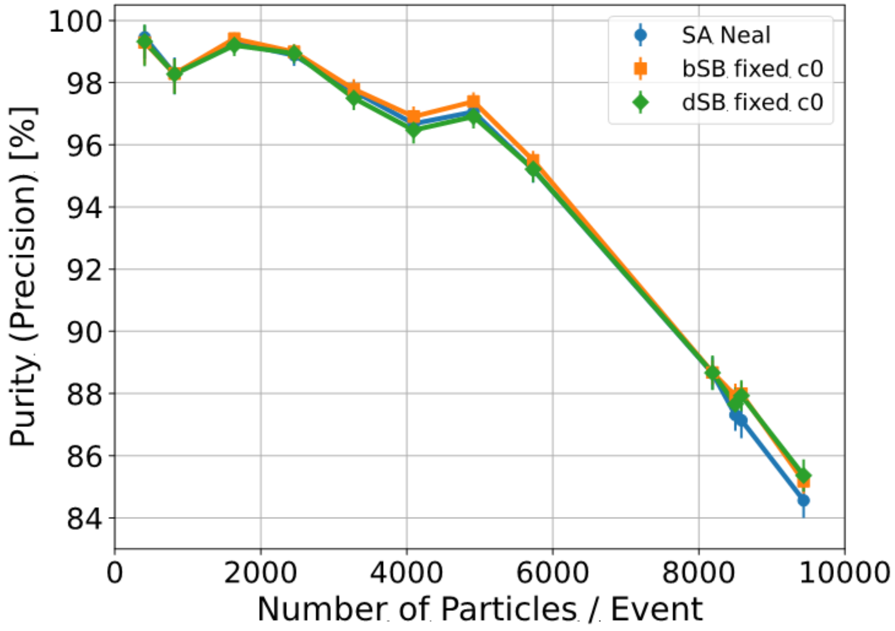
Ising Energy w QAIAs



- **Ballistic simulated bifurcation can find <u>the lowest Ising energy with the smallest fluctuation</u>** for all events considered.
- Discrete simulated bifurcation provides slightly degraded energy prediction to bSB & D-Wave Neal, though the impact on the track reconstruction performance is not significant (see next slide).

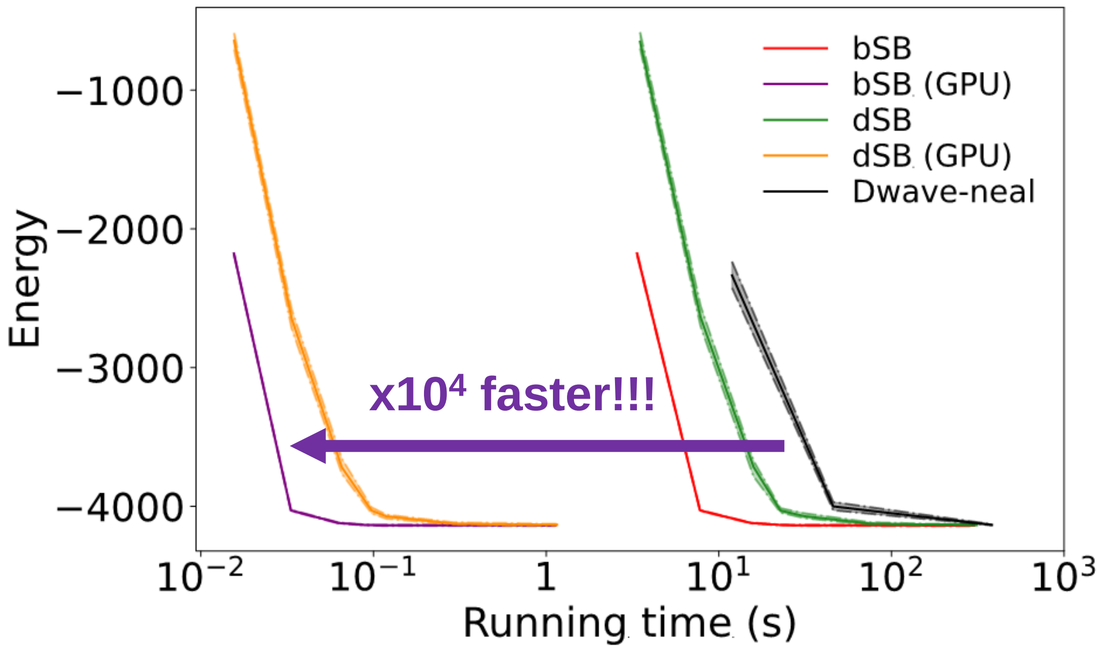
Track Efficiency & Purity w/ QAIAs





- Simulated bifurcation provides **comparable or slightly better performance than D-Wave Neal.**
- **Track efficiency stays over 95%** for all dataset up to the highest HL-LHC conditions
- Purity degrades with track multiplicity but **>90% for <6000 particles, >84% even for ~10000 particles**.

Computation Speed



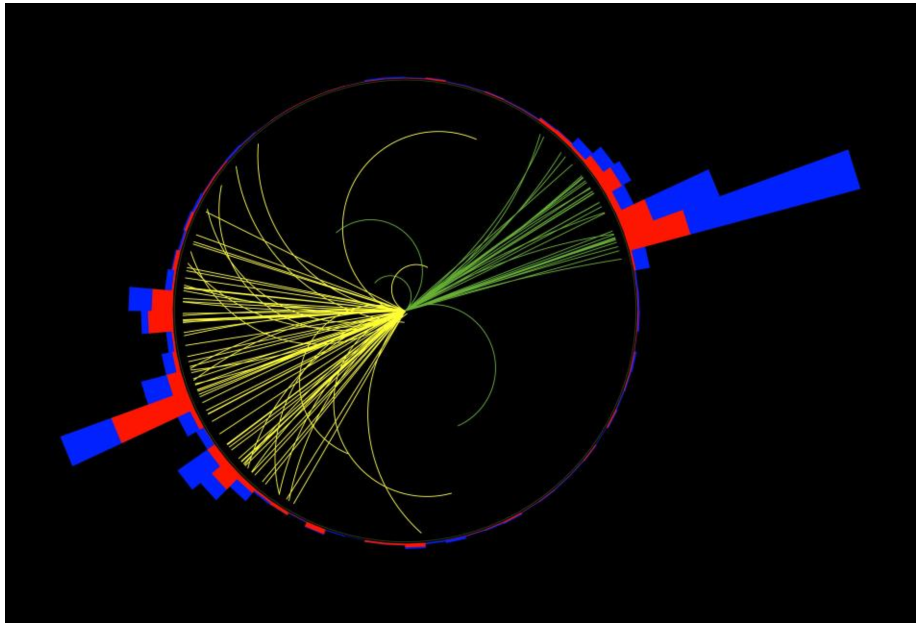
Only 1 CPU/GPU used respectively

Data Information		Time to target [s]				
# of particles	QUBO size	bSB	bSB (GPU)	dSB	dSB (GPU)	D-Wave Neal
409	778	0.007	0.021	0.032	0.092	0.060
818	1431	0.012	0.019	0.293	0.478	0.169
1637	2904	0.012	0.019	0.293	0.478	0.169
2456	4675	0.014	0.017	–	–	0.479
3274	6945	0.032	0.022	–	–	1.229
4092	10295	0.005	0.022	0.015	0.065	0.030
4912	14855	0.027	0.016	–	–	2.165
5730	22022	0.109	0.042	–	–	3.853
8187	67570	0.488	0.028	–	–	404.297
8500	78812	1.899	0.108	–	–	785.732
8583	80113	1.321	0.067	–	–	93.782
9435	109498	3.884	0.140	–	–	1366.808

- Ballistic simulated bifurcation provides **4 orders of magnitude speed-up (23min → 0.14s)** at most, compared to D-Wave Neal (moreover D-Wave qbsolv is 2 orders of magnitude slower than Neal).
  → More speed-up expected with larger data size.
- Unlike D-Wave Neal, **simulated bifurcation can effectively run w/ multiple processing & GPU → Perfect match with HEP computing environment!!**

Jet Reconstruction

H. Okawa, X.-Z. Tao, Q.-G. Zeng, M.-H. Yung,
paper in preparation

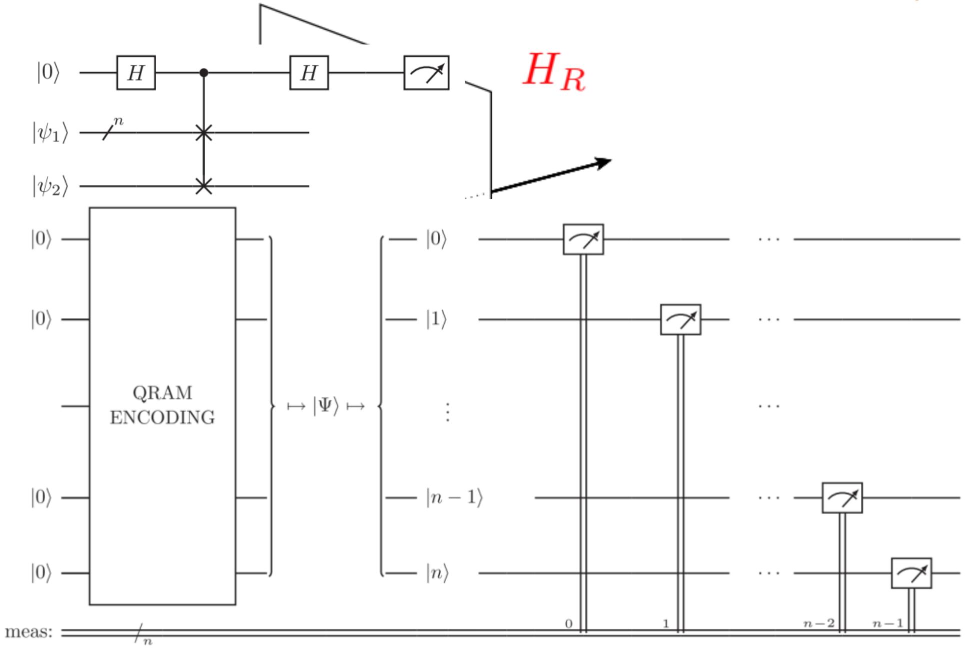


Existing Studies (Iterative Methods)

- Jet reconstruction is a clustering problem. Quantum algorithms may bring in acceleration.
- A few algorithms were considered to replace the traditional iterative calculation. Expected to bring in speed-up, but still at a conceptual stage.

Grover search

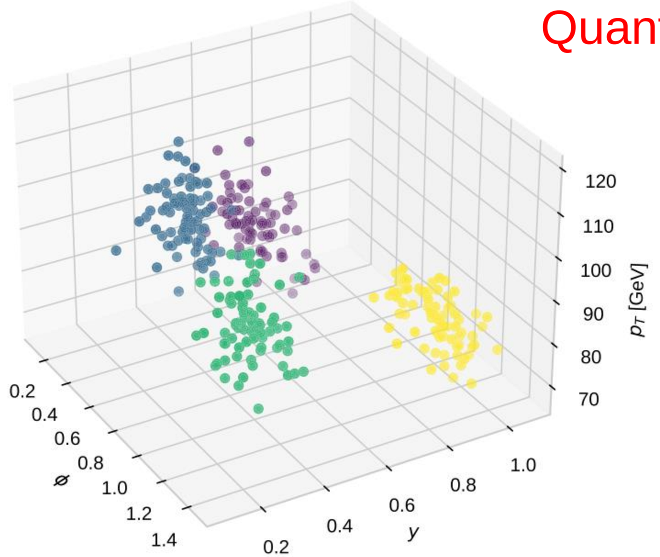
J.J. Martinez de Lejarza, L. Cieri, G. Rodrigo, PRD 106 036021 (2022),
A. Wei, P. Naik, A.W. Harrow, J. Thaler, PRD 101, 094015 (2020)

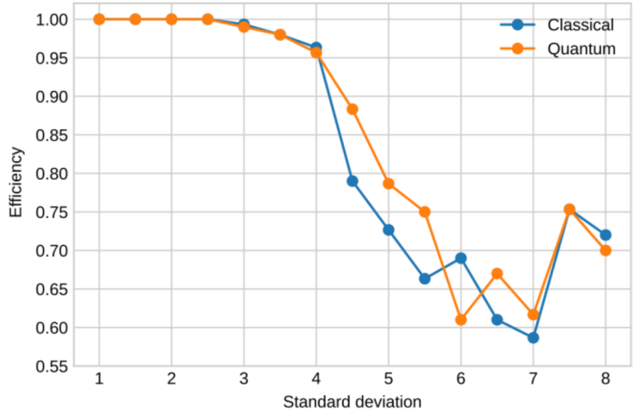


Quantum K-means, quantum affinity propagation (AP), quantum k_t

J.J. Martinez de Lejarza, L. Cieri, G. Rodrigo, PRD 106 036021 (2022),
D. Pires, P. Bargassa, J. Seixas, Y. Omar, arXiv:2101.05618 (2021).

Quantum K-means

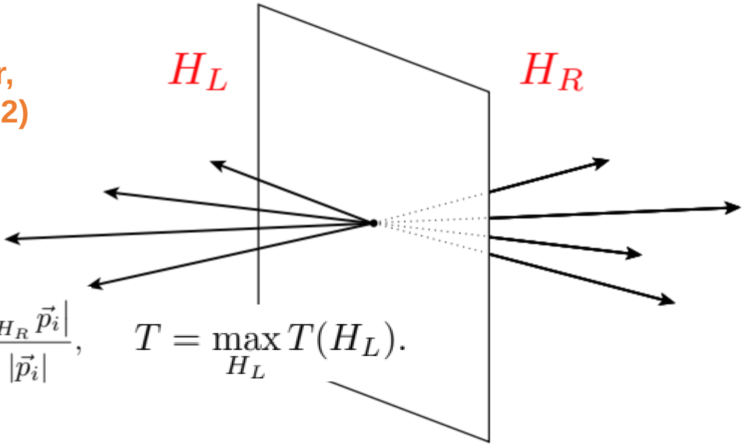




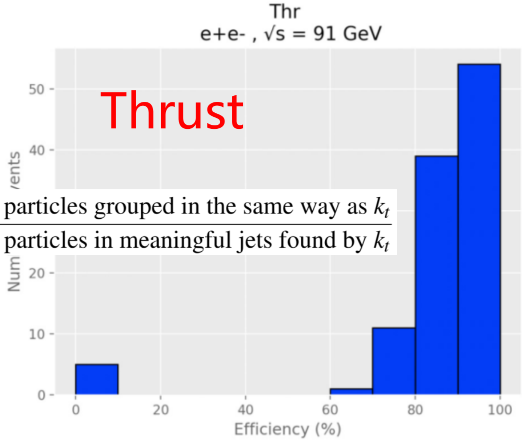
Existing Studies (Global Methods)

Quantum Annealing (Thrust or Angle-based)

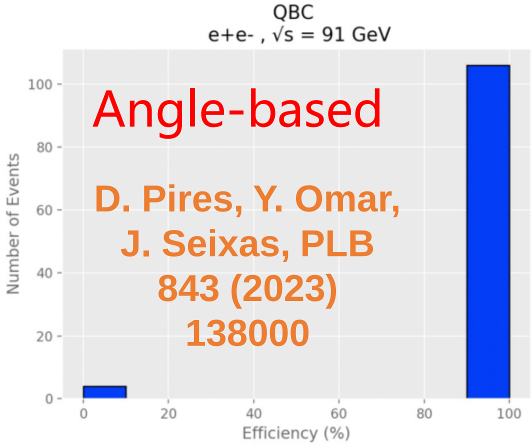
A. Delgado, J. Thaler,
PRD 106, 094016 (2022)



$$T(H_L) = \frac{2\left|\sum_{i\in H_L}\vec{p}_i\right|}{\sum_{i=1}^{N}|\vec{p}_i|} = \frac{2\left|\sum_{i\in H_R}\vec{p}_i\right|}{\sum_{i=1}^{N}|\vec{p}_i|}, \qquad T = \max_{H_L} T(H_L).$$



$$\varepsilon = \frac{\#\text{ of particles grouped in the same way as } k_t}{\#\text{ of particles in meaningful jets found by } k_t}$$



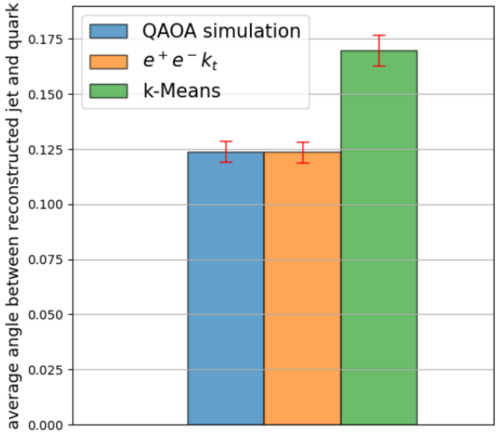
D. Pires, Y. Omar, J. Seixas, PLB 843 (2023) 138000

- Jet reconstruction can also considered as a QUBO problem.
- D. Pires et al.: Angle-based method has better performance than the Thrust-based method, but **does not work for multijet (N_{jet}>2) events so far.**
- Y. Zhu et al.: Used small-size dataset & evaluated average angle w/ QAOA.

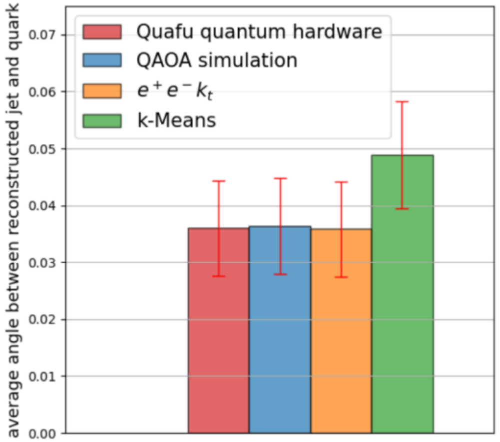
Y. Zhu, W. Zhuang, C. Qian, Y. Ma, D.E. Liu, M. Ruan and C. Zhou, arXiv:2407.09056

Quantum Gates (e.g. QAOA)

30-particle data ($e^+e^- \to ZH \to \nu\nu ss$)



6-particle data ($e^+e^- \to ZH \to \nu\nu ss$)



QUBO Formulation in This Study

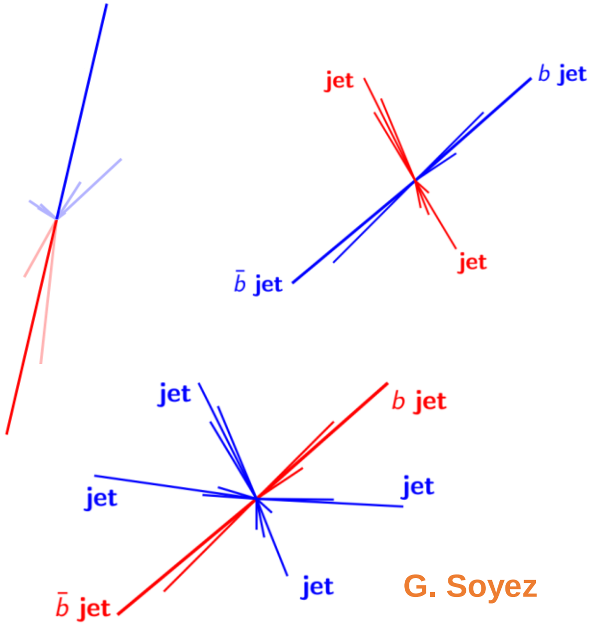
QUBO Formulation

$$O_{\mathrm{QUBO}}^{\mathrm{multijet}}(x_i) = \sum_{n=1}^{n_{\mathrm{jet}}} \sum_{i,j=1}^{N_{\mathrm{input}}} Q_{ij} x_i^{(n)} x_j^{(n)} + \lambda \sum_{i=1}^{N_{\mathrm{input}}} \left(1 - \sum_{n=1}^{n_{\mathrm{jet}}} x_i^{(n)}\right)^2 ,$$

$$Q_{ij} = 2\min(E_i^2, E_j^2)(1 - \cos\theta_{ij}).$$ **[ee-k_t distance]**

$$Q_{ij} = -\frac{1}{2}\cos\theta_{ij}$$ **[angle-based]**

D. Pires, Y. Omar, J. Seixas, PLB 843 (2023) 138000

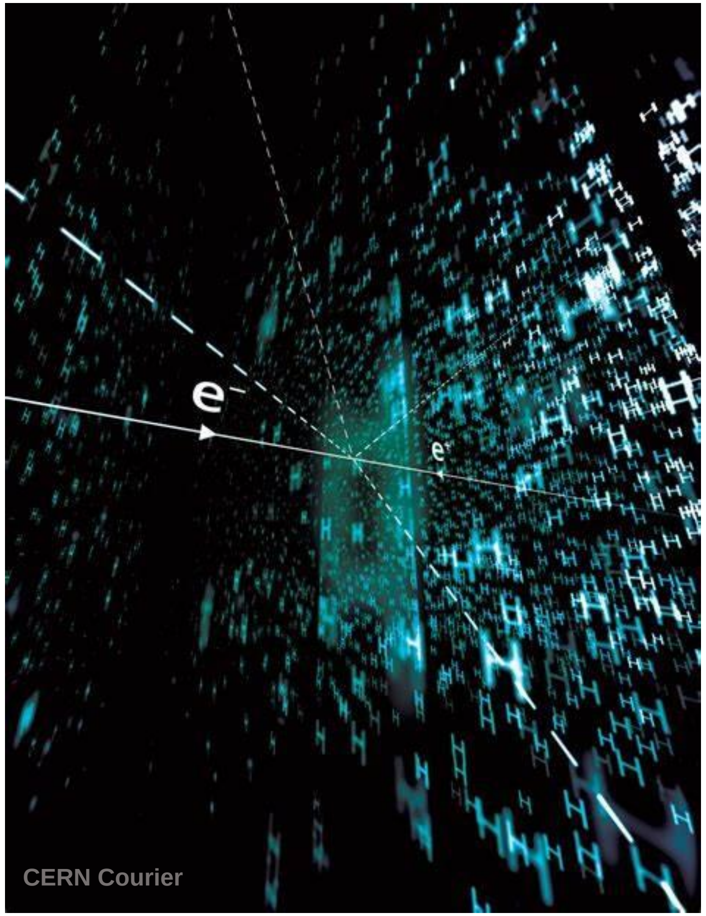


G. Soyez

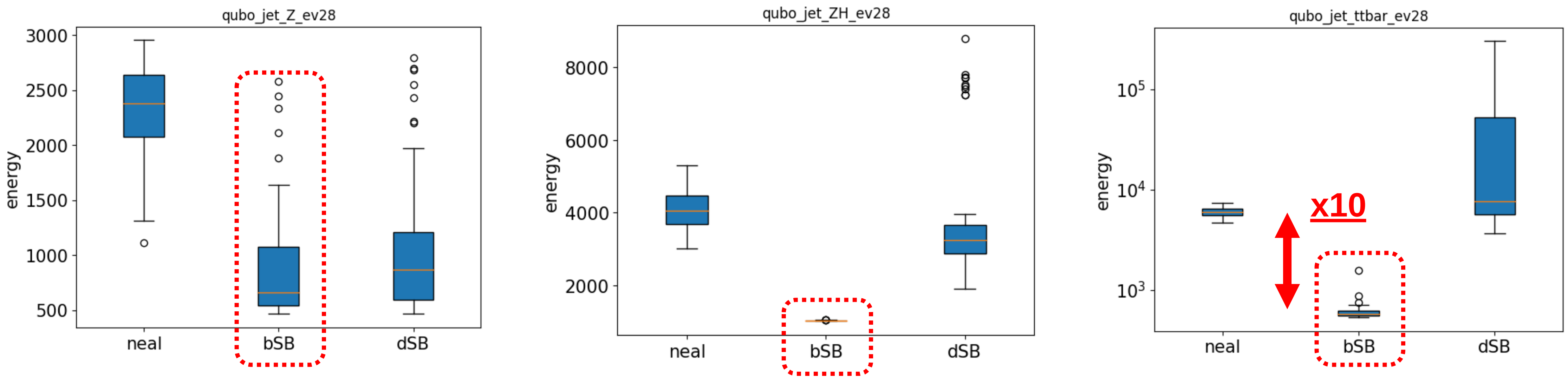
- **Exclusive jet finding with the ee-k_t algorithm** is the baseline at CEPC & other e+e- future Higgs factories.
- **We adopt the same distance in the QUBO formulation. <u>QUBO is designed for general jet multiplicity beyond dijet.</u>**
- Performance is also compared with the angle-based method from a previous study.

Dataset

- Three sets of e+e- collision events are generated to consider various jet multiplicity: $Z \rightarrow q\bar{q}$ (√s=91 GeV, 2 jets), $ZH \rightarrow q\bar{q}b\bar{b}$ (√s=240 GeV, 4 jets), $t\bar{t} \rightarrow b\bar{b}q\bar{q}\bar{q}q$ (√s=360 GeV, 6 jets)
- **Delphes card with the CEPC 4th-detector concept** is used for the fast simulation.
  → Thanks to Gang Li, Shudong Wang and Xu Gao for feedback!
- Jets are reconstructed from the particle flow candidates.



Minimum Ising Energy

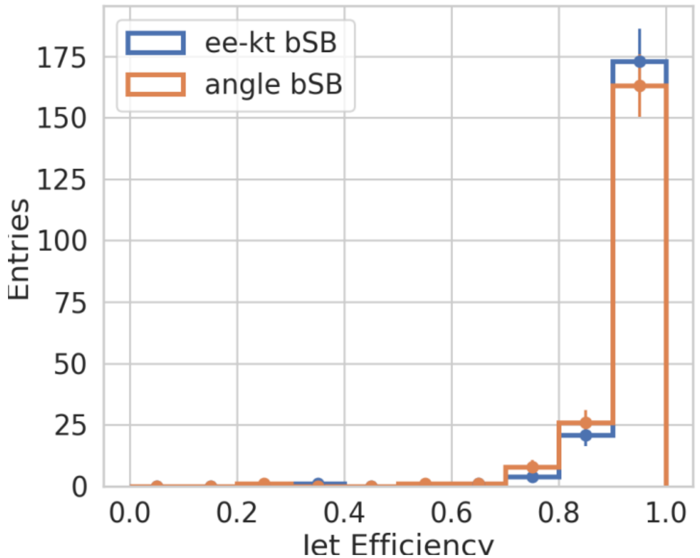


- QUBO for jet reconstruction is fully connected unlike the track reconstruction QUBO, which is largely sparse.
- Ballistic simulated bifurcation finds the lowest energy with the smallest fluctuation.
- **Performance is especially outstanding for complex QUBOs → bSB can find x10 lower minimum energy for the ttbar events!**

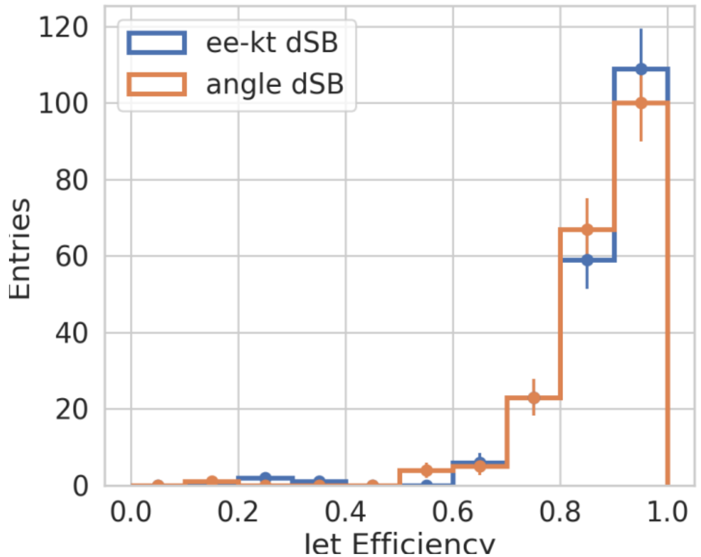
Efficiency (Z→qq: 2 jets)

$$\varepsilon = \frac{\text{\# of particles grouped in the same way as } k_t}{\text{\# of particles in meaningful jets found by } k_t}$$

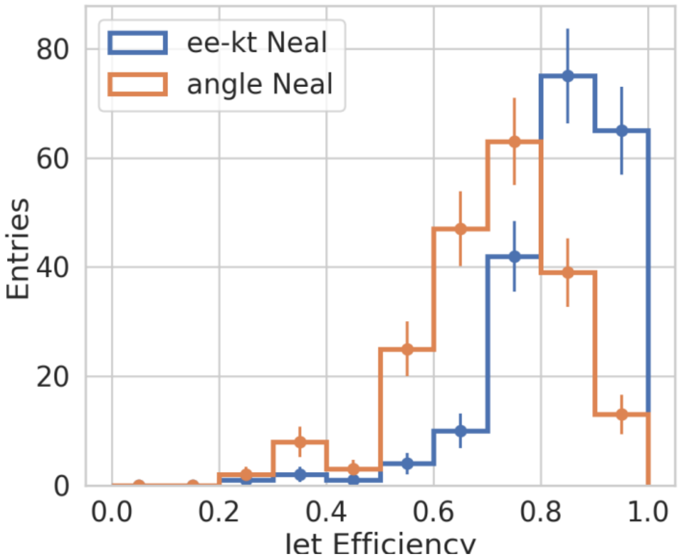
bSB



dSB

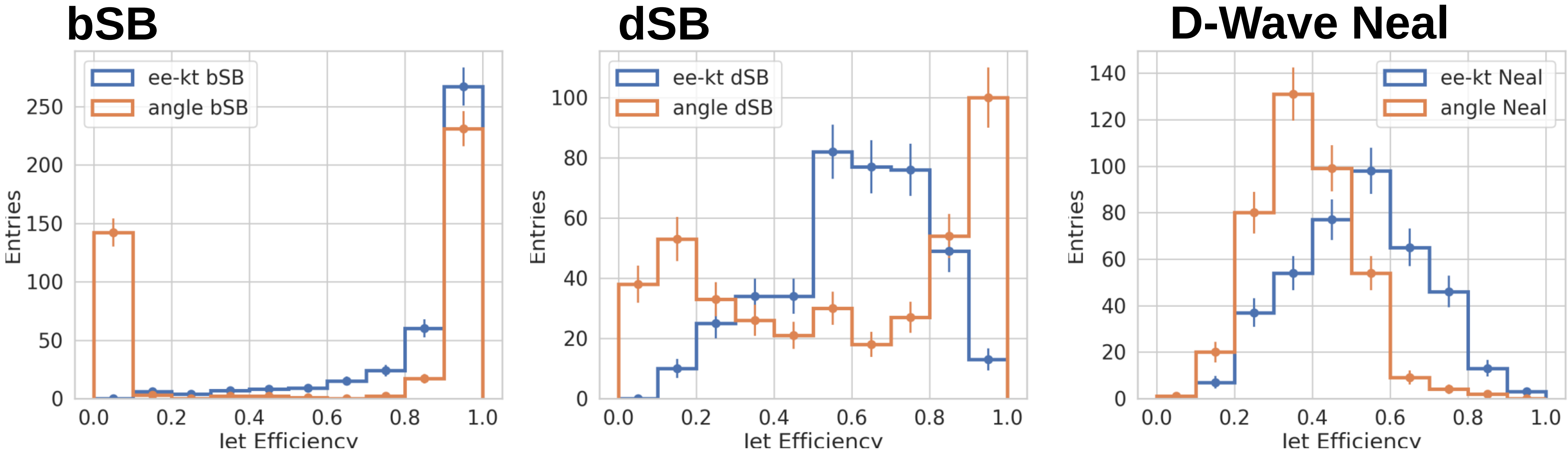


D-Wave Neal



- Most jet reconstruction w/ quantum approaches adopts the above-defined efficiency as performance metric; i.e. compatibility of jet assignment w/ the traditional ee-k_t jet finding.
- **bSB provides the highest efficiency. D-Wave Neal has visibly degraded performance already in dijet events. dSB also has lower efficiency than bSB.**
- **The ee-k_t approach performs better than the angle-based method for all cases.**

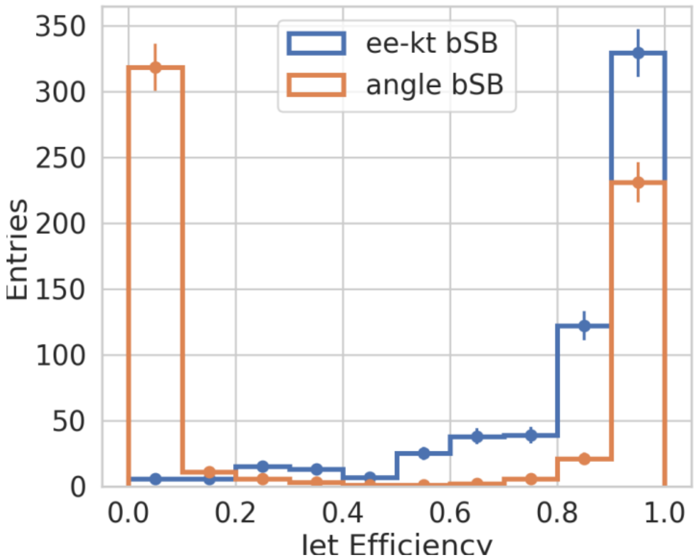
Efficiency (ZH→qqbb: 4 jets)



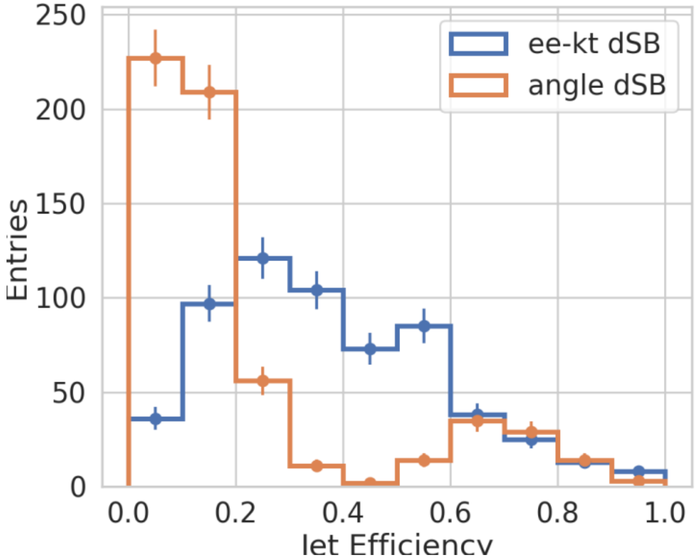
- **Angle-based method does not work for N_{jet}>2**; angles are very likely inappropriate for multijet conditions.
- **dSB & D-Wave Neal cannot reconstruct jets properly regardless of the distance adopted → because of the non-optimal predicted energy**
- **Only bSB maintains reasonable performance. There is still room for further optimization.**

Efficiency (tt→bbqqqq: 6 jets)

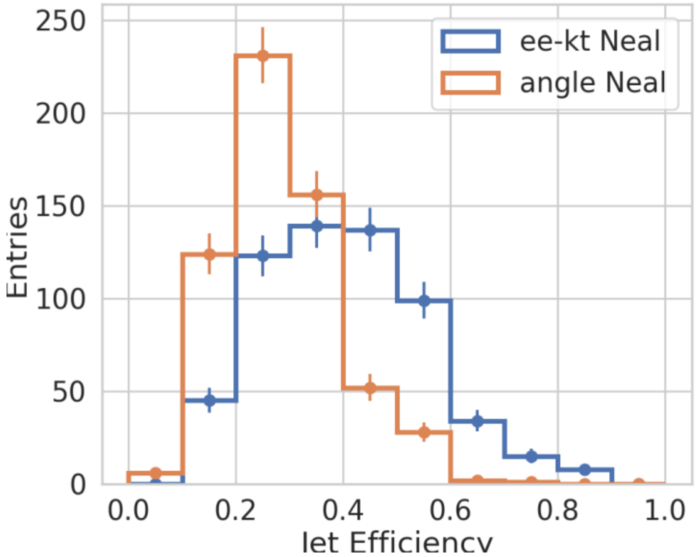
bSB



dSB

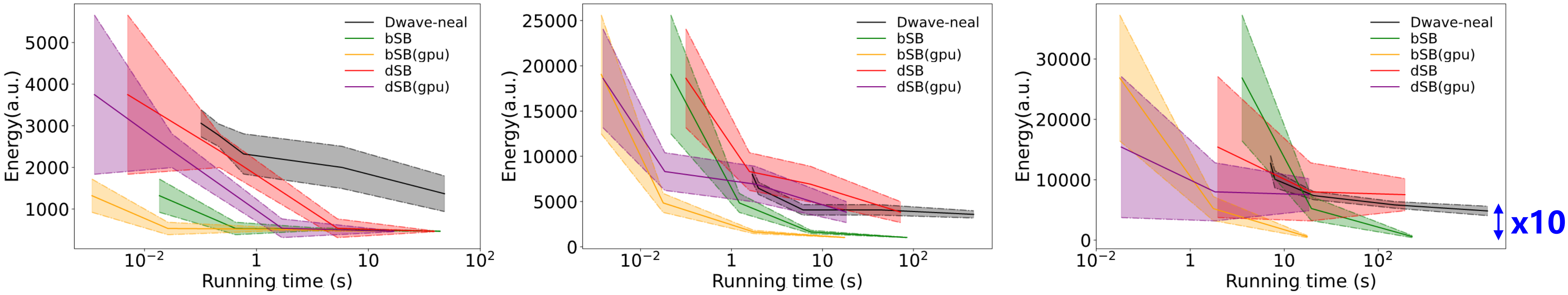


D-Wave Neal



- **Angle-based method does not work for N_{jet}>2**; angles are very likely inappropriate for dense conditions.
- **dSB & D-Wave Neal cannot reconstruct jets properly regardless of the distance adopted → because of the non-optimal predicted energy**
- **Only bSB maintains reasonable performance. There is still room for further optimization.**

Computation Speed

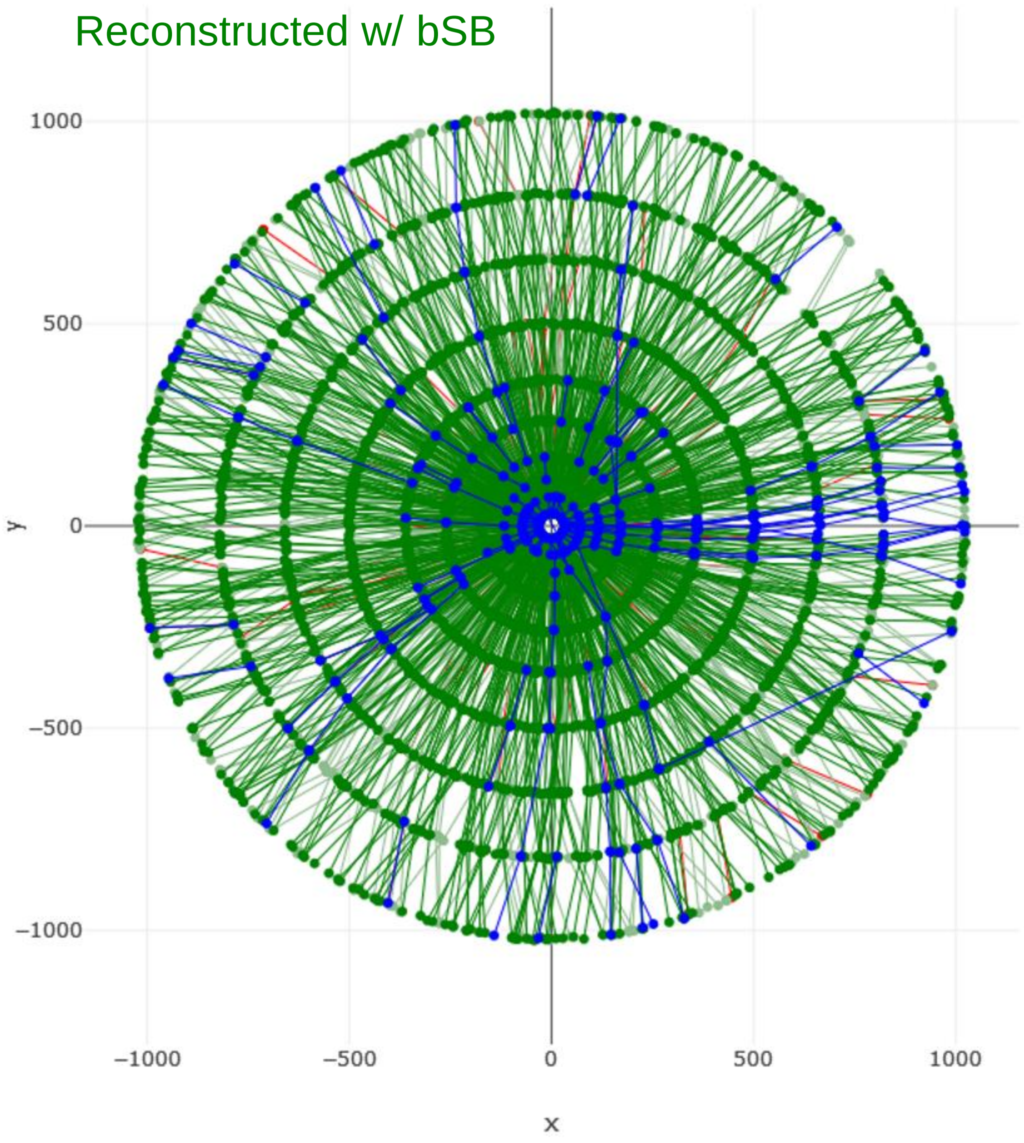


- Time-to-target was evaluated for 3 processes.
- **<u>D-Wave Neal cannot reach reasonable energy, regardless of the running time.</u>**
- **dSB is slow in energy convergence & less successful than bSB for energy prediction.**
- **<u>bSB significantly outperforms dSB & Neal.</u>**

	bSB [s]	bSB (GPU) [s]	dSB [s]	dSB (GPU) [s]	D-neal [s]
Z (eff.=0.9)	0.129	0.073	0.262106	0.188894	5.550219
ZH (eff.=0.5)	2.467985	1.5853	3.948689	2.71392	4.295124
$t\bar{t}$ (eff.=0.5)	3.894546	2.2297	33.05592	21.48662	---

Summary

- Tracking & jet reconstruction are CPU-consuming reconstruction tasks at the LHC & HL-LHC.
- Quantum-annealing-inspired algorithms (QAIAs) are promising approaches for near-term implementations. Ballistic simulated bifurcation (bSB) is particularly quite powerful.
- Presented recent results on track & jet reconstruction w/ QAIA.
- Tracking:
  - bSB can directly handle very large datasets including the densest conditions at the HL-LHC.
  - **bSB provides four orders of magnitude speed-up** at most (& more speed-up expected w/ larger dataset) from D-Wave Neal & can already be considered for implementation.
- Jet reconstruction:
  - Only bSB can predict reasonable energy for jet reconstruction QUBOs.
  - Angle-based QUBO does not work for multijet, but ee-k_t distance QUBO can successfully reconstruct multijet events. **First successful demonstration of multijet reconstruction w/ QUBO.**
  - Further optimization ongoing.



Thank you for listening!

感谢聆听！