Quantum-Annealing-Inspired Algorithms for Future Colliders

The 10th China LHC Physics Conference, November 14-17, 2024

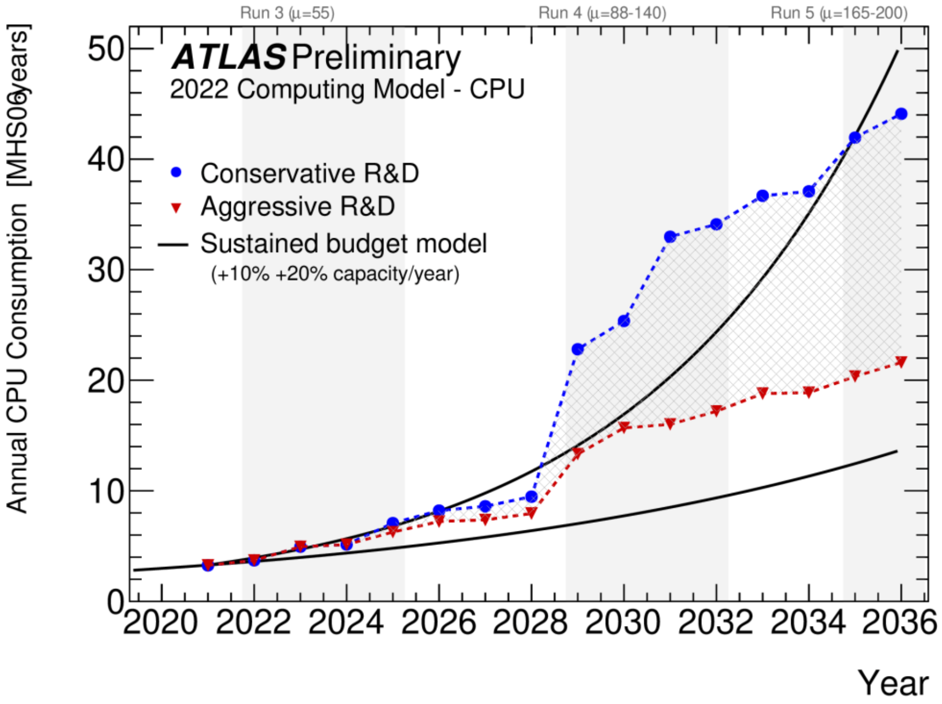
大川(Okawa) 英希(Hideki)

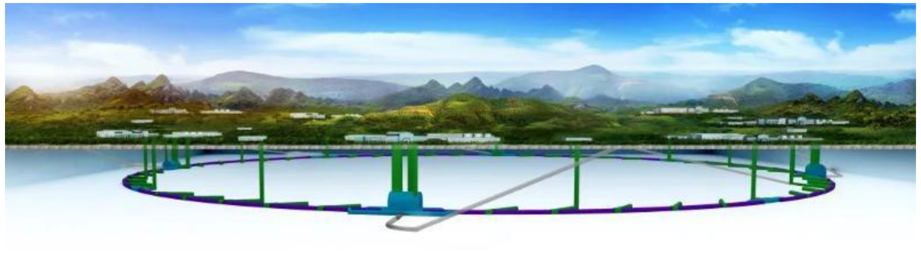
Institute of High Energy Physics, Chinese Academy of Sciences

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

Picture credits: Qianhui Shi, Yan Liang

Reconstruction at Future Colliders

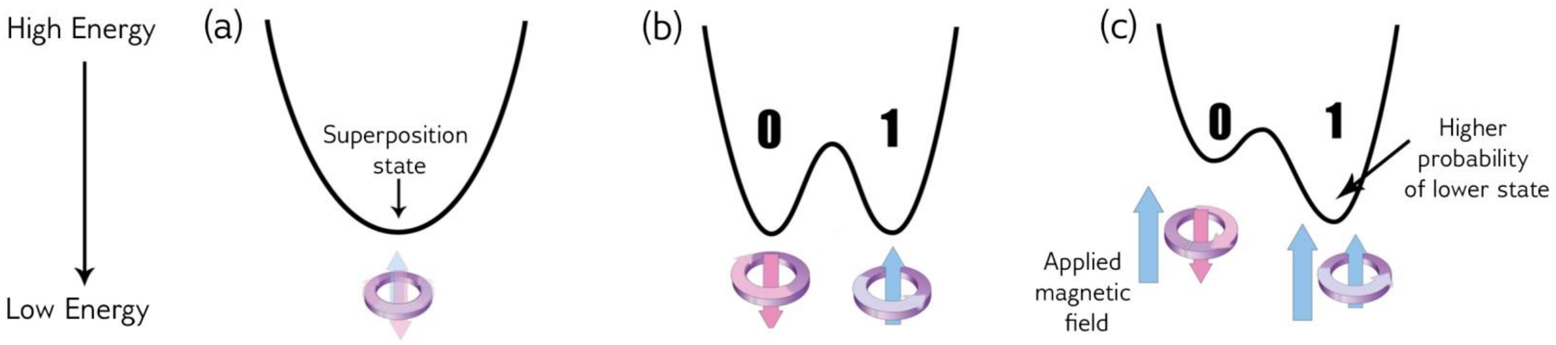




- At HL-LHC & CEPC Z-pole operation, **we will enter the exa-byte era.**
- At the HL-LHC, CPU time exponentially increases with pileup, leading to increase in annual computing cost by x10-20.
- **CEPC Z-pole data taking may experience similar computing challenges.**
- **Along w/ detector simulation, reconstruction is very CPU-consuming.**
- **We may benefit from quantum algorithms.**

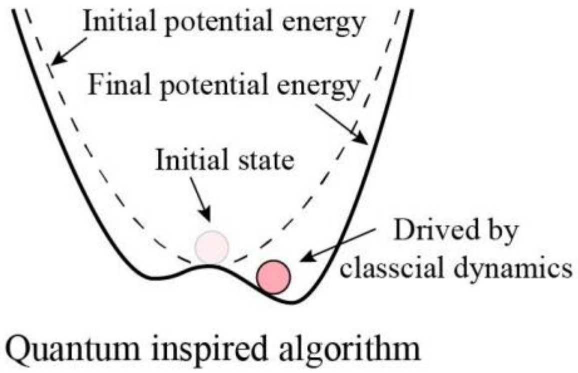
Quantum-Annealing-Inspired Algorithms (QAIAs)

Quantum Annealing



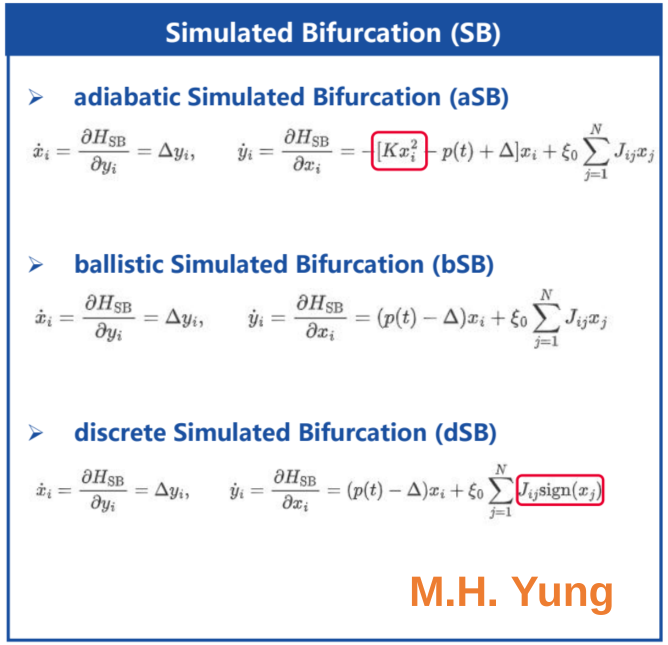
From D-Wave

Quantum-inspired



Quantum inspired algorithm

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



Simulated Bifurcation (SB)

- adiabatic Simulated Bifurcation (aSB)

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

- ballistic Simulated Bifurcation (bSB)

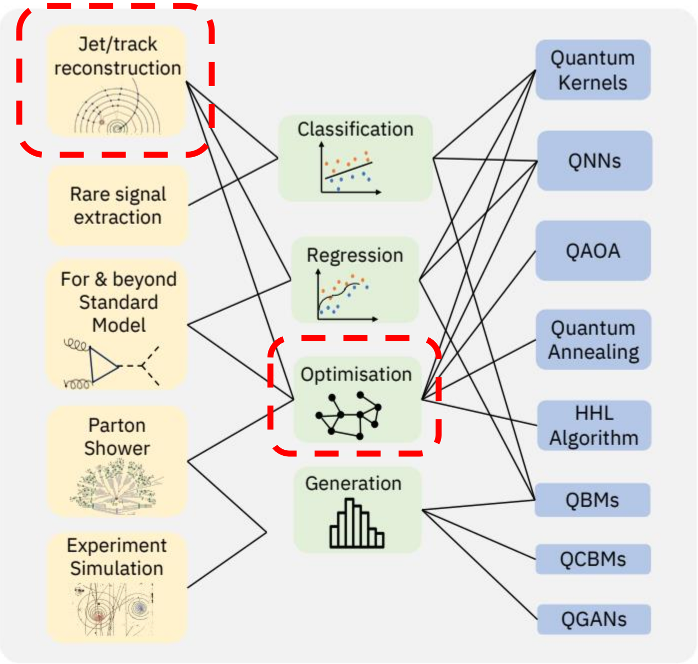
$$\dot{x}_i = \frac{\partial H_{\mathrm{SB}}}{\partial y_i} = \Delta y_i, \qquad \dot{y}_i = \frac{\partial H_{\mathrm{SB}}}{\partial x_i} = (p(t) - \Delta) x_i + \xi_0 \sum_{j=1}^{N} J_{ij} x_j$$

- discrete Simulated Bifurcation (dSB)

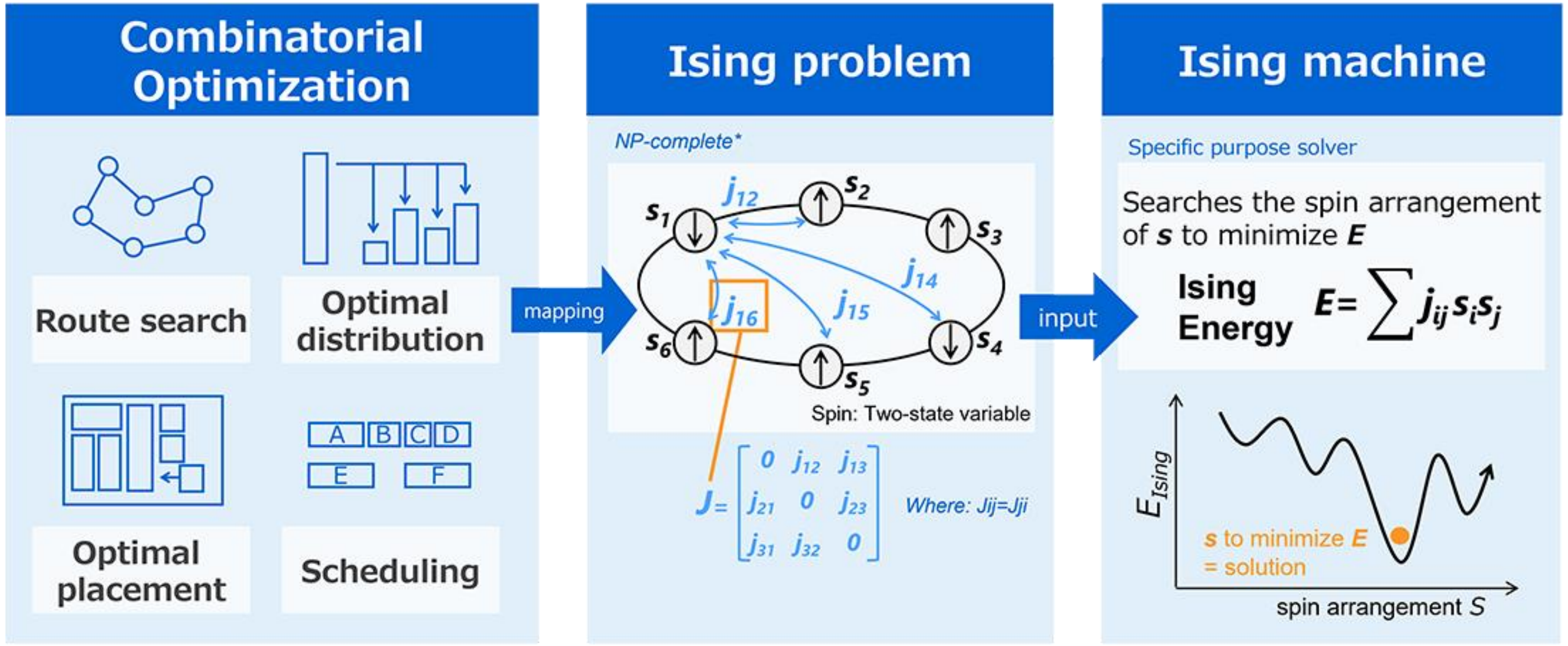
$$\dot{x}_i = \frac{\partial H_{\mathrm{SB}}}{\partial y_i} = \Delta y_i, \qquad \dot{y}_i = \frac{\partial H_{\mathrm{SB}}}{\partial x_i} = (p(t) - \Delta) x_i + \xi_0 \sum_{j=1}^{N} \boxed{J_{ij} \mathrm{sign}(x_j)}$$

M.H. Yung

Combinatorial Optimization Problem



QC4HEP



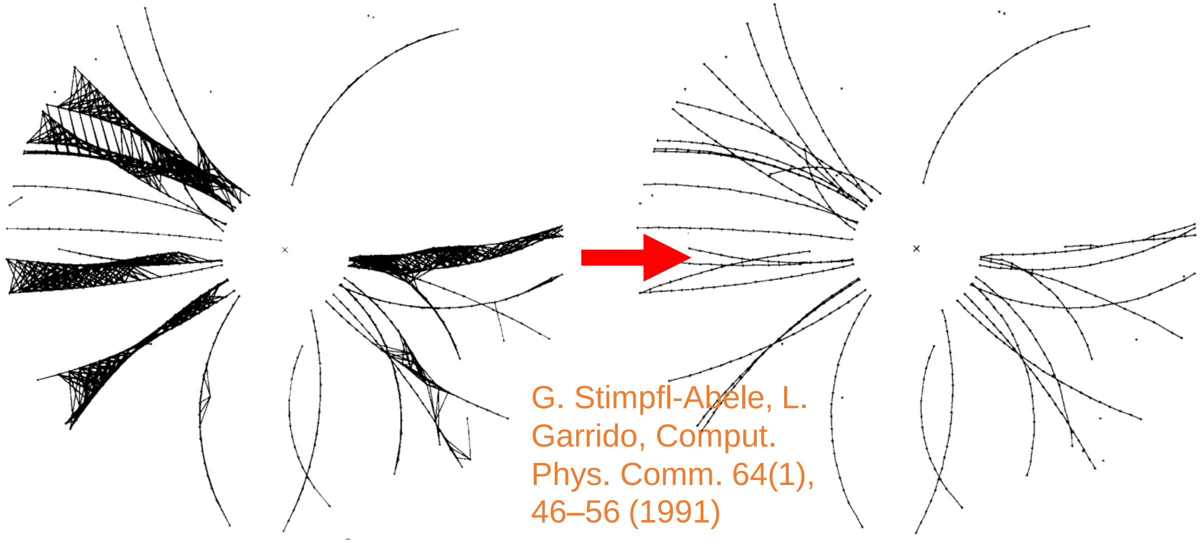
©TOSHIBA

- 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 (or quadratic unconstrained binary optimization; QUBO) problems**. The ground state of an Ising Hamiltonian/QUBO is designed to provide the answer.

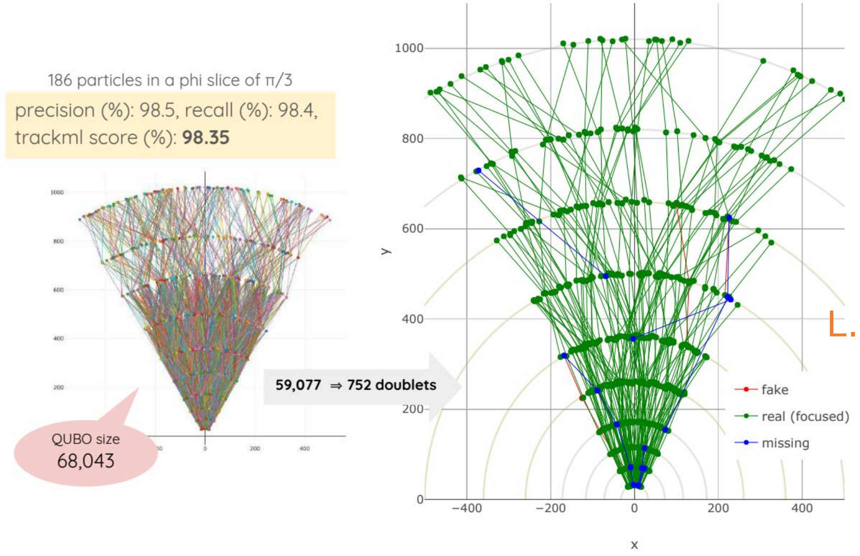
  Ising [±1 spins], QUBO [0/1 binaries]
- **Track & jet reconstruction can also be formulated as Ising/QUBO problems.**

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., H. Okawa, etc.)



G. Stimpfl-Abele, L. Garrido, Comput. Phys. Comm. 64(1), 46–56 (1991)



L. Linder

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.

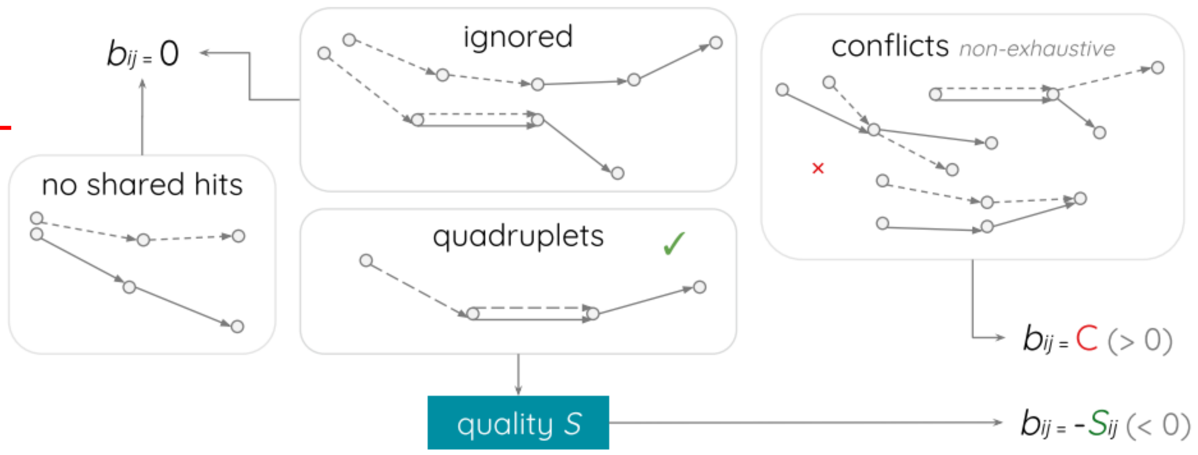
$$O(a,b,T) = \underbrace{\sum_{i=1}^{N} a_i T_i}_{\text{Quality of triplets}} + \underbrace{\sum_{i}^{N} \sum_{j<i}^{N} b_{ij} T_i T_j}_{\text{Compatibility b/w triplet pairs}}$$

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

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.

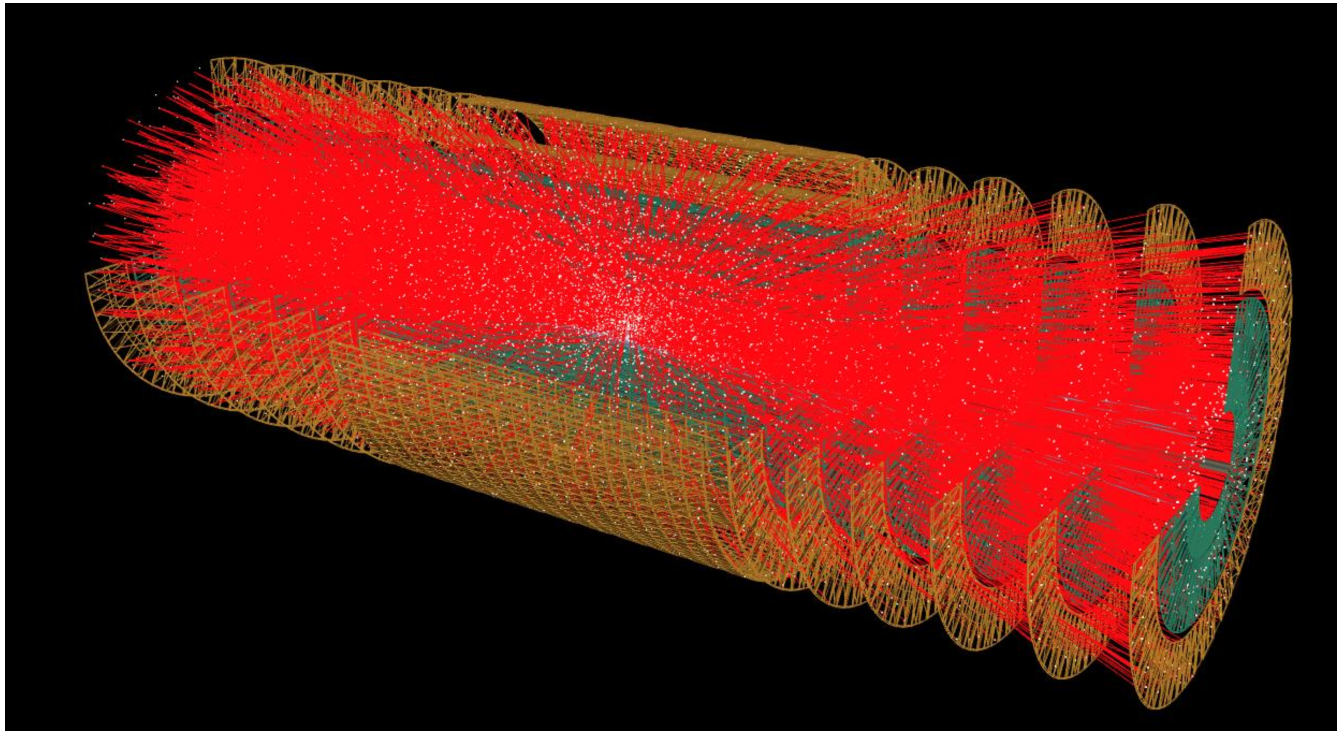


Minimizing QUBO is equivalent to searching for the ground state of the Hamiltonian.

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)

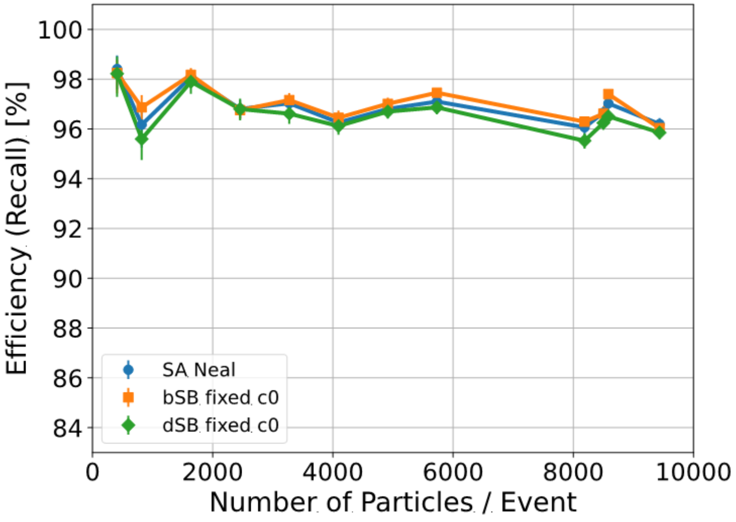


Thanks to Andreas Salzburger for suggestions and discussions!

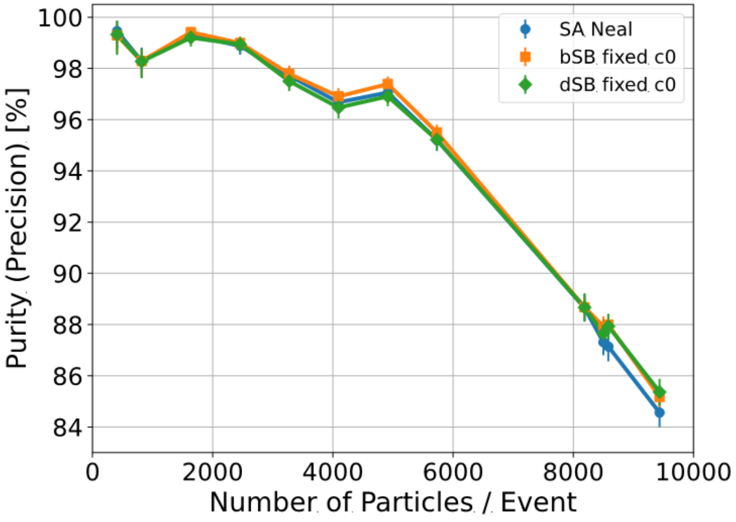
H. Okawa, Q.G. Zeng, X.Z. Tao, M.H. Yung, Springer Comput. Softw. Big Sci. 8, 16 (2024)

Reco. Performance & Comp. Speed

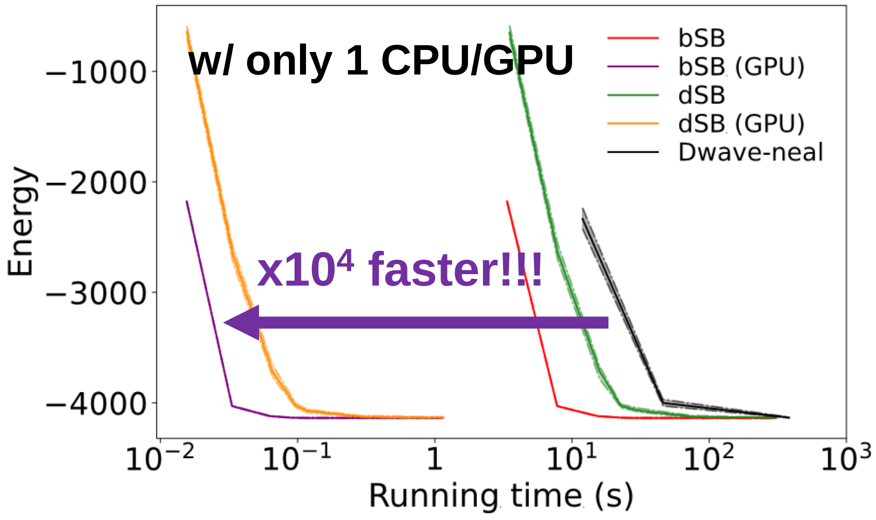
Efficiency



Purity



Ising Energy vs computing time

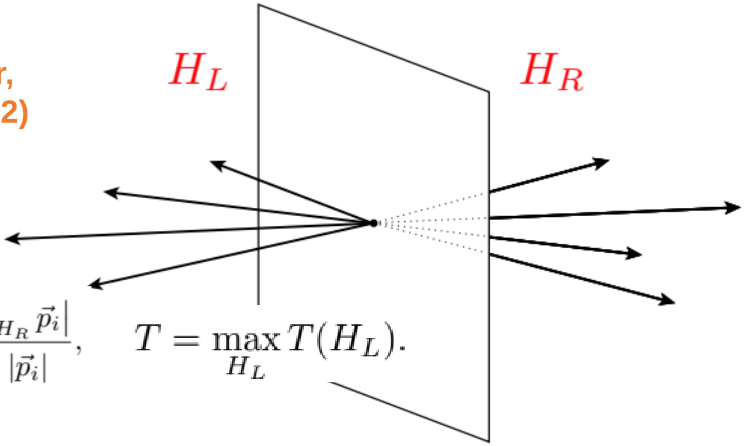


- Simulated bifurcation provides **compatible or slightly better efficiency & purity than D-Wave Neal:** efficiency ~ 95%, purity >90% for <6000 particles, >84% even for ~10000 particles.
- Ballistic simulated bifurcation provides **4 orders of magnitude speed-up (23min → 0.14s) from D-Wave Neal** for HL-LHC data **(& D-Wave qbsolv is even 2 orders of magnitude slower than Neal)**.
- **Simulated bifurcation can effectively run w/ multiple processing, GPU & FPGA → Perfect match with HEP environment!!**

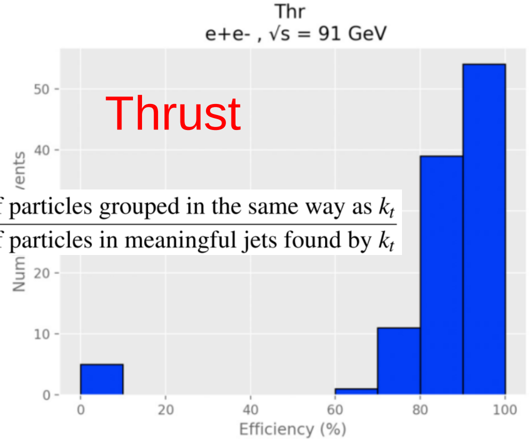
Jet Reconstruction as Ising Problem

Quantum Annealing (Thrust or Angle-based)

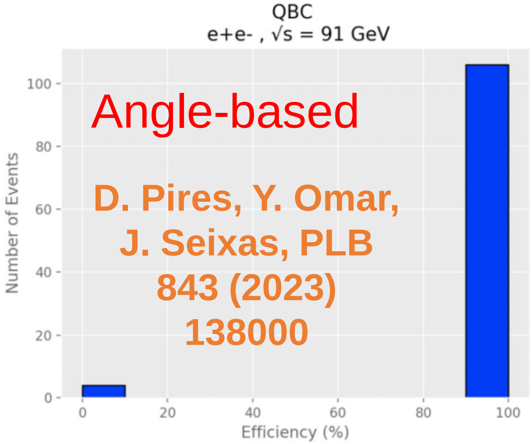
A. Delgado, J. Thaler, PRD 106, 094016 (2022)



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



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

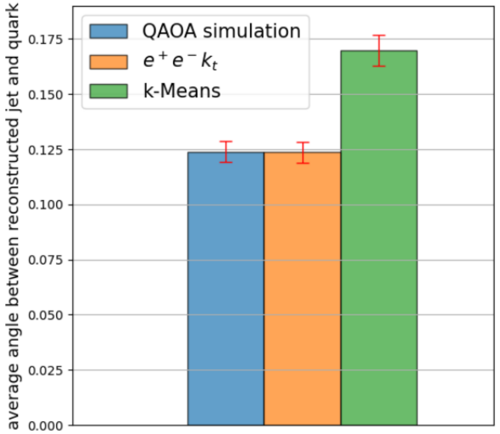


D. Pires, Y. Omar, J. Seixas, PLB 843 (2023) 138000

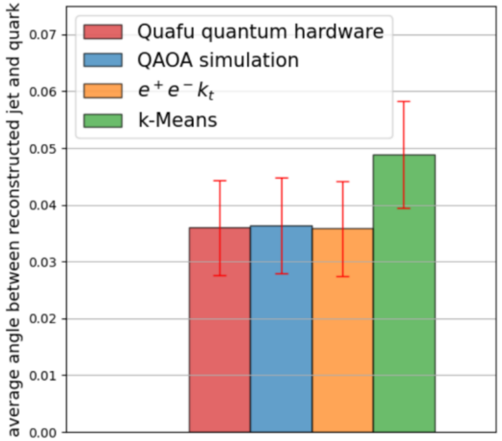
Quantum Gates (e.g. QAOA)

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

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



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



- Jet reconstruction can also be considered as a QUBO problem. (There are also other quantum approaches; see backup)
- 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.

Multijet Reconstruction (Our Study)

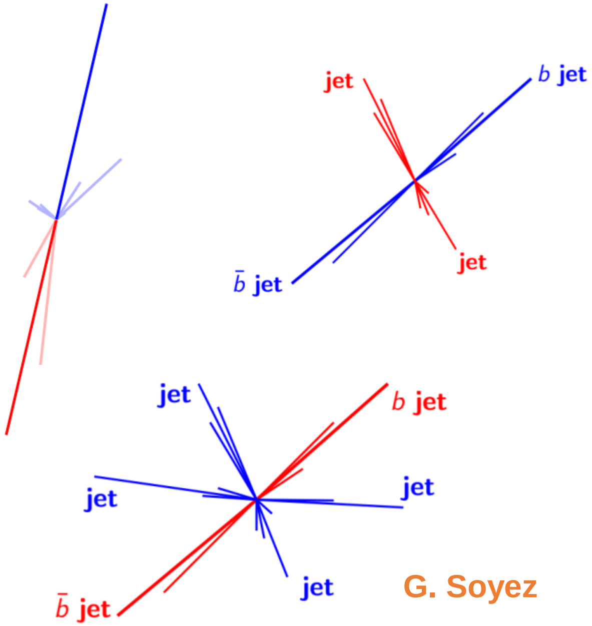
QUBO Formulation

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

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

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

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

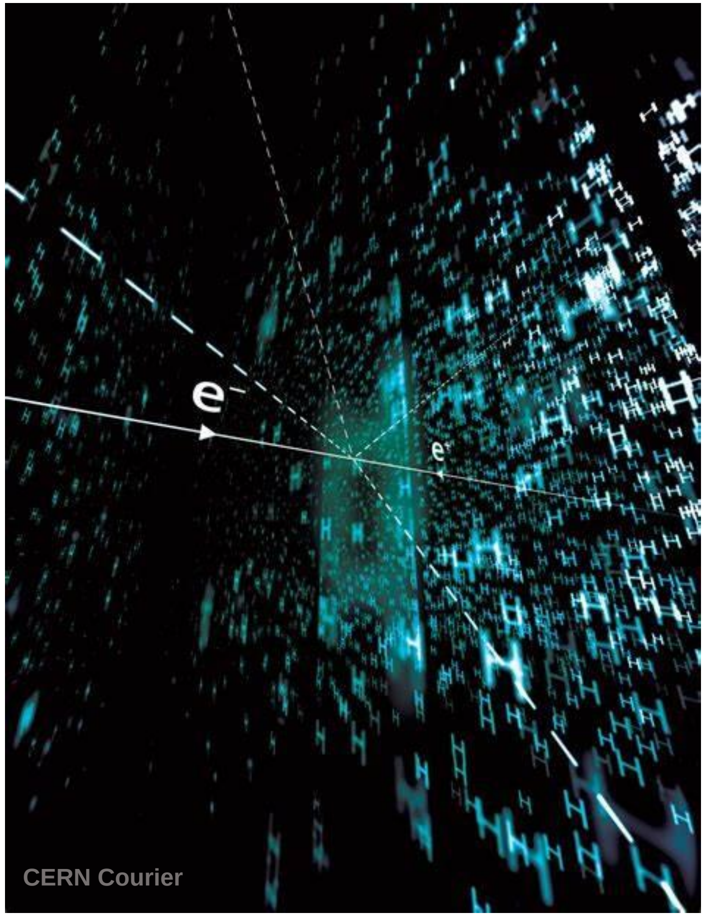


G. Soyez

- **Exclusive jet finding (n_{jet} fixed) with the ee-k_t algorithm** is the baseline at CEPC & other e+e- future Higgs factories.
- **We adopt the same ee-k_t distance in the QUBO formulation. QUBO is designed for general jet multiplicity beyond dijet.**
- The angle-based method is also shown for comparison **[D. Pires et al PLB 843 (2023) 138000].**

Dataset

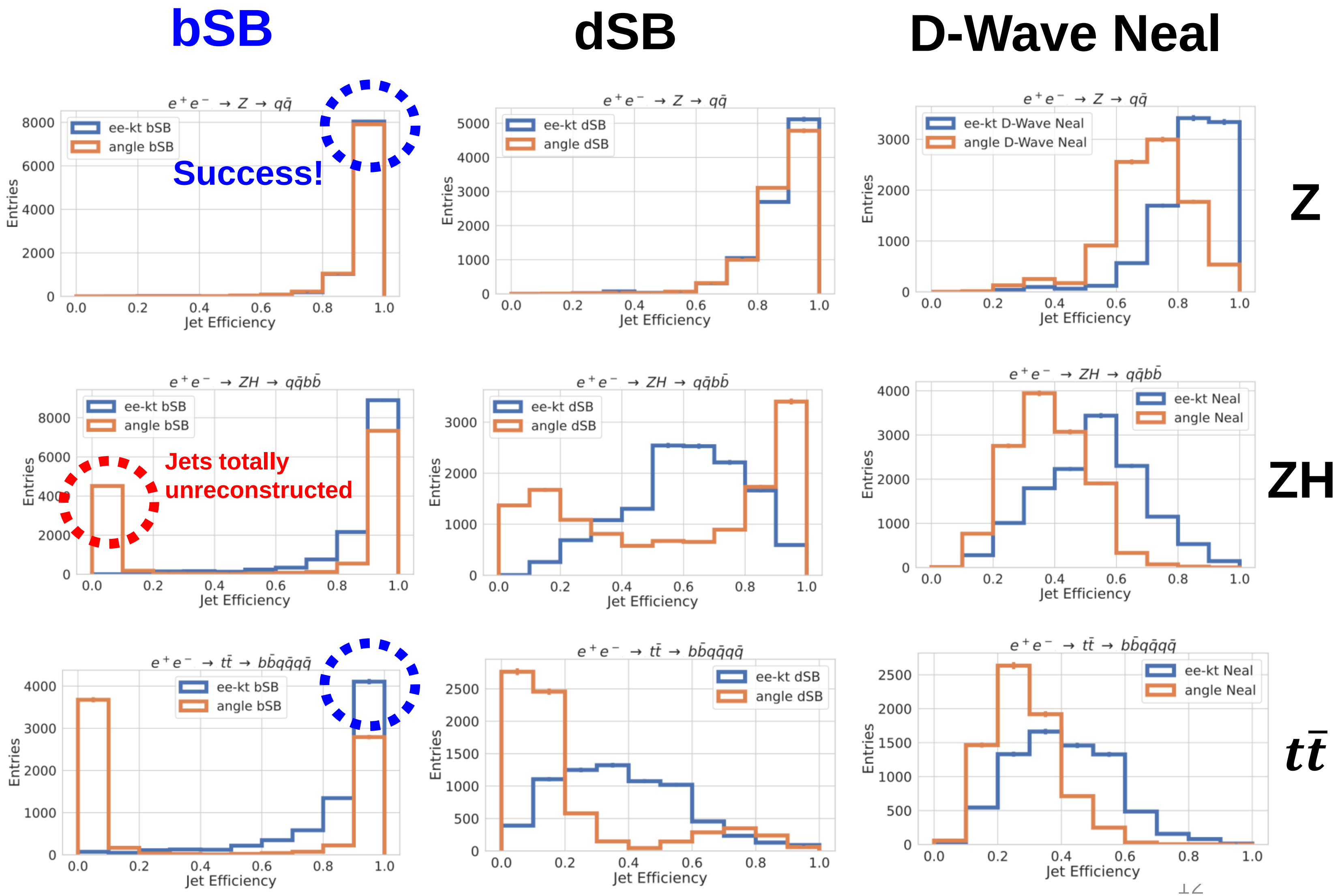
- Three sets of e+e- collision events are generated to consider various jet multiplicity:
  - **$Z \rightarrow q\bar{q}$ (√s=91 GeV, 2 jets),**
  - **$ZH \rightarrow q\bar{q}b\bar{b}$ (√s=240 GeV, 4 jets)**
  - **$t\bar{t} \rightarrow b\bar{b}q\bar{q}\bar{q}q$ (√s=360 GeV, 6 jets)**
- **Delphes card 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**.



Resemblance to FastJet

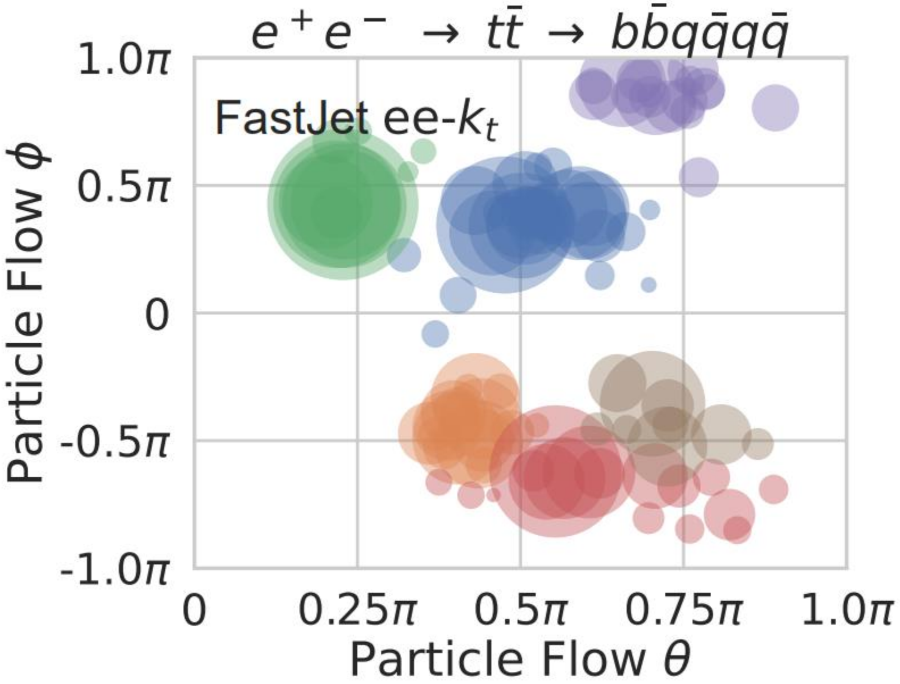
$$\varepsilon = \frac{\#\text{ of particles grouped in the same way as } k_t}{\#\text{ of particles in meaningful jets found by } k_t}$$

- Resemblance/efficiency = **compatibility of jet assignment w/ the traditional FastJet ee-k_t jet finding.**
- **Angle-based method only works for dijet. → misses many jets**
- **bSB provides the highest efficiency & can only reconstruct multijet events.**
- **D-Wave Neal has visibly degraded performance already in dijet events. dSB also has lower efficiency than bSB & cannot handle multijet.**

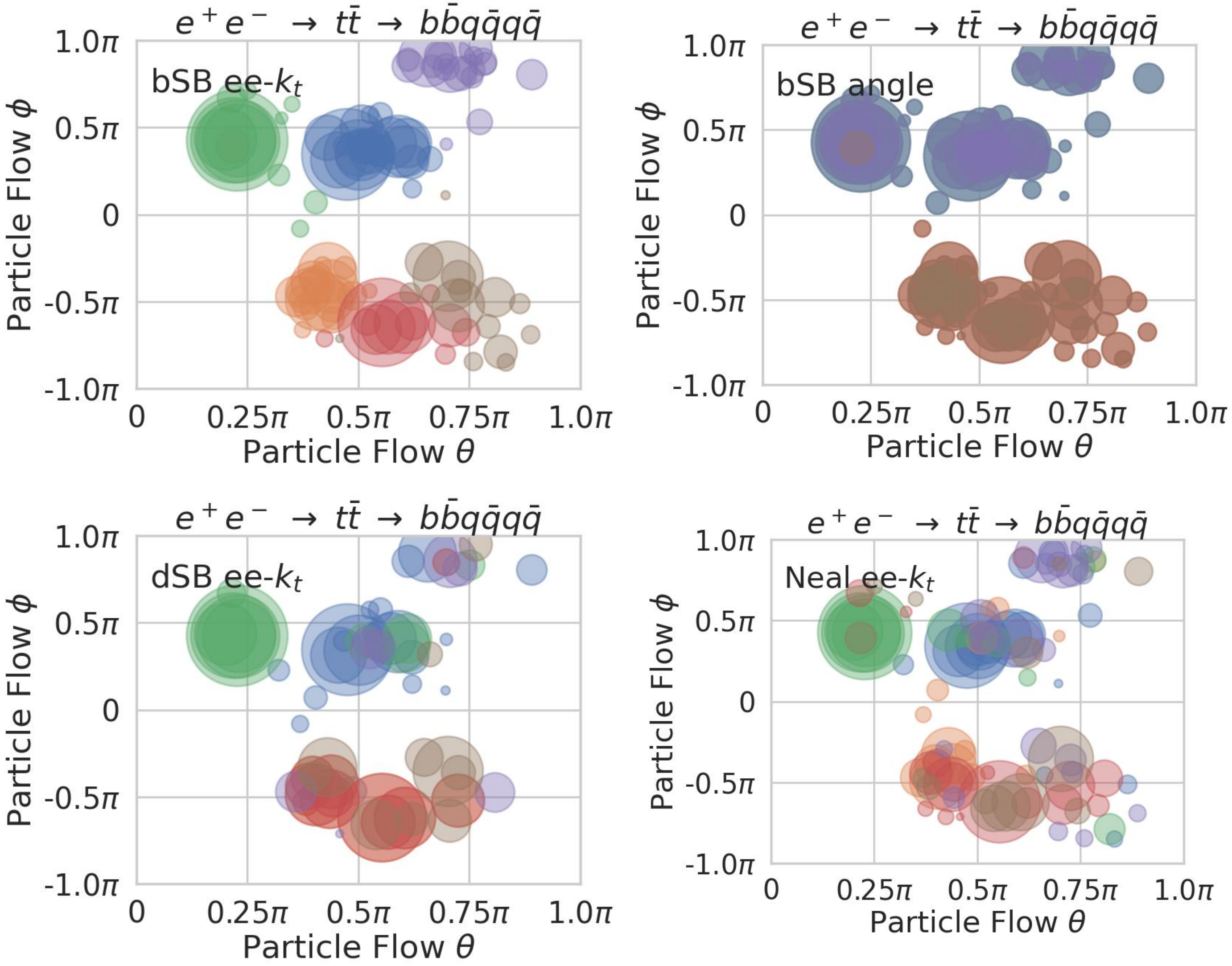


Event Displays ($t\bar{t}$)

Baseline (FastJet)

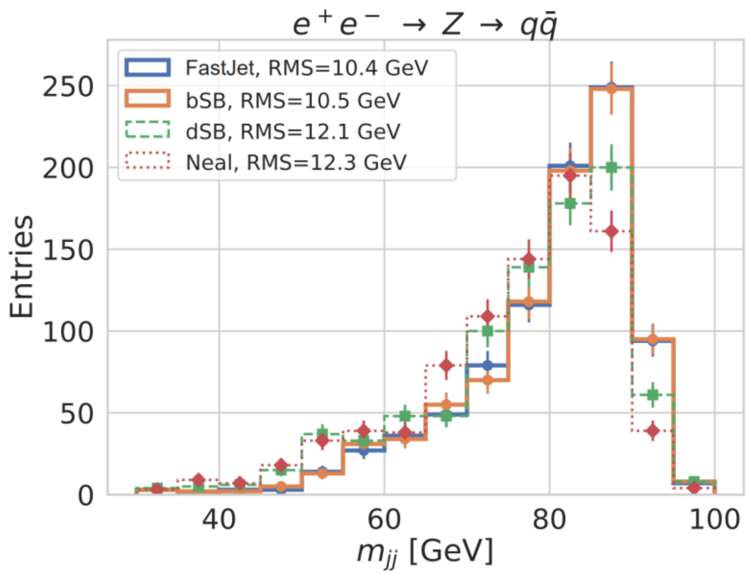


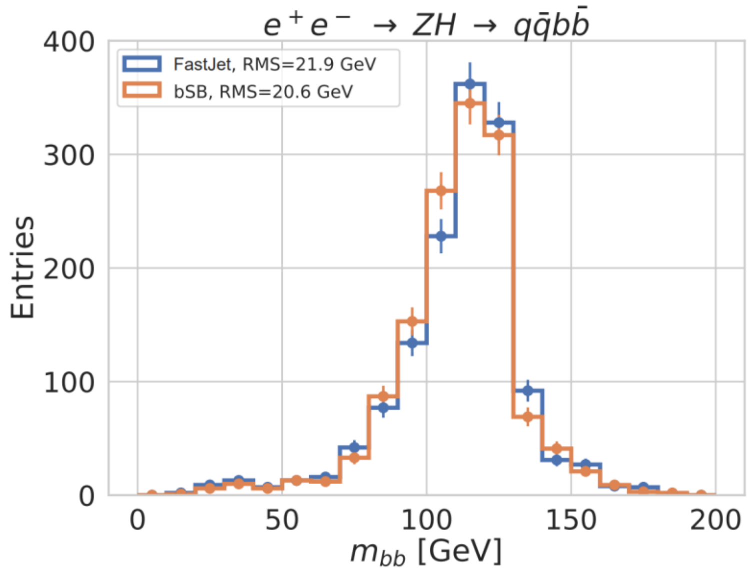
QAIAs

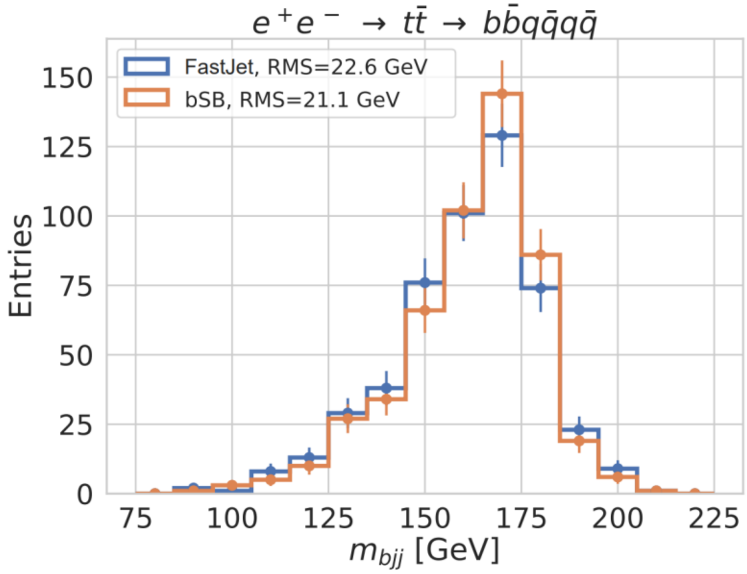


- **Only bSB w/ ee-k_t QUBO can reasonably reconstruct all jets.**
- **Other approaches misses some jets and/or PFlows are totally mixed up.**

Impact on Invariant Mass







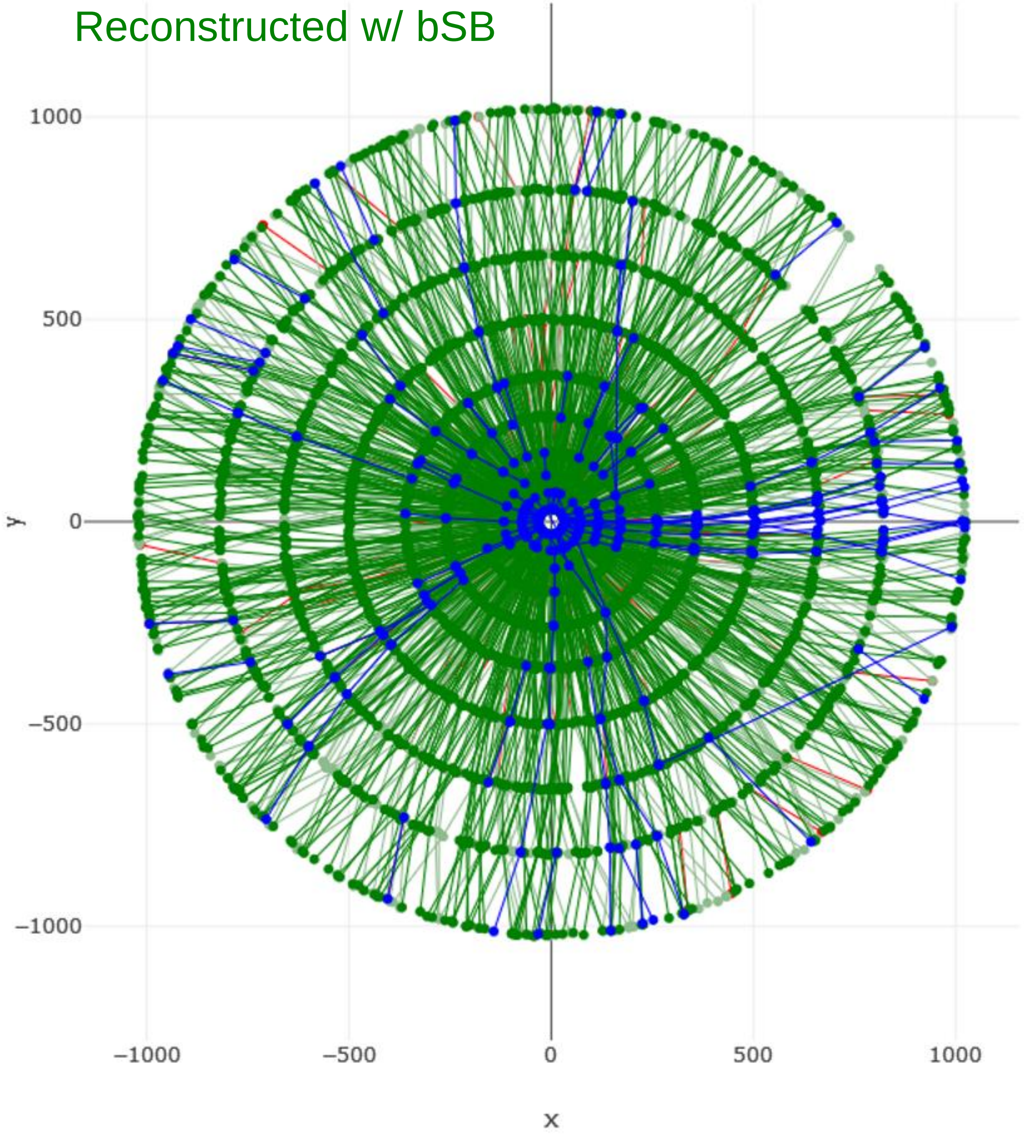
- **As FastJet is NOT the 'TRUE' answer, resemblance to it is not the decisive performance metric.** → Z, Higgs and top quark mass resolutions are evaluated.
- **bSB improve mass resolution for multijet! (& comparable resolution for Z)**
- **dSB & Neal already has ~20% degradation in Z mass resolution & unable to properly reconstruct jets in multijet events (thus not shown for ZH & $t\bar{t}$)**

Further reading:

- **H. Okawa, et al., Springer Comput. Softw. Big Sci. 8, 16 (2024)**
- **H. Okawa, et al., arXiv:2410.14233 (2024)**

Summary

- Tracking & jet reconstruction are CPU-consuming 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.**
- Tracking:
  - **This the world's first application of simulated bifurcation to high energy physics.**
  - 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 be considered for implementation **NOW**.
- Jet reconstruction:
  - **World's first successful demonstration of multijet reconstruction w/ QUBO.**
  - 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.



Thank you for listening!

感谢聆听！

Quantum Approaches

Quantum Gates

- Uses quantum logic gates. General-purposed
- IBM, Google, Xanadu, IonQ, Origin Quantum, QuantumCTek, etc.

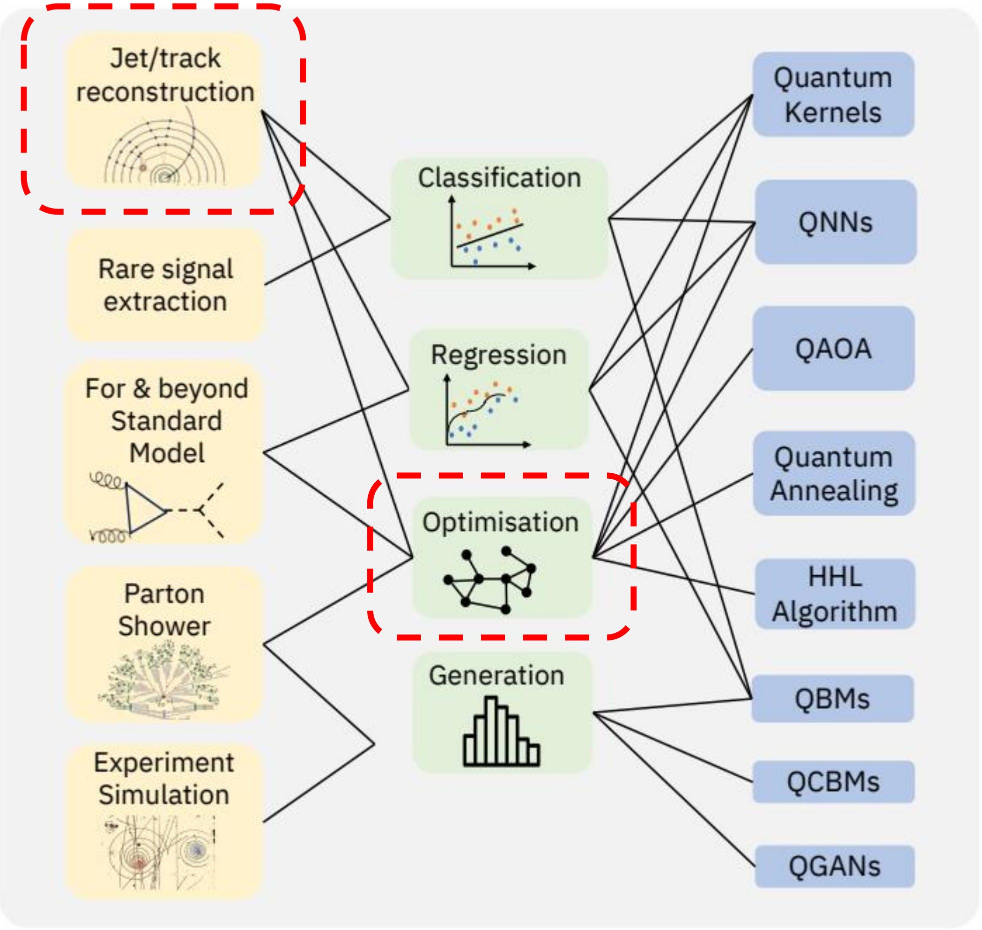
Ising machines

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

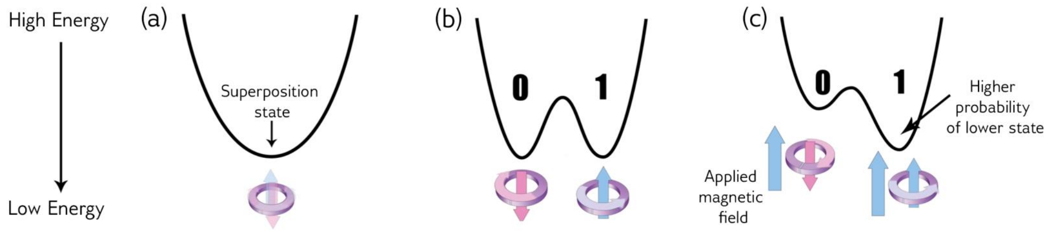
- Inspired by quantum annealing.
- Simulated annealing, simulated coherent Ising machine, simulated bifurcation, etc.



QC4HEP White Paper

Quantum Approaches

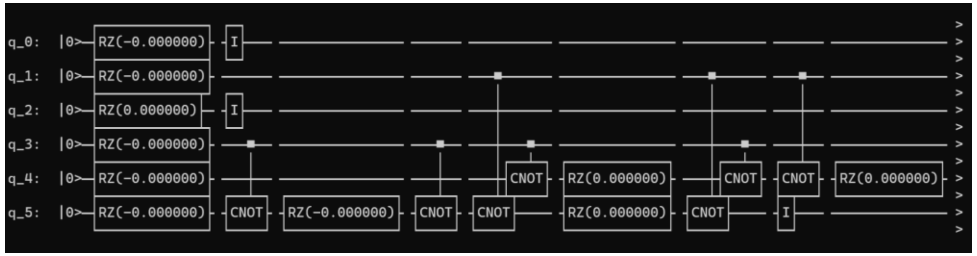
Quantum annealing


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.

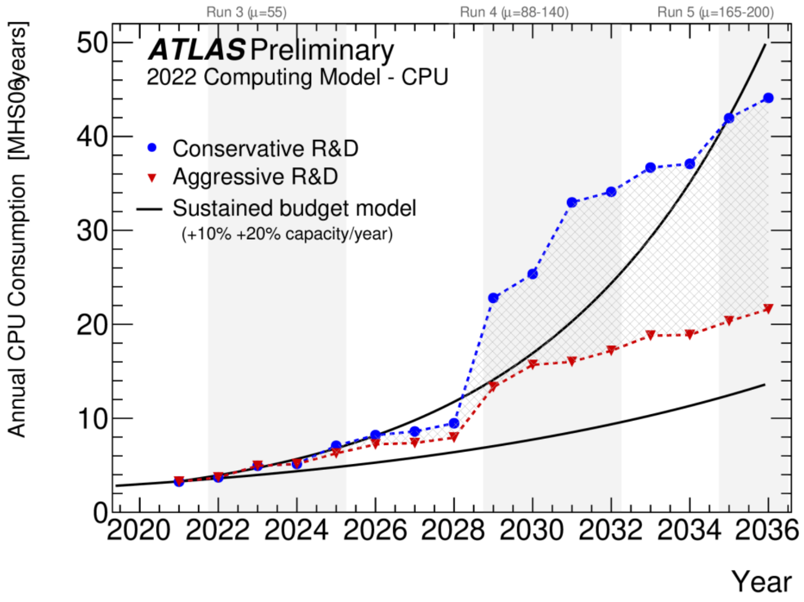
Quantum Gates

QAOA circuit implemented in Origin Quantum

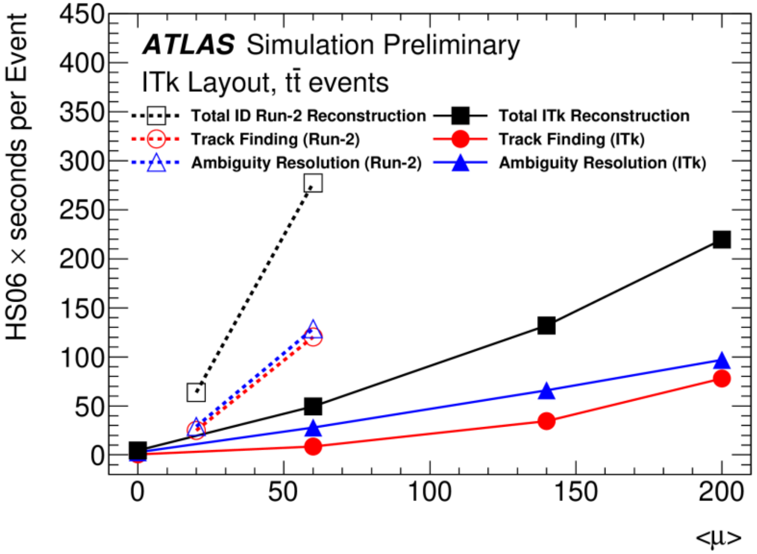


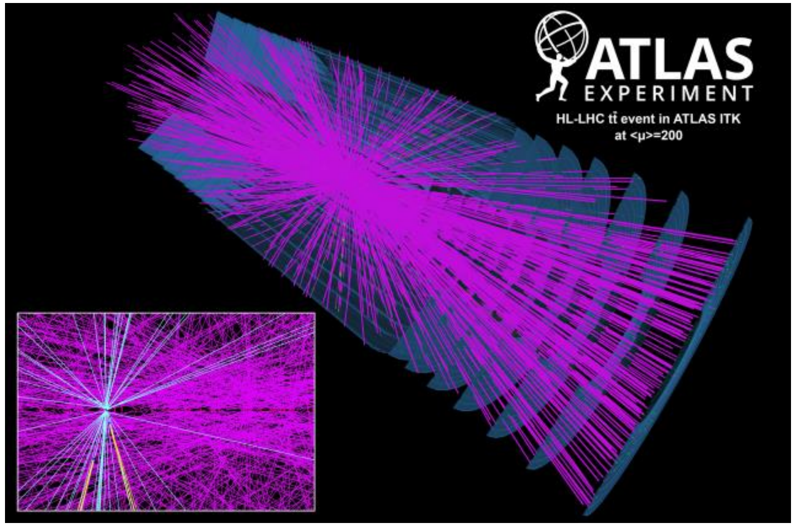
- Quantum gate machines are universal, and can also solve Ising problems with variational circuits: e.g. Variational Quantum Eigensolver (VQE), Quantum Approximate Optimization Algorithm (QAOA), etc.
- Pros: Universal computing, a few platforms available in China
- Cons: Number of qubits is much less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

Reconstruction at LHC & HL-LHC



ATL-PHYS-PUB-2019-041





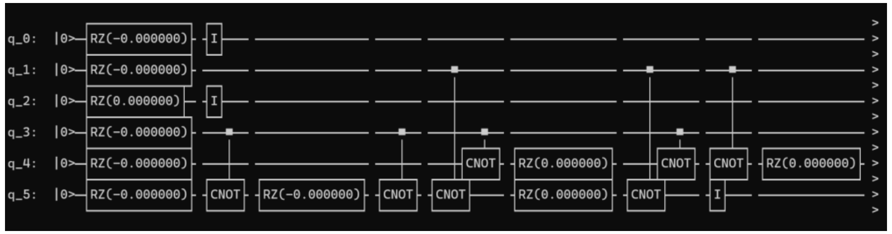
- At the HL-LHC, <u>**CPU time exponentially increases with pileup**</u>, leading to increase in annual computing cost by x10-20.
- **Tracking is the most CPU-consuming reconstruction task.**
- **Jet reconstruction is also known to be CPU-intensive.**
- GPU & ML-based approaches are actively investigated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

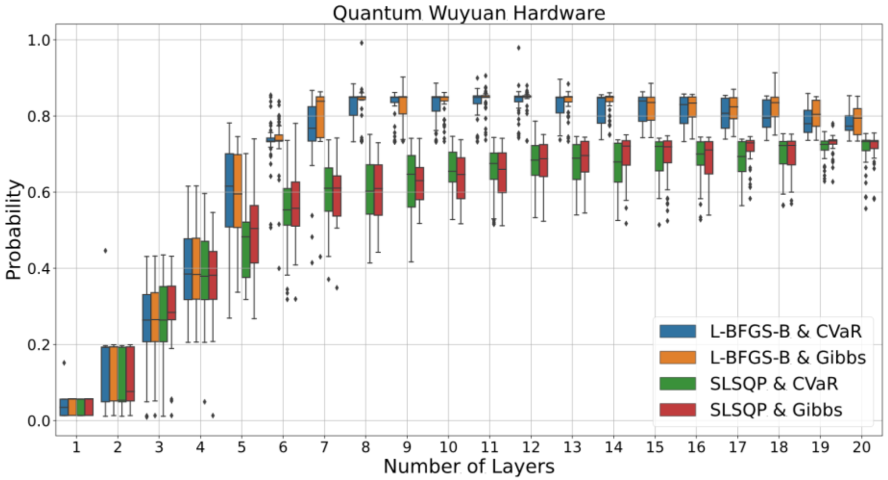
	Run 1	Run 2	HL-LHC
μ	21	40	150-200
Tracks	~280	~600	**~7-10k**

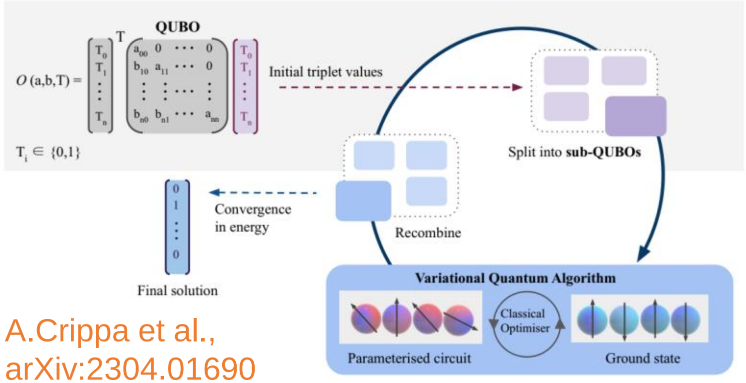
Previous Study w/ Quantum Gates

- Thorough optimization of QAOA in terms of # of layers, optimizers & loss functions.
- 6-qubit hardware (Origin Quantum Wuyuan) & simulator are used.

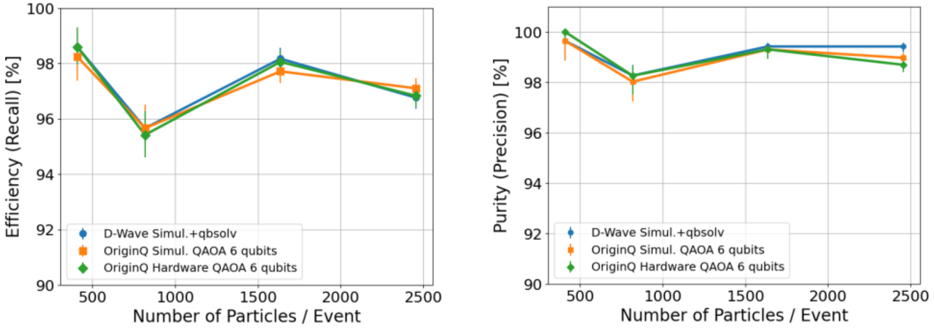
- Used a theoretically robust sub-QUBO method to split the problem into 6-qubit size
- Comparable performance obtained w/ the previous D-Wave studies (F. Bapst et al. Comp. Soft. Big Sci. 4 (2019) 1)





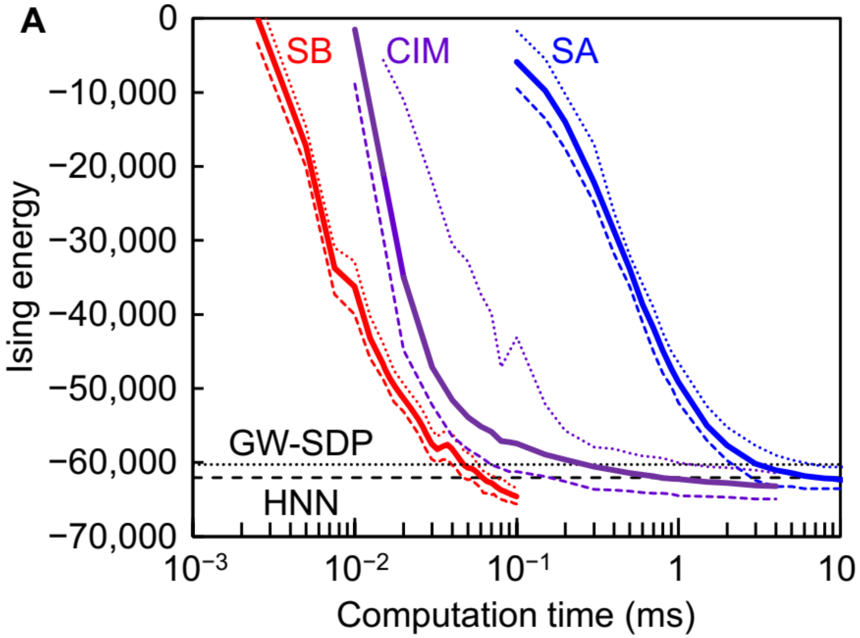


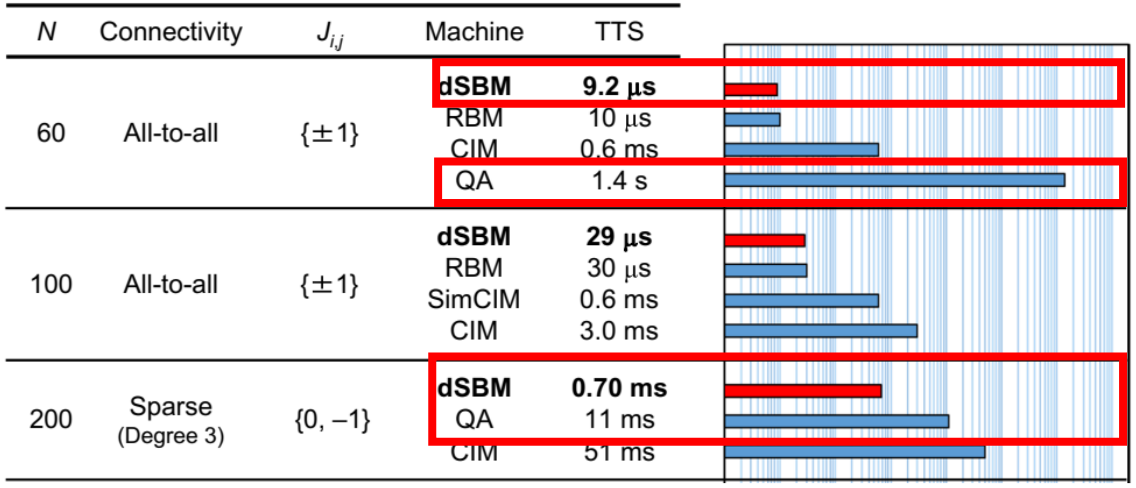
A.Crippa et al., arXiv:2304.01690



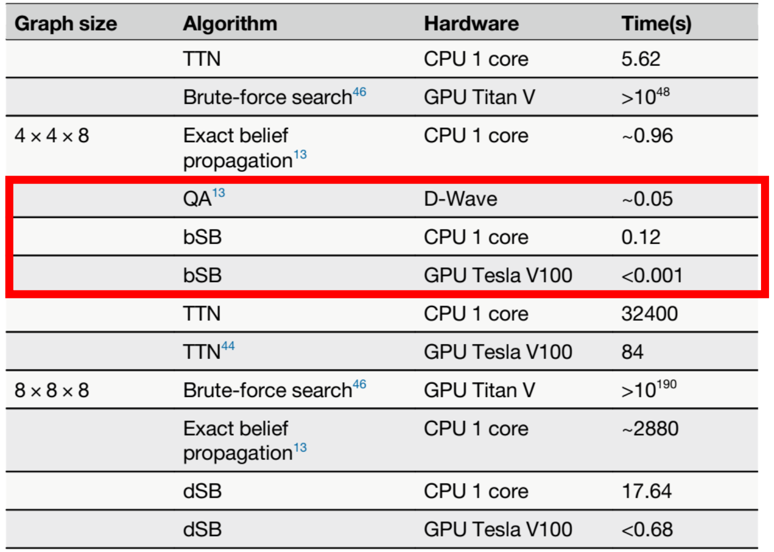
Simulated Bifurcation (SB)

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**	**9.2 μs**
			RBM	10 μs
			CIM	0.6 ms
			QA	1.4 s
100	All-to-all	{±1}	**dSBM**	**29 μs**
			RBM	30 μs
			SimCIM	0.6 ms
			CIM	3.0 ms
200	Sparse (Degree 3)	{0, −1}	**dSBM**	**0.70 ms**
			QA	11 ms
			CIM	51 ms

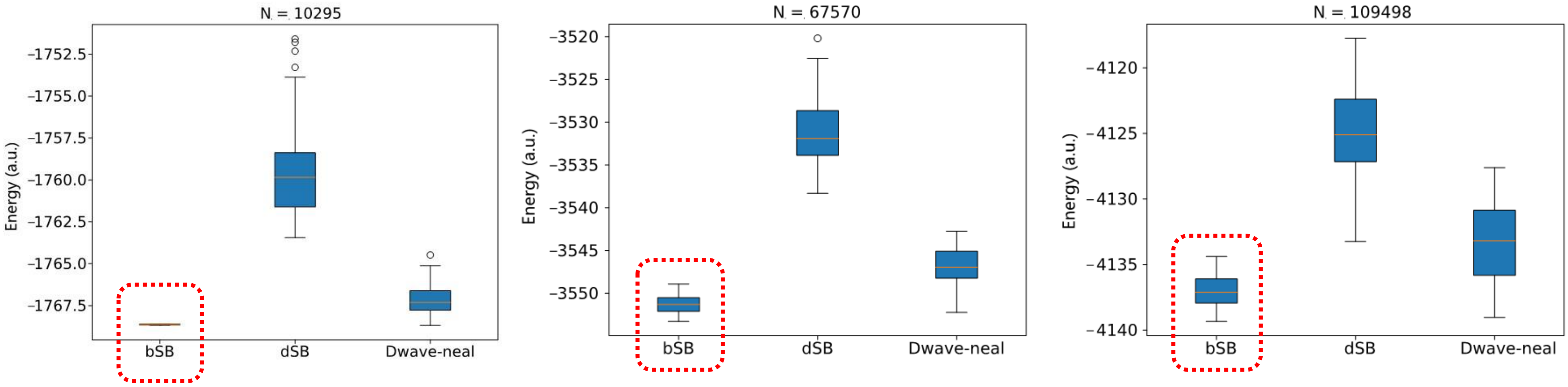


Graph size	Algorithm	Hardware	Time(s)
4 × 4 × 8	TTN	CPU 1 core	5.62
	Brute-force search[46]	GPU Titan V	$>10^{48}$
	Exact belief propagation[13]	CPU 1 core	~0.96
	QA[13]	D-Wave	~0.05
	bSB	CPU 1 core	0.12
	bSB	GPU Tesla V100	<0.001
8 × 8 × 8	TTN	CPU 1 core	32400
	TTN[44]	GPU Tesla V100	84
	Brute-force search[46]	GPU Titan V	$>10^{190}$
	Exact belief propagation[13]	CPU 1 core	~2880
	dSB	CPU 1 core	17.64
	dSB	GPU Tesla V100	<0.68

- **SB is known to outperform other quantum-inspired algorithms as well as quantum annealing (QA) for some problems**
- Our previous study: track reconstruction w/ SB → 4 orders of magnitude speed-up from SA. **H. Okawa, Q.G. Zeng, X.Z. Tao, M.H. Yung, Springer Comput. Softw. Big Sci. 8, 16 (2024)**

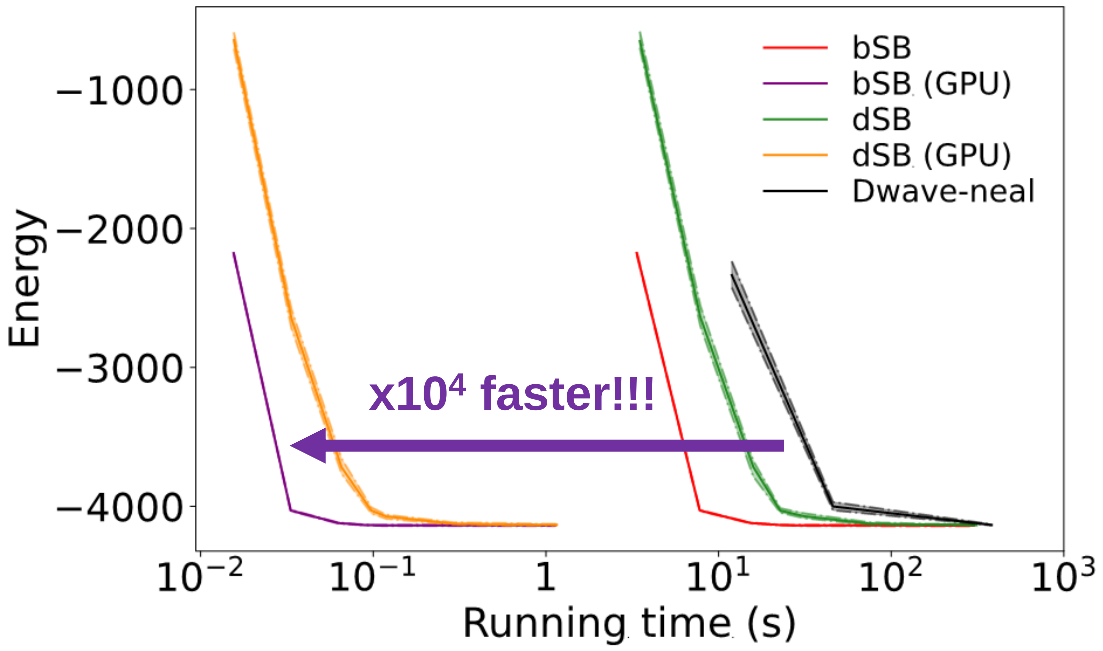
Q.G. Zeng et al., Comm. Phys. (2024) 7:249

Ising Energy w QAIAs

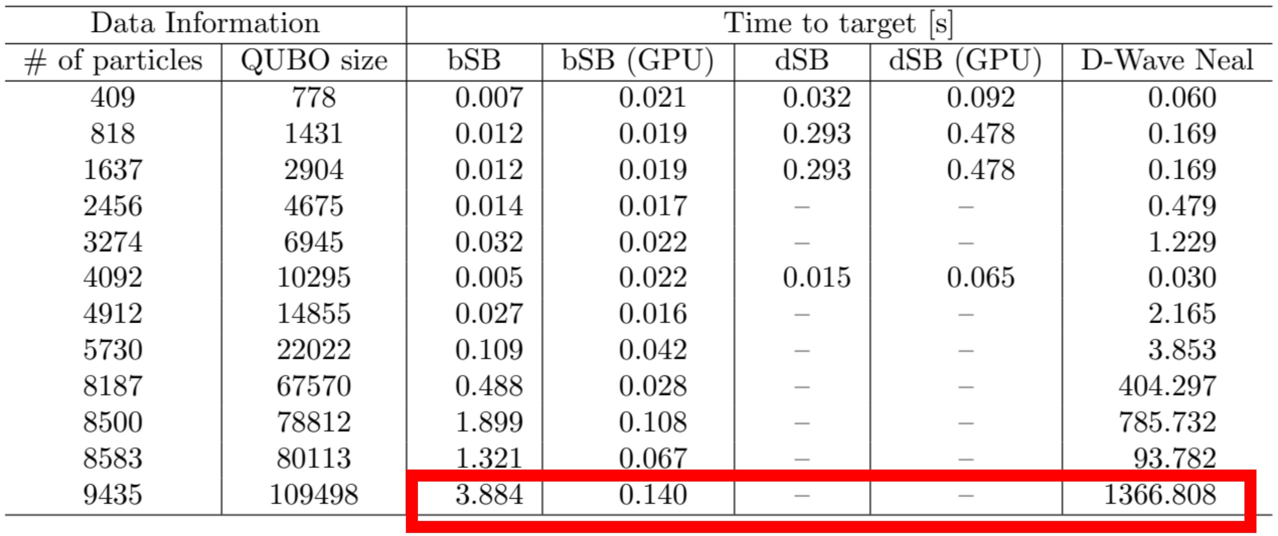


- **Ballistic simulated bifurcation can find the lowest Ising energy with the smallest fluctuation** for all events considered.
- Discrete simulated bifurcation provides slightly degraded energy prediction to bSB & D-Wave Neal, though the impact on the track reconstruction performance is not significant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

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

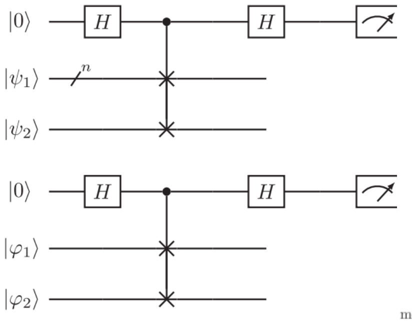
- Ballistic simulated bifurcation provides **4 orders of magnitude speed-up (23min → 0.14s) from D-Wave Neal** at most **(D-Wave qbsolv is even 2 orders of magnitude slower than Neal)**.
  → More speed-up expected with larger data size.
- Unlike D-Wave Neal, **simulated bifurcation can effectively run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

Quantum Jet Reconstruction (Iterative)

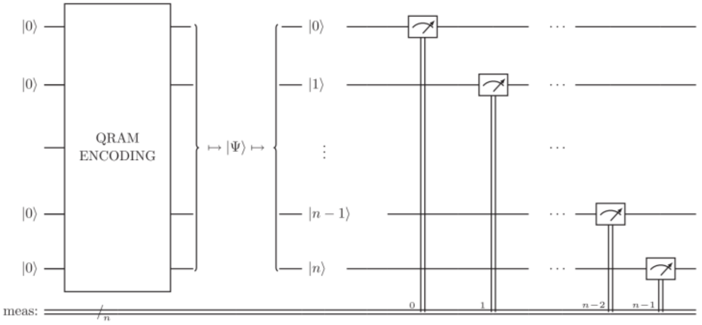
- Jet reconstruction is a clustering problem. Quantum algorithms may bring in acceleration.
- A few algorithms were considered to replace the traditional iterative calculation. Expected to bring in speed-up, but still at a conceptual stage.

Quantum K-means， Quantum Affinity Propagation (AP)， Quantum k_t

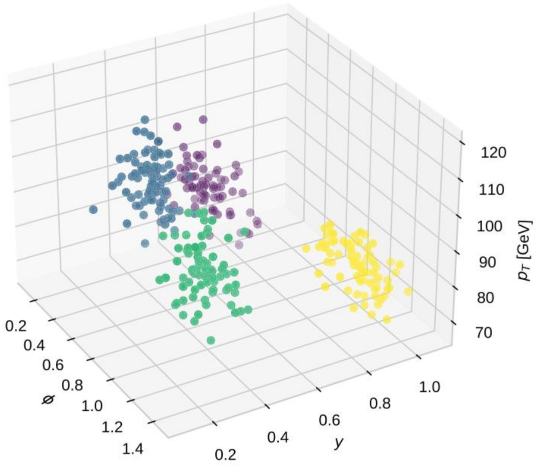
Computes distance in Minkowski space

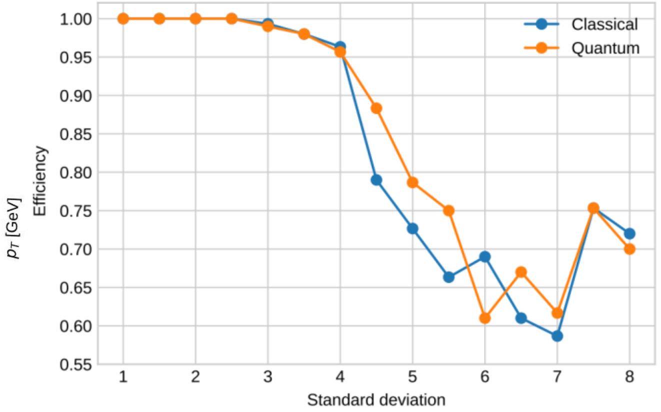


Assigns each particle to the nearest centroid



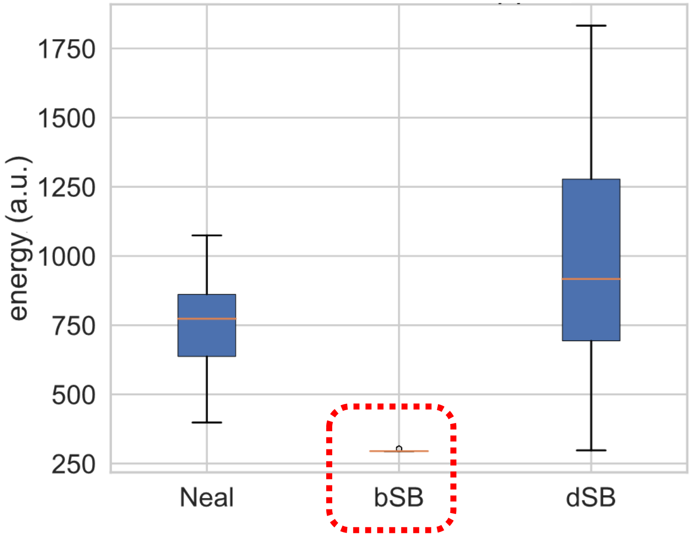
J.J. Martinez de Lejarza, L. Cieri, G. Rodrigo, PRD 106 036021 (2022)



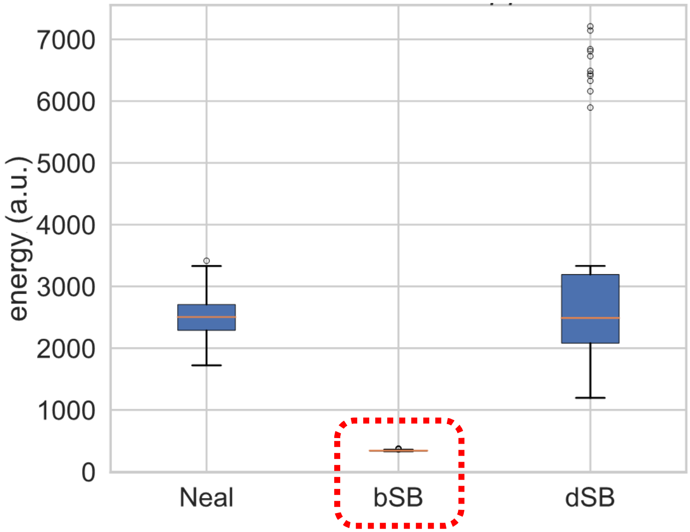


- Similar studies： Grover search **A. Wei, P. Naik, A.W. Harrow, J. Thaler, PRD 101, 094015 (2020),** quantum K-means **D. Pires, P. Bargassa, J. Seixas, Y. Omar, arXiv:2101.05618 (2021).**

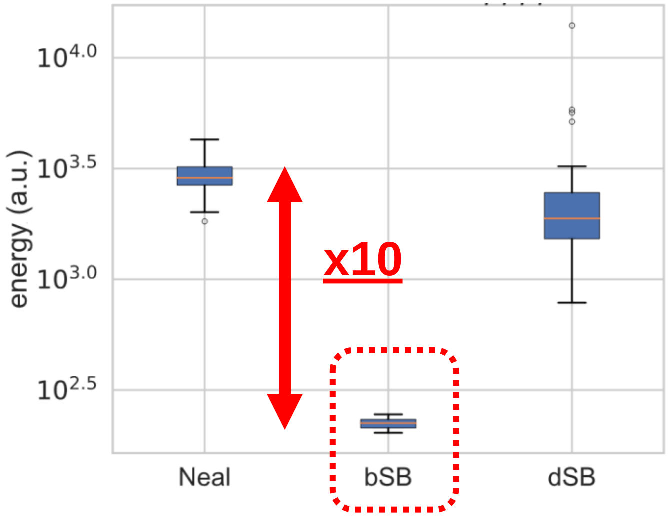
Ising Energy Prediction



$e^+e^- \rightarrow Z \rightarrow q\bar{q}$



$e^+e^- \rightarrow ZH \rightarrow q\bar{q}b\bar{b}$

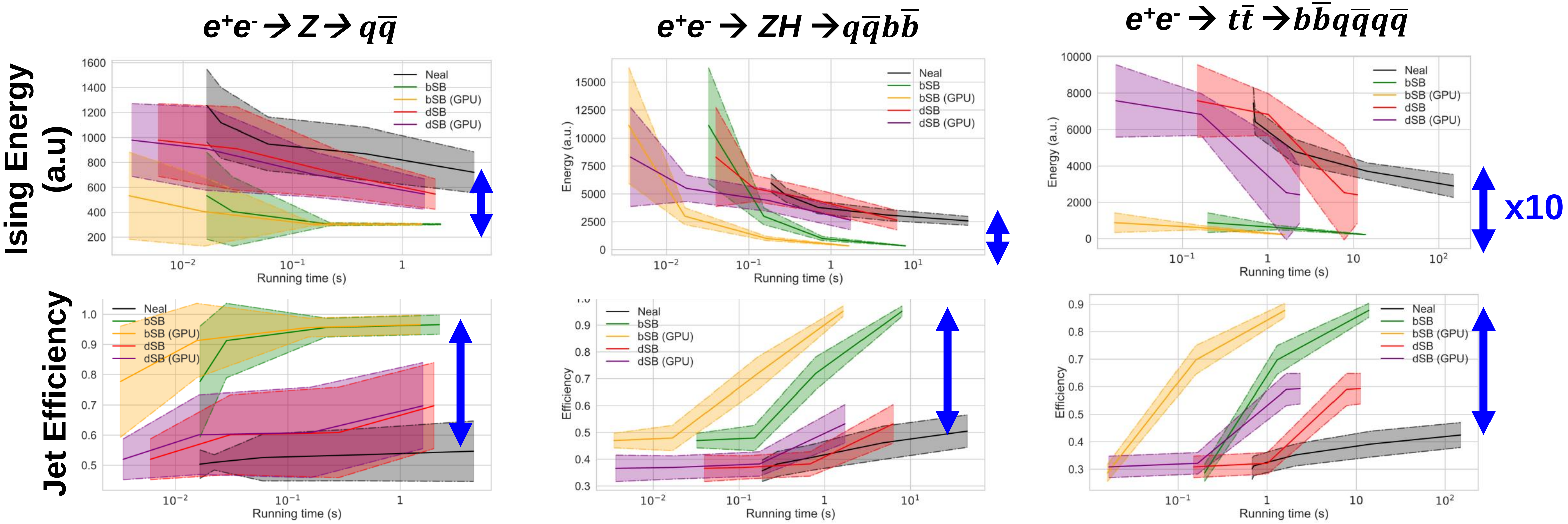


$e^+e^- \rightarrow t\bar{t} \rightarrow b\bar{b}q\bar{q}q\bar{q}$

- **Fully-connected QUBOs are difficult to solve**; it is known that quantum annealing hardware is not good at solving them so far.
    - This is in contrast to track reconstruction, in which the QUBOs are largely sparse.
- **Ballistic SB (bSB) predicts energy lowest with the smallest fluctuation.**
- **Performance is especially outstanding for 6-jet QUBOs → bSB can find x10 lower minimum energy for the all-hadronic $t\bar{t}$ events!**

Computation Speed

Only 1 CPU/GPU used



- **Ising solvers usually continue to improve energy prediction w/ running time.**
- **bSB significantly outperforms dSB & Neal (& an order of maganitude speed-up w/ GPU)**
- **D-Wave Neal is trapped in a local minimum (x10 worse energy prediction for $t\bar{t}$). dSB is slower in energy convergence & less successful than bSB for energy prediction.**