



# 量子退火启发算法 在未来对撞机中的应用

第29届Mini-workshop on the frontier of LHC, 2024年12月14-15日 大川(Okawa) 英希(Hideki) 中国科学院高能物理研究所 与曾庆国,陶贤哲,翁文康合作(深圳量子科学与工程研究院)

### 第二次量子革命:新时代的黎明

#### 谷歌(2019)

#### Article

Quantum supremacy using a programmable superconducting processor



(2019) 505



#### 九章(2020)

#### QUANTUM COMPUTING

#### Quantum computational advantage using photons

Han-Sen Zhong<sup>1,2</sup>\*, Hui Wang<sup>1,2</sup>\*, Yu-Hao Deng<sup>1,2</sup>\*, Ming-Cheng Chen<sup>1,2</sup>\*, Li-Chao Peng<sup>1,2</sup>, Yi-Han Luo<sup>1,2</sup>, Jian Qin<sup>1,2</sup>, Dian Wu<sup>1,2</sup>, Xing Ding<sup>1,2</sup>, Yi Hu<sup>1,2</sup>, Peng Hu<sup>3</sup>, Xiao-Yan Yang<sup>3</sup>, Wei-Jun Zhang<sup>3</sup>, Hao Li<sup>3</sup>, Yuxuan Li<sup>4</sup>, Xiao Jiang<sup>1,2</sup>, Lin Gan<sup>4</sup>, Guangwen Yang<sup>4</sup>, Lixing You<sup>3</sup>, Zhen Wang<sup>3</sup>, Li Li<sup>1,2</sup>, Nai-Le Liu<sup>1,2</sup>, Chao-Yang Lu<sup>1,2</sup>†, Jian-Wei Pan<sup>1,2</sup>†

Quantum computers promise computers. Boson sampling the quantum computational indistinguishable single-mode connectivity and random mat using 100 high-efficiency sing hypotheses exploiting therma quantum computer, *Jiuzhang* space dimension of 10<sup>30</sup> and strategy and supercomputers

Science 370 (2020) 1460



• 第一次量子革命:激光、晶体管、核磁共振等

# ・第二次量子革命:能够识别、控制单个量子。<br/> 商用量子计算机的到来。<br/> 在生成随机数<br/> 时证明了量子霸权。正在走向实际应用。

### 中等规模带噪声的量子计算(NISQ)时代



- ・现在是中等规模带噪声的量子计算时代 (>50量子比特)。
- 超导容错量子计算机要约100万量子比特。 → 谷歌上周发布了新芯片Willow。 "a breakthrough" 随着量子比特数量的增加,逻辑错误呈指数级减少。

第29届LHC Mini-workshop

# 量子计算机和伊辛机







Quantum inspired algorithm

量子计算机

### 高能物理中的应用



关于量子模拟还有C.W. Bauer et al., PRX Quantum 4 (2023) 2, 027001

大川英希

# 组合优化问题

翻译来©TOSHIBA



- ・组合优化问题是非确定性多项式(NP)完全问题→不能确定是否在多项式时间内找到答案。
- 可以表示为伊辛问题。社会中的许多应用。伊辛机在合理的时间内求解绝对或近似的基态。
- 在高能物理中, 径迹和喷注重建等许多任务也可以表述为伊辛问题。

# 真实量子计算机



- 量子退火通过绝热定理和量子隧穿寻找伊辛 哈密顿量的全局最小值。
- 量子比特数量 (5000+) 比量子门多, 但并 非所有量子比特都适用于fully-connected graphs, 也不擅长解决这种问题。

- 量子门可以用变分线路解决伊辛问题:
  - 例如: 变分量子本征求解器 (Variational Quantum Eigensolver, VQE), 量子近似优化算 法 (Quantum Approximate Optimization Algorithm, QAOA) 等
- 它通过用<u>经典优化器</u>更新变分参数来搜索伊
   辛哈密顿量的全局最小值。

# 量子退火启发算法

#### 量子退火启发算法通过微分方程的经典时间演化寻找最小能量



- ・模拟分叉算法 (SB) 模仿Kerr-nonlinear parametric振子的量子绝热演化,表现出分叉 现象,可用于描述伊辛自旋。
- 根据如何考虑自旋(x<sub>i</sub>)的连续处理,存在三种 算法(x<sub>i</sub>): aSB, bSB, dSB

- ・模拟退火 (SA) 在solution空间中使用随机移动。
- 在每个随机移动中,如果获得较低的能量ΔE<0, 则自动接受。
- 如果ΔE>0,则仅根据玻尔兹曼概率上接受: P(ΔE) = exp(-ΔE/kT).

### 径迹重建为组合优化问题

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



# 基于QUBO的径迹重建算法流程



- 根据硅探测器的粒子撞击信息,在每个事例中定义QUBO哈密顿量。
- 预测的最低能量将决定应保留哪些triplets (保留:  $T_i=1$ , 放弃:  $T_i=0$ )。
- 连接保留的triplets将提供径迹。

# 真实量子计算机 (量子退火,量子门)





- 将QUBO分为sub-QUBO (矩阵分割计算方法)
- Ising求解精度没有下降,但计算速度下降了几个数量级



第29届LHC Mini-workshop

### 量子退火启发算法



#### ・<u>为了高能物理,全球首次用模拟分叉算法。可以直接处理HL-LHC数据(没有量子比特数量的限制)</u>

- 径迹重建性能很好(在HL-LHC最稠密条件,还保持效率>95%,纯度>85%)。
- ・ 比模拟退火算法(D-Wave Neal)快一万倍 (23分钟 → 0.14秒)! 这不是未来的算法,在正在运行的实验上可以
   用! (欧洲的LHC,中国的BESIII实验等)
- ・模拟分叉算法可以并行运行和使用GPU和FPGA。与高能物理计算环境完美匹配!

# 喷注重建为组合优化问题

#### 量子退火机 (基于Thrust或角度)



#### 量子门 (量子近似优化算法QAOA) 📲

朱永峰<sup>a</sup>, 庄伟峰<sup>b</sup>, 钱辰<sup>b</sup>, 马运恒<sup>b</sup>, 刘东 <sup>b,c</sup>, 阮曼奇<sup>d</sup>, 周辰<sup>a</sup>, arXiv:2407.09056 <sup>a</sup>北大, <sup>b</sup>北京量子院, <sup>c</sup>清华, <sup>d</sup>高能所

30粒子简化数据 (e+e-→ZH→vvss) 6粒子简化数据 (e+e-→ZH→vvss)





- 喷注重建也可以被视为组合优化(QUBO)问题
- 量子退火:角度方法超越Thrust方法,但<mark>多喷</mark> <u>注事例(N<sub>jet</sub>>2)中的性能差。</u>
- 量子门QAOA:使用简化数据来评估平均角度。

### 多喷注重建作为组合优化问题



- ・考虑exclusive喷注重建(=之前固定喷注数)。我们采用ee-k<sub>t</sub>算法,它是中国CEPC和其他 未来正负电子对撞机的标准。
- 在QUBO公式中采用了ee-k<sub>t</sub>距离。x<sub>i</sub><sup>(n)</sup>=1表示第i个particle flow (pflow)属于第n个喷注。
- •考虑了基于量子角度方法进行比较。[D. Pires et al PLB 843 (2023) 138000].





和模拟退火 (既往研究) 量子角度方法 • 算法无法处理多喷注事例。



大川英希





·另外,我们的量子退火启发算法能够提高喷注能量分辨率。

大川英希

第29届LHC Mini-workshop

### 总结

参考文献:

- H. Okawa, Spinger CCIS 2036 (2024) 272, arXiv:2310.10255
- H. Okawa, et al., Springer Comput. Softw. Big Sci. 8, 16 (2024)
- H. Okawa, et al., arXiv:2410.14233 (2024)
- 高能物理界在积极启动量子计算应用研究!
- ・真实量子计算应用还没有进入实践阶段,但量子退火启发算法已经开始提供 创新!弹道模拟分叉算法(bSB)的性能尤其强大。
- 在高能对撞机, 径迹和喷注重建是计算资源消耗任务。
- 径迹重建:
  - ・全球首次在高能物理中使用模拟分叉算法。
  - ・<u>可以直接处理HL-LHC数据,比模拟退火算法计算速度快一万倍!</u>
- 喷注重建
  - ・全球首次成功使用QUBO重建多喷注事例!
  - ・在CEPC快速模拟事例中,改善了多喷流事例的不变质量分辨率 (顶夸克~7%)
- 敬请期待!





### **Google Quantum Al Roadmap**

# Our quantum computing roadmap

Our focus is to unlock the full potential of quantum computing by developing a large-scale computer capable of complex, error-corrected computations. We're guided by a roadmap featuring six milestones that will lead us toward top-quality quantum computing hardware and software for meaningful applications.



### **Difficulty & Usefulness**

#### Random circuit sampling (RCS): in context

To date, no quantum computer has outperformed a supercomputer on a commercially relevant application. Our latest research is a step towards that direction.

**Google Quantum Al** 



IBM Technology Atlas

#### Quantum roadmap

The future of computing is quantum-centric. Updated October 2024 completed pushed to next year on target

	2024	2025	2026	2027	2029	2033+
Quantum journey	Introduce parallelization of quantum computations.	Demonstrate quantum- centric supercomputing.	Automate and increase the depth of quantum circuits.	Scale quantum computing,	Deliver a fully error-corrected system.	Deliver quantum-centric supercomputers with 1,000's of logical qubits.
Strategy overview	<ul> <li>We will improve the quality and speed of quantum circuits to allow running 5,000 gates with parametric circuits.</li> <li>Wind the first quantum- computer by integ processors, middle quantum commun also enhance the quantum circuits.</li> </ul>		We will enable quantum circuits with 7,500 gates through circuit quality improvement.	We will scale qubits, electronics, infrastructure, and software to reduce footprint, cost, and energy usage. The quality of quantum circuits will improve to allow running 10,000 gates.	We will bring users a quantum system with 200 qubits capable of running 100 million gates.	Beyond 2033, quantum-centric supercomputers will include thousands of qubits capable of running 1 billion gates, unlocking the full power of quantum computing.
Why this matters to our clients and the world	rs md Qiskit Primitives with error mitigation will provide the foun- dation platform where algorithm and application developers can focus on the workflows and get the best quality out of the quantum hardware. ♥ We will make computing easient abstracting quantum fur Qiskit patterns, or way for domain left the best quality out of the quantum fur hardware. ♥ We will make computing easient abstracting quantum fur Qiskit patterns, or way for domain left the best quality out of the quantum fur the best quality out of the quantum fur Qiskit patterns, or way for domain left the best quality out of the quantum fur Qiskit patterns, or way for domain left the best quality out of the quantum fur Qiskit patterns, or way for domain left the best quality out of the quantum fur Qiskit patterns, or way for domain left the best quality out of the quantum fur Qiskit patterns, or way for domain left the best quality out of the quantum fur Qiskit patterns, or way for domain left the best quality out of the quantum fur Qiskit patterns, or way for domain left the best quality out of the quantum fur Qiskit patterns, or way for domain left the best quality out of the quantum fur Qiskit patterns, or way for domain left the patterns of the quantum fur Qiskit patterns or the quantum		By running circuits with more gates, clients can expand their use case exploration. Circuit mapping collections will simplify mapping use cases to quantum circuits.	Scaled quantum systems will allow users to run larger computations. Multiple computing resources will be seamlessly combined to optimally handle workflows and extend the computational reach of quantum systems.	Users will be able to run large- scale problems using high-rate quantum error correction.	Quantum computers running algorithms using thousands of logical qubits are expected to enable general applications in security, chemistry, machine learning, and optimization.
The technology or innovations that will make this possible	<ul> <li>Built-in error mitigation will automatically determine the best method to reduce the effect of noise.</li> <li>Transpiler services will optimally rewrite circuits for hardware, taking advantage of AI.</li> <li>Watson Code Assistant will help users write Qiskit code to program quantum systems.</li> </ul>	We will demonstrate a quantum node as part of a network that incorporates classical and quantum communication. Resource management tools will enable system partitioning, manage quantum and classical workflows, and parallel execu- tion. Qiskit will provide libraries of quantum functions and high- er-level APIs for faster algorithm and application development.	To improve performance and allow running more complex algorithms, we will enable the decomposition of quantum circuits into shorter circuits, run these in parallel using multiple quantum chip processors, stitch these circuits back together with classical hardware, and automate the process of mapping use cases to quantum circuits.	Intelligent orchestration will analyze workflows to identify the optimal resource allocation (QPUs, communication, and classical resources) for the task. Qiskit will orchestrate approaches to handle errors to provide noise-free outputs to the users. Working alongside clients, we will build and make available use-case- specific libraries in Qiskit.	A novel and efficient error correction code will extend the computational reach of quantum resources. The system will have low-level dedicated classical hardware and a compiler for quantum-centric supercomputing.	Efficient logical decoding will enable 2,000 qubits working in a distributed 100,000-qubit machine. The middleware will include distributed software tools to manage noise-free quantum computations working seamlessly with classical computations. Qiskit will include general purpose quantum computing libraries to simplify the work of developers.
How these advancements will be delivered to IBM clients and partners	Multiple higher-quality 100+ qubit Heron processors will be connected using classical communication.	Pre-built Qiskit functions and optimized libraries will become available. Our multi- quantum-chip Flamingo system, comprised of processors each made from multiple chips, will be demonstrated.	These achievements in 2026 are slated to be integrated with our modular Flamingo systems to allow users to run circuits with up to 7,500 gates and 1,000+ qubits.	The performance of our Flamingo systems will improve to allow users to run circuits with up to 10,000 gates and 1,000+ qubits.	The Starling system will be available to clients. It will be a modular, error- corrected quantum-centric supercomputer with 200 qubits capable of running a total of 100 million gates.	Our 100,000-qubit Blue Jay system will define 2,000 qubits capable of running a total of 1 billion gates. The middleware will integrate this system into ever more powerful quantum-centric supercomputers.

# 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<sub>T</sub>>1 GeV in the barrel are considered.
- QUBO is computed event by event using <u>hepqpr-qallse framework</u>.

#### 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!

### Dataset

- Three sets of e+e- collision events are generated to consider various jet multiplicity:
  - $Z \rightarrow q\overline{q}$  ( $\sqrt{s}=91$  GeV, <u>2 jets</u>),
  - $ZH \rightarrow q\overline{q}b\overline{b}$  ( $\sqrt{s}=240$  GeV, <u>4 jets</u>)
  - $t\overline{t} \rightarrow b\overline{b}q\overline{q}\overline{q}q$  ( $\sqrt{s}=360$  GeV, <u>6 jets</u>)
- Delphes card with the CEPC 4<sup>th</sup>-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.



### **Reconstruction at LHC & HL-LHC**

![](_page_24_Figure_1.jpeg)

![](_page_24_Picture_2.jpeg)

- 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

# **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.) & tripletbased (F. Bapst et al.) approaches; quantum gate machines (L. Funcke et al., <u>H. Okawa</u>, etc.)

![](_page_25_Figure_4.jpeg)

# Simulated Bifurcation (SB)

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

>1048

~0.96

~0.05

0.12

< 0.001

32400

~2880

17.64

<0.68

84 >10<sup>190</sup>

![](_page_26_Figure_2.jpeg)

Ν	Connectivity	$J_{i,j}$	Machine	TTS		
60	All-to-all	{±1}	d <b>SBM</b> RBM CIM QA	<b>9.2 μs</b> 10 μs 0.6 ms 1.4 s		
100	All-to-all	{±1}	dSBM RBM SimCIM CIM	<b>29 μs</b> 30 μs 0.6 ms 3.0 ms		
200	Sparse (Degree 3)	{0, -1}	dSBM QA CIM	<b>0.70 ms</b> 11 ms 51 ms		
		Grap	h size Al ⊤	<b>gorithm</b>	Hardware CPU 1 core	<b>Time(s)</b>

Brute-force search<sup>46</sup>

Brute-force search<sup>46</sup>

Exact belief

QA<sup>13</sup>

bSB

bSB

TTN

TTN<sup>44</sup>

dSB dSB

Exact belief

propagation<sup>13</sup>

propagation<sup>13</sup>

 $4 \times 4 \times 8$ 

 $8 \times 8 \times 8$ 

- SB is known to outperform quantum annealing (QA) and other quantum-annealing-inspired algorithms for some problems.
- We have brought SB into high energy physics for the first time in the world!

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

GPU Tesla V100

GPU Titan V

CPU 1 core

CPU 1 core

CPU 1 core

GPU Titan V

CPU 1 core

CPU 1 core

GPU Tesla V100

GPU Tesla V100

D-Wave

### **Quantum-Inspired**

$$\begin{aligned} H_{\text{SB}}(\boldsymbol{x},\boldsymbol{y},t) &= \sum_{i=1}^{N} \frac{\Delta}{2} y_{i}^{2} + \sum_{i=1}^{N} \left[ \frac{K}{4} x_{i}^{4} + \frac{\Delta - p(t)}{2} x_{i}^{2} \right] - \frac{\xi_{0}}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} J_{ij} x_{i} x_{j} \\ \dot{x}_{i} &= \frac{\partial H_{\text{SB}}}{\partial y_{i}} = \Delta y_{i} \\ \dot{y}_{i} &= -\frac{\partial H_{\text{SB}}}{\partial x_{i}} = - \left[ K x_{i}^{2} - p(t) + \Delta \right] x_{i} + \xi_{0} \sum_{j=1}^{N} J_{i,j} x_{j} \end{aligned}$$

- "Quantum-annealing-inspired" algorithms search for ground • state through the classical time evolution of differential equations.
- Simulated bifurcation (SB) emulates quantum adiabatic • evolution of Kerr-nonlinear parametric oscillators, exhibiting bifurcation phenomena.
- Three variants exist depending on how one handles the • continuous treatment of the spins  $(x_i)$ : aSB, bSB, dSB 大川英希 第29届LHC Mini-workshop

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

![](_page_27_Figure_7.jpeg)

# Ising Energy w QAIAs

![](_page_28_Figure_1.jpeg)

- Ballistic simulated bifurcation can find <u>the lowest Ising energy with</u> 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**

![](_page_29_Figure_1.jpeg)

- Ballistic simulated bifurcation provides <u>4 orders of magnitude speed-up (23min → 0.14s) from D-</u> <u>Wave Neal</u> 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 <u>w/ multiple processing, GPU &</u> <u>FPGA</u> → 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<sub>t</sub>

![](_page_30_Figure_4.jpeg)

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).

大川英希

# **Ising Energy Prediction**

![](_page_31_Figure_1.jpeg)

- 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 tt events!

### **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.
- <u>bSB w/ ee-k<sub>t</sub> provides the</u> <u>highest efficiency & can</u> <u>reconstruct multijet events.</u>
- Angle-based method only works for dijet. → misses many jets
- D-Wave Neal has degraded performance already in dijet.
- dSB also has lower efficiency than bSB & cannot handle multijet.

![](_page_32_Figure_7.jpeg)

大川英希

![](_page_33_Figure_0.jpeg)

- Ising solvers usually continue to improve energy prediction w/ running time.
- <u>bSB significantly outperforms dSB & Neal</u> (& 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.

第29届LHC Mini-workshop