量子电路切割在经典态矢量模拟中的适用性与局限性

Mitsuhiro Matsumoto[1], * Shinichiro Sanji[1] 和 Takahiko Satoh[2]

[1] *Technology Laboratory, PwC Consulting LLC, Japan*

[2] *Faculty of Science and Technology, Keio University, Japan*

电路切割将大型量子电路划分为较小的子电路，这些子电路可以独立执行，并通过经典后处理重新组合。在采用全态重构的经典态矢量模拟中，运行时间取决于一种权衡：子电路规模缩小所带来的收益，与数量呈指数增长的子电路及全态重构所产生的开销之间的权衡。对于等分，我们推导出切割次数的阈值条件；低于该阈值时，切割能够缩短墙钟时间。态矢量实验验证了预测的加速边界，适用范围最高达 24 个量子比特；而最高达 30 个量子比特的运行时间分解则揭示了交叉点：在 $q \approx 18$ 和 $q \approx 22$ 处，合并的耗时分别超过预处理和子电路模拟。作为实践指南，我们表明，在 10分钟实际运行时间预算 下，双向切割相较于不切割的模拟，可将最大可行量子比特数提高 4 至 6 个。

arXiv:2603.01443v1 [quant-ph] 2 Mar 2026

I. 引言

量子计算有望在特定计算问题上实现超越经典计算的指数级加速 [1]。然而，当前的含噪声中等规模量子（NISQ）设备仍受到可用量子比特的数量和质量的严重限制 [2]。为克服这些限制，电路切割是一种将大型量子电路划分为较小子电路并独立执行的技术，已成为一种可将有效量子电路规模扩展至近期设备限制之外的方法 [3–5]。

现有关于电路切割的研究主要聚焦于降低分区所带来的开销 [6,7]。在经典模拟方面，混合薛定谔–费曼方法¨和基于张量网络的方法 [9, 10] 利用量子电路分区，使模拟能够突破朴素的内存限制；分区规模与子电路数量呈指数增长之间的根本权衡也已广为人知。经典模拟对于量子算法的验证、基准测试和工作流程层面的性能研究仍不可或缺，尤其是在关注端到端运行时间时。然而，对于采用全态重构的经典态矢量模拟的具体执行流程，端到端墙钟开销何时会超过子电路规模缩小所带来的收益，也就是相应的切割数，目前鲜少受到关注。针对切割阈值的封闭形式指导原则，有助于合理分配有限的经典计算资源，并明确电路切割在墙钟时间方面何时具有实际效益。

此外，电路切割也作为一种工作流程得到研究：它将大型量子电路拆分为多个子电路，再通过经典后处理将它们重新组合起来 [11]。实际的经典模拟前沿同样受到高性能态矢量模拟器的影响 [12–14], 张量网络后端 [15], 以及大规模分区演示 [16]。这些进展促使我们对电路切割进行端到端的实际运行时间分析，而这正是本文的重点。

本研究聚焦于带有全态重构的经典态矢量模拟，该场景出现在量子设备的基准测试、调试和验证中。切割会减小子电路规模，但使子电路数量呈指数增长；在此设定下，必须将结果合并为完整的 2^q维状态向量，并产生额外的 $O(2^q)$合并成本。因此，判断切割是否有益，需要对切割工作流程的全部阶段进行端到端的实际运行时间分析。据我们所知，针对这一特定设定的定量实际运行时间分析尚未得到系统报告。

我们分析了经典态矢量模拟中电路切割的实际运行时间效率，并提出三项主要贡献：

首先，我们将总切割时间分解为预处理、子电路模拟和合并，并分析它们的相对缩放关系。这表明，随着量子比特数增加，会出现两个瓶颈交叉点：合并超过预处理，以及合并超过子电路模拟。在我们的实现中，这些交叉点出现在中等量子比特数处（在我们的设置中约为 $q \approx 18$ 和 $q \approx 22$ ）。其次，我们据此成本模型推导出在等分为 n 个分段的条件下，切割能够产生净加速的最大总切割数 c_{total} 所需满足的条件。该阈值的形式为 $c_{\text{total}} \propto q$ ，并且近似与深度无关。最多24个量子比特的态矢量实验证实了所预测的阈值。

第三，在10分钟实际运行时间预算下，与不进行切割的模拟相比，双向切割可将最大可行量子比特数提高4至6个量子比特，且深度越大，提升越明显。

我们首先在第 II 节回顾电路切割方法及其计算特性。在第 III 节介绍执行时间模型及其区间结构，并推导等分为

* mitsuhiro.matsumoto@普华永道.com

个 n 分段的切割阈值条件。我们在第 V 节给出数值结果，并通过经典态矢量模拟验证预测趋势。最后，我们总结研究发现，并在第 VI 节作出结论。

II. 背景

A. 电路切割

电路切割将大型量子电路分解为一组较小的子电路，并通过经典后处理重新组合其结果。实践中，主要有两种代表性方法：(i)线切割 （通过插入测量-制备信道来切断“线”）和 (ii)门切割 （将跨越不同分区的非局域纠缠门替换为分解后的局域操作）。本文聚焦于经典状态向量场景下的门切割，其开销结构与硬件场景有本质区别。

1. *Quantum hardware setting*

在量子硬件上，电路切割会产生采样开销，且该开销随切割位置数量呈指数增长。对于门切割，每个非局域门都会通过准概率分解（QPD）替换为局域信道 [5]。针对多个双量子比特酉算子的切割，最优采样开销已有研究刻画 [17]。对于线切割，则通过插入测量-制备信道来切断量子比特线 [4]。切割 m 条并行线的采样成本按 $\widetilde{\mathcal{O}}(4^m/\epsilon^2)$ 增长（忽略多对数因子），其信息论下界为 $\Omega(2^m/\epsilon^2)$ [18]。近期研究探讨了在辅助量子比特使用量等不同操作约束下实现最优或近最优开销的线切割方案 [6, 7]。

2. *Classical state-vector setting*

当子电路通过态向量模拟进行经典模拟时，可以使用另一种分解。任意双量子比特酉算子 U 都允许算子施密特分解 [19]

$$U = \sum_{\ell=1}^{k} s_\ell \, (A_\ell \otimes B_\ell), \tag{1}$$

其中 A_ℓ 和 B_ℓ 是单量子比特算子， s_ℓ 是复系数， k 是 U的施密特秩。对于双量子比特门，施密特秩仅取 $k = 2$ （如CZ门和CX门等受控酉算子）或 $k = 4$（如SWAP等所有其他非局域门） [19]。

由于经典模拟器可以直接访问完整的态矢量，因此可以直接利用式（1）的线性性质：每一项都会产生一个独立的子电路态矢量，结果则以复向量的形式求和。每个切割点会产生 k 项，且没有采样开销。

例如，CZ门（$k = 2$）允许分解 [20]

$$\mathrm{CZ} = \frac{1}{1+i}\big(S \otimes S + i\, S^\dagger \otimes S^\dagger\big), \tag{2}$$

其中 $S = \mathrm{diag}(1, i)$是相位门。每一项都是单量子比特酉算子的张量积，因此在分区边界处切割一个CZ门会将其替换为两个子电路，分别对相应分段施加 S 或 $S^\dagger$ 。CX门通过CX $= (I \otimes H)$CZ $(I \otimes H)$相关联，并且在局部基变换的意义下允许相同的两项分解。当切割 c个量子门时，子电路数量按 k^c 增长（随切割次数呈指数增长）。

B. 经典模拟

薛定谔式（态矢量）模拟需要为一个包含 q 个量子比特的系统存储大小为 2^q 的态矢量。指数级的内存占用很快就会成为主要瓶颈。一个广为人知的参考点是“49 量子比特壁垒”，这促使人们采用分解/切片和存储感知策略，以突破朴素内存限制，实现更大规模的精确模拟 [10]。除精确模拟外，近似方法会有意牺牲准确性来换取规模扩展。一种方向是有损态矢量压缩，例如基于量化或压缩的方法，在控制由此产生的误差/保真度损失的同时，减少每个振幅所占的比特数 [21, 22]。另一种方向是针对结构化基准测试（例如 Sycamore 风格的随机量子电路），以目标保真度 近似生成样本，从而提供可扩展的经典基线，而无需计算完整的精确态 [9]。

对于具有有利结构（例如纠缠有限或树宽较小）的量子电路，张量网络模拟器 [23] 可能比状态向量模拟更有效。若干大规模近似采样结果也可以很自然地用张量网络术语来表述。本文聚焦于状态向量领域，并量化电路切割在何种情况下能够通过缩小子电路规模来换取切割开销，从而降低经典墙钟运行时间。

III. 执行时间模型

我们对单个计算节点上采用顺序子电路执行、且不使用门融合或并行化时的电路切割墙钟时间进行建模。在经典态矢量场景下，切割与重组过程能够精确还原未切割量子电路的完整态矢量（达到机器精度），因此不会产生保真度损失，

我们只关注执行时间的权衡。

A. 定义与成本分解

为评估电路切割是否能减少经典模拟时间，我们将总切割时间分解为

$$T_{\mathrm{cut}} = T_{\mathrm{pre}} + T_{\mathrm{sub}} + T_{\mathrm{merge}}, \tag{3}$$

其中， T_{pre} 是生成所有子电路及其相关系数的预处理时间， $T_{\mathrm{sub}} = \sum_{i=1}^{N_{\mathrm{sub}}} T^{(i)}$ SV是子电路模拟总时间（$T^{(i)}$ SV 是第 i 个子电路的态矢量模拟时间）， T_{merge}是通过合并子电路结果来重构完整态矢量的合并时间。比较基准为 $T_{\mathrm{orig}} = O(2^q D)$，即不进行切割时模拟原始电路的墙钟时间。

我们考虑一个包含 q 个量子比特、划分为 n 个片段的量子电路。令 c_i 表示与片段 i 相邻的所有边界上的量子门切割总数， q_i 表示片段 i 中的量子比特数， $c_{\max} =$ 表示 $\max_i c_i$。子电路总数为 $N_{\mathrm{sub}} = \sum_{i=1}^{n} k^{c_i}$，其中 k 是切割量子门的施密特秩。

各阶段的计算成本如表 I 所示。这里， D 表示原始电路的深度， $D_{\mathrm{seg},\ i} (\le D)$ 表示分区后片段 i的深度。我们记作 $D_{\mathrm{seg}} = \max_i D_{\mathrm{seg},\ i}$。由于未应用门融合，当量子门均匀分布在各片段中时，下界 $D_{\mathrm{seg}} \ge D/n$趋于达到，因此

$$\frac{D}{n} \le D_{\mathrm{seg}} \le D. \tag{4}$$

预处理阶段通过针对每个子电路遍历一次原始电路，生成所有 N_{sub} 子电路。每次子电路模拟都对包含 2^{q_i} 个振幅的缩减态矢量进行操作，但必须针对全部 N_{sub} 项重复执行。合并阶段通过对 $k^{c_{\max}}$ 个张量积求和，重构完整的 2^q维态矢量。如果每次子电路模拟后都丢弃态矢量，其内存成本为$O(2^{q_{\max}})$；如果同时保留所有态矢量，则为 $O(N_{\mathrm{sub}} \cdot 2^{q_{\max}})$，其中 $q_{\max} = \max_i q_i$。各阶段的伪代码见附录B。

B. 切割成本的区间结构

式 (3) 中的三个成本组成部分对量子电路参数的依赖关系在定性上各不相同。为确定哪个组成部分主导 T_{cut}，我们根据表 I 中的缩放关系比较这三个组成部分。

为得到闭式比值表达式，我们现在专门考虑等分（$q_i = q/n$ 对所有 i， $D_{\mathrm{seg},\ i} = D_{\mathrm{seg}}$ 对所有 i）。

表 I. 计算成本汇总。 q：量子比特总数， n：分段数， c_i：与片段 i 相邻的切割次数， $c_{\max} = \max_i c_i$， k：切割门的施密特秩， D：原始电路深度， D_{seg}：最大子电路深度， $N_{\mathrm{sub}} = \sum_i k^{c_i}$：子电路总数。

阶段	Time	内存
预处理	$O(N_{\mathrm{sub}}\, D)$	$O(N_{\mathrm{sub}}\, D)$
子电路模拟	$O(N_{\mathrm{sub}}\, 2^{q_{\max}}\, D_{\mathrm{seg}})$	$O(2^{q_{\max}})$
合并	$O(k^{c_{\max}}\, 2^q)$	$O(2^q)$

1. T_{pre} *vs* T_{sub}

根据表 I，

$$\frac{T_{\mathrm{pre}}}{T_{\mathrm{sub}}} = \frac{D}{2^{q/n}\, D_{\mathrm{seg}}}. \tag{5}$$

该比值与 k 和 c_i无关。由公式（4）可得，

$$\frac{1}{2^{q/n}} \le \frac{T_{\mathrm{pre}}}{T_{\mathrm{sub}}} \le \frac{n}{2^{q/n}}. \tag{6}$$

由于 $2 \le n \le q$ 和 $2^{q/n} \ge 2$，上界 $n/2^{q/n}$会随 q/n增大而趋于零，因此只要q/n 足够大， $T_{\mathrm{pre}} \ll T_{\mathrm{sub}}$ 就成立。对于较小的 q/n （例如 $q/n \lesssim 5$），两者可能相当。

2. T_{pre} *vs* T_{merge}

由表 I 可得，

$$\frac{T_{\mathrm{pre}}}{T_{\mathrm{merge}}} = \frac{N_{\mathrm{sub}}}{k^{c_{\max}}} \cdot \frac{D}{2^q}. \tag{7}$$

第一项满足

$$\frac{N_{\mathrm{sub}}}{k^{c_{\max}}} = \sum_{i=1}^{n} k^{c_i - c_{\max}} \le n, \tag{8}$$

因为每一项$k^{c_i - c_{\max}} \le 1$。当所有片段的切割次数相同时，该界达到饱和值；随着切割分布变得越不均匀，该值便逐渐趋近于 1。无论哪种情况，这一因子都是 $O(n)$，不会引入对 q 或 c的指数依赖。因此，该比值由 $D/2^q$控制。当 $2^q \gg D$时，无论 n如何， $T_{\mathrm{pre}} \ll T_{\mathrm{merge}}$ 都成立。对于较小的 q ，或者在 D 和 2^q 相当的极深电路中， T_{pre} 可能与 T_{merge}相当。结合第 III B 1 节的结果，只要 q/n 和 q 都足够大， T_{pre}就是 T_{cut} 中最小的部分。

3. T_{sub} *vs* T_{merge}

根据表 I，

$$\frac{T_{\mathrm{sub}}}{T_{\mathrm{merge}}} = \frac{N_{\mathrm{sub}}}{k^{c_{\max}}} \cdot \frac{D_{\mathrm{seg}}}{2^{q(1-1/n)}}. \tag{9}$$

根据式（8），第一个因子为 $O(n)$。因此，区间边界由以下因素决定

$$\frac{D_{\text{seg}}}{2^{q(1-1/n)}} \approx 1. \tag{10}$$

当 $D_{\text{seg}} \gg 2^{q(1-1/n)}$ （ q 较小或 D_{seg}较大）时， T_{sub}占主导地位（子电路主导区间）。在相反的极限情况下 $D_{\text{seg}} \ll 2^{q(1-1/n)}$ （ q 较大或 D_{seg}较小），T_{merge} 占主导地位（合并主导区间）。

4. *Crossover structure*

前述比值表明，在深度固定时，增加 q 会导致两个交叉点： T_{merge} 超过 T_{pre}和 T_{sub}。一旦 T_{merge} 超过二者，它就成为 T_{cut}的主导组成部分。由于 T_{merge} 不会因子电路规模缩小而受益，因此在此区间内，电路切割的优势受到限制。在第 IV 节中，我们将此成本模型用于推导定量的阈值条件，并在第 V 节中通过实际运行时间测量对其进行验证。

IV. 阈值分析

我们推导切割次数的条件，在满足这些条件时，电路切割能够缩短经典模拟时间。我们考虑一个包含 q 个量子比特、深度为 D 的量子电路，将其划分为 n 个大小相等的分段，每个分段包含 q/n 个量子比特，每个边界处有 c 次切割，总计 $c_{\text{total}} = c\,(n-1)$ 次切割。我们假设量子门均匀分布在各分段中，因此每个子电路的深度为 $D_{\text{seg}} \approx D/n$。

A. 成本比较框架

总切割时间 $T_{\text{cut}} = T_{\text{pre}} + T_{\text{sub}} + T_{\text{merge}}$由三个部分组成，其缩放关系见表 I。其中， T_{pre} 从不占主导地位。第 III B 节中的比率分析表明， $T_{\text{pre}}/T_{\text{sub}}$ 和 $T_{\text{pre}}/T_{\text{merge}}$ 都会随 q增加而降低，而第 V C 节中的运行时间分解证实，在所有测试配置中， T_{pre} 都不会超过 T_{cut}并成为主导项。因此，我们在阈值分析中省略 T_{pre}。

其余两个组成部分涉及性质上不同的运算。 T_{sub} 和 T_{orig} 均使用同一模拟器执行逐量子门状态向量模拟，区别仅在于态矢量的规模和量子门的数量。这些差异已由表 I 中的显式因子 $2^{q/n}$ 和 D_{seg} 所体现，因此 T_{sub} 与 T_{orig} 的数量级比较是可靠的。我们由此二者推导出阈值。相比之下， T_{merge} 执行向量化张量积和向量加法，这与逐量子门模拟属于不同类型的计算。这两类运算的相对开销取决于实现，且并未被 $O(\cdot)$标度所涵盖。因此， T_{merge} 未纳入理论阈值，但在实证评估中予以计入。

B. 等分阈值

1. *Threshold under uniform gate distribution*

在本节的等分设置下，根据表 I 比较 T_{sub} 和 T_{orig}，

$$N_{\text{sub}} \cdot 2^{q/n} \cdot \frac{D}{n} \;<\; 2^{q} \cdot D, \tag{11}$$

化简为

$$\frac{N_{\text{sub}}}{n} \;<\; 2^{q(1-1/n)}. \tag{12}$$

深度 D 相互抵消，因此该条件仅取决于q、 n、 c和 k。

现在令 $k = 2$ （CZ 门切割），并推导出明确的阈值。

a. 二分情形（$n = 2$）。当 $N_{\text{sub}} = 2 \cdot 2^{c_{\text{total}}}$时，式 (12) 给出

$$c_{\text{total}} < \frac{q}{2}. \tag{13}$$

b. 一般 $n \geq 3$。在N_{sub}中，内部项占主导地位，得到 $N_{\text{sub}} \approx (n-2) \cdot 2^{2c_{\text{total}}/(n-1)}$。代入式 (12) 并取对数，得到

$$c_{\text{total}} < \frac{q(n-1)^2}{2n} + \frac{(n-1)\log_2\!\left(\frac{n}{n-2}\right)}{2}. \tag{14}$$

在（q， c_{total}）平面中的主斜率为（$n-1$）$^2/(2n)$，且随 n 增大。我们将式 (14) 中的第二项记为 δ。对于实际相关的 n 取值（最高约为 10）， δ 仍接近 1。5。

表 II 总结了 n =2、3 和 4 的阈值参数。这些阈值可保证 $T_{\text{sub}} < T_{\text{orig}}$，并为选择切割次数提供了实用指南。 T_{merge} 的额外贡献会随着 $O(k^{c_{\max}} \cdot 2^{q})$ 增长，但未计入这些阈值中，并且在 q 较大时可能变得显著。

表 II. $k = 2$ （CZ 门切割）的子电路模拟阈值 $c_{\text{total}} < c_{\text{total}}^{\max}$。该阈值具有 $c_{\text{total}} <$ 斜率 $\times\, q + \delta$ 的形式。这些条件将 T_{sub} 与 T_{orig} 进行比较，且不包括 T_{merge}。

n	斜率	δ
2	1/2	0
3	2/3	$\log_2 3 \approx 1.58$
4	9/8	3/2

2. Effect of non-uniform gate distribution

当量子门未均匀分布在各个区段时，某些子电路包含的量子门会多于其他子电路，从而导致 $D_{\mathrm{seg}} > D/n$。对于一般的 D_{seg}，条件 $T_{\mathrm{sub}} < T_{\mathrm{orig}}$ 变为

$$\frac{N_{\mathrm{sub}}}{2^{q(1-1/n)}} \cdot \frac{D_{\mathrm{seg}}}{D} < 1. \tag{15}$$

由于 $D/n \le D_{\mathrm{seg}} \le D$，第二个因子 D_{seg}/D 的范围从 $1/n$（均匀分布）到 1（最坏情况）。对于 $n = 2$，c_{total}阈值比公式（13）少 1 次切割。对于$n \ge 3$，降幅为 $(n-1)\log_2(n)/2$，对于 $n = 3$ 和 $n = 4$，该值分别为 1.6 和 3.0。均匀深度阈值代表最有利的情况。对于较小的 n，降幅不大，但会随 n 增大而增加，因此对于较大的 n，应考虑实际的门分布。

本节推导的阈值将 T_{sub} 与 T_{orig} 进行比较，且不包含 T_{merge}。T_{merge} 对总切割时间的贡献将在第 V C 节通过实证量化。

V. 评估

我们在经典状态向量模拟器上验证阈值条件，并报告运行时间分解。

A. 实验设置

1. Simulation environment

所有实验均在一台运行 Python 3.10.8 和 Qiskit 2.2.3 的单节点计算机（Apple M3 Max，128 GB RAM）上进行。状态向量模拟使用 Statevector 类（qiskit.quantum_info）完成。所有运行的墙钟时间均在同一台机器上测量。对于运行时间分解测量（图 3），每次运行之间都会显式释放内存，以避免量子比特数较大时预处理时间增长。除非另有说明，每种配置均测量十次，并报告平均值。

2. Circuit family and partitioning

我们使用合成电路族进行阈值验证。图 1 展示了一个代表性实例。对于两次或更多次切割，在移除初始哈达玛门后，通过追加该量子电路的额外副本来扩展该量子电路。这种交替单量子比特门层与双量子比特门层的重复块结构，反映了实际变分与模拟量子电路的布局，例如变分量子本征求解器中使用的 Ans¨atze [24, 25]，QAOA 量子电路 [26]，以及 Trotter 化哈密顿量模拟 [27]。在所有这些情形中，增加重复层数会同时增加电路深度以及跨越分区边界的双量子比特门数量。

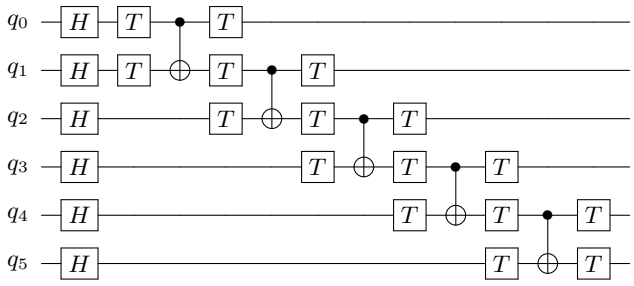

图 1. 一个可通过一次切割进行分区的原始电路。对于两次或更多次切割，在移除初始哈达玛门后，通过在该电路后追加其副本来扩展电路。

我们将电路等分为 n 个子电路，每个子电路包含 q/n 个量子比特，因为这种方法可将模拟时间降至最低，附录 A1 已对此进行了验证。本节其余部分的所有实验均采用等分方式。

3. Runtime measurement and evaluation

我们测量 T_{cut} 及其三个组成部分 T_{pre}、T_{sub} 和 T_{merge}，定义见公式 (3)。对于 $(q,\ c_{\mathrm{total}})$ 的每种组合，我们评估相对加速比

$$\Delta = (T_{\mathrm{cut}} - T_{\mathrm{orig}})/T_{\mathrm{orig}}, \tag{16}$$

其中 $\Delta < 0$ 表示加速。

B. 阈值条件的验证

使用第 IV.A 节所述的实验设置，图 2 中的热图展示了 $n = 2$ 和 $n = 3$ 条件下 $(q,\ c_{\mathrm{total}})$ 平面上的相对加速比 Δ。在这两种情况下，加速区域（$\Delta < 0$，蓝色）与减速区域（$\Delta > 0$，红色）之间的边界总体上符合第 IV 节推导的理论阈值。

为了量化测量边界，我们使用加权支持向量机（SVM）分类器为每个面板拟合线性决策边界，如图 2 中的实线所示。SVM 拟合的边界为 $c_{\mathrm{total}} = 0.70\,q - 7.30$ $(n = 2)$ 和 $c_{\mathrm{total}} = 1.00\,q - 11.50$ $(n = 3)$，而理论阈值（式 (13) 和式 (14)）分别为$c_{\mathrm{total}} < q/2$ $(n = 2)$ 和 $c_{\mathrm{total}} < 2q/3 + \log_2 3$ $(n = 3)$。比较两者可知，测量边界的斜率更大，截距也更负。

这种差异源于理论阈值只比较 T_{sub} 和 T_{orig}，而忽略了 T_{pre} 和T_{merge}。在基准电路中，深度 D 随 c_{total}增加。由于 $T_{\mathrm{merge}} = O(k^{c_{\max}} \cdot 2^q)$ 与深度无关，而 $T_{\mathrm{orig}} = O(2^q \cdot D)$，因此随着 c_{total}（进而 D）增大，比值 $T_{\mathrm{merge}}/T_{\mathrm{orig}}$会下降，使测得的斜率变得更陡。

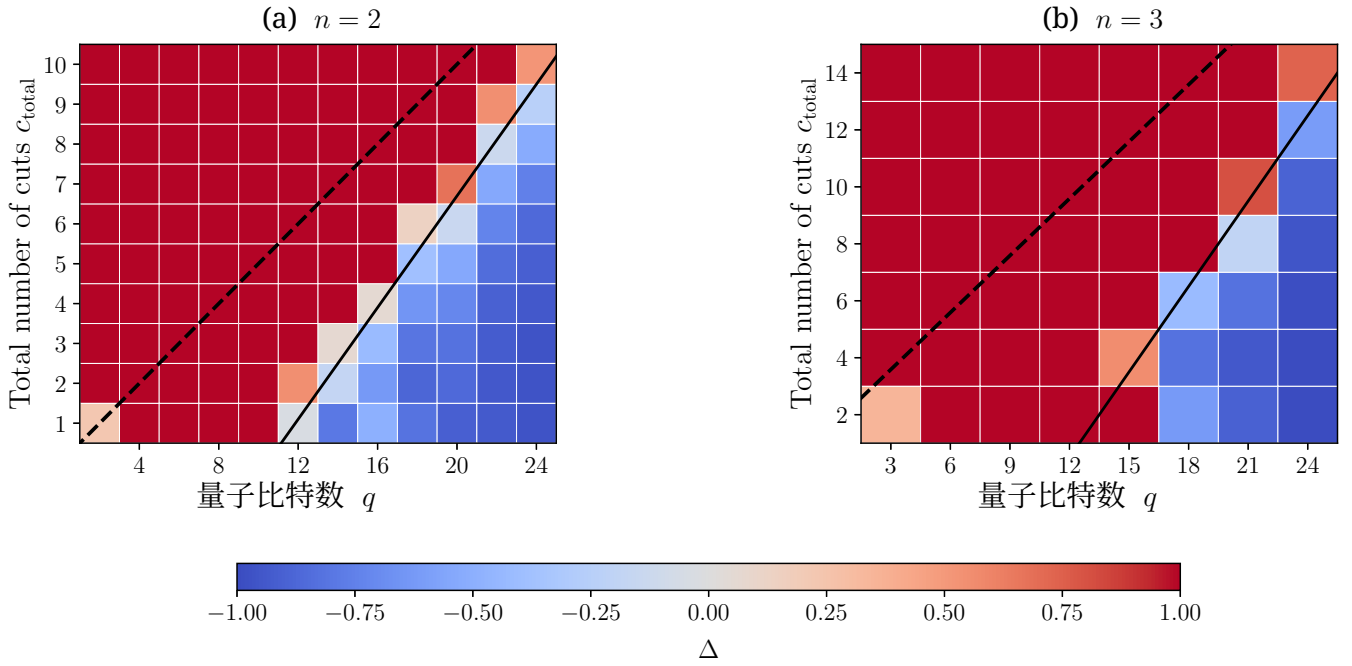

图 2. 测得的相对加速比 $\Delta = (T_{\text{cut}} - T_{\text{orig}})/T_{\text{orig}}$ 在 $(q,\, c_{\text{total}})$ 平面上，对应于 (a) 双向 ($n = 2$, $q = 2, 4, \ldots, 24$, $c_{\text{total}} = 1$ 10 $n = 3$ $q = 3$ 6 24 $c_{\text{total}} = 2$ 4 14 ,...,) 与 (b) 三向 (,,,...,,,,...,) 等分。蓝色表示加速 ($\Delta < 0$)；红色表示减速 ($\Delta > 0$)。虚线所示为第 IV 节推导的理论阈值。实线所示为由加权支持向量机分类器对测量数据拟合的线性边界。两条边界均证实，加速区域随 q 线性增长，但测量边界比理论预测更陡，详见第 V 节 B。

与此同时，$T_{\text{pre}} = O(N_{\text{sub}}\, D)$ 随 q 的增长速度远低于 T_{sub}，形成一个几乎恒定的偏移量，从而降低截距。对于 $n = 2$ 和 $n = 3$，线性形式$c_{\text{total}} \propto q$ 都得以保留，这证实理论阈值捕捉到了正确的缩放关系。

C. 运行时间分解与瓶颈交叉 vers

为识别电路切割的主要成本因素，我们将总切割时间分解为三个阶段：预处理（T_{pre}）、子电路模拟（T_{sub}）和合并（T_{merge}）。我们在图3中报告了双向切割在 $c_{\text{total}} = 6$条件下这三个分量随 q 的变化。每个数据点表示十次独立运行的平均值。

对于小型量子电路（$q \lesssim 16$），预处理和子电路模拟占切割总时间的大部分，而合并成本可忽略不计（$< 10^{-1}$ 秒）。然而，T_{merge} 随 q增长最为陡峭，因为合并步骤根据子电路结果重构完整的 2^q维态矢量。这导致两个交叉点:

- **第一交叉点**（$q \approx 18$）：$T_{\text{合并}}$超过 $T_{\text{预处理}}$，表明合并相对于预处理不再是次要开销。

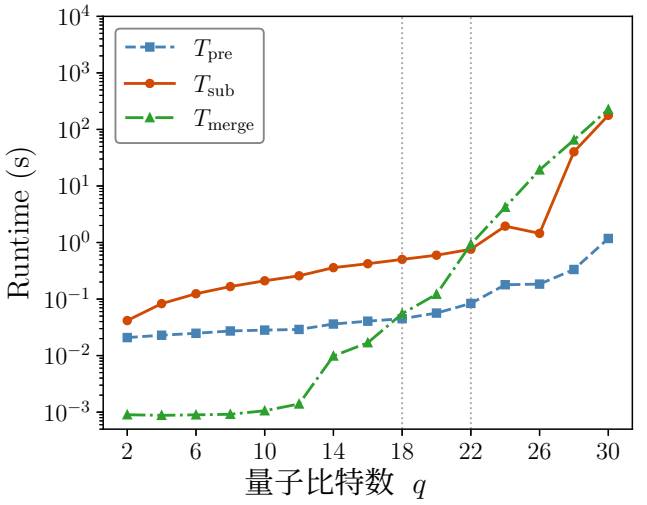

图3. 含$c_{\text{total}} = 6$ 次切割的双向电路切割的运行时间分解，作为量子比特数 q的函数。图中展示三个阶段：预处理（T_{pre}）、子电路模拟（T_{sub}）和合并（T_{merge}）。垂直虚线标出两个交叉点：约在 $q \approx 18$处，合并相对于预处理不再是次要开销；约在 $q \approx 22$处，合并超过子电路模拟。

- **第二个交叉点**（$q \approx 22$）：$T_{\text{合并}}$超越$T_{\text{子电路模拟}}$，成为三者中成本最高的阶段

阶段。对于 $q = 30$，合并约占总切割时间的 56%，子电路模拟占 43%，而预处理所占比例可以忽略不计（< 1%）。

在附录 A 的图 7 中，其他切割数量也呈现出相同的定性趋势。

这些交叉点表明，针对不同规模的量子电路，应采用不同的优化目标。对于小型量子电路（在我们的设置中为 $q \lesssim 18$ ），总切割时间主要由预处理和子电路模拟决定，标准优化策略（例如门融合、并行子电路模拟）最为有效。对于较大量子电路（在我们的设置中为 $q \gtrsim 22$），优化合并步骤（例如采用稀疏或压缩的状态向量表示 [21, 22], ）最能缩短总切割时间。

D. 深度扫描与运行时间细分

由于在均匀门分布下比较 T_{sub} 与 T_{orig} 时，D 会抵消，因此方程（13）和（14）中的阈值不包含电路深度 D。因此，加速边界预计在不同深度下保持稳定。下文的深度扫描将对此预期进行实证检验。

然而，在实际的状态向量模拟中，墙钟时间也取决于量子电路深度（或等价地，量子门的应用次数）。为量化这一影响并评估阈值边界的稳健性，我们通过在基准量子电路的顶部和底部量子比特上附加 10^p 层单量子比特门（交替使用 T和 X），同时保持（q，c_{total}）不变，进行合成深度扫描。量子电路图见附录 A 中的图 8。

我们针对每种深度配置提取（q，c_{total}）平面上的 $\Delta = 0$ 边界，并将它们叠加绘制于图 4 中。具体而言，对于每个深度 D，我们计算 $\Delta(q, c_{\text{total}}; D) = (T_{\text{cut}}(q, c_{\text{total}}; D) - T_{\text{orig}}(q; D)) / T_{\text{orig}}(q; D)$，并使用加权支持向量机分类器拟合线性决策边界。每条拟合线下方的区域表示切割有益。所有测试深度（p =0，2，3，4）下得到的边界彼此接近，证实加速与减速之间的转变在定性上保持稳定。深度增加会带来两种相互竞争的效应：T_{orig} 随深度增加，使切割开销相对减小；与此同时，T_{sub} 也随深度增加，部分抵消了这一优势。这些效应大体相互抵消，因此加速边界基本保持稳定。这种稳定性指的是整体边界。在单独的对于不同的配置，深度仍可能使天平倾向于切割，详情见附录 A 中的图 9。

为提供实际操作指南，我们还估算了在固定墙钟时间预算内，原始模拟和双向切割各自可行的最大量子比特数，如图5所示。我们将时间限制设为

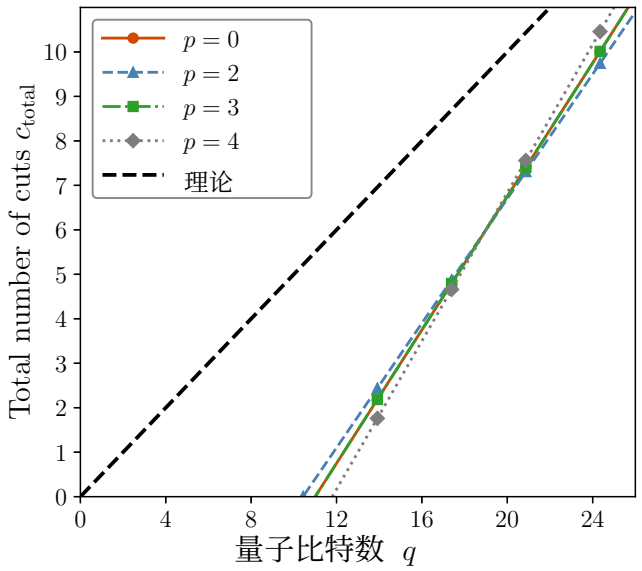

图 4. 四种深度配置下，双向分区（$n = 2$）在（q，c_{total}）平面上的加速边界（p =0，2，3，4，分别对应 $D \propto 1$，10_2，10_3，10_4 个额外量子门层）。每条彩色线表示在相应深度下通过加权支持向量机拟合得到的 $\Delta = 0$ 边界。每条线以下的区域切割更有利。黑色虚线表示第 IV 节推导出的理论阈值。在所有测试深度下，这些边界彼此接近，证实基于量子比特数的阈值即使对于深度各异的量子电路，也能提供合理的一阶指导。

每次模拟运行10分钟。选择这一时间限制是因为：(i) 这对迭代式量子电路设计和验证工作流程而言切实可行；(ii) 所有报告的结果都完全基于实测模拟时间，不涉及任何外推。每种配置要么在时限内完成，要么超时，因此可行性边界是精确的。每种配置均独立运行三次。各次运行的结果一致，图中报告的是平均值。在这10分钟预算内，原始（未切割）模拟在所有深度配置下最多可处理22–26个量子比特（$p = 0$，2、3、4），而双向切割将可行范围扩展至28–30个量子比特，增加了4至6个量子比特，如图5中的箭头标注所示。阴影区域突显了切割在所有深度下带来的增益。

此外，随着深度从 $p = 0$ 增至 $p = 4$，收益从 4 个量子比特时的水平增至 6 个量子比特时的水平，这与图 9（附录 A）所示的观察结果一致：T_{cut} 随深度增长的速度低于T_{orig}。

E. 适用范围与局限

本节的实验在测试范围内证实了四种定性趋势（阈值边界见$q \leq 24$，运行时间分解见 $q \leq 30$ ）：（i）

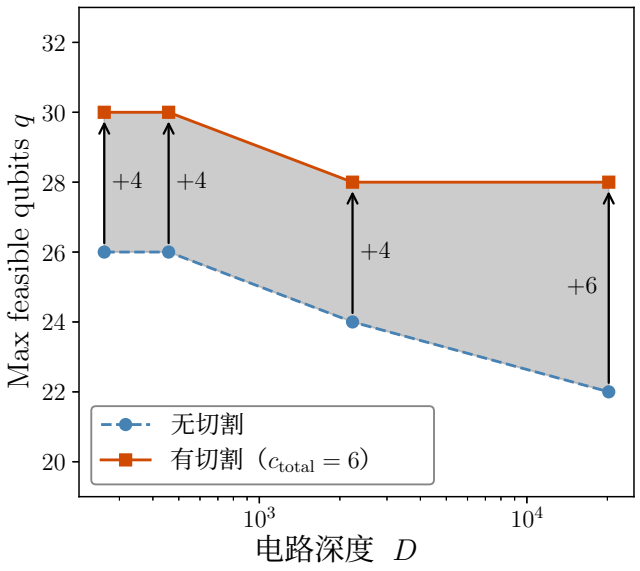

图5. 在10分钟实际运行时间预算内可达到的最大可行量子比特数，作为电路深度 D的函数。蓝色虚线表示原始（未切割）模拟，橙色实线表示采用 $c_{\text{total}} = 6$的双向切割。阴影区域突出了切割带来的改进。标注箭头标示各深度下的量子比特数增益：$p = 0$，2，3为4个量子比特，$p = 4$为6个量子比特，将可行范围从22–26（无切割）扩展至28–30（有切割）。

加速边界与 q呈线性关系；（ii）在量子比特数较大时，T_{merge} 成为主要成本组成部分；（iii）在很宽的量子电路深度范围内，边界基本保持稳定；（iv）在实际运行时间预算固定的情况下，电路切割能够扩展可行模拟范围。这四种趋势都源于状态向量模拟的指数级扩展，预计在不同平台和模拟器实现中仍然成立。不过，定量数值——包括支持向量机拟合斜率和截距、瓶颈交叉点以及可行性提升——取决于我们使用的单节点环境、基准电路族和模拟器实现，因此应针对每个目标平台重新校准。运行时间分解方法以及在（q，c_{total}）平面上的阈值分析可直接应用于其他环境，并为此类重新校准提供框架。

执行时间按照电路深度 D 来讨论，遵循模拟复杂度研究中的惯例。对于 Qiskit Statevector 等顺序门应用模拟器，墙钟时间随门数 G 而非 D变化；但在阈值分析和数值实验所假设的均匀门分布下，两者是一致的。

VI. 结论

我们研究了在何种实际条件下，量子电路切割能够加速带有全态重构的经典态矢量模拟。我们的主要贡献总结如下。

(i) 将总切割时间分解为预处理、子电路模拟和合并后，揭示了两个瓶颈交叉点：合并先超过预处理，随后超过子电路模拟。在我们的实现中，这些交叉点分别出现在 $q \approx 18$ 和$q \approx 22$附近。

（ii）根据该成本模型，我们推导出切割能够产生净加速的最大切割数 c_{total} 所需满足的条件。该阈值的形式为 $c_{\text{total}} \propto q$，并且近似与深度无关。最多24个量子比特的状态向量实验证实了所预测的阈值。

（iii）作为实际示例，我们估算了在10分钟实际运行时间预算下可行的最大量子比特数。与不进行切割的模拟相比，双向切割将该上限扩展了4至6个量子比特，且增益随深度增大而增加。

这些结果也为状态向量模拟中众所周知的49量子比特壁垒 [10]提供了一种视角。量子电路切割并不能无条件地突破这一壁垒，因为切割数的指数开销意味着只有需要较少切割次数的量子电路才能实现净加速。然而，阈值条件提供了一种实用的预筛选标准：给定量子电路的（q, c_{total}）参数，即可预先判断切割是否可能产生净加速，而无需投入完整模拟。

实验验证涵盖了阈值边界的 $q \leq 24$ 以及运行时间分解的 $q \leq 30$ 。超出这一范围后，阈值预测依赖于对数量级缩放关系的外推，而内存限制成为单节点状态向量模拟的实际瓶颈。本文报告的定量交叉点和可行性提升仅适用于我们的单节点环境，因此应针对每个目标平台重新校准。我们预计，阈值的函数形式和区间结构在不同环境中仍会保持。

本研究聚焦于全态重构，此时 $O(2^q)$ 合并成本不可避免。若只需计算期望值，则可用对子电路期望值进行经典求和来替代合并阶段，从而消除 $O(2^q)$ 成本，并改变区间结构。将当前框架扩展到这一情形是自然的下一步。

还有若干值得探索的方向。首先，可通过扩展至非对称和拓扑感知划分来完善阈值模型；并且，可借助并行执行子电路和更高效的合并算法，进一步放宽有效阈值。其次，将这一框架应用于特定应用电路，例如量子化学和组合优化中涉及的电路，将有助于在实际工作流程中预先评估

电路切割能否带来净加速。第三，若将量子执行成本纳入考量，包括测量次数和量子态层析开销，就能将分析扩展至量子—经典交叉区间。这需要为量子侧单独建立成本模型，因此留待未来工作。

致谢

本文基于一项由新能源与工业技术开发组织（NEDO）委托开展的项目 JPNP23003 所取得的成果。作者感谢 Junya Nakamura、Tsuyoshi Kitano、Hana Ebi、Hideaki Kawaguchi、Takaharu Yoshida、Hiroki Kuji 和 Shigetora Miyashita 所进行的富有成效的讨论。TS 获得日本科学技术振兴机构（JST）项目资助，项目编号为 JPMJPF2221。

附录 A：补充图表

本附录汇集了支持性可视化图表，为正文中的论点提供更多证据。内容包括等分的最优性、不同切割数量下的运行时间分解，以及深度扫描的详细信息。

1. 等分的最优性

我们验证了等分能够最大限度地缩短总模拟时间。将一个 20 量子比特量子电路划分为两个子电路，并将上方子电路中的量子比特数量从 1 逐步调整到 19。对于每种配置，在 1 至 5 次切割的情况下分别测量模拟时间 20 次，并计算均值和标准差。如图 6 所示，当量子电路被等分为两半（每个子电路含 $q/2 = 10$ 个量子比特）时，模拟时间比达到最小。

2. 不同切割数量下的运行时间分解

图 7 展示了四种具有代表性的切割数量（$c = 3$、6、9、12）下双向切割的运行时间分解，其中包括图 3 中的 $c = 6$ 情况。所有切割数量下均呈现出与图 3 相同的定性趋势： T_{merge} 随着 q 增长最快，从而产生两个交叉点，与第 VC 节中所述的交叉点类似。随着 c 增加，交叉点略微向较大的 q 移动，但运行时间分解的整体结构保持不变。

3. 深度扫描：量子电路与运行时间分解

第 VD 节深度扫描中使用的深度增强电路如图 8 所示。10^p 层交替的 T

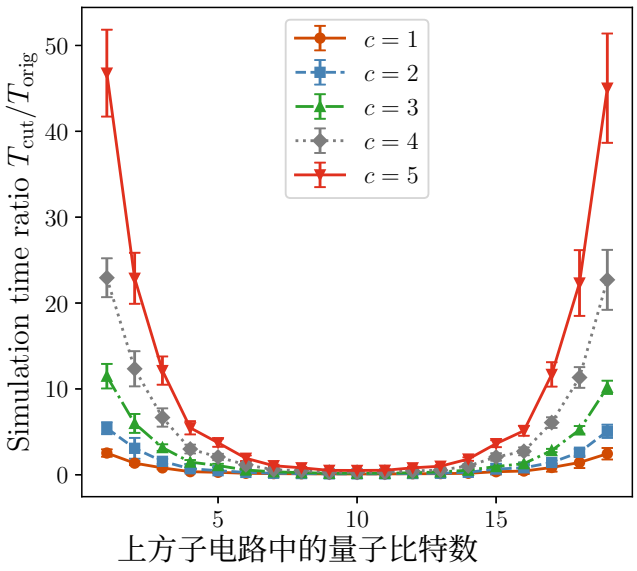

图 6. 模拟时间比 $T_{\mathrm{cut}}/T_{\mathrm{orig}}$ 随上子电路中量子比特数的变化关系，针对一个被划分为两个子电路、具有 $c = 1, \ldots\ldots, 5$ 次切割的 20 量子比特量子电路。对于所有切割数量，最小值均出现在 $q/2 = 10$ ，这证实等分可使总模拟时间最小。

和 X 量子门被附加到基准电路的顶部和底部量子比特上，同时量子比特数量 q 和切割 c 保持不变。

我们报告了代表性配置（$q = 24$、 $c = 10$、二路划分）下运行时间分解 $\{T_{\mathrm{pre}}, T_{\mathrm{sub}}, T_{\mathrm{merge}}\}$随电路深度 D 的变化，见图 9。在三个分量中， T_{pre} 和 T_{sub} 随 D 近似线性增长，而 T_{merge} 近似与深度无关。因此， T_{cut} 内的主导分量在$D \approx 750$附近从 T_{merge} 转变为 T_{sub} 。由于 T_{merge} 不随深度增长， T_{cut}的增长速度慢于 T_{orig}，并且 T_{orig} 在 $D \approx 3000$附近超过 T_{cut}。

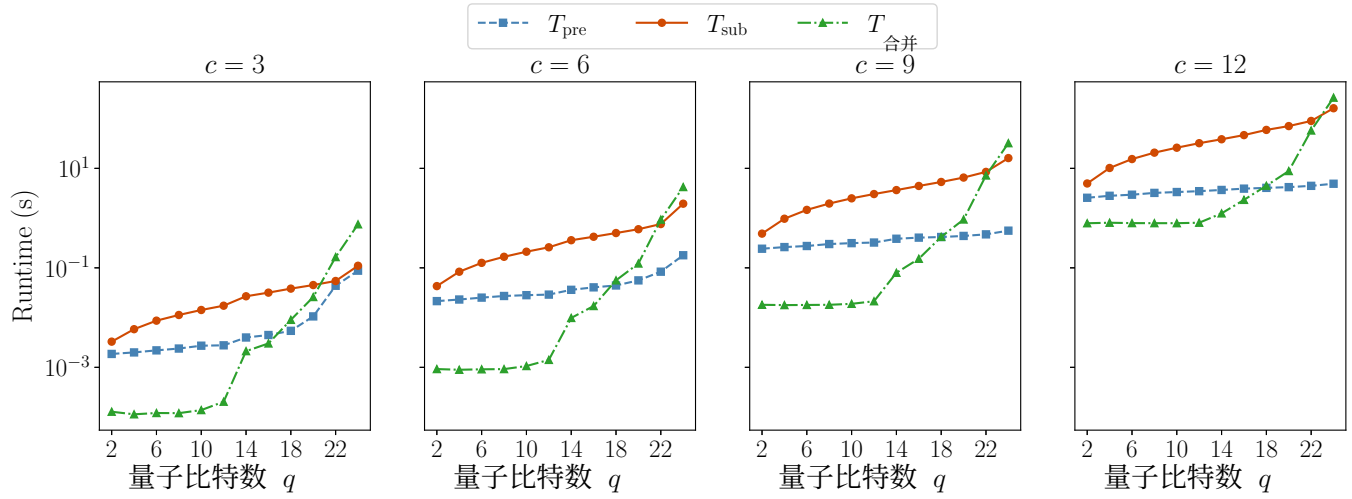

图 7. 双向切割（$n = 2$）在四个代表性切割数量 $c = 3, 6, 9, 12$ 下的运行时间分解。每个面板展示了三个成本组成部分：预处理（T_{pre}）、子电路模拟（T_{sub}）和合并（T_{merge}），它们作为量子比特数量 q 的函数。在所有切割数量下均观察到与图 3 相同的定性趋势： T_{merge} 随 q 增长最为陡峭，产生了与第 V C 节中所识别的类似的两个交叉点。 $c = 6$ 面板对应于正文中的图 3，并出于便于比较而纳入。

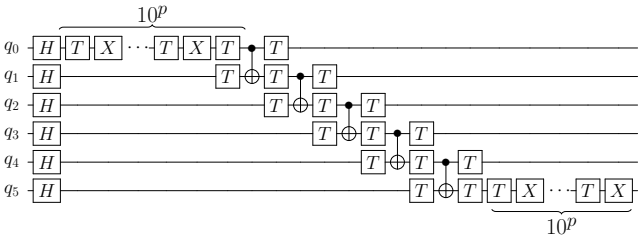

图 8. 深度扫描中使用的深度增强电路。将 10^p 层交替的 T 与 X 量子门附加到基准电路的顶部和底部量子比特上。量子比特数量 q 和切割 c 保持固定，仅改变电路深度。

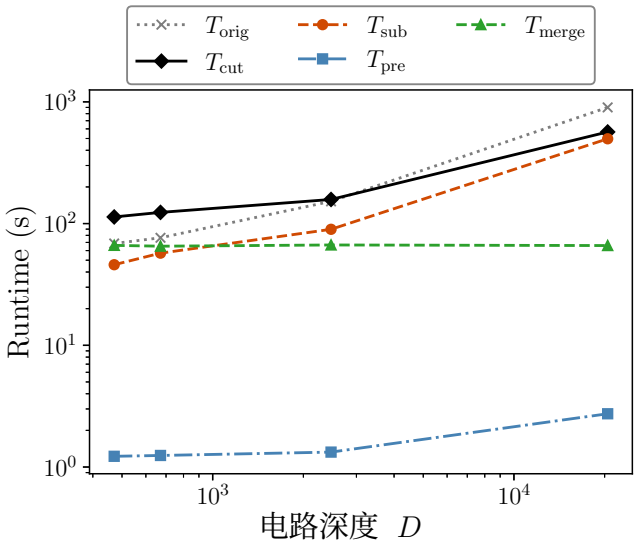

图 9. 针对具有 10 次切割、划分为 2 个分段的 24 量子比特量子电路，运行时间分解作为电路深度 D 的函数。总切割时间（T_{cut}）主要由子电路模拟时间（T_{sub}）构成，二者随深度呈现出相似的缩放趋势。相比之下，合并开销（T_{merge}）大致保持恒定，而预处理成本（T_{pre}）对深度的依赖较弱。在 $D \approx 750$ 附近， T_{cut} 内的主要瓶颈从 T_{merge} 转移到 T_{sub}。无切割的原始模拟时间（T_{orig}）比 T_{cut} 增长更快，并在 $D \approx 3000$ 附近超过后者，表明即使对于在较小深度下无法获得加速的配置，电路切割在足够大的深度下也可能变得有利。

附录B：伪代码和计算成本

本附录给出了表 I 中成本表达式所依据的伪代码。所有伪代码均假设采用 CZ 门切割，并使用 $k = 2$。在预处理阶段（算法1）中，首先将分区边界上的每个 CX 门通过 CX $= (I \otimes H)$ CZ $(I \otimes H)$ 转换为 CZ 门，然后再进行分解。

算法1预处理

输入：包含 q 个量子比特的原始电路 $\mathcal{C}$，量子比特分区 $\{G_1, \ldots, G_n\}$ 输出：片段 $i = 1$、.....、n 的子电路 $\{S_i^{(r)}\}$ 及分解索引 $r = 1$、.....、2^{c_i}，其中 c_i 是与片段 i 相邻的所有分区边界上的切割总数。合并索引表 $\mathcal{M}$。1：扫描 $\mathcal{C}$，识别所有跨越分区边界的 CX 门 2：对每个片段 $i = 1$、.....、n 执行 3：在每个边界门处拆分 $\mathcal{C}$，得到门列表片段$F_i^{(0)}, F$ **(1)** $_i$、.....、 $F(c_i)$，这些片段仅包含 G_i 中的量子比特 4：对每种组合 b $\in \{0,1\}^{c_i}$ 执行5 $S_i^{(\mathbf{b})} \leftarrow F_i^{(0)}$: : 对 6 $j = 1$、.....、c_i 执行 7：如果 $b_j = 0$，则 8：将 HSH（或 S）门追加到切割量子比特 9：否则 10：将 $HS^\dagger H$（或 $S^\dagger$）门追加到切割量子比特 11：结束 if 12：追加片段 $F_i^{(j)}$ 13：结束 for 14：结束 for 15：结束 for 16：构建合并索引表 $\mathcal{M}$，将每个全局组合索引映射到各片段索引 b

复杂度。 时间： $O(N_{\text{sub}} \cdot D)$，其中 $N_{\text{sub}} = \sum_{i=1}^{n} 2^{c_i}$ 是子电路的总数。当每个边界具有相同的切割次数 c时，边缘片段具有 $c_i = c$，内部片段具有 $c_i = 2c$，并且对于 $n = 2$, $N_{\text{sub}} = 2 \cdot 2^c$。内存：$O(N_{\text{sub}} \cdot D)$。

算法 2子电路模拟

输入： 来自算法1的子电路 $\{S_i^{(r)}\}$

输出： 状态向量 $\{|\psi_i^{(r)}\rangle\}$，每个向量的维度为 2^{q_i}

1: **对于** 每个片段 $i = 1$, ..., n **执行**
2: **对于** 每个分解索引 $r = 1$, ..., 2^{c_i} **执行**
3: $|\psi_i^{(r)}\rangle \leftarrow$ 状态向量模拟$(S_i^{(r)})$ ▷ $O(2^{q_i} D_{\text{seg},\ i})$
4: **结束循环**
5: **结束循环**

复杂度。 时间： $O\big(\sum_{i=1}^{n} 2^{c_i} \cdot 2^{q_i} \cdot D_{\text{seg},\ i}\big)$，其中 q_i 是量子比特的数量，而 $D_{\text{seg},\ i}$ 是片段 i 的深度（例如，对于含有 c 个切割的均匀二分划分， $O(2 \cdot 2^c \cdot 2^{q/2} \cdot D_{\text{seg}})$）。
内存：每个子电路占用 $O(2^{q_{\max}})$，如果保留所有态矢量用于合并，则最多占用 $q_{\max} = \max_i q_i$。

算法 3合并

输入： 来自算法 2 的状态向量 $\{|\psi_i^{(r)}\rangle\}$，合并索引表 $\mathcal{M}$

输出： 重构的完整状态向量 $|\Psi\rangle$，维度为 2^q

1: 根据 CZ 分解计算 $\{\alpha_m\}$ ▷ $O(2^{c_{\max}})$
2: $|\Psi\rangle \leftarrow \mathbf{0}_{2^q}$
3: **对于** 每种组合 $m = 1$ $, \ldots, 2^{c_{\max}}$ **执行**
4: 查找各片段索引（r_1, ... , r_n） $\leftarrow \mathcal{M}[m]$
5: $|\phi_m\rangle \leftarrow |\psi^{(r_1)}{}_1 \quad\rangle \otimes \cdots \otimes |\psi^{(r_n)n} \quad\rangle$ ▷ $O(2^q)$
6: $|\Psi\rangle \leftarrow |\Psi\rangle + \alpha_m |\phi_m\rangle$ ▷ $O(2^q)$
7: **结束循环**

复杂度。 时间： $O(2^{c_{\max}} \cdot 2^q)$，其中 $c_{\max} =$ 最大值$_i\, c_i$（$2^{c_{\max}}$ 个张量积，每个产生一个 2^q维向量；向量；例如， $O(2^c \cdot 2^q)$对应 $n = 2$）。内存： $O(2^q)$（一个累积累积向量 $|\Psi\rangle$ 和一个临时向量 $|\varphi_m\rangle$）。

[1] J. M. Shalf 和 R. Leland，《超越摩尔定律的计算》，Computer **48**，14（2015）。[2] J. Preskill，《NISQ时代及其后的量子计算》，Quantum **2**，79（2018），arXiv:1801.00862 [量子物理]。[3] S. Bravyi、G. Smith 和 J. A. Smolin，《权衡经典和量子计算资源》，《物理评论X》**6**，021043（2016）。[4] T. Peng、A. W. Harrow、M. Ozols 和 X. Wu，《在小型量子计算机上模拟大型量子电路》，Phys. Rev. Lett. **125**, 150504 2020 1904 00102 ( ), [ ] arXiv: . 量子物理 . [5] K. Mitarai 和 K. Fujii，《通过采样单量子比特操作构造虚拟双量子比特门》，New J. Phys. **23**，023021（2021），arXiv:1909.07534 [量子物理]。

[6]L. Brenner、C. Piveteau 和 D. Sutter，借助经典通信实现最优线切割（2023），arXiv:2302.03366 [量子物理]。[7] H. Harada、K. Wada 和 N. Yamamoto，无需辅助量子比特的双重最优并行线切割，PRX Quan- tum **5**，040308（2024），arXiv:2303.07340 [量子物理]。[8] I. L. Markov、A. Fatima、S. V. Isakov 和 S. Boixo，量子霸权比看起来既更近，也更远（2018），arXiv:1807.10749 [量子物理]。[9] F. Pan、K. Chen 和 P. Zhang，求解 Sycamore量子电路的采样问题，Phys. Rev. Lett.**129**，090502（2022），arXiv:2111.03011 [量子物理]。[10] E. Pednault、J. A. Gunnels、G. Nannicini、L. Horesh、T. Magerlein、E. Solomonik 和 R. Wisnieff，在量子电路模拟中突破49量子比特壁垒

（2017），arXiv:1710.05867 [量子物理]。[11] W. Tang、T. Tomesh、M. Suchara、J. Larson 和 M. Martonosi，CutQC：利用小型量子计算机进行大型量子电路评估，载于 *Proceedings of the 26th ACM International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS '21)* 2021 ()。

[12] T. Jones 、 A. Brown 、 I. Bush 和 S. C. Benjamin，QuEST 与量子计算机的高性能模拟，Scientific Reports **9**，10736（2019）。[13] B. Villalonga、S. Boixo、B. Nelson、C. Henze、E. Rieffel、R. Biswas 和 S. Mandra，一款用于验证和基准测试真实硬件上实现的量子电路的灵活高性能模拟器（2018），arXiv:1811.09599 [量子物理]。[14] Y. Suzuki、S. Uno、R. Raymond、T. Tanaka、T. Onodera 和 N. Yamamoto，Qulacs：一款用于研究的快速多功能量子电路模拟器，Quantum **5**，559（2021）。[15] T. Nguyen、D. Lyakh、E. Dumitrescu、D. Clark、J. Larkin 和 A. McCaskey，用于在百亿亿级计算机上模拟量子电路的张量网络量子虚拟机（2021），arXiv:2104.10523 [量子物理]。[16] F. Arute *et al.*，利用可编程超导处理器实现量子霸权，Nature **574**，505（2019）。[17] L. Schmitt、C. Piveteau 和 D. Sutter，使用多个双量子比特酉算子切割量子电路，Quantum **9**，1634（2025）。[18] A. Lowe、M. Medvidovic、A. Hayes、L. J. O'Riordan、T. R. Bromley、J. M. Arrazola 和 N. Killoran，利用随机测量快速进行量子电路切割，Quantum **7**，934（2023），arXiv:2207.14734 [量子物理]。[19] S. Balakrishnan 和 R. Sankaranarayanan，算子施密特分解与双量子比特的几何边缘——量子比特量子门，《Quantum Information Processing》**10**，449（2011）。[20] S. C. Marshall、C. Gyurik 和 V. Dunjko，《小型量子计算机上的高维量子机器学习》，《Quantum》 **7**，1078 2023 2203 13739 ()，[ ] arXiv: 。量子物理。[21]N. Huffman、D. Pavlichin 和 T. Weissman，《用于薛定谔式量子模拟的有损压缩》（2024），arXiv:2401.11088 [量子物理]。[22] X.-C. Wu、S. Di、F. Cappello、H. Finkel、Y. Alexeev 和 F. T. Chong，《用于量子电路模拟的振幅感知有损压缩》（ 2018 1811 05140 ），[ ] arXiv: 。量子物理。[23] I. L. Markov 和 Y. Shi，《通过收缩张量网络模拟量子计算》，《SIAM计算杂志》**38**，963（2008）。[24] A. Peruzzo、J. McClean、P. Shadbolt、M.-H. Yung、X.-Q. Zhou、P. J. Love、A. Aspuru-Guzik 和 J. L. O'Brien，《光子量子处理器上的变分本征值求解器》，《Nature Communications》**5**，4213（2014）。[25] A. Kandala、A. Mezzacapo、K. Temme、M. Takita、M. Brink、J. M. Chow 和 J. M. Gambetta，《适用于小分子和量子磁体的硬件高效变分量子本征求解器》，《Nature》**549**，242（2017）。[26]E. Farhi、J. Goldstone 和 S. Gutmann，《量子近似优化算法》（2014），arXiv:1411.4028 [量子物理]。[27] A. M. Childs、Y. Su、M. C. Tran、N. Wiebe 和 S. Zhu，《具有对易子标度的 Trotter 误差理论》，《Physical Review X》**11**，011020（2021）。