arXiv:2305.01064v1 [quant-ph] 1 May 2023

关于 Google 量子霸权主张的问题和关注问题

Gil Kalai、Yosef Rinott 和 Tomer Shoham

2023年5月3日

摘要

2019年10月，《自然》发表了一篇论文 [6] ，介绍了在 Google 开展的一项实验工作。该论文声称在一台 53量子比特量子计算机上实现了量子（计算）霸权。此后，我们一直在开展一项长期研究，探讨 Google 实验的各项统计方面问题。在 [30] 中，我们研究了 Google 的统计框架，认为该框架非常严谨，并提出了一些技术改进建议。本文基于统计分析，阐述了对 Google 2019 年实验的三个主要关注问题。第一个关注问题是，数据与谷歌噪声模型（或任何其他特定模型）不符。第二个关注问题是，一个用于先验估算保真度的关键简单公式似乎包含了出人意料的独立性假设，但它却给出了非常准确的预测。第三个关注问题涉及校准过程的统计性质。[1]

[1]本研究得到ERC资助项目834735的支持。感谢Carsten Voelkmann提出许多修改意见和有益建议。

1 引言

2019 年的论文《使用可编程超导处理器实现量子霸权》 [6] 声称，Google 的 Sycamore 量子计算机在约 200 秒内完成了某项计算；而据 Google 团队估计，最先进的经典超级计算机需要约 10,000 年才能完成同一项计算。Google 的 Sycamore 量子计算机执行了一项采样任务；也就是说，它从某个离散概率分布中生成长度为 53、带有显著噪声的随机比特串，该分布以所有此类 2^{53} 比特串为支撑集。（比特串是由 0 和 1 组成的向量。）Google 执行的这项特定采样任务称为随机电路采样（简称 RCS）。2021年，一支来自中国科学技术大学的团队使用其祖冲之号超导处理器 [33, 35], 重复了 Google 2019 年的实验，并声称取得了比 Google 实验更强的量子优势。

在 Google 实验中，Sycamore 量子计算机针对约 1000 个随机电路生成了样本，每个样本由数十万个二进制字符串组成。对于每个样本，研究人员计算了一个称为线性交叉熵保真度估计器 的特定统计量，记作 $\mathcal{F}_{XEB}$。根据这一估计器的数值，人们得出结论：量子计算机生成的样本代表了“量子霸权”。至于 Google 的量子霸权主张，此后已有多个研究团队 [28, 27, 34, 13, 14, 29, 11]提出了经典算法，其运行速度比 Google 论文中采用的算法快十个数量级。例如，Pan、Chen 和 Zhang [27]于 2021年实现了这一点。

将 Google 的量子霸权主张置于显微镜下审视

本文提出的统计分析或可揭示 Google 实验数据的质量与可靠性，以及该实验所采用的统计方法。第一作者 Kalai 自 2019 年 9 月（当时 [6] 和 [7] 遭到泄露）以来，便一直对 Google 实验提出一些关注问题。在 Google 论文发表几个月后，我们启动了一项长期项目，研究 Google 实验的各个统计方面，并仔细审视 Google 论文。值得一提的是，Google 的量子霸权主张似乎驳斥了 Kalai 关于量子计算的理论（[15, 16, 18]），以及 Kalai 的具体预测：NISQ 系统无法展示“量子霸权”。这

事实上，这影响了 Kalai 对 Google 量子霸权主张的评估，并可能使其产生偏差。（近期改进的经典算法在很大程度上驳斥了 Google 的量子霸权主张，因此 Google 的结果不再能驳斥 Kalai 的理论。）

下面简要回顾我们之前发表的论文。在第一篇论文 [30]中，我们主要研究了保真度估计量，并初步比较了样本的经验分布与谷歌噪声模型的分布。我们最近发表的论文 [20]详细描述了 Google 实验的相关数据和信息，并列出了一些确认、反驳、弱点和关注问题。

本文探讨三个核心关注问题。在第 2 节中，我们简要介绍 Google 实验的背景及其部分统计要素。在第 3 节中，我们阐述第一个关注问题：数据与谷歌噪声模型（或任何其他特定模型）不符。在第 4 节中，我们阐述第二个关注问题：某个关键先验估计（“公式（77）”）对 $\mathcal{F}_{XEB}$ 保真度估计量展现出惊人的预测能力。在第 5 节中，我们讨论第三个关注问题：双门校准的卓越效果及其其他统计性质。

几个具有普遍意义的问题

审视一项科学研究必然需要一丝不苟、锱铢必较，但我们认为还有几个问题具有普遍意义。

(a) 如何运用统计方法分析从含噪量子计算机获得的样本，并找到适当的模型来描述经验数据?

(b) 我们认为，某个预测模型中的统计独立性假设可能并不现实。研究这一问题可以采用怎样的科学框架和方法?

(c) 我们发现，针对一个多变量函数的局部优化过程（即分别优化每个变量的过程）竟然能够达到临界点，这令人惊讶。还可以用哪些工具来研究这一问题?

(d) 可以用哪些工具来研究某种经验行为是否具有非平稳性，甚至是否本质上不可预测?

(e) 审视重要科学著作时，恰当的方法与伦理规范是什么？理论学家（像我们这样的人）与实验学家之间的鸿沟又该如何弥合?

2 Google 的量子霸权主张

2.1 简要背景

现在，我们将简要概述 Google 实验、谷歌噪声模型、Google 的 $\mathcal{F}_{XEB}$ 线性交叉熵保真度估计量，以及 Google 在 [7] 中提出的公式（77），该公式用于根据电路各组件的保真度预测电路的保真度。更多详情请参阅 [30, 6, 7]。

Google 实验基于构建一台含有 n 个超导量子比特的量子计算机（电路），该计算机执行 m 轮计算。计算通过单量子比特门和双量子比特门完成。计算结束时，对量子比特进行测量，得到一个长度为 n 的 0 和 1 组成的字符串。最终实验针对 $n = 53$ 和 $m = 20$。该实验使用了 1113 个单量子比特门和 430 个双量子比特门。对于该实验，Google 团队生成了一个样本，其中包含三百万个长度为 53 的 0-1 向量。

Google 实验中使用的电路具有以下结构。量子比特排列在平面网格上，因此可以用两个坐标标识单个量子比特，例如量子比特 (3，3)。电路包含两种层：一种是作用于成对（相邻）量子比特的双量子比特门层。每个这样的双量子比特门层之后，都会有另一层随机选取的单量子比特门，作用于每个量子比特。（双量子比特门层的数量称为电路的深度。）单量子比特门层构成了实验中“可编程”的部分。双量子比特门层在整个实验中都按照某种固定模式排列。对于量子比特数在 12 到 53 之间且深度为 $m = 14$ 的电路，采用了 EFGH 模式；之后，对于 $n = 53$ 个量子比特以及深度为 $m = 12$、14、16、18、和 20 的电路，则采用了新的 ABCDCDAB 模式。每个字母（如 E）都对应一组固定的双量子比特门，这些门同时作用于量子比特；见图 1。（更多详情请参见 [20]。）作为实验的一部分，Google 团队还研究了电路的两个简化版本——“删减电路”和“补丁电路”，它们是通过移除“完整”电路中的部分双量子比特门构造的。具体而言，在补丁电路中，量子比特被划分为两个彼此不相互作用的独立补丁。

[2]Google 实验中，针对 900 多个电路的全部随机性，取决于为十个大型随机电路选择的随机单量子比特门。用于最大完整电路的第 k 次实验（$k = 0, 1, ..., 9$）的同一组随机单量子比特门（这些电路含有 $n = 53$ 个量子比特，深度为 $m = 20$），也用于所有类型的电路（即完整电路、省略电路和补丁电路）、各种模式、量子比特数 n 以及深度 m。

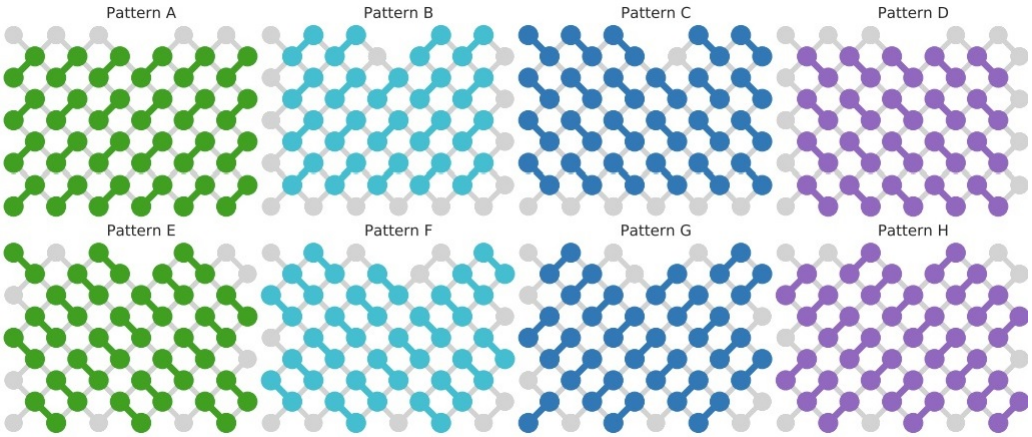

图 1：来自 [7] 的图 S25。 $n = 53$ 中单个双量子比特门层的不同模式。

每个包含 n 个量子比特的电路 C 都描述了一个针对长度为 n 的 0-1 向量的概率分布 $\mathbf{P}_C(x)$。（实际上，它描述的是一个 2^n 维的复振幅向量；对于每个 0-1 向量 x，都有一个对应的振幅 $z(x)$，并且 $\mathbf{P}_C(x) = |z(x)|^2$。）量子计算机能够依据概率分布 $\mathbf{P}_C(x)$ 进行采样，但会受到相当程度的噪声影响。当 n 和 m 不太大时，经典模拟可以计算出振幅本身（进而得出概率 $\mathbf{P}_C(x)$）。Google 的量子霸权主张基于这样一个事实：随着 n 和 m 增大，这些经典模拟很快就会变得不可行。

当 C 是随机电路时，概率分布 $\mathbf{P}_C(x)$ 的表现类似于波特–托马斯分布；也就是说，各个概率的表现仿佛是从指数分布中抽取的随机统计独立值（随后再进行归一化）。当然，波特–托马斯分布的一个实例取决于 2^n 个随机变量，而 $\mathbf{P}_C(x)$ 的 2^n 个值则取决于电路 C，其描述需要 n 个参数的多项式数量。

Google 基本噪声模型是

$$\mathbf{N}_C(x) = \phi \mathbf{P}_C(x) + (1 - \phi)2^{-n}, \tag{1}$$

其中 ϕ 是保真度，这是一个大致描述样本质量的参数。粗略来说，谷歌噪声模型假设，如果所有

如果量子计算机的组件运行时没有错误，那么将根据 $\mathbf{P}_C(x)$ 抽取一个比特串；如果发生错误，量子计算机则会生成一个均匀随机分布的比特串。

备注：对于含噪量子电路，如果 ρ 是理想态的密度矩阵（即无噪声量子计算机运行量子电路 C 时的状态），而 ψ 是含噪态的密度矩阵，那么保真度 ϕ 定义为

$$\phi = \langle \psi \, | \rho | \, \psi \rangle. \tag{2}$$

我们注意到，无法直接从量子计算机生成的样本分布中读取保真度 ϕ 本身。给定无限的样本量（或足够大的样本量，使经验分布能够很好地近似含噪分布）和无限的计算能力，寻找估计保真度的最佳方法是一个有趣的问题。另一个与含噪中等规模量子实验密切相关的重要问题是：在样本量仅为多项式级别（关于 $(1/\phi)$）的情况下，寻找估计保真度的最佳方法。

根据其噪声模型（以及分布 $\mathbf{P}_C$ 属于波特–托马斯分布这一事实），Google 论文描述了一种名为线性交叉熵估计量的统计量，记作 $\mathcal{F}_{XEB}$（有时也简记为 XEB）。量子计算机生成一串 $\tilde{}\mathbf{x}$ 共 N 个比特串 $\tilde{}\mathbf{x} = (\tilde{}\mathbf{x}^{(1)}, \tilde{}\mathbf{x}^{(2)}, \ldots, \tilde{}\mathbf{x}^{(N)})$ 后，保真度的估计量 $\mathcal{F}_{XEB}$ 按如下方式计算：

$$\mathcal{F}_{XEB}(\tilde{\mathbf{x}}) = \frac{1}{N} \sum_{i=1}^{N} 2^n \mathbf{P}_C(\tilde{\mathbf{x}}^{(i)}) - 1. \tag{3}$$

计算 $\mathcal{F}_{XEB}$ 需要了解采样比特串对应的$\mathbf{P}_C(x)$。

谷歌量子优越性声明还基于以下先验预测：根据各个组件的错误概率预测电路的保真度（ [7]中的公式（77））：

$$\hat{\phi} \;=\; \prod_{g \in \mathcal{G}_1} (1 - e_g) \prod_{g \in \mathcal{G}_2} (1 - e_g) \prod_{q \in \mathcal{Q}} (1 - e_q). \tag{4}$$

这里， $\mathcal{G}_1$ 是单量子比特门集合， $\mathcal{G}_2$ 是双量子比特门集合， $\mathcal{Q}$ 是量子比特集合。对于门 g，公式中的术语 e_g 指的是单个门 g发生错误的概率。对于量子比特 q， e_q 是测量量子比特 q时发生读出错误的概率。

Google 的优越性论文 [6] 对最终的 53量子比特样本提出了两个关键主张。

（A）其样本的保真度 ϕ 高于 $1/1000$。

(B) 在经典超级计算机上生成一个保真度相近的样本需要 10,000 年。

对于关于 ϕ 的值的主张 (A)，其论证依赖于一种外推论证，其中包含两个要素。一个要素是在经典可处理区域开展的数百次实验，也就是概率分布 $\mathbf{P}_C$ 可以由经典计算机计算，并且可以直接测试量子计算机性能的区域。另一个要素是用于预测保真度的理论公式（4）。根据论文，整个电路的保真度与公式（4）（ [7] 中的公式（77））给出的预测高度一致，偏差低于 10–20%。在经典可处理区域报告的实验超过 700 次，其中包括在简化量子电路上开展的实验（这类电路更容易用经典计算机模拟）。这些实验支持这样一种主张：公式（77）给出的保真度预测确实非常稳健，且适用于量子霸权区域中的 53 量子比特电路。

关于经典计算难度的主张（B），Google 团队主要依据他们所使用的一种特定算法的运行时间进行外推。他们还援引计算复杂度方面的支持，以论证当前任务在渐近意义上难以完成。

备注： 1. 主张（B）以及由此延伸的整个量子霸权主张，已在多个研究团队 [28, 27, 34, 13, 14, 29, 11] （包括其他团队）的研究中基本被推翻；这些研究展示了经典算法，其速度比 Google 论文中使用的算法快十个数量级。例如，Pan、Chen 和 Zhang [27]于 2021年实现了这一点。另请参阅第2.3节。

2. Google 对完整电路和删减电路使用的、由（3）给出的 $\mathcal{F}_{XEB}$ 保真度估算器并不适用于补丁电路。对于补丁电路，则分别对每个补丁使用 $\mathcal{F}_{XEB}$ 保真度估算器，再将两个估算结果相乘。

2.2 分布与大小偏倚分布

区分以下几者很有用：

(i) 大小为 2^n 的空间上的波特–托马斯分布 $\mathbf{P}_C(x)$，(ii) 所有可能的 2^n 比特字符串上的实际概率分布 $\mathbf{P}_C(x)$，以及 (iii) 大小偏倚分布，即当按照 Google 噪声模型 (1) 对 x 进行采样时，概率的实际分布（$\mathbf{P}_C(x)$）。

在第一项中，对于每个比特串 x，我们都有一个概率 $\mathbf{P}_C(x)$。在第二项中，我们只将这些 2^n 概率视为一个实数随机集合。所有 2^n 比特串对应的 2^n 概率 $\mathbf{P}_C(x)$ 的表现，就像从指数随机变量中抽取的一个大小为 2^n 的归一化样本。

基于 Google 噪声模型 (1) 的大小偏倚分布（条目 (iii)）表现得像具有密度函数的真实分布。

$$\phi x e^{-x} + (1-\phi)e^{-x}. \tag{5}$$

Google 比特串的实证大小偏倚分布与 (5) 非常吻合（见下文图3、[7] 中的图 S32 和 [30]中的图9）。正如我们在 [30] 中解释的，大小偏倚分布对于“完整” 2^n 分布的重大变化具有稳定性。例如，如果你只对 1 的个数为素数的比特串进行采样，这不会改变大小偏倚分布，但会对原始的 Porter–Thomas 分布造成重大改变，使大多数概率变为零。

2.3 Gao 等人的工作

Gao、Kalinowski、Chou、Lukin、Barak 和 Choi [11] 研究了随机电路采样，并考察了含噪量子电路的三个量：第一个量是由公式 2 定义的保真度 ϕ，第二个量是线性交叉熵估计量 $\mathcal{F}_{XEB}$，第三个量是记作 $p_{no\ err}$ 的无错误概率。Gao 等人的一个基本观察是，对门施加退极化噪声后，得到的分布与理想分布呈正相关（因此会得到正的 $\mathcal{F}_{XEB}$ 值）。对一组双量子比特门施加这种退极化噪声，使电路分成两部分，就能得到一种“补丁”电路；对于这种电路，可以分别在每个补丁上进行计算，从而快速构造经典算法来生成具有

$\mathcal{F}_{XEB}$高值的样本。因此，该论文表明，存在一些“对抗性”方法，即使无法计算各个振幅，也能生成具有较高 $\mathcal{F}_{XEB}$ 值的样本。

论文 [11] 描述了门错误效应为何会导致与理想分布呈正相关，以及为何一般情况下会出现严格不等式的其他原因：

$$\mathcal{F}_{XEB} > \phi > p_{no\ err}. \tag{6}$$

备注： 1. 在谈到他们用于快速生成具有较高 $\mathcal{F}_{XEB}$值的样本的算法时，Gao 等人写道：“值得注意的是，我们算法的 XEB 值通常会随着量子电路规模增大而提高，而噪声量子设备的 XEB 值则会迅速下降。当量子比特数增加、而电路深度和每个门的错误率保持不变时，这种缩放趋势依然成立。”换言之，从渐近角度看，当 n 很大时，[11] 的对抗性经典方法优于量子计算机。

2. 关于与当前实验的比较，Gao 等人写道：“具体而言，我们提出了一种高效的经典算法，仅用几秒钟便可在一台 GPU 机器上实现较高的 XEB 值，即达到最先进实验所得结果的 2–12%。”对于 Google 的终极量子优越性实验（$n = 53, m = 20$），Gao 等人（[11], 表 I）在一台 GPU 机器上用 0.6 秒生成了样本，其 $\mathcal{F}_{XEB}$保真度为 0.000185，相当于 Google 所宣称的 0.00224 保真度的 8%。

3. 允许其算法运行更长时间，可能会在现有经典计算机上超越 Google 的 $\mathcal{F}_{XEB}$ 值。我们还注意到，Gao 等人的方法采用了“补丁电路”，而且也可能基于“删减电路”设计算法，从而为对抗性方法实现高效算法，并超越 Google 所声称的保真度。

2.4 本文的一些发现

1. Google 实验的样本与 Google 噪声模型以及任何其他特定噪声模型之间存在很大差距。（第 3.2 节。）经验分布与模型之间的差距是不对称的。（第 3.5.2 节。）

2. 经验行为存在尚未得到解释的大幅波动。因此，有证据表明，Google 噪声模型与均匀分布之间的距离更小（当

量子比特数量为 $n > 16$ 时），而实验样本与 Google 噪声模型之间的距离更大。（第3.3节、第3.4.1节。）

3. 样本的经验行为并非平稳。（第3.5.1节。）4. 尽管经验分布并非平稳，$\mathcal{F}_{XEB}$ 保真度沿样本保持稳定。此外，在 Sycamore 量子计算机后续实验中报告的、会导致错误骤增的“高能事件”，在 2019 年量子霸权实验中无法检测到。（第3.5.1节。）5. 公式（77）对交叉熵基准测试保真度估计的预测能力在统计上令人惊讶：其中隐含的独立性假设认为，量子计算机等系统的各组成部分彼此独立；然而，众所周知，这些系统会受到噪声以及与其环境相互作用所导致错误的影响，这一假设尤其令人瞩目。（第 4.1 节。）6. 公式（77）对补丁电路的预测与实验数据之间存在系统性偏差；Google 对这一偏差的解释并不令人信服。（第 4.7、5.4 节。）7. 补丁电路中两个补丁的保真度表现迥异，这似乎与公式（77）相矛盾。（第 4.6 节。）8. 实验的成功完全取决于校准调整所产生的巨大影响。（第 5.2 节。）不同双量子比特门的校准调整效果差异很大，即使是同一个双量子比特门在不同位置出现时，效果也不尽相同。9. 校准调整的效果出奇地好，尤其考虑到所述校准具有局部性。从数学上说，我们看到一个局部优化过程到达了一个依赖数百个参数的函数的临界点。（第 5.3、5.5 节。）

2.5 对未来量子霸权实验的简要建议

对于未来的实验，我们建议按以下步骤进行，并明确区分各个步骤。

步骤 A：根据针对 1 量子比特和 2 量子比特电路的实验校准电路。

步骤 B：创建完全随机化的随机电路集合，为每个电路选择新的随机单量子比特门。（此步骤可以按照 [30] 和 [20] 的第 5.1 节所建议的方式盲目执行。）

步骤 C：运行量子计算机，生成比特串并计算保真度。

在 2019 年量子优越性实验中，十个电路使用了相同的随机单量子比特门，这些门在校准阶段之前就已选定。一种新的校准变体（“并行交叉熵基准测试”）与该实验同期开发，并在生成样本几天后完成（见 [20] 的第 2.4 节）。

我们还提出：a) 为少数几个电路生成大量样本（1000 万至 5000 万个比特串），以便研究经验分布。b) 详细说明公式（77）中所使用的所有组件的保真度。c) 提供校准实验的原始数据。这将使我们能够对校准过程进行部分核查，而不会泄露校准过程本身。（参见 [20]的第 5.2 节。）

3 关切一：数据与 Google 模型（或任何其他特定噪声模型）不符

3.1 采样计算问题

采样问题的常见定义如下，该定义引自 Aaronson 和 Chen [2] （另见 Lund 等人 [24]）：

> “目标是从所需概率分布中，精确或近似地采样一个 n 位字符串的问题。”

Bouland 等人 [9] 对基于采样计算任务来证明量子霸权的提案框架作了如下描述：

> “量子霸权提案有一个共同框架。计算任务是从某个实验上可行的量子过程或算法（针对某个给定输入）的输出分布 D 中进行采样。要确立量子霸权，我们必须证明其困难性（也就是说，不存在能够从任何接近 D 的分布中进行采样的高效经典算法），并进行验证（也就是说，某个算法能够检验实验设备是否从接近 D 的输出分布中进行了采样）。”

在本节中，我们将表明，Google 量子霸权实验产生的随机比特串并不符合这些定义。[3] 量子计算机产生样本的经验分布与 Google 噪声模型以及我们所知的任何其他特定模型都大不相同。我们还将研究 Google 实验提供的样本所具有的其他有趣统计性质。

3.2 $n = 12,\ 14$ 的经验分布

对于实验电路 C，给定比特串 x，我们用 $O(x)$ 表示样本中 x 出现的次数。经验分布 是一个

[3] Aaronson 和 Gunn [3]以及 Gao 等人 [11] 研究了量子霸权在多大程度上能够基于高$\mathcal{F}_{XEB}$的采样计算困难性，而不涉及是否实现了某个特定采样任务的问题。

定义在比特串集合 x 上的离散概率分布，由 $O(x)/N$ 给出，其中 N 是样本大小。回顾一下， n 表示量子比特的数量， $M = 2^n$ 表示比特串的数量。在 [30] 中，我们使用皮尔逊卡方检验统计量，将出现次数与 Google 噪声模型进行比较。皮尔逊卡方检验统计量为

$$\mathcal{X}^2 = \sum_{i=1}^{M} \frac{(O(x_i) - N\mathbf{N}_C(x_i))^2}{N\mathbf{N}_C(x_i)}, \tag{7}$$

及其在固定 M 且较大 N 时的渐近分布为具有 $M-1$ 个自由度的卡方分布（或者，如果估计了 ϕ ，则为 $M-2$ ）。($\mathbf{N}_C$ 由 (1) 定义。) 对于 $n = 12$，我们得到的 χ^2-值约为 40,000，而根据 $\mathbf{N}_C(x)$，样本的期望值约为 4,000；参见表 3。结论是，Google 模型 (1) 未能充分描述数据，因此应探索其他模型。经验分布与 Google 噪声模型之间的巨大差异从图 2 中也显而易见。

我们还考虑了其他几种距离度量，用于比较出现次数与 Google 的噪声模型以及其他噪声模型。

$$L_1 = \sum_{i=1}^{M} |N\mathbf{N}_C(x_i) - O(x_i)|, \qquad L_2 = \sqrt{\sum_{i=1}^{M} (N\mathbf{N}_C(x_i) - O(x_i))^2},$$

$$KL = \sum_{i=1}^{M} \frac{O(x_i)}{N} \log\left(\frac{O(x_i)/N}{\mathbf{N}_C(x_i)}\right),$$

以及皮尔逊相关系数。（KL 指库尔贝克-莱布勒散度。）结果见表 3。

在 [30], 第 6 节中，我们考虑了将读出错误纳入考量的改进误差模型。当我们采用最佳的改进误差模型（非对称读出噪声模型）时，噪声模型与经验分布之间的卡方距离略有改善：对于 $n = 12, 14$，该数值分别从 41,190 和 54,286 降至 36,786 和 50,093（表 3）。对于 $n = 12, 14$，均匀分布样本与 Google 噪声模型之间的卡方距离，仅略大于经验样本与 Google 噪声模型之间的距离。我们预计，对于更大的 n ，量子计算机生成的样本将会是

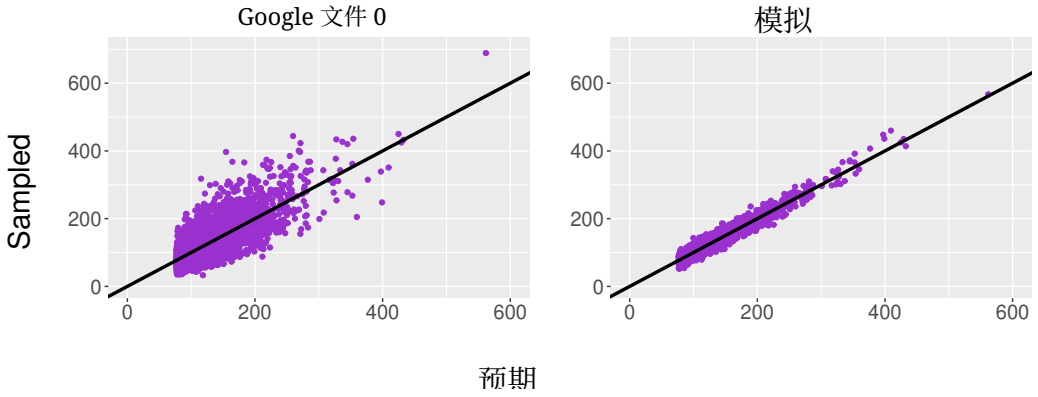

图 2：左侧散点图展示了 Google 样本（文件 0）中理论频率与经验频率的对比，其中包含 $n = 12$。右侧散点图展示了根据谷歌噪声模型 (1) 得到的理论频率与我们模拟的经验频率的对比，其中包含 $\phi = 0.3701$。

它与 Google 模型（以及我们所知的任何特定模型）的差异，比从均匀分布中抽取的样本更大（见第3.3节中的说明）。若能利用详细的门错误找到与经验分布更吻合的模型，将会很有意思；这可以通过对含噪电路进行模拟来实现。

3.3 估计量 T 与经验方差

在 [30] 中，我们基于出现次数考虑了另一种用于估算保真度的估计量 T（此估计不需要知道 $\mathbf{P}_C(x_i)$。）由于 exp(1)变量的期望方差为 1，对于保真度水平 ϕ，根据 Google 噪声模型得到的分布的期望方差为 $\phi^2 \cdot 2^{-2n}$。对于大小为 N 的样本，样本的期望方差会大于分布的方差，但可以根据样本的方差估算分布的方差。考虑到抽样大小，用于估算 ϕ^2 的无偏（归一化）估计量 T^2 定义如下：

$$T^2 := \frac{M(M+1)}{(N^2-N)(M-1)} \left(\sum_{i=1}^{M} O(x_i)^2 - N - (N^2-N)/M \right). \qquad (8)$$

在 Google 噪声模型下， T^2 是 ϕ^2的无偏估计量。预计读出和门错误的影响会使 T 增加几个百分点。对于第 6 节中考虑的详细读出模型，情况就是如此[30] （表 1）；在对门错误作出一些合理统计假设的情况下，门错误的贡献也同样如此。我们可以使用（8）根据经验数据估计T^2 ，并且在 $n \leq 22$时，这会给出相当好的估计；这已在 [30] 第4.7节中完成，结果表明，与 Google 噪声模型相比，Google 比特串的 T 值要高得多。例如，从表 1 中可以看到，对于 $n = 12,18, 22, T$与 Google 模型值相比大约高出 30%、50% 和 80%。 （这表明，量子计算机产生的实际含噪信号的方差远大于 Google 噪声模型给出的方差；例如，对于 $n = 22$， T^2 比 $\mathcal{F}_{XEB}$ 估计量大三倍以上。）这种额外方差代表了一种重要的噪声形式（下文将其记为$\mathbf{N}_T$），但目前对其了解甚少。

备注： 如果 T^2 确实是方差的良好估计量，那么对于不太小的 n 值，均匀分布相比经验分布更“接近”由 Google 噪声模型描述的分布。当 n 大于 14 时，我们只能在以下（合理）假设下估计 ℓ_2-距离：导致大方差的附加噪声 $\mathbf{N}_T$与 $\mathbf{P}_C$不相关，并且进一步假设 T^2 （归一化后）能够估计方差。均匀分布与 Google 噪声模型给出的分布之间的（归一化） ℓ_2 距离的期望值为 ϕ （对于随机电路，该值集中在 ϕ 附近）。在我们的假设下，经验分布与 Google 模型所描述的分布之间的 ℓ_2 距离为 $\sqrt{T^2 - \phi^2}$；如果 $T > \sqrt{2}\phi$，该值大于 ϕ。

这表明，对于 $n > 14$ ，来自均匀分布的样本将比 Sycamore 量子计算机生成的实验样本更接近（并且随着 n 增大而接近得多）Google 噪声模型。（这一结论不仅适用于 Google 噪声模型，也适用于我们所知的任何特定噪声模型。）[4]

[4]Greg Valiant 和 Paul Valiant [32] 的研究探讨了如何使用规模次线性的样本来估计各种统计参数，例如熵、方差，以及经验分布与模型之间的距离，这项研究或许与此相关。

n	(77)	Google 文件		模拟-Google 模型模拟-读出模型			
		平均交叉熵基准测试	平均 T	平均交叉熵基准测试	平均 T	平均交叉熵基准测试	平均 T
12	0.3862	0.3701	0.4689	0.3727	0.3712	0.3721	0.3742
14	0.3320	0.3298	0.4392	0.3316	0.3306	0.3315	0.3342
16	0.2828	0.2721	0.3917	0.2720	0.2722	0.2719	0.2755
18	0.2207	0.2442	0.3557	0.2442	0.2442	0.2442	0.2479
20	0.1875	0.2184	0.3210	0.2181	0.2179	0.2182	0.2216
22	0.1554	0.1650	0.2989	0.1651	0.1642	0.1649	0.1685
24	0.1256	0.1407	0.2838	0.1407	0.1312	0.1408	0.1264

表 1：公式（77）的预测值、平均 $\mathcal{F}_{XEB}$ 和平均 T。每次模拟均基于 100 次重复，每个 Google 文件进行 10 次。模拟中 Google 模型的 ϕ 取为表中报告的平均 $\mathcal{F}_{XEB}$，而对称读出模型的 $\phi_r o$ 为 $\phi/(1-0.038)^n - \phi$，其中 ϕ 为平均 $\mathcal{F}_{XEB}$。

3.4 经验大小偏倚分布

对于（无噪声的）波特–托马斯分布，其大小偏倚分布是一个密度函数为 xe^{-x} 的实值分布。在谷歌噪声模型下，当保真度水平为 ϕ 时，大小偏倚分布为 $\phi x e^{-x} + (1-\phi)e^{-x}$。（见第 2.2 节。）给定样本 $\tilde{\mathbf{x}}^{(i)}$，我们将这些实数称为

$$\{\mathbf{P}_C(\tilde{\mathbf{x}}^{(i)})\},$$

作为经验大小偏倚分布。尽管 Google 样本的经验分布与 Google 模型大不相同，但经验大小偏倚分布的整体表现与模型非常接近。（这一点在下文的图3和 [7]中的图 S32 中有所展示。）当n 较大时，大多数比特串在样本中只出现一次，因此我们无法区分来自 Google 文件的经验大小偏倚分布与基于 Google 模型进行模拟得到的分布。当 n较小时，我们可以通过考察不同尺度下的不同行为来呈现 Google 数据的表现，而直方图（图3）中的良好吻合也体现了平滑作用。我们还注意到，当保真度较低时，经验大小偏倚分布主要由均匀分布中的样本决定，因此经验大小偏倚分布与理论的一致，并未传递多少信息，只能说明噪声与“信号”不相关。

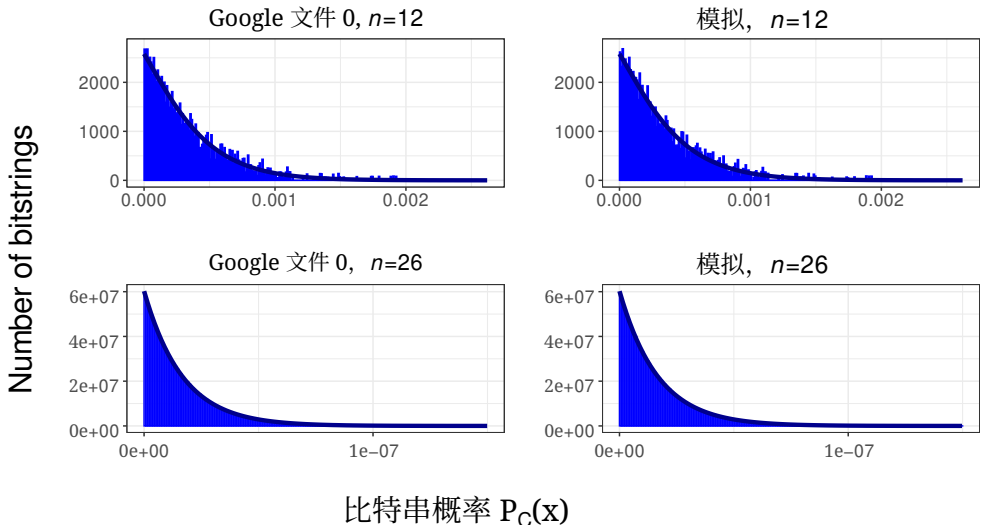

图3： $\{P_C(\tilde{\mathbf{x}}_i)\}_{i=1}^N$ 的直方图，其中包含来自 Google 文件编号0（$n = 12$）（左上）和 $n = 26$（左下）的 $N = 500{,}000$ 个 $\{\tilde{\mathbf{x}}_i\}_{i=1}^N$ 。右侧是根据谷歌噪声模型生成的样本，其中包含 $\phi = 0.3701$（右上）和 $\phi = 0.1024$（右下）。单元格数量为200。深蓝色曲线表示大小偏倚分布的渐近密度。

3.4.1 经验分布在不同尺度上的表现

我们按照 $P_C(x_i)$ 的值对所有比特串排序，并根据 $P_C(x_i)$ 的值将比特串分成 s 组，然后考察每组对应的比特串数量。例如，对于 $s = 100$ ，我们考察样本中 $P_C(x_i)$ 位于前 1%、从前往后介于 10% 和 11% 之间，或从前往后介于 49% 和 50% 之间的比特串，这些数量与 Google 模型完全一致。也就是说，除抽样误差外，这些数量与通过模拟从 Google 模型抽取样本时的表现相同。然而，当我们考察这些比特串的经验方差时，发现方差非常大（远超抽样效应）。这一点在图4中有所展示。在图5中，我们放大查看了 12 个量子比特和 Google 的第一个比特串文件（文件 0）中， x 的值介于中位数和 0.55 分位数之间的比特串。（这样的比特串约有 200 个。）我们比较了每个比特串出现的频率

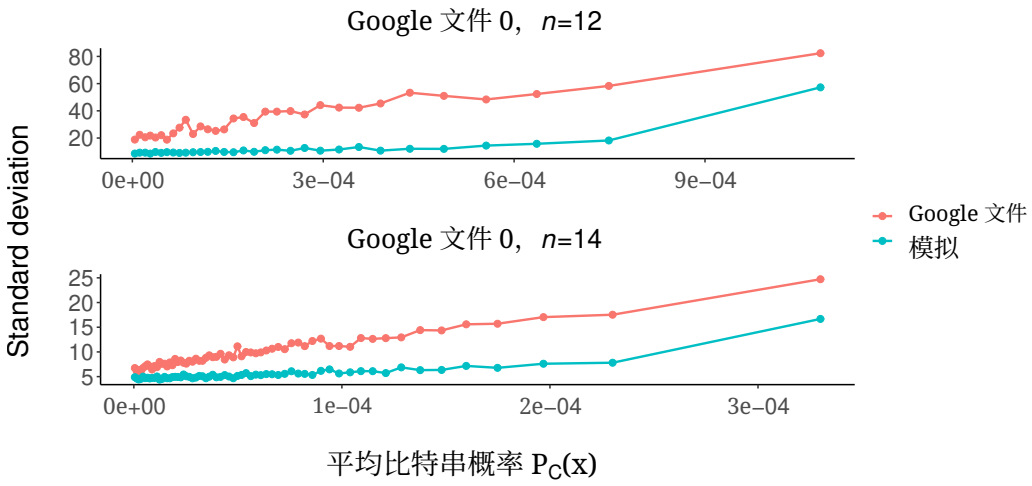

图4：我们根据$\mathbf{P}_C(x)$的值，将比特串划分为每组128个（$n = 12$）和256个（$n = 14$），并计算每组中经验分布的标准差。蓝线表示基于谷歌噪声模型的模拟结果，其中$\phi = 0.3701$ 对应 $n = 12$，$\phi = 0.3298$ 对应 $n = 14$。红线表示 Google 数据的表现。

Google 比特串文件中的比特串与通过模拟得到的文件中的比特串。

3.5 对经验分布的详细研究

3.5.1 分布的非平稳性

对于 $n = 12$，以及量子计算机产生的每个样本量为500,000的比特串文件，我们将样本随机分成两半，并计算两半中比特串出现次数的 ℓ_1 距离。我们将这些 ℓ_1 距离与按采样顺序排列的比特串前后两半之间的 ℓ_1 距离进行了比较。在图6中可以看到，按采样顺序划分得到的两个经验分布之间的距离，显著大于随机划分的结果。这一发现表明样本具有非平稳行为。这种行为可能与一种更强的性质有关，即经验分布中的一个较大成分具有噪声敏感性。（噪声敏感性以及后续

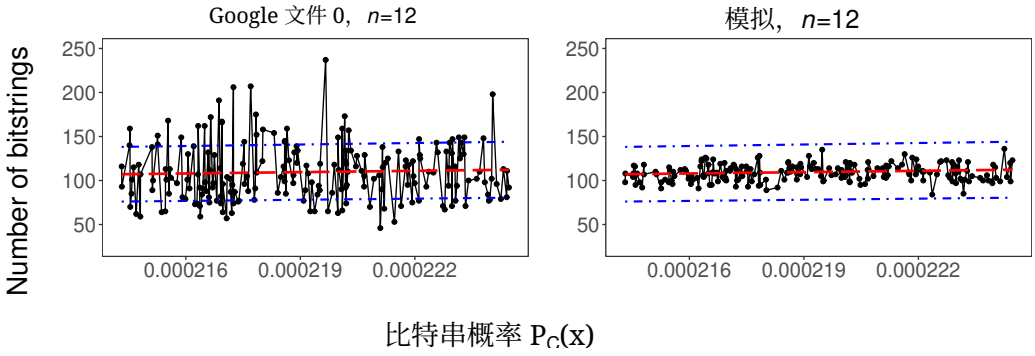

图5：放大观察振幅介于中位数与 0.55 分位数之间的比特串的经验频率。左图显示 Google 文件 0 中比特串的经验出现次数，n=12。右图基于一次模拟，使用 $\phi = 0.3862$。红线表示 Google 噪声模型中比特串数量的期望值，蓝色虚线表示与期望值相差 3 个标准差的位置。

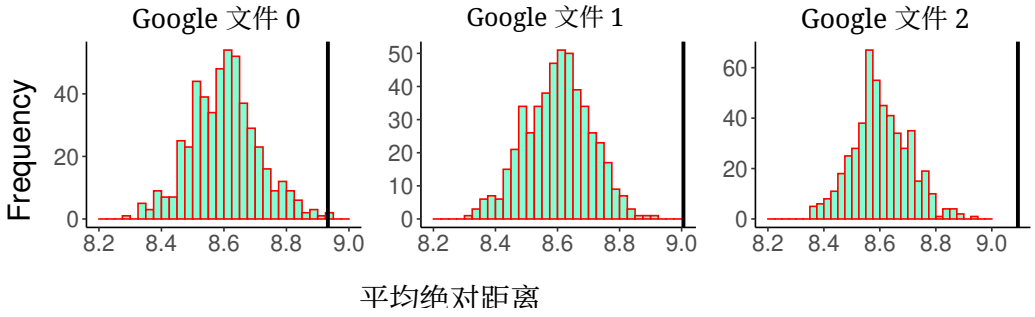

图6：比较 Google 样本的两半：黑色竖线表示按照采样顺序将样本分成两半后，比特串出现次数之间的 ℓ_1 距离。直方图显示了将比特串随机划分为两半时，比特串出现次数之间的 ℓ_1 距离。

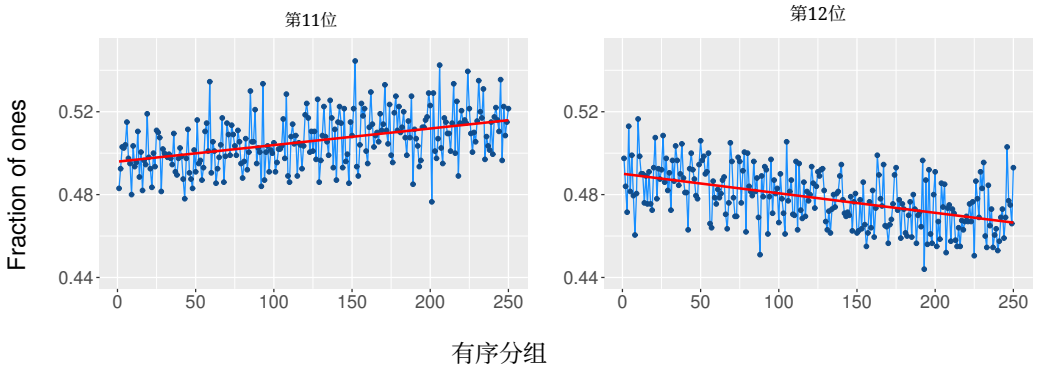

图7：1的比例发生漂移。我们按照采样顺序，将 500,000 个比特串分成 250 组，每组包含 2,000 个比特串。对于每组，我们计算比特串中某个位置上的比特为“1”的比例。该图显示了一个电路（文件 0）中位置 11 和 12 处的 1 的比例，其中 $n = 12$。（红线是对数据点进行线性回归拟合得到的结果。）不同电路中的趋势一致，但比特串中不同位置的趋势各不相同。

数据的不可预测性已在 Kalai 的研究成果 [16, 15], 中得到预测，其研究可追溯至 Kalai 和 Kindler [19]。）我们观察到，某些比特中“1”的百分比会随采样顺序而变化，呈现出特定的非平稳行为（图7）。

人们原本会预期，这些发现应当会导致保真度下降，但我们并未发现两半数据之间的 $\mathcal{F}_{XEB}$-保真度（或其他保真度指标）存在显著差异。我们还注意到，近期使用 Sycamore 量子计算机 [26, 4] 开展的实验中，数据曾出现错误率突然升至远高于正常水平、随后又逐渐恢复正常的阶段。（这可能是由宇宙射线引起的。）相比之下，我们并未观察到数据中出现 $\mathcal{F}_{XEB}$-保真度显著降低的阶段（对于 $n = 12$）。

3.5.2 表面上的不对称性

当我们将经验分布与 Google 噪声模型进行比较时，会观察到一些不对称性： $O_C(x) - \mathbf{N}_C(x) \cdot N$ 的正值绝对值往往大于负值。这一点可以通过

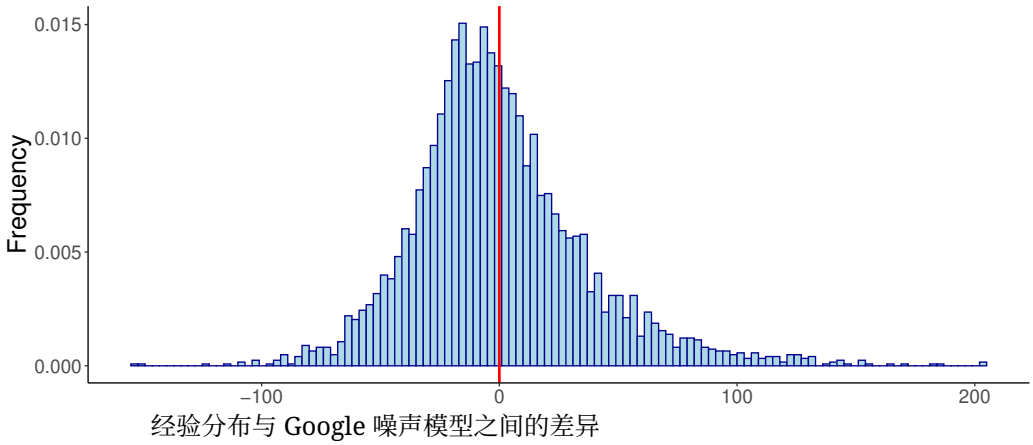

图 8：经验分布与 Google 噪声模型给出的值之间差异的直方图，$n = 12$, Google 文件 0，$\phi = 0.3701$。（其他 Google 文件的表现也类似。）

观察图 2 左侧对角线上方和下方的点来发现。图 8 显示，$O_C(x) - \mathbf{N}_C(x) \cdot N$的值的分布是不对称的。经验分布与模型之间的差距为何会出现表观不对称性？是否确实有必要对此作出解释？

对于我们的参数，基于 Google 噪声模型的样本相对于分布的抽样误差所导致的偏差是对称的（图 2 右侧）。我们可以预期，更准确的噪声模型将导致正确的波特–托马斯分布与大量不相关的波特–托马斯分布混合。这个图景表明，个体差异 $O_C(x) - \mathbf{N}_C(x) \cdot N$的表现类似于高斯分布，因此是对称的。可以通过模拟进一步检验这一点。

3.6 Google 团队对数据与模型之间差异的回应

Google 团队一直将其噪声模型视为对极其复杂实验的一种高度简化的模型，并指出“它往往奏效，但并非总是如此。”他们强调（参见 [7]第四节），$\mathcal{F}_{XEB}$ 并不依赖 Google 特有的噪声模型，因此适用范围要广泛得多。

（事实上，我们的论文 [30]也证实了这一点。）他们提出， [30] 第 8 节中针对 12 个量子比特电路所做的皮尔逊卡方检验未能通过，可能是读出偏差所致，并补充道：“如果按汉明重量对概率进行分组，就能一眼看出这种偏差。但这或许还不够。也可以尝试进行含噪声模拟，对于 12 个量子比特而言，这并不难。无论如何，要用简单的噪声模型来匹配各个实验概率并不容易。”他们提议逐步研究更贴近实际的噪声模拟，以理解这种表观不对称性。（2021年，Google推出了一款含噪电路模拟器，或许能够实现这一提议；IBM、NASA 等机构也开发了含噪模拟器。）

关于过大的经验方差，Google 团队指出，他们曾针对一个类似参数进行过类似的统计研究（称为散斑纯度基准测试 （SPB），见 [7]第 VI.C.4 节），研究对象是较小的 1 量子比特和 2 量子比特电路，但并未将这项研究扩展到大型电路。Google 团队还指出，我们发现样本的经验行为并非平稳，这与他们在 [7]第 VIII、H 和 I 节中的研究有关。

3.7 噪声建模：读出错误与傅里叶分析

样本的一般模型可描述如下：

$$\mathbf{N}_C^* = \phi \mathbf{P}_C + \phi_{ro}\mathbf{N}_{ro} + (1-\phi_g)\mathbf{N}_g + \mathbf{N}_T. \tag{9}$$

术语 $\phi \mathbf{P}_C$ 表示由电路 C 乘以保真度所描述的理想分布，这就是我们希望在数据中检测到的“主信号”。术语$\mathbf{N}_{ro}$ 表示在不存在门错误的条件下读出错误的影响。 ϕ_g 表示不存在任何门错误的概率，而 $\phi_{ro} =: \phi_g - \phi$。$\mathbf{N}_g$ 表示存在门错误时，门错误和读出错误共同产生的影响。术语 $\mathbf{N}_T(x)$对应于另一种影响显著但尚未得到充分理解的噪声形式。（它可以被视为前三项的某种较大波动（期望值为零）。）假设 $\mathbf{N}_T$ 与 $\mathbf{P}_C$ 及其他各项不相关是合理的。在 Google 噪声模型（1）下关于参数 ϕ 的统计估计量，在形式为（9）的更详细模型下仍然适用（ [7]中已经对此作了强调）。在 [30] （第 6 节）中，我们提出了两个模型

用于读出错误，并基于门错误的影响与读出错误的影响在统计上相互独立这一假设，提出了用于估计读出错误总体影响的相应估计量。在随后正在进行的工作中 [21] ，我们开发了能够快速计算这些估计量的傅里叶工具，同时还根据傅里叶层级改进了保真度估计量。[5]

3.8 其他问题

3.8.1 各种噪声模型类别

为了理解 Google 数据中的某些现象，我们还考虑了几大类噪声模型。以下是一个例子（来自 [30]）：设$\mathbf{Q}$ 为定义在长度为 n 的比特串上的概率分布，并考虑对每个比特串 x， 2^n~正实数 γ_x 。接下来考虑 $\mathbf{Q}(x) = \gamma_x \mathbf{Q}(x)$~，并对其进行归一化，使得 $\mathbf{Q}$ 成为一个概率分布。一个特殊情况是，我们从定义在 $[0,1]$ 上的概率分布 D 开始，并针对每个比特串 x，按照 D 随机选择 γ_x ，且这些选择在统计上相互独立。当我们令 $\mathbf{Q}(x) = \mathbf{P}_C(x)$ 并应用这一过程时，会得到一个截然不同但具有相同大小偏倚分布的分布，这展示了不对称性（第 3.5.2 节）以及图4中描述的行为。（我们没有理由认为这种类型的模型适用于 NISQ 系统。）

3.8.2 补丁电路的数据

补丁电路是由两个彼此不相互作用的部分组成的简化电路。（Google 团队于2022年6月公布了补丁电路的比特串和电路描述。）补丁电路的数据为比较经验分布与理想概率以及各种噪声模型提供了新的机会，而且可用于6量子比特电路。对这些数据的初步研究显示，它们具有与本节所述相似的性质。

[5]关于随机电路采样中傅里叶–沃尔什展开的早期研究，参见 Boixo 等人。 [8], Gao 和 Duan [10], 以及 Kalai [15]。另请参阅 Kalai 和 Kindler[19] 关于玻色子采样的一项早期相关研究。Aharonov 等人 [5] 是一篇近期的相关论文。

3.8.3 估算保真度的其他方法

实验数据与 Google 噪声模型不符，这一事实再次引出了估算保真度这一有趣的问题。一种估算保真度的方法，是将经验分布与理想分布之间的距离 D 视为基本参数，并以 Google 噪声模型为依据，将 D 转换为保真度估算器。（这仅在 $n = 12$，14 时才可行。）我们尝试了分布之间的各种距离定义，（不出所料）得到的保真度估算值低于使用线性交叉熵得出的估算值。

3.8.4 预测噪声

我们提出如下问题：给定 $\mathbf{P}_C$ 的值，假设我们想要预测噪声的影响。例如，我们想根据比特串的前半部分预测后半部分的行为，或者根据一个电路的比特串，预测另一个电路的比特串的行为。当然，对于按照 Google 噪声模型抽取的样本，这种预测是不可能的，因为两部分之间的差异代表了抽样误差。我们未能找到根据一个电路预测另一个电路行为的方法，但能够根据比特串的前半部分（部分地）预测其后半部分。

3.9 第3节总结

Google 量子优越性实验的样本与 Google 噪声模型之间存在很大差距。事实上，这些样本与我们所知的任何噪声模型都相去甚远。有证据表明，Google 噪声模型与均匀分布之间的距离（当量子比特数为 $n > 16$ 时）小于实验样本与 Google 噪声模型之间的距离。我们研究了经验分布的其他性质，例如它在不同尺度下的表现、其非平稳特性以及傅里叶表现；对于主要来自其他含噪中等规模量子实验的数据，以及来自含噪电路模拟器的数据，还有许多工作有待完成。

4 问题二：公式（77）

4.1 公式（77）的预测能力

如前所述，Google 论文 [7]（第 2.1 节中的方程（4））中的公式（77）基于电路各组件的保真度，对电路的保真度进行了估算：

$$\hat{\phi} \;=\; \prod_{g\in\mathcal{G}_1}(1-e_g)\prod_{g\in\mathcal{G}_2}(1-e_g)\prod_{e\in\mathcal{Q}}(1-e_q).$$

Google 论文声称，该公式以 10%–20% 的精度估算电路的保真度 ϕ，并将其认定为电路无错误概率（$p_{\mathbf{no\ err}}$）。这种惊人的一致性被 [6] 和 [7] 视为系统规模扩大时不会出现额外退相干物理效应的迹象。它本身也被视为一项重大的科学新发现。我们担忧的是，这个公式具有如此强的预测能力，令人惊讶；它意味着错误之间不存在相互作用，表明错误完全独立，也意味着不存在影响多个量子比特的共同环境效应。本节将讨论这一问题。这个问题曾在 [18] 中提出（也曾在 [17, 12] 及更早的科学博客讨论中提出），并在 [30, 20] 中有所提及。

备注： 1）迄今为止，公式中 e_q 和 e_g 的具体数值尚未公开。 (77)

2）基于平均保真度的公式（4）的近似为

$$\hat{\phi}^{*} = (1-0.0016)^{|\mathcal{G}_1|}(1-0.0062)^{|\mathcal{G}_2|}(1-0.038)^{n}. \tag{10}$$

3) Google 团队提出了以下近似方法：不考虑单量子比特门，而是将其贡献计入双量子比特门的综合保真度（称为双量子比特门周期）：

$$\hat{\phi}^{**} = (1-0.0093)^{|\mathcal{G}_2|}(1-0.038)^{n}. \tag{11}$$

公式（11）可以看作是在单量子比特门数量约为双量子比特门数量两倍这一假设下应用公式（10），而实际上，单量子比特门的数量要多得多。表 4 比较了公式（77）（据 Google 报告）的预测、上述简化预测，以及实验电路的平均保真度。正如预期，公式（77）比公式

（10）和（11）。与公式（10）相比，公式（11）的预测效果显著更好。

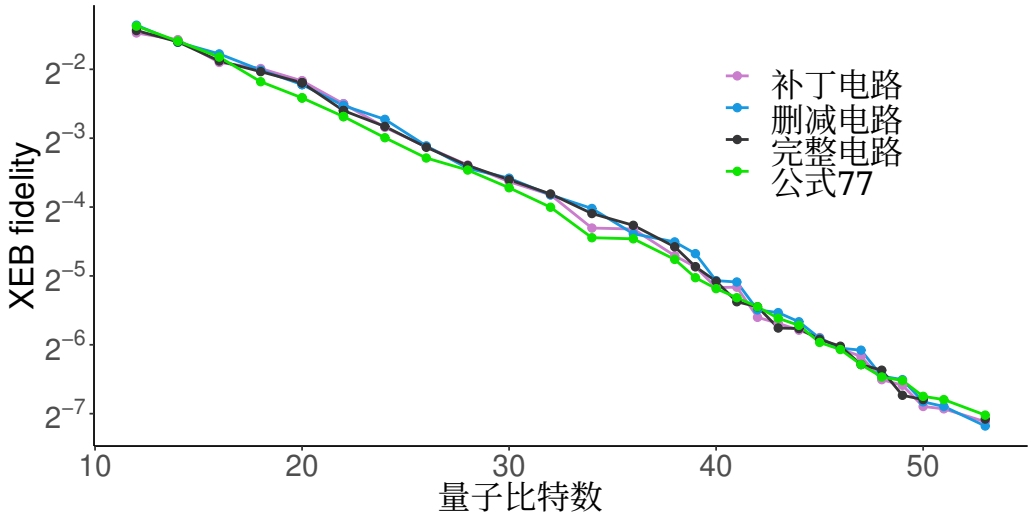

图9：完整电路、省略电路和补丁电路的平均 XEB 保真度与 (77) 的比较。数据来自表5

4.2 Google 团队的解释和统计论证

4.2.1 解释

Google 论文 [6]将公式（77）的预测能力视为系统规模扩大时不存在额外退相干物理的证明。以下是 Google 论文 [7]中的一段引文

> “显然，测得的保真度与预测的保真度非常吻合，偏差最多仅为 10-20%。鉴于此处的序列涉及数十个量子比特和～ 1000 量子门，这种吻合程度有力地证明了数字误差模型的有效性。测量结果与预测结果高度吻合，这一点还能进一步增强这一结论。

> 完整电路、补丁电路和删减电路的保真度。尽管这三种方法的门序列只有细微差异，却可能导致系统的计算复杂度以及子系统间的纠缠程度截然不同。通过这些不同方法测得的保真度彼此一致，且与根据各个门预测的保真度相符，这有力地证明了数字误差模型所作假设的正确性。此外，即使存在量子纠缠，这些假设依然成立。数字误差模型得到验证，具有至关重要的意义，尤其是对量子纠错而言。自量子纠错领域的首篇论文发表以来，人们普遍假设量子噪声不存在空间或时间相关性。我们的数据表明，现有量子处理器能够实现这一特性。”

此外，Google 团队认为公式（77）的预测能力是一项重要而令人惊叹的发现。以下是 John Martinis 于2019年11月1日的一段引述[25]:

> “现在我想指出，这些数据最令人惊叹之处是什么；如果你想用一种有趣的方式来思考，为什么我们这些量子科学家都能保住工作，甚至还能获得更多资金（那就更好了）。答案就在这些数据里，看看这条黑线……它只是根据我之前展示给你们的一量子比特和二量子比特同时数据得出的一个简单预测……你把所有这些保真度以完全经典的方式相乘，这就是高中阶段的概率论，而它却能预测这个极其复杂的系统中正在发生什么——其中进行着各种量子计算。对我来说，这绝对令人惊叹：如此简单的错误模型竟然如此有效。我们拿到数据时，真正让人惊讶的就是这一点。当然，如果情况更糟，我今天就不会站在这里和你们讲话了。”

随着用于计算振幅的高效经典算法问世，以及 Google 的量子霸权主张逐渐衰退，Google 团队成员和其他研究人员都认为，错误的独立性……

这使得公式（77）的预测能力成为 Google 实验中最重要的剩余科学成果，并将其所体现的统计独立性视为有力证据，表明量子容错所要求的错误统计独立性假设已经得到满足。

4.2.2 统计论证

Google 团队（2019年11月）通过一项基于以下三个要素的统计计算，为公式（77）（方程（4））非凡的预测能力提供了依据。

1. 对单个读出错误概率和门错误概率的估计是准确的。Google 团队报告称，他们对单个量子比特错误概率和门错误概率的估计，其错误率为 20%。 2. 对单个错误概率的估计不存在偏差；也就是说，这些估计不存在系统性错误。 3. 门错误和读出错误在统计上相互独立。

基于这些假设，Google 对公式（77）预测的相对偏差（粗略）估计为

$$DEV = 0.2 \cdot (\sqrt{n} \cdot 0.038 + \sqrt{|\mathcal{G}_1|} \cdot 0.0016 + \sqrt{|\mathcal{G}_2|} \cdot 0.0063). \qquad (12)$$

例如，对于 $n = 53$ 和 $m = 14$，单量子比特门的数量为 $|\mathcal{G}_1| = 795$ ，双量子比特门的数量为 $|\mathcal{G}_2| = 301$，根据公式12得出约为 8.6%。

这一估算的解释是，一旦我们接受评估实际保真度时的错误是无偏的（或非系统性的），其累积效应就会像组件数量的平方根那样增长（如简单随机游走）。例如，对于量子比特错误，单个量子比特发生错误的概率平均为 0.038(1 ± 0.2)。对于 n 个量子比特，(1 ± 0.2) 项的累积效应是将平均错误乘以 20%，得出 $0.2 \cdot \sqrt{n} \cdot 0.0\overline{38}$。

4.3 我们对公式（77）背后统计假设的看法

在我们看来，公式（77）背后最不合理的假设是统计独立性假设。假设量子计算机中数百甚至数千个组件的故障在统计上相互独立，是不合理的

。接下来，我们来讨论有偏差错误与无偏差错误的问题。下面的计算表明，即使偏差很小，也会对保真度产生很大影响。固定一个电路 C。假设对于 C 的每个组件，错误率以概率（ $0.5+t$ ）高出 20%，以概率（ $0.5-t$ ）低 20%。公式（77）的偏差（当 t 很小时）将为:

$$(1-2t)DEV + 2t \cdot (n \cdot 0.038 + |\mathcal{G}_1| \cdot 0.0016 + |\mathcal{G}_2| \cdot 0.0063),$$

其中， DEV是公式12中的值。对于 $n = 53$， $m = 14$和 $t = 0.05$ ，这给出的结果大约为50%。最后，我们注意到，在第3.5.1节中研究的非平稳行为可能会削弱公式（77）的统计论证。具体而言，1的比例发生的漂移可能与 ± 20%的估计值以及无偏差假设相冲突。

备注： 1） [6] 的方法已经假设错误之间存在某种依赖关系。也就是说，对于计算机的每个组件，当该组件独立工作时，我们有一个错误估计值；而当该组件作为更大规模计算的一部分工作时，我们有一个更高的错误估计值；公式（77）使用的正是这些更高的错误估计值。（假设更高的错误估计值适用于更大电路中的组件是合理的，但实际情况并非如此。）正如我们已经提到的，我们没有关于不同组件失效概率的数据，尤其没有关于错误估计值本身所声称的 ± 20%错误背后的统计数据。

2) 我们指出，错误之间的正相关实际上会带来更高的保真度。

3) Google 对公式（77）的统计论证中的三个要点假设， $\mathcal{F}_{XEB}$ 是 $p_{no\ err}$的良好估计量。相比之下，Gao 等人[11] 证明（另见第2.3节），存在系统性差异，并且

$$\mathcal{F}_{XEB} > p_{no\ err}.$$

造成这一差距的原因有多种，其中之一是两个错误可能相互抵消，产生“没有错误”的效果。Gao 等人在论文中预期，对于 Google 和中国科学技术大学的电路，“交叉熵基准测试的XEB值对保真度的高估仅为几个百分点”，但这一预期差距值得进一步理解。 $\mathcal{F}_{XEB}$ 与 $p_{no\ err}$ 之间的巨大差异将进一步成为质疑公式（77）的理由。因此，使用模拟进一步研究 $\mathcal{F}_{XEB}$ 与 $p_{no\ err}$ 之间的差异将很有意义。

4.4 对我们担忧的其他回应

几位研究人员为公式（77）的预测能力（以及所需的统计独立性）提出了各种可能的解释：Peter Shor（在一篇博客讨论中 [31]）、Boaz Barak 在 [12], 以及其他人提出的一种解释是，对于随机电路，公式（77）成功所需的统计独立性在典型情形下是合理的。我们看不出这一说法有何依据，并指出，公式（77）对非随机电路（例如补丁电路和删减电路）的预测也非常准确。

Adam Bouland（私人通信，2020年）提出了一个明确的模型，表明公式（77）也与各种形式的相关噪声相容：当某个特定的单量子比特门发生故障时，随机泡利算符会作用于电路中的所有量子比特，而不只是被该门作用的量子比特。Bouland 补充道："所以，我认为对他们结果更合理的批评是，他们的度量指标过于宽松（因为它掩盖了噪声中的相关性），而不是公式好得令人难以置信（我觉得很难这样说，除非质疑其实验结果的有效性，而这将是一个严重的指控）。"我们同意，公式（77）也支持各种相关噪声模型，但我们不认为这一评论与我们对公式（77）预测能力的担忧有直接关系。尽管将（77）的预测能力所体现出的统计独立性视为量子容错的乐观依据是合理的，但 Bouland 的例子表明，这并不足以排除可能导致量子容错失效的某些相关性（当然，这也不是必要条件）。

Greg Kuperberg（私人通信，2020年）区分了有运动部件的系统（如飞机）和没有运动部件的系统（如 Sycamore 芯片），并认为，对于有运动部件的系统，我们的担忧是合理的，但对于 Sycamore 芯片则不然；它"只有缩略图大小，没有运动部件（电子除外），并由稀释制冷机屏蔽。"

几位研究者指出，对公式（77）的担忧，实际上是在质疑 Google 数据和实验的完整性。例如，Adam Bouland 和 Dorit Aharonov 在 [17] 中提出了这一点。Aaronson [1] 认为，对公式（77）预测能力的担忧，等同于暗示 Google 的数据遭到篡改或伪造；但他也补充说，基于公式（77）作出的成功预测，恰恰是一次优秀实验本应得出的结果！

我们指出，在本文中，我们认为，对公式（77）感到担忧或对此感到惊讶，并不等于提出指控。

4.5 确认与复现

4.5.1 Kalachev、Panteleev 和 Yung 对公式（77）预测能力的确认

Kalachev、Panteleev 和 Yung 的近期研究为 Google 关于保真度的主张以及 Google 公式（77）的预测能力提供了有力支持 [13]。他们提出了改进的振幅计算算法，并计算了近两百个电路样本的振幅（$38 \leq n \leq 53$），而 Google 团队并未计算这些电路的振幅。在所有这些情况下，$\mathcal{F}_{XEB}$ 都与公式（77）的预测完全吻合。Liu 等人 [23]近期取得了另一项非凡的确认，他们成功计算出 Google 最大规模的 53 量子比特、深度为 20 的电路中的一个电路的振幅。$\mathcal{F}_{XEB}$ 保真度与公式（77）吻合。（当然，正如第 2.1 节所述，这些进展在很大程度上驳斥了 Google 的量子霸权主张。）

这些确认缓解了人们对公式（77）的担忧，因为它的预测能力适用于那两百个电路——Google 团队声称他们甚至没有计算其振幅——以及数十个他们根本无法计算振幅的电路。

4.5.2 复现与类似实验

中国科学技术大学对 Google 实验 [33, 35] 的近距离复现，也为 Google 的公式（77）的预测能力提供了独立支持。我们没有研究这些实验的数据，这仍是一项有趣的挑战。除中国科学技术大学的复现外，我们所知最接近的随机电路采样实验是 Kim 等人 [22]，（IBM 的一组研究人员）开展的。该论文介绍了最多包含六个量子比特（且深度较大）的电路。我们未能找到其他学术团体或公司关于量子计算机进行随机电路采样的报告和数据。

4.6 单个补丁和组合电路的公式（77）

请回想一下，补丁电路是通过删除作用于两个补丁中量子比特的双量子比特门，从完整电路中得到的。因此，两个补丁中的每一个都可以视为独立的随机电路。Google 团队于 2022年6月提供了补丁电路的数据（EFGH、$n = 53$、 $m = 14$ 类型的电路除外，这些数据仍未提供）。我们计算了补丁电路中两个补丁的保真度，并观察到补丁之间的保真度存在很大差距。例如，对于 $n = 40$，第一个补丁的保真度为 0.139，而第二个补丁为 0.220。对于 $n = 53$、$m = 20$ 实验，一部分的保真度为 0.026，而另一部分为 0.070。（参见附录中的表 6、表 7 和图 13。）相比之下，我们基于平均保真度所采用的近似（公式（10）和（11））会得出两个补丁的相近数值，因此并不适用。若要进一步研究这一点，我们需要论文中报告的、用于根据公式（77）计算预测值的单个量子比特保真度和门保真度数值。

4.7 补丁电路的公式（77）存在系统性偏差

要在 Google 论文中建立外推论证，就必须证明补丁、被省略部分和完整电路的线性交叉熵估计量一致；[6, 7]强调了这种一致性的重要性（例如，[6] 中的图4以及下面的引述）。然而，关于公式（77）的另一个疑虑（Google 论文本身也提出了这一点）是：公式（77）预测补丁电路的 $\mathcal{F}_{XEB}$ 保真度会比完整电路高 10%，但实际上，在数百个电路中，补丁电路与完整电路的保真度会一致。公式（77）之所以预测补丁电路的保真度高于完整电路，是因为补丁电路的双量子比特门数量少得多，而且被省略的双量子比特门不会计入乘积公式。

[7] 第 VIII.B 节中的以下引文提出了这个问题：

> “对于所研究的每种系统规模，我们发现，补丁 XEB 和完整 XEB 得出的保真度彼此吻合良好，典型偏差为保真度本身的 $\sim 5\%$（我们

> 将 34 个量子比特时最差情况下的 10% 偏差归因于两个数据集之间的暂时系统波动；交错测量保真度数据中也观察到了这一波动）。
>
> 从理论上讲，由于门数量略有减少，人们预计补丁电路的保真度会比完整电路高出 ∼10%。我们发现，补丁电路的表现略低于预期；我们认为，这是因为双量子比特门的幺正算子是针对完整操作而非补丁操作进行优化的。无论如何，补丁电路与完整电路之间的一致性表明，补丁电路可以很好地估算完整电路的表现。鉴于这两种方法产生的纠缠存在显著差异，这一点相当令人瞩目。这些结果让我们得以很好地预览此前讨论的三种状态下的系统性能。

作者确实指出，公式（77）预测补丁电路的保真度会更高（约高出 10%），并解释说，对公式（77）的系统性违背反映了“针对完整操作而非补丁操作进行优化”。这一解释显然指的是校准过程，但我们认为它缺乏说服力。我们将在第 5.4 节详细讨论这一问题。

第 4 节总结

公式（77）令人瞩目的预测能力在统计学上颇为惊人：量子计算机各组件之间隐含的独立性尤为突出。补丁电路与完整电路之间实验保真度的高度一致，显示出一种无法解释的、系统性偏离公式（77）预测的现象。另一方面，确认[13, 23]和复现 [33, 35] 为 Google 论文中的主张提供了支持。仍有许多问题有待探索。例如，(i) 利用含噪电路的模拟，可以估算公式（77）两边的差值大小（Gao 等人 [11]）；(ii) 公式（77）中的各个数值可用于研究补丁电路中两个补丁各自不同的保真度。

5 问题三：校准过程

5.1 校准过程及其局部性质

我们所说的校准过程，是指 Google 量子霸权实验中一种至关重要的方法。该方法基于多次运行单量子比特和双量子比特量子电路，根据 Sycamore 量子计算机的实际运行情况调整实验电路的定义。Google 论文（[7] 第 VI.B.4 节）将校准过程描述为一种优化算法，其“优化搜索空间 [是] [维] $\sim 100^{4N}$ [，其] 远大于由 N 个量子比特构成的处理器的希尔伯特空间维数，而后者为 2^N。”

需要明确的是，校准并非拧紧 Sycamore 上的螺丝，而是对程序进行更改。我们可以将校准过程视为对模型的更改，以大幅减少某些系统性噪声。例如，如果我们发现某个本应施加 90 度旋转的单量子比特门，实际上总是施加 80 度旋转，那么我们不会改变该单量子比特门的工程设计，而是会更改电路的定义。

校准包括对电路中每个双量子比特门 g 进行特定修改。如果双量子比特门 g 作用于量子比特 x 和 y，那么

- 双量子比特门 g 的参数经过了修改，此外，
- 在电路的定义中加入了两对单量子比特门（[7] 中称为“Z-旋转”）。每对门中，一个作用于量子比特x，另一个作用于量子比特 y；一对在执行双量子比特门 g 之前加入，另一对则在其执行之后加入。

电路中的单量子比特门没有进行其他修改。（更多详情请参见 [20]第 3 节。）这些校准修改是在补丁、省略电路和完整电路上，以**EFGH** 模式进行的所有实验中同时完成的。（对于**ABCDCDAB**模式下不可验证的完整电路，则单独运行了校准过程。）

Google 团队向我们解释说，对于给定的双量子比特门 g，大部分 Z 旋转是为了补偿“相互作用的量子比特因空闲频率不同而产生的相位变化”，其数值由固定的量子比特频率和量子门时长决定。（此外，在较新的

实验中，这一部分已被移除。）暂且不考虑 Z 旋转的这一部分，这种校准方法会修正实验电路相对于其所代表的随机电路中的小幅系统误差。

针对作用于量子比特 x 和 y 的双量子比特门 g 所进行的校准调整，是基于涉及这两个量子比特的多个单量子比特和双量子比特电路运行的数据。多个单量子比特和双量子比特量子电路并非单独运行，而是并行运行；每次运行都包含相同类型的双量子比特门层，其中，与实验电路一样，这些双量子比特门层之间插入了随机单量子比特门层。我们注意到，应用相同类型的双量子比特门层并在这些层之间插入随机单量子比特门时，每个量子比特至多与另一个量子比特相互作用；因此，我们可以并行运行许多双量子比特电路（以及单量子比特电路）。该校准方法基于运行仅包含一种类型双量子比特门层的这些电路，并调整双量子比特门以最大化 $\mathcal{F}_{XEB}$ 值。似乎同一算法也可用于双量子比特门层类型不同的情形，尤其是可直接用于实验电路本身。

如 [20], 所述，Google 团队根据在小型电路（包含 1 个和 2 个量子比特）上收集的数据，为每个双量子比特门制定修改方案时所采用的具体流程，属于商业机密。我们也没有作为该流程输入的小型电路实验的实验数据。在 [20] 中，我们提出了有关校准过程的其他几个非统计问题，并在此提及其中两个。第一个问题是，校准过程削弱了“可编程量子计算机”这一主张。（我们的统计研究为这一问题提供了佐证。）第二个问题是，校准过程的改进与实验交错进行，而且，针对**EFGH** 电路的最后一刻校准流程（见 [20]第 4.5 节）带来了显著改进。

校准过程的关键特性是其局部性，而且是以下这种严格意义上的局部性：对于电路中描述所有双量子比特门的大量参数，并不进行一般性的优化；相反，对作用于量子比特 x 和 y 的双量子比特门 g 所作的修改，是基于涉及 x 和 y 的单量子比特及双量子比特电路的运行结果，主要旨在消除这一特定双量子比特门的系统噪声。在本节其余部分，我们将对校准过程进行统计研究，并分享一些出人意料的统计发现，尤其考虑到校准的局部性。

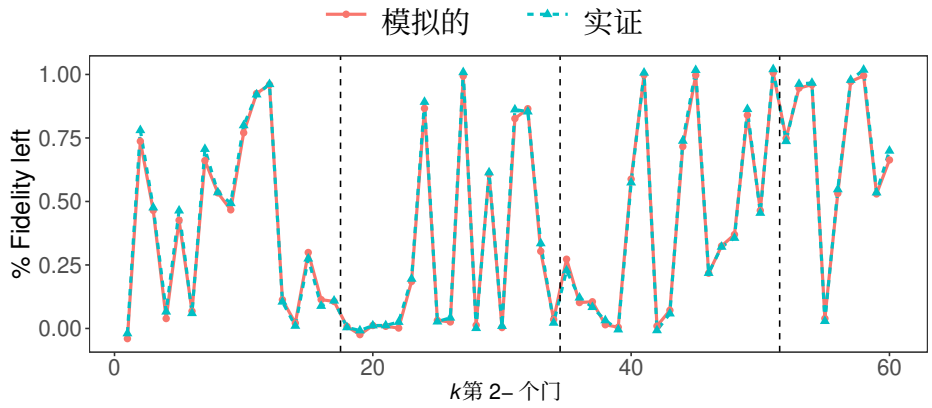

图10：校准的惊人效果 I。对于 Google 的文件 0 中带有 $n = 12$ 的电路，移除第 k 个双量子比特门的校准所产生的影响。请注意，同一个双量子比特门会沿着电路周期性出现，如黑色竖直虚线所示。这里我们移除了校准的所有组成部分：单量子比特门旋转和双门调整。

5.2 结果完全依赖于校准过程 S

如果我们从一个实验随机电路开始，不进行单量子比特门和双量子比特门的修改，并计算实验比特串的 $\mathcal{F}_{XEB}$ 保真度，就会发现保真度骤降至零。此外，即使我们只应用与某个特定双量子比特门的一次出现相对应的修改，对保真度的影响仍然很大：在许多情况下影响显著，有些情况下保真度会骤降至接近零，另一些情况下影响则较小（图10）。关于这一点，Google 团队提请我们参阅 [7]，的第 VII.E 节，其中解释了每个双量子比特门对应量子比特的四次 Z 旋转所校正的系统误差：“ Z 旋转有两个来源。首先，它们捕捉双量子比特门期间量子比特频率变化所导致的相移。其次，它们补偿相互作用的量子比特空闲频率不同所导致的相位变化。后者使定义 Z 旋转的三个参数取决于门的施加时刻。相比之下，对于给定的量子比特对， θ 和 ϕ 不依赖于周期。”

5.3 校准效果似乎好得不合理

在图 10 中，我们还比较了电路 60 个双量子比特门中每个门对应的两个量。第一个量表示在不对单个双量子比特门应用校准时，剩余保真度所占的比例。这些保真度用蓝色小三角形表示。第二个量假设校准能完美描述电路：我们根据校准后的电路进行采样，然后计算取消双门校准时的保真度。这些保真度用红色小圆圈表示。红色实线与蓝色虚线完美吻合，表明双门校准非常有效，堪称“恰到好处”。校准通过抵消各个双量子比特门上的系统误差来实现“噪声消除”；令人惊讶的是，校准过程竟然实现了完美的噪声消除。从数学上讲，各个双量子比特门的改进都近乎完美，这意味着校准过程在不同双量子比特门的众多参数 $\mathcal{F}_{XEB}$构成的函数中达到了（近似的）临界点。若是通过全局优化过程，同时调整所有这些参数来最大化$\mathcal{F}_{XEB}$，那么达到临界点就不那么令人惊讶了。

5.4 Google 对补丁与完整电路之间 $\mathcal{F}_{XEB}$估算结果一致性的解释

在第4.7节中，我们讨论了这样一个疑虑：尽管公式（77）预测补丁电路的 $\mathcal{F}_{XEB}$ 保真度将比完整电路的保真度高出10%，但实际上，在数百个电路中，补丁电路与完整电路的保真度却是一致的。Google 论文 [7] 对此给出了解释，认为公式（77）的系统性偏离反映了“针对完整操作而非补丁操作进行优化”，显然指的是本节讨论的校准过程。涉及量子比特 x和 y 的双量子比特门 G 的校准，是基于多次运行涉及这两个量子比特的单量子比特和双量子比特电路来进行的。具体做法是交替执行相同类型的双量子比特门层，以及夹在这些门层之间的随机单量子比特门层。“完整”操作与“补丁”操作之间的区别在于，前者包含少数几个不与量子比特 x 和 y直接相互作用的双量子比特门，而这种差异似乎不太可能产生任何影响。

5.5 双门调整似乎异常有效

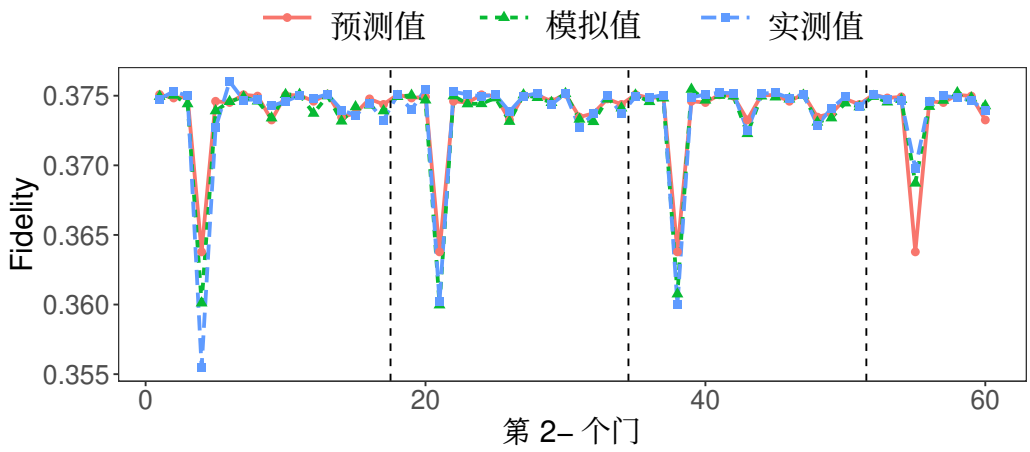

图11：校准 II 的显著效果。移除电路中涉及第 k 个双量子比特门的双门调整所产生的影响。（同一个双量子比特门会沿着电路周期性出现。）涉及量子比特 (3,3) 和 (3,4) 的双量子比特门始终有很大影响。

在本小节中，我们保留校准过程中添加的 Z-旋转不变，并研究调整双量子比特门定义所产生的影响。Google 团队声称，对于使用 $n = 53$ 和 $n = 14$的实验，移除双门调整会使保真度降低一半，并给出了公式（13）和（14）（见下文）来估算这一影响。图11 中，移除首次实验 $n = 12$中 60 个双量子比特门各自的双门调整对保真度的影响以蓝色方块表示。小绿三角表示通过模拟得出的同一影响（每个双量子比特门进行 50 次模拟），并假设调整是完美的。红色圆圈表示通过下文公式（13）得出的影响（同样假设调整是完美的）。经验测得的影响与假设调整达到最优时的结果非常接近，表明校准“恰好达到了预期效果”。从技术角度来说，当我们将 $\mathcal{F}_{XEB}$ 视为双量子比特门参数的函数时，校准得到的参数接近该函数的一个临界点。

n	12	14	16	18	20	22	24	26
经验数据	0.793	0.789	0.784	0.782	0.785	0.766	0.757	0.712
模拟	0.781	0.773	0.770	0.766	0.763	0.744	0.732	0.676
理论	0.818	0.813	0.808	0.803	0.799	0.783	0.772	0.725

表2：双门调整的影响，按十个文件取平均：基于经验数据、基于模拟（假设调整完美），以及基于公式（14）（假设调整完美）。

我们还比较了移除所有双门调整的影响。表2第一行描述的是经验样本相对于移除双门调整后的电路的保真度，与其相对于校准后的量子电路的保真度之间的比值（按十个文件取平均）。第二行描述的是在假设双门调整完美的情况下，基于模拟得出的这一比值。第三行描述的是在假设双门调整完美的情况下，基于公式（13）和（14）得出的这一比值。图 12 分别展示了十个 Google 文件的情况。

备注： 1. 校准中还有一个我们不了解的特性：对于**EFGH** 电路，涉及量子比特 (3,3) 和 (3,4) 的双量子比特门的调整效果要明显得多。移除这个双量子比特门调整的影响——该门在电路中出现四次——就是图11中的四个大凹谷。

2. 以下是 Google 团队给出的双门调整效果公式。假设双门调整是完美的，那么移除单个双门（其参数为 θ_i 和 ϕ_i）的校准所导致的保真度下降估计为

$$\psi_i = \big(4 + \cos(2(\theta_i - \pi/2)) + \cos(\phi_i - \pi/6) + 2 * \cos(\theta_i - \pi/2) * (1 + \cos(\phi_i - \pi/6))\big)/10, \quad (13)$$

整体保真度下降幅度的估计值，是将所有双量子比特门对应的这些量相乘得到的（基于不同调整之间相互独立的统计假设）：

$$R_0 = \prod_{i=1}^{|\mathcal{G}_2|} \psi_i. \quad (14)$$

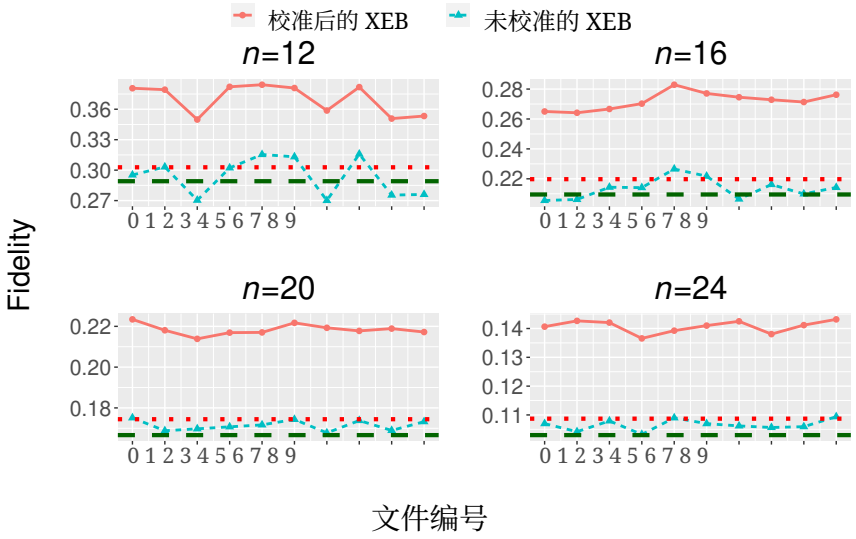

图 12：移除双门调整（对所有双门）的效果，与基于模拟和 Google 估算程序得出的理想效果相比。 ϕ 取平均值 F_{XEB}。红色虚线表示理论上的下降，深绿色表示模拟。

5.6 第 5 节总结

校准过程会考虑双量子比特门的系统误差，并对电路的定义进行某些调整。这些调整是局部的，即涉及量子比特 x 和 y 的双量子比特门，其调整主要取决于这两个量子比特上的单门和双门电路的结果。关于校准过程的一些统计发现包括：(i) 即使只有一个双量子比特门，其影响也很大；(ii) 不同双量子比特门的影响差异很大，甚至同一双量子比特门在不同位置出现时，影响也各不相同；(iii) 双门校准的效果显著。

我们注意到，这些发现加剧了校准过程与 Google 关于“可编程量子计算机”的主张之间的矛盾。鉴于校准过程具有局部性，其效果尤其令人惊讶：从数学角度来说，我们看到一个局部优化过程到达了一个依赖数百个参数的函数的临界点。

6 总结

在本文中，我们主要从统计学角度研究了三个问题。第一个问题已在 [30] 中得到展示：Google 样本与 Google 噪声模型，或我们研究过的任何特定噪声模型之间存在巨大差距。这一发现削弱了 [6] 中关于成功完成采样任务的主张，即针对某个特定分布实现近似采样；同时也可能让人对 NISQ 系统完成采样任务的潜力产生怀疑。

第二个问题涉及 Google 论文（4）中的公式（77），其形式简单，类似于独立性。再加上实验过程中涉及的噪声和校准，Google 的保真度估计值在数百次不同实验中似乎都相当接近公式（77），这令人惊讶。科学中确实会出现令人惊讶的现象，并可能带来重要的科学发现；但从统计学角度来看，对于量子计算机等已知易受噪声以及与环境相互作用所致错误影响的系统，其中各组件之间隐含的独立性十分引人注目。补丁电路中公式（77）存在的系统性偏差，以及两个补丁的 $\mathcal{F}_{XEB}$ 保真度之间的巨大差异，也引出了相关问题。Kalachev 等人 [13] 和 Liu 等人 [23] 做出的出色验证，通过表明公式（77）能够很好地近似数百个电路的结果，缓解了这些关注问题；Google 团队曾表示，他们并未计算这些电路的振幅。中国科学技术大学对 Google 实验进行了高度吻合的复现，也缓解了我们对公式（77）和 Google 实验的关注。

本文讨论的第三个问题是 Google 实验中的校准过程，其细节已申请专利，且并未完整公开。电路需要校准这一事实，以及校准调整所产生的巨大影响，都削弱了这样一种说法：Google 的实验描述的是一台可编程量子计算机。校准过程近乎完美的有效性令人惊讶，尤其是考虑到其局部性。

我们陈述了客观事实和研究发现，并让读者自行解读，或者更确切地说，请读者留意我们的关注问题，并等待未来实验提供更多实验数据。我们希望，我们对 [30, 20] 和本文中 Google 实验的详细分析，将有助于学术界评估 Google 的量子霸权主张 [6, 7], ，以及在实验性含噪中等规模量子计算这一极为活跃的领域中提出的其他影响深远的主张。

参考文献

[1]S. Aaronson，《Shtetl Optimized》上的一条评论，2020年2月。https://scottaaronson.blog/?p=4608#comment-1830402 。[2] S. Aaronson 和 L. Chen，量子霸权实验的复杂性理论基础，第 *32* 届计算复杂性会议论文集，*CCC '17*（拉脱维亚里加）第 1–67 页，2017年。arXiv:1612.05903。[3] S. Aaronson 和 S. Gunn，关于伪造线性交叉熵基准测试的经典难度，2019年，arXiv:1910.12085。[4] R. Acharya 等人，通过扩展表面码逻辑量子比特来抑制量子错误，《自然》614（2023年），676–681。[5] D. Aharonov、X. Gao、Z. Landau、Y. Liu、U. Vazirani，一种用于噪声随机电路采样的多项式时间经典算法，2022年，arXiv:2211.03999。[6] F. Arute 等人，使用可编程超导处理器实现量子霸权，《自然》574（2019年），505–510。[7]F. Arute 等人（2019年），《使用可编程超导处理器实现量子霸权》的补充信息，2019年，arXiv:1910.11333。[8] S. Boixo、V. N. Smelyanskiy 和 H. Neven，噪声混沌量子电路采样的傅里叶分析，2017年，arXiv:1708.01875。[9] A. Bouland 等人、B. Fefferman、C. Nirkhe、C. 等人，关于量子随机电路采样的复杂性与验证，《自然·物理》15（2019年），159–163。[10] X. Gao 和 L. Duan，噪声量子计算的高效经典模拟，2018年，arXiv:1810.03176。[11] X. Gao、M. Kalinowski、C.-N. Chou、M. D. Lukin、B. Barak 和 S. Choi，线性交叉熵作为量子优势度量的局限性，2021年，arXiv:2112.01657。

[12]S. Irani（主持人），量子霸权专题讨论，耶路撒冷希伯来大学，2019年12月。网址：https://youtu.be/ Yb7uIGBynU 。[13] G. Kalachev、P. Panteleev 和 M.-H. Yung，采用多张量收缩验证量子电路的交叉熵基准测试，2021年，arXiv:2108.05665。[14]G. Kalachev、P. Panteleev、P. F. Zhou 和 M.-H. Yung，以有界保真度对随机量子电路进行经典采样，2021年，arXiv:2112.15083。[15] G. Kalai，数学、计算与游戏的三个谜题，载于国际数学家大会论文集，2018年，里约热内卢，第I卷，2018年，第551–606页。[16] G. Kalai，反对量子计算机的论证，载于：M. Hemmo 和 O. Shenker（编），量子、概率与逻辑：*Itamar Pitowsky* 的工作及其影响，施普林格（2020年），第399–422页，arXiv:1908.02499。[17] G. Kalai，谷歌量子霸权演示，录像讲座，2019年12月。网址：https://youtu.be/p18P1y8GD9U。[18] G. Kalai，反对量子计算机的论证、量子自然定律与谷歌的量子霸权主张，载于：洲际学院项目《定律：刚性与动力学》（M. J. Hannon 和 E. Z. Ra- binovici 编），世界科技出版公司（即将于2023年出版）。arXiv:2008.05188。[19] G. Kalai 和 G. Kindler，高斯噪声敏感性与玻色子采样，2014年，arXiv:1409.3093。[20]G. Kalai、Y. Rinott 和 T. Shoham，Google 的量子霸权主张：数据、文档与讨论，2022年，arXiv:2210.12753。[21] G. Kalai、Y. Rinott 和 T. Shoham，通过随机电路采样展示量子优势：傅里叶展开与统计学，进行中的手稿。[22] J.-S. Kim、L. S. Bishop、A. D. Corcoles、S. Merkel、J. A. Smolin、S. Sheldon，用于对多量子比特系统中的噪声进行分类的硬件高效随机电路，《物理评论*A*》104（2021年），022609。[23] Y. Liu 等人，用精确快速的张量网络收缩验证量子霸权实验，2022年，arXiv:2212.04749。

[24] A. P. Lund、M. J. Bremner 和 T. C. Ralph，量子采样问题、玻色子采样与量子霸权，《自然·物理》期刊，*Quantum Inf*3（2017年），15。[25]J. Martinis，使用可编程超导处理器实现量子霸权，录像讲座，加州理工学院，2019年11月。https://youtu.be/FklMpRiTeTA 。[26] M. McEwen 等人，从大型超导量子比特阵列中解析宇宙射线造成的灾难性错误突发，《自然·物理》18（2022年），107–111。[27] F. Pan、K. Chen 和 P. Zhang，求解Sycamore量子电路的采样问题，《物理评论快报》129（2022年），090502。[28] F. Pan 和 P. Zhang，模拟Sycamore量子霸权电路，2021年，arXiv:2103.03074。[29] E. Pednault、J. A. Gunnels、G. Nannicini、L. Horesh 和 R. Wisni- eff，利用辅助存储模拟深层54量子比特Sycamore电路，2019年，arXiv:1910.09534。[30] Y. Rinott、T. Shoham 和 G. Kalai，量子霸权演示的统计方面，《统计科学》37（2022年），322–347。[31]P. Shor，关于“组合数学及更多”的评论，2019年11月。网址：https://gilkalai.wordpress.com/2019/11/13/gils-collegial-quantum-supremacy-skepticism-faq/#comment-61697 。[32] G. Valiant 和 P. Valiant，估计不可见数据：熵及其他属性的改进估计量，《美国计算机学会会刊》64（2017年），1–41。[33] Y. Wu 等人，使用超导量子处理器实现强量子计算优势。《物理评论快报》127（2021年），180501。[34] Y. Zhou、E. M. Stoudenmire 和 X. Waintal，什么限制了量子计算机的模拟？，2020年，arXiv:2002.07730。[35] Q. Zhu 等人，通过60量子比特、24周期随机电路采样实现量子计算优势，《科学通报》，第67卷（2022年），240–245。arXiv:2109.03494。

附录：补充信息

$n = 12$ **各种分布之间的距离**，14

n=12						
模型 1	模型 2	$\mathcal{X}^2$	L_1	L_2	KL	Cor
Google 噪声模型	Google 样本	41190	112072	2370	0.040	0.77
	基于 Google 噪声模型的模拟	4090	35569	707	0.004	0.97
	基于对称噪声模型的模拟	5555	41110	813	0.006	0.96
	基于非对称噪声模型的模拟 12342		61462	1221	0.012	0.92
	基于均匀分布的模拟	56297	140502	2983	0.058	~ 0
Google 样本	对称噪声模型	40018	110498	2343	0.039	0.78
	非对称噪声模型	36786	105468	2252	0.035	0.80

n=14						
模型 1	模型 2	$\mathcal{X}^2$	L_1	L_2	KL	Cor
Google 噪声模型	Google 样本	54286	129436	1339	0.052	0.69
	基于 Google 噪声模型的模拟	16381	71385	707	0.017	0.88
	基于对称噪声模型的模拟	17685	74058	732	0.018	0.87
	基于非对称噪声模型的模拟 25347		88504	878	0.025	0.82
	基于均匀分布的模拟	58825	139950	1473	0.060	~ 0
Google 样本	对称噪声模型	52903	127792	1323	0.052	0.70
	非对称噪声模型	50093	124112	1290	0.049	0.72

表 3：Google 样本、Google 噪声模型、对称噪声模型和非对称噪声模型与其他模型及其所依据的模拟之间的距离。Google 样本的距离取十个 Google 文件的平均值。每次模拟基于 100 次重复。 ϕ 模拟中的取值为平均 $\mathcal{F}_{XEB}$，对于 $n = 12$为 0.3701，对于 $n = 14$为 0.3298。 ϕ_{ro}，即对称读出系数，为 $\phi/((1-0.038)^n) - \phi$，而非对称读出模型的系数为 $\phi_{ro} + \phi$；两种情况下， ϕ 均为平均值$\mathcal{F}_{XEB}$。

与公式（77）及其简化版本相比的 $\mathcal{F}_{XEB}$-保真度。数值由补充数据提供，见 [6]。

n	m	电路类型	单门数量	双门数量	公式（10）	公式（11）	（77）-Google	大道 $\mathcal{F}_{XEB}$
12	14	EFGH	180	60	0.3242	0.3586	0.3862	0.3701
14	14	EFGH	210	70	0.2687	0.3023	0.332	0.3298
16	14	EFGH	240	80	0.2228	0.2548	0.2828	0.2721
18	14	EFGH	270	94	0.1801	0.2069	0.2207	0.2442
20	14	EFGH	300	105	0.1483	0.1728	0.1875	0.2184
22	14	EFGH	330	116	0.1222	0.1443	0.1554	0.1650
24	14	EFGH	360	130	0.0988	0.1171	0.1256	0.1407
26	14	EFGH	390	140	0.0819	0.0987	0.1024	0.1141
28	14	EFGH	420	150	0.0679	0.0832	0.09072	0.0949
30	14	EFGH	450	164	0.0549	0.0676	0.07594	0.0823
32	14	EFGH	480	175	0.0452	0.0564	0.06236	0.0713
34	14	EFGH	510	186	0.0372	0.0471	0.04592	0.0586
36	14	EFGH	540	200	0.0301	0.0383	0.04535	0.0520
38	14	EFGH	570	210	0.0249	0.0322	0.03693	0.0419*
39	14	EFGH	585	217	0.0224	0.0291	0.03073	0.0343*
40	14	EFGH	600	224	0.0202	0.0262	0.02751	0.0296*
41	14	EFGH	615	227	0.0186	0.0245	0.02514	0.0241*
42	14	EFGH	630	234	0.0167	0.0221	0.0229	0.0229*
43	14	EFGH	645	241	0.015	0.0199	0.02043	0.0185*
44	14	EFGH	660	248	0.0135	0.0179	0.01901	0.0184*
45	14	EFGH	675	255	0.0122	0.0161	0.016	0.0165*
46	14	EFGH	690	259	0.0111	0.0150	0.01485	0.0154*
47	14	EFGH	705	266	0.0100	0.0135	0.01281	0.0128*
48	14	EFGH	720	273	0.009	0.0122	0.01134	0.0121*
49	14	EFGH	735	280	0.0081	0.0109	0.01088	0.0094*
50	14	EFGH	750	287	0.0073	0.0099	0.009337	0.0090*
51	14	EFGH	765	294	0.0065	0.0089	0.008978	-
53	14	EFGH	795	301	0.0055	0.0077	0.007666	0.0074*
53	12	ABCDCDAB	689	258	0.0086	0.0115	0.01206	0.0147*
53	14	ABCDCDAB	795	301	0.0055	0.0077	0.007925	0.0094*
53	16	ABCDCDAB	901	344	0.0036	0.0052	0.005308	0.0060*
53	18	ABCDCDAB	1007	387	0.0023	0.0035	0.003555	-
53	20	ABCDCDAB	1113	430	0.0015	0.0023	0.002335	0.0019**

公式 (77)、(10) 和 (11) 对完整电路的预测能力。基于以下内容的补充数据 [6]。* 由 Kalachev 等人计算（[13] 以及私人通信）。** 在 [23]中针对一个电路计算。

$\mathcal{F}_{XEB}$ 对于完整、省略和补丁电路，以及公式（77）

n	电路类型	m	补丁	省略	full	预测（77）
12	EFGH	14	0.3604	0.3898	0.3701	0.3862
14	EFGH	14	0.3364	0.3282	0.3298	0.3320
16	EFGH	14	0.2679	0.2920	0.2721	0.2828
18	EFGH	14	0.2520	0.2481	0.2442	0.2207
20	EFGH	14	0.2225	0.2144	0.2184	0.1875
22	EFGH	14	0.1767	0.1735	0.1650	0.1554
24	EFGH	14	0.1394	0.1511	0.1407	0.1256
26	EFGH	14	0.1146	0.1157	0.1141	0.1024
28	EFGH	14	0.0954	0.09234	0.0949	0.0907
30	EFGH	14	0.0807	0.0836	0.0823	0.0759
32	EFGH	14	0.0707	0.0706	0.0713	0.0624
34	EFGH	14	0.0506	0.0616	0.0586	0.0459
36	EFGH	14	0.0501	0.0478	0.0520	0.0454
38	EFGH	14	0.0383	0.0440	0.0419*	0.0369
39	EFGH	14	0.0342	0.0391	0.0343*	0.0307
40	EFGH	14	0.0278	0.0298	0.0296*	0.0275
41	EFGH	14	0.0279	0.0294	0.0241*	0.0251
42	EFGH	14	0.0206	0.0222	0.0229*	0.0229
43	EFGH	14	0.0194	0.0216	0.0185*	0.0204
44	EFGH	14	0.0181	0.0197	0.0184*	0.0190
45	EFGH	14	0.0168	0.0167	0.0165*	0.0160
46	EFGH	14	0.0150	0.0151	0.0154*	0.0149
47	EFGH	14	0.0139	0.0148	0.0128*	0.0128
48	EFGH	14	0.0110	0.0114	0.0121*	0.0113
49	EFGH	14	0.0104	0.0110	0.0094*	0.0109
50	EFGH	14	0.0084	0.0088	0.0090*	0.0093
51	EFGH	14	0.0082	0.0084	-	0.0090
53	EFGH	14	0.0072	0.0069	0.0074*	0.0077
53	ABCDCDAB	12	0.0131	0.0139	0.0147*	0.0121
53	ABCDCDAB	14	0.0085	0.0090	0.0094*	0.0079
53	ABCDCDAB	16	0.0054	0.0058	0.0060*	0.0053
53	ABCDCDAB	18	0.0033	0.0039	-	0.0036
53	ABCDCDAB	20	0.0022	0.0022	0.0019**	0.0023

表5：完整电路、删减电路和补丁电路的保真度，与 Google 的公式（77）进行比较，$n = 12 - 53$。补丁电路和删减电路的数值、基于公式（77）的预测，以及完整电路的数值（针对$n < 38$），均由以下文献的补充数据提供：[6]。* 由 Kalachev 等人计算（[13] 及私人通信）**一个电路。计算见 [23]

补丁电路的数据

下表展示了我们基于 Google 团队提供的数据和程序所进行的计算。（与表5中的$\mathcal{F}_{XEB}$-值存在一些细微差异。）

比特串	第一个补丁			第二个补丁			组合电路		
	n	**XEB**	公式（10）	n	**XEB**	公式（10）	交叉熵基准测试的乘积	公式（10）	公式（77）
12	6	0.5771	0.6022	6	0.5547	0.6022	0.3201	0.3626	0.3862
14	6	0.5677	0.6022	8	0.5759	0.4991	0.3269	0.3006	0.3320
16	8	0.4928	0.4991	8	0.5506	0.4991	0.2713	0.2491	0.2828
18	9	0.4707	0.4488	9	0.5323	0.4488	0.2506	0.2014	0.2207
20	9	0.4722	0.4488	11	0.4798	0.3697	0.2266	0.1659	0.1875
22	11	0.3825	0.3697	11	0.4635	0.3697	0.1773	0.1367	0.1554
24	12	0.3293	0.3324	12	0.4213	0.3324	0.1387	0.1105	0.1256
26	14	0.2670	0.2755	12	0.4279	0.3324	0.1142	0.0916	0.1024
28	14	0.2695	0.2755	14	0.3509	0.2755	0.0946	0.0759	0.0907
30	15	0.2470	0.2477	15	0.3270	0.2477	0.0808	0.0614	0.0759
32	17	0.2131	0.2041	15	0.3326	0.2477	0.0709	0.0506	0.0624
34	17	0.2068	0.2041	17	0.2437	0.2041	0.0504	0.0417	0.0459
36	18	0.1898	0.1835	18	0.2632	0.1835	0.0500	0.0337	0.0454
38	19	0.1611	0.1650	19	0.2426	0.1691	0.0391	0.0279	0.0369
39	20	0.1432	0.1511	19	0.2436	0.1691	0.0349	0.0256	0.0307
40	20	0.1394	0.1511	20	0.2197	0.1521	0.0306	0.0230	0.0275
41	21	0.1044	0.1393	20	0.2177	0.1521	0.0227	0.0212	0.0251
42	21	0.1149	0.1393	21	0.1930	0.1367	0.0222	0.0190	0.0229
43	22	0.0972	0.1253	21	0.1841	0.1367	0.0179	0.0171	0.0204
44	22	0.0955	0.1253	22	0.1640	0.1230	0.0157	0.0154	0.0190
45	23	0.0879	0.1126	22	0.1643	0.1230	0.0144	0.0138	0.0160
46	23	0.1004	0.1126	23	0.1509	0.1126	0.0152	0.0127	0.0149
47	24	0.0876	0.1013	23	0.1493	0.1126	0.0131	0.0114	0.0128
48	24	0.0931	0.1013	24	0.1347	0.1038	0.0125	0.0105	0.0113
49	25	0.0688	0.0911	24	0.1448	0.1038	0.0100	0.0095	0.0109
50	25	0.0723	0.0911	25	0.1264	0.0934	0.0091	0.0085	0.0093
51	25	0.0717	0.0911	26	0.1170	0.0839	0.0084	0.0076	0.0090

表 6：补丁电路的保真度，电路类型为 **EFGH**。

m	第一个补丁			第二个补丁			组合电路		
	n	**XEB**	公式（10）	n	**XEB**	公式（10）	**交叉熵基准测试**的乘积	公式（10）	公式（77）
12	27	0.0821	0.095	26	0.1563	0.1027	0.0131	0.0098	0.0121
14	27	0.0647	0.0769	26	0.1325	0.0839	0.0085	0.0065	0.0079
16	27	0.0486	0.0623	26	0.1113	0.0682	0.0054	0.0042	0.0053
18	27	0.0354	0.0504	26	0.0855	0.0554	0.0033	0.0028	0.0036
20	27	0.0261	0.0409	26	0.0697	0.0453	0.0022	0.0019	0.0023

Ta表 7：补丁电路 $n = 53$ 的保真度，电路类型**ABCDCDAB**。

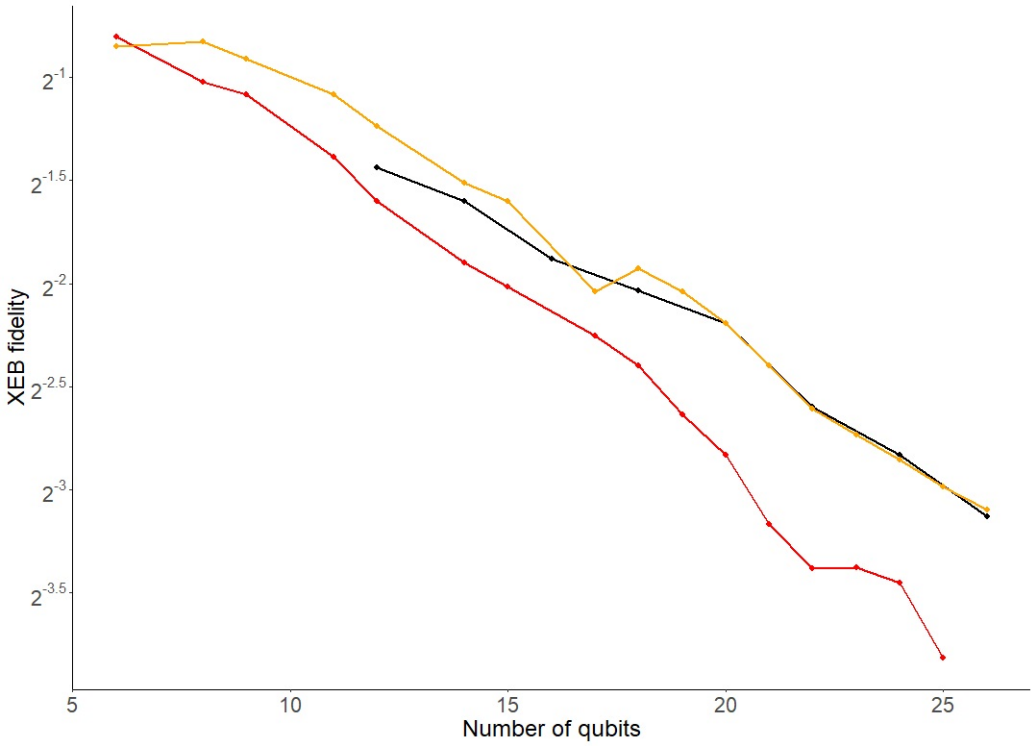

图 13：第一个补丁（红色）、第二个补丁（橙色）和完整电路（电路类型 **EFGH**）的平均 XEB 保真度对比。第一个补丁的表现明显不同。依据表 6。

Gil Kalai，耶路撒冷希伯来大学、爱因斯坦数学研究所，以及赖希曼大学埃菲·阿拉齐计算机科学学院。gil.kalai@gmail.com。

Yosef Rinott，耶路撒冷希伯来大学、费德曼理性研究中心和统计学系。yosef.rinott@mail.huji.ac.il。

TomerShoham，耶路撒冷希伯来大学、费德曼理性研究中心和计算机科学系。tomer.shohamm@gmail.com。