arXiv:2302.04144v1 [quant-ph] 8 Feb 2023

量子计算机的能量估算基准测试

Andreas J. C. Woitzik[1], * Lukas Hoffmann, [1] Andreas Buchleitner, [1,2] 和 Edoardo G. Carnio
[1,2]1弗赖堡阿尔伯特-路德维希大学物理研究所，赫尔曼-赫德尔街 3 号，79104，德国弗赖堡2EUCOR量子科学与量子计算中心，弗赖堡阿尔伯特-路德维希大学，赫尔曼-赫德尔街 3 号，79104，德国弗赖堡
（日期：编译于 2023年2月9日星期四 01:31）

要认证量子计算机的性能，需要采用标准化测试。我们提出一种简单的能量估算基准测试，其灵感源自量子化学。利用这一基准测试，我们对德国埃宁根 IBM Quantum System One 的噪声结果进行统计表征。我们发现，基准测试结果与该设备报告的量子门误差和读出误差几乎没有相关性。在时间分辨分析中，我们对设备进行数小时监测，发现基准测试结果存在两小时周期的振荡以及一些离群值，而这些现象无法用设备报告的状态加以解释。随后我们表明，已实施的测量误差缓解技术无法消除这些振荡，这表明理论基准结果统计数据与实际结果之间的偏差并非仅源于设备的测量噪声。

I. 引言

选择何种物理平台来编码信息，决定了计算规则 [1]。基于这一前提，量子计算机于 20 世纪 80 年代被提出 [2]，如今已成为研究热点。这类设备的目标应用涵盖量子系统模拟 [3, 4] ，以及更高效地求解某些优化问题 [5]。已有研究证明，若干算法 [6, 7]，（例如格罗弗提出的算法 [8], ）在理想（即无噪声）量子计算机上的运行速度优于经典计算机。然而，当前量子硬件 [3] 的规模和噪声水平尚不足以支持可靠地运行有用算法并实现可观测的加速。目前，人们正投入大量精力改进硬件相关方面 [9–12], 以及算法层面 [3, 7, 13, 14]。为了比较当前易出错量子设备的可靠性水平，需要基准测试，即用于评估其性能的预定义任务 [15]。正因如此，过去几年间，人们对面向量子比特数量较少、错误率较高的量子硬件的基准测试兴趣显著增加；对于这类硬件，量子纠错 [6] 仍遥不可及。量子信息处理领域存在不同类型的基准测试。早在 20 世纪 90 年代末，人们便提出了态制备基准测试的初步构想，用于衡量将量子比特寄存器制备到给定目标态的保真度 [16]。一些基准测试针对特定应用，例如肖尔算法 [17], ；另一些则关注门误差 [18], 、固定宽度和深度的人工电路 [19], ，或寄存器大小的可扩展性 [20, 21]。当前量子计算硬件上的热门算法包括量子化学计算[4], ，这也启发了相关基准测试 [22], ，其中就包括我们在此提出并应用于 IBM Quantum System One 机器上的基准测试。

我们的工作安排如下：首先定义基准，介绍一种标准的测量误差缓解技术，并在第 II 节介绍位于德国埃宁根的 IBM Quantum System One。在第 IIIA 节中，我们将介绍数据收集，并在第 III B 节对其进行整体分析。在第 IIIC 节中，我们将解析基准测试结果随时间的变化，并揭示意想不到的动态特征。在第 IV 节中，我们将总结有关该机器噪声统计数据的结论。

* andreas.woitzik@physik.uni-freiburg.de

II. 方法论与术语

A. $|W\rangle$-基准

我们的基准围绕着制备一个由三个量子比特（即二能级系统）构成的 $|W\rangle$态 [23] 展开：

$$|W\rangle = \frac{1}{\sqrt{3}}(|001\rangle + |010\rangle + |100\rangle), \tag{1}$$

其中， $|0\rangle$ 和 $|1\rangle$ 分别是泡利 Z 算符的 $+1$ 和 -1 本征值对应的本征态。虽然通常研究该态是为了探究其（多体）纠缠性质 [24, 25], ，但在此基准中，它是具有周期性边界条件、跃迁项为 t 且位点相互作用为零的一维三量子比特无自旋费米-哈伯德模型的一个本征态。该模型的哈密顿量在二次量子化表象下写作

$$H_{FH} = -t \sum_{\langle i,j \rangle} \left(\hat{c}_i^\dagger \hat{c}_j + \hat{c}_j^\dagger \hat{c}_i \right), \tag{2}$$

其中，求和遍历所有相邻位点，$\hat{c}_i, \hat{c}_i^\dagger$ 分别是费米子湮灭与产生算符。下文中，所有能量均以 t为单位。

经过Jordan-Wigner变换后， [26], 自旋无关的费米-哈伯德哈密顿量 H_{FH} 对于 n 个格点可表示为

$$H_{FH} = -\frac{1}{2}\left(Y_1 \tilde{Z} Y_n + X_1 \tilde{Z} X_n + \sum_{j=1}^{n-1} Y_j Y_{j+1} + X_j X_{j+1} \right), \tag{3}$$

其中， X_j， Y_j， Z_j是第 j 个子空间上的泡利算符（也称为σ_x， σ_y， σ_z）（并隐式地扩展为在所有其他子空间上作用的恒等算符 $\mathbb{I}_{m\neq j}$ ）。为方便记号，我们令 $\tilde{Z} = \prod_{j=2}^{n-1} Z_j$。我们将泡利算符与恒等算符的（张量）积称为泡利串。

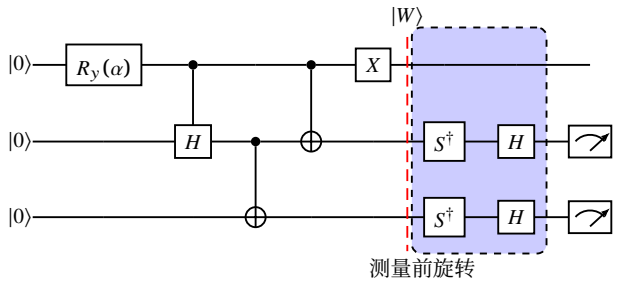

图1：用于制备 $|W\rangle$ 式（1）中的态（左侧，直到红色虚线处），以及确定泡利串 $\langle W|Y_2Y_3|W\rangle$ 的期望值所需的测量前旋转（右侧，位于虚线框内）。门 R_y（α）描述了绕布洛赫球上相关量子比特的 α 轴旋转 y，在本例中为 $\alpha = 2\arccos(1/\sqrt{3})$。_对于测量前旋转，$S$ 描述了绕单量子比特布洛赫球 $\pi/2$ 轴旋转，并且 z 是阿达马门。由此可得 H。$Y_j = SHZ_jHS^\dagger$。

具体而言，对于本文研究的三量子比特情形，所得问题即所谓的费米子三角形 [27]，其哈密顿量为

$$H_{FT} = -\frac{1}{2}(Y_1Z_2Y_3+X_1Z_2X_3+Y_1Y_2+X_1X_2+Y_2Y_3+X_2X_3). \quad (4)$$

直接计算表明，$|W\rangle$ 是 H_{FT}的基态，其本征值为

$$E = \langle W|H_{FT}|W\rangle = -2. \quad (5)$$

我们提出的基准包括制备量子比特寄存器：从初始态 $|000\rangle$开始，制备为 $|W\rangle$态，然后根据目标函数进行测量$\langle W|H_{FT}|W\rangle$，即其能量期望值。在完全隔离且受控的量子比特寄存器中，我们预期得到–2，如式（5）所示。然而，在量子硬件上实际执行时，每个门都会产生误差。单个量子门误差对能量的影响会在制备和测量过程中不断累积。因此，目标函数相对于理论期望值的偏差将反映制备和测量中的不完善之处 [28]。

我们使用图 $|W\rangle$ 所示的量子电路制备该态。图 1。制备电路需要经过转译，即转换为一系列（通常更长的）门，这些门可在相应的量子硬件上物理实现 [29]。对于测量，要计算式（4）中泡利串的期望值，其中涉及 X_j 或 Y_j，需要额外的（测量前）门，将状态旋转到设备的测量基底（通常由 Z_j的本征基定义）。图 1展示了计算期望值 $\langle Y_2Y_3\rangle$所需的测量前门。

B. 测量误差缓解

“误差缓解”是指这样一类技术：假设真实量子硬件的实际输出数据受到可用典型模型描述的系统误差影响，通过这些技术可以从算法执行结果中推断出正确结果。通常，误差缓解技术会对原始输出数据进行后处理 [30]，；这与真正的量子纠错不同，后者使用多个物理量子比特编码一个逻辑量子比特 [31]。误差缓解技术既可以针对量子门误差 [32]，，也可以针对初始化或测量误差 [33]。由于后者在小型系统中占主导地位 [34]，，我们仅讨论测量误差缓解的影响。对于更大的系统，量子门误差对系统中累积误差的贡献可能同样显著，甚至占主导地位。

在我们的分析中，我们使用 Qiskit（IBM 为其量子设备开发的软件开发套件）中实现的测量误差缓解流程 [35]，该流程基于量子探测器层析术 [36]。在此过程中，所有计算基态 $|\psi_j\rangle$都会在量子设备上制备并测量，从而得到一个重叠矩阵 $\Lambda_{ij} = \langle\psi_i|\psi_j\rangle$，用于表征测量过程（探测器）。在不存在测量误差的情况下，Λ 是单位矩阵；否则，它会包含非对角元素。如果 Λ 可逆，则可将其逆矩阵作用于设备得到的结果直方图$\mathbf{p}^{\text{exp}}$，以恢复理想的（准）分布$\mathbf{p}_{\text{ideal}} = \Lambda^{-1}\mathbf{p}^{\text{exp}}$。不可避免的是，在此流程中，完全确定一个稠密 Λ 所需的采样操作次数（例如计算期望值）会随寄存器大小呈指数增长，这是误差缓解技术的一个基本限制，相关讨论见例如参考文献 [37]。

C. 平台

下文中，我们将使用字符串 ibmq_ehningen 指代位于德国埃宁根的 IBM Quantum System One 设备。该设备属于 IBM 的 Falcon 系列，拥有 27 个量子比特，搭载变体 r5.11[38]。可通过网页接口与该设备通信，计算则由调度器管理，并在延迟时段运行。

而字符串 ibmq_qasm_simulator 则指代一种虚构设备（由 IBM 的 Qiskit 提供），它以量子电路形式模拟量子计算，但具有完美的读出和量子门操作。

III. 数值结果

A. 数据收集

现在，我们来定义描述数据收集过程的术语。首先，目标函数在式（5）中的一种实现，是将哈密顿量中每个泡利串各自的期望值相加得到的（即，

式（4）中的六个串）。每个期望值都通过一次实验确定。一次实验包含 1024 次测量发射，每次测量发射都包括初始化初始三量子比特态 $|000\rangle$，在所选平台（模拟器或量子硬件）上执行图1所示酉算符的转译版本，进行测量前旋转，并在设备的测量基上进行测量。每次测量发射都会产生一个三位比特串，用于描述对 $Z_1Z_2Z_3$ 的测量结果；该结果对应于量子电路产生的、经过制备和旋转（在测量前阶段进行，见图1）的态。因此，一次实验会收集结果统计的直方图，并根据量子力学态模型的规定，利用该直方图计算给定泡利串的期望值。随后，将式（4）中各泡利串对应的期望值相加，得到由一次实现确定的能量期望值（5）。

在每个作业中，我们会连续数小时，在一个或多个量子比特三元组上反复评估目标函数 $\langle W|H_{FH}|W\rangle$。这些三元组是随机选取的三个相连量子比特的子集（请注意，同一个三元组可能会在不同作业中接受探测）。我们将某个作业中一个三元组随时间变化的数据称为一个时间序列。每当生成一批新的结果时，我们都会存储由 50 次实现组成的数据包、相应的各个泡利串的期望值、时间戳，以及机器接口提供的设备属性。我们还关注读出误差和门误差；这些数据仅在校准 [39]时更新。由于作业期间不会进行校准，因此这些数值在每个时间序列中均保持不变。平均错误率 $\epsilon = 1 - F$（其中 F 是设备接口报告的平均门保真度 [40, 41], ）约为：单量子比特门为 10^{-4}，双量子比特门为 10^{-2}，测量为 10^{-2}。

所有数据采集于2022年二月、三月以及九月至十二月，共计72个时间序列中的124950次实现。在下文中，我们首先讨论所有实现次数的分布，然后讨论其随时间的波动。

B. 整个数据集的统计数据

在图 2（a）中，我们比较了在ibmq_ehningen（蓝色）和ibmq_qasm_simulator （绿色）上采集的实现次数直方图。ibmq_qasm_simulator 的直方图高度对称，并在 −2处达到峰值。00，标准差为0。03。由于该平台模拟的是无误差门，因此该直方图的宽度源于使用有限数量的测量发射估计每个泡利串的2/3 期望值时不可避免的舍入误差。这些微小偏差在对各期望值求和以获得预期能量时会进一步传播，从而可能导致结果低于 −2，但并不真正违反变分原理。

ibmq_ehningen 的直方图峰值约为 −1。83，且明显偏斜，尾部延伸至高达 −1的数值。与ibmq_qasm_simulator 的偏差同样可以追溯到各个泡利串的统计数据。如表I 所示，在从一个示例作业中选取的三量子比特组中，与正（负）号相加的比特串出现率，除两种例外外，均一致地低于（高于）精确率。这些系统性偏移意味着，每个泡利串的期望值总是达不到2/3，因此，与ibmq_qasm_simulator相比，采样能量的分布向较高值偏移（这是由于哈密顿量（4）的整体负号所致）。此外，当所有比特字符串的出现频率相同时，即当寄存器处于（或退相干至）最大混合态时，采样能量趋向于0。

现在，我们分别尝试厘清门错误和读出错误的影响。为模拟读出错误（“0”被误读为“1”，反之亦然），我们考虑任意一个 n 量子比特 [6]发生比特翻转 X 的概率 p：

$$\Lambda = \bigotimes_{j=1}^{n} \left[(1-p)\mathbb{I}_j + pX_j\right] = (1-np)\mathbb{I} + p\sum_{j=1}^{n} X_j + O(p^2), \tag{6}$$

其中，只要 p，我们就使用了 $(1-p)^n \approx (1-np)$，而恒等算符 $\mathbb{I}$ 作用于所有量子比特。在这个简单模型中，我们可以通过对矩阵 Λ 的对角元素取平均来估计 $1-np$（参见 II B）；这些元素在每组 50 次实现之前及每组实现中均进行估计。对于表 I 中呈现的统计数据，我们得到 $p \approx 1.1\%$，这与报告的错误率一致。在图 2(c) 中可以看到，这一估计与总读出误差的分布相符，即各量子比特读出误差之和，其大小与 np相当。如果我们将 np 视为泡利串期望值上的误差（以标准差衡量），并将这些期望值视为独立同分布的随机变量，则得到传播读出误差 $3 \times 1.1\% \times \sqrt{6} \approx 8\%$；这与通过测量误差缓解得到的图 2(b) 直方图偏移相符。鉴于上述统计假设以及任何误差缓解方法所依据的假设，这一数值应视为粗略估计。

总门误差，如图2（d）所示，是一次实现中所有相关门误差的总和，即每个实验的制备和（预）测量阶段所产生的误差之和，见图1。一般而言，该数值主要由受控门（双量子比特门）的误差决定，这些门出现在量子电路的制备阶段。因此，我们假设总门误差决定了图2（a、b）中直方图的宽度。这一结论受到以下观察的启发：图2（b）中对读出误差进行缓解会使直方图发生显著偏移，但几乎不会影响其宽度或整体形状，尤其是在向较高数值延伸的尾部。

根据图 2(c,d)，很难确立总读出误差或门误差与基准测试结果之间明确的定量关系。不过，值得注意的是，根据接口设计，在持续数小时的作业期间，设备返回的这些量化指标始终为固定值。换言之，在两次校准之间，设备的状态在形式上被视为 静态 状态。因此，在下一节中，我们将通过分析目标函数的值是否随时间变化以及如何变化，来检验这一假设。

比特串\泡利串	Y_1Y_2	X_1X_2	Y_2Y_3	X_2X_3	精确	$Y_1Z_2Y_3$	$X_1Z_2X_3$	精确	E
+000	0.41(2)	0.41(2)	0.41(2)	0.41(2)	5/12	0.32(1)	0.31(1)	1/3	
−001			0.091(9)	0.099(9)	1/12	0.022(5)	0.019(5)	0	
−010	0.095(9)	0.095(9)	0.093(9)	0.10(1)	1/12	0.088(9)	0.095(9)	1/12	
+011			0.40(2)	0.39(2)	5/12	0.080(9)	0.081(8)	1/12	
−100	0.089(9)	0.10(1)			1/12	0.015(4)	0.015(4)	0	
+101					0	0.31(1)	0.32(1)	1/3	
+110	0.40(2)	0.39(1)			5/12	0.079(8)	0.091(9)	1/12	
−111					0	0.087(9)	0.066(8)	1/12	
$\langle\cdots\rangle$	0.63(3)	0.60(3)	0.63(3)	0.60(3)	2/3	0.58(3)	0.61(3)	2/3	−1.83(3)

表 I： `ibmq_ehningen`中量子比特 11、14 和 13 的结果统计（作业于 2022-09-19 22:47:16 开始，持续 7.5 小时）。数据包含 31 个数据包，每个数据包有 50 个采样能量 $E = \langle H_{FT}\rangle$，这些能量来自公式 (4)。第一列列出图 1中量子电路的可能测量结果；符号（比特串中“1”的个数为偶数时为正，为奇数时为负）表示计算期望值时各项相加所需的符号 $\langle\cdots\rangle$ 见最后一行。中间各列报告了顶行所列每个泡利串对应的各种可能测量结果的出现频率（括号内为标准差）。在分别涉及 2 个和 3 个量子比特的泡利串右侧，给出了量子力学预测的精确概率。表格右下角计算了能量的期望值。对于每种可能的测量结果和每个泡利串，数值及其不确定度分别是从收集到的直方图中提取的平均值和标准差。

C. 时间分辨特征

如上一节所述，每次测量发射的结果本质上具有随机性。因此，我们预期会观察到、实际上大多数情况下也确实观察到的基线行为，是与图 2(a,b) 中直方图相符的波动。然而，沿时间轴解析数据，也揭示了我们在上述整体分析中无法观察到的意外现象。在图 3 中，我们绘制了能量密度直方图（纵坐标）随记录时间戳（横坐标）的变化。可以辨别出：

1. 振荡。 大多数时间序列都可以定性地观察到平均能量的周期性振荡，其振幅小于密度直方图的典型宽度（例如见图 3(b)）。然而，相当一部分时间序列（在分析的72个时间序列中有22个）呈现出振幅大得多的振荡。图 3(a) 展示了一个例子，我们使用 SciPy 的 [43] 最小二乘法拟合了一个形式为$y_0+A\sin[t/(2\pi T)+\phi_0]$的模型，并得出周期为 $T \approx 2$ 小时。这些振荡在2022年春季和秋季均有观测到，且与量子比特分配无关。定性估算发现，所有受影响时间序列的周期均为1至 2小时。在少数情况下，信号还呈现出额外的正斜率或负斜率，表明可能存在时间尺度更长的振荡。

2. 离群值。 在若干情况下（72 次中有 21 次），我们观察到这样的时间序列：每包 50 个样本会产生异常偏高的目标函数值（相较于长时间范围内的整体行为）。图 3(b) 展示了一个例子。特别是，我们在以下作业中观察到了这一现象：我们按顺序从不同的三量子比特组收集数据（采用轮换方案）；对于具有一个或更多共同量子比特的三元组，离群值会出现在相同的时间区间，但其他（不相关的）三元组则不会出现。这一观察表明，在作业期间，某些事件影响了一个或多个量子比特，但不一定影响整个设备。值得注意的是，此类事件会影响整包数据，使其密度直方图偏向较高的能量值，但不会使其分布变宽。

3. 延迟。少数时间序列（共影响 72 个受测寄存器中的 33 个）显示出时间分辨率不一致，例如，时间序列开始时的采样较为稀疏 [图3(c)], ，或在末尾 [图 3(b)]。我们无法确定这一现象是由量子设备响应缓慢、其周边硬件响应缓慢，还是其他原因造成的。如上所述，我们假设在任何作业运行期间都不会进行校准 [39]。

4. 结果恒定。2022年春季的数日间，有少数作业对每组三个受测量子比特都产生了完全恒定的结果。图 3(d) 展示了一个示例。这种情况几乎可以肯定是 IBM 一侧的软件错误所致，此后再未发生。不过，由于设备未报告任何异常，如果我们只是将统计数据汇总到直方图中，就不会发现这一错误。这些作业的数据未纳入任何统计分析。

这些随时间变化的特征与量子设备接口提供的名义信息不一致。如上文所述，大多数离群值似乎是由影响至少一个量子比特的事件造成的；如果这些事件被确认为外部事故 [44], ，这或许足以成为在统计分析中排除离群值的理由。

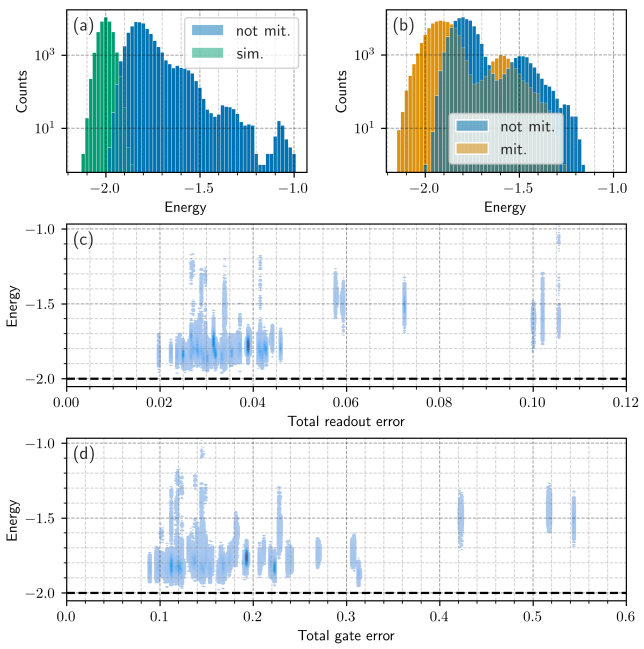

图 2：（a）比较二月和三月 2022年在 ibmq_ehningen 上采样的未经缓解能量直方图（蓝色，“not mit.”）与在 ibmq_qasm_simulator 上采样的 39750 个能量直方图（绿色，“sim.”）。（b）比较 2022年九月至十二月期间在 ibmq_ehningen 上采样的能量直方图（蓝色，“not mit.”）与相应的经缓解值（橙色，“mit.”）。（c、d）

在ibmq_ehningen上测得的采样能量密度图，横坐标分别按照以下因素进行区分：(c) 总读出误差和 (d) 总门误差。对于每个作业，这些误差均通过将在计算时由设备接口报告的相应误差求和来计算。位于 –2 处的虚线表示由公式（5）得出的预期值。颜色越深（越浅），表示记录到具有相应总门误差或读出误差的给定能量样本的集中度越高（越低）。改编自 [42]。

D. 测量误差的缓解

为辨别第 III. C 节中观察到的动力学是否由读出误差引起，我们比较（测量误差）缓解后的数据与未经缓解的数据。

当我们仅在作业开始时计算一次矩阵 Λ （参见 IIB），然后使用该矩阵对整个时间序列进行后处理时，我们观察到（图 4）采样能量系统性地向较低值偏移，这与图 2（b）中直方图的偏移相符。虽然这一结果在预期之中，但它假定每个作业期间设备属性保持静态；这一假定与观察到的周期性振荡和离群值相矛盾，而这些现象在缓解处理后仍按预期持续存在。

为减轻随时间变化的误差，我们因此在每组 50 个样本之前，使用 Λ 矩阵对设备进行表征。这种“动态”缓解方法虽说代价高昂，但对于我们的三量子比特测试案例而言是可行的。然而，与上述“静态”缓解方法相比，差异微乎其微（见图4）：振荡和离群值都未受影响，不过由于缓解后的能量直方图预计会变宽，前者可能不那么明显 [45]。

由于通过 Λ 矩阵，我们可以针对每 50 个能量样本的数据包获得设备的快照，因此我们在图 4(b) 中绘出了图 p 所示作业的比特翻转概率随时间变化的估计值4(a)。正如第 IIIB 节所述，该量可作为每个量子比特平均读出测量误差的替代指标。可以观察到若干特征：(i) 设备报告的平均读出误差始终略低于估计的 p；(ii) 在时间戳 03:00 之后，p 也呈现出振荡，不过其时间尺度较短（与图 3(a) 相比），约为 1 小时；(iii) 在 01:30 左右出现了一个意外的尖峰，导致对相应直方图的缓解幅度异常大；(iv) p 随时间变化的特征并未显示出 03:00 和 06:00 左右离群值的迹象，因此对这些离群值的缓解程度与其他时间分箱相当。

总而言之，读出误差的静态和动态缓解都会使采样能量整体偏移约 5%–8%，但无法校正图 3. 中观察到的时间依赖特征。

IV. 结论

本文提出了一种受量子化学启发的简单能量估算基准测试，并将其应用于位于德国埃宁根的 IBM Quantum System One。根据基准测试结果，我们得出以下观察结论：

1. 与仅受量子力学噪声影响的预期值相比，采样能量（即基准测试结果）的直方图发生了偏移且分布更宽 [图 2(a)]；然而，这一趋势似乎与报告的总读出误差 [图 2(c)] 或量子门误差 [图 2(d)] 仅有微弱相关性。
2. 尽管设备的各项属性在整个作业期间名义上保持不变，基准测试结果仍显示出超出典型随机波动的动态特征，例如周期约为 2 小时的振荡 [图 3(a)] 和离群值 [图 3(b)]。3. 测量误差缓解缩小了采样能量与预期能量之间的偏差，但并未减小直方图的离散程度 [图 2(b)]，也未消除其时间依赖性 [图 4(a)]。

4. 动态缓解，即基于 Λ 针对每个数据采样包单独估算的矩阵，无法校正所观察到的现象 [图 4(a)],因此，这些现象不可能仅由测量误差造成。事实上，平均读出错误率本身也呈现出意料之外的时间依赖特征，且似乎与图 3 中观察到的动态无关。

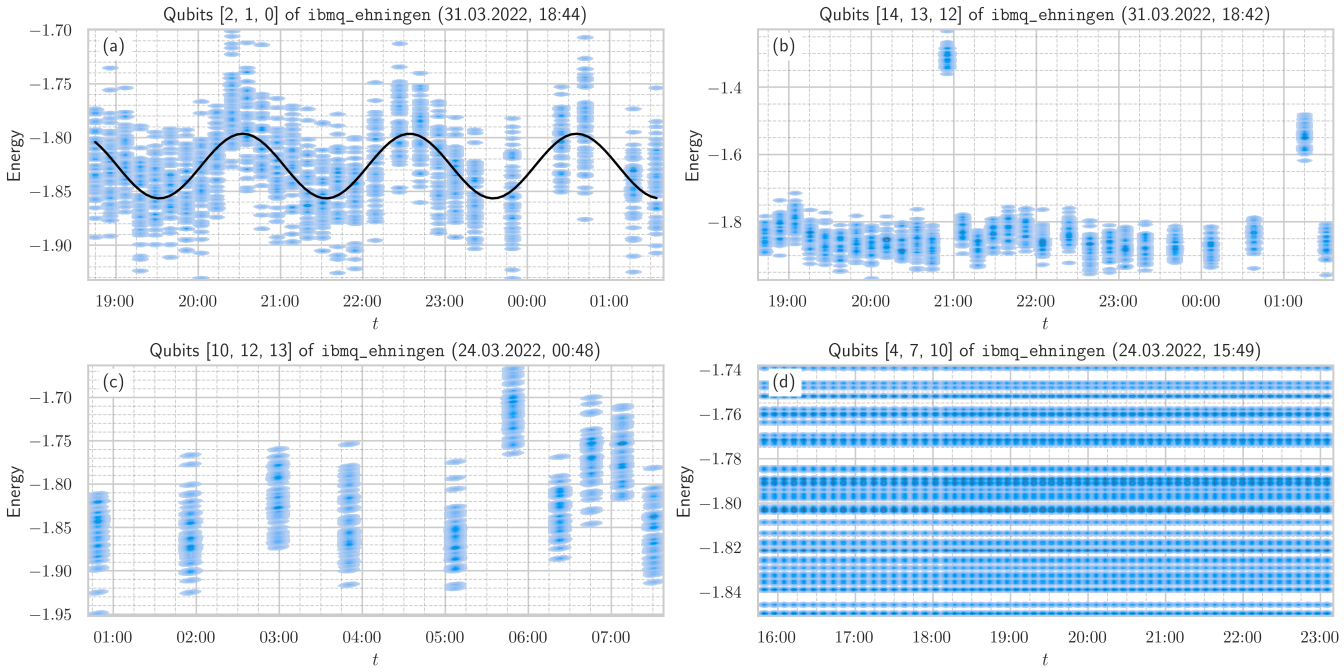

图 3： (5) 所采样基态能量随时间的变化，呈现出不同现象：(a) 周期约为两小时 [121。8(9) 分钟的振荡（如图中示例所示）],(b) 离群值，(c) 延迟以及 (d) 恒定结果。(a) 中的黑线通过对数据进行正弦拟合以作参考。图中所示数据均未经缓解。改编自 [42]。

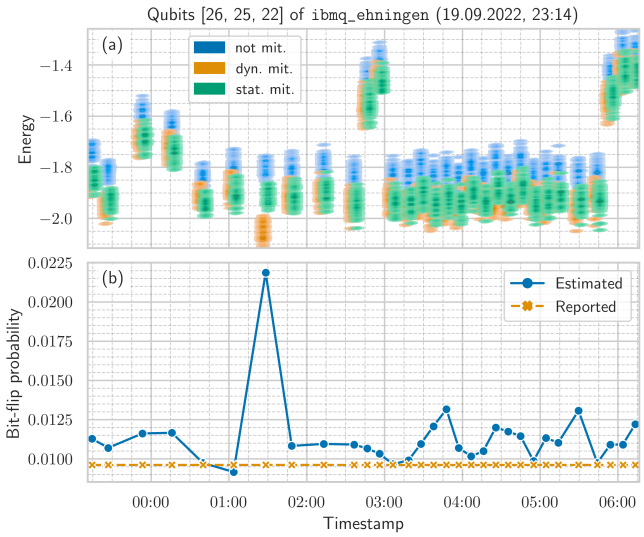

图 4：(a) 随时间以数据包形式采样得到的能量密度直方图；有关数据采集的详情，请参阅正文。原始数据以蓝色表示，静态缓解数据（每个作业仅计算一次校准矩阵）以绿色表示（为便于辨认，略向右移），动态缓解数据（每个数据包都计算校准矩阵）以橙色表示（略向左移）。(b) 蓝色表示根据缓解矩阵估算的比特翻转概率p，如公式 (6) 所示；橙色表示机器报告的每个量子比特的平均比特翻转错误。

仅凭观察结果 3 和 4 即可得出结论：当前范式假设校准之间的机器属性保持静态，这不足以确定机器输出的真实不确定性。必须仔细考量错误率的动态变化，以及错误在不同电路和实验中的累积情况。尤其是，需要通过收集多个时间点的结果来厘清意料之外的时间依赖性。遵循实验物理学实践，收集适当的统计数据，在使用（含噪）量子设备时尚未成为成熟做法，这无疑是因为其成本高昂。然而，如果信息编码所依托的物理平台是含噪或无序系统，就必须在合理的时间序列内记录系综统计，才能对可控可靠性作出定量论断。

致谢

我们感谢 Daniel Urban、Andreas Ketterer 和 Thomas Wellens 分享他们对 IBM QuantumSystem One 的见解。

我们感谢 Dominik Lentrodt 对稿件提出宝贵意见。

A. W. 感谢康拉德·阿登纳基金会提供经费支持。E. G. C. 感谢 Georg H. Endress基金会以及“SiQuRe”项目（巴登-符腾堡州量子计算能力中心）提供的支持。

由巴登-符腾堡州经济、劳动和旅游部资助。

[1] R. Landauer，《今日物理》44，23（1991）。[2] R. P. Feynman，《国际理论物理学杂志》21，467（1982）。[3] K. Bharti, A. Cervera-Lierta, T. H. Kyaw, T. Haug, S. Alperin-Lea 等，《Reviews of Modern Physics》94，015004（2022）。[4] Y. Cao, J. Romero, J. P. Olson, M. Degroote, P. D. Johnson 等，《化学评论》119，10856（2019）。[5] R. Orús、S. Mugel 和 E. Lizaso，《物理评论》4，100028（2019）。[6] M. A. Nielsen 和 I. L. Chuang，《量子计算与量子信息》（剑桥大学出版社，2012）。[7] A. Montanaro，npj Quantum I nformation 2，1（2016）。[8] L. K. Grover，载于第二十八届美国计算机协会计算理论年度研讨会论文集（1996），第212–219页。[9] N. P. de Leon, K. M. Itoh, D. Kim, K. K. Mehta, T. E. Northup 等，《科学》372，eabb2823（2021）。[10] M. Kjaergaard, M. E. Schwartz, J. Braumüller, P. Krantz, J. I.-J. Wang 等，《凝聚态物理年度评论》11，369（2020）。[11] R. Blatt 和 C. F. Roos，《自然·物理》8，277（2012）。[12] C. D. Bruzewicz, J. Chiaverini, R. McConnell 和 J. M. Sage，《应用物理评论》6，021314（2019）。[13] A. Callison 和 N. Chancellor，Physical Review A 106，010101（2022）。[14] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo 等，《自然综述：物理》3，625（2021）。[15] K. Huppler，载于性能评估与基准测试技术会议（施普林格，2009），第18–30页。[16] E. Knill, R. Laflamme, R. Martinez 和 C.-H. Tseng，《七比特量子计算机上的猫态基准》（1999），arXiv:quant-ph/9908051。[17] M. Amico, Z. H. Saleem 和 M. Kumph，Physical Review A 100，012305（2019）。[18] K. Michielsen, M. Nocon, D. Willsch, F. Jin, T. Lippert 和 H. De Raedt，《计算物理通讯》220，44（2017）。[19] R. Blume-Kohout 和 K. C. Young，Quantum 4，362（2020）。[20] E. Magesan, J. M. Gambetta 和 J. Emerson，Physical Review A 85，042311（2012）。[21] A. Cornelissen, J. Bausch 和 A. Gilyén，《基于门的量子计算机的可扩展基准测试》（ 2021 2104 10698 ），[ ] arXiv: . quant-ph . [22] A. J. McCaskey, Z. P. Parks, J. Jakowski, S. V. Moore, T. D. Morris 等，npj Quantum Information 5, 99 (2019). [23] W. Dür、G. Vidal 和 J. I. Cirac，Physical Review A 62, 062314 (2000). [24] F. Mintert、A. R. Carvalho、M. Kuś 和 A. Buchleitner，Physics Reports 415, 207 (2005). [25] O. Gühne 和 G. Tóth，Physics Reports 474, 1 (2009). [26] P. Jordan 和 E. Wigner，Zeitschrift für Physik 47, 631 (1928). [27] A. J. C. Woitzik、P. K. Barkoutsos、F. Wudarski、A. Buchleitner 和 I. Tavernelli，Physical Review A 102, 042402 (2020). [28] J. Brugger、C. Seidel、M. Streif、F. A. Wudarski、C. Dittel 和 A. Buchleitner，Physical Review A 105, 042605 (2022). [29] IBM Quantum，转译器 API [计算机软件] (2022). [30] Z. Cai、R. Babbush、S. C. Benjamin、S. Endo、W. J. Huggins 等，arXiv:2210.00921 (2022). [31] B. M. Terhal，Reviews of Modern Physics 87, 307 (2015). [32] A. He、B. Nachman、W. A. de Jong 和 C. W. Bauer，Physical Review A 102, 012426 (2020). [33] H. Kwon 和 J. Bae，IEEE Transactions on Computers 70, 1401 (2020). [34] P. D. Nation、H. Kang、N. Sundaresan 和 J. M. Gambetta，PRX Quantum 2, 040326 (2021). [35] M. S. ANIS、Abby-Mitchell、H. Abraham、AduOffei、R. Agar- wal、G. Agliardi 等，Qiskit：用于量子计算的开源框架 (2021). [36] F. B. Maciejewski、Z. Zimborás 和 M. Oszmaniec，Quantum 4, 257 (2020). [37] R. Takagi、S. Endo、S. Minagawa 和 M. Gu，npj Quantum Information 8, 114 (2022). [38] IBM Quantum，IBM 量子处理器类型。[39] IBM 文档——关于校准作业 (2023). [40] J. Emerson、R. Alicki 和 K. Życzkowski，Journal of Optics B: Quantum and Semiclassical Optics 7, S347 (2005). [41] E. Magesan、J. M. Gambetta 和 J. Emerson，Physical Review A 85, 042311 (2012). [42] L. Hoffmann，《含噪中等规模量子设备上的 VQE—— 使用混合量子算法对含噪量子硬件进行基准测试》，学士学位论文，弗赖堡阿尔伯特-路德维希大学 (2022). [43] P. Virtanen、R. Gommers、T. E. Oliphant、M. Haberland、T. Reddy 等，Nature Methods 17, 261 (2020). [44] J. M. Martinis，npj Quantum Information 7, 90 (2021). [45] K. Temme、S. Bravyi 和 J. M. Gambetta，Physical Review Letters 119, 180509 (2017).