



中华人民共和国国家标准

GB/T XXXXX—XXXX

量子计算服务平台 第2部分：性能评估

Quantum computing service platform—
Part 2: Performance evaluation

（征求意见稿）

（本草案完成时间：2025 年 7 月 18 日）

在提交反馈意见时，请将您知道的相关专利连同支持性文件一并附上。

XXXX—XX—XX 发布

XXXX—XX—XX 实施

国家市场监督管理总局
国家标准化管理委员会 发布

目 次

前言 II

引言 III

1 范围 1

2 规范性引用文件 1

3 术语和定义 1

4 缩略语 1

5 性能评估体系描述 2

6 量子计算机系统性能评估 3

 6.1 量子比特数 3

 6.2 连通性 3

 6.3 比特相干性 3

 6.4 操作保真度 3

 6.5 操作时长 4

7 量子资源经典模拟器性能评估 4

 7.1 经典模拟器数量 4

 7.2 比特规模 4

 7.3 最大线路深度 5

 7.4 量子线路模拟效率 5

8 服务性能评估 5

 8.1 任务响应时间 5

 8.2 任务并发度 5

 8.3 负载均衡恢复时间 6

 8.4 可用服务数量 6

 8.5 吞吐量 6

 8.6 新建连接速率 7

参考文献 8

前 言

本文件按照GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

本文件是GB/T XXXXX《量子计算服务平台》的第2部分。GB/T XXXXX已经发布了以下部分：

——第1部分：架构与功能要求

——第2部分：性能评估

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别这些专利的责任。

本文件由全国量子计算与测量标准化技术委员会（SAC/TC 578）提出并归口。

本文件起草单位：

本文件主要起草人：

引 言

量子计算服务平台作为一种在线提供量子计算资源和工具的平台,通过云计算技术将量子算力以更灵活、更便捷的方式提供给用户。目前,国内外已有许多企业和研究机构推出了量子计算服务平台,以此推动量子计算产业落地和生态建设,已成为业界共识。制定量子计算服务平台相关国家标准,在量子计算服务平台的架构和功能要求基础上,为量子计算服务平台提供性能评估指标以及必要的评估方法,可以进一步推动国内量子计算服务平台功能完善和性能提升,形成推进产业发展合力。

GB/T XXXXX《量子计算服务平台》旨在规定量子计算服务平台的架构、功能要求、技术参数、测试方法、可靠性试验、检验规则等要求,依照系统主要功能模块划分,标准目前拟由两个部分构成。

——第1部分:架构和功能要求。规定量子计算服务平台的架构和功能要求等要求。

——第2部分:性能评估。规定量子计算服务平台性能评估体系、评价指标和测试方法等要求。

随着量子计算技术和服务平台技术与产业的发展,未来可能还会包含更多部分。

量子计算服务平台 第2部分：性能评估

1 范围

本文件描述了量子计算服务平台的性能评估体系及评估方法，包括量子计算机系统性能、量子资源经典模拟器性能、平台服务性能等评估指标。

本文件适用于量子计算服务平台性能评估、测试，也可作为量子计算服务平台的研制、生产、实用等提供参考。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T 42565-2023 量子计算 术语和定义

GB/T XXXXX-XXXX 量子计算系统性能测试方法

GB/T XXXXX-XXXX 量子计算服务平台 第1部分：架构与功能要求

3 术语和定义

GB/T 42565-2023和GB/T XXXXX-XXXX（量子计算服务平台 第1部分：架构与功能要求）界定的以及下列术语和定义适用于本文件。

3.1

量子计算 quantum computing

一种基于量子力学原理，利用量子比特进行信息运算和存储，并运行量子算法的新型计算。

[来源：GB/T 42565-2023，3.12，有修改]

3.2

量子计算服务平台 quantum computing service platform

实现量子计算资源按需提供服务化交付的软硬件集合，由量子计算机硬件资源、量子计算服务、资源管理与调度功能、以及辅助的经典信息通信基础设施（包括计算、存储、通信网络等）等构成。

[来源：GB/T 42565-2023，6.1，有修改]

3.3

量子资源经典模拟器 classical simulators for quantum resources

一种使用经典计算资源对量子计算机行为和功能进行模拟的设备或系统。

3.4

量子计算云服务提供商 quantum computing cloud service provider

提供量子计算服务平台和计算服务的参与方。

4 缩略语

下列缩略语适用于本文件。

CPU：中央处理器（Central Processing Unit）

DDoS：分布式拒绝服务（Distributed Denial of Service）

GPU：图形处理器（Graphics Processing Unit）

HTTP：超文本传输协议（Hypertext Transfer Protocol）

HTTPS：超文本传输安全协议（Hypertext Transfer Protocol Secure）

5 性能评估体系描述

本文件采用量子计算机系统性能、量子资源经典模拟器性能、服务性能等特性表征量子计算服务平台的性能，每个特性包含若干评价指标。本文件用性能评价指标来评估对这些特性的要求，具体性能评价指标见图1所示。性能评估体系描述如下：

- a) 量子计算机系统性能：用于衡量量子计算服务平台所基于的量子计算机系统的特性，包括量子比特数、连通性、比特相干性、操作保真度、操作时长等指标；
- b) 量子资源经典模拟器性能：用于衡量量子计算服务平台所基于的经典模拟器资源的特性，包括经典模拟器数量、比特规模、最大线路深度、量子线路模拟效率等指标；
- c) 服务性能：用于衡量量子计算服务平台能够提供的云计算服务的特性，包括任务响应时间、任务并发度、负载均衡恢复时间、可用服务数量、吞吐量、新建连接速率等指标。

注1：本文件考虑云服务的底层硬件支持、真实物理实现，评价指标涉及量子计算云服务提供商、量子计算云服务客户以及第三方监管机构。本评价体系中只考虑量子计算机系统的通用性指标，对于超导、离子阱、光量子、中性原子等体系的特殊指标不包含在本评价体系中。

注2：由于考虑了底层硬件的规范，本评价体系中的细则涉及了较多专业的测试方法。具体测试方法见GB/T XXXXX-XXXX（量子计算系统性能测试方法）。

注3：本文件中暂不考虑非门线路模型量子计算机系统的特殊指标。

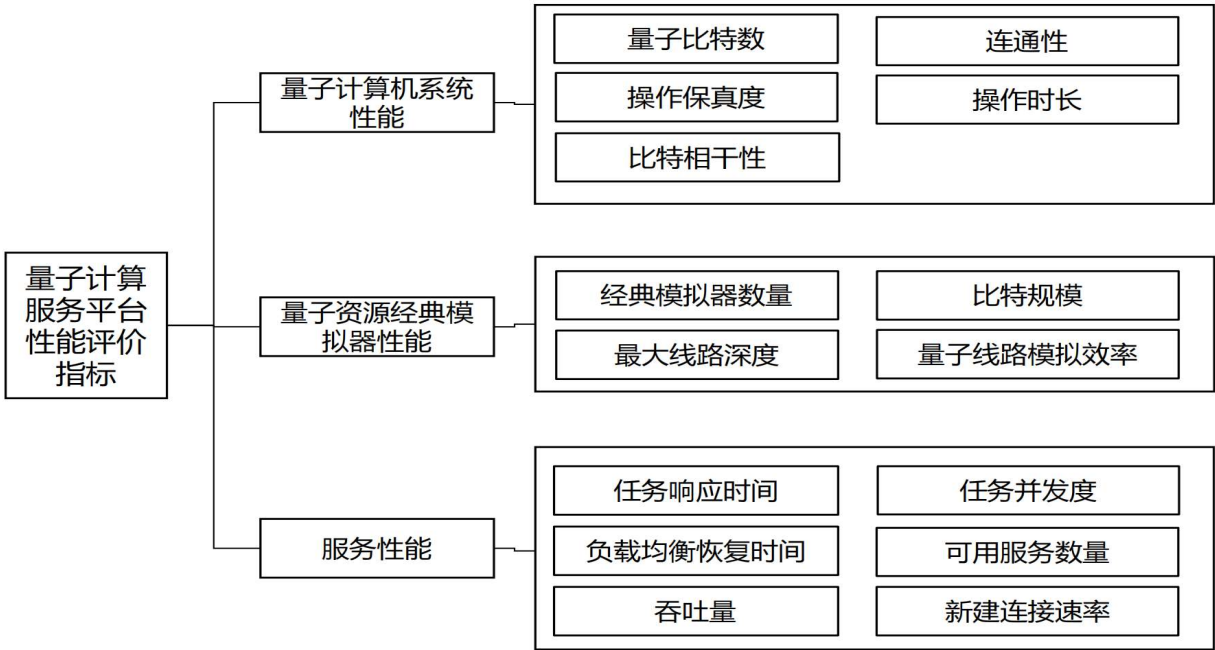


图1 量子计算服务平台性能评价指标

6 量子计算机系统性能评估

6.1 量子比特数

- a) 定义描述：指一个量子处理器包含的物理量子比特总数量。
 - b) 评估方法：具体测试方法见《GB/T XXXXX-XXXX 量子计算系统性能测试方法》中 6.1.1 节。
- 注：量子比特可分为数据量子比特和辅助量子比特。数据量子比特：用来编码量子信息的量子比特；辅助量子比特：为辅助数据量子比特完成计算任务添加的量子比特。

6.2 连通性

- a) 定义描述：量子比特的连通性指单个量子比特能够直接产生纠缠的量子比特数量；单个量子处理器的连通性指统计所有量子比特的连通性后通过平均值进行表征。
- b) 评估方法：具体测试方法见《GB/T XXXXX-XXXX 量子计算系统性能测试方法》中 6.1.2 节。

6.3 比特相干性

6.3.1 纵向弛豫时间

- a) 定义描述：表征量子比特从激发态（高能态）自发弛豫到基态（低能态）所需的时间尺度，反映量子比特与环境之间能量耗散的速率，也称为能量弛豫时间，通常表示为 T_1 。
- b) 评估方法：具体测试方法见《GB/T XXXXX-XXXX 量子计算系统性能测试方法》中 6.1.4 节。

6.3.2 横向弛豫时间

- a) 定义描述：表征量子比特保持其量子态相位相关信息的时间尺度，反映量子比特相位相干性随时间的衰减速率，也称为相位弛豫时间，其中包含了一部分能量弛豫的影响，通常表示为 T_2 。
- b) 评估方法：具体测试方法见《GB/T XXXXX-XXXX 量子计算系统性能测试方法》中 6.1.4 节。

6.3.3 纯相位弛豫时间

- a) 定义描述：表征量子比特的状态在横向弛豫的基础上去除能量弛豫影响后的纯相位弛豫特征时间，通常表示为 T_ϕ 。
- b) 评估方法：通过测量得到的纵向弛豫时间 T_1 和横向弛豫时间 T_2 值，利用公式（1）计算得到纯相位弛豫时间 T_ϕ ：

$$\frac{1}{2T_1} + \frac{1}{T_\phi} = \frac{1}{T_2} \dots\dots\dots (1)$$

式中：

- T_1 ——纵向弛豫时间；
- T_2 ——横向弛豫时间；
- T_ϕ ——纯相位弛豫时间。

6.4 操作保真度

6.4.1 读取保真度

- a) 定义描述：用于衡量量子比特的测量结果与量子比特真实状态相符的概率，确定了量子比特的测量准确度。
- b) 评估方法：具体测试方法见《GB/T XXXXX-XXXX 量子计算系统性能测试方法》中 6.2.1 节。

6.4.2 单比特门保真度

- a) 定义描述：用于表征对一个量子比特进行单比特门操作之后，实际操作效果与理论预期效果之间相似程度的度量，是量化单比特门操作好坏的关键指标。
- b) 评估方法：具体测试方法见《GB/T XXXXX-XXXX 量子计算系统性能测试方法》中 6.2.1 节。

6.4.3 两比特门保真度

- a) 定义描述：用于表征对两个量子比特进行两比特门操作之后，实际操作效果与理论预期效果之间相似程度的度量，是量化两比特门操作好坏的关键指标。
- b) 评估方法：具体测试方法见《GB/T XXXXX-XXXX 量子计算系统性能测试方法》中 6.2.1 节。

6.5 操作时长

6.5.1 单比特门操作时长

- a) 定义描述：用于衡量系统中执行单比特门操控所需的时间，并且要求单比特门保真度要达到 98% 以上。这里不限制单比特门的类型。
- b) 评估方法：具体测试方法见《GB/T XXXXX-XXXX 量子计算系统性能测试方法》中 6.2.2 节。

6.5.2 两比特门操作时长

- a) 定义描述：用于衡量系统中执行两比特门操控所需的时间，并且要求两比特门保真度要达到 95% 以上。这里不限制两比特门的类型，建议选择 CZ 门。
- b) 评估方法：具体测试方法见《GB/T XXXXX-XXXX 量子计算系统性能测试方法》中 6.2.2 节。

7 量子资源经典模拟器性能评估

7.1 经典模拟器数量

- a) 定义描述：用于衡量量子计算服务平台能够提供的可并行执行的经典模拟器数目。
- b) 评估方法：在限定模拟计算方式下，评估系统中每一个单个模拟器执行量子模拟时候可支持的最高量子比特数（基本由内存容量上限决定），记作 N_m 。可以将运行的量子比特数目标设定为可并行执行的量子比特数目，为 $N_m - 2$ ，同时开启多个模拟器并让其运行随机线路。此时系统能够支持的最多并行运行个数，没有显示内存报错、排队等候等状况，能够同时运行的模拟器数目则为经典模拟器数量。

示例：一个经典模拟器默认最高的全振幅可模拟量子比特数目为 35，测试时候将并行开启比特数目为 33 个量子比特的全振幅模拟计算，线路为随机采样的线路库。此时系统能够支持的最多并行运行个数，则为经典模拟器数量。理论上可以同时支持 4 个经典模拟器，但实际上考虑到调度系统架构的调度和辅助存储能力等，通常最多支持 2 个经典模拟器。那么在一个量子计算服务平台中部署了若干个经典模拟器，均可以同时进行并行测试，最终系统上同时并行运行的数目，则为最多经典模拟器数目。

7.2 比特规模

- a) 定义描述：用于衡量量子计算服务平台提供的经典模拟器可以接受的最高的量子比特数目任务，同时不支持问题拆分或者子空间拆分的近似求解方式。对于某一种经典模拟器，支持的最大量子比特数越大，表示计算能力越强。

- b) 评估方法：针对量子计算服务平台提供的某一种经典模拟器，构建一个 N 比特和 D 层深度的随机基准测试量子电路；初始化经典模拟器，选择执行后端（CPU 或者 GPU），对量子电路进行模拟和采样；在不同比特数下进行模拟测试；逐渐增加比特数，直至出现内存不足的运行报错，不报错 的最大比特数即为比特规模。

7.3 最大线路深度

- a) 定义描述：用于衡量量子计算服务平台提供的经典模拟器所支持的模拟线路的最大深度。
- b) 评估方法：针对量子计算服务平台提供的某一种经典模拟器，构建一个 N 比特和 D 层深度的随机基准测试量子电路；初始化经典模拟器，选择执行后端（CPU 或者 GPU），对量子电路进行模拟和采样；在不同线路深度下进行模拟测试；逐渐增加线路深度，直至出现内存不足的运行报错或执行时间显著上升以致无法完成模拟任务，不出问题的线路深度值即为最大线路深度。

7.4 量子线路模拟效率

- a) 定义描述：用于衡量量子计算服务平台提供的经典模拟器运行一个随机基准测试线路的运行效率，时间越短表示运行效率越高。
- b) 评估方法：针对量子计算服务平台提供的某一种经典模拟器，构建一个 N 比特和 D 层深度的随机基准测试量子电路；初始化经典模拟器，选择执行后端（CPU 或者 GPU），对量子电路进行模拟和采样；输出量子线路模拟时间，作为量子线路模拟效率。

8 服务性能评估

8.1 任务响应时间

- a) 定义描述：用于衡量量子计算服务平台对计算任务的解析调度和任务执行时间，时间越短表示服务平台性能越强。
- b) 评估方法：量子计算服务平台在对任务请求进行响应之后，对任务进行解析、调度、任务切换和实际执行时间进行记录和反馈。按照公式（2）计算：

$$T_{response} = T_{parse} + T_{schedule} + T_{handoff} + T_{operation} \dots\dots\dots (2)$$

式中：

$T_{response}$ ——任务响应时间；

T_{parse} ——任务解析时间，即平台解析任务的时间；

$T_{schedule}$ ——任务调度时间，即平台为任务分配量子资源和经典资源的时间；

$T_{handoff}$ ——量子计算与经典计算之间的切换时间；

$T_{operation}$ ——量子计算和经典计算任务准备完毕后，实际执行的开始时间。

注： 这里不包括用户提交任务到服务平台之后排队等待的时间。

8.2 任务并发度

- a) 定义描述：用于衡量量子计算服务平台同时处理多个量子计算任务的能力，反映了系统在同一时间内所能承载的最大任务个数。
- b) 评估方法：采用日志分析的方法，收集和分析云平台的运行日志，包括任务请求日志、系统运行日志等，统计单位时间内处理的任务请求数量、任务执行时间等信息。按照公式（3）计算：

$$C = QPS_{peak} \times T \dots\dots\dots (3)$$

式中：

C ——量子计算服务平台任务并发度；

QPS_{peak} ——峰值时间每秒任务请求数；

T ——每个请求平均耗时。

对于新搭建的量子计算服务平台，没有足够的数据支撑用于计算 QPS_{peak} 时，可以根据系统资源量对并发度进行估算，按照公式（4）计算：

$$C = \frac{R_{total}}{R_{avg}} \dots\dots\dots (4)$$

式中：

C ——量子计算服务平台任务并发度；

R_{total} ——量子计算服务平台总资源量；

R_{avg} ——平均每个任务所需资源量。

8.3 负载均衡恢复时间

- a) 定义描述：用于衡量量子计算服务平台在涉及高并发和资源分配的任务中，在节点故障或负载不均匀的情况下，系统恢复到均衡状态所需的时间。
- b) 评估方法：使用云服务平台自带的监控工具，实时监测负载均衡器的运行状态。在故障发生时，这些工具能够及时记录故障发生的时间点以及系统恢复到正常负载均衡状态的时间点，通过计算两者之间的时间差，得到负载均衡恢复时间。按照公式（5）计算：

$$T_{total} = T_d + T_s + T_r \dots\dots\dots (5)$$

式中：

T_{total} ——负载均衡恢复时间；

T_d ——故障检测时间：指从负载均衡器或后端服务器出现故障到系统检测到该故障的时间；

T_s ——流量切换时间：在检测到故障后，负载均衡器将流量重新分配到其他可用服务器所需的时间；

T_r ——服务恢复时间：对于出现故障的服务器或服务，恢复到正常运行状态所需的时间。

8.4 可用服务数量

- a) 定义描述：用于衡量量子计算服务平台所能提供的可用服务资源数量，直接决定了平台的吞吐能力和用户的使用体验。
- b) 评估方法：记录在一段时间内，服务平台能同时支持的最大任务数、总的计算资源数量以及在故障发生时快速恢复所需最短时间。

8.5 吞吐量

8.5.1 HTTP 吞吐量

- a) 定义描述：用于衡量量子计算服务平台在单位时间内安全处理 HTTP 请求的能力，反映其在高并发场景下的针对于抗 DDoS 攻击能力等安全能力实现的性能表现。
- b) 评估方法：开启量子计算服务平台的应用防火墙等相关安全攻击防护功能，使用测试仪表或工具测试量子计算服务平台在无应用层业务传输失败且正常识别并阻断攻击情况下的 HTTP 吞吐量（页面大小分别在 64KB、32KB、16KB、1KB 等条件下）。

8.5.2 HTTPS 吞吐量

- a) 定义描述：用于衡量量子计算服务平台在启用安全加密与防护机制下，高效处理加密请求并维持如安全校验等业务稳定的综合能力。
- b) 评估方法：开启量子计算服务平台相关安全攻击防护功能，使用测试仪表或工具测试量子计算服务平台在无应用层业务传输失败且正常识别并阻断攻击情况下的 HTTPS 吞吐（页面大小分别在 64KB、32KB、16KB、1KB 等条件下）。分别测试 TLS_ECDHE_RSA_WITH_AES_256_GCM_SHA384(TLS1.2)、
TLS_ECDHE_RSA_WITH_AES_128_CBC_SHA256 (TLS1.2) TLS_AES_128_GCM_SHA256 (TLS1.3)、ECC_SM4_SM3(GMTLS)等密码套件。

8.6 新建连接速率

8.6.1 HTTP 新建连接速率

- a) 定义描述：用于衡量量子计算服务平台在抵御高频连接型攻击（如 CC 攻击、HTTP Flood）时，快速建立并安全处理合法请求的防御效能。
- b) 评估方法：使用测试仪表或工具测试量子计算服务平台在无应用层业务传输失败情况下的 HTTP 新建连接速率（页面大小 64 KB）。

8.6.2 HTTPS 新建连接速率

- a) 定义描述：用于衡量量子计算服务平台在高强度加密握手与安全防护压力下，快速建立可信通信通道并抵御资源耗尽攻击的综合能力。
- b) 评估方法：使用测试仪表或工具测试量子计算服务平台在无应用层业务传输失败情况下的 HTTPS 新建连接速率（页面大小 64 KB）。分别测试 TLS_ECDHE_RSA_WITH_AES_256_GCM_SHA384 (TLS1.2)、TLS_ECDHE_RSA_WITH_AES_128_CBC_SHA256 (TLS1.2)、TLS_AES_128_GCM_SHA256 (TLS1.3)、ECC_SM4_SM3 (GMTLS) 等密码套件。

参 考 文 献

- [1] GB/T 36326-2018 信息技术 云计算 云服务运营商通用要求
 - [2] GB/T 37738-2019 信息技术 云计算 云服务质量评价指标
-