

T/ISC

团 体 标 准

T/ XXXX—XXXX

AI 算力场景下高带宽内存芯片性能适配技术规范

Technical specification for performance adaptation of high-bandwidth memory chips
in AI computing power scenarios

（征求意见稿）

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

中国互联网协会 发 布

目 次

前言 II

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 技术要求 1

6 性能适配指标 2

7 测试方法 4

附录 A（资料性） 测试报告模板 8

参考文献 9

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。

本文件由中国互联网协会归口。

本文件起草单位：

本文件主要起草人：

AI 算力场景下高带宽内存芯片性能适配技术规范

1 范围

本文件规定了高带宽内存芯片（HBM）在人工智能训练和推理场景下的技术要求、性能适配指标和测试方法。

本文件适用于搭载高带宽内存（HBM）的AI加速卡及其HBM存储子系统的设计、选型、测试与验收。采用上述AI加速卡的AI服务器和智算中心在集成、选型和验收时可参照使用。规范性引用文件

本文件没有规范性引用文件。

2 术语和定义

下列术语和定义适用于本文件。

2.1

高带宽内存芯片 high-bandwidth memory chip

通过硅通孔（TSV）和微凸块技术实现多层堆叠、具有高带宽和低功耗特点的DRAM芯片。

2.2

存算比 memory-to-compute ratio

AI算力系统配置的高带宽内存标称峰值带宽与规定计算精度下标称稠密峰值算力的比值。

2.3

软错误率 soft error rate; SER

在规定工况及观测层级下，单位具有有效检测覆盖的受测存储容量与单位有效时间内发生的ECC后不可纠正软错误事件数。本文件表5的SER以单个AI计算模组或加速卡内HBM控制器ECC解码端可见的不可纠正软错误事件为统计对象；不计GPU缓存、寄存器、主机内存或接口错误，不将事件数等同错误位数。器件内部纠错屏蔽而未在该端显现的事件不在本指标观测范围内，本指标不应表述为器件原始SER。

3 缩略语

下列缩略语适用于本文件。

AI：人工智能（Artificial Intelligence）

HBM：高带宽存储器（High Bandwidth Memory）

HBM3：高带宽存储器第三代（High Bandwidth Memory 3）

HBM3e：高带宽存储器第三代增强版（High Bandwidth Memory 3 enhanced）

HBM4：高带宽存储器第四代（High Bandwidth Memory 4）

LLM：大语言模型（Large Language Model）

TSV：硅通孔（Through-Silicon Via）

BER：误码率（Bit Error Rate）

ECC：纠错码（Error Correcting Code）

4 技术要求

4.1 存算协同架构要求

表1存算比应以同一测量单元、同一工作档位的标称HBM峰值带宽与规定精度下的稠密峰值算力计算。通用AI训练、大模型训练和多模态/视频模型训练应采用BF16或FP16稠密峰值算力；大模型推理和多模态/视频模型推理应采用FP8稠密峰值算力。不应计入结构化稀疏对应的峰值算力。报告应列明应用类型、计算阶段、计算精度、带宽、算力及各自来源，不应混合不同精度或不同设备的数值。其他精度、量化方式或稀疏计算方式应记录其计算口径并单独评价，不直接套用表1数值。存算比符合表1不应代替应用容量与实际负载性能的评价。

表 1 存算协同架构要求

应用类型	计算阶段及精度口径	最低存算比（GB/s • TFLOPS ⁻¹ ）
通用AI训练	BF16或FP16稠密训练	≥3.0
大模型训练（LLM）	BF16或FP16稠密训练	≥3.0
大模型推理	FP8稠密推理	≥1.5
多模态/视频模型训练	BF16或FP16稠密训练	≥3.0
多模态/视频模型推理	FP8稠密推理	≥1.5
注：结构化稀疏峰值算力与稠密峰值算力的统计口径不同。		

4.2 工作环境要求

在JEDEC规范允许范围内，确保长期可靠性。应记录所依据的JEDEC文件编号、版本及适用器件，以及制造商规定的工作温度、电压和刷新条件。试验期间应记录实际条件；出现超限时应标记对应时段及其对结果的影响。

4.3 信号完整性要求

- 应符合以下规定：
- a) BER≤10⁻¹²（链路级）；
 - b) BER≤10⁻¹⁵（ECC 后）；
 - c) HBM 数据接口数据率应不低于 6.4 Gb/s/pin。

5 性能适配指标

5.1 带宽性能

第6章指标以搭载被测高带宽存储芯片的单个AI计算模组或加速卡为测量单元。被测芯片代际、容量、额定数据速率、计算精度、软件栈、散热配置及实际工作档位应保持一致并记录。涉及存储器功耗、结温和信号完整性的项目，应同时标明器件级测点与本测量单元的对应关系；不应以整卡功耗、壳温或主机接口错误计数直接替代HBM器件指标。

带宽性能应符合表2要求。

表 2 带宽性能

指标	要求
单任务持续有效带宽利用率	Copy、Scale、Add、Triad各项均应≥70%
带宽时间序列稳定性	应报告固定采样窗口下的P5、P50、P95、最小值、最大值和变异系数（CV）
多任务带宽影响与连续性	应报告各任务及合计有效带宽、相对单任务基线的比例、各任务最低进度及最长无进展时间

单任务持续有效带宽利用率为单任务有效统计阶段内实测平均有效带宽与理论峰值带宽的比值。带宽时间序列稳定性按7.2规定的固定采样窗口统计并报告。多任务带宽影响与连续性按同一测量单元内多任务结果相对相应单任务基线的比例、各任务结果及最长无进展时间进行报告。理论峰值带宽按被测AI计算模组或加速卡的HBM额定带宽确定。理论峰值带宽应按测试前确定的额定工作档位固定计算，并在单任务与多任务比较中保持一致；不应随测试中的热降频或功率限制降低分母。应同步记录实际频率和限频事件。

5.2 延迟性能

延迟性能应符合表3要求。

表 3 延迟性能

指标	要求
----	----

指标	要求
GPU类AI加速卡设备侧HBM全局内存基准平均访问延迟	≤800 cycles
延迟分布	应报告P50、P95、P99及最大值
延迟测试条件	应报告实际工作频率、计时器来源、HBM代际及数据率、访问位置、数据集规模、TLB状态、访问步长、访问粒度和并发配置

GPU类AI加速卡设备侧HBM全局内存基准平均访问延迟应在设备侧测试程序中测量。访问数据集应大于L2缓存容量，访问位于HBM可寻址全局内存，TLB应预热，访问步长、访问粒度和事务形成方式应固定。≤800cycles仅适用于上述基准条件；cycles指GPU设备侧计时周期，不等于HBM时钟周期。随机访问、顺序访问及不同并发深度的延迟分布仍应分别测试和报告，但不以小于等于800cycles作为统一硬判定。延迟应从设备侧发出存储访问请求计至相应数据可供该请求使用，不计主机任务提交和主机与设备之间的数据传输时间。

5.3 单位比特能耗

单位比特能耗应符合表4要求。

表 4 能效比

技术节点	能耗
HBM3	≤5 pJ/bit
HBM3e	≤4.5 pJ/bit
HBM4	≤4 pJ/bit

单位比特能耗为规定负载下被测对象的动态功耗与实测有效带宽的比值。动态功耗为负载状态平均功耗与相同环境、配置和工作档位下空闲状态平均功耗之差。功耗测量边界应与带宽测量边界一致，并在测试报告中说明。功率测量应覆盖产生所测带宽的HBM器件供电边界，不包含AI计算核心等非HBM负载；应列明所覆盖的电源轨及未被测量的部分。仅列明未测电源轨不能代替完整边界测量；未覆盖的HBM功耗不应按零计入，测量边界不完整时不应据此作表4符合性判定。

5.4 可靠性

可靠性指标应在规定的环境条件、工作电压、数据速率和刷新配置下进行评价。可靠性应符合表5要求。

表 5 可靠性

指标	要求
连续稳定运行时间	≥72 h；不应出现不可纠正错误、功能失效或测试中断，应记录可纠正错误次数
ECC后不可纠正软错误率（SER）	≤10 ⁻¹³ errors/（bit·h），按7.5规定的单侧95%置信上限判定
可靠性资料	应提供适用对象、失效定义、评估或预测方法、任务剖面、环境条件和适用边界

5.5 热与功耗约束

热与功耗约束应符合表6要求。

表 6 热与功耗约束

指标	要求
结温	≤95℃
热梯度	≤20℃
满载功耗与散热条件	应报告功耗测量边界、总功耗、环境温度、冷却方式、工作负载、运行频率和测试持续时间

结温为满载稳态时HBM器件的最高结温；热梯度为同一HBM器件最高温测点与最低温测点的温差。测点位置、温度采集方式和温度类型应在测试报告中说明。多器件配置应逐器件计算温差，并以最不利结

果判定；不应以跨器件平均值掩盖局部超限。温差计算应采用同一时段、相同温度类型且具有可比较校准关系的测点，不应将表面温度与结温相减；结温估算应列明模型适用范围和误差。满载功耗与散热条件应按表6要求报告。

6 测试方法

6.1 测试负载

应包含以下内容：

- d) LLM 训练（Transformer）：模拟 GPT、LLaMA 等稠密 Transformer 模型训练，序列长度 4096~8192，含周期性检查点保存/恢复；
- e) LLM 推理（KV Cache 访问）：模拟大模型推理的 KV Cache 反复读写，批量大小与并发请求数按典型部署配置；
- f) 稀疏访问模式：模拟 MoE 等稀疏模型的路由访问特征，按稀疏索引随机访存；
- g) 高并发访存：模拟多实例或多任务并发访存，并发线程数 ≥ 32 ；如涉及多卡，应分别记录并评价各卡结果。

6.2 带宽测试

测试方法如下：

- a) 使用适用于被测 AI 加速器的 STREAM 类带宽测试程序或具有等效访存模式的设备侧测试工具，分别执行 Copy、Scale、Add 和 Triad 等访存模式；访问数据集应大于末级缓存容量；有效带宽应按完成的读写有效字节总量除以对应时间计算，Copy 和 Scale 计一次读、一次写，Add 和 Triad 计两次读、一次写，不计重复统计量及未完成访问；应校验各访存模式的运算结果，记录编译选项、元素类型、数组规模及计数规则；不应把被优化删除或未产生有效 HBM 访问的运算计入有效流量；
- b) 每一访存模式在单任务和多任务状态下均应测试不少于 3600s，前 600s 为预热期且不计入统计；其后全部连续有效统计阶段应纳入计算，不应只截取波动较小的时段；单任务与多任务各模式的结果应分别记录，不应以各模式合计 3600 s 代替单项时长；
- c) 在有效统计阶段，按 1s 固定采样窗口记录时间戳、有效带宽和负载状态。各采样窗口应等长且连续；有负载而无完成访问的窗口应记为零带宽，采集缺失应单独标记，不应作为零值或正常值填补；GB 按 10^9 字节计，原始完成字节数及窗口起止时刻应保存；采集缺失影响指标完整性时，应复测并保留原始异常记录。测试终止时不足一个完整窗口的数据应单独保留，不应混入统计；
- d) 分别在单任务和多任务并发状态下进行测试。多任务并发测试时，应记录并发任务的数量、类型和配置。单任务有效带宽和多任务有效带宽应在相同访存模式、访问粒度、数据规模、测试时间及设备工作状态下测量；多任务带宽应采用同一统计窗口内所有任务完成的有效字节总量计算，并同时保存各任务结果；各任务应在共同统计阶段持续施加负载；任务提前结束或未实际运行时不应仍按所配置并发数评价。单任务和多任务的总数据规模及各任务所占范围应分别记录；应分别记录各任务的有效完成量及最长无有效访问完成区段，不应仅以合计带宽掩盖某一任务未获得服务的情况；
- e) 单任务持续有效带宽利用率按实测平均有效带宽与理论峰值带宽之比计算。带宽时间序列稳定性应报告 P5、P50、P95、最小值、最大值和变异系数（CV）。多任务带宽影响与连续性应报告各任务及合计有效带宽、相对单任务基线的比例、各任务最低进度及最长无进展时间；
- f) 各访存模式的单任务持续有效带宽利用率均应符合表 2 的规定。带宽时间序列稳定性和多任务带宽影响与连续性应按表 2 要求报告。不同访存模式的测试结果应分别报告，不应以各模式测试结果的平均值代替单项结果；理论峰值带宽或平均有效带宽为零时，不应以零值代入相应分母给出合格结论，应记录原因并复核；
- g) 测试报告中应记录被测对象、测试工具及版本、理论峰值带宽及其数据来源、访存模式、访问粒度、数据规模、测试持续时间、预热时间、采样窗口、设备工作档位、单任务基线、并

发任务组合、有效带宽统计结果及表 2 规定的报告结果。测试原始数据应予保存，并能够支持测试结果复核。

6.3 延迟测试

6.3.1 延迟测试应分别采用基准访问、随机访问和顺序访问模式。基准访问应满足 6.3 规定的 HBM 全局内存、数据集大于 L2 缓存、TLB 预热、固定访问步长、固定访问粒度和固定事务形成方式等条件。随机访问和顺序访问的测试应保持访问粒度、随机地址生成规则和数据集规模一致；应验证计时器开销、请求完成依赖关系及缓存控制效果，并保存验证记录。

6.3.2 并发访问深度依次设置为 1、32 和 128，每种配置的测试时间应不少于 600s；并发深度指被测单元中已发出且尚未完成的访问请求数，应报告配置深度及实际在途请求统计；不应以线程数直接代替。实际工作频率、计时器来源、HBM 代际及数据率、访问位置、数据集规模、TLB 状态、访问步长、访问粒度和并发配置均应记录。

6.3.3 基准访问平均延迟按有效统计阶段全部访问延迟样本的算术平均值计算；延迟分布应报告 P50、P95、P99 及最大值。随机访问和顺序访问结果应分别报告，各访问深度应分别统计。超时、访问失败和采样丢失应单独计数并保留，不应将其当作正常完成样本。有效样本数为零或平均访问延迟不大于零时，不应据此判定合格，应复核计时及采集结果。

6.3.4 基准访问平均延迟应符合表 3 的规定。测试报告中应记录访问模式、访问位置、访问粒度、数据集规模、缓存控制策略、TLB 状态、访问步长、并发访问深度、实际工作频率、计时器来源、HBM 代际及数据率、样本数量及延迟分布结果。出现未完成访问、数据错误或测量失效时，应报告对应异常，不应仅凭成功访问样本的分位数作出该测试完整符合的结论。

6.4 能耗测试

单位比特能耗按式（1）计算：

$$E = 1000 \frac{P}{B} \dots\dots\dots (1)$$

式中：

E ——单位比特能耗，单位为皮焦每比特（pJ/bit）；

P ——规定负载下被测对象的动态功耗，为负载状态平均功耗与空闲状态平均功耗之差，单位为瓦（W）；

B ——与功耗测量同步取得的平均有效带宽，单位为吉比特每秒（Gbit/s）。

功耗测量应与带宽测试同步进行。负载状态和空闲状态应采用相同的环境条件、软硬件配置、散热配置及设备工作档位。按式（1）计算单位比特能耗，结果应符合表4的规定。测试报告中应记录功率测量位置、测量边界、空闲平均功耗、负载平均功耗、动态功耗、满载总功耗、平均有效带宽及计算结果。带宽由GB/s换算为Gbit/s时应乘以8；动态功耗为负或同步带宽为零时，应复核测量，不应据此判为能耗合格。功耗与带宽应采用同一负载窗口；不应将最低功耗窗口与最高带宽窗口组合。应报告测量不确定度及接近限值时所采用的判定规则；动态功耗差值未超过测量分辨能力时，不应报告为已证明零能耗。

6.5 可靠性测试

6.5.1 连续稳定运行测试：按照 7.1 规定的代表性负载连续运行不少于 72h，全程监测运行状态、ECC 状态、错误日志、数据完整性及有效带宽。测试期间不应出现不可纠正错误、功能失效或测试中断；应记录可纠正错误次数。连续运行期间应覆盖 7.1 所列各类负载，记录负载顺序、持续时间、切换时刻及各类负载的运行结果；不应以单一未说明负载代替全部代表性负载。异常中断应单独报告，不应将中断前后的运行时长相加记为一次连续 72 h。

6.5.2 软错误率测试：应按第 3 章规定的 HBM 控制器 ECC 解码端观测口径记录不可纠正软错误事件，并记录 ECC 配置、计数器含义及检测覆盖能力。应保存事件时间、地址或可用定位信息、症状、去重依据及复核记录；经复核排除持续性故障后方可归类为软错误。原因未明事件应单独报告，不应当作零事件、从异常记录中删除或据此判定 SER 符合。

6.5.3 SER 试验应在正常供电、按器件要求刷新及 ECC 启用的条件下进行。测试前应固定并记录样品、受测地址及容量、数据模式、访问与巡检方式、温度、电压、刷新和 ECC 配置、地点及海拔、辐射环境资料、事件检测和定位能力。对受测容量的不可纠正错误检测覆盖应有证据；仅记录可分配容量或无日

志输出不应视为已完成全容量有效观测。未持续受测或未被有效监测的容量、时间不应计入暴露量。辐照加速试验的数据不应直接并入普通运行试验的容量时间统计。

6.5.4 每个受测段的容量、观测层级和工况应保持确定且不变；变化时应分段记录，不同工况分别评价。各样品应分别判定，不应以跨卡合计结果代替单卡符合性结论。

6.5.5 SER符合性判定采用单侧95%置信上限。令 N 为各有效受测段去重后的不可纠正软错误事件数之和， E 为各段具有有效检测覆盖的受测容量与有效时间乘积之和，单位为 $\text{bit}\cdot\text{h}$ 。采用恒定事件率的泊松模型时，置信上限由自由度为 $2(N+1)$ 的 χ^2 分布的0.95分位数除以 $2E$ 得到； N 为零时，上限为 $-\ln(0.05)/E$ ，约为 $2.996/E$ 。应同时报告点估计与上限，仅当上限不大于 10^{-13} errors/($\text{bit}\cdot\text{h}$)时，方可判定该项符合。

6.5.6 试验方案应事前确定观测条件、固定结束时点或有效暴露量、接受事件数及异常终止规则。零事件方案所需有效暴露量应不小于 2.996×10^{13} $\text{bit}\cdot\text{h}$ ；允许非零事件的方案应按对应置信上限反算更大的计划暴露量。不应在观察结果后反复延长、截取或选择数据直至取得合格。发生外部测量故障时应保留原记录，按事前规则处理无效暴露及复测；不应删除样品真实异常后重新计为零事件。

6.5.7 应说明事件率恒定及事件独立性假设的适用依据。存在持续性缺陷、显著聚集或工况漂移而无法支持该模型时，不应强行按泊松模型判定。模型不适用、原因未明事件未解决、检测覆盖不足或有效暴露量不足时，应报告无法证明符合；统计上限高于限值时，该次试验不应判为符合。零事件不表示真实SER为零。

6.5.8 72 h连续运行试验仅在同时满足本项观测口径、检测覆盖、工况和事前统计方案时，方可贡献对应SER暴露量；不应自动以全部标称容量乘72 h代替有效暴露量。SER统计达标不能替代72 h连续运行要求；72 h未出现不可纠正错误也不能替代SER统计判定。SER结论仅适用于所列条件及观测层级，不应外推为所有运行环境下的器件原始错误率。

6.5.9 测试报告中应记录被测对象、样品数量、环境条件、工作电压、数据速率、刷新配置、测试负载、累计测试时间、可纠正和不可纠正错误记录、计算方法及合格判定。制造商应提供被测AI加速卡及其HBM存储子系统的可靠性资料，包括适用对象、失效定义、评估或预测方法、任务剖面、环境条件和适用边界。

6.6 热与功耗测试

应采用能够反映HBM结温的片上传感器，或经适用热模型及校准关系换算的测温方法；使用热电偶或红外热像仪时，应记录测点、表面温度及换算依据，不应将表面温度直接作为结温。测试环境温度为 $25\text{ }^{\circ}\text{C}\pm 2\text{ }^{\circ}\text{C}$ ，满载运行不少于30 min，并持续记录至温度及功耗进入稳定状态后，记录最高结温、同一器件温差与满载总功耗。应保存稳定状态的时间序列及判据。按表6判定结温 $\leq 95\text{ }^{\circ}\text{C}$ 、热梯度 $\leq 20\text{ }^{\circ}\text{C}$ ；应报告功耗测量边界、满载总功耗、环境温度、冷却方式、工作负载、运行频率和测试持续时间。器件制造商规定的结温上限低于 $95\text{ }^{\circ}\text{C}$ 时，还应满足该结温上限；制造商另行规定的壳温或环境温度范围，应在对应测点分别满足，不应直接替代结温限值。

6.7 信号完整性测试

使用高速示波器进行眼图测试，验证高速接口（ $\geq 6.4\text{ Gb/s/pin}$ ）的信号质量；开展链路级BER测试，记录链路级 $\text{BER}\leq 10^{-12}$ 、ECC后 $\text{BER}\leq 10^{-15}$ 的满足情况，结果与5.3条要求逐项对应。眼图结果不应代替误码统计。链路级和ECC后统计应分别记录观测位置、错误位数、受测总比特数、时间、数据速率和数据模式；不应以ECC事件数直接充当错误位数。零误码时应报告受测比特数和所用置信水平下的误码率上限。链路级与ECC后统计不应混合分母；合并多个通道时应同时保存各通道结果和暴露量，不应把合计统计结果表述为已证明每条通道达到相同上限。误码观测不应仅选取工作期间的有利时段。受测总比特数应来自所报告通道和观测位置的实际有效观测；不应仅以额定速率乘以墙钟时间代替未持续发送或未持续监测期间的暴露量。受测比特数为零时不应计算误码率或判定符合。

6.8 检验规则

6.8.1 测试样品抽样比例 $\geq 5\%$ （同一批次不少于3个样品）。

6.8.2 测试过程中任一适用技术要求或性能指标不满足第5章或第6章要求时，应复核测试配置、仪器状态和原始数据；复核后仍不符合要求的，判定该项不合格并如实记录。复测不应覆盖首次结果；报告应区分样品真实失效、外部试验条件失效和采集故障，并列明原始结果、纠正措施及复测结果。未完

成的必需项目不应按合格计算。符合性判定规则及其所用统计模型、置信水平和测量不确定度处理方式应在测试前列入测试方案；不应根据测试结果事后更换规则以取得合格结论。SER 应执行 7.5 规定的单侧 95%置信上限及逐样品判定，不应另以点估计或其他置信水平替代。

6.8.3 连续稳定运行时间和 ECC 后不可纠正软错误率应分别按 6.4 及 7.5 规定的方法进行评价。

6.9 测试报告

测试完成后应形成正式测试报告，报告格式参见附录A。报告应区分实测结果、制造商资料及统计推算结果；测试缺项、测量失效或条件不符合时，应逐项说明，不应据此形成全部要求符合的结论。采用统计合并、删失结果或制造商可靠性资料时，应列明适用对象和工况，不应将单次短时通过、无观测事件或未测试的后续代际表述为已获得完整验证。

附 录 A
(资料性)
测试报告模板

- A.1 样品信息：型号、批次、HBM 代际、容量、数量、额定数据速率和制造商。
- A.2 测试平台：AI 计算模组或加速卡、主机、驱动、运行库、测试工具、计算精度、频率和散热配置。
- A.3 测量边界：带宽和延迟测量层级、功率测点、空闲基线、温度测点及 ECC 状态。结温换算依据、功率所含电源轨及测量不确定度。SER 控制器 ECC 观测位置、计数器定义、内部纠错屏蔽范围及检测覆盖证据。
- A.4 测试条件：负载、访问模式、访问粒度、并发配置、预热时间、测试时间、1 s 采样窗口及稳态窗口。固定带宽分母、实际频率及限频事件、各负载的连续运行时段和实际在途深度。判定规则及其预先确定的统计模型、置信水平和测量不确定度处理方式。同组比较配置、应用容量需求与分片配置、计时器验证、缓存控制验证和 TLB 状态记录。SER 地点及环境、数据与巡检方式、固定观测计划、接收事件数和异常终止规则。
- A.5 测试结果：原始数据文件标识、计算中间量、带宽、延迟、能耗、可靠性、热与功耗及信号完整性结果。误码统计暴露量、SER 分段容量和时间、各通道误码暴露量、失败及超时数量、测量不确定度和判定规则。逐任务完成量及最长无进展时间，动态功耗与满载总功耗分别列示。带宽时间序列的 P5、P50、P95、最小值、最大值和 CV。延迟分布的 P50、P95、P99 及最大值。SER 分类与去重事件数、原因未明事件、有效暴露量、点估计、单侧 95% 置信上限及泊松模型适用依据。
- A.6 结论：各项合格判定、异常情况、复核过程及最终结论。

参 考 文 献

- [1] GB/T 9813.3 计算机通用规范 第3部分：服务器
 - [2] GB/T 26236.1 信息技术 软件资产管理 第1部分：过程
 - [3] JESD270-4 High Bandwidth Memory (HBM4) DRAM
 - [4] JESD238 High Bandwidth Memory 3 (HBM3) DRAM
 - [5] JESD235 High Bandwidth Memory (HBM) DRAM
 - [6] IEEE 754 浮点算术标准
 - [7] OCP NIC 3.0 Specification (OCP规范 电气规范)
 - [8] UCIe 1.1 Specification (针对合封die之间的数据接口的标准)
-