

2025 年全国大学生计算机系统能力大赛

智能计算创新设计赛（先导杯）

面向 CPU-DCU 异构平台的 GMRES 优化

赛题技术方案

一、评价方式的基本说明

第 1 条 广义最小残差法（Generalized Minimal Residual, GMRES）广泛应用于计算流体力学和电磁仿真等领域，是科学计算中求解大型稀疏线性方程组 $Ax=b$ 的主流迭代算法之一，适用的问题比较广，其中 A 为稀疏矩阵， b 通常称为右端向量。与传统的直接解法相比，GMRES 通过 Krylov 子空间投影逐步逼近方程的解，从而减少计算量并提高数值稳定性。

第 2 条 GMRES 算法的计算效率通常依赖于硬件加速器的并行计算能力，但由于其迭代依赖、不规则访存等问题，使 GMRES 在加速器上的性能受到限制。本赛题提供重启动 GMRES 算法在 CPU 上的串行实现，要求各参赛队综合运用各种知识：包括但不限于线性算子的 DCU 移植、混合精度迭代算法设计、SpMV 算子加速、算子融合等），实现 GMRES 算法在国产加速卡上的优化，在相对残差满足给定阈值的前提下实现更快速、更高效的问题求解。

第 3 条 本赛题考察 GMRES 在 DCU 上的移植和优化，要求只使用国产加速卡加速 GMRES；本赛题禁止修改基准测试中计时相关代码、右端向量初始化代码、最大迭代次数、残差阈值；禁止调用成熟线性代数库（包括但不限于 MAGMA、GINKGO 等）的 GMRES 实现；本赛题不考虑迭代算法预处理器/预条件子的优化，故禁止参赛队使用任何预处理；考虑到冷启动问题，本赛题允许选手在计时之前添加 warmup 代码；稀疏线性方程组的系数矩阵需以 CSR 格式压缩存储，该矩阵 A 和右端向量 b 传输到 DCU 端的开销，以及结果向量 x 传输到 CPU 端的开销可不算在计时范围内；若要采用其他的稀疏矩阵压缩格式，需从 DCU 端的 CSR 矩阵开始转换，且格式转换的代码必须包含到计时范围内。

第 4 条 除本技术方案特别要求、规定和禁止的事项外，各参赛队可自行决定优化采用的方案、代码优化等细节。在 DCU 移植的过程中，允许参赛队调用 hipBLAS 和 hipSPARSE 中的线性算子或手工实现内核函数，不允许调用除此之外的其他库。

二、初赛评分标准

第 5 条 残差验证。

1. 在残差验证中，稀疏线性系统的系数矩阵均采用大赛给定的稀疏矩阵，右端向量均采用大赛给定源码初始化。
2. 根据赛事提供的硬件平台，运行残差验证数据集，残差阈值设置为 $1E-6$ ，如果迭代最终相对残差小于或等于残差阈值，则验证通过。

第 6 条 性能测试。

1. 性能验证数据集和残差验证数据集相同。
2. 比赛共 80 个测试用例，每个用例满分为 1.25。
3. 对于每个测试用例，在给定最大迭代次数下：
 - (1) 如果残差验证未通过，则参赛队在该测试用例上的得分为 0；
 - (2) 如果残差验证通过，则记录参赛队在该测试用例上的时间为 T 。设该测试用例的基准时间为 T_{baseline} ，按下列公式计算参赛队在该测试用例上的得分。若 $T > T_{\text{baseline}}$ ，则参赛队在该测试用例上的得分为 0。

$$(1 - T / T_{\text{baseline}}) \times 1.25$$

最终得分为 80 个测试用例的得分总和。

第 7 条 初赛总成绩 100 分，按性能测试得分为准。

三、决赛评分标准

第 8 条 比赛内容与初赛一致。

第 9 条 残差测试与性能测试方法与初赛一致。参赛队团队协作及现场答辩的评分标准在决赛阶段另行公布。

第 10 条 决赛总成绩 100 分，各分项成绩权重如下：

1. 性能测试成绩：80%
2. 参赛队团队协作、设计文档及现场答辩：20%

四、 参赛作品提交

第 11 条 各参赛选手初赛、决赛阶段需要在大赛平台提交完整的设计内容：

1. 参赛队均需提供完整的 GMRES 算法优化源码，以进行评测。
2. 参赛队需提供优化文档。内容包括但不限于基本算法介绍、设计思路和方法、算法优化、详细算法设计与实现、实验结果与分析、程序代码模块说明、详细程序代码编译说明、详细代码运行使用说明等。

第 12 条 参赛队必须严守学术诚信。如果需要使用第三方 IP 或者借鉴他人的部分源码，必须在设计文档和源代码的头部予以明确说明。

五、 竞赛平台与测试程序

第 13 条 大赛提供的竞赛平台和测试程序包括：

1. 代码托管平台，支持各参赛队的群体协作与版本控制。
2. 竞赛评测系统，根据参赛队的申请从代码托管平台获取指定版本，并加载基准测试程序，自动进行残差及性能测试。
3. 基于基准测试数据集，用于在目标硬件平台上对参赛队提交的代码进行性能评测。

六、 软硬件系统规范

第 14 条 如果在目标平台上进行算子手工开发和优化，可使用 HIP 编程接口和核函数语言。HIP 编程是目前进行 DCU 应用开发最主要的方式，为方便编译 HIP 代码，DTK 中提供了脚本工具 HIPCC，在内部帮助开发者完成预编译器、编译器、链接器的参数配置与调用。

第 15 条 使用大赛指定的硬件和编译环境编译参赛队提交的工程源码：

1. 加速单元：一块 K100, 64 GB
2. 软件栈：DTK 25.04, GCC 12.2.0

七、 大赛网站与参考资源

第 16 条 系统能力大赛主办方官网：<https://pra.educg.net>

第 17 大赛网站提供多种软件开发工具及设计资料，包括但不限于下列内容：

1. **HIP** 程序设计语言规范、文法及说明。
2. **DTK** 软件使用说明
3. 竞赛平台（代码提交方式、代码提交示例、竞赛平台使用）说明文档。
4. 性能基准测试程序及其文档。
5. 参赛队远程调试及本地调试指南。