Solver2026

第九届解法器快速算法及应用研讨会

会议手册

Program & Abstracts

2026年11月27日—30日

广东·深圳

solver-conference.cn



承办单位

中国科学院深圳先进技术研究院
SHENZHEN INSTITUTES OF ADVANCED TECHNOLOGY
CHINESE ACADEMY OF SCIENCES

香港中文大學（深圳）
The Chinese University of Hong Kong, Shenzhen

湖南韶峰应用数学研究院
Hunan Shaofeng Institute for Applied Mathematics

深圳国际工业与应用数学中心
SHENZHEN INTERNATIONAL CENTER FOR INDUSTRIAL AND APPLIED MATHEMATICS

赞助单位

HUAWEI　METASTONE 是石科技　中科曙光 Sugon　解得科技

组织委员会

徐小文（北京应用物理与计算数学研究所）
张晨松（中国科学院数学与系统科学研究院）
舒适（湘潭大学）
谭光明（中国科学院计算技术研究所）
刘伟峰（中国石油大学（北京））
薛巍（清华大学）

本地组织委员会

陈荣亮（中国科学院深圳先进技术研究院）
龚世华（香港中文大学（深圳））
蒋毅（中国科学院深圳先进技术研究院）
林增（中国科学院深圳先进技术研究院）
覃善林（中国科学院深圳先进技术研究院）
王东（香港中文大学（深圳））
王筱平（香港中文大学（深圳））
徐磊（中国科学院深圳先进技术研究院）
闫争争（中国科学院深圳先进技术研究院）

程序委员会共同主席

刘伟峰（中国石油大学（北京））
谢和虎（中国科学院数学与系统科学研究院）
陈荣亮（中国科学院深圳先进技术研究院）

程序委员会委员

安恒斌（北京应用物理与计算数学研究所）
陈荣亮（中国科学院深圳先进技术研究院）
崔涛（中国科学院数学与系统科学研究院）
冯春生（湘潭大学）
高振（中国海洋大学）
何鑫（中国科学院计算技术研究所）
胡少亮（中物院高性能数值模拟软件中心）
荆燕飞（电子科技大学）
李胜国（国防科技大学）
刘伟峰（中国石油大学（北京））
谭光明（中国科学院计算技术研究所）
谢和虎（中国科学院数学与系统科学研究院）
徐小文（北京应用物理与计算数学研究所）
薛巍（清华大学）
岳孝强（湘潭大学）
张晨松（中国科学院数学与系统科学研究院）
张硕（中国科学院数学与系统科学研究院）
钟柳强（华南师范大学）
周振亚（北京华大九天科技股份有限公司）
朱圣鑫（北师港浸大（UIC））

Solver 会场位置指引

中国科学院深圳先进技术研究院

A 栋 5 楼 504 会议室



地址：广东省深圳市南山区西丽深圳大学城学苑大道 1068 号

地铁：乘 7 号线至深大丽湖站，从 D 口出站，步行至研究院。

打车：导航至“中国科学院深圳先进技术研究院东门”。

第九届解法器快速算法及应用研讨会

2026 年 11 月 27 日至 30 日　广东 · 深圳

会议议程

日期	会议安排	地点
11 月 27 日（周五）	报到	
11 月 28 日（周六）	学术报告、特别主题论坛、SolverChallenge 论坛	深圳先进院 A 栋 5 楼 504 会议室
11 月 29 日（周日）	学术报告、赞助商报告、会议总结与会旗交接	深圳先进院 A 栋 5 楼 504 会议室
11 月 30 日（周一）	离会	

会场地址

中国科学院深圳先进技术研究院 A 栋 5 楼 504 会议室

广东省深圳市南山区西丽深圳大学城学苑大道 1068 号。乘地铁 7 号线至深大丽湖站，从 D 口出站，步行至深圳先进院；乘车可导航至深圳先进院东门。

车辆接送

会务组将安排大巴统一接送（酒店-先进院），车牌信息等在酒店注册报到处查询

出发时间：

11.28 上午 8:00 璟峯酒店-先进院，下午 18:50 先进院-餐厅（石鼓路光峰科技总部大厦）

11.29 上午 8:00 璟峯酒店-先进院，下午 17:30 先进院-璟峯酒店

签到与会务联系

11 月 27 日：16:00–18:00 注册报到（璟峯酒店大堂）

11 月 28 日、29 日：08:15–08:30 签到（会场门口）

会务咨询：陈荣亮 rl.chen@siat.ac.cn；张烨 ye.zhang3@siat.ac.cn

WiFi 信息

账号：A504

密码：siat@504

会议议程

11 月 28 日　星期六

时间	报告人	报告单位	报告题目
08:15–08:30			签到 / 报到
开场及学术报告　主持人：谭光明（中国科学院计算技术研究所）			
08:30–08:45			开场致辞
08:45–09:00	邹海峰	深圳市大数据研究院	AMG 算法强连通阈值智能调优及应用
09:00–09:15	朱子睿	湘潭大学	基于 MPI 进程分组策略的 setup 和 solve 异步并发的代数多重网格预条件迭代方法
09:15–09:30	钟柳强	华南师范大学	拟线性 H(curl)-椭圆问题混合有限元离散系统的两网格法
09:30–09:45	赵宇波	中国科学院深圳先进技术研究院	NML / NForge：把物理问题编译为可交付的专业求解软件
09:45–10:00	夏建林	南京大学	Superfast and Stable Eigenvalue and Singular Value Decompositions for Rank-Structured Matrices
10:00–10:30			茶歇交流、合影
特别主题论坛　主持人：陈荣亮（中国科学院深圳先进技术研究院）			
10:30–12:00			Solver 与 AI 的双向奔赴 嘉宾：许进超〔香港中文大学（深圳）〕、蔡小川（澳门大学）、Luca Pavarino（意大利帕维亚大学）等
12:00–13:30			午餐及午休
学术报告　主持人：舒适（湘潭大学）			
13:30–13:45	何鑫	中国科学院计算技术研究所	x-solver 在面向结构力学场景的异构求解技术
13:45–14:00	岳孝强	湘潭大学	Synchronization of Parareal over Multiple Coarse Steps
14:00–14:15	袁璠	国防科技大学	基于区域分解的分布式异步并行不完全 LU 分解方法
14:15–14:30	叶欣瑶	西南财经大学	Toward a Unified Iterative Solver Library for Complex Symmetric Linear Systems
14:30–14:45	杨肖剑	国防科技大学	针对多项式预条件的对角块内存感知优化及应用
14:45–15:00	阳王东	湖南大学	AI 赋能解法器的三个层面
15:00–15:15	徐飞	北京工业大学	Low-rank Approximation for Nonlocal Nonlinear Eigenvalue Problems
15:15–15:30	薛巍	清华大学	结构化代数多重网格的混合精度优化
15:30–16:00			茶歇交流
学术报告及赞助商报告　主持人：薛巍（清华大学）			
16:00–16:15	辛起	香港中文大学（深圳）	复杂几何区域上 Stokes 方程的几何稳健预条件方法
16:15–16:30	谢炎	深圳市大数据研究院	高频 Helmholtz 和时谐 Maxwell 方程的可扩展迭代求解：RAS-PML 区域分解与学习型子域加速
16:30–16:45	夏青	华为技术有限公司	面向普适科学计算场景的高性能数值计算引擎——华为天筹数值计算求解器
16:45–17:00	汪震	湘潭大学	Learning HTL-ORAS Algorithmic Components and Neural Solvers for ADR and Linear Elasticity Problems
17:00–17:15	王菁原	澳门大学	Data-Enhanced Preconditioning for a Sequence of Linear Systems
17:15–17:30	王川	香港浸会大学	大批量矩阵计算智能算法及其应用
17:30–17:45	苏怡丹	北京邮电大学	混合精度预条件下基于 QR 分解的块共轭梯度法舍入误差分析
17:45–18:00			华为赞助商报告
SolverChallenge 论坛　主持人：崔涛（中国科学院数学与系统科学研究院）			
18:00–18:45			颁奖、特等奖队伍报告、创新奖队伍快闪报告； SolverChallenge 竞赛：六年回顾与展望（报告人待定）
19:00–21:00			晚餐

会议议程

11 月 29 日　星期日

时间	报告人	报告单位	报告题目
08:15–08:30			签到 / 报到
		学术报告　主持人：冯春生（湘潭大学）	
08:30–08:45	宋健飞	浙江大学	HiSLU：面向动态稀疏线性系统的高性能增量求解器
08:45–09:00	宋博	西北工业大学	Space-time Parallel Methods for Wave Propagation
09:00–09:15	罗力	澳门大学	A Learnable Multigrid Framework via Graph Convolutions for Solving PDEs
09:15–09:30	李文瑄	中国石油大学（北京）	LinglongAMG：面向 GPU 的内存有效代数多重网格求解器
09:30–09:45	刘建军	中国科学院数学与系统科学研究院	高性能预条件子 HIPO 及其应用
09:45–10:00	林圣乐	湖南大学	从不规则到缓存友好：LeSpGEMM 的设计与优化
10:00–10:30			茶歇交流
		学术报告及赞助商报告　主持人：张晨松（中国科学院数学与系统科学研究院）	
10:30–10:45	李良	电子科技大学	ZiweiSolver：面向电磁工业软件的线性系统求解器
10:45–11:00	靳喜	成都信息工程大学	基于 MRRR 三对角特征值求解器的高性能批量对称矩阵特征值分解
11:00–11:15	贾朝蓬	清华大学	高可扩展的高阶模板问题多重网格求解器
11:15–11:30	胡议月	国防科技大学	TRIDENT: An End-to-End Streaming Accelerator for TriSpGEMM.
11:30–11:45	黄思路	北京应用物理与计算数学研究所	JXPAMG 解法器进展
11:45–12:00			是石科技赞助商报告
12:00–13:30			午餐及午休
		学术报告　主持人：刘伟峰〔中国石油大学（北京）〕	
13:30–13:45	张汉	清华大学	高温气冷堆耦合计算 TaiHang 软件的解法器进展
13:45–14:00	付书彬	宁波东方理工大学	A Massively Parallel Three-grid Preconditioner for the High-frequency Helmholtz Equation
14:00–14:15	范智博	北京应用物理与计算数学研究所	CFUA-AMG: 基于粗细变量分裂的非光滑聚集并行代数多重网格算法
14:15–14:30	杜皓	北京应用物理与计算数学研究所	稀疏线性代数方程组随机算法：现状与差距
14:30–14:45	丁彦灵	西南财经大学	A Robust Alternative Minimum Residual Method for Complex Symmetric Linear Systems
14:45–15:00	陈骏豪	北京邮电大学	代数多重网格求解器的自动精度选择
15:00–15:15	程贺琳	中国石油大学（北京）	What Frontier LLMs Can (and Cannot) Optimize in GPU SpMV
15:15–15:30	边浩东	青海大学	PANA：面向多核 CPU 的 SpMV 细粒度自适应负载均衡设计
15:30–16:00			茶歇交流
		学术报告及赞助商报告　主持人：徐小文（北京应用物理与计算数学研究所）	
16:00–16:15	安恒斌	北京应用物理与计算数学研究所	A Shell-Informed Smoothed Aggregation Multigrid Preconditioner for Linear Systems from Reissner–Mindlin Shells
16:15–16:30			中科曙光赞助商报告
16:30–16:45			解得科技赞助商报告
		总结与闭幕　主持人：谢和虎（中国科学院数学与系统科学研究院）	
16:45–17:00			会议总结：徐小文
17:00–17:15			闭幕与会旗交接：Solver26、Solver27 组委会

报告摘要

11 月 28 日报告摘要

AMG 算法强连通阈值智能调优及应用

邹海峰（深圳市大数据研究院） 2026 年 11 月 28 日 08:45–09:00

大规模稀疏线性系统的高效求解是工程数值模拟中的关键问题。代数多重网格（AMG）算法的强连通阈值直接影响粗化策略与多层结构，进而影响求解效率和收敛性能，传统依赖经验的阈值选择方式难以适应不同问题的特征。本报告将介绍 AMG 算法强连通阈值智能调优的基本原理，重点讨论基于深度学习的矩阵特征提取与阈值预测方法，并结合流体计算和油藏数值模拟中的典型算例，分析其在提升求解效率与增强算法稳健性方面的应用效果。

基于 MPI 进程分组策略的 setup 和 solve 异步并发的代数多重网格预条件迭代方法

朱子睿（湘潭大学） 2026 年 11 月 28 日 09:00–09:15

代数多重网格方法（AMG）是求解大型稀疏线性代数方程组最有效的算法之一，其主要工作包括先后执行的 setup 和 solve 两阶段。随着超级计算机性能不断提升，并行的规模越来越大，AMG 面临着算法可扩展性和并行可扩展性的严峻挑战。针对上述问题，本报告提出基于 MPI 分组策略的 setup 与 solve 异步并发的代数多重网格模型。模型将 setup 和 solve 分配至不同的进程组上进行异步并发，这种模式既可通过合理设置构造组进程数保留高质量 setup，又能通过异步并发获得两阶段的时间重叠收益。基于这个模型，我们设计并实现了 MPI ACAMG 算法。理论分析和数值实验表明本算法具有更好的算法可扩展性和并行可扩展性。例如针对空间局部各向异性的三维扩散问题，相比于传统方法，我们不仅降低了 solve 阶段的迭代次数，还在 64 进程和 128 进程下分别获得了 4.74 和 19.36 的加速比。

拟线性 H(curl)-椭圆问题混合有限元离散系统的两网格法

钟柳强（华南师范大学） 2026 年 11 月 28 日 09:15–09:30

针对拟线性 H(curl)-椭圆问题, 首先, 给出相应的混合有限元离散化方法, 其中两个有限元空间分别采用第一类线性棱有限元空间和节点协调型线性有限元空间, 并证明了该有限元解在 H(curl) 范数下的的最优误差估计; 其次, 设计和分析了两类求解上述离散系统的两网格法: 第一类算法为经典的两网格方法, 第二类算法为牛顿迭代校正的两网格方法, 并分别给出相应的收敛性分析; 最后, 数值实验分别验证了混合有限元解的最优误差估计和两类两网格法的收敛阶.

NML / NForge：把物理问题编译为可交付的专业求解软件

赵宇波（中国科学院深圳先进技术研究院） 2026 年 11 月 28 日 09:30–09:45

科学与工程计算中，大量工作仍卡在「已有模型与方法，却难以及时、可追溯地变成可维护的求解程序」。通用仿真软件覆盖常见任务，却不易承接新本构、新耦合与可审查交付；通用代码生成能写出程序，却还难稳定满足任务指标与适用条件。本报告介绍一种以数学结构显式表达为入口、以求解程序生成为出口的问题编译思路：把区域、未知场、方程、本构、边界与初值，以及可替换的离散、积分与时间推进，组织为可检查的计算对象，并据此生成求解程序与求解流程，并可编译为软件。报告将从解法器软件生产的角度，讨论形式化表达、程序生成与结果检验如何衔接，说明我们的当前进展，以及在部分专业问题族上的应用取向，希望与解法器同行就「可生成、可检验、可复用」的求解软件生产交换看法。

Superfast and Stable Eigenvalue and Singular Value Decompositions for Rank-Structured Matrices

夏建林（南京大学） 2026 年 11 月 28 日 09:45–10:00

For an important class of matrices, including banded matrices, Toeplitz matrices (in Fourier space), some kernel matrices, some discretized matrices, some transformation matrices, and many more, there are rank-structured representations or approximations. It is feasible to design so-called superfast eigenvalue decompositions for the symmetric case and superfast SVDs for the nonsymmetric or rectangular case. The idea is to use divide-and-conquer strategies, followed by accelerations with the fast multipole method, so as to reach nearly O(n) complexity. In this talk, we show some key strategies to overcome the complexity bottleneck from the original O(n^3) to nearly O(n), and moreover, several numerical stability challenges.

x-solver 在面向结构力学场景的异构求解技术

何鑫（中国科学院计算技术研究所） 2026 年 11 月 28 日 13:30–13:45

本报告将围绕结构力学典型场景中涉及的矩阵类型，介绍中科院计算技术研究所自研的 x-solver 迭代求解算法和异构加速技术，以及与商业结构力学仿真软件中迭代求解器的对比。

Synchronization of Parareal over Multiple Coarse Steps

岳孝强（湘潭大学） 2026 年 11 月 28 日 13:45–14:00

Synchronization between the coarse and fine solvers is a key feature of Parareal. The synchronization window length ΔT is typically set equal to the coarse step size Δt_c and under this setting there are extensive convergence analysis and applications of Parareal. A single window can also contain multiple coarse steps (i.e., $\Delta T = J_c \Delta t_c$ with $J_c > 1$) without any implementation difficulty. In this talk, we investigate whether using multiple coarse steps per synchronization window improves the speedup or not. For given fine step size and synchronization window number, we prove for two situations (ΔT fixed or Δt_c fixed) that the convergence factor of Parareal decays as J_c grows. For fixed ΔT an increase in J_c naturally leads to faster convergence but there is a trade-off due to the increased coarse solver cost. We show that in this case an optimal J_c (>1) exists which leads to better speedup compared to $J_c=1$. For fixed Δt_c, an increase in J_c does not cause relative cost increase of the coarse solver compared to the fine solver and thus a smaller convergence factor for $J_c>1$ implies solution computation on longer time intervals with better (at least the same) speedup versus $J_c=1$. We support our findings by numerical results. This is a joint work with Martin J. Gander and Shu-Lin Wu.

基于区域分解的分布式异步并行不完全 LU 分解方法

袁璠（国防科技大学） 2026 年 11 月 28 日 14:00–14:15

大规模稀疏线性方程组的高效求解是科学与工程计算中的关键问题。不完全 LU（ILU）分解是一类广泛使用的预条件技术，但其固有的数据依赖及进程间通信开销限制了并行可扩展性。本报告介绍一种基于区域分解的分布式异步并行 ILU 方法——DAS-ILU。该方法通过面向数据依赖的区域划分与重排序，将计算节点划分为可独立处理的内部节点和需要异步更新的分隔节点，使分隔区域规模较常规方法缩小约一半；结合局部 ILU 分解、异步不动点迭代及优化的数据交换策略，在避免显式构造 Schur 补的同时，降低分解与三角求解阶段的计算和通信开销。此外，该方法支持点块结构稀疏矩阵，并可作为代数多重网格的光滑子。结构力学、计算流体力学及油藏模拟等测试表明，DAS-ILU 相较 Chow–Patel 细粒度并行 ILU 和 HYPRE ILU 的求解加速比分别最高达到 2.07 和 4.11，

在 3,584 核、约 3.58 亿未知量的测试中展现出良好的可扩展性，并对具有挑战性的非对称及不定系统保持较强的鲁棒性，为大规模科学计算提供了一种高效的并行预条件方法。

Toward a Unified Iterative Solver Library for Complex Symmetric Linear Systems

叶欣瑶（西南财经大学） 2026 年 11 月 28 日 14:15–14:30

Large sparse complex symmetric systems Ax = b, where A equals its transpose A^T but not its conjugate transpose A^H, arise in applications such as frequency-domain wave propagation and electromagnetic scattering. A number of Krylov subspace methods exploit this structure, including COCG, COCR, QMR_SYM, CSQMOR, BiCGCR, SCBiCR, CS_MRTR and CS_MrR, but they are scattered across the literature with different calling conventions, and, to our knowledge, widely used numerical platforms such as MATLAB and Baltamatica (北太天元) provide no built-in solver that exploits the complex symmetric structure. We present an extensible iterative solver library, with a reference implementation in MATLAB, that provides these methods behind a common calling interface modeled on MATLAB's built-in solvers, with optional preconditioning, a common flag protocol for convergence, stagnation and breakdown, and a test suite for edge cases. Using the library, we compare the methods with one another and with MATLAB's general-purpose solvers on complex symmetric systems from Helmholtz-type discretizations and structural dynamics, measuring matrix-vector products, storage and true residuals. We are working on a port to Baltamatica and plan a C/C++ implementation, so that these methods become available beyond MATLAB.

针对多项式预条件的对角块内存感知优化及应用

杨肖剑（国防科技大学） 2026 年 11 月 28 日 14:30–14:45

Krylov 子空间方法广泛用于大型稀疏线性系统和特征值问题。它们的效率通常受限于在多项式和复合预处理方案中重复计算稀疏算子的过程，其中矩阵幂核（MPK）是一类重要的内存受限操作。我们提出了对角块 MPK（DBMPK），这是一种轻量级的操作符级优化，利用对角块的数据重用，同时保持规则的并行执行。DBMPK 通过块级核融合实现局部数据重用，同时避免了激进重用策略引入的全局依赖。与现有的 MPK 优化技术相比，DBMPK 在不同的稀疏模式和多项式阶数下通常能提供更可预测的性能。此外，我们还将 DBMPK 扩展到典型的多项式和复合预处理框架，展示了它在独立 MPK 计算之外的适用性。在 x86 和 ARM 高性能计算平台上的实验结果表明，DBMPK 在广泛的稀疏矩阵上能提供稳定的性能提升。在实际迭代求解器中集成 DBMPK 后，复合预处理器的速度可提升最多 23%，大型多重网格工作负载的性能约提升 17%。

AI 赋能解法器的三个层面

阳王东（湖南大学） 2026 年 11 月 28 日 14:45–15:00

利用 AI 技术来求解方程是发展的大趋势，在发展的过程中，随着 AI 的成熟与能力的增强，AI 赋能求解存在三个层面，一个是 AI 辅助，利用 AI 技术优化传统求解方法，充分发挥传统求解方法的能力。二是 AI 增强，利用 AI 扩展传统求解方法，能够发现更多更好新的求解方法。三是 AI 原生，利用 AI 来直接求解方程，不在依赖传统求解方法。

Low-rank Approximation for Nonlocal Nonlinear Eigenvalue Problems

徐飞（北京工业大学） 2026 年 11 月 28 日 15:00–15:15

Solving nonlocal nonlinear eigenvalue problems, such as those arising in Hartree-Fock theory, is fundamentally challenged by the prohibitive computational costs associated with the nonlocal Fock

exchange operator. To address this, we introduce a generalized low-rank approximation framework for the nonlocal operator. By leveraging low-rank decomposition and adjustable variables, the proposed method guarantees high accuracy for occupied orbitals while strictly preserving Hermiticity and structural consistency with the exact operator. This low-rank approximation strategy effectively eliminates the computational bottlenecks caused by nonlocality, enabling the efficient solution of large-scale nonlocal nonlinear eigenvalue problems.

结构化代数多重网格的混合精度优化

薛巍（清华大学） 2026 年 11 月 28 日 15:15–15:30

低精度计算硬件日益普及，如何利用其加速有计算精度需求的科学计算应用仍面临挑战。面向稀疏线性系统求解问题，本报告汇报在结构化代数多重网格预条件中利用半精度计算加速求解的最新成果。基于稀疏矩阵格式、科学应用数值特征及多重网格特点，给出四条半精度计算引擎使用准则，以及相应的优化实现技术。在 ARM 与 X86 平台上的实验验证了所提出原则和优化实现技术的有效性。

复杂几何区域上 Stokes 方程的几何稳健预条件方法

辛起（香港中文大学（深圳）） 2026 年 11 月 28 日 16:00–16:15

复杂区域中几何尺度的退化与局部尺度的显著差异，常导致 Stokes 方程离散系统的预条件效果恶化和迭代次数增长。本报告介绍一种基于边界距离的加权分块预条件方法，重点讨论其对几何参数变化的稳健性。该方法利用距离加权压力范数刻画局部几何尺度，构造压力 Schur 补的预条件算子，无需针对不同几何预先指定全局特征尺度。结合 Taylor–Hood 有限元、代数多重网格与 FGMRES，报告通过谱分析和数值实验，考察通道变窄、孔隙尺度变化、局部间隙收缩及多尺度结构共存时的求解表现。实验结果表明，在所测试的几何变化和网格加密过程中，该方法的预条件谱与迭代次数保持较为稳定；尤其在局部尺度差异显著的复杂区域中，相较传统压力质量矩阵及经参数调优的常系数预条件方法表现出更强的稳健性。

高频 Helmholtz 和时谐 Maxwell 方程的可扩展迭代求解：RAS-PML 区域分解与学习型子域加速

谢炎（深圳市大数据研究院） 2026 年 11 月 28 日 16:15–16:30

高频 Helmholtz 和时谐 Maxwell 方程离散后通常形成大规模、强不定线性系统，其高效可扩展求解是电磁波数值模拟中的重要问题。本报告介绍一种基于 RAS-PML 区域分解的并行迭代求解方法，通过在子域边界引入 PML 吸收层，提高局部问题对波传播的刻画能力，并结合 Krylov 迭代实现大规模并行求解。在此基础上，进一步研究多层学习方法与区域分解的融合，将 MGCFNN 与 RAS-PML 求解框架相结合，利用其多层结构提升高频 Maxwell 方程的求解效率。报告将重点讨论 RAS-PML 方法的收敛特性与并行可扩展性，以及 MGCFNN 与经典区域分解方法的融合与加速效果

面向普适科学计算场景的高性能数值计算引擎——华为天筹数值计算求解器

夏青（华为技术有限公司） 2026 年 11 月 28 日 16:30–16:45

数值计算求解器是现代工业软件的核心引擎，是连接数学模型与工程应用的关键桥梁，广泛服务于计算机辅助工程（CAE）、物理仿真、优化设计与控制系统等科学计算领域。当前数值求解技术主要涵盖线性方程组求解、非线性方程组求解与特征值问题求解三大核心方向，其技术能力直接决定了多物理场耦合仿真、大规模稀疏矩阵计算等复杂工程问题的求解精度与效率。然而，在实际工业场景的规模化应用中，该领域仍面临一系列严峻挑战：复杂物理模型的高保真离散与高效求解

之间的深层矛盾、大规模乃至超大规模问题带来的计算性能瓶颈、极端工况下算法的数值稳定性控制，以及求解策略灵活性与系统集成复杂度之间的平衡等，这些问题严重制约了数值求解器在工业级科学计算场景中的广泛落地。面对上述挑战，华为天筹数值计算求解器系统性融合超参数优化、求解策略自适应嵌入与 AI 辅助算法发现等前沿技术，在稀疏线性系统直接法与迭代法求解、非线性方程组高性能求解以及大规模特征值分析等核心能力上取得了重要突破。通过自适应算法选择机制、混合精度计算策略与异构并行加速框架等关键技术的深度协同，天筹求解器在流场仿真、结构力学分析、散热设计等典型科学计算场景中展现出卓越的计算效率与数值鲁棒性。本报告将系统介绍天筹数值计算求解器的技术架构、核心算法设计及其关键实现方案，并对未来超大规模科学计算场景下的求解效率突破、AI 与传统数值算法的协同增强以及云化平台化部署等关键演进方向进行展望。

Learning HTL-ORAS Algorithmic Components and Neural Solvers for ADR and Linear Elasticity Problems

汪震（湘潭大学） 2026 年 11 月 28 日 16:45–17:00

Deep learning offers new opportunities to improve the efficiency of iterative solvers for large-scale partial differential equations (PDEs). This talk presents learning-enhanced optimized restricted additive Schwarz (ORAS) and hybrid two-level ORAS (HTL-ORAS) methods for finite element discretizations of advection–diffusion–reaction (ADR) and three-dimensional linear elasticity problems.

First, a neural network is developed to learn an optimized interface transmission parameter for ORAS. Numerical experiments demonstrate that the learned parameter significantly accelerates both the ORAS iteration and its preconditioned flexible generalized minimal residual (FGMRES) solver, with good generalization across variations in physical parameters. Second, lightweight neural networks are introduced to learn reduced coarse spaces and restriction operators for HTL-ORAS. Two strategies are investigated: one constructs basis functions with prescribed subdomain supports, while the other learns a small number of dense global basis functions. Numerical results show that the resulting two-level solvers substantially improve convergence over the learned one-level ORAS method while reducing the coarse-space dimension and computational cost compared with the ideal-coarse-space approach. These results highlight the potential of learning-based techniques to enhance the efficiency of domain decomposition methods for large-scale PDE systems.

Data-Enhanced Preconditioning for a Sequence of Linear Systems

王菁原（澳门大学） 2026 年 11 月 28 日 17:00–17:15

We develop a runtime data enhanced preconditioner for accelerating the solution of a sequence of linear systems. The key idea is to reuse information generated during previous iterative solves to identify slowly converging features that are not effectively treated by an existing preconditioner. These features are then used to construct a low-dimensional enhancement to the base preconditioner. The learned enhancement can be reused across related systems and updated when needed, making the approach well suited to applications arising from time-dependent, stochastic, and other parameter-dependent PDE problems. Numerical experiments on several PDE models show that the proposed method can reduce both the number of Krylov iterations and the overall computing time, particularly when the original preconditioner is only moderately effective.

大批量矩阵计算智能算法及其应用

王川（香港浸会大学） 2026 年 11 月 28 日 17:15–17:30

大批量矩阵运算（如矩阵求逆和奇异值分解）是科学与工程领域中的基础计算任务，并广泛存在于无线通信、信号处理和科学计算等实际应用中。传统数值方法通常对批量中的每个矩阵进行独立求解，忽略了不同矩阵之间潜在的相关性，从而产生大量重复计算。事实上，许多实际数据在批量维度上具有明显的光滑性和低秩性，为减少冗余计算提供了可能。针对这一问题，我们提出一种大批量矩阵计算框架，利用矩阵运算结果之间的光滑性和低秩性，构建低秩连续表示，从而以少量基础矩阵运算快速得到大批量矩阵的计算结果，显著减少重复求解。实验结果表明，在 DeepMIMO 真实无线信道矩阵上，所提方法在满足精度需求的同时能够显著降低大批量矩阵计算时间；在矩阵求逆、SVD、EVD、矩阵函数及矩阵方程等任务中，相较于 NumPy、SciPy 中的经典数值实现可取得一个数量级以上的总体加速。

混合精度预条件下基于 QR 分解的块共轭梯度法舍入误差分析

苏怡丹（北京邮电大学） 2026 年 11 月 28 日 17:30–17:45

针对具有多个右端项的对称正定线性方程组，块共轭梯度法可以在外层迭代保持较高精度的同时，采用低精度预条件以降低计算开销。本报告研究基于 QR 分解的块共轭梯度法在这一混合精度框架下的舍入误差传播。通过建立加权误差恒等式，推导同时考虑预条件误差与外层迭代舍入误差的局部前向误差界；在适当的收缩性与局部误差假设下，进一步建立误差递推关系，并给出指定迭代步数内的前向误差上界。针对混合精度迭代块 Jacobi 预条件，推导显式的预条件误差界，并将其纳入上述分析框架。此外，将分析推广至经典预条件共轭梯度法。最后，通过 CPU 和 GPU 平台上的数值实验，考察不同内层计算精度对误差传播及求解精度的影响，展示理论界的适用性与保守性。

For symmetric positive definite linear systems with multiple right-hand sides, block conjugate gradient methods can employ low-precision preconditioning to reduce computational cost while retaining higher precision in the outer iteration. This talk examines rounding error propagation in a QR-based block conjugate gradient method within this mixed-precision framework. Starting from a weighted error identity, we derive a local forward error bound that accounts for both preconditioning errors and rounding errors in the outer iteration. Under suitable contraction and local error assumptions, we establish an error recurrence and an upper bound on the forward error over a prescribed number of iterations. For mixed-precision iterative block Jacobi preconditioning, we derive an explicit preconditioning error bound and incorporate it into the analysis. We also extend the framework to the classical preconditioned conjugate gradient method. Numerical experiments on CPU and GPU platforms examine the effects of different inner precisions on error propagation and solution accuracy, illustrating both the applicability and the conservatism of the theoretical bounds.

报告摘要

11 月 29 日报告摘要

HiSLU：面向动态稀疏线性系统的高性能增量求解器

宋健飞（浙江大学） 2026 年 11 月 29 日 08:30–08:45

科学计算与工程仿真中，常需反复求解一系列数值及稀疏结构发生变化的线性系统。传统稀疏直接求解器在每轮求解中重新执行完整流程，难以充分利用相邻系统之间的关联，产生大量重复计算。本报告介绍一种面向动态稀疏线性系统的高性能增量求解器，通过分析矩阵变化在分隔树与消去树中的依赖关系，确定各阶段的更新范围，实现重排序、符号分解和数值分解的增量更新。该求解器支持矩阵数值变化以及节点、边的增删，通过复用未受影响的结构与分解结果，在保持直接求解精度的同时降低重复求解开销。

Space-time Parallel Methods for Wave Propagation

宋博（西北工业大学） 2026 年 11 月 29 日 08:45–09:00

As high-performance computing architectures continue their evolution toward many-core designs, the efficient utilization of substantial parallelism through scalable algorithms becomes increasingly imperative. In this talk, a novel class of space-time parallel methods is proposed that layers time parallelization together with a parallel spatial solver for linear and nonlinear wave equations. The parallel spatial solver utilizes nonoverlapping domain decomposition to divide a spatial domain into multiple subdomains, and applies a parallel Schwarz waveform relaxation (SWR) iteration relied on the first order absorbing boundary conditions to find consistent solutions. The parallel time integrator employed belongs to the family of parallel exponential propagation (ParaExp), which allows for the efficient computation of solutions to hyperbolic problems. Two established algorithms are combined in our developed space-time algorithms called ParaExp-SWR, to add parallel scalability in time. It will be shown that the resulting ParaExp-SWR iterations can be analyzed using standard SWR analysis based on energy estimates, and that the ParaExp-SWR iterations will converge at approximately the same rate as the SWR iterations. Furthermore, compared to the SWR solvers, the ParaExp-SWR methods enhance parallel scaling properties in time. Numerical examples including validation models and Maxwell's equations demonstrate the convergence performance, parallel speedup and efficiency of our methods.

A Learnable Multigrid Framework via Graph Convolutions for Solving PDEs

罗力（澳门大学） 2026 年 11 月 29 日 09:00–09:15

This paper presents a novel framework that integrates learnable graph convolutions with the geometric multigrid method for solving partial differential equations (PDEs). The

discretization of PDEs is first represented as a graph structure, enabling the application of graph convolutions to enhance the multigrid performance. By incorporating graph convolutions into the multigrid components such as smoothing and inter-grid transfer operators, we develop a learnable multigrid that can adaptively optimize its performance based on the underlying problem characteristics. In this framework, the graph convolutions are embedded directly within the multigrid cycle, effectively transforming the entire multigrid solver into a specialized neural network architecture, rather than combining a classical solver with surrogate models. The learnable multigrid framework is lightweight in terms of parameter count. Numerical experiments demonstrate the effectiveness of the proposed approach in solving some challenging problems,

showing improved convergence rates compared to traditional multigrid methods. The generalizability of the learned parameters across different problem settings, including varying source terms, coefficients, geometries, and mesh sizes, is also investigated with proper weight-sharing and transfer-learning strategies.

LinglongAMG：面向 GPU 的内存有效代数多重网格求解器

李文瑄（中国石油大学（北京）） 2026 年 11 月 29 日 09:15–09:30

LinglongAMG 是一种面向 GPU 的内存高效代数多重网格（AMG）求解器，可用于求解大型稀疏线性方程组或构建预条件器。针对 AMG 构建（Setup）阶段细粒度 kernel 过多、中间数组反复读写、片上数据局部性不足，以及 Galerkin 三矩阵乘积中间矩阵存储和哈希操作开销较大的问题，LinglongAMG 融合满足依赖约束的 kernel，减少冗余的全局内存传输；设计利用共享内存和寄存器复用的数据结构，增强片上数据局部性；采用一步式 Galerkin 乘积，避免显式存储中间矩阵，并减少哈希操作。此外，LinglongAMG 支持通过 NVLink 互连的多 GPU 执行。

高性能预条件子 HIPO 及其应用

刘建军（中国科学院数学与系统科学研究院） 2026 年 11 月 29 日 09:30–09:45

HIPO 是一个面向分布式异构并行计算的预条件框架。该框架基于异构编程模型，兼容海光、沐曦等国产 GPU 平台，并依托 MPI/NCCL 实现跨节点分布式计算。HIPO 支持任意精度、任意类型的实复矩阵的线性方程组求解及特征值计算，同时提供混合精度运算功能。HIPO 内置了经典型与聚合型两大类 AMG 预条件子，并支持用户根据应用层特征灵活设计自定义预条件算子。报告最后介绍 HIPO 在复双曲猜想计算机验证和轨迹优化方面的应用。

从不规则到缓存友好：LeSpGEMM 的设计与优化

林圣乐（湖南大学） 2026 年 11 月 29 日 09:45–10:00

稀疏矩阵-矩阵乘（SpGEMM）广泛应用于科学计算、图分析和机器学习，但不规则访存导致数据复用不足、累加效率受限。我们提出了一种局部性增强的 SpGEMM 算法（LeSpGEMM）：通过轻量 C-LSH 聚类增强行间局部性，并结合 L2 缓存预算与簇紧致度，在 SIMD 块累加器和哈希累加器之间自适应选择。在三种多核 CPU 平台上的实验表明，LeSpGEMM 相较六种先进实现取得稳定加速，同时保持较低的预处理开销。

ZiweiSolver：面向电磁工业软件的线性系统求解器

李良（电子科技大学） 2026 年 11 月 29 日 10:30–10:45

电磁 CAE 工业软件中常需要求解结构复杂的线性系统，其求解效率、内存开销与数值可靠性等直接影响工业软件的性能，是电磁工业软件的核心模块。本报告介绍团队开发的 ZiweiSolver(紫微矩阵求解器)，针对电磁计算产生的不同性态线性方程组，系统性地开发了多波前直接求解器(加入了块低秩压缩和蝴蝶形压缩策略)，代数多重网格(AMG)算法，代数区域分解(ADDM)算法以及各种适合的 Krylov 子空间迭代算法等。目前实现了基于 OpenMP 的共享内存并行版本，大部分中大规模及以上场景，开发的求解器在内存消耗和 CPU 时间上超越了 MKL 库。未来一方面持续设计开发更高效的求解算法，另一方面将针对国产硬件、GPU 及分布式存储等使用情景进行加速。

基于 MRRR 三对角特征值求解器的高性能批量对称矩阵特征值分解

靳喜（成都信息工程大学） 2026 年 11 月 29 日 10:45–11:00

对称矩阵特征值分解（EVD）是许多科学计算和数据密集型应用中的基础计算内核，这些应用通常需要批量求解大量相互独立的中小规模问题。现有 GPU 批量求解器主要采用 Jacobi 迭代，或先

进行三对角化，再使用分治法（D&C）或隐式移位 QL 算法。Jacobi 需要反复扫描稠密矩阵，D&C 涉及递归特征向量合并和大量数据搬移，而 QL 中旋转的顺序依赖则限制了并行性。为解决这些限制，本文提出一种面向 GPU 的批量对称 EVD 求解器，将规模自适应 Householder 三对角化与基于多重相对鲁棒表示（MRRR）的批量三对角特征对求解器相结合。三对角化阶段对小矩阵采用片上融合的非分块核函数，对较大矩阵采用面板分块方法。批量 MRRR 利用矩阵间、特征值间和特征向量间的并行性，并采用单层 n 路多分法与片上融合初始化。据我们所知，这是首次将 GPU 批量 MRRR 集成到端到端稠密对称 EVD 求解器中。在 NVIDIA H100 PCIe GPU 上，本文求解器相较于 cusolverDnXsyevBatched、cusolverDnDsyevjBatched 和 EigenG-Batched 最高加速比分别达到 3.59×、14.73× 和 130.66×，同时保持可靠的数值精度。

高可扩展的高阶模板问题多重网格求解器

贾朝蓬（清华大学） 2026 年 11 月 29 日 11:00–11:15

在量子化学、流体力学等领域的偏微分方程数值求解中，高阶有限差分离散能提升模拟精度，但会生成高阶星型模板，如半径为 6 的 37 点模板。现有结构化多重网格库缺乏高阶模板支持，而通用多重网格又易引发粗层算子膨胀，制约大规模扩展。本工作提出面向高阶模板的结构化多重网格求解器：推导聚合粗化下的非零模式演化规律，提出以非光滑聚合为主的粗化策略；在可提供近似模板时，设计近似算子预条件机制，支持预条件子使用近似模板；针对系数具有平移不变性的场景引入无矩阵格式，并结合 LX2 处理器设计缓存分块、SVE 滑动窗口向量化及向量-标量协同优化。在“灵晟”超算上的千万核级的密度泛函理论泊松方程测试中，从 8192 进程扩展至 327,680 进程时，求解阶段弱可扩展效率达 48.8%，相比领域特定基线获得 2.42x 加速。

TRIDENT: An End-to-End Streaming Accelerator for TriSpGEMM.

胡议月（国防科技大学） 2026 年 11 月 29 日 11:15–11:30

Triple Sparse General Matrix-Matrix Multiplication (TriSpGEMM) is a major performance bottleneck in scientific computing and graph analytics. Existing domain-specific accelerators support only Sparse General Matrix-Matrix Multiplication (SpGEMM). Performing TriSpGEMM by sequencing two SpGEMM kernels introduces a global synchronization barrier, forces intermediate data to be written to and read from off-chip memory, and leads to frequent pipeline stalls, all of which significantly degrade performance. A natural solution is to design a fused architecture that executes TriSpGEMM directly, but such a design must overcome three key challenges, i.e., pipeline overflow caused by highly irregular and bursty sparse dataflows, locality inefficiency stemming from the trade-off between input reuse and output aggregation, and severe head-of-line blocking due to variable memory access latencies. In this work, we present TRIDENT, a transient-data-centric accelerator for fused TriSpGEMM processing. TRIDENT integrates a hierarchical flow-control mechanism, a hybrid dataflow co-designed with a windowed sparse format, and an out-of-order decoupled execution scheme to ensure pipeline stability, locality efficiency, and latency tolerance. Comprehensive evaluations demonstrate that TRIDENT achieves an average of 3.56× performance speedup over state-of-the-art SpGEMM accelerators, and up to 49.6× and 6.3× improvements over CPU and GPU baselines, respectively.

JXPAMG 解法器进展

黄思路（北京应用物理与计算数学研究所） 2026 年 11 月 29 日 11:30–11:45

作为求解大规模稀疏线性方程组的核心算法之一，代数多重网格（AMG）在科学计算与工程仿真领域具有重要应用价值。JXPAMG 是一个自主研发的并行代数多重网格解法器，已在激光聚变、

复杂流动、结构力学、电子学系统等实际应用领域广泛使用。本报告介绍 JXPAMG 在算法研发、性能优化及应用方面的最新进展。

高温气冷堆耦合计算 TaiHang 软件的解法器进展

张汉（清华大学） 2026 年 11 月 29 日 13:30–13:45

高温气冷堆是一个复杂的多物理、多尺度、多系统耦合的复杂核动力系统，对耦合求解的稳定性、精度与计算效率均有严苛要求，基于 Newton-Krylov 方法是最具前景的耦合计算框架。此外，工程计算中，模型几何简化、物理本构的不确定性，要求利用实测数据标定优化关键参数；反应堆运行阶段的在线状态实时跟踪需求，也为求解器带来新挑战。本报告介绍课题组高温气冷堆耦合计算程序 TAIHANG 的最新进展、验证结果与工程应用，以及在 NK 框架下进一步开发 3D-Var 数据同化模块和降阶模型模块的思路和初步结果。

A Massively Parallel Three-grid Preconditioner for the High-frequency Helmholtz Equation

付书彬（宁波东方理工大学） 2026 年 11 月 29 日 13:45–14:00

Accurate simulation of three-dimensional time-harmonic wave propagation over many wavelengths requires control of phase error and efficient solution of large indefinite systems. We develop a three-grid solver based on the compact 27-point interpolated optimized finite-difference (IOFD) discretization. Its wavenumber-dependent stencil supports a fine-grid resolution of six points per shortest wavelength and an unshifted physical correction on the 2h grid at only three points per shortest wavelength. The method retains unshifted IOFD operators on the h and 2h grids, while a complex-shifted 2h–4h auxiliary cycle preconditions a factorization-free iterative approximation of the coarse inverse. Restricting the shift to this auxiliary cycle preserves the propagative character of the coarse correction. Comparison with the outgoing Green function confirms phase and relative-amplitude accuracy on a sequence of meshes up to 10240^3, exceeding one trillion unknowns, with the largest problem spanning approximately 1704 wavelengths per coordinate. The same fixed solver configuration retains robust convergence across smooth, discontinuous, high-contrast, and geophysical velocity models and exhibits scalable parallel performance. In particular, a 2560^3 problem spanning approximately 425 wavelengths in each coordinate direction is solved in 43.8 seconds on just 64 NVIDIA A100 GPUs.

CFUA-AMG: 基于粗细变量分裂的非光滑聚集并行代数多重网格算法

范智博（北京应用物理与计算数学研究所） 2026 年 11 月 29 日 14:00–14:15

针对大规模稀疏线性系统的求解，我们提出了一种基于粗细变量分裂的非光滑聚集型代数多重网格算法 CFUA-AMG，其主要思想是利用经典 AMG 粗化过程的粗细变量分裂（即 C/F 分裂）识别强连接结构，指导传统 UA-AMG 聚集根节点的选择和聚集的形成，提升聚集粗空间的质量，改善 UA-AMG 算法的收敛性。结合经典 AMG 理论分析框架，我们证明了 CFUA-AMG 算法对于行和非负的对称正定 M-矩阵，其两层网格收敛因子与矩阵规模无关。我们将 CFUA-AMG 集成到自主 AMG 解法器 JXPAMG 中，通过几类典型算例验证了算法的有效性。与现有常用聚集算法相比，对模型问题算例，平均可以获得 1.29 倍的加速比；对激光聚变应用中的辐射扩散方程算例，平均可以获得 2.04 倍的加速比；对复杂流动应用中的不可压 Navier-Stokes 方程算例，平均可以获得 1.94 倍的加速比。CFUA-AMG 在三类算例中的迭代次数稳定，表现出良好的收敛健壮性和问题适应性。

稀疏线性代数方程组随机算法：现状与差距

杜皓（北京应用物理与计算数学研究所） 2026 年 11 月 29 日 14:15–14:30

随机算法近年来发展迅速，展现出实际应用潜力，为大规模线性代数方程组的求解提供了新的技术途径。本报告汇总稀疏线性代数方程组随机算法的研究现状，对当前主要的三类随机算法进行了总结，分析了这些算法的特点、计算复杂度和面临的问题。在此基础上，本报告对当前随机算法的研究现状进行了评估，指出了离实际应用需求的差距，对面向大规模实际应用的随机算法研究进行了展望。

A Robust Alternative Minimum Residual Method for Complex Symmetric Linear Systems

丁彦灵（西南财经大学） 2026 年 11 月 29 日 14:30–14:45

For solving indefinite complex symmetric linear systems, existing Krylov subspace solvers such as COCG and COCR often suffer from severe numerical instability and true residual divergence. In this paper, we propose an alternative minimum residual method, termed cs_MrR, based on Rutishauser's coupled two-term recurrences. By directly updating the solution via coupled displacement steps, cs.MrR avoids the explicit generation of direction vectors. We prove that cs.MrR is mathematically equivalent to cs.MRTR, COCR, and BiCGCR in exact arithmetic. In finite-precision arithmetic, cs_MrR effectively mitigates catastrophic cancellation and reliably keeps the true residual at the minimal level. Numerical experiments demonstrate the superior stability and robustness of cs.MrR over existing methods.

代数多重网格求解器的自动精度选择

陈骏豪（北京邮电大学） 2026 年 11 月 29 日 14:45–15:00

混合精度计算有望降低代数多重网格（AMG）的存储与计算开销，但精度过低也可能导致收敛变慢或求解失败。不同问题和多重网格层次结构对精度的要求有所不同，精度配置需要结合具体的数值性质和求解要求确定。本报告讨论 AMG 中的自动精度选择问题。

我们结合理论分析与数值实验，考察不同存储精度和运算精度对 AMG 收敛行为的影响，讨论舍入误差、浮点数表示范围及残差缩放的作用。在此基础上，我们介绍自动精度选择框架的设计思路，讨论如何利用求解前可获得的问题特征和数值分析信息选择精度配置。报告结合已有算例说明相关数值现象，并讨论精度选择中仍需解决的问题。

What Frontier LLMs Can (and Cannot) Optimize in GPU SpMV

程贺琳（中国石油大学（北京）） 2026 年 11 月 29 日 15:00–15:15

随着大模型逐步融入代码开发流程，用户可以通过网页对话或 Agent 工具，以自然语言提出需求，由模型自主选择优化策略，完成代码编写、测试与迭代。然而，对于性能高度依赖数据特征和硬件行为的稀疏算子，大模型的优化能力究竟能够达到何种水平、性能收益主要来自哪些技术层面，仍有待系统研究。我们以稀疏矩阵向量乘法（SpMV）这一解法器关键算子为切入点，研究前沿大模型优化 GPU 代码的能力及其边界。围绕内核调优、工作映射和存储格式三个层次，我们构建了结合性能测量与源码分析的评估框架，在 RTX 5090 和 A100 上，选取 22 个 SuiteSparse 矩阵开展八项实验及四项补充分析，涵盖开放式优化、迭代反馈、分层约束、格式探索和成熟方法改进等场景。结果表明，大模型能够通过内核调优和工作映射实现性能提升，部分生成实现的性能达到甚至超过成熟方法。然而，即使明确要求设计新的存储格式，所分析的实现仍主要复用或组合现有技术，尚未形成具有实质创新的存储格式核心。生成条件与模型配置会影响具体性能表现，但未能稳定突破上述能力边界。基于这些发现，我们建议算子开发者与调用者充分利用大模型辅助代码实

现、性能调优和受约束的策略搜索，而将新存储格式设计及突破性技术创新交由领域专家主导。对于解法器开发者，还需进一步验证算子层面的性能提升能否转化为端到端应用收益。

PANA：面向多核 CPU 的 SpMV 细粒度自适应负载均衡设计

边浩东（青海大学） 2026 年 11 月 29 日 15:15–15:30

稀疏矩阵向量乘（SpMV）作为各类科学与工程计算中的核心算子，其并行性能高度依赖于矩阵稀疏结构与硬件架构特性。然而，现有的静态划分策略通常仅从计算或访存的单一视角切分任务，无法应对非规则计算中的复杂负载失衡问题，严重制约了 SpMV 在现代多核 CPU 上的性能提升。本报告介绍了我们在并行 SpMV 动态负载均衡方面的最新工作——PANA。该工作在微指令级别实现了细粒度的任务切分与自适应调度。一方面，该机制全面考量各核心处理的微指令类型与数量，突破了传统方法仅依赖简单计算或访存统计的局限，实现了对计算与访存的权衡；另一方面，利用各核心执行时间的实时反馈指导任务重分配，兼具良好的跨平台可移植性与出色的负载均衡效果。实验结果表明，PANA 在主流多核 CPU 平台上实现了高效的动态负载均衡，性能显著超越现有的代表性并行 SpMV 方案。

A Shell-Informed Smoothed Aggregation Multigrid Preconditioner for Linear Systems from Reissner–Mindlin Shells

安恒斌（北京应用物理与计算数学研究所） 2026 年 11 月 29 日 16:00–16:15

Discretizations of Reissner–Mindlin shells produce large ill-conditioned linear systems where standard AMG often fails. We develop a shell-informed smoothed aggregation AMG (SI-SAMG) preconditioner. By deriving the six zero-strain modes from shell kinematics and embedding them into the tentative prolongator using aggregate-wise QR factorizations, SI-SAMG effectively handles the shell's low-energy modes. We prove that the condition number of the preconditioned system is independent of the finest mesh size. Numerical experiments demonstrate its superior robustness over standard AMG methods and satisfactory parallel strong scaling.

SOLVER26 参会人员

姓名	单位
安恒斌	北京应用物理与计算数学研究所
安泽田	湘潭大学
敖渝焱	湘潭大学
边浩东	青海大学
蔡小川	澳门大学
曾宏伟	中国石油大学（北京）
曾柳铭	深圳市大数据研究院
常高	山西科技学院
陈槟	中国科学院深圳先进技术研究院
陈骏豪	北京邮电大学
陈明卿	浙江师范大学
陈启明	中国科学院深圳先进技术研究院
陈荣亮	中国科学院深圳先进技术研究院
陈惜源	湘潭大学
陈永田	国防科技大学
陈泽童	中国科学院深圳先进技术研究院
陈周纬宇	湘潭大学
谌骞骞	中国科学院深圳先进技术研究院
程鼎博	电子科技大学
程贺琳	中国石油大学（北京）
程载恒	中国科学院深圳先进技术研究院
崔涛	中国科学院数学与系统科学研究院
代炜琪	北京应用物理与计算数学研究所
戴越旺	国防科技大学计算机学院
单瑞彤	吉林大学
单钰鑫	中国科学院深圳先进技术研究院
邓丽萍	湖南大学
丁肇辉	华为技术有限公司
丁彦灵	西南财经大学
杜皓	北京应用物理与计算数学研究所
范民涛	四川大学数学学院
范智博	北京应用物理与计算数学研究所
方小中	湖南大学
费玲	中国科学院深圳先进技术研究院
冯春生	湘潭大学
冯婷婷	国防科技大学
付书彬	宁波东方理工大学
高振	中国海洋大学
高紫嫣	湘潭大学
龚世华	香港中文大学（深圳）
顾芸珏	香港中文大学
郭诚欣	中国科学院计算技术研究所

姓名	单位
郭昭彤	中国科学院数学与系统科学研究院
韩志龙	电子科技大学
何家希	同济大学数学科学学院
何米禾	北京应用物理与计算数学研究所
何少则	是石科技
胡彬辉	电子科技大学
胡靖超	浙江大学
胡静颖	中国科学院深圳先进技术研究院
胡少亮	北京应用物理与计算数学研究所
胡议月	国防科技大学
黄成雪	国防科技大学
黄丹晨	中国科学院数学与系统科学研究院
黄东强	国家超级计算无锡中心
黄嘉	厦门大学数学科学学院
黄顺森	香港中文大学（深圳）
黄思路	北京应用物理与计算数学研究所
季威	深圳蜂巢云泊软件技术有限公司
贾朝蓬	清华大学
贾海鹏	中国科学院计算技术研究所
江颖	中山大学
蒋廷桢	华为技术有限公司
蒋毅	中国科学院深圳先进技术研究院
金晨浩	中国科学院数学与系统科学研究院计算数学与科学工程计算研究所
金颖	国防科技大学
金洲	浙江大学
靳喜	成都信息工程大学
阚宇剑	四川大学
康豪	电子科技大学
柯艺芬	福建师范大学
乐嘉诚	湘潭大学
李家伟	国防科技大学
李静萱	电子科技大学
李凯	国防科技大学
李坤	西南财经大学
李良	电子科技大学
李琪	湘潭大学
李胜国	国防科技大学
李世顺	信阳师范大学
李思团	中国科学院数学与系统科学研究院
李维鑫	电子科技大学
李文瑄	中国石油大学（北京）
李易达	中国石油大学（北京）
李正光	吉林大学
郦奕宸	中国科学院深圳先进技术研究院

姓名	单位
梁戊佳	中国石油大学(北京)
林茂	北京大学重庆大数据研究院
林圣乐	湖南大学
林学磊	哈尔滨工业大学（深圳）
林增	中国科学院深圳先进技术研究院
刘成庆	青海大学
刘定干	四川大学
刘继伟	湖南大学
刘建界	国防科技大学计算机学院
刘建军	中国科学院数学与系统科学研究院
刘芮华	湖南大学
刘盛	电子科技大学
刘帅	湘潭大学
刘伟峰	中国石油大学（北京）
刘研	国防科技大学
刘雅琪	中科曙光集团
刘毅文	湖南大学
刘颖智	中国科学院深圳先进技术研究院
刘云	湘潭大学
龙胜男	中科曙光集团
卢欣	中国石油大学（北京）
卢玥辰	中国石油大学（北京）
陆棋枫	清华大学
罗传泰	电子科技大学
罗力	澳门大学
罗湘然	国防科技大学
罗晓轲	北京应用物理与计算数学研究所
吕小敬	中国船舶科学研究中心
马楚鹏	大湾区大学
马嘉豪	深圳大数据研究院
毛润彰	北京应用物理与计算数学研究所
聂存云	湖南工程学院
欧瑜	中国科学院数学与系统科学研究院
泮忆铭	北京应用物理与计算数学研究所
彭浩	电子科技大学
彭洁	衡阳师范学院
戚瑞轩	湖南大学
钱敞	华为技术有限公司
邱昊中	国防科技大学
邱汝康	湖南大学
任宇康	电子科技大学
戎梦飞	中国石油大学（北京）
沈瑞刚	桂林电子科技大学
石清元	新加坡国立大学

姓名	单位
舒适	湘潭大学
宋博	西北工业大学
宋家乐	中国石油大学（北京）
宋健飞	浙江大学
苏怡丹	北京邮电大学
孙海冰	电子科技大学
孙景昱	北京应用物理与计算数学研究所
孙祥	中国海洋大学
覃善林	中国科学院深圳先进技术研究院
唐鸣	桂林电子科技大学
唐仕兵	江苏师范大学
滕子珺	复旦大学
田甜	香港中文大学（深圳）
田智鲲	湖南工程学院
童嘉博	澳门科技大学
涂祥琦	中国科学院深圳先进技术研究院
涂学民	美国堪萨斯大学
万嘉文	中国科学院深圳先进技术研究院
万瑞兴	电子科技大学
汪宏宇	电子科技大学
汪云婷	中国科学院计算技术研究所
汪震	湘潭大学
王川	香港浸会大学
王纯欣	电子科技大学
王达	中国科学院深圳先进技术研究院
王建春	中国船舶科学研究中心
王龙	华为技术有限公司
王梦麒	中科曙光集团
王菁原	澳门大学
王俊仙	湘潭大学
王磊	桂林电子科技大学
王淞	清华四川能源互联网研究院
王向阳	国防科技大学
王雪	山东大学
王野	上海思朗科技股份有限公司
王运涛	中国科学院计算技术研究所
王真淦	北京邮电大学
王震	中国科学院数学与系统科学研究院
王子林	青海大学
王梓涵	北京应用物理与计算数学研究所
温瑞萍	太原师范学院
吴蕊利	北京应用物理与计算数学研究所
伍湘君	中国气象局地球系统数值预报中心
伍星	天津师范大学

姓名	单位
武辉	湘潭大学
夏晨阳	深圳蜂巢互联
夏佳硕	中国科学院数学与系统科学研究院
夏建林	南京大学
夏青	华为技术有限公司
肖映雄	湘潭大学
谢富伟	湘潭大学
谢和虎	中国科学院数学与系统科学研究院
辛起	香港中文大学（深圳）
熊子慷	湘潭大学数学与计算科学学院
徐飞	北京工业大学
徐嘉祁	中国科学院深圳先进技术研究院
徐磊	中国科学院深圳先进技术研究院
徐盼盼	宁夏大学
徐童垚	国防科技大学
徐小文	北京应用物理与计算数学研究所
徐昕	浙江大学
许进超	香港中文大学（深圳）
许秋燕	宁夏大学
薛凤	青海大学
薛巍	清华大学
闫森	中国科学院深圳先进技术研究院
闫争争	中国科学院深圳先进技术研究院
阳王东	长沙学院
杨晨琰	国防科技大学
杨晟业	国防科技大学
杨海鹏	电子科技大学
杨荣华	湘潭大学
杨少峰	中国科学院计算技术研究所
杨爽	中国空气动力研究与发展中心计算空气动力研究所
杨肖剑	国防科技大学
杨峥	中国科学院深圳先进技术研究院
叶常青	北京应用物理与计算数学研究所
叶国烽	中科曙光集团
叶欣瑶	西南财经大学
于沛楠	清华大学
于悦	香港中文大学（深圳）
余愿	湘潭大学
袁璠	国防科技大学
张晨松	中国科学院数学与系统科学研究院
张丛庆	湘潭大学
张峻玮	吉林大学数学学院
张力月	北京邮电大学
张俐	电子科技大学

姓名	单位
张梦晨	福州大学
张庆阳	国防科技大学
张硕	中国科学院数学与系统科学研究院
张思佳	吉林大学数学学院
张思哲	上海思朗科技股份有限公司
张文博	福建师范大学
张烨	中国科学院深圳先进技术研究院
张莹桥	中国空气动力研究与发展中心 计算空气动力研究所
张玉印	电子科技大学
张悦嘉	复旦大学
张子怡	中国石油大学（北京）
张祖铭	北京应用物理与计算数学研究所
赵芮锋	北京邮电大学
赵瑞峰	思朗科技
赵新月	上海财经大学数学学院
郑成瑞	四川大学
郑晋凯	深圳市大数据研究院
郑荣振	湘潭大学
郑世良	电子科技大学
钟柳强	华南师范大学
钟宇豪	电子科技大学
周翰林	中国科学院深圳先进技术研究院
周泓泽	国防科技大学计算机学院
周立平	湖南科技学院
周明媚	北京大学重庆大数据研究院
周翔	深圳市大数据研究院
周雪梅	桂林电子科技大学
周振亚	北京华大九天科技股份有限公司
朱力杰	厦门大学数学科学学院
朱婉婉	广西师范大学
朱芷莹	浙江大学
朱子睿	湘潭大学
邹海峰	深圳市大数据研究院
邹航	国防科技大学
Luca Pavarino	意大利帕维亚大学

中国科学院深圳先进技术研究院介绍

中国科学院深圳先进技术研究院（简称“深圳先进院”）由中国科学院、深圳市人民政府和香港中文大学于 2006 年 2 月共同建立。作为国家战略科技力量主力军，深圳先进院始终锚定国家需求、科技前沿和产业发展的最大公约数，着力打通从科技制高点到科技成果转化融合点的“两点一线”，深化与国际学术体系及粤港澳大湾区产业生态对接，逐步走出一条紧扣国家战略、适配湾区发展，以多学科交叉为特色、科教产才“四位一体”融合发展的创新路径，已建设成为国际上有影响力、国内有地位、地方有贡献的国家科研机构。

深圳先进院聚焦“四个率先”和“两加快一努力”目标任务，围绕抢占科技制高点策划实施重大攻坚任务，充分发挥体系化建制化优势，围绕“IT+BT+MT”形成多学科交叉融合研究体系，已建设先进集成技术研究所、生物医学与健康工程研究所、先进计算与数字工程研究所、生物医药与技术研究所、脑认知与脑疾病研究所、合成生物学研究所、先进电子材料研究所、碳中和技术研究所、科学仪器研究所（筹）、材料人工智能研究所等 10 个内设研究机构，以重点领域、重大任务、重大平台为依托，持续增强体系化攻关能力，努力在日益激烈的国际科技竞争中抢占科技制高点。

历经 20 年的奋进发展，深圳先进院人员规模已达 4935 人（含员工 2612 人、学生 2323 人），打造出一支平均年龄 34.65 岁的国际化人才队伍；拥有海归人才 795 人，6 位外籍学者获得中国政府友谊奖，77 位学者入选 2026 全球前 2%顶尖科学家榜单；累计发表论文超 2.6 万篇，申请专利 1.88 万件，专利转化率达 29.79%，累计孵化和培育企业 2406 家，持股企业 397 家；建有 4 个全国重点实验室、2 个国家级创新中心、2 个市级重大科技基础设施、3 个市级基础研究机构、117 个省市创新载体。其中，牵头建设医学成像科学与技术系统全国重点实验室、定量合成生物学全国重点实验室，参与共建集成电路材料全国重点实验室、脑认知与类脑智能全国重点实验室；国家高性能医疗器械创新中心、国家生物制造产业创新中心持续深化协同创新；深圳先进电子材料国际创新研究院、深圳合成生物学创新研究院、深港脑科学创新研究院三大基础研究机构加速突破；深圳合成生物研究、脑解析与脑模拟两项重大科技基础设施已入驻光明科学城并启动运营，有力吸引粤港澳大湾区乃至全球的高端科研资源加速聚集。

立足粤港澳大湾区国际科技创新中心及大湾区综合性国家科学中心建设的历史机遇，深圳先进院秉持“做国家最需要的科研，创世界未见过的技术”的核心价值观，在深圳这片改革开放的前沿热土上持续探索体制机制创新，强化“抢”的意识、“高”的标准，力争建设成为需求导向、综合交叉类的世界一流科研机构，为我国向全球科技开拓者、引领者加速转变，实现高水平科技自立自强和建设科技强国，作出国家战略科技力量主力军应有的重大贡献！

先进计算与数字工程研究所介绍

中国科学院深圳先进技术研究院先进计算与数字工程研究所（以下简称数字所”）为先进院的核心研究单元之一，致力于面向计算机科学和人工智能技术与应用快速发展趋势，构建国际一流的科研团队和平台，在数据智能领域从基础到应用层面做出原创性、前瞻性的重大贡献，培养高层次及科技研发人才，服务粤港澳大湾区建设世界一流科创中心。

数字所拥有 20 余个国家及省市级实验平台，建有国家网格华南节点(中国科学院超算深圳分中心)等计算平台，科研设备累计超过 3000 台(套),拥有高性能 GPU 服务器累计超 100 台，CPU 服务器超 200 台等。

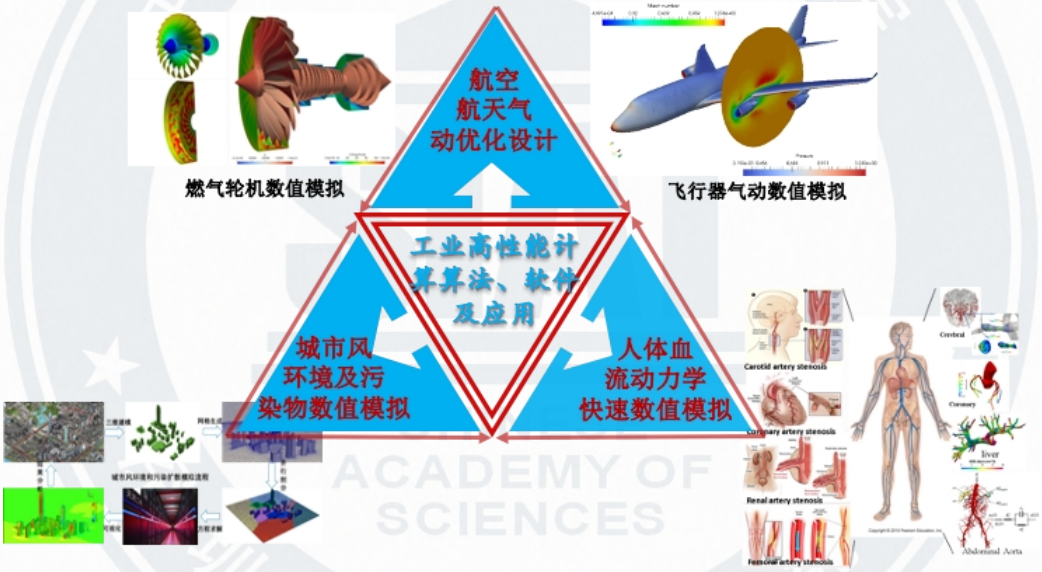
近年来在大数据与云计算、人工智能技术与应用、高性能计算等方向，汇聚了一支海归为主，高水平、国际化的师资队伍，科研队伍 500 余人，与国际诸多著名大学和研究院建立了紧密的学术交流与合作关系，是我国先进计算与数字工程领域对外合作与交流的重要窗口。

工程与科学计算研究室介绍

工程与科学计算实验室成立于 2010 年 2 月，是中国科学院深圳先进技术研究院重点发展的核心研究单元之一。实验室拥有一支由研究员、工程师、博士后、研究生、访问学者组成的高水平专业科研队伍，其中研究员 2 名，副研究员 4 名，助理研究员 3 名，高级工程师 1 名，70%成员具有海外学习和工作经历。研究室成员长期从事高性能并行计算、数值分析、软件工程、医学工程等方面的研究工作，具有坚实的理论基础和应用实践经验，学科交叉性强。

实验室跟踪当前国际最前沿的数值计算方法进展，以高可扩展性并行计算能力为目标，开展工程与科学计算中具有领先意义和应用价值的计算方法及应用研究，为工程与科学计算领域中的多尺度、多领域大规模计算问题提供支撑技术、实现方法和计算软件。主要研究方向涵盖以下领域：大型科学计算中的高可扩展性并行求解方法和软件、航空航天飞行器气动和气弹分析、高速列车和汽车的空气动力学优化设计、人体血流动力学数值模拟、空气污染预报和溯源、高层建筑的高精度结构力学分析。

研究室在这些领域开展了大量的基础算法和应用研究，研发了多套面向工程应用问题的完全具有自主知识产权的软件包，在神威太湖之光、天河二号、曙光星云等多台国产超级计算机上实现了上万核的并行计算测试。同时，研究室先后承担二十余项国家重大重点项目。



图：研究室主要研究方向

香港中文大学（深圳）介绍

香港中文大学（深圳）是一所经教育部批准，参照中外合作办学条例设立的大学。大学以创建一所立足中国、面向世界的一流研究型大学为己任，致力于培养具有国际视野、中华传统和社会担当的创新型高层次人才。大学的办学特色是国际化的氛围、中英并重的教学环境、书院制传统、通识教育、新型交叉学科设置和以学生为本的育人理念。

目前，来自世界各地的 10000 多名优秀学子正在港中大（深圳）求学。自 2014 年成立以来，经过十年的发展，大学学科建设已逐步完善，现有经管学院、理工学院、人文社科学院、数据科学学院、医学院、音乐学院、公共政策学院和人工智能学院八个学院以及一个研究生院。

大学注重国际交流和学生国际视野的培养，目前，港中大（深圳）已与 160 多所境外名校开展实质性交流与合作，遍布世界 30 多个国家和地区，各类国际合作项目超过 290 个。在校学生中具有境外学分学习经历的比例超过 60%，在国内高校中名列前茅。

大学面向全球招聘，已建立一支具国际水平的师资队伍，截至目前，已面向全球招聘引进了 700 多名国际知名优秀学者和研究人员，其中包括诺贝尔奖得主 5 名，各国院士超 50 名。目前引进的教师 100%具有在国际一流大学执教或研究工作经验，是一批具有国际视野、富有创新精神和教书育人热忱的优秀教师。

香港中文大学（深圳）建校以来，以创新驱动为指引，以粤港澳大湾区需求为导向，研究领域契合了国家战略发展需求，涵盖了地方新兴产业和未来产业发展方向，包括人工智能与机器人、大数据与数据科学、信号与信息处理、类脑研究、生物医药、生物信息、先进材料、新型电子元器件、智能电网、新型网络、金融物流等。

梧桐树下听凤鸣，神仙湖畔论古今，香港中文大学（深圳）将立定脚跟，砥砺前行，继续追求卓越的教学与研究水平，为中国培养国际化的创新型领袖人才，致力于发展成为享誉世界的中国一流学府。

湖南韶峰应用数学研究院介绍

湖南韶峰应用数学研究院是湖南国家应用数学中心入驻湘潭国家高新区的独立法人新型研发机构，2021 年 12 月 22 日时任省委常委、湘潭市委书记张迎春等领导揭牌成立。研究院由湖南国家应用数学中心和湘潭高新技术产业开发区、湘潭市大数据和产业创新发展中心、湘潭智城联合信息科技有限公司联合共建，是湖南科技创新体系的重要组成部分，是组织高水平科学研究、搭建产学研融通平台、培养高层次复合型人才、孵化高新技术企业的科技创新平台。研究院 2022 年 8 月入选湖南省新型研发机构，2023 年 4 月入选湖南省制造业数字化转型应用服务商，2025 年获批湖南省博士创新站、湖南省“数字新基建”标志性项目实施单位，2026 年入选湖南省制造业数字化转型人工智能服务商。

研究院拥有国家杰青、国家优青、长江学者、万人计划、湖南省科技领军人才、省杰青等专兼职科研人员 157 人，联合共建了高性能与智能计算中心、智能可信导航与定位湖南省重点实验室、智慧司法与数字治理湖南省重点实验室、政务算力网络湖南省工程研究中心等 12 个科创平台，自主研发了韶峰天工 CAX 智能工作流协作平台、韶峰天玑低空具身数智平台、韶峰天问数智教育平台等 3 个 AI 底座平台，自主建设了北斗高精度融合导航定位技术创新中心、智能制造及工程机械数字化设计创新中心等 6 个创新中心，与湘电集团、润建股份、湖南征信、湖南移动等重点行业企业签署 20 余项战略合作协议，与中国船舶、湖南航天、江南集团、三一集团、华曙高科等 90 余家企事业单位建立了合作关系。

研究院面向国家重大战略需求、面向湖南“4×4”现代化产业体系和 13 条重点产业链发展需要、面向湘潭“三优三特”现代化产业集群，重点开展高性能与智能计算、共性算法设计、通用模型构建、工业软件开发和智能导航应用等方向的科技研发，着力推进算力、算法、算据、算网“四算融合”和政产学研用合作，努力打造集“技术研发、成果转化、科技金融、人才培养、国际合作”五大功能为一体的科技创新平台，赋能区域经济社会高质量发展。



深圳国际工业与应用数学中心介绍

深圳国际工业与应用数学中心于 2021 年 9 月正式启动，2022 年 12 月正式注册成为深圳市大数据研究院福田分院，由中国科学院院士马志明教授担任中心主任，香港中文大学（深圳）王筱平教授担任常务副主任，现已聚集多位国际一流的工业与应用数学人才。中心将瞄准粤港澳大湾区数字经济行业创新的关键问题，围绕人工智能理论、方法及应用，智能工业软件核心求解器开发，科学计算与应用数学的理论及应用三大重点研究领域开展科学研究和技术攻关。中心致力于吸引和聚集国际顶尖人才和团队，打造国际学术品牌，促进应用数学国际合作，并协同工业界面向国家战略解决一些系列核心技术问题。



Ascend

昇腾AI创新大赛2026

非同凡响 我要去闯

算子挑战赛

S10赛季

昇腾AI创新大赛-算子挑战赛
是面向AI开发者打造的赛事，旨在培养一批精通Ascend C算子开发的开发者，
鼓励开发者基于昇腾AI算力、CANN的基础能力进行深度创新与实践，
加速AI与行业融合，促进开发者能力提升。

比赛日程

报名时间：26年10月27日12:00 — 27年2月1日 12:00

作品提交：26年10月27日12:00 — 27年2月1日 12:00

年度总决赛（预计）： 27年5月

奖项设置

奖项	奖项描述	含税金额（人民币）	数量
一等奖	每题排名第1名	1万元	共计5个奖项
二等奖	每题排名第2名	5千元	共计5个奖项
三等奖	每题排名第3名	2千元	共计5个奖项
优秀奖	每题排名第4-10名	1千元	共计35个奖项

扫码立即报名

Kunpeng

鲲鹏创新大赛2026

非同凡响 我要去闯

高性能计算全球挑战赛

S3赛季

鲲鹏创新大赛-高性能计算全球挑战赛
是面向高性能计算行业基于鲲鹏计算平台的并行计算工程挑战赛事，
旨在培养参赛人员的并行计算思想和并行计算系统能力，
推动建立鲲鹏超算新生态，促进开发者能力提升。

比赛日程

报名时间： 26年10月20日12:00-27年2月1日12:00

作品提交： 26年10月30日12:00-27年2月1日12:00

年度总决赛（预计）： 27年5月

奖项设置

奖项	奖项描述	含税金额（人民币）	数量
一等奖	每题排名第1名	50000	2
二等奖	每题排名第2-3名	20000	4
三等奖	每题排名第4-6名	10000	6
优秀奖	每题排名第7-10名	5000	8

扫码立即报名

是石科技介绍

是石科技成立于 2021 年，是国内领先的独立第三方全栈算力服务商。公司聚焦“超智融合”，业务涵盖大规模集群全栈服务与并行优化服务等，可以为科学工程计算、工业仿真和人工智能应用提供异构算力组织、软件适配、性能调优及运营服务。目前，公司运营 10 余个异构算力集群，纳管算力资源 20000+P。

公司研发人员占比超过 70%，核心人员均为高性能计算、人工智能等领域资深专家及清华大学、北京大学、北京航空航天大学等重点高校学者。团队曾参与全球高性能计算应用领域最高奖“戈登·贝尔奖”项目、吴文俊人工智能科技进步特等奖项目及 CCF HPC China 超算最佳应用项目。

是石科技产品包括国产 Token 优化工厂、Meta-Infer高效推理引擎、TokenFleet 分布式调度框架等。国产 Token 优化工厂贯通任务、芯片、模型与集群调度，兼容 10 余种国产算力芯片，适配 20 余个主流模型，日均 Token 吞吐量达万亿级。Meta-Infer 通过内核补齐、算子融合、计算与通信重叠、并行与容量调优及缓存复用提升推理执行效率；TokenFleet面向多模型、多租户和混合负载，开展请求路由、跨机房容量协同与队列治理。

成立以来，公司已服务各行业头部客户 200 余家，覆盖头部大模型公司、头部互联网大厂及航空航天、生物制药、新能源等领域企业，并在高端制造、工业仿真、气候气象、海洋模式、流体力学等场景开展并行优化实践。截至 2026 年，公司累计订单超 30 亿元；成立 5 年来，已连续盈利 4 年。

国产 Token 优化工厂 · 拓元 Vectron

项目介绍

企业级大模型推理生产平台，以模型优化、推理加速与智能调度，将异构 GPU 算力转化为稳定、可计量的 Token 服务，让同等算力产出更多有效 Token，帮助企业降低推理成本、提升服务性能、加快 AI 应用落地。

产品优势



痛点解决

- **成本高且不可控——**按 Token 计费，提升资源利用率，减少 GPU 闲置与高峰溢价
- **相应蛮、高峰期卡——**智能调度与推理加速协同，秒级响应，稳定承载高并发
- **数据安全与合规压力——**支持私有化部署、数据不出域，支持等保、ISO 27001 等合规体系
- **无法定制优化——**开放调度策略与模型参数，按业务场景定制推理方案

场景落地与部署价值

- **科研与高端工业智能化**



场景挑战：高端制造、航空航天、生物医药、新能源及科研院所引入代码辅助优化、科学计算智能体与多模态推理时，面临国产芯片适配难、长上下文显存压力大等瓶颈。

平台价值：向下兼容 10 余款国产芯片、向上适配 20 余个主流模型，支持定制化调优与高稳长上下文推理，打通从底层算力优化到行业/科研 AI 应用的“最后一公里”。

- **金融、政务与医疗等强合规核心业务**



场景挑战：敏感数据要求严格不出域，核心业务对稳定性、自主可控与合规认证门槛高。

平台价值：提供私有化部署与专属集群部署，本地化国产算力自主可控、独享 GPU 资源与独立调度策略，支持等保及 ISO 27001 合规体系，稳定承载高并发秒级响应。

- **高并发弹性业务与轻量快速验证**



场景挑战：编程应用、文生视频、企业智能体等业务流量波峰波谷显著，自建集群易现 GPU 闲置或高峰卡顿溢价。

平台价值：灵活选用混合云架构（本地保敏感数据、云端弹性扩容）或公有云服务（开箱即用、按 Token 计费、免运维），提供从 POC 验证到大规模生产的平滑路径。

联系我们 CONTACT US

扫码关注公众号

扫码关注视频号

企业电话：4006882032

HAI，计算成就将来

www.meta-stone.com

中科曙光介绍

中科曙光是我国核心信息基础设施领军企业，2014 年在上交所上市（股票代码:603019）。早在上世纪 90 年代初，在 863 计划支持下成立的国家智能计算机研究开发中心，开启了曙光系列高性能计算机的研制历程。从曙光一号、曙光 1000 到曙光 7000，再到“硅立方”系列、scaleX 超节点系列与 AI 超集群，30 多年来曙光不断实现技术突破，多次刷新性能记录。其高性能计算机广泛应用于石油、气象、教育、政府、企业、科研、生物、金融、电信和互联网等众多行业，成为支撑大规模数值模拟、工业仿真、大模型训练推理及 AI for Science 等应用创新的关键基础设施。当前，曙光已成为亚洲最大的高性能计算机生产和制造商之一，2015 年进入全球 TOP500 份额前三，并十年九次获得中国 HPC TOP100 市场份额第一。

在高端计算之外，曙光同样在存储领域拥有 20 余载深厚积累。作为公司核心产品线之一，曙光存储依托近千人的研发团队和 300 余项核心技术发明专利，推出的 ParaStor 分布式存储与 FlashNexus 集中式全闪存，以世界级的领先性能代表着国内存储技术的最高水平。目前，曙光存储已为全国数千家用户管理着数 EB 的存储空间。凭借自主研发与超大规模部署经验，其在先进计算、通信、人工智能、气象环保、能源、金融、生物信息等领域广泛应用，市场份额持续领先，能够充分满足新型数据中心对敏捷高效、安全可靠存储的需求。

作为中国高性能计算领域的领军者，曙光秉承技术创新与客户至上理念，在全国各省市均设有分支机构，拥有 3 大智能制造生产基地、5 大研发中心，并在 50 多个城市部署了城市云计算中心。未来，曙光将持续专注于核心技术研发，携手用户与合作伙伴共建应用生态，以科技创新助力“数字中国”建设，驱动经济高质量发展，让国家、社会、个人共享计算与数据创造的价值。



解得科技介绍

解得科技是一家专注于面向 GPU 的大规模高性能解法器软件研发与产业化的科技公司，聚焦科学与工程计算核心需求，连接先进计算芯片与工业软件，构建自主可控、高性能、高可扩展的基础求解软件体系。

公司主要产品包括 PanguLU 直接解法器、LinglongAMG 代数多重网格解法器和 XianzhiKrylov 迭代解法器，适配国内外主流 GPU 平台，支持从单 GPU 到大规模多 GPU 并行计算，并可通过标准接口集成至 CAE、EDA 及各类工业仿真软件，服务装备制造、能源工程、航空航天、汽车工业以及电子与芯片设计等关键领域。



笔记页

中国科学院 SIAT CHINESE ACADEMY OF SCIENCES 深圳先进技术研究院

Solver2026

第九届解法器快速算法及应用研讨会

2026年11月27日—30日 · 深圳

solver-conference.cn

承办单位

中国科学院深圳先进技术研究院
SHENZHEN INSTITUTES OF ADVANCED TECHNOLOGY
CHINESE ACADEMY OF SCIENCES

香港中文大學（深圳）
The Chinese University of Hong Kong, Shenzhen

湖南韶峰应用数学研究院
Hunan Shaofeng Institute for Applied Mathematics

深圳国际工业与应用数学中心
SHENZHEN INTERNATIONAL CENTER FOR INDUSTRIAL AND APPLIED MATHEMATICS

赞助单位

HUAWEI

METASTONE 是石科技

中科曙光 Sugon

解得科技

