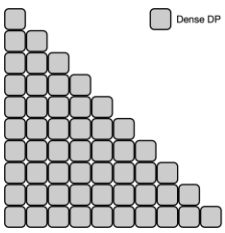
Hierarchical Computations on Manycore Architectures Powered by the PaRSEC Runtime System

HiCMA Powered by PaRSEC

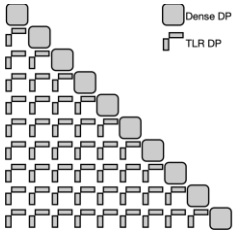
Qinglei Cao, Rabab Alomairy, George Bosilca, Sameh Abdulah, Hatem Ltaief, and David E. Keyes

The Hierarchical Computations on Manycore Architectures (HiCMA) library, which is powered by the PaRSEC runtime system, extends dense linear algebra libraries toward hierarchical computations on distributed many-core, heterogeneous architectures. The core idea of HiCMA is to develop fast linear algebra computations on dense, low-rank, mixed-precision, and sparse data formats for scientific and AI applications, while satisfying specified numerical accuracy and leveraging massively parallel heterogeneous hardware.

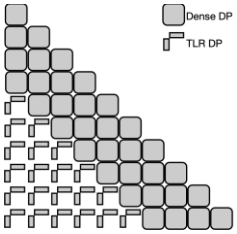
SUPPORTED MATRIX COMPUTATIONS



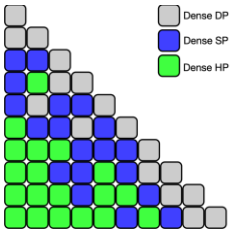
Dense; FP64



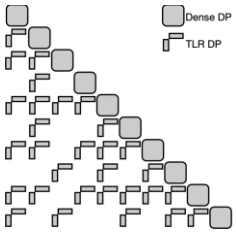
Diag Dense + TLR; FP64



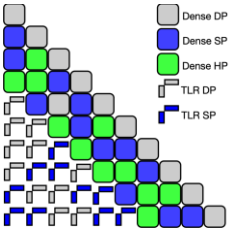
Band Dense + TLR; FP64



Dense; Adaptive Tile-Centric MP

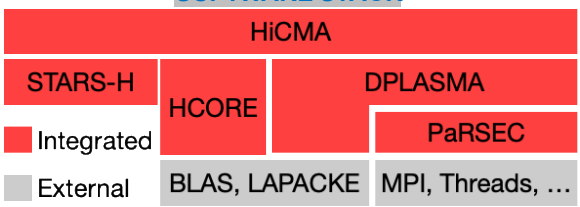


Diag Dense + TLR + Sparse; FP64

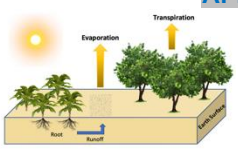


Band Dense + TLR; Adaptive Tile-Centric MP

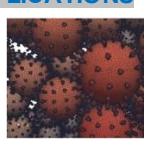
SOFTWARE STACK



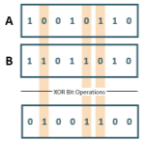
APPLICATIONS



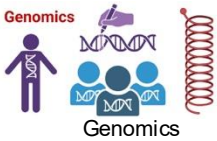
Geostatistical Modeling and Prediction



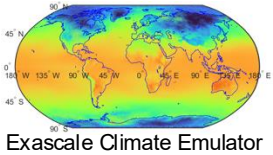
3D Unstructured Mesh Deformation



Hamming Distance



Genomics



Exascale Climate Emulator

CURRENT FEATURES

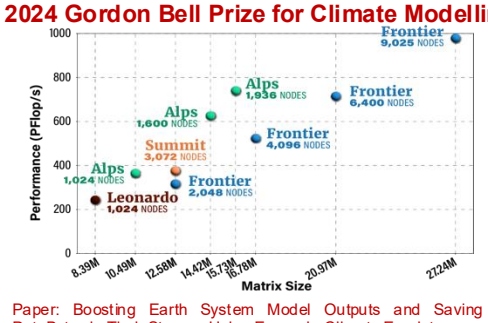
- **Matrix Computations:** Cholesky factorization, Matrix-matrix multiplication, Symmetric rank k update
- **Tile-Centric Mixed-Precision** (FP64, FP32, FP16, FP8, INT8, 1Bit), **Tile Low-Rank** (TLR) Approximation, and **Sparse** Computations
- **Memory Layout:** LAPACK and DPLASMA/tile
- **Hardware Supports:** CPUs and GPUs (Nvidia and AMD), Shared and distributed-memory environments

FUTURE DIRECTIONS

- Julia Integration
- LU and QR Factorizations
- Tile Low-Rank on GPUs
- Sparse Direct Solvers
- NVIDIA CUTLASS
- Precision Emulation

GITHUB REPO

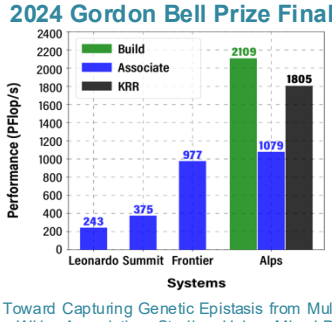
2024 Gordon Bell Prize for Climate Modeling



Paper: Boosting Earth System Model Outputs and Saving PetaBytes in Their Storage Using Exascale Climate Emulators

Stochastically capture spatio-temporal variability, accelerate computation, improve emulator accuracy, and yield insights on climate change and extremes.

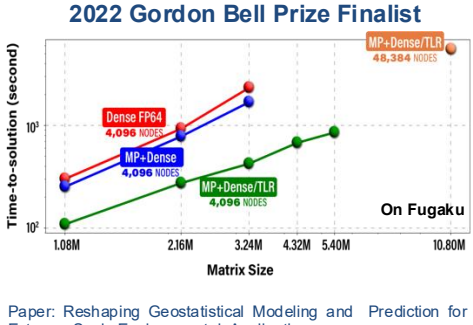
2024 Gordon Bell Prize Finalist



Paper: Toward Capturing Genetic Epistasis from Multivariate Genome-Wide Association Studies Using Mixed-Precision Kernel Ridge Regression

Enhance Kernel Ridge Regression (KRR) and achieve a performance improvement of five orders of magnitude over the state-of-the-art CPU-only REGENIE.

2022 Gordon Bell Prize Finalist



Paper: Reshaping Geostatistical Modeling and Prediction for Extreme-Scale Environmental Applications

Extend the capability of space-time geostatistical modeling using algebraic approximations and illustrate application-expected accuracy worthy of double precision from majority low-precision computations and low-rank matrix approximations.

A collaboration of

KAUST
THE UNIVERSITY OF TENNESSEE KNOXVILLE
Massachusetts Institute of Technology
SAINT LOUIS UNIVERSITY
NVIDIA

With support from

CINECA
U.S. DEPARTMENT OF ENERGY
OAK RIDGE National Laboratory
RIKEN
CSCS
NSF