Highly Productive, High Performance Application Frameworks for Post Petascale Computing

Naoya Maruyama, Tokyo Tech (PI)
Takayuki Aoki, Tokyo Tech (Co PI)
Kenjiro Taura, U-Tokyo (Co PI)
Kenji Yasuoka, Keio (Co PI)

Project Overview

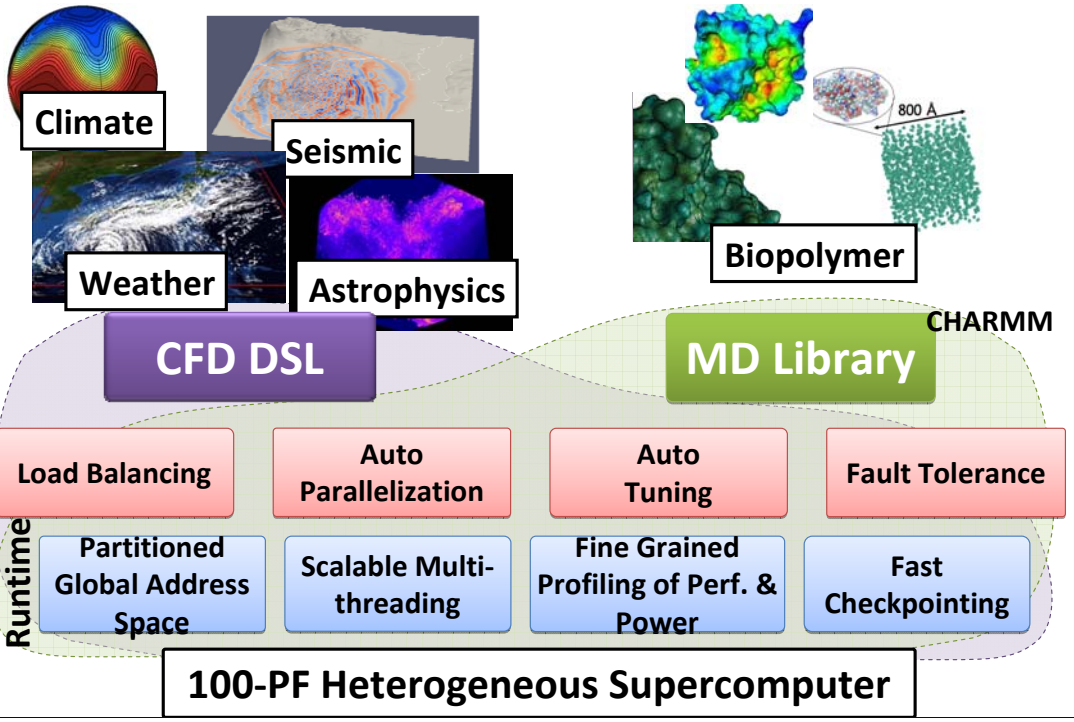
Emergence of large-scale heterogeneous supercomputers → Low productivity due to lack of programming model

Domain-specific application frameworks for post peta-scale systems that address:

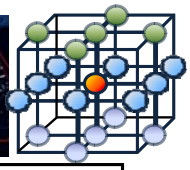
1. Productivity: Auto-parallelization and global memory
2. Performance: Comparable to manually tuned code
3. Resilience: Transparent and ultra-fast checkpointing
4. Power efficiency: Hardware-software cooperative power management

→ Demonstration using TSUBAME3 by developing frameworks for CFD and MD

Framework Overview

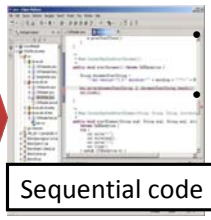


Domain-Specific Programming Model for Heterogeneous Supercomputer



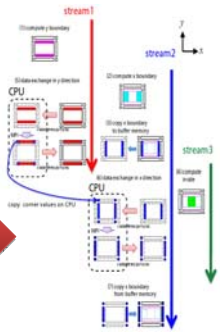
Modeling & discretization

$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} = -\frac{1}{\rho} \nabla P - 2\Omega \times \mathbf{u} - \Omega \times (\Omega \times \mathbf{r}) + \mathbf{g} + \mathbf{F}$$



Sequential code

- CPU/GPU hybrid programming
- Memory hierarchy optimization
- Scalability optimization by overlapping communication and computation
- ....



Domain-Specific Language

- Declarative
- Portable
- Shared memory model
- Sequential execution model

Backend

- Automatic parallelization
- Model-based performance and power optimization
- Auto-tuning-based optimization
- Hybrid code

Runtime

- Fault tolerance (checkpointing)
- Model refinement with runtime information
- Further auto tuning