Adaptive Plasma Physics Simulations: Dealing with Load Imbalance using Charm++

19th Annual Workshop on Charm++ and Its Applications

Diego Jiménez, Iván Vargas, **Esteban Meneses**

Advanced Computing Laboratory
Costa Rica High Technology Center
{djimenez,emeneses}@cenat.ac.cr

School of Physics
Costa Rica Institute of Technology
ivargas@tec.ac.cr

CeNAT
Centro Nacional de Alta Tecnología

Arenal Lake, Costa Rica

Plasma Physics Laboratory for Fusion Energy and its Applications

Costa Rica Institute of Technology

Outline

Background

Adaptive Plasma Simulation

Experimental Evaluation

Discussion

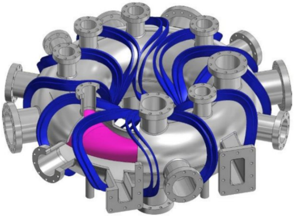
Diego Jiménez, Iván Vargas, **Esteban Meneses**

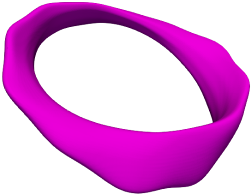
Stellarator Costa Rica 1 (SCR-1)

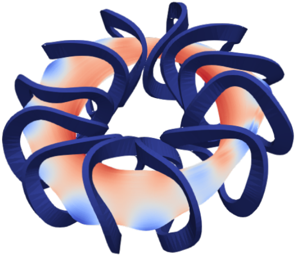
Magnetic confinement plasma reactor

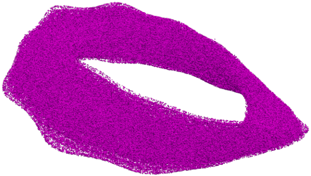
Simulation and Modeling

Biot-Savart Solver for Computing and Tracing Magnetic Field Lines (BS-SOLCTRA)







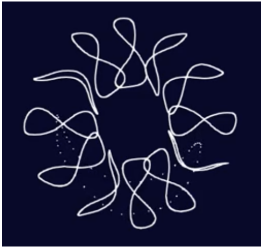


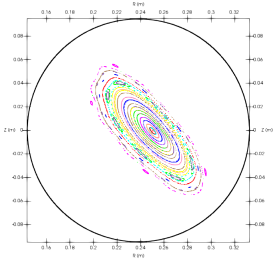
Diego Jiménez, Iván Vargas, **Esteban Meneses**

Plasma Physics Simulations

BS-SOLCTRA code

- Field-line tracing method
- Renewed interest for divertor research (deposition patterns of power and matter)
- C language with MPI, OpenMP, and AVX512
- Evaluates how modular coil designs affect confinement



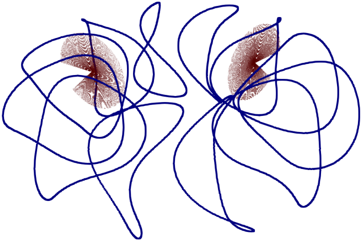


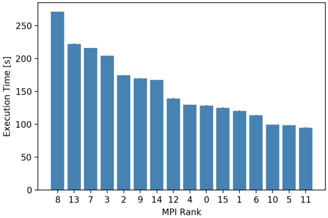
CeNAT

Load Imbalance

Particle divergence in BS-SOLCTRA

- Input scenario: 63,488 particles distributed among 16 MPI Ranks
- Non-deterministic runtime particle divergence

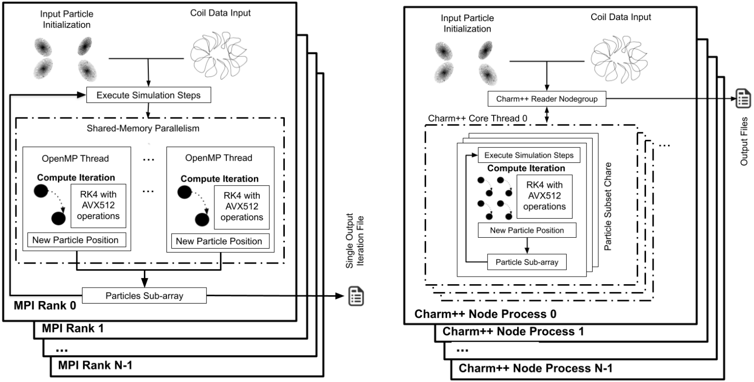




$$\Lambda = \frac{L_{max}}{L_{avg}} - 1 = 0.75$$

Adaptive Plasma Physics Simulations

Migrating from MPI+X to Charm++



Diego Jiménez, Esteban Meneses, VI Vargas. **Adaptive Plasma Physics Simulations: Dealing with Load Imbalance using Charm++**. Practice and Experience in Advanced Research Computing (PEARC), July, 2021.

CeNAT

Experimental Evaluation
Setup

- Two KNL systems:
  - Kabré Supercomputer (CeNAT): debugging and performance analysis
  - Theta Supercomputer (ALCF): performance and scalability experiments
- Charm++ SMP build:
  - 1 comm. thread per Charm++ SMP process
  - 62 PEs + 1 comm. thread per KNL node

Experimental Evaluation

Ground Zero Comparison

- **Ground Zero Comparison**: performance overhead from purely porting application to Charm++
  - Non-divergent replicated particle as input (15,872 particles)
  - 4 KNL nodes
  - 10k iterations, no load-balancing

Implementation	**Average Exec. Time (s)**	**Standard Deviation**	**Coefficient of Variation**
MPI+X Static	323.777	2.842	0.009
MPI+X Dynamic	329.671	1.136	0.003
Charm++ SMP	349.313	9.431	0.027

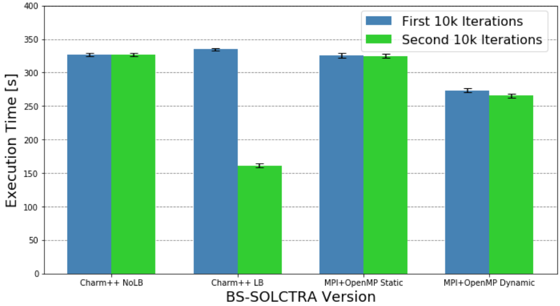
- Charm++ overhead: 7.89%
- Still profitable: 75% load imbalance

CeNAT

Experimental Evaluation

Load Imbalance Base Comparison

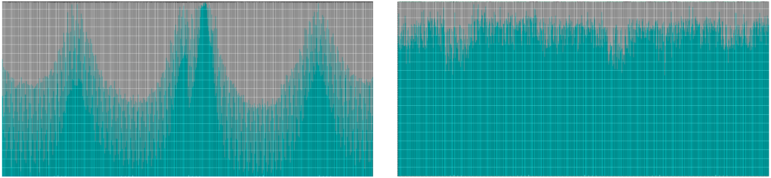
- Next step: use imbalanced scenario as input and apply balancing strategy
- Setup:
  - 16 KNL nodes
  - 63,488 particles (3,968 ppn)
  - 8:1 virtualization ratio in Charm++ implementation
  - 20k iterations, balancing strategy on the 10k mark



Diego Jiménez, Iván Vargas, **Esteban Meneses**

Experimental Evaluation

Load Imbalance Base Comparison



Charm++ RTS: Greedy Load Balancing Strategy

45.2% $\longrightarrow$ 80.2% average CPU usage

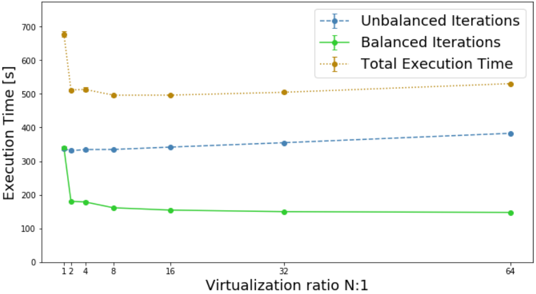
Charm++ deals with both inter and intra-node imbalance

Diego Jiménez, Iván Vargas, **Esteban Meneses**

Experimental Evaluation

Virtualization Ratio

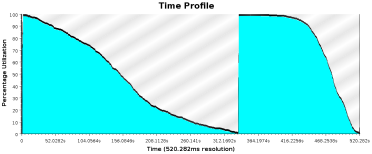
- Over-decomposition
- Setup:
  - $20k$ iterations
  - First $10k$ iterations: Unbalanced
  - Second $10k$ iterations: balanced



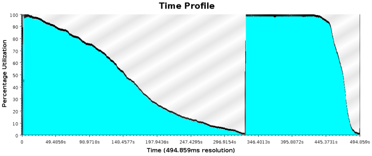
Diego Jiménez, Iván Vargas, **Esteban Meneses**

Experimental Evaluation

Virtualization Ratio



2:1 Virtualization Ratio



8:1 Virtualization Ratio

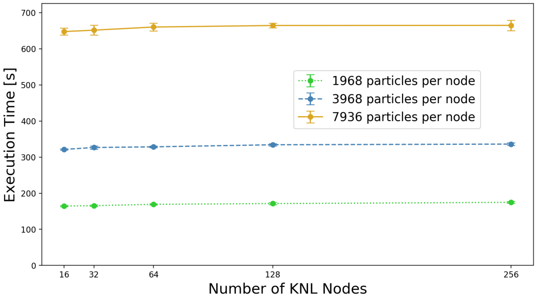
- Gentle decline in usage after load balancing
- Bigger chares affect runtime's ability to distribute work evenly

- Steeper decline in usage after load balancing
- Smaller amount of particles per chare, better work distribution

Experimental Evaluation

Scaling Charm++

- MPI+OpenMP version exhibits good weak scaling
- Setup:
  - Scale from 16 to 256 nodes
  - Three different problem sizes per node: 1984 ppn, 3968 ppn and 7936 ppn
  - $10k$ iterations, no load balancing
  - Random uniform particle distribution
  - 8:1 virtualization ratio



Experimental Evaluation

Checkpoint/Restart Capabilities

Problem Size	Checkpoint Time (s)	Std. Deviation	CV	Checkpoint Size [B]
1984 ppn	5.112	0.978	0.191	7,8M
3968 ppn	4.303	0.825	0.191	7,9M
7936 ppn	4.129	1.193	0.289	9.9M

- Added minimum checkpointing capabilities
- Initial exploration:
  - 8 nodes
  - Checkpoint time remains roughly constant
  - Some variability - vulnerability of file systems

Diego Jiménez, Iván Vargas, **Esteban Meneses**

CeNAT

Porting to Charm++

Lessons learned

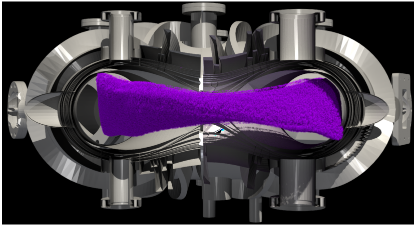
1. MPI to Charm++ migration:
   - MPI ranks are mapped to Charm++ node processes in the SMP build
   - Data is partitioned into chares that are assigned to SMP worker threads
   - Determining best virtualization ratio is important in maximizing performance
2. Load-imbalance:
   - Charm++ deals with intra and inter-node imbalance
   - Several parameters must be tuned: frequency, algorithm, virtualization ratio
3. SMP and non-SMP Charm++ builds:
   - SMP exhibited non-negligible performance variability in some experiments.
   - non-SMP creates heavyweight processes per each compute-core, more stable in performance.
   - In the SMP build, one core has to be sacrificed for each SMP process. This could affect compute intensive applications.

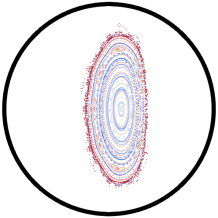
CeNAT

Scalability

Higher granularity results

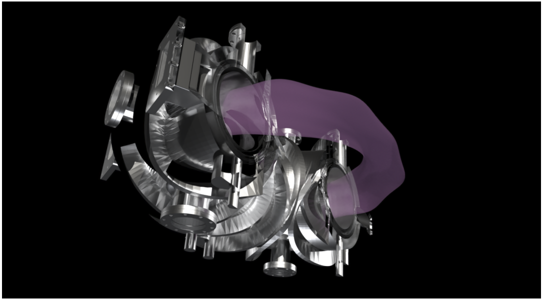
Nodes	Average (s)	Std. Dev. (s)	Coeff. Var.	Speedup	Efficiency
64	2808.80	35.03	0.012	1.00	1.00
128	1421.45	15.44	0.011	1.99	0.99
256	729.95	7.62	0.010	3.85	0.96
512	391.71	3.64	0.009	7.17	0.90
1024	206.24	4.66	0.023	13.62	0.85





Scientific Visualizations

A new computer graphics model



Luis Campos-Duarte, Diego Jiménez, Esteban Meneses, Ricardo Solano-Piedra, Esteban Pérez, VI Vargas, Ernesto Rivera-Alvarado. **Towards Photorealistic Visualizations for Plasma Confinement Simulations**. Practice and Experience in Advanced Research Computing (PEARC), July, 2021.

CeNAT

Scientific Collaboration

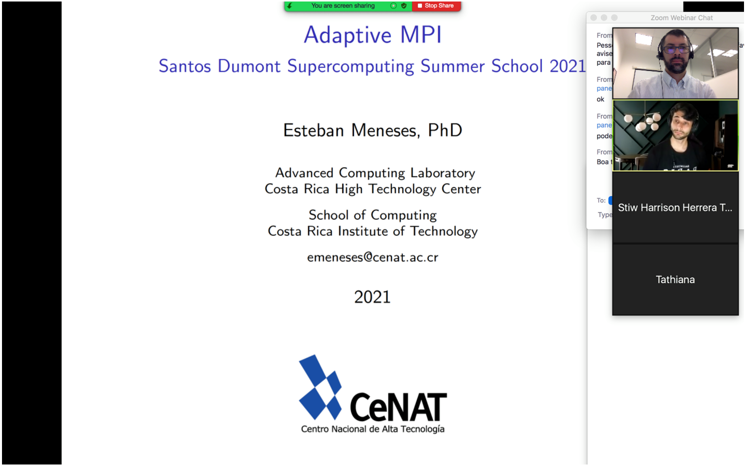
Particle-in-cell code

- Joint project with Max Planck Computing and Data Facility (MPCDF): Erwin Laure and Markus Rampp
- *Advancing plasma physics computer simulations with the latest high performance computing techniques*
  - Exploration of task parallelism
  - Exploration of performance-portability libraries
- GEMPIC: Geometric ElectroMagnetic Particle-In-Cell Methods
- Porting AMReX library to AMPI

CeNAT

Scientific Training

Teaching Adaptive MPI



Acknowledgements

This research used resources of the Argonne Leadership Computing Facility (ALCF), which is a DOE Office of Science User Facility supported under Contract DE-AC02-06CH11357.

The research program hosting this effort is partially supported by a machine allocation on Kabré supercomputer at the Costa Rica National High Technology Center (CeNAT).

CeNAT

Concluding Remarks

Sucessfully ported MPI+X application to Charm++

- Adaptive Charm++ runtime system improves resource utilization (1.64 speedup)
- Checkpointing was added for fault-tolerance and split execution capabilities
- **Future work**:
  - Study performance variability in Charm++ SMP build
  - Automatic load balancing depending on input scenario
  - Split execution to understand effect of dynamic variation in electric current on confinement device operations

Thank you!

emeneses@cenat.ac.cr

CeNAT
Centro Nacional de Alta Tecnología