# Saving Energy by Exploiting Residual Imbalance on Iterative Applications

Edson L. Padoin, **Laércio L. Pilla,**
Márcio Castro, Philippe O. A. Navaux
and Jean-François Méhaut

Federal University of Rio Grande do Sul – BR
CEA/LIG/University of Grenoble – FR

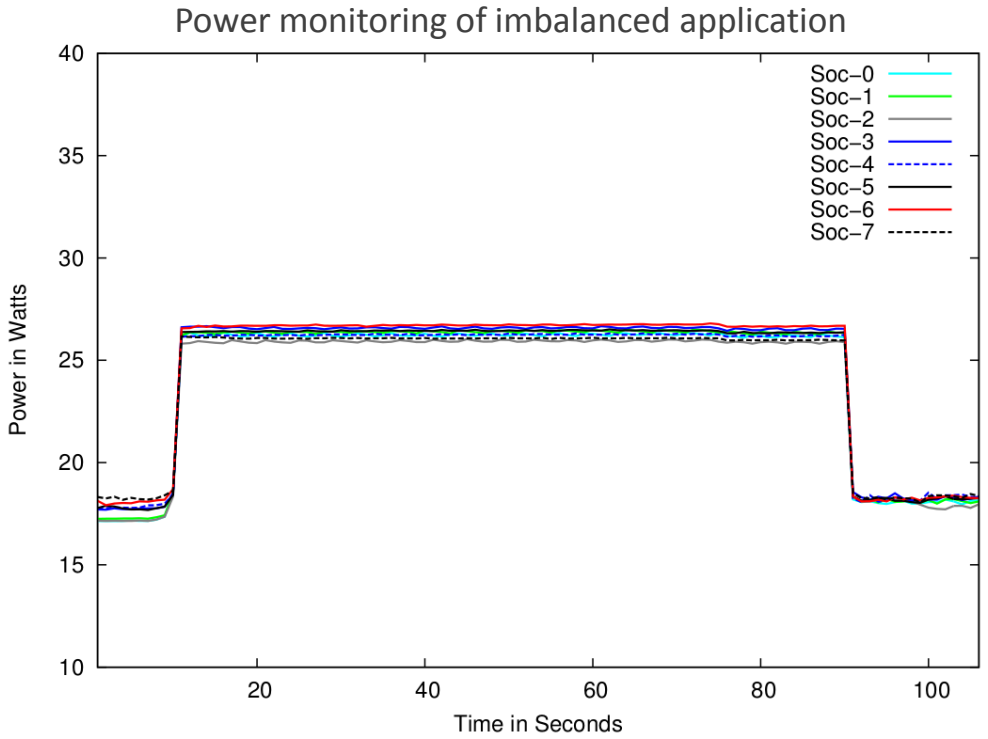



# Motivation

- **Energy consumption**
  - Key issue to build **Exascale systems**
  - DARPA limits consumption in 20 MWatt
- Current top performance HPC systems
  - **PFlops** while consuming **MWatts**
    - Tianhe-2 - #1 Top500
      - 33.8 PFlops
      - 17.8 MWatt
      - 1.9 GFlops/Watt

# Motivation

- **Load imbalance**
  - –May pass unnoticed
  - –All sockets
    - Constant power demand
    - Similar energy consumption
    - **Energy waste**



# Motivation

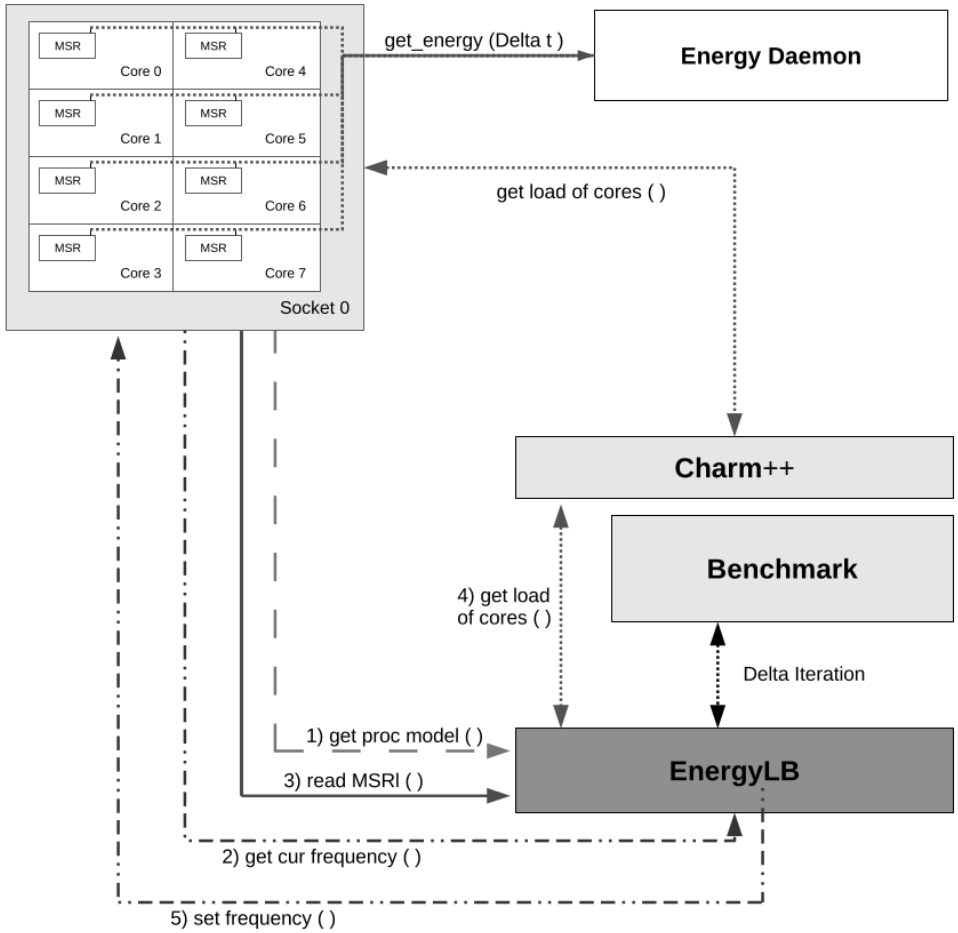
- **Goal**
  - Minimize **energy consumption** while considering **core workload**
    - Performance with less power
- **Proposal**
  - Combine **dynamic load balancing** with **DVFS**
    - Energy Daemon + Charm++/AMPI

# Agenda

- Motivation
- Energy Daemon and EnergyLB
- Experiments
  - First Prototype
  - Second Prototype
- Concluding Remarks

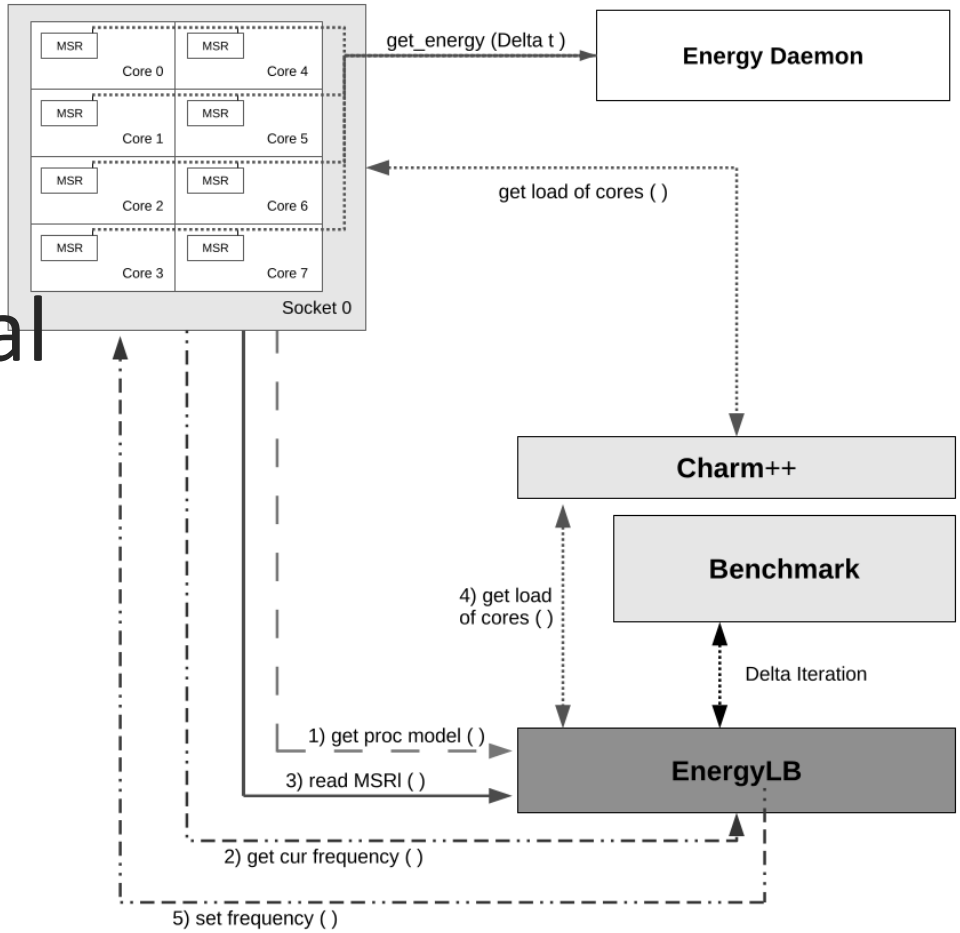
# Energy Daemon and EnergyLB

- **Energy Daemon**
  - Gets/computes
    - Energy
    - Power
  - Traces execution
  - Works on Intel and ARM



# Energy Daemon and EnergyLB

- **EnergyLB**
  - Charm++ module
    - LB framework
  - Computes residual imbalance
  - Controls clock frequency
    - Less loaded, less frequency



# Energy Daemon and EnergyLB

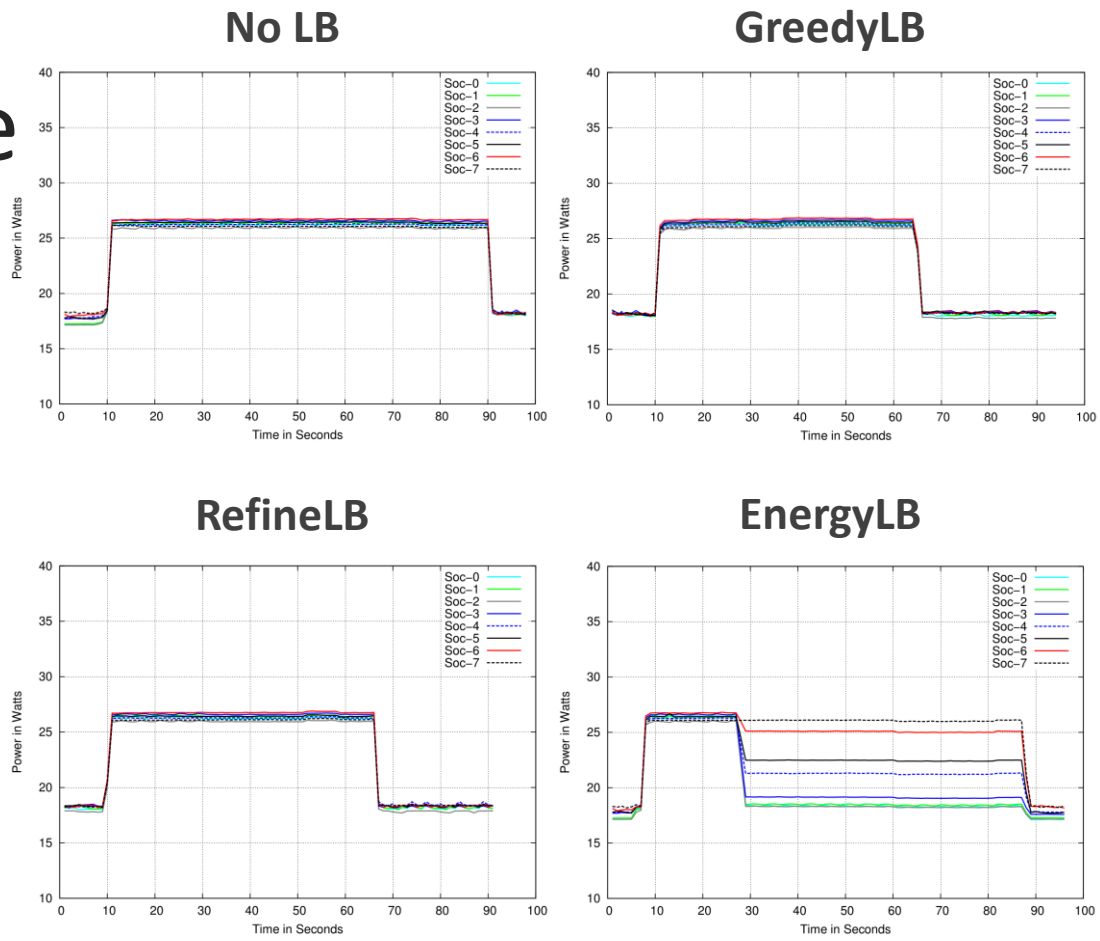
- **EnergyLB**
  - Current prototypes
    - **Centralized version**
      - Attached to other LB to DVFS
      - Uses one core per socket
    - **Hierarchical version**
      - Root load balancer
      - Another LB per socket
      - Uses all cores
  - EnergyLB does not migrate tasks

# Experiments

- Platform: **SGI UV2000**
  - 24 Intel Xeon E5-4640 8 cores (192 cores)
    - Clock frequency range
      - maximum: 2.4 GHz
      - minimum: 1.2 GHz
    - DVFS control per socket
  - 756 GB of DDR3 memory

# Experiments

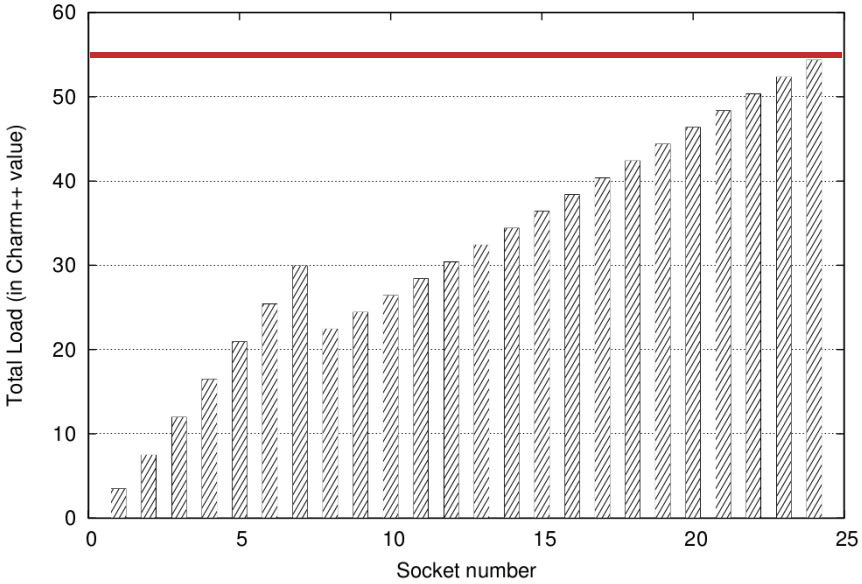
- First prototype
  - LB_test
  - 500 chares
  - 250 iterations
  - LB: every 50 iterations



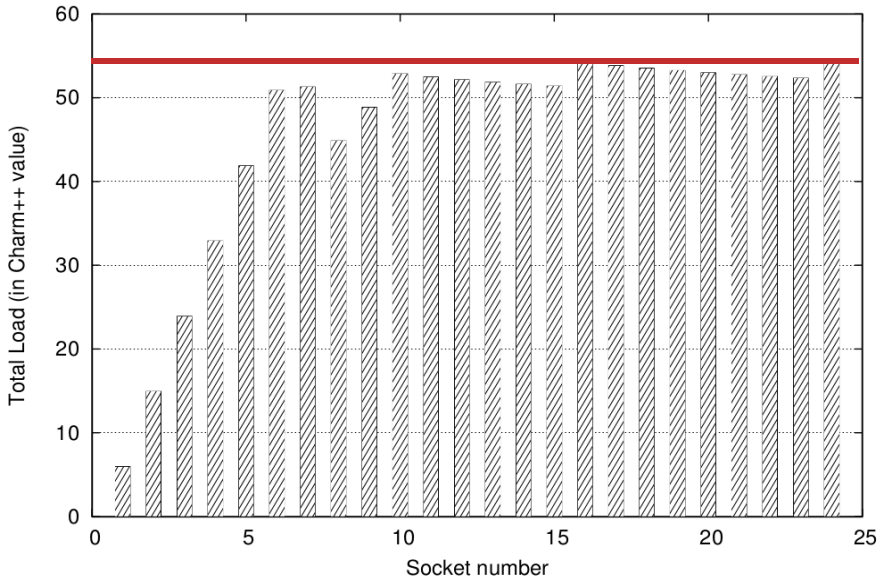
# Experiments

- First prototype
  - POA_bench
    - More parameters

**Before EnergyLB**



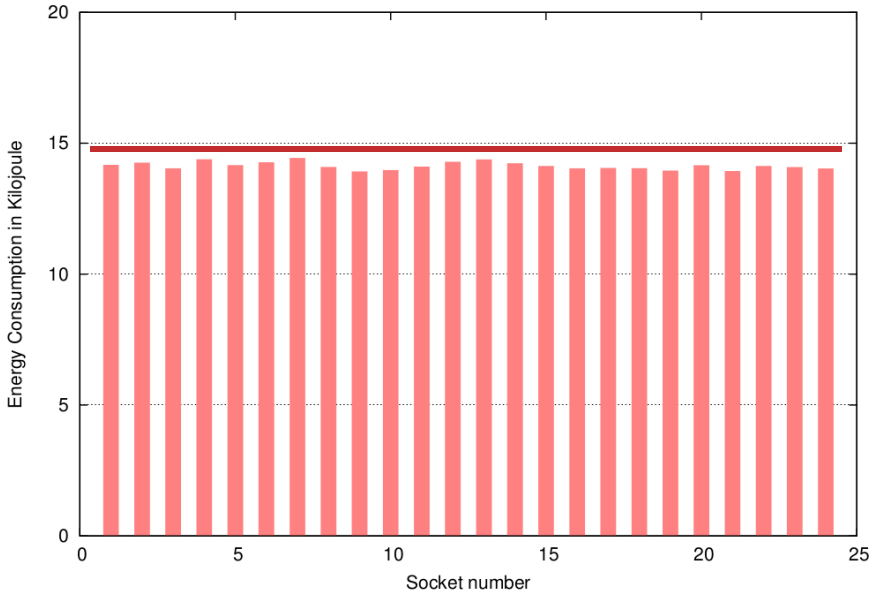

**After EnergyLB**



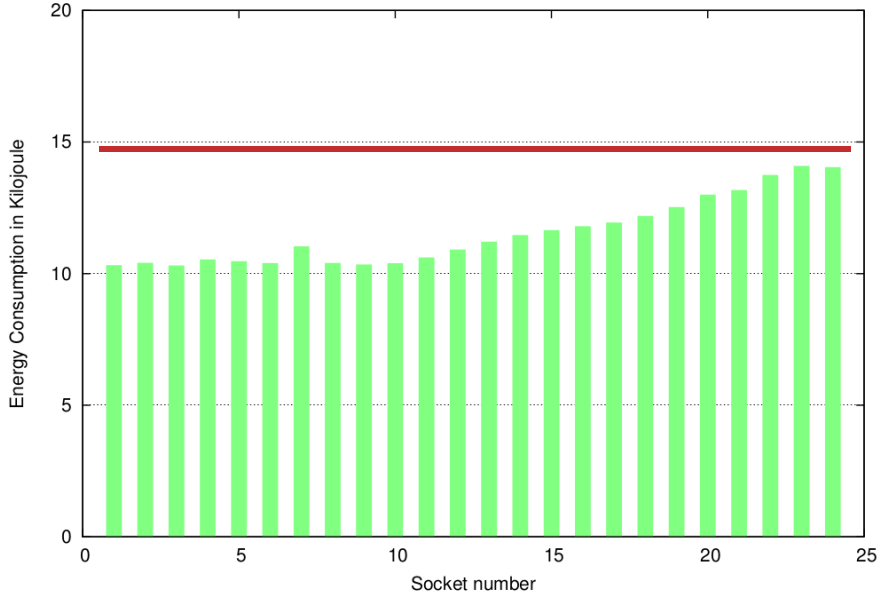
# Experiments

- First prototype
  - –POA_bench
    - More parameters

**Before EnergyLB**

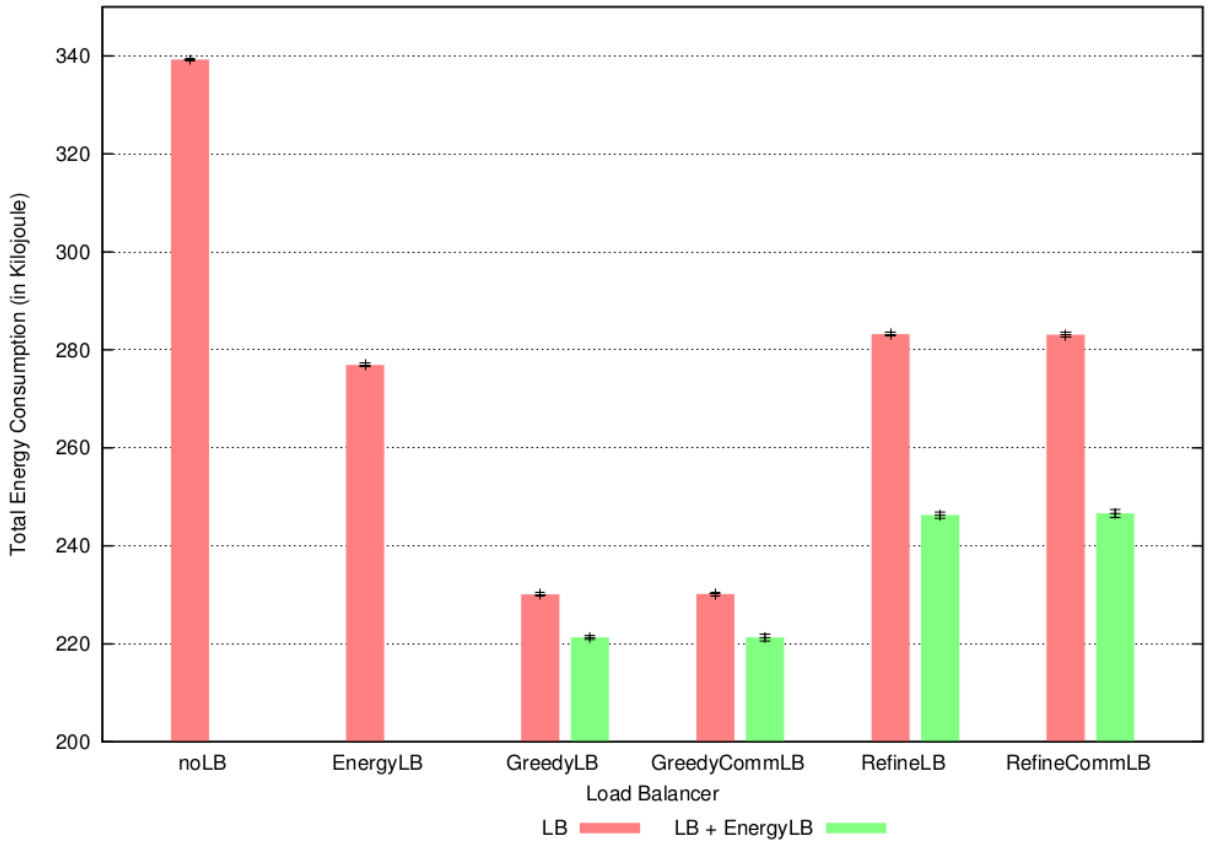



**After EnergyLB**



# Experiments

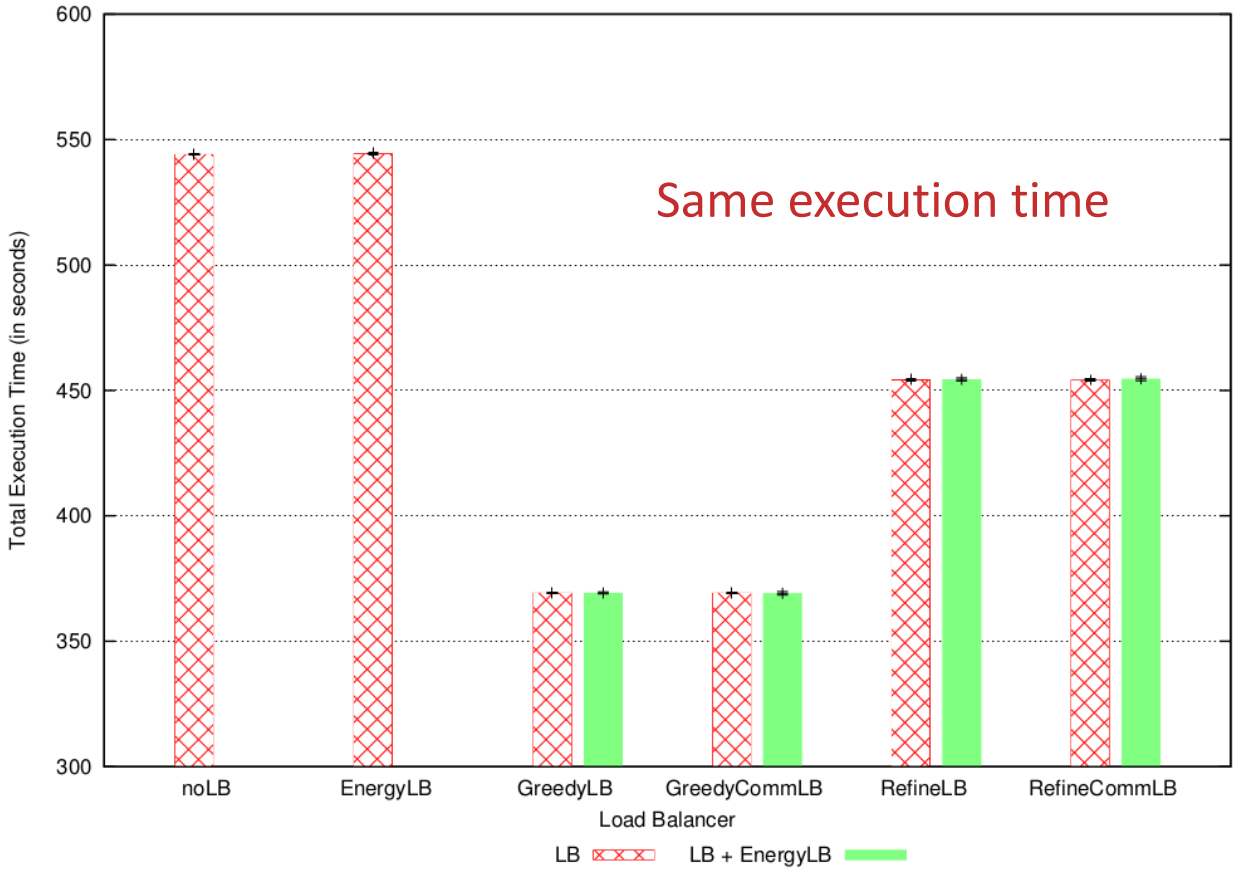
- First prototype
  - POA_bench




Saves between 4% and 13%

# Experiments

- First prototype
  - POA_bench



# Experiments

- Second prototype
  - **Ondes3D**
    - Seismic wave simulator
    - MPI -> AMPI
      - 512 VPs on 192 cores
    - Dynamic load imbalance

# Experiments

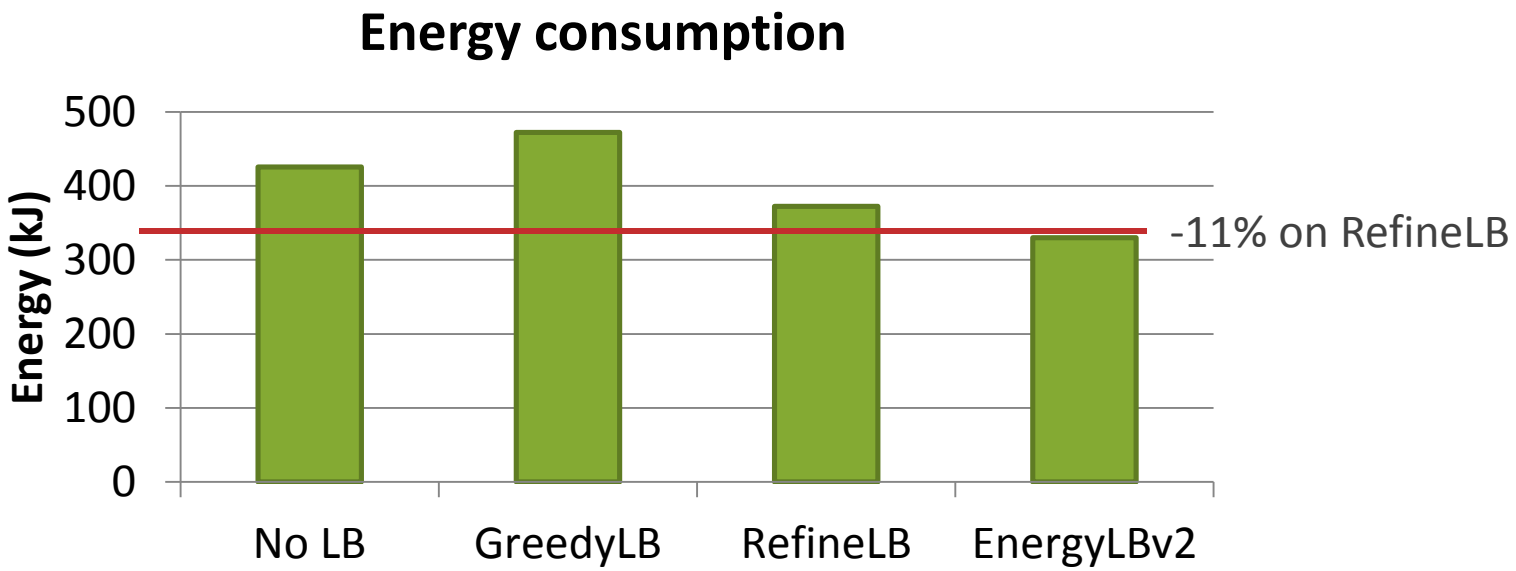
- Second prototype

| | Power (W) | % to no LB | Energy (Kj) | % to no LB | Time (s) | % to no LB |
|---|---|---|---|---|---|---|
| No LB | 49.6 | --- | 425.8 | --- | 357.7 | --- |
| GreedyLB | 50.3 | +1.45% | 472.5 | +10.98% | 391.3 | +9.39% |
| RefineLB | 52.7 | +6.31% | 372.2 | -12.59% | 294.1 | -17.78% |
| EnergyLB +RefineLB | 38.7 | -22.02% | 330.1 | -22.46% | 355.7 | -0.56% |

# Experiments

- Second prototype

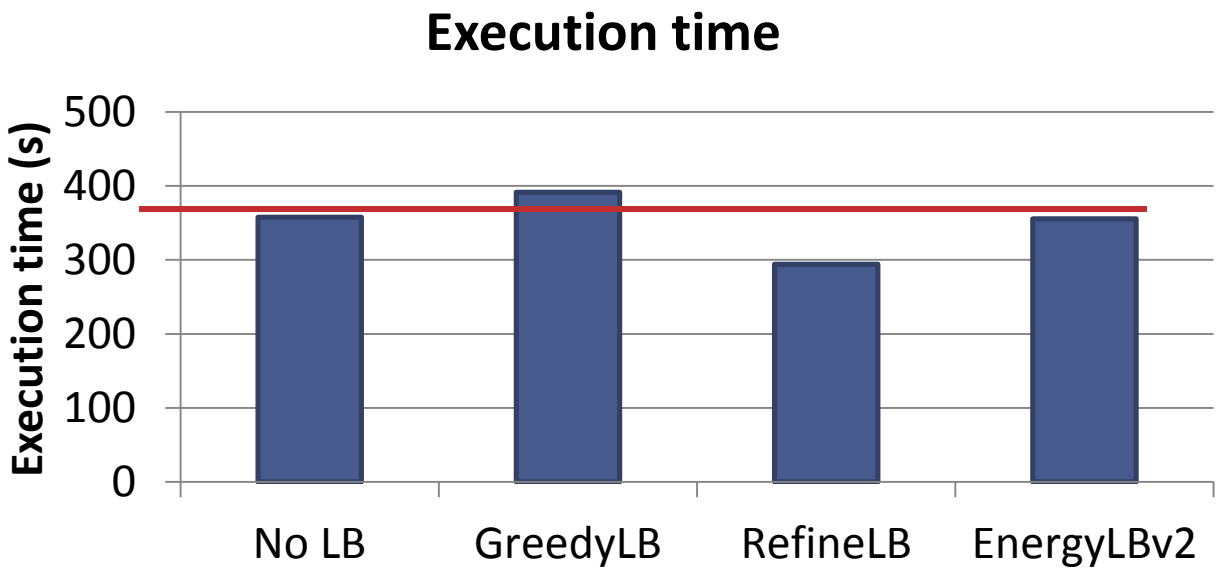
| | Power (W) | % to no LB | Energy (Kj) | % to no LB | Time (s) | % to no LB |
|---|---|---|---|---|---|---|
| No LB | 49.6 | --- | 425.8 | --- | 357.7 | --- |
| GreedyLB | 50.3 | +1.45% | 472.5 | +10.98% | 391.3 | +9.39% |
| RefineLB | 52.7 | +6.31% | 372.2 | -12.59% | 294.1 | -17.78% |
| EnergyLB +RefineLB | 38.7 | -22.02% | 330.1 | -22.46% | 355.7 | -0.56% |



# Experiments

- Second prototype

| | Power (W) | % to no LB | Energy (Kj) | % to no LB | Time (s) | % to no LB |
|---|---|---|---|---|---|---|
| No LB | 49.6 | --- | 425.8 | --- | 357.7 | --- |
| GreedyLB | 50.3 | +1.45% | 472.5 | +10.98% | 391.3 | +9.39% |
| RefineLB | 52.7 | +6.31% | 372.2 | -12.59% | 294.1 | -17.78% |
| EnergyLB +RefineLB | 38.7 | -22.02% | 330.1 | -22.46% | 355.7 | -0.56% |



# Concluding Remarks

- **Goal**
  - Minimize **energy consumption** while considering **core workload**
    - Performance with less power
- **Proposal**
  - Combine **dynamic load balancing** with **DVFS**
    - Energy Daemon + Charm++/AMPI
    - **Two prototypes**

# Concluding Remarks

- **First prototype**
  - Energy gains between 4% and 13%
- **Second prototype**
  - Energy gains of 22% but no performance gains

# Concluding Remarks

- **Future work**
  - **Evolve hierarchical prototype**
    - Migrate tasks between sockets
  - **More experiments**
    - More platforms
      - Heterogeneous
    - More applications

# Saving Energy by Exploiting Residual Imbalance on Iterative Applications

Edson L. Padoin, **Laércio L. Pilla,**
Márcio Castro, Philippe O. A. Navaux
and Jean-François Méhaut

Federal University of Rio Grande do Sul – BR
CEA/LIG/University of Grenoble – FR