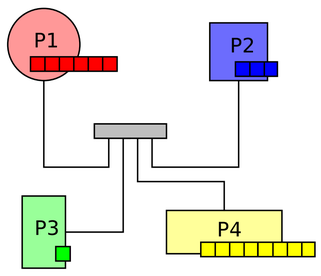
Design and Implementation of Parallel Algorithms for Modern Heterogeneous Platforms Based on Functional Performance Models

David Clarke

Heterogeneous Computing Laboratory, UCD, Dublin

Platform and Application

Generalised Heterogeneous Platform



Data Parallel Application

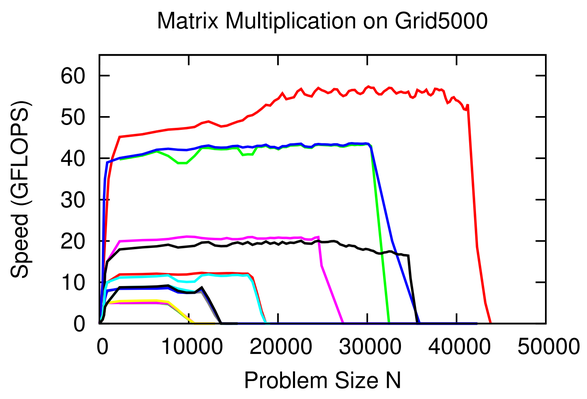
```
...

while(...) {
 compute_parallel(data, size);
 syncronise (data);
}

...
```

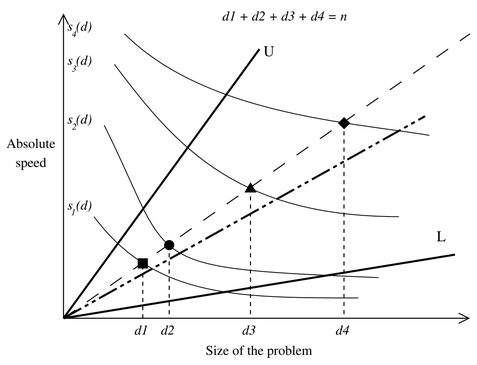
Traditional load Balancing

- Traditionally, processor performance is defined by a constant number.
- Computational units are partitioned as $d_i = N \times \frac{s_i}{\sum_{j=1}^{p} s_j}$.
- In reality, speed is a function of problem size.



Partitioning Algorithms

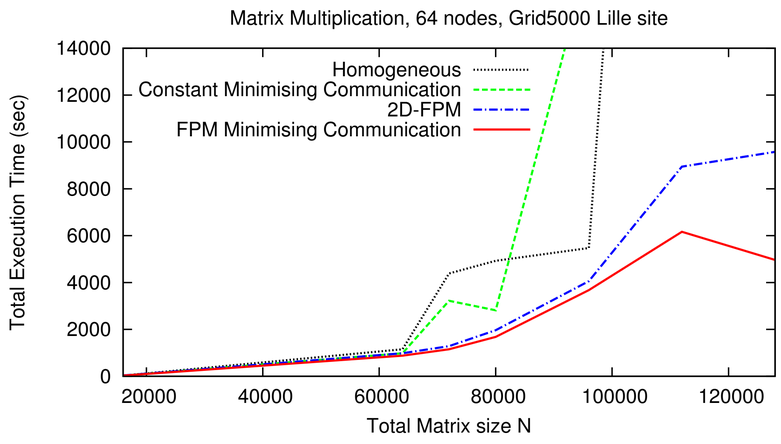
- Input: FPMs and N. Output: $d_1, d_2, \ldots, d_p$.
- Load Balanced when: $\frac{d_1}{s_1(d_1)} \approx \frac{d_2}{s_2(d_2)} \approx \ldots \approx \frac{d_p}{s_p(d_p)}$, $d_1 + d_2 + \ldots + d_p = N$
- Geometric Partitioning Algorithm
  - Points $\left(d_i, s_i(d_i)\right)$ lie on a line passing through the origin when $\frac{d_i}{s_i(d_i)} = constant$.
  - Value of N determines the slope.
  - Algorithm iteratively bisects solution space to find values d_i.



- Numerical Partitioning Algorithm
  - Solves the system of nonlinear equations: $F(x) = \begin{cases} n - \sum_{i=1}^{p} x_i \\ \frac{x_i}{s_i(x_i)} - \frac{x_1}{s_1(x_1)}, \ 2 \le i \le p \end{cases}$

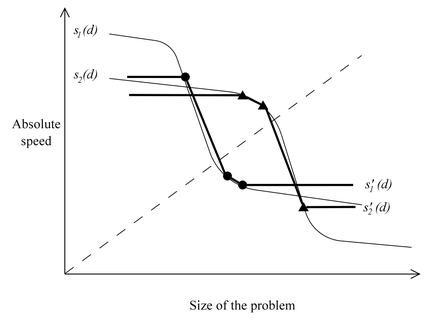
Two-Dimensional Matrix Partitioning with 1D FPMs

- Height and width combined into one parameter, area $d_i = m_i \times n_i$.
- Square areas are benchmarked $m = n = \sqrt{d}$.
- Partition with 1D FPM algorithm to find area of rectangles (geometric or numerical).
- Use communication volume minimising algorithm to compute ordering and shape of these rectangles.



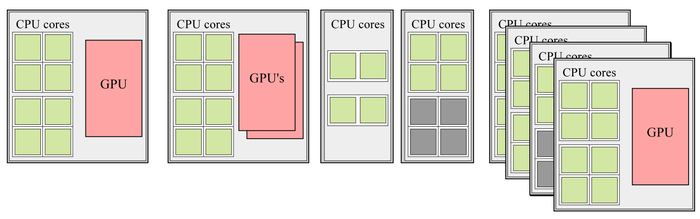
Dynamically Built Models

- Functional performance models are application and platform specific.
- On a stable platform they can be built at compile time.
- When the execution of an application is unique, an approximation may be built dynamically in the relevant areas at runtime.

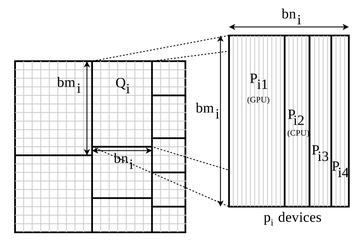


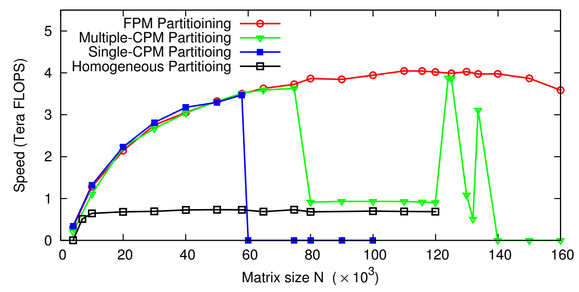
Partitioning for Hierarchical Heterogeneous Platforms

- Heterogeneous cluster of CPU+GPU nodes.
- Hierarchy in platform $\rightarrow$ hierarchy in partitioning



- Nested parallelism
  - *inter-node partitioning algorithm (INPA)*
  - *inter-device partitioning algorithm (IDPA)*
  - IDPA is nested inside INPA

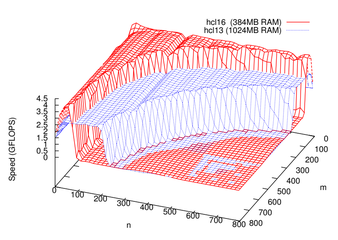




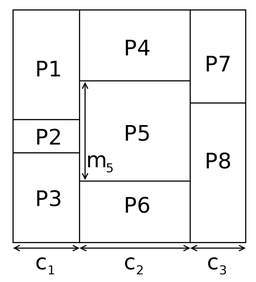
90 nodes, 432 CPUs, 12 GPUs from Grid5000 Grenoble site

Current Work: Two-Dimensional Partitioning with 2D FPMs

- Numerical 2D partitioning algorithm
  - Solve as a constrained non-linear minimisation problem
  - Minimise: $\sum_{j=1}^{q} \sum_{i=1}^{p_j} t_{ij}$, with $q+1$ constraints: $n_1 + n_2 + \ldots + n_q = N$ $m_{1j} + m_{2j} + \ldots + m_{p_j} = M$, for $1 \le j \le q$



- 3-step column based algorithm
  1. Extract 1D model from 2D with $n = c$ for $1 \le c \le N$. Partition to find optimum time, add point $s_{col}(c)$ to model.
  2. Use 1D partitioning find optimum column widths.
  3. Find optimum heights for each processor within the columns.



Software, Applications & Platforms

- Software: FuPerMod
  - Contains all presented algorithms.
  - Based on system and mathematical software: C/C++, MPI, Autotools, GNU Scientific Library, Boost C++ libraries, BLAS, CUDA Toolkit
  - Designed for easy integration with existing heterogeneous applications.
- Applications
  - Signal Processing Systems group, INESC-ID, Lisbon, Portugal
    - Using FPMs to partition database requests on heterogeneous platform.
    - Extended FPMs to overlap communications, $\times 4$ speedup for FFT.
  - Division of Scientific Computing, Uppsala University, Sweden
    - Upcoming collaboration to load balance multiphase flow simulations.
- Platforms
  - Grid'5000, 10 sites, 1260 nodes, France
  - HCL Cluster, local 16 node experimental cluster.
  - SiPS Cluster, 4 node CPU+GPU, Lisbon, Portugal

- Lastovetsky, A., and R. Reddy, **"Distributed Data Partitioning for Heterogeneous Processors Based on Partial Estimation of their Functional Performance Models"**, HeteroPar 2009
- Lastovetsky, A., and R. Reddy, **"Two-dimensional Matrix Partitioning for Parallel Computing on Heterogeneous Processors Based on their Functional Performance Models"**, HeteroPar 2009
- Clarke, D., A. Lastovetsky, and V. Rychkov, **"Dynamic Load Balancing of Parallel Computational Iterative Routines on Highly Heterogeneous HPC Platforms"**, PPL 2011
- Clarke, D., A. Lastovetsky, and V. Rychkov, **"Dynamic Load Balancing of Parallel Computational Iterative Routines on Platforms with Memory Heterogeneity"**, Heteropar 2010
- Rychkov, V., D. Clarke, and A. Lastovetsky, **"Using Multidimensional Solvers for Optimal Data Partitioning on Dedicated Heterogeneous HPC Platforms "**, PaCT 2011
- Clarke, D., A. Lastovetsky, and V. Rychkov, **"Column-Based Matrix Partitioning for Parallel Matrix Multiplication on Heterogeneous Processors Based on Functional Performance Models"**, HeteroPar 2011
- Lastovetsky, A., R. Reddy, V. Rychkov, and D. Clarke, **"Design and implementation of self-adaptable parallel algorithms for scientific computing on highly heterogeneous HPC platforms"**, PARCO (under review)
- Clarke, D., A. Ilic, A. Lastovetsky, L. Sousa, **"Hierarchical Partitioning Algorithm for Scientific Computing on Highly Heterogeneous CPU + GPU Clusters"** EuroPar 2012 (under review)