Optimal Partitioning Shapes for Parallel Matrix-Matrix Multiplication with Heterogeneous Processors

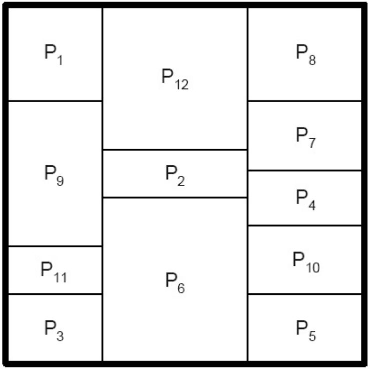
Ashley DeFlumere

Heterogeneous Computing Laboratory, UCD, Dublin

May 11, 2012

Motivation and Goals

- Traditionally, all processors assigned rectangles to compute
- Finding the optimal rectangular partition is difficult
- Is the rectangular shape optimal?



Methods

5 different MMM Algorithms:

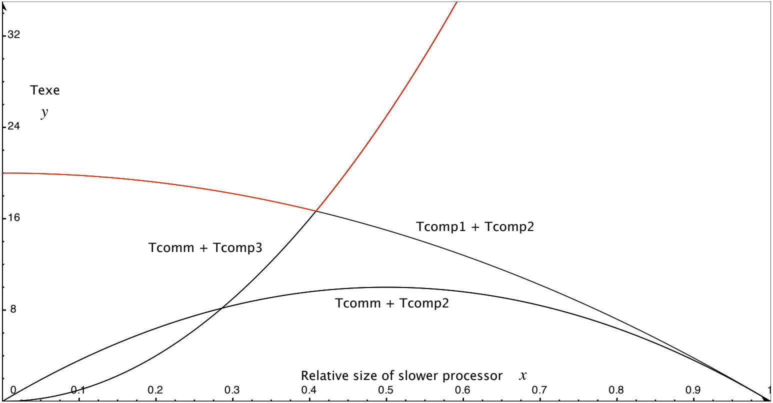
- 2 Barrier, communication then computation
- 2 Overlap, some immediate computation
- 1 Parallel, k-steps overlap all communication and computation

Models:

- Constant Performance Model
- Hockney Model of Communication

Finding the Partition Size - PCO

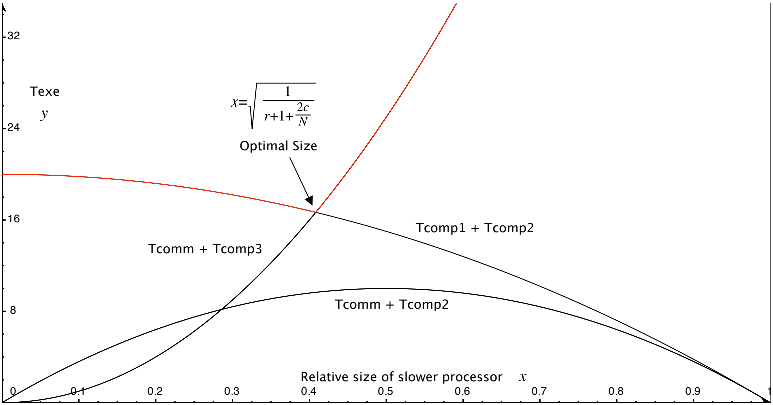
$$Texe = \max(\max(\max(Tcomm), Tcomp_1) + Tcomp_2\,,\ \max(Tcomm) + Tcomp_3)$$



Finding the Partition Size - PCO

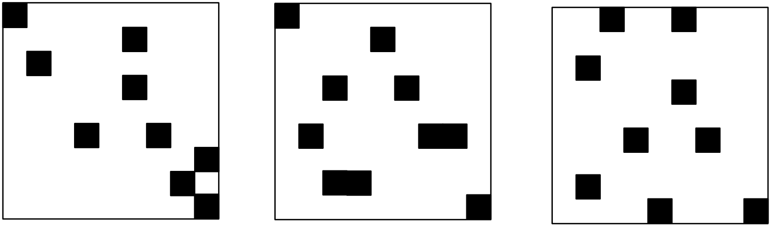
$$Texe = \max(\max(\max(Tcomm), Tcomp_1) + Tcomp_2\,,\ \max(Tcomm) + Tcomp_3)$$

$$Texe = \max\left(\max\left(\max\left(\frac{2x}{N} - \frac{2x^2}{N}, \frac{2x^2}{N}\right), \frac{(1-x)^2}{c}\right) + 2\frac{(x-x^2)}{c}, \max\left(\frac{2x}{N} - \frac{2x^2}{N}, \frac{2x^2}{N}\right) + \frac{x^2}{\frac{c}{a}}\right)$$



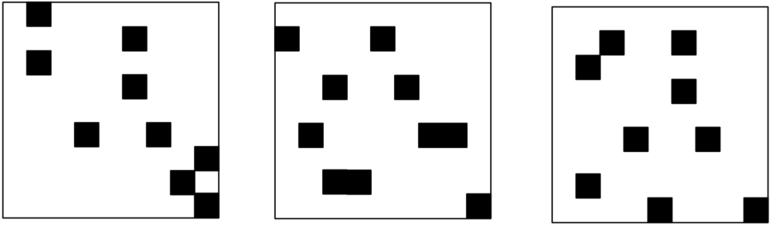
Finding Partition Shapes

- Consider all possible shapes, reduce using Push Technique



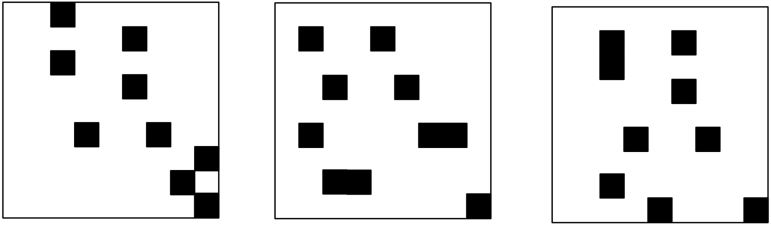
Finding Partition Shapes

- Consider all possible shapes, reduce using Push Technique



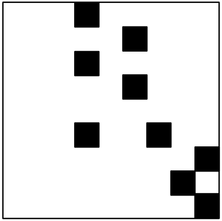
Finding Partition Shapes

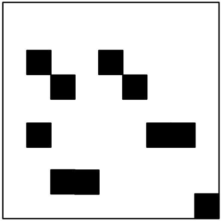
- Consider all possible shapes, reduce using Push Technique

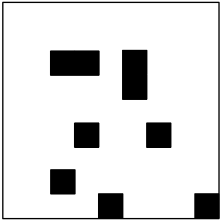


Finding Partition Shapes

- Consider all possible shapes, reduce using Push Technique

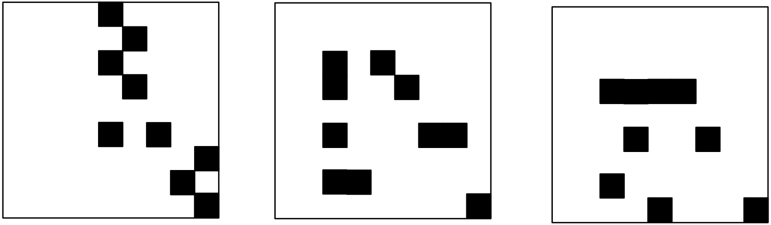






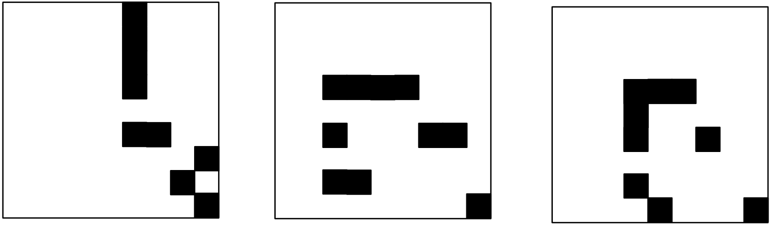
Finding Partition Shapes

- Consider all possible shapes, reduce using Push Technique



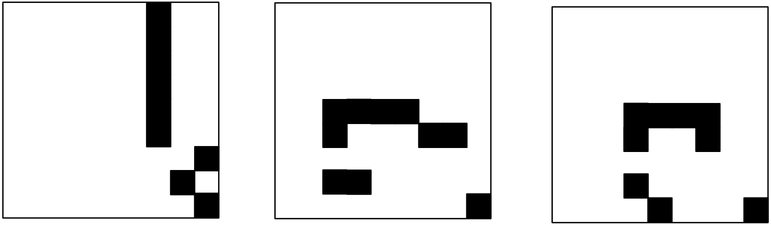
Finding Partition Shapes

- Consider all possible shapes, reduce using Push Technique



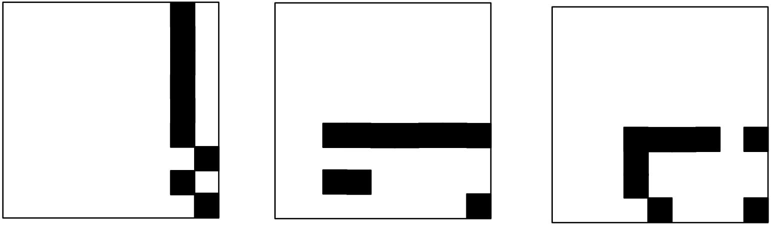
Finding Partition Shapes

- Consider all possible shapes, reduce using Push Technique



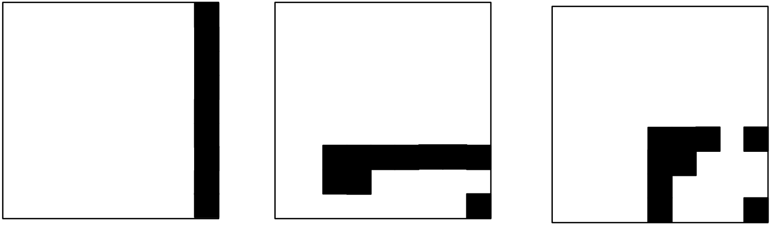
Finding Partition Shapes

- Consider all possible shapes, reduce using Push Technique



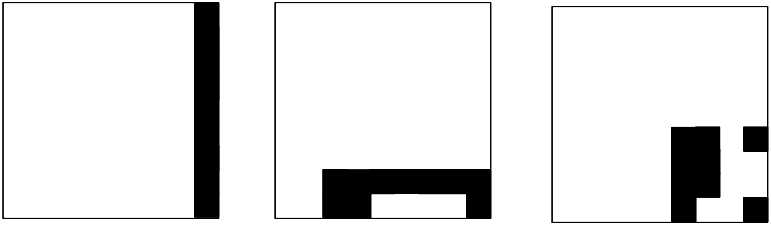
Finding Partition Shapes

- Consider all possible shapes, reduce using Push Technique



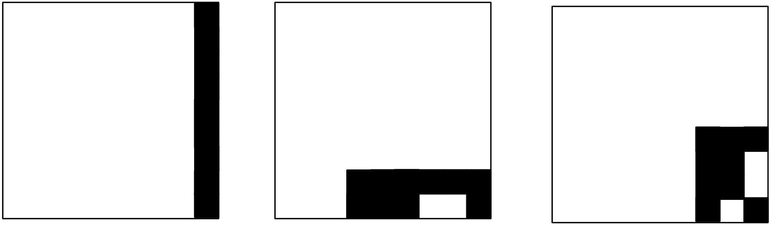
Finding Partition Shapes

- Consider all possible shapes, reduce using Push Technique



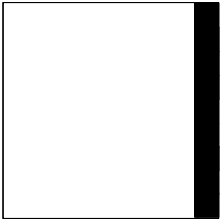
Finding Partition Shapes

- Consider all possible shapes, reduce using Push Technique

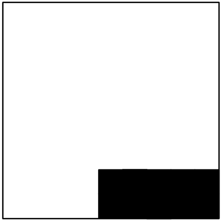


Finding Partition Shapes

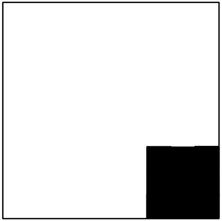
- Consider all possible shapes, reduce using Push Technique



Straight-Line

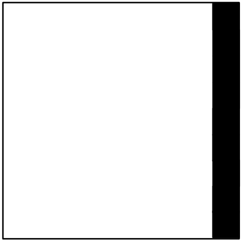


Rectangle-Corner

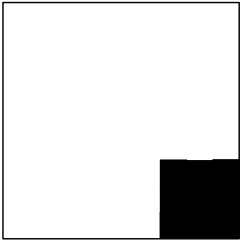


Square-Corner

Analyzing Partition Shapes



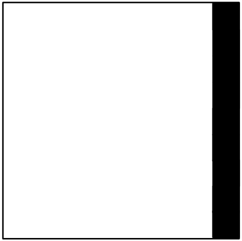
Straight-Line



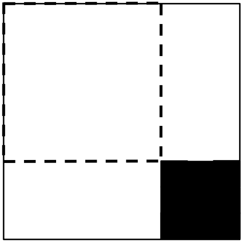
Square-Corner

Analyzing Partition Shapes

- Square-Corner optimal for all Overlap algorithms



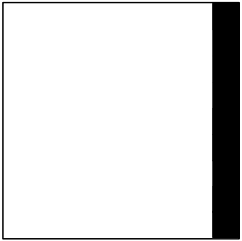
Straight-Line



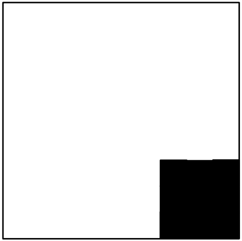
Square-Corner

Analyzing Partition Shapes

- Square-Corner optimal for all Overlap algorithms
- Square-Corner optimal for other algorithms when processor speed ratio > 3, Straight-Line optimal when speed ratio < 3



Straight-Line



Square-Corner

Ongoing and Future Work

Current Work:

- Case of 3 heterogeneous processors
- Proving Push Technique does not increase execution time
- Analyzing canonical candidate shapes for optimal shape

Future work:

- Arbitrary number of heterogeneous processors

Publications to Date:

- A. DeFlumere and A. Lastovetsky, “Theoretical Results on Optimal Partitioning for Matrix-Matrix Multiplication with Two Processors,” School of Computer Science and Informatics, University College Dublin, Technical Report UCD-CSI-2011- 09, September 2011.
- A. DeFlumere, A. Lastovetsky, and B. A. Becker, “Partitioning for Parallel Matrix-Matrix Multiplication with Heterogeneous Processors: The Optimal Solution,” 21st International Heterogeneity in Computing Workshop (HCW 2012), IEEE Computer Society, May 2012.