SparseBLAS Products in UPC: an Evaluation of Storage Formats

Jorge González-Domínguez*, Óscar García-López, Guillermo L. Taboada, María J. Martín, Juan Touriño

Computer Architecture Group
University of A Coruña (Spain)
{jgonzalezd,oscar.garcia,taboada,mariam,juan}@udc.es

International Conference on Computational and Mathematical Methods in Science and Engineering
CMMSE 2011

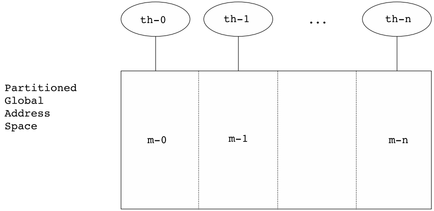
UPC: a Suitable Alternative for HPC in Multi-core Era

Programming Models:

- Traditionally: Shared/Distributed memory programming models
- Challenge: hybrid memory architectures
  - PGAS (Partitioned Global Address Space)

PGAS Languages:

- UPC -> C
- Titanium -> Java
- Co-Array Fortran -> Fortran



UPC Compilers:

- Berkeley UPC
- GCC (Intrepid)
- Michigan TU
- HP, Cray and IBM UPC Compilers

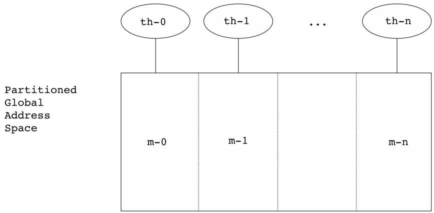
UPC: a Suitable Alternative for HPC in Multi-core Era

Programming Models:

- Traditionally: Shared/Distributed memory programming models
- Challenge: hybrid memory architectures
  - PGAS (Partitioned Global Address Space)

PGAS Languages:

- UPC -> C
- Titanium -> Java
- Co-Array Fortran -> Fortran



UPC Compilers:

- Berkeley UPC
- GCC (Intrepid)
- Michigan TU
- HP, Cray and IBM UPC Compilers

Studied Numerical Operations

BLAS Libraries

- Basic Linear Algebra Subprograms
- Specification of a set of numerical functions
- Widely used by scientists and engineers
- SparseBLAS and PBLAS (Parallel BLAS)

Studied Routines

- *usmv*: Sparse Matrix-Vector Product ($\alpha * A * x + y = y$)
- *usmm*: Sparse Matrix-Matrix Product ($\alpha * A * B + C = C$)

Studied Storage Formats

Elements ordered by rows

- Coordinate
- Compressed Sparse Row (CSR)
- Block Sparse Row (BSR)
- Skyline with lower matrices

Elements ordered by columns

- Compressed Sparse Column (CSC)
- Skyline with upper matrices

Elements ordered by diagonals

- Diagonal

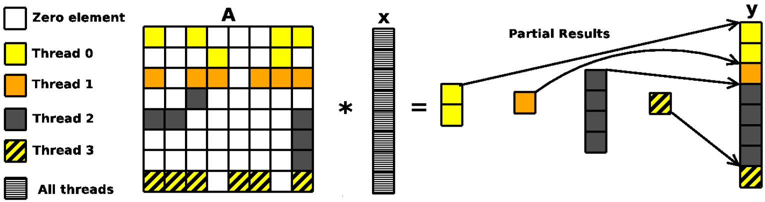
Syntax

$$\alpha * A * x + y = y$$

Structures

- α -> Scalar
- A -> Sparse matrix
- x -> Dense vector
- y -> Dense vector

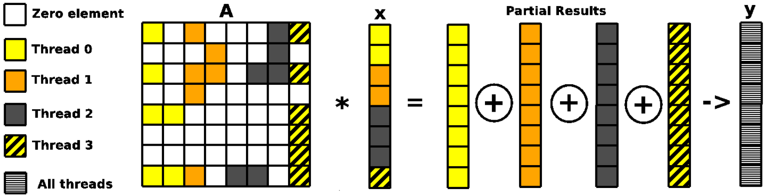
Distribution by rows



Characterstics

- Well balanced computational workload in multiplication
- Unbalanced computational workload in final additions
- Gathering of data only with one copy per thread

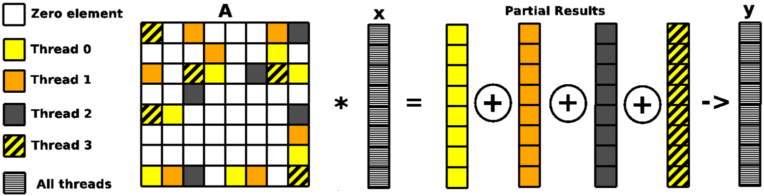
Distribution by columns



Characterstics

- Well balanced computational workload
- Gathering of data with one reduce per vector element

Distribution by diagonals



Characterstics

- Unbalanced computational workload
- Gathering of data with one reduce per vector element

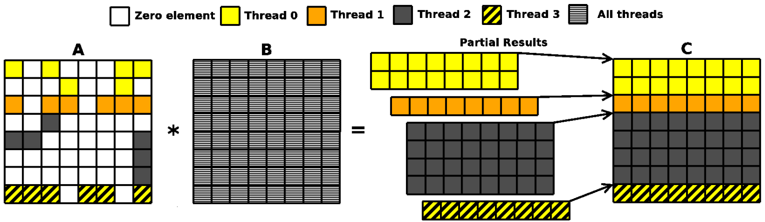
Syntax

$$\alpha * A * B + C = C$$

Structures

- α -> Scalar
- A -> Sparse matrix
- B -> Dense matrix
- C -> Dense matrix

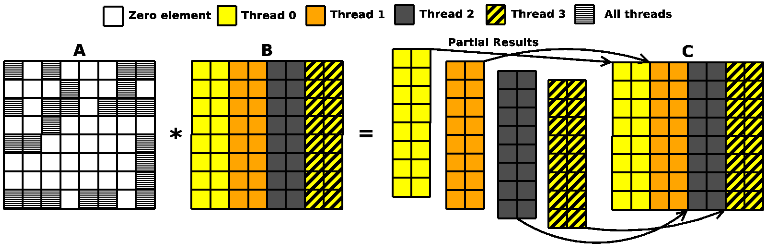
Distribution by rows



Characterstics

- Well balanced computational workload in multiplication
- Unbalanced computational workload in additions
- Gathering of data only with one copy per thread

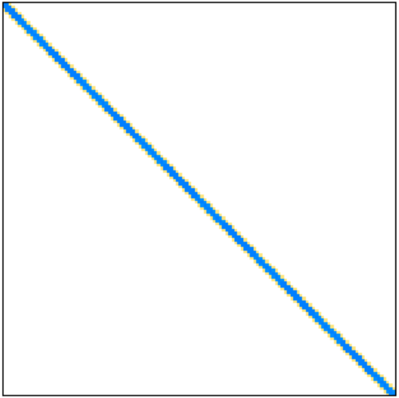
Distribution by columns



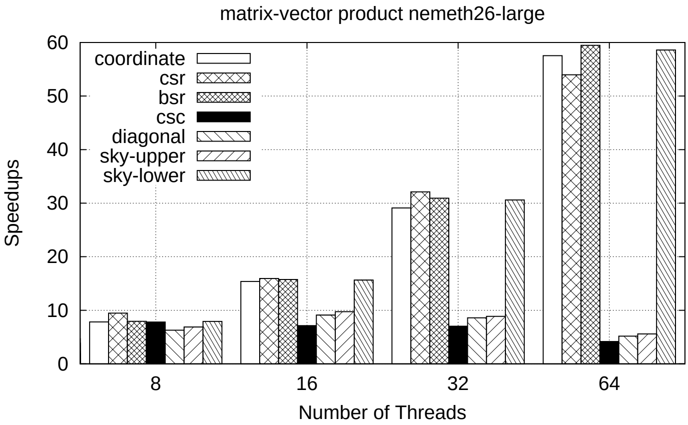
Characterstics

- Well balanced computational workload in multiplication
- Well balanced computational workload in additions
- Gathering of data with one copy per thread and row

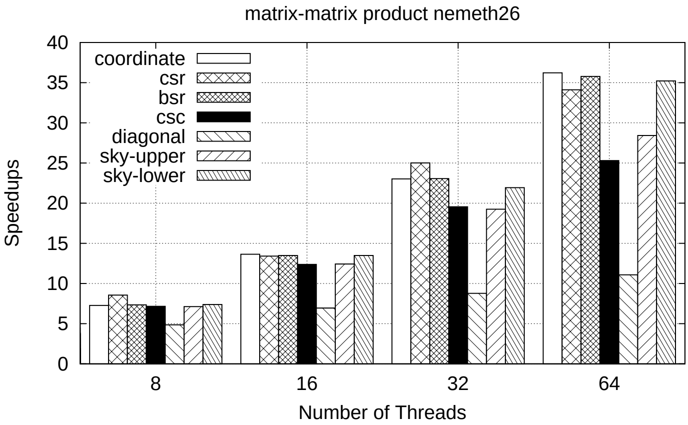
Experimental Results with a Regular Matrix (I)



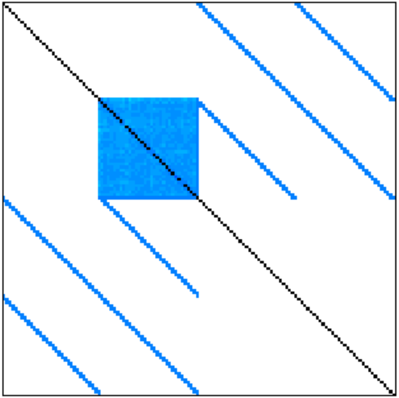
Experimental Results with a Regular Matrix (II)



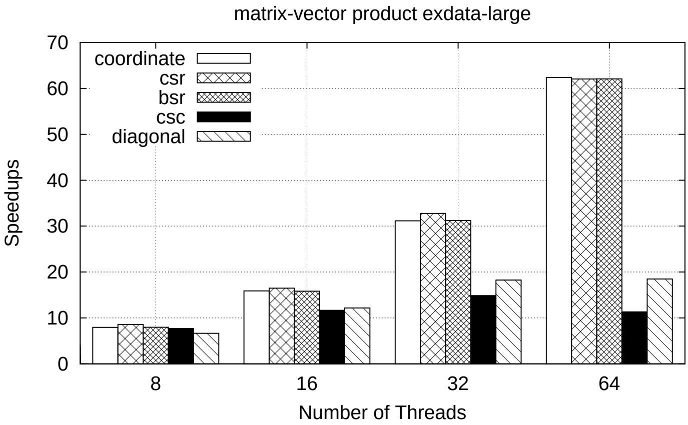
Experimental Results with a Regular Matrix (and III)



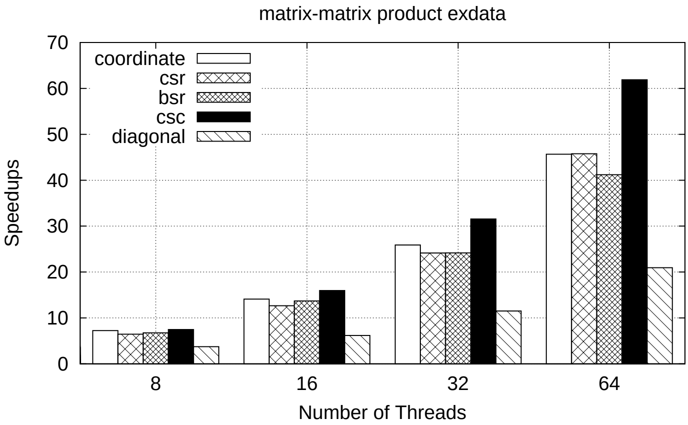
Experimental Results with an Irregular Matrix (I)



Experimental Results with an Irregular Matrix (II)



Experimental Results with an Irregular Matrix (and III)



Main Conclusions

Summary

- High speedups for both routines
- Best approach:
  - Sparse matrix-vector product -> by rows
  - Sparse matrix-matrix product
    - If regular sparse matrix -> by rows
    - If irregular sparse matrix -> by columns

Future Work

- Study the impact of performing each distribution

SparseBLAS Products in UPC: an Evaluation of Storage Formats

Jorge González-Domínguez*, Óscar García-López, Guillermo L. Taboada, María J. Martín, Juan Touriño

Computer Architecture Group
University of A Coruña (Spain)
{jgonzalezd,oscar.garcia,taboada,mariam,juan}@udc.es

International Conference on Computational and Mathematical Methods in Science and Engineering
CMMSE 2011