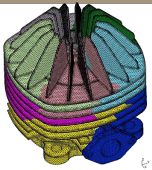
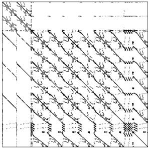
Distributed-memory BLR Factorization for Large-Scale Systems and Applications

P. Amestoy[1] A. Buttari[2] J.-Y. L'Excellent[3] T. Mary[4]

[1]INP-IRIT [2]CNRS-IRIT [3]INRIA-LIP [4]University of Manchester

SIAM PP'18, Tokyo, March 7-10

Context

 →  → 

Linear system $Ax = b$

A is large and sparse

Direct methods

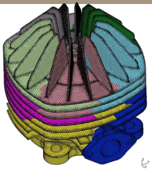
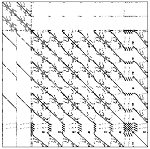
Factorize $A = LU$ and solve $LUx = b$

☺ Numerically reliable

☹ Computational cost

Theo Mary (contact: theo.mary@manchester.ac.uk)

Context

 →  → 

Linear system $Ax = b$

A is large and sparse

Direct methods

Factorize $A = LU$ and solve $LUx = b$

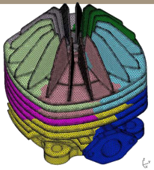
☺ Numerically reliable

☹ Computational cost

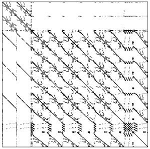
Objective of this work:
reduce the cost of sparse direct solvers ...
...while maintaining their numerical reliability

Theo Mary (contact: theo.mary@manchester.ac.uk)

Context



→



→

Large scale applications

- Target size is $n \sim 10^9$ for sparse
- $O(n^{4/3})$ memory complexity and $O(n^2)$ flop complexity
  Practical example on a 1000^3 27-point Helmholtz problem:
  15 ExaFlops and 209 TeraBytes for factors!

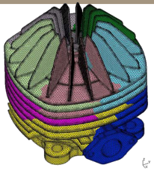
⇒ Need to reduce the asymptotic complexity

Large scale systems

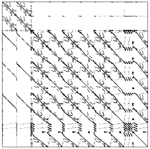
Increasingly large numbers of cores available, need to efficiently make use of them by designing parallel algorithms

Theo Mary (contact: theo.mary@manchester.ac.uk)

Context



→



→

Large scale applications

- Target size is $n \sim 10^9$ for sparse
- $O(n^{4/3})$ memory complexity and $O(n^2)$ flop complexity
  Practical example on a 1000^3 27-point Helmholtz problem:
  15 ExaFlops and 209 TeraBytes for factors!

$\Rightarrow$ Need to reduce the asymptotic complexity

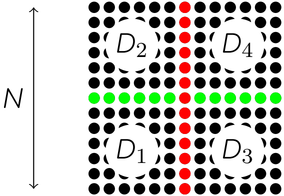
Large scale systems

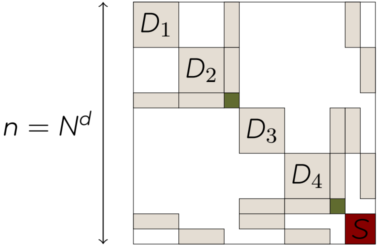
Increasingly large numbers of cores available, need to efficiently make use of them by designing parallel algorithms

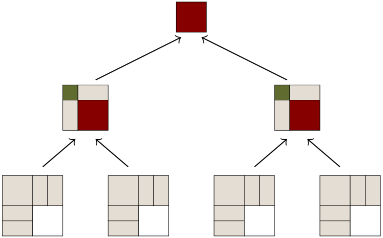
These two objectives are not necessarily compatible

Theo Mary (contact: theo.mary@manchester.ac.uk)

Introduction





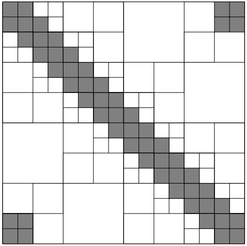


3D problem complexity

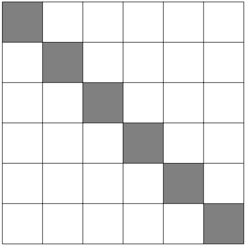
→ Flops: $\mathcal{O}\left(n^2\right)$, mem: $\mathcal{O}\left(n^{4/3}\right)$

- George. *Nested dissection of a regular finite element mesh*, SIAM J. Numer. Anal., 1973.

$\mathcal{H}$ and BLR matrices

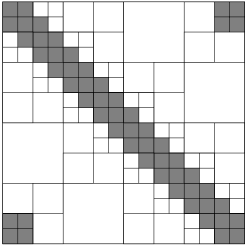


$\mathcal{H}$-matrix



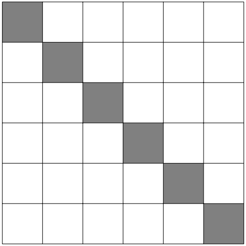
BLR matrix

$\mathcal{H}$ and BLR matrices



$\mathcal{H}$-matrix

- $O(n^{2/3}r)$ memory and $O(n^{2/3}r^2)$ flop complexity
- Complex, hierarchical structure

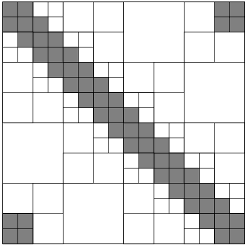


BLR matrix

- $O(nr^{1/2})$ memory and $O(n^{4/3}r)$ flop complexity
- Simple, flat structure

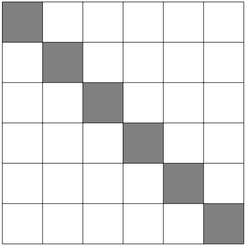
Theo Mary (contact: theo.mary@manchester.ac.uk)

$\mathcal{H}$ and BLR matrices



$\mathcal{H}$-matrix

- $O(n^{2/3}r)$ memory and $O(n^{2/3}r^2)$ flop complexity
- Complex, hierarchical structure



BLR matrix

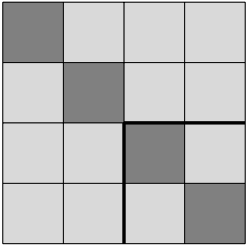
- $O(nr^{1/2})$ memory and $O(n^{4/3}r)$ flop complexity
- Simple, flat structure

Find a good comprise between complexity and performance

- Easy to handle numerical pivoting
- No global order between blocks $\Rightarrow$ flexible data distribution
- Small blocks $\Rightarrow$ can fit on single shared-memory node

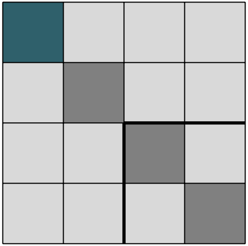
Theo Mary (contact: theo.mary@manchester.ac.uk)

BLR factorization: FCSU+LUAR variant



- FCSU:

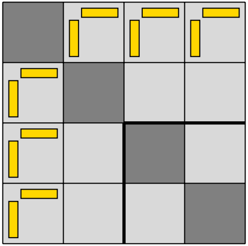
BLR factorization: FCSU+LUAR variant



- FCSU: Factor,

Theo Mary (contact: theo.mary@manchester.ac.uk)

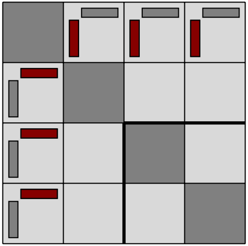
BLR factorization: FCSU+LUAR variant



- FCSU: Factor, Compress,

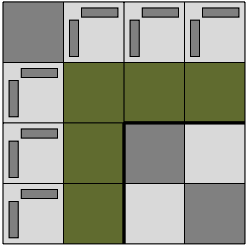
Theo Mary (contact: theo.mary@manchester.ac.uk)

BLR factorization: FCSU+LUAR variant



- FCSU: Factor, Compress, Solve,

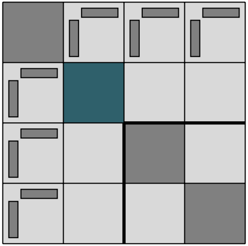
BLR factorization: FCSU+LUAR variant



- FCSU: Factor, Compress, Solve, Update

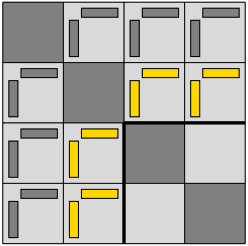
Theo Mary (contact: theo.mary@manchester.ac.uk)

BLR factorization: FCSU+LUAR variant



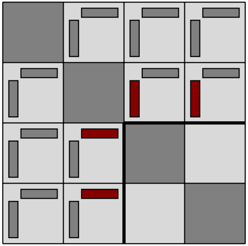
- FCSU: Factor, Compress, Solve, Update

Theo Mary (contact: theo.mary@manchester.ac.uk)



- FCSU: Factor, Compress, Solve, Update

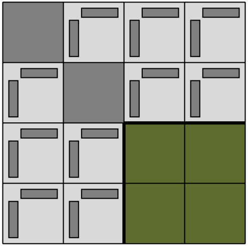
BLR factorization: FCSU+LUAR variant



- FCSU: Factor, Compress, Solve, Update

Theo Mary (contact: theo.mary@manchester.ac.uk)

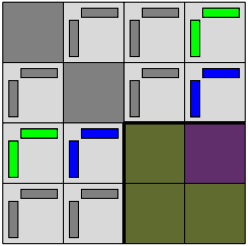
BLR factorization: FCSU+LUAR variant



- FCSU: Factor, Compress, Solve, Update

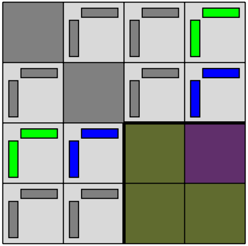
Theo Mary (contact: theo.mary@manchester.ac.uk)

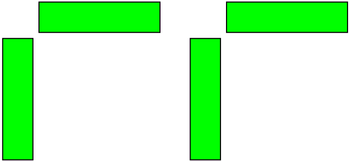
BLR factorization: FCSU+LUAR variant



- FCSU: Factor, Compress, Solve, Update

BLR factorization: FCSU+LUAR variant

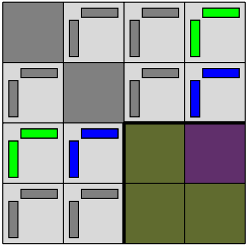


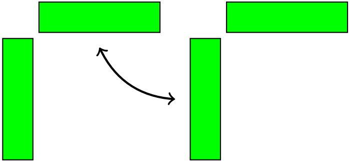


- FCSU: Factor, Compress, Solve, Update

Theo Mary (contact: theo.mary@manchester.ac.uk)

BLR factorization: FCSU+LUAR variant

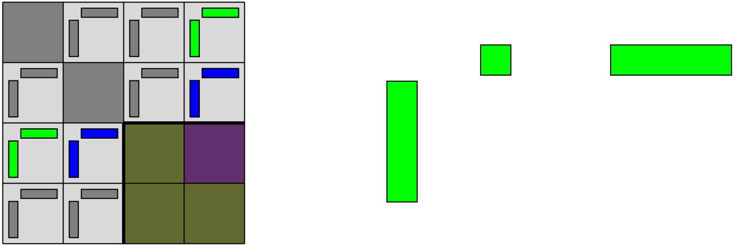




- FCSU: Factor, Compress, Solve, Update

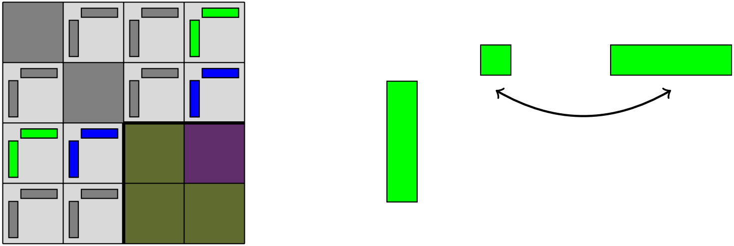
Theo Mary (contact: theo.mary@manchester.ac.uk)

BLR factorization: FCSU+LUAR variant



- FCSU: Factor, Compress, Solve, Update

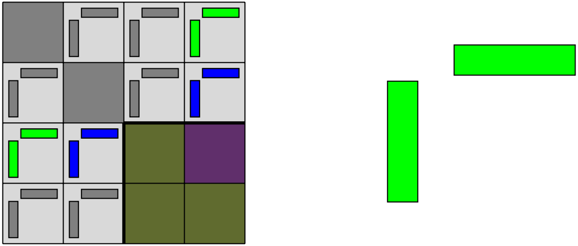
BLR factorization: FCSU+LUAR variant



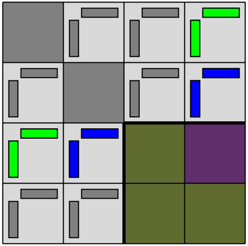
- FCSU: Factor, Compress, Solve, Update

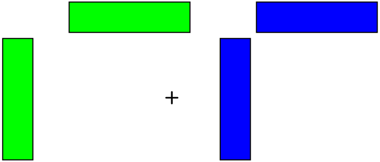
Theo Mary (contact: theo.mary@manchester.ac.uk)

BLR factorization: FCSU+LUAR variant



- FCSU: Factor, Compress, Solve, Update

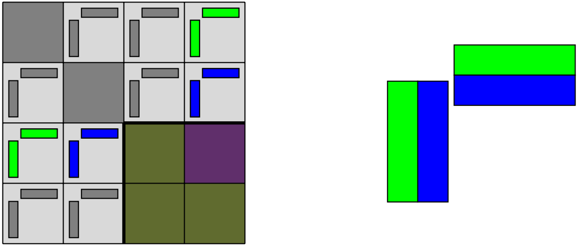




- FCSU: Factor, Compress, Solve, Update

Theo Mary (contact: theo.mary@manchester.ac.uk)

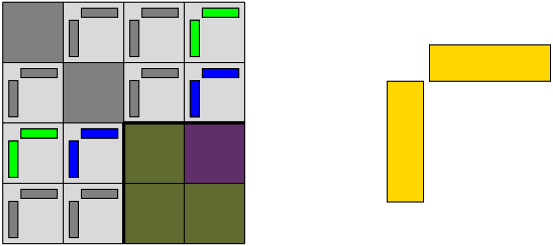
BLR factorization: FCSU+LUAR variant



- FCSU: Factor, Compress, Solve, Update
- LUAR: Low-rank Updates Accumulation

Theo Mary (contact: theo.mary@manchester.ac.uk)

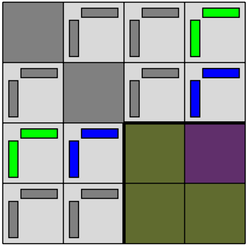
BLR factorization: FCSU+LUAR variant

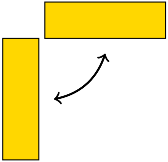


- FCSU: Factor, Compress, Solve, Update
- LUAR: Low-rank Updates Accumulation and Recompression

Theo Mary (contact: theo.mary@manchester.ac.uk)

BLR factorization: FCSU+LUAR variant

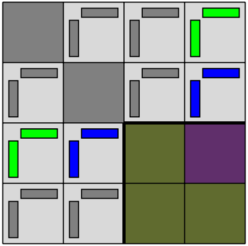


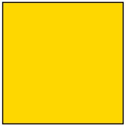


- FCSU: Factor, Compress, Solve, Update
- LUAR: Low-rank Updates Accumulation and Recompression

Theo Mary (contact: theo.mary@manchester.ac.uk)

BLR factorization: FCSU+LUAR variant

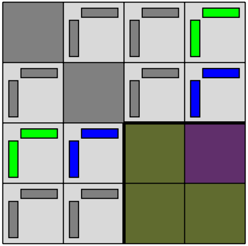


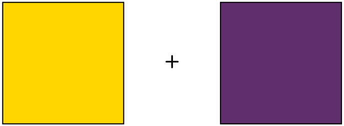


- FCSU: Factor, Compress, Solve, Update
- LUAR: Low-rank Updates Accumulation and Recompression

Theo Mary (contact: theo.mary@manchester.ac.uk)

BLR factorization: FCSU+LUAR variant

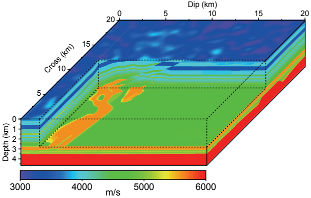




- FCSU: Factor, Compress, Solve, Update
- LUAR: Low-rank Updates Accumulation and Recompression

Theo Mary (contact: theo.mary@manchester.ac.uk)

Experimental Setting: Applications



3D Seismic Modeling

Helmholtz equation

Single complex (c) arithmetic

Unsymmetric *LU* factorization

Required accuracy: $\varepsilon = 10^{-3}$

Credits: SEISCOPE

matrix	n	nnz	flops	storage
10Hz	17M	446M	2.6 PF	0.7 TB
15Hz	58M	1523M	29.6 PF	3.7 TB
20Hz	130M	3432M	150.0 PF	11.0 TB

Full-Rank statistics

- Amestoy, Brossier, Buttari, L'Excellent, Mary, Métivier, Miniussi, and Operto. *Fast 3D frequency-domain full waveform inversion with a parallel Block Low-Rank multifrontal direct solver: application to OBC data from the North Sea*, Geophysics, 2016.

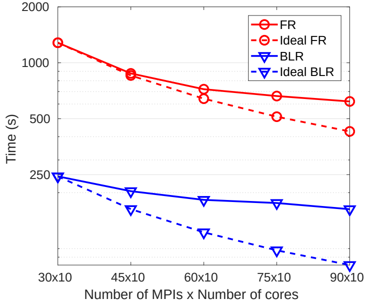
Experimental Setting: Systems

1. Experiments on matrices 10Hz and 15Hz are done on the `eos` supercomputer at the CALMIP center of Toulouse (grant P0989):
   - Two Intel(r) 10-cores Ivy Bridge @ 2,8 GHz
   - Peak per core is 22.4 GF/s
   - 64 GB memory per node
   - Infiniband FDR interconnect

2. Experiments on matrix 20Hz are done on the `occigen` supercomputer at the CINES center of Montpellier:
   - Two Intel(r) 12-cores Haswell @ 2,6 GHz
   - Peak per core is 41.6 GF/s
   - 128 GB memory per node
   - Infiniband FDR interconnect

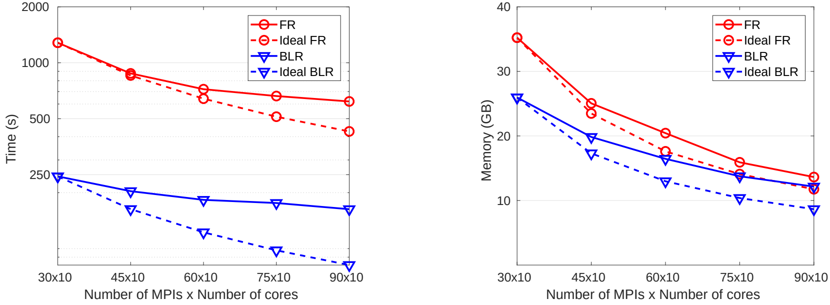
Theo Mary (contact: theo.mary@manchester.ac.uk)

Strong scalability analysis (matrix 10Hz)



Three challenges to improve the scalability of the BLR factorization:

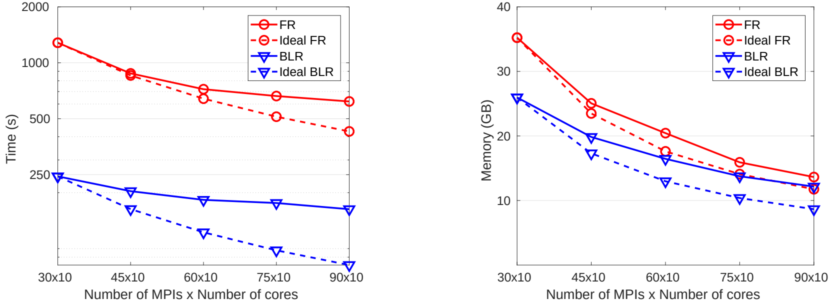
Strong scalability analysis (matrix 10Hz)



Three challenges to improve the scalability of the BLR factorization:

1. The communications challenge: flops reduced by 12.8 but volume of comms only by $2.2 \Rightarrow$ higher weight of comms

Strong scalability analysis (matrix 10Hz)

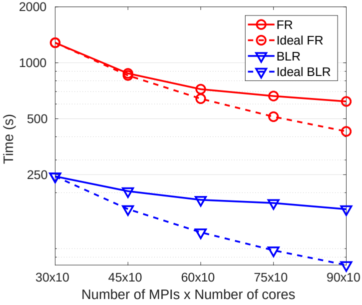


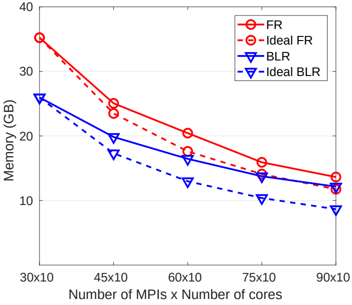
Three challenges to improve the scalability of the BLR factorization:

1. The communications challenge: flops reduced by 12.8 but volume of comms only by $2.2 \Rightarrow$ higher weight of comms
2. The load imbalance challenge: ratio between most and less loaded processes increases from 1.3 (FR) to 2.6 (BLR)

Theo Mary (contact: theo.mary@manchester.ac.uk)

Strong scalability analysis (matrix 10Hz)





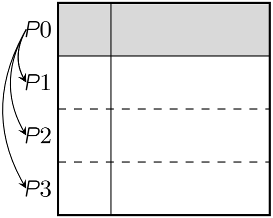
Three challenges to improve the scalability of the BLR factorization:

1. The communications challenge: flops reduced by 12.8 but volume of comms only by $2.2 \Rightarrow$ higher weight of comms
2. The load imbalance challenge: ratio between most and less loaded processes increases from 1.3 (FR) to 2.6 (BLR)
3. The memory challenge

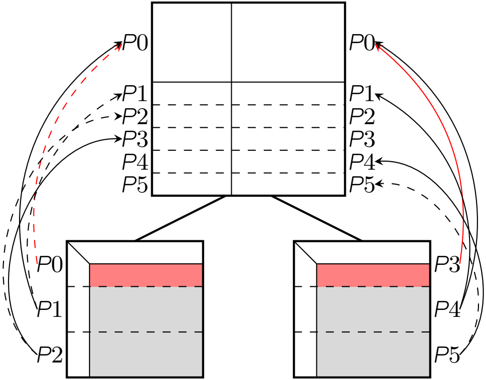
Theo Mary (contact: theo.mary@manchester.ac.uk)

The communications challenge

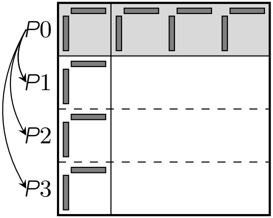
Type of messages



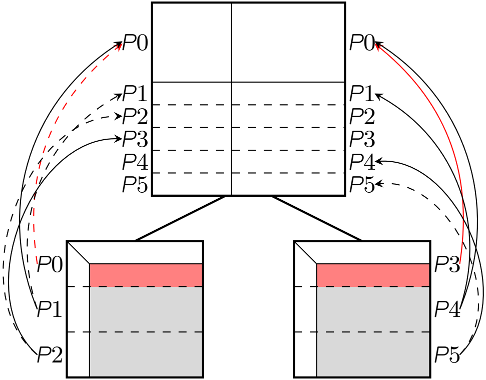
LU messages



CB messages



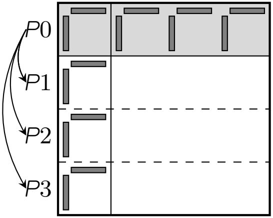
LU messages



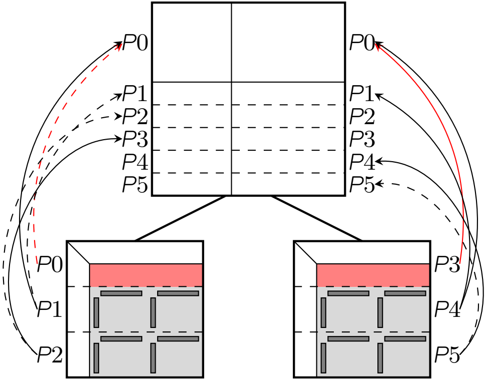
CB messages

- Volume of *LU* messages is reduced by compressing the factors
  - ☺ Reduces operation count, communications, and memory consumption

Type of messages



LU messages

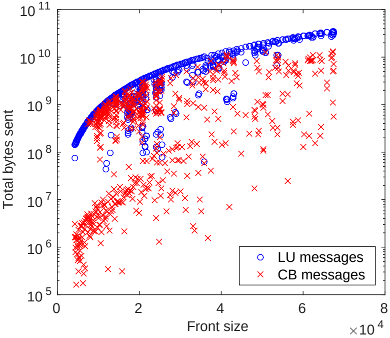


CB messages

- Volume of *LU* messages is reduced by compressing the factors
  - ☺ Reduces operation count, communications, and memory consumption
- Volume of CB messages can be reduced by compressing the CB
  - ☺ Reduces communications and memory consumption
  - ☹ Increases operation count unless assembly is done in LR

Theo Mary (contact: theo.mary@manchester.ac.uk)

Communication analysis

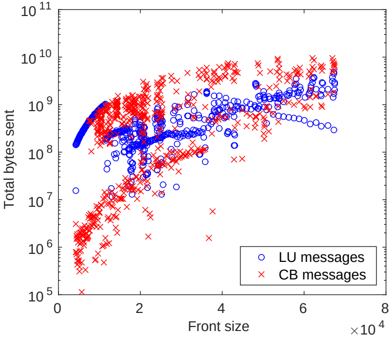


- FR case: *LU* messages dominate

Theoretical communication bounds

	$\mathcal{W}_{LU}$	$\mathcal{W}_{CB}$	$\mathcal{W}_{tot}$
FR	$\mathcal{O}\left(n^{4/3}p\right)$	$\mathcal{O}\left(n^{4/3}\right)$	$\mathcal{O}\left(n^{4/3}p\right)$

Theo Mary (contact: theo.mary@manchester.ac.uk)

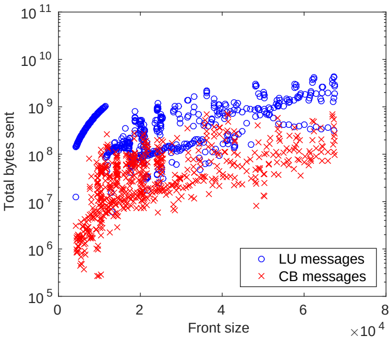


- FR case: *LU* messages dominate
- BLR case: CB messages dominate ⇒ underwhelming reduction of communications

Theoretical communication bounds

	$\mathcal{W}_{LU}$	$\mathcal{W}_{CB}$	$\mathcal{W}_{tot}$
FR	$\mathcal{O}\left(n^{4/3}p\right)$	$\mathcal{O}\left(n^{4/3}\right)$	$\mathcal{O}\left(n^{4/3}p\right)$
BLR (CB_{FR})	$\mathcal{O}\left(nr^{1/2}p\right)$	$\mathcal{O}\left(n^{4/3}\right)$	$\mathcal{O}\left(nr^{1/2}p + n^{4/3}\right)$

Communication analysis



- FR case: *LU* messages dominate
- BLR case: CB messages dominate ⇒ underwhelming reduction of communications

⇒ CB compression allows for truly reducing the communications

Theoretical communication bounds

	$\mathcal{W}_{LU}$	$\mathcal{W}_{CB}$	$\mathcal{W}_{tot}$
FR	$\mathcal{O}\left(n^{4/3}p\right)$	$\mathcal{O}\left(n^{4/3}\right)$	$\mathcal{O}\left(n^{4/3}p\right)$
BLR (CB_{FR})	$\mathcal{O}\left(nr^{1/2}p\right)$	$\mathcal{O}\left(n^{4/3}\right)$	$\mathcal{O}\left(nr^{1/2}p + n^{4/3}\right)$
BLR (CB_{LR})	$\mathcal{O}\left(nr^{1/2}p\right)$	$\mathcal{O}\left(nr^{1/2}\right)$	$\mathcal{O}\left(nr^{1/2}p\right)$

Theo Mary (contact: theo.mary@manchester.ac.uk)

Performance impact of CB compression

matrix	10Hz	15Hz	20Hz
order	17 M	58 M	130 M
cores	900 Ivy Bridge	900 Ivy Bridge	2,400 Haswell
computer	eos (CALMIP)	eos (CALMIP)	occigen (CINES)
factor flops (FR)	2.6 PF	29.6 PF	150.0 PF
⇒ BLR (CB_{FR})	0.1 PF (5.3%)	1.0 PF (3.3%)	3.6 PF (2.4%)
⇒ BLR (CB_{LR})	0.2 PF (6.1%)	1.1 PF (3.7%)	3.9 PF (2.6%)
factor time (FR)	601	5,206	n/a
⇒ BLR (CB_{FR})	123 (4.9)	838 (6.2)	1,665
⇒ BLR (CB_{LR})	213 (2.8)	856 (6.1)	2,641
CB_{LR} time impact	+73%	+2%	+58%
comm. volume (FR)	5.3 TB	29.6 TB	n/a
comm. volume (CB_{FR})	1.7 TB (3.2)	13.3 TB (2.2)	79.8 TB
comm. volume (CB_{LR})	0.6 TB (9.1)	1.2 TB (23.2)	8.6 TB

⇒ CB compression becomes increasingly critical?

Theo Mary (contact: theo.mary@manchester.ac.uk)

Performance impact of CB compression

matrix	10Hz	15Hz	20Hz
order	17 M	58 M	130 M
cores	900 Ivy Bridge	900 Ivy Bridge	2,400 Haswell
computer	eos (CALMIP)	eos (CALMIP)	occigen (CINES)
factor flops (FR)	2.6 PF	29.6 PF	150.0 PF
⇒ BLR (CB_{FR})	0.1 PF (5.3%)	1.0 PF (3.3%)	3.6 PF (2.4%)
⇒ BLR (CB_{LR})	0.2 PF (6.1%)	1.1 PF (3.7%)	3.9 PF (2.6%)
factor time (FR)	601	5,206	n/a
⇒ BLR (CB_{FR})	123 (4.9)	838 (6.2)	1,665
⇒ BLR (CB_{LR})	213 (2.8)	856 (6.1)	2,641
CB_{LR} time impact	+73%	+2%	+58%
comm. volume (FR)	5.3 TB	29.6 TB	n/a
comm. volume (CB_{FR})	1.7 TB (3.2)	13.3 TB (2.2)	79.8 TB
comm. volume (CB_{LR})	0.6 TB (9.1)	1.2 TB (23.2)	8.6 TB

⇒ CB compression becomes increasingly critical?

Theo Mary (contact: theo.mary@manchester.ac.uk)

Performance impact of CB compression

matrix	10Hz	15Hz	20Hz
order	17 M	58 M	130 M
cores	900 Ivy Bridge	900 Ivy Bridge	2,400 Haswell
computer	eos (CALMIP)	eos (CALMIP)	occigen (CINES)
factor flops (FR)	2.6 PF	29.6 PF	150.0 PF
⇒ BLR (CB_{FR})	0.1 PF (5.3%)	1.0 PF (3.3%)	3.6 PF (2.4%)
⇒ BLR (CB_{LR})	0.2 PF (6.1%)	1.1 PF (3.7%)	3.9 PF (2.6%)
factor time (FR)	601	5,206	n/a
⇒ BLR (CB_{FR})	123 (4.9)	838 (6.2)	1,665
⇒ BLR (CB_{LR})	213 (2.8)	856 (6.1)	2,641
CB_{LR} time impact	+73%	+2%	+58%
comm. volume (FR)	5.3 TB	29.6 TB	n/a
comm. volume (CB_{FR})	1.7 TB (3.2)	13.3 TB (2.2)	79.8 TB
comm. volume (CB_{LR})	0.6 TB (9.1)	1.2 TB (23.2)	8.6 TB

⇒ CB compression becomes increasingly critical?

Theo Mary (contact: theo.mary@manchester.ac.uk)

The memory challenge

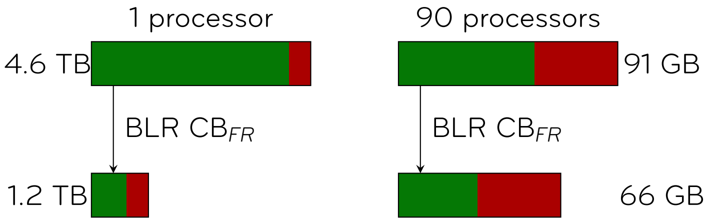
Memory scalability analysis

Memory consumption on matrix 15Hz: **factors** + **active memory** (**CB** + **active front**)

Theo Mary (contact: theo.mary@manchester.ac.uk)

Memory scalability analysis

Memory consumption on matrix 15Hz: **factors** + **active memory** (**CB** + **active front**)

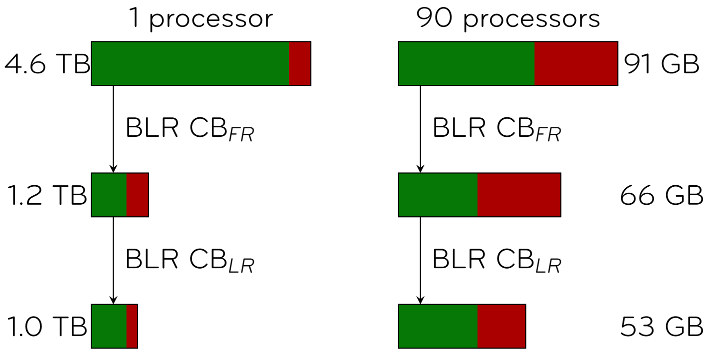


- Factors compression (19% of FR) leads to important gains, but the BLR solver inherits the poor scalability of the active memory

Theo Mary (contact: theo.mary@manchester.ac.uk)

Memory scalability analysis

Memory consumption on matrix 15Hz: **factors** + **active memory** (**CB** + **active front**)

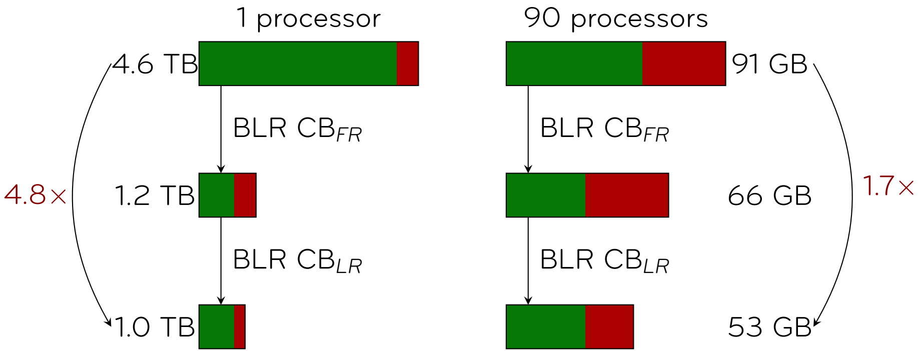


- Factors compression (19% of FR) leads to important gains, but the BLR solver inherits the poor scalability of the active memory
- CB compression (7% of FR) slightly attenuates this issue

Theo Mary (contact: theo.mary@manchester.ac.uk)

Memory scalability analysis

Memory consumption on matrix 15Hz: **factors** + **active memory** (**CB** + **active front**)

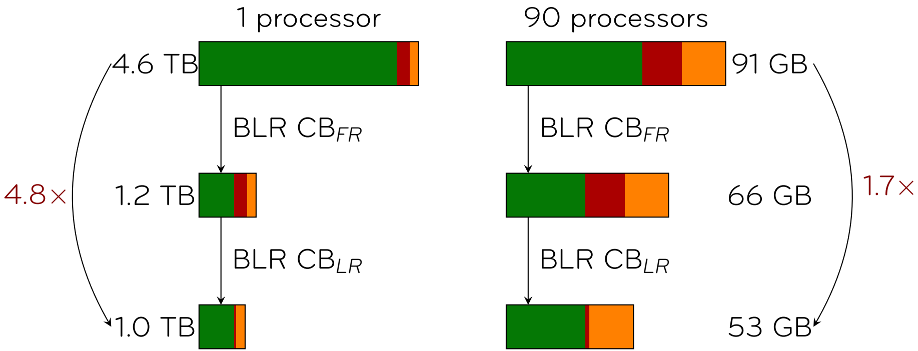


- Factors compression (19% of FR) leads to important gains, but the BLR solver inherits the poor scalability of the active memory
- CB compression (7% of FR) slightly attenuates this issue

Theo Mary (contact: theo.mary@manchester.ac.uk)

Memory scalability analysis

Memory consumption on matrix 15Hz: **factors** + **active memory** (**CB** + **active front**)



- Factors compression (19% of FR) leads to important gains, but the BLR solver inherits the poor scalability of the active memory
- CB compression (7% of FR) slightly attenuates this issue
- Storage for the active front becomes critical

Conclusion

Summary: a distributed-memory BLR solver...

...to reduce time to solution

- On 58 millions problem, 6× time gains on 900 cores
- Much room left for improvement (30× flops potential!)

...to reduce memory consumption

- On 58 millions problem, 40% memory gains on 900 cores
- Thanks to CB compression: 25% → 40%
- Also much room left for improvement (80% gain in sequential!)

...to solve larger problems

- 130 millions problem on 2400 cores in less than an hour
- What do we need to go one order of magnitude larger?

Theo Mary (contact: theo.mary@manchester.ac.uk)

Perspectives

Improving the memory scalability

- Active front becomes dominant and limits memory scalability:
  - Switch to fully-structured (matrix-free) implementation?
  - Panel by panel allocation and compression
- Memory aware mappings: map critical fronts on more processes to improve memory scalability

Improving the load balance

- How to deal with the unpredictability of low-rank compression?
- Can we do more than heuristics?
- Dynamic scheduling and asynchronicity will be important

Improving the asymptotic complexity

- Multilevel BLR format: add just a few more levels

Theo Mary (contact: theo.mary@manchester.ac.uk)

References

Publications

- Theo Mary. *Block Low-Rank Multifrontal Solvers: Complexity, Performance, and Scalability*, PhD thesis, 2017.
- Amestoy, Buttari, L'Excellent, and Mary. *On the Complexity of the Block Low-Rank Multifrontal Factorization*, SIAM J. Sci. Comput., 2017.
- Amestoy, Buttari, L'Excellent, and Mary. *Performance and Scalability of the Block Low-Rank Multifrontal Factorization on Multicore Architectures*, under review in ACM Trans. Math. Soft., 2017.
- Amestoy, Brossier, Buttari, L'Excellent, Mary, Métivier, Miniussi, and Operto. *Fast 3D frequency-domain full waveform inversion with a parallel Block Low-Rank multifrontal direct solver: application to OBC data from the North Sea*, Geophysics, 2016.
- Shantsev, Jaysaval, de la Kethulle de Ryhove, Amestoy, Buttari, L'Excellent, and Mary. *Large-scale 3D EM modeling with a Block Low-Rank multifrontal direct solver*, Geophysical Journal International, 2017.

Software

- MUMPS 5.1.2

Theo Mary (contact: theo.mary@manchester.ac.uk)

Thank you for your attention

Slides available here:

personalpages.manchester.ac.uk/staff/theo.mary/