Leveraging Computational Astronomy Applications on Massively Parallel Hardware Systems

Hatem Ltaief

Senior Research Scientist
Extreme Computing Research Center, KAUST, Saudi Arabia

IPAM workshop: New Architectures and Algorithms
November 26-30, 2018

Outline

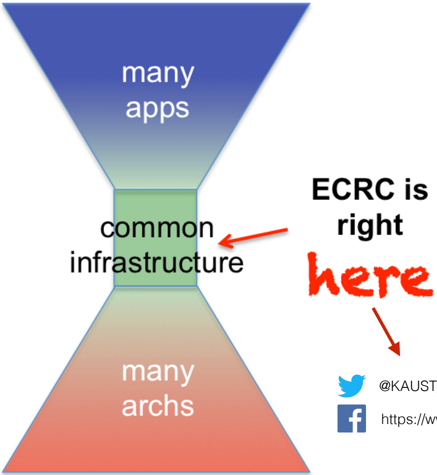
Students/Collaborators

- Extreme Computing Research Center @ KAUST
  K. Akbudak, R. AlOmairy, A. Charara, D. Keyes, A. Mikhalev, E. A. Gonzalez Fisher, and D. Sukkari
- L'Observatoire de Paris, LESIA
  N. Doucet, E. Gendron, D. Gratadour, and A. Sevin
- Subaru Telescope, National Astronomical Observatory of Japan
  O. Guyon and J. Lozi
- Mathematical Institute, University of Oxford, UK
  Y. Nakatsukasa
- Innovative Computing Laboratory @ UTK
  PLASMA/MAGMA/PaRSEC Teams
- INRIA/INP Bordeaux, France
  Runtime/HiePACS Teams

Vendors

- NVIDIA GPU Research Center
- Intel Parallel Computing Center
- Cray Center of Excellence

The Hourglass Revisited

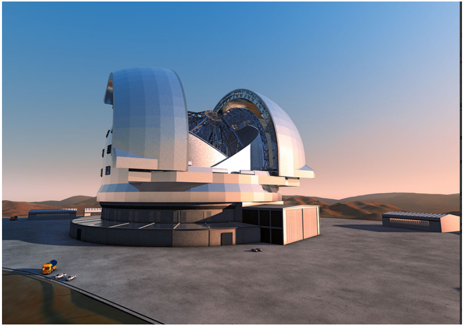


@KAUST_ECRC

https://www.facebook.com/ecrckaust

Exciting Time for Astronomy at KAUST/ECRC!

- Supporting two major worldwide ground-based astronomy efforts:

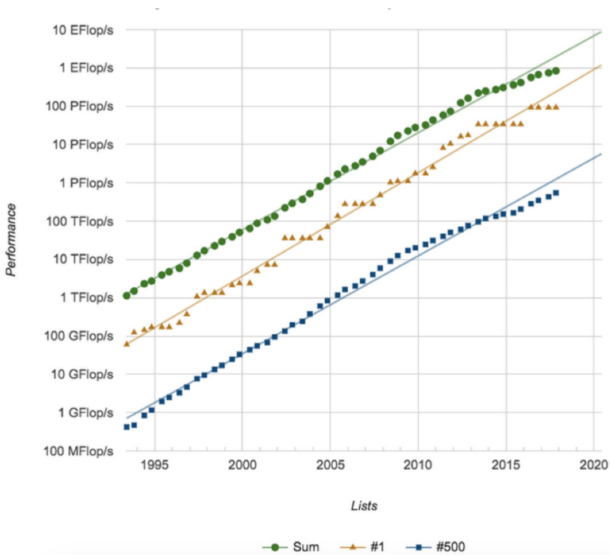


The E-ELT Telescope

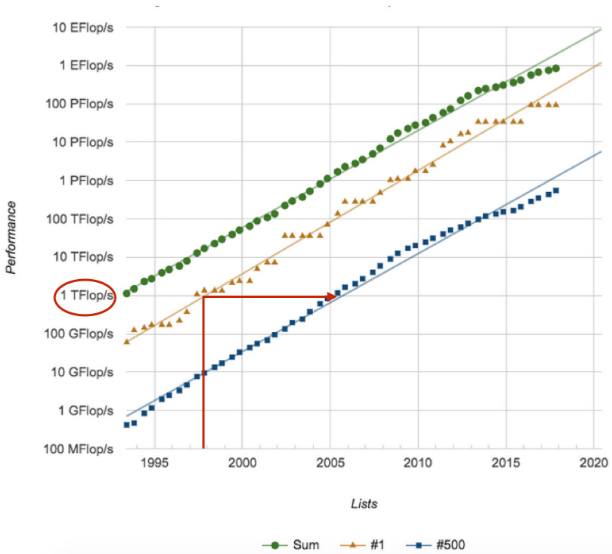
The Subaru Telescope

Outline

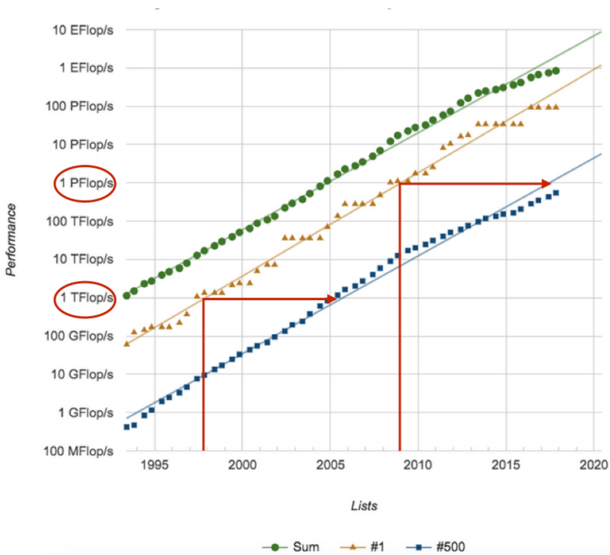
High Performance Computing: The Top500 List



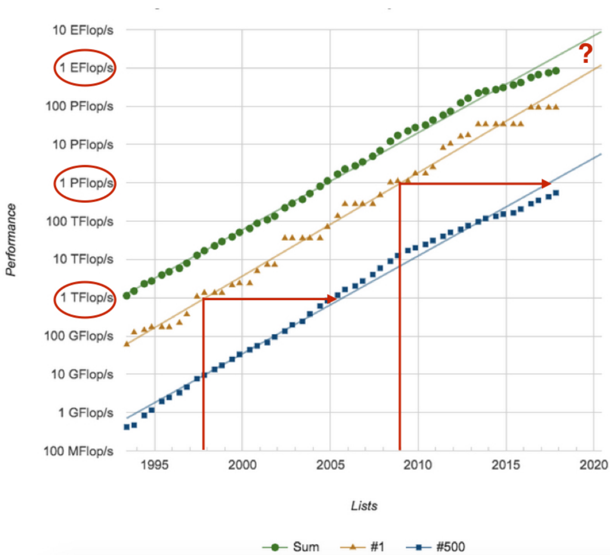
High Performance Computing: The Top500 List



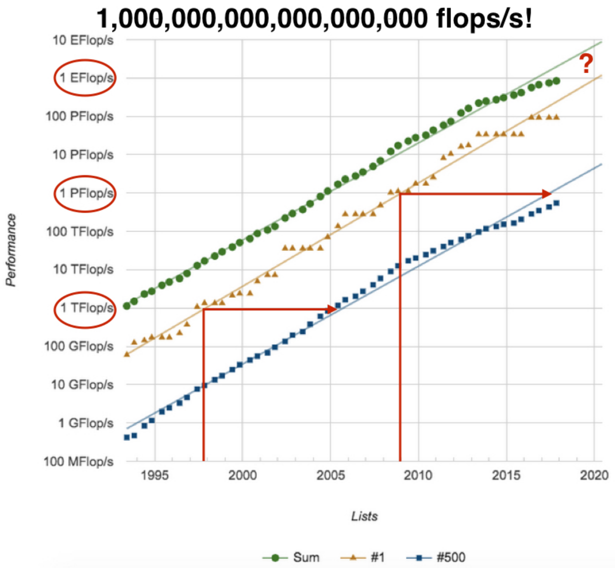
High Performance Computing: The Top500 List



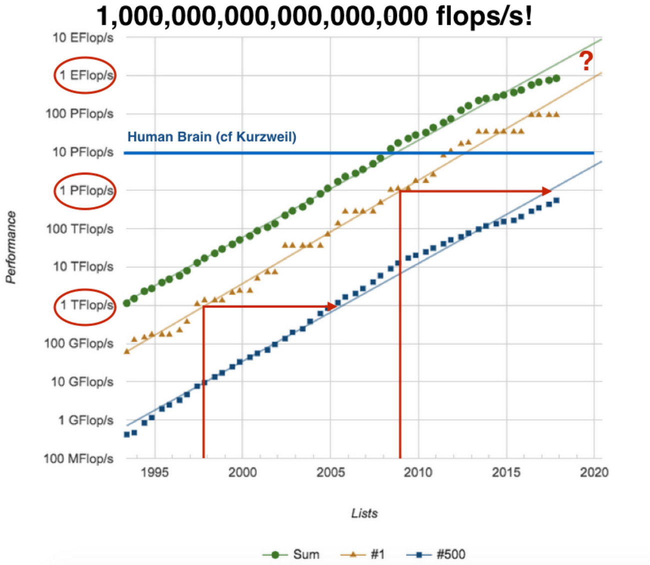
High Performance Computing: The Top500 List



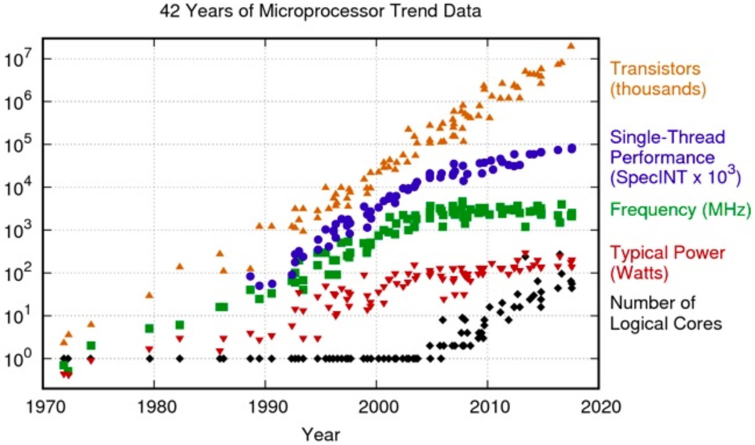
High Performance Computing: The Top500 List



High Performance Computing: The Top500 List



It is getting *Moore and Moore* hot here!



Original data up to the year 2010 collected and plotted by M. Horowitz, F. Labonte, O. Shacham, K. Olukotun, L. Hammond, and C. Batten
New plot and data collected for 2010-2017 by K. Rupp

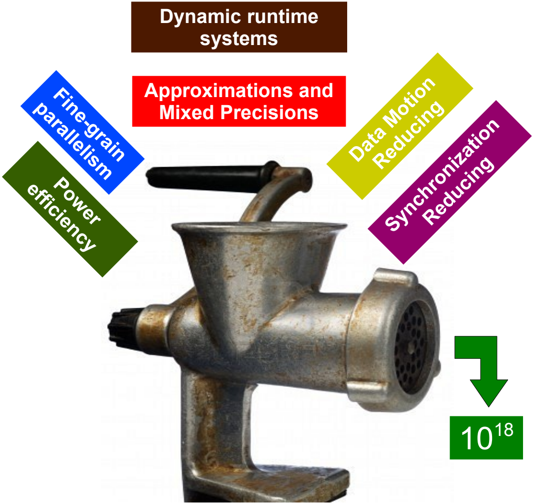
https://www.karlrupp.net/2018/02/42-years-of-microprocessor-trend-data/

Hardware Trends: Energy / Data Movement Matters!

	2011	2018
DP FLOP	100 pJ	10 pJ
DP DRAM Read	4800 pJ	1920 pJ
Local interconnect	7500 pJ	2500 pJ
Cross system	9000 pJ	3500 pJ

John Shalf, LBNL

Algorithmic Recipes For Exascale Computing



Outline

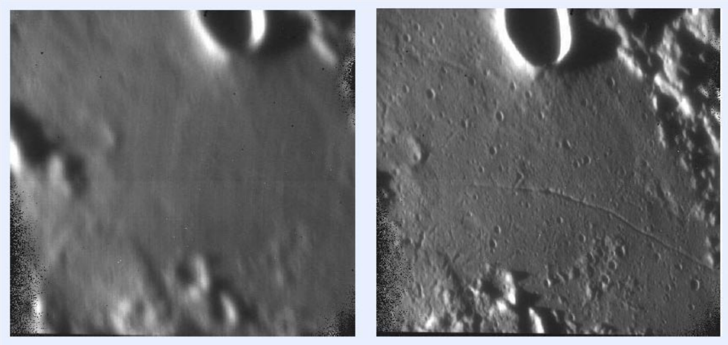
The effect of atmospheric turbulence

The sun observed with a compact camera

- Disturbs the trajectory of light rays (wavefront perturbations)
- Reduces astronomical images quality

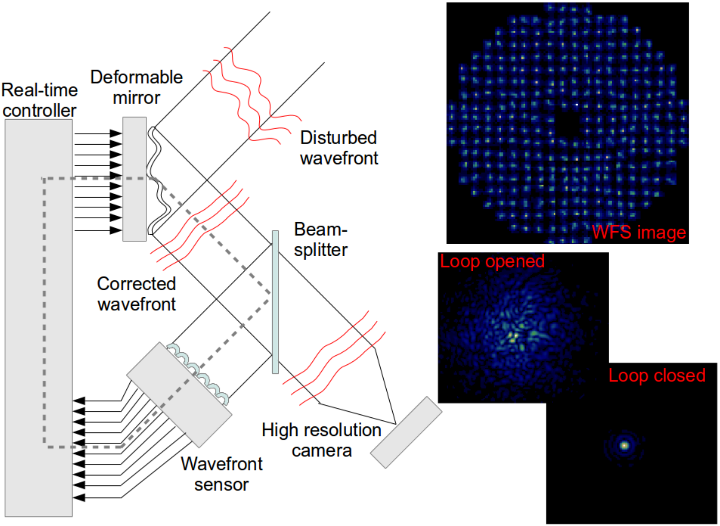
Adaptive Optics (AO)

- AO: technique used compensate in real-time the wavefront perturbations providing a significant improvement in resolution

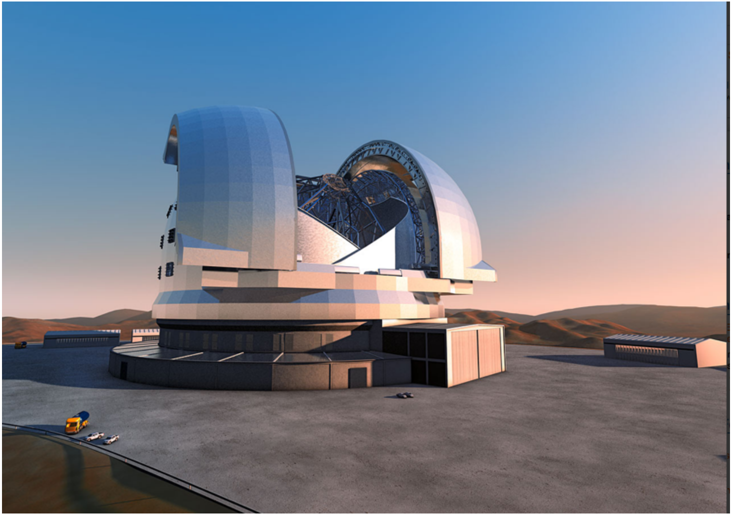


The moon observed with a 8m telescope (left: no AO, right: with AO)

How AO works


Real-time
controller
Deformable
mirror
Disturbed
wavefront
Beam-
splitter
Corrected
wavefront
High resolution
camera
Wavefront
sensor
WFS image
Loop opened
Loop closed

The World's Biggest Eye on The Sky



Credits: ESO (http://www.eso.org/public/teles-instr/e-elt/)

And here comes the linear algebra...

- Multiple-Object AO approach (MOAO)
- Compute the tomographic reconstructor matrix using covariance matrix between direction specific *truth* sensor and other sensors and the inverse of measurements covariance matrix
- $R' = C_{tm} \cdot C_{mm}^{-1}$
- Factorize and Solve for R' with C_{mm}, a 100k x 100k matrix, is extremely compute intensive
- At the core of system operations (soft real-time, should be achieved in seconds to update the real-time box)
- Also a critical component for the numerical simulation of the system behavior (observation forecast for today's design studies)

The Subaru Telescope

And here comes *again* the linear algebra...

- Compute the pseudo inverse A^+

$$AA^+A = A \, , A \in \mathbb{R}^{m \times n} (m \geq n)$$

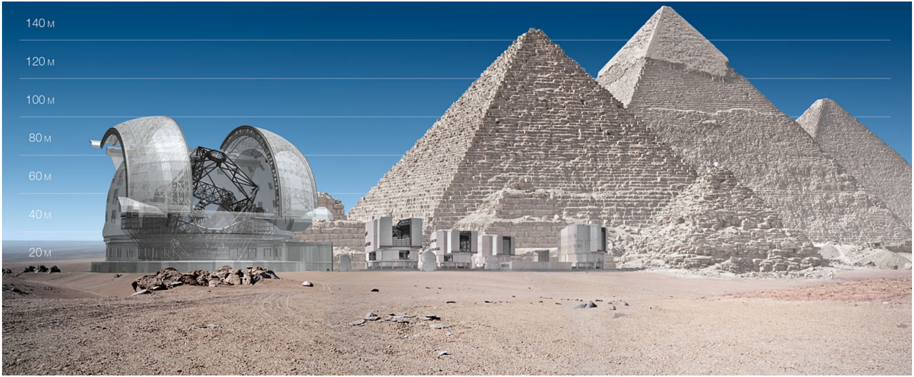
- The numerical challenge of the pseudo inverse are twofold:
  - Numerical: dealing with rectangular matrix which may engender numerical instabilities
  $$A^\top A \; = \; V \Lambda V^\top$$
  - Computational: high algorithmic complexity, it should still be able to keep up with the overall throughput of the AO framework
- Using SVD: $A \; = U \Sigma V^\top$ then:

$$A^+ \; = \; V \Sigma^{-1} U^\top$$

- Only most significant singular values with their associated singular vectors are required ($\approx$**10%**)

Outline

The World's Biggest Eye on The Sky



Credits: ESO (http://www.eso.org/public/teles-instr/e-elt/)

- The largest optical/near-infrared telescope in the world.
- It will weigh about 2700 tons with a main mirror diameter of 39m.
- Location: Chile, South America.

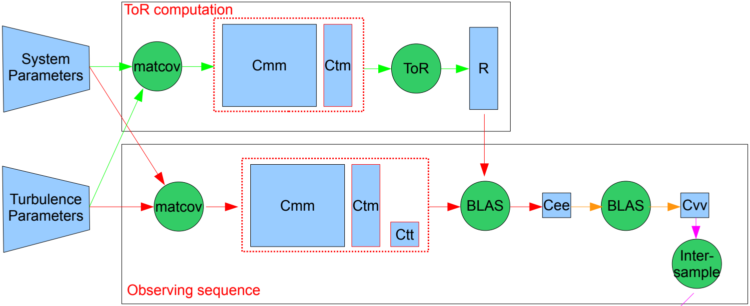
The Top 10 (present and future) Radio and Optical Ground-based Telescopes

Rank	**Name**	**Location**	**Diameter**	**Cost**	**Year**
10	Large Synoptic Survey Telescope (LSST)	Chile	8.4m	450 million	2014
9	South African Large Telescope (SALT)	South Africa	9.2m	36 million	2005
8	Keck	USA	10m	100 million	1996
7	Gran Telescopio Canarias (GTC)	Spain	10.4m	130 million	2009
6	Aricebo Observatory	Puerto Rico	305m	9.3 million	1963
5	Atacama Large Millimeter Array (ALMA)	Chile	12m	1.4 billion	2013
4	Giant Magellan Telescope (GMT)	Chile	24.5m	2.2 billion	2024
3	Thirty Meter Telescope (TMT)	USA	30m	1.4 billion	2030
2	Square Kilometer Array (SKA)	Australia	*90m*	2 billion	2020
1	**European Extremely Large Telescope (E-ELT)**	**Chile**	**39m**	**1.3 billion**	**2024**

Consortium: multiple nation initiatives

Src: http://www.space.com/14075-10-biggest-telescopes-earth-comparison.html

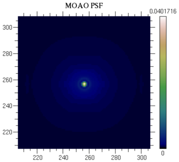
Global Workflow Chart



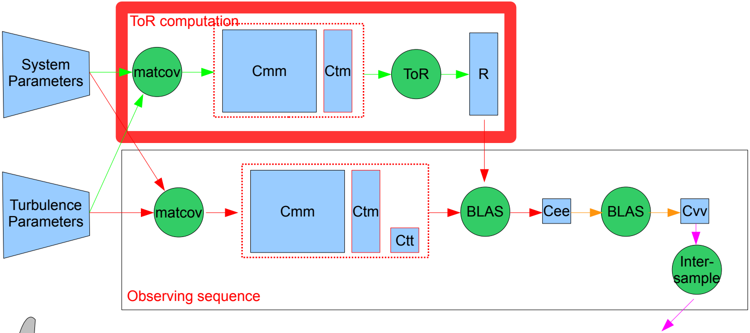
$$R = C_{tm}.C_{mm}^{-1}$$

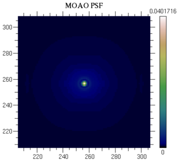
$$C_{ee} = C_{tt} - C_{tm}R^{t} - RC_{tm}^{t} + RC_{mm}R^{t}$$

$$C_{vv} = D^{\dagger}C_{ee}D^{\dagger t}$$



Global Workflow Chart



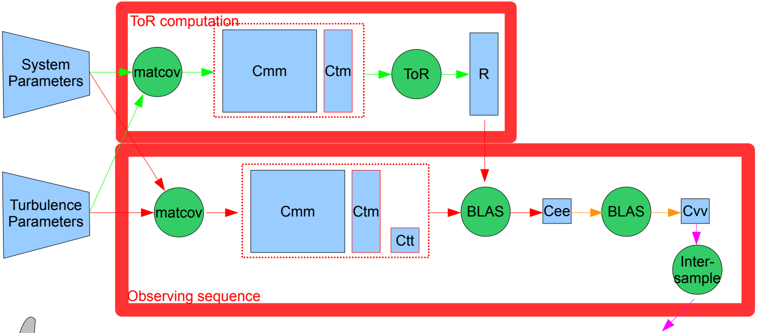


$$R = C_{tm}.C_{mm}^{-1}$$

$$C_{ee} = C_{tt} - C_{tm}R^t - RC_{tm}^t + RC_{mm}R^t$$

$$C_{vv} = D^{\dagger}C_{ee}D^{\dagger t}$$

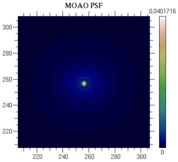
Global Workflow Chart



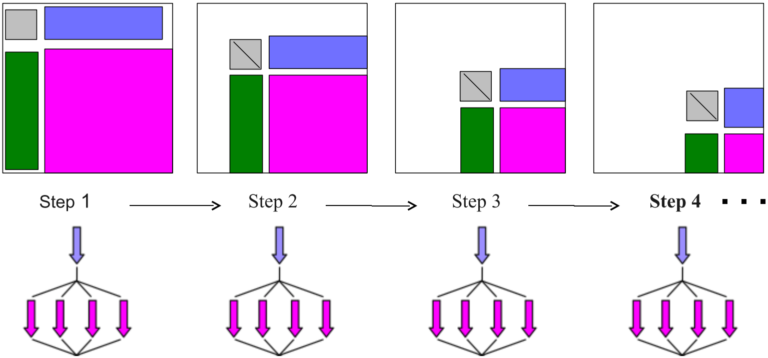
$$R = C_{tm}.C_{mm}^{-1}$$

$$C_{ee} = C_{tt} - C_{tm}R^t - RC_{tm}^t + RC_{mm}R^t$$

$$C_{vv} = D^{\dagger}C_{ee}D^{\dagger t}$$

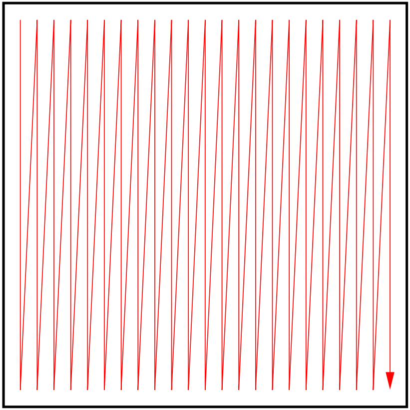


Blocked Algorithms: Fork-Join Paradigm

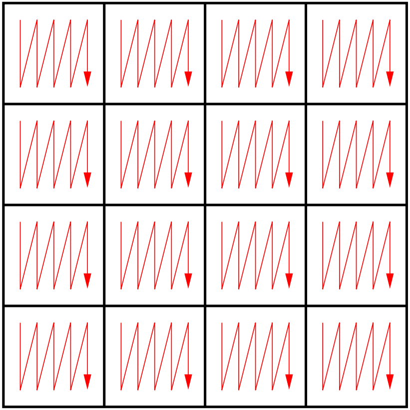


Tile Algorithms: Asynchronous Many-Tasks Execution

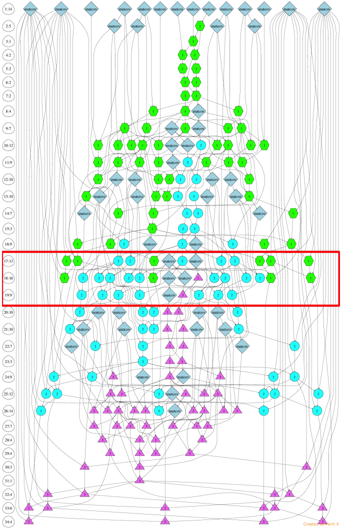
LAPACK: column-major format



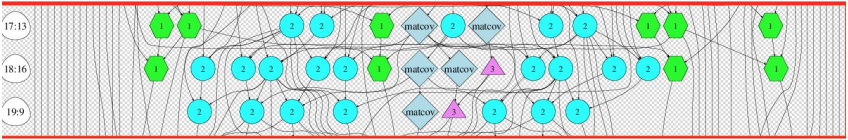
PLASMA/CHAMELEON: tile format



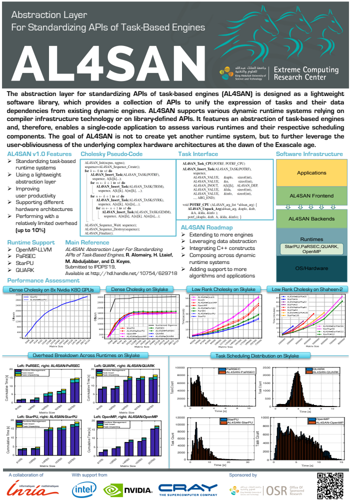
Directed Acyclic Graph for MOAO



Zooming in...

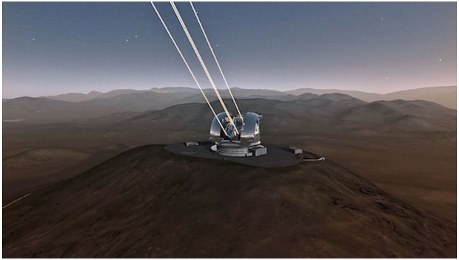


AL4SAN: Abstraction Layer For Standardizing APIs of Task-Based Engines – https://github.com/ecrc/al4san

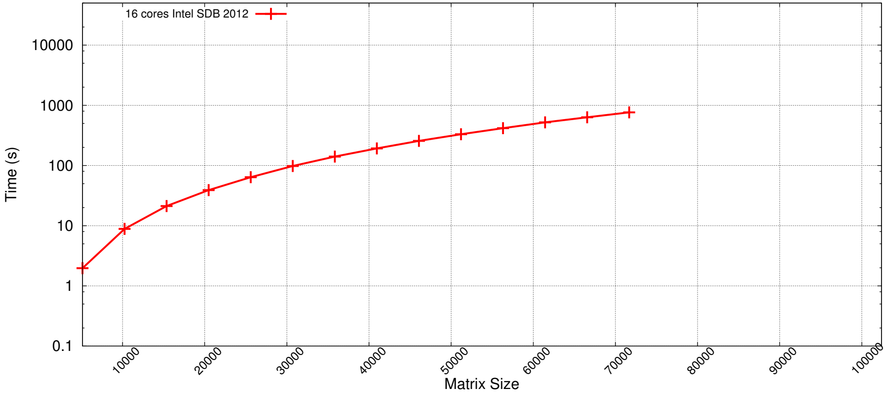


E-ELT Application Apparatus

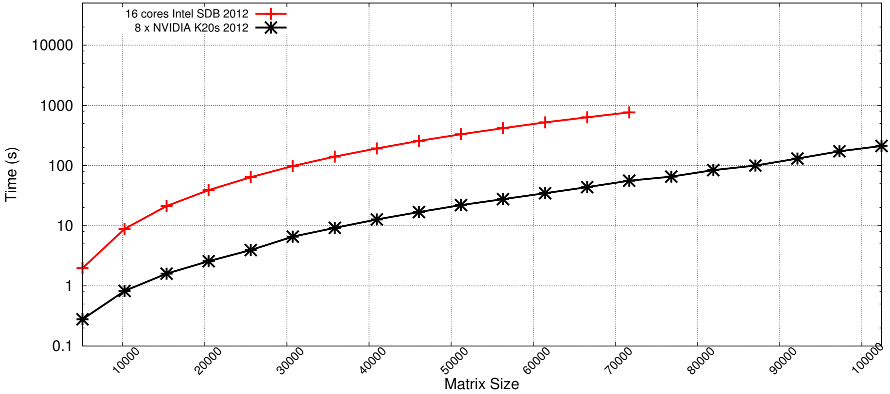
- Diameter telescope: 39m
- Number of measurements: up to 100K
- Number of measurements of the true sensor: 10240
- Number of actuators: 5120
- Performance of the ToR computation



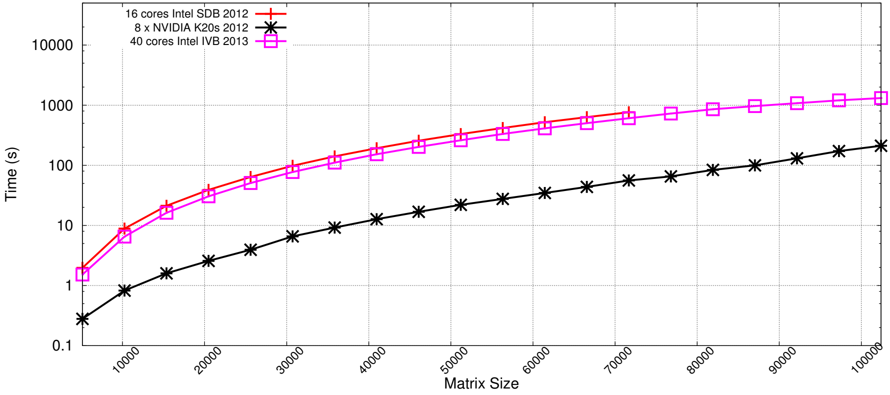
Performance Evolution of the ToR Computations



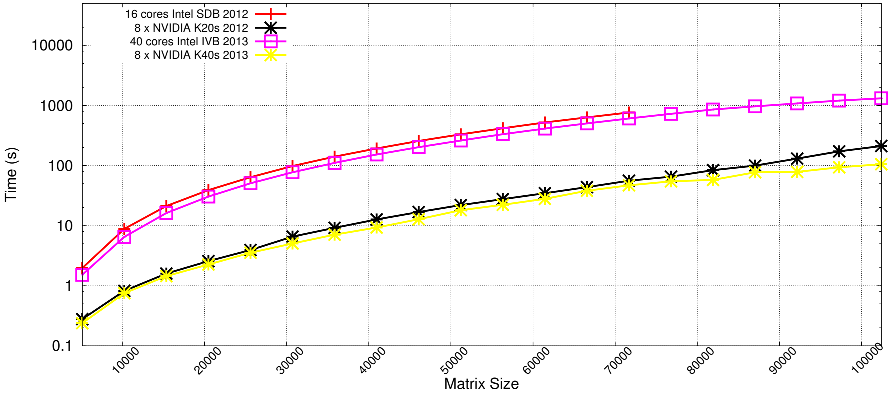
Performance Evolution of the ToR Computations



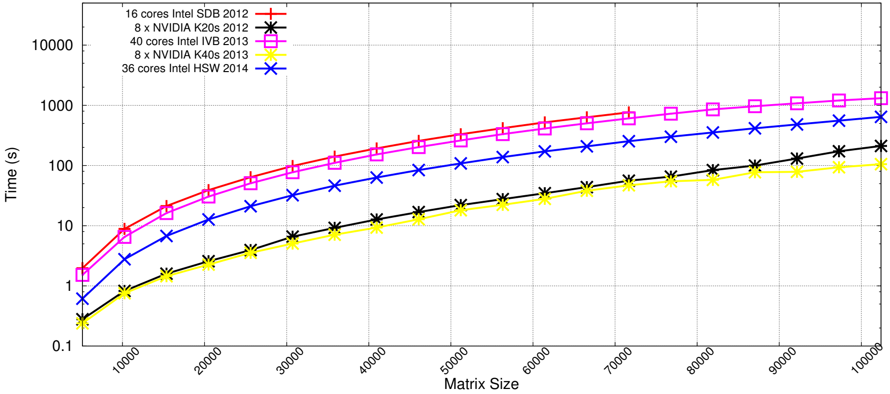
Performance Evolution of the ToR Computations



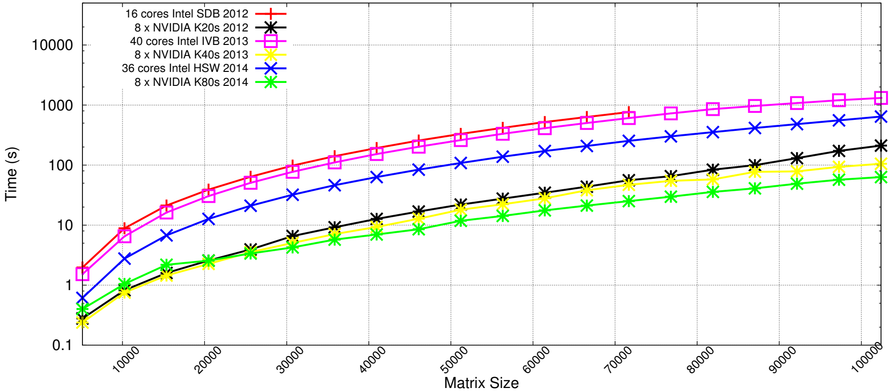
Performance Evolution of the ToR Computations



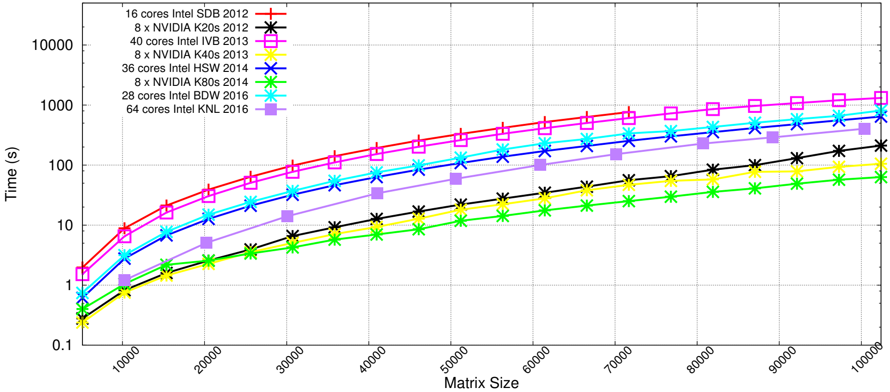
Performance Evolution of the ToR Computations



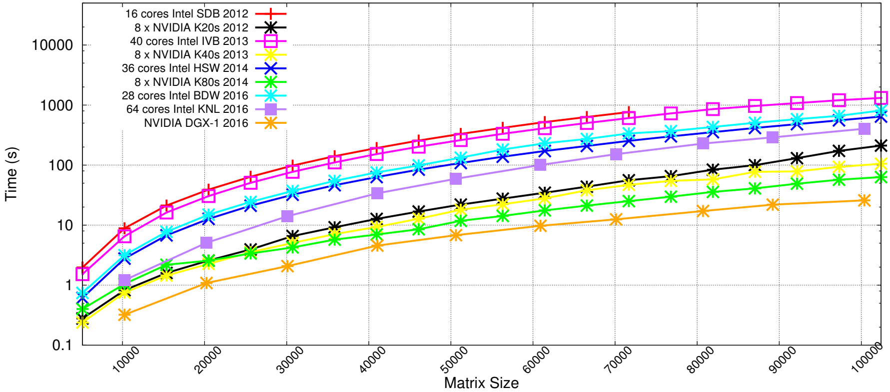
Performance Evolution of the ToR Computations



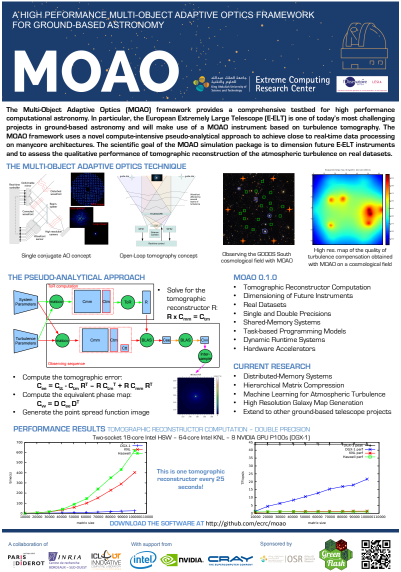
Performance Evolution of the ToR Computations



Performance Evolution of the ToR Computations

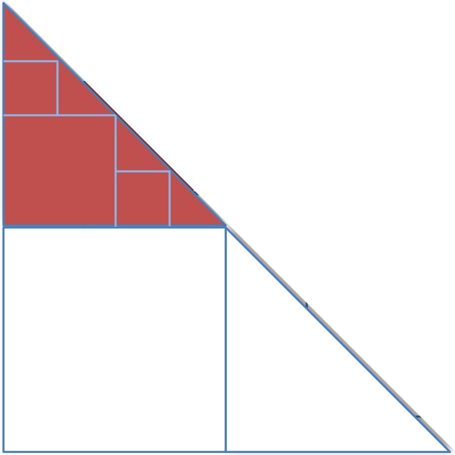


MOAO Software Release – https://github.com/ecrc/moao

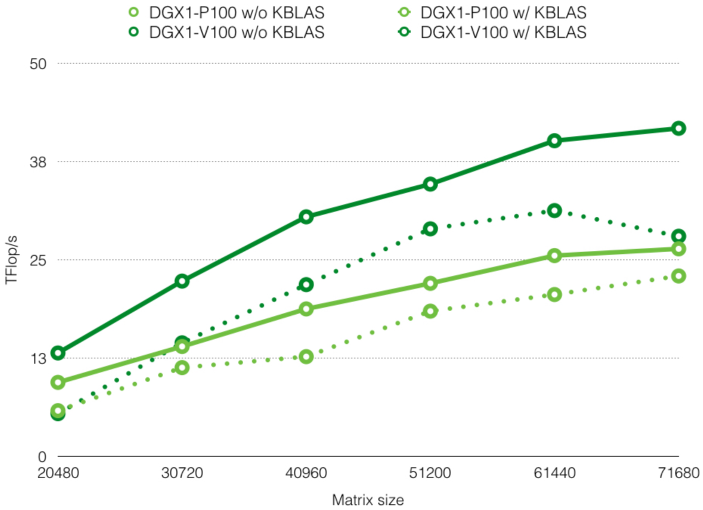


New DPOTRF @ KBLAS – https://github.com/ecrc/kblas

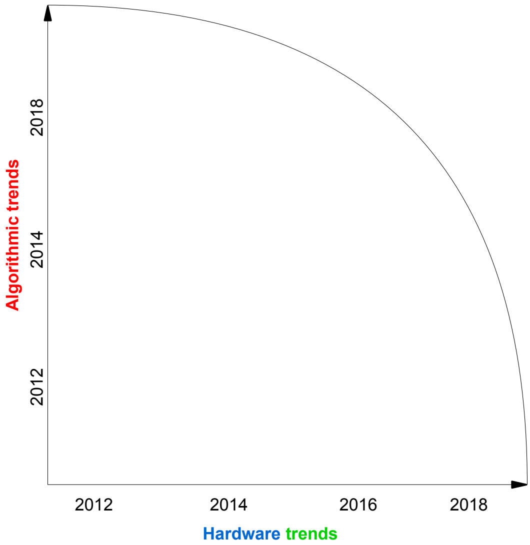
GPU-resident using recursive formulation



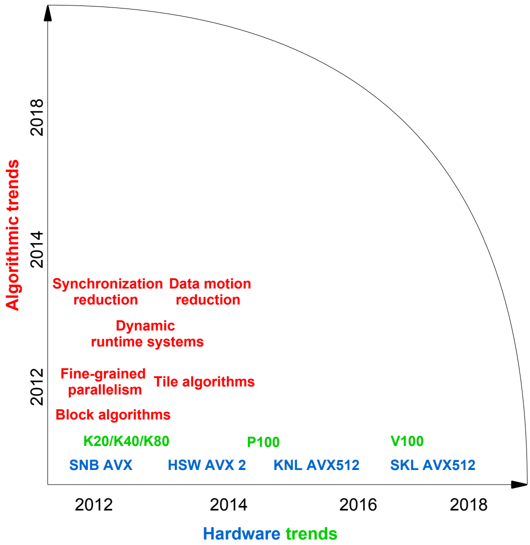
ToR Performance using GPU-resident DPOTRF@KBLAS



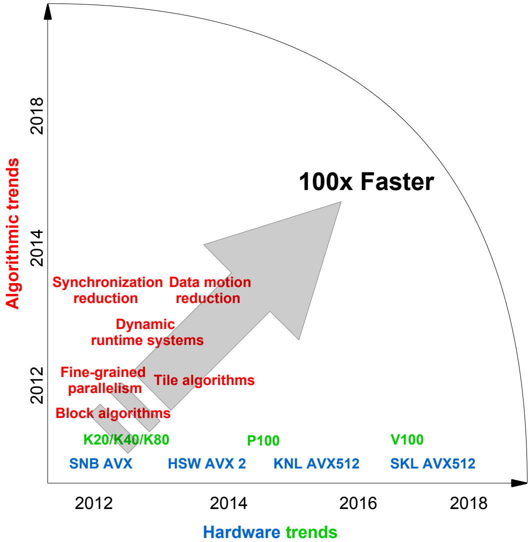
ToR Performance Evolution



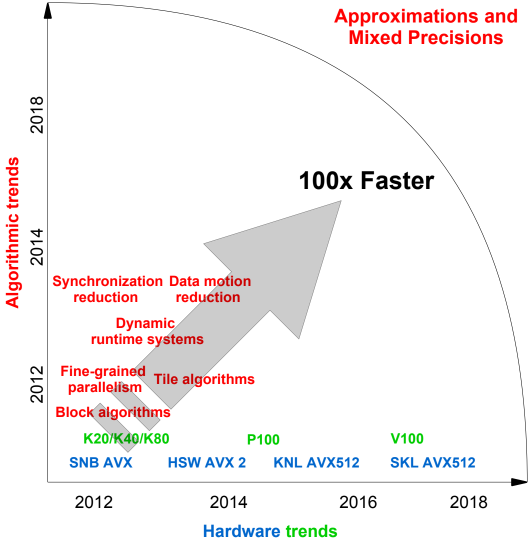
ToR Performance Evolution



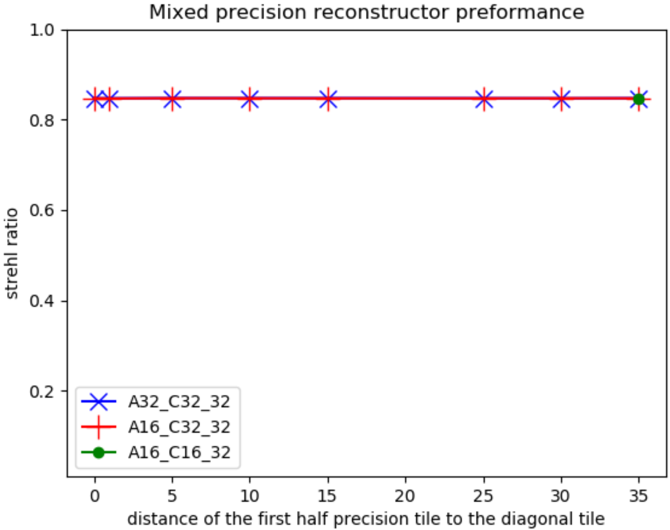
ToR Performance Evolution



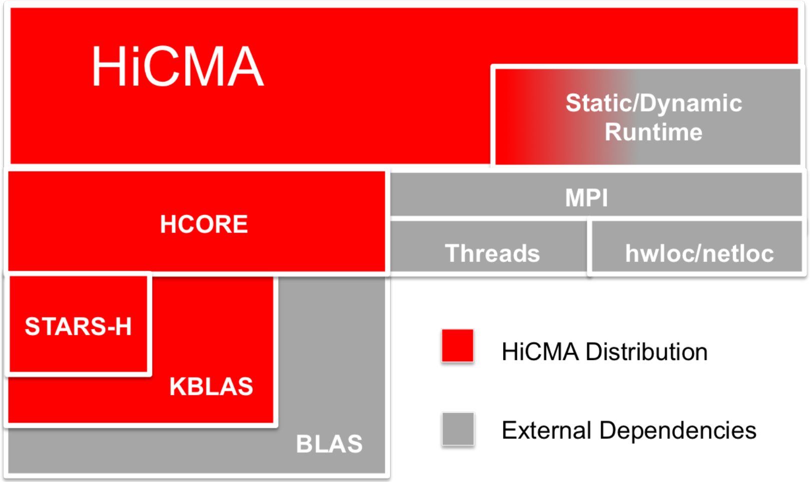
ToR Performance Evolution



Accuracy of Mixed Precisions ToR: Brute Force

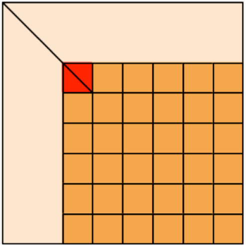


Hierarchical Computations on Manycore Architectures



Available at http://github.com/ecrc/hicma

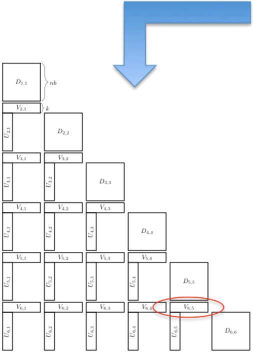
Dense Linear Algebra Renaissance



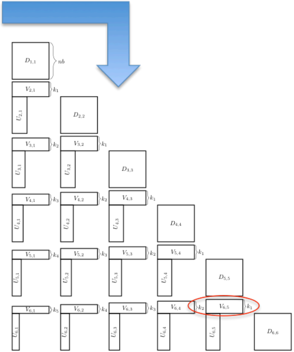
dense tiles
Cholesky: $O(n^3)$

↓

tile low rank
Cholesky: $O(kn^2)$

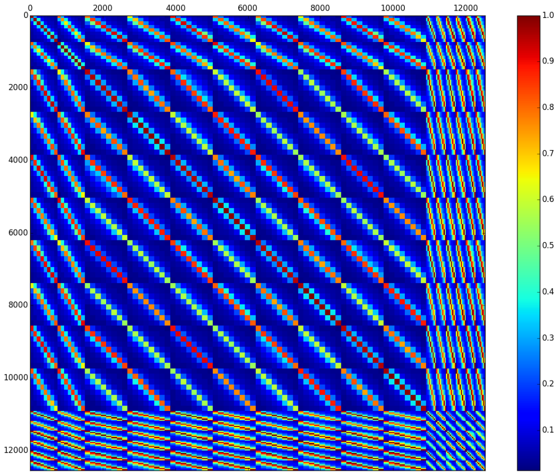


Fixed ranks
Preconditioners
Performance oriented

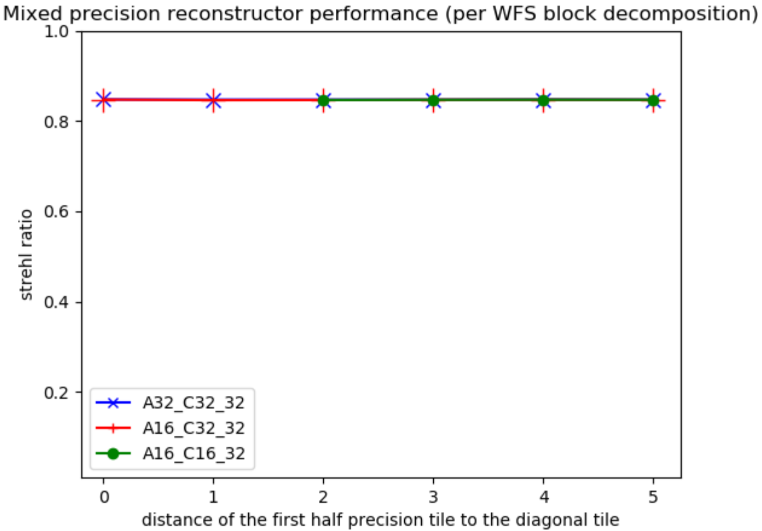


Fixed accuracy
Variable ranks
Dense/Sparse Direct Solvers

X-Ray (SVD) of the Covariance Matrix



Accuracy of Mixed Precisions ToR: Smarter Heuristic



Estimated Performance of Mixed Precisions ToR

Work in progress:

- On NVIDIA V100 GPUs:
  - dgemm achieves about 6.4 Tflop/s on single V100
  - sgemm achieves about 14 Tflop/s on single V100
  - hgemm achieves about 27 Tflop/s on single V100
  - hgemm (w/ tensor cores) reaches about 85 Tflop/s on single V100
- Single precision ToR performance: **42** Tflop/s on 8 V100s
  - That is a ToR at ELT scale computed every 25 seconds
- Speedup factor of **6** between sgemm and hgemm tensor cores
- **6 x 42 TeraOps/s = 252 PetaOps/s on 8 V100s**
  - That is a ToR at ELT scale computed every 5 seconds
- Probably one of the first real applications amenable for tensor cores usage outside of the traditional AI workloads

Outline

And here comes *again* the linear algebra...

- Compute the pseudo inverse A^+

$$AA^+A = A\ , A \in \mathbb{R}^{m \times n}(m \geq n)$$

- The numerical challenge of the pseudo inverse are twofold:
  - Numerical: dealing with rectangular matrix which may engender numerical instabilities

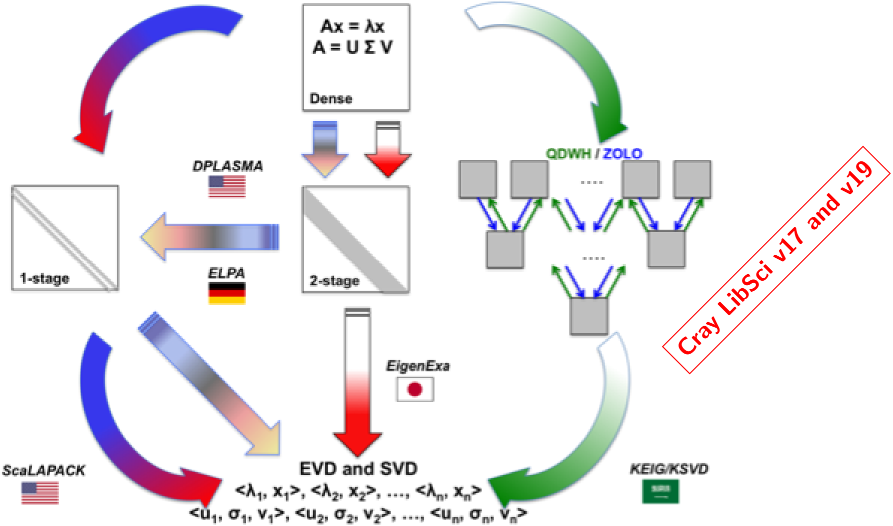
$$A^\top A \ = \ V \Lambda V^\top$$

  - Computational: high algorithmic complexity, it should still be able to keep up with the overall throughput of the AO framework
- Using SVD: $A \ = U \Sigma V^\top$ then:

$$A^+ \ = \ V \Sigma^{-1} U^\top$$

- Only most significant singular values with their associated singular vectors are required ($\approx$**10%**)

The Big Picture



What is The Polar Decomposition?

- The polar decomposition:

$$\mathbf{A} \ = \ \mathbf{U_p H} \ , \ A \in \mathbb{R}^{m \times n} (m \geq n),$$

where U_p is an orthogonal matrix and $H = \sqrt{A^\top A}$ is a symmetric positive semidefinite matrix

- The polar decomposition is a critical numerical algorithm for various applications, including aerospace computations, chemistry, factor analysis

QDWH Polar Decomposition Algorithm

- The QR-Dynamically Weighted Halley iterations:

$$X_0 = A/\alpha, \begin{bmatrix} \sqrt{c_k} X_k \\ I \end{bmatrix} = \begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix} R,$$
$$X_{k+1} = \frac{b_k}{c_k} X_k + \frac{1}{\sqrt{c_k}} \left(a_k - \frac{b_k}{c_k} \right) Q_1 Q_2^\top, \ k \geq 0$$

- The iterative procedure converges:

$$A = U_p H,$$

where, $U_p U_p^\top = I_n$, H is symmetric positive semidefinite
- Backward stable algorithm for computing the polar decomposition
- Based on conventional computational kernels, i.e., Cholesky/QR factorizations (≤ 6 iterations for double precision) and GEMM

Numerical Algorithm

Algorithm 1 Pseudo-Inverse using the QDWH-Based Partial SVD.

Compute the polar decomposition $A = U_p H$ using QDWH
Calculate $[Q\ R] = QR(U_p + Id)$
Find the index $ind = min(find(abs(diag(R)) < threshold))$
Extract $\tilde{Q} = Q(:, ind : end)$
Reduce the original matrix problem $\tilde{A} = A \times \tilde{Q}$
Compute the SVD of the reduced matrix problem $\tilde{A} = U\Sigma\tilde{V}^T$
Compute the right singular vectors $V = \tilde{Q}^T \times \tilde{V}$
Calculate the pseudo-inverse $A^+ = V\Sigma^{-1}U^T$

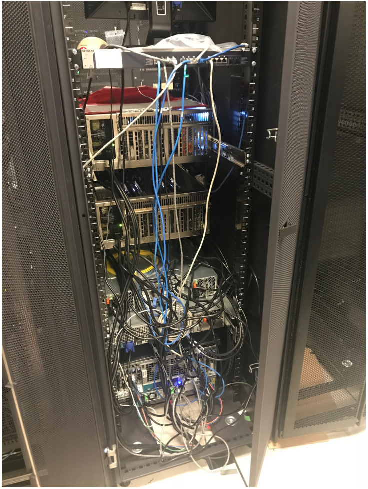
Algorithmic Complexity

	Standard SVD	QDWH-based **Full** SVD	QDWH-based **Partial** SVD
Algorithmic complexity	$22Nn^3$	$43Nn^3$	QDWH: $(4+1/3)Nn^3$ x $\#it_{Chol}$ QR and GEMM: $4/3Nn^3 + 2sNn^2 + 2Nns^2$ SVD: $22s^3$

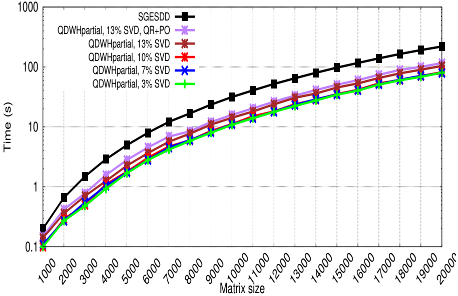
Where, Nn is the matrix size, and s is the number of the selected singular values/vectors ($s << Nn$)

Here is the beast from outside...

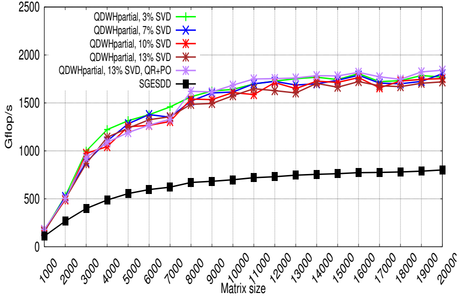
Here is the beast from inside...



Synthetic Ill-Conditioned matrices, K40



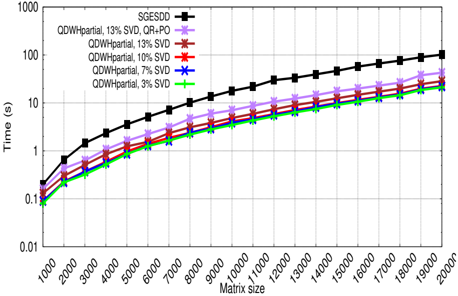
(a) In Seconds.



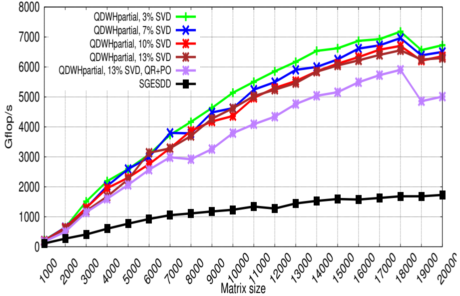
(b) In Gflops/s.

Up to **3X** speedup, 1.8Tflop/s, 45% of the theoretical peak performance

Synthetic Ill-Conditioned matrices, P100



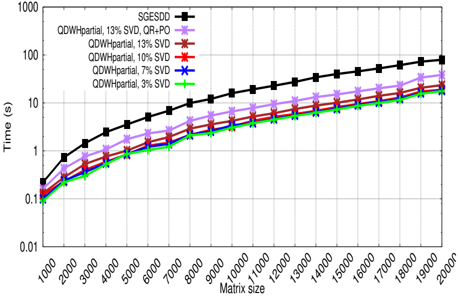
(c) In Seconds.



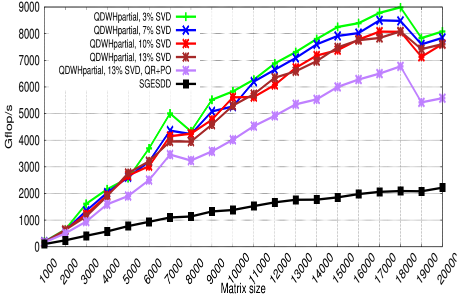
(d) In Gflops/s.

Up to **4X** speedup, 7Tflop/s, 75% of the theoretical peak performance

Synthetic Ill-Conditioned matrices, V100



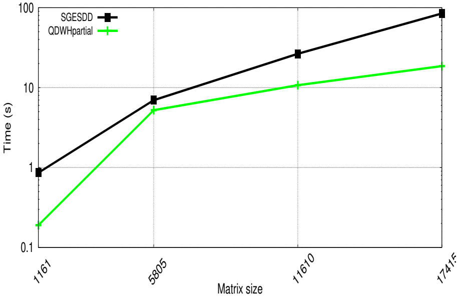
(e) In Seconds.



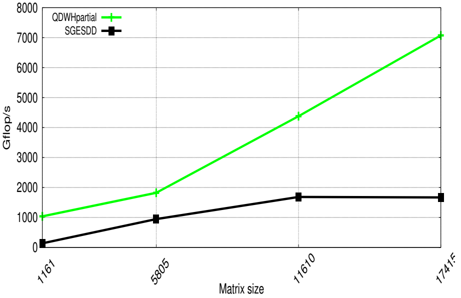
(f) In Gflops/s.

Up to **5X** speedup, 9Tflop/s, 65% of the theoretical peak performance

Real Observational Datasets from Subaru, V100



(g) In Seconds.



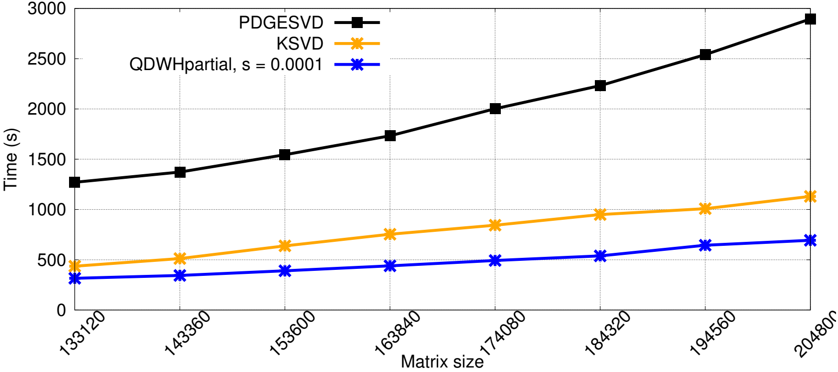
(h) In Gflops/s.

Up to **4X** speedup

Porting to our Cray XC40 *Shaheen 2.0* Supercomputer

Compute Node x 6174 with dual socket 16-core Intel Haswell
$200,000$ cores total

Scaling up using 30,000 cores

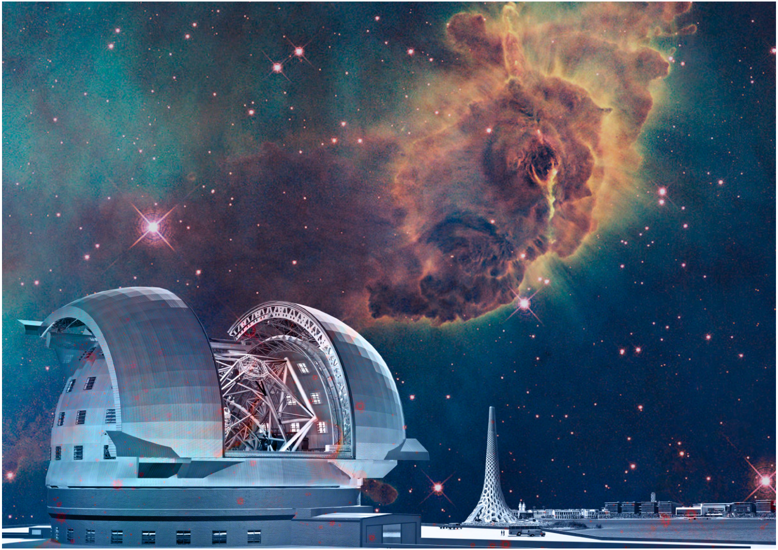


Outline

Summary

- The astronomical game is about outsmarting the atmospheric turbulence
- AO requires massively parallel hardware systems
- Efficient task-based programming model
- Dense, tightly-connected GPU-based compute node (i.e., *DGX-1*) for real-time processing
- Leveraging the data sparsity of the covariance matrix operator
- Exploiting the Toeplitz matrix structure (from cubic to quadratic to log linear algorithmic complexity)
- AO community effort for software standardization (w/ D. Gratadour and O. Guyon)
- More seriously, when can we have BLAS running on Quantum Computers?

Bringing Astronomy Back Home ;-)



Courtesy from CEMSE Communications, KAUST

Questions?

Thank you ☺