STARS-H: a High-Performance $\mathcal{H}$-Matrix Market Library for Large-Scale Systems

A. Mikhalev H. Ltaief D. Keyes

Extreme Computing Research Center
Computer, Electrical and Mathematical Sciences and Engineering Division
King Abdullah University of Science and Technology

April 23, 2018

What is STARS-H?

STARS-H stands for

Software for
Testing
Accuracy
Reliability and
Scalability of
-
Hierarchical computations.

Key reason

Every software for hierarchical matrices lives its own life: different programming system languages, different input format, different ways to store data and different target applications and hardware. This makes it difficult to compare results, reuse them in other software and present a "fair" list of best practices. **Salvation is standardization**.

Why Hierarchical Matrices?

Hierarchical or $\mathcal{H}$-matrices are based on a simple geometric idea: interactions between groups of far-away objects can be described by a small amount of parameters. Example: celestial bodies. Such matrices are algebraic analogs of the famous Fast Multipole Method (FMM). Just like sparse matrices, they appear in many problems during discretization process.

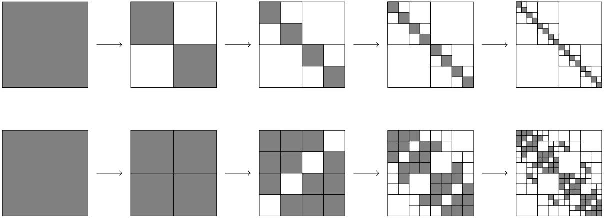


Figure: Examples of mosaic partitioning into $\mathcal{H}$-matrix.

Why Hierarchical Matrices? N-body problems!

N-body problem (e.g. astrophysics, electrostatics): given "particles" $\{x_i\}$ with "charges" $\{q_i\}$, find "potentials" $\{p_i\}$, such that

$$p_i = \sum_j f(x_i, x_j) q_j,$$

where $f(x_i, x_j)$ is a "potential" at x_i, created by a "particle" in x_j with a unit charge. If f is asymptotically smooth, then:

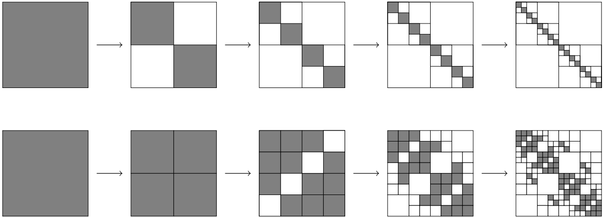


Figure: Examples of mosaic partitioning into $\mathcal{H}$-matrix.

Why Hierarchical Matrices? Boundary integrals!

1. Assume boundary problem:

$$\begin{cases} Lu(x) = 0, & x \in \Omega, \\ Du(x) = \phi(x), & x \in \partial\Omega. \end{cases}$$

2. Let $G(x, y)$ be Green's function of operator L in a free space, then:

$$u(x) = \int_{\partial\Omega} G(x, y) w(y) dy,$$

with some **weight** function $w(y)$.

3. Kernel $K(x, y)$, defined as:

$$K(x, y) = DG(x, y),$$

is, usually, asymptotically smooth.

4. Get **boundary integral equation**

$$\int_{\partial\Omega} K(x, y) w(y) dy = \phi(x), x \in \partial\Omega,$$

which is equivalent to initial problem in item 1.

Discretizations of Boundary Integral Equation with asymptotically smooth kernels lead to hierarchical matrices (e.g. single/double layer potential).

Why Hierarchical Matrices? Fractional derivatives!

Fractional derivatives appear in different non-stationary problems, where long-term system evolution follows integer PDE, but short-term evolution does not!

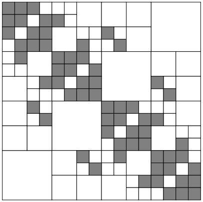
One of many definitions of a fractional derivative

Fractional derivative of power α with n as nearest integer greater than α is following (RiemannLiouville fractional derivative):

$$ {}_aD_t^{\alpha} f(t) = \frac{d^n}{dt^n} {}_aD_t^{-(n-\alpha)} f(t) = \frac{1}{\Gamma(n-\alpha)} \frac{d^n}{dt^n} \int_a^t \frac{f(\tau)}{(t-\tau)^{\alpha+1-n}} d\tau. \quad (1) $$

Fractional derivatives lead to dense matrices even for uniform discretizations.

Why Hierarchical Matrix Market? Application-wise.



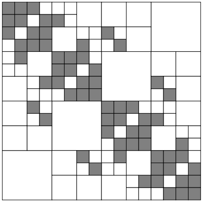
There are many formats of hierarchical matrices: e.g., HODLR, HSS, $\mathcal{H}$ and $\mathcal{H}^2$. The variety of software libraries is even greater: hlib, hlibpro, h2tools, strumpack, hicma, kifmm, pvfmm, ...
Every library and every format have its own advantages and disadvantages.

Assume you have your own application. An opportunity and challenge is to select such software library, that fits given hardware best.

Goal of STARS-H

Provide comparison of a set of software libraries for a given application on a given hardware.

Why Hierarchical Matrix Market? Computation-wise.



There are many formats of hierarchical matrices: e.g., HODLR, HSS, $\mathcal{H}$ and $\mathcal{H}^2$. The variety of software libraries is even greater: hlib, hlibpro, h2tools, strumpack, hicma, kifmm, pvfmm, ...
Every library and every format have its own advantages and disadvantages.

Assume you have your own software library. An opportunity and challenge is to compare against other software, since there is no standard set of tests.

Goal of STARS-H

Provide a set of applications to standardize comparison of hierarchical libraries.

$\mathcal{H}$-Matrix Market in an HPC Framework

Serving as an $\mathcal{H}$-matrix market means not only problem setting, but also $\mathcal{H}$-matrix generation, depending on

1. **clustering discrete elements**
2. partitioning (format, admissibility)
3. accuracy threshold

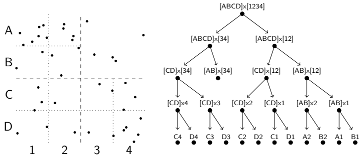


Figure: Clusters of "particles".

Not every hierarchical library is about approximation from dense. A lot of research is about operations with hierarchical matrices.

Goal of STARS-H

Enable standardized input for computations of different $\mathcal{H}$-matrix libraries on every supported hardware and do it in an HPC manner.

Distribution of Ranks for TLR Format: Synthetic Kernels

8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	9	8	8
8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8
8	8	9	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8

(a) $\frac{\sin(kr)}{r}$, $k = 1$

26	27	28	28	27	27	28	28	27	27
27	27	28	28	28	27	28	28	28	27
28	28	28	28	28	27	29	29	28	28
28	28	28	27	27	27	28	29	28	28
27	28	28	27	25	26	27	27	27	27
27	27	27	27	26	25	27	28	28	27
28	28	29	28	27	27	27	28	28	28
28	28	29	29	27	28	28	28	28	28
27	28	28	28	27	28	28	28	27	27
27	27	28	28	27	27	28	28	27	26

(b) $\frac{\sin(kr)}{r}$, $k = 10$

107	111	108	87	78	91	75	79	70	61
111	113	114	91	88	106	96	98	82	70
108	114	120	112	108	98	87	104	98	79
87	91	112	112	109	74	66	87	96	75
78	88	108	109	106	82	74	98	106	91
91	106	98	74	82	106	109	108	88	78
75	96	87	66	74	109	112	112	91	87
79	98	104	87	98	108	112	120	114	108
70	82	98	96	106	88	91	114	113	111
61	70	79	75	91	78	87	108	111	107

(c) $\frac{\sin(kr)}{r}$, $k = 50$

250	250	250	114	104	250	103	111	97	82
250	250	250	115	115	250	250	250	115	97
250	250	250	250	250	250	111	250	250	111
114	115	250	250	250	90	79	111	250	103
104	115	250	250	250	101	90	250	250	250
250	250	250	90	101	250	250	250	115	104
103	250	111	79	90	250	250	250	115	114
111	250	250	111	250	250	250	250	250	250
97	115	250	250	250	115	115	250	250	250
82	97	111	103	250	104	114	250	250	250

(d) $\frac{\sin(kr)}{r}$, $k = 100$

250	115	69	35	27	33	23	24	21	18
115	250	108	43	36	79	36	33	26	21
69	108	250	104	69	69	37	48	33	24
35	43	104	250	111	29	23	37	36	23
27	36	69	111	250	38	29	69	79	33
33	79	69	29	38	250	111	69	36	27
23	36	37	23	29	111	250	104	43	35
24	33	48	37	69	69	104	250	108	69
21	26	33	36	79	36	43	108	250	115
18	21	24	23	33	27	35	69	115	250

(e) $\frac{\cos(kr)}{r}$, $k = 1$

250	120	76	43	36	40	31	33	30	26
120	250	115	51	44	86	44	42	34	30
76	115	250	111	76	77	46	56	42	33
43	51	111	250	116	37	33	46	44	31
36	44	76	116	250	46	37	77	86	40
40	86	77	37	46	250	116	76	44	36
31	44	46	33	37	116	250	111	51	43
33	42	56	46	77	76	111	250	115	76
30	34	42	44	86	44	51	115	250	120
26	30	33	31	40	36	43	76	120	250

(f) $\frac{\cos(kr)}{r}$, $k = 10$

250	15	6	5	4	4	4	4	4	3
15	250	15	6	5	4	4	4	4	4
6	15	250	15	6	5	4	4	4	4
5	6	15	250	15	6	5	4	4	4
4	5	6	15	250	15	6	5	4	4
4	4	5	6	15	250	15	6	5	4
4	4	4	5	6	15	250	15	6	5
4	4	4	4	5	6	15	250	15	6
4	4	4	4	4	5	6	15	250	15
3	4	4	4	4	4	5	6	15	250

(g) Cauchy

250	10	10	10	10	10	10	10	10	10
10	250	10	10	10	10	10	10	10	10
10	10	250	10	10	10	10	10	10	10
10	10	10	250	10	10	10	10	10	10
10	10	10	10	250	10	10	10	10	10
10	10	10	10	10	250	10	10	10	10
10	10	10	10	10	10	250	10	10	10
10	10	10	10	10	10	10	250	10	10
10	10	10	10	10	10	10	10	250	10
10	10	10	10	10	10	10	10	10	250

(h) Random TLR

Distribution of Ranks for TLR Format: Real kernels

250	115	68	34	27	32	22	23	20	16
115	250	107	43	36	79	35	33	25	20
68	107	250	104	69	69	36	48	33	23
34	43	104	250	111	28	23	36	35	22
27	36	69	111	250	37	28	69	79	32
32	79	69	28	37	250	111	69	36	27
22	35	36	23	28	111	250	104	43	34
23	33	48	36	69	69	104	250	107	68
20	25	33	35	79	36	43	107	250	115
16	20	23	22	32	27	34	68	115	250

(a) $\frac{1}{r}$, $\tau = 10^{-9}$

250	104	64	33	27	33	25	25	23	19
104	250	97	39	34	72	36	35	27	23
64	97	250	94	64	62	34	45	35	25
33	39	94	250	100	27	21	34	36	25
27	34	64	100	250	34	27	62	72	33
33	72	62	27	34	250	100	64	34	27
25	36	34	21	27	100	250	94	39	33
25	35	45	34	62	64	94	250	97	64
23	27	35	36	72	34	39	97	250	104
19	23	25	25	33	27	33	64	104	250

(b) Exp., $\tau = 10^{-9}$

250	77	69	48	44	53	42	38	34	33
77	250	77	47	48	68	56	57	46	34
69	77	250	75	68	55	43	55	57	38
48	47	75	250	75	33	27	43	56	42
44	48	68	75	250	40	33	55	68	53
53	68	55	33	40	250	75	68	48	44
42	56	43	27	33	75	250	75	47	48
38	57	55	43	55	68	75	250	77	69
34	46	57	56	68	48	47	77	250	77
33	34	38	42	53	44	48	69	77	250

(c) Sqr.exp., $\tau = 10^{-9}$

250	89	65	40	35	43	35	33	30	28
89	250	84	40	40	69	47	46	37	30
65	84	250	82	64	54	37	47	46	33
40	40	82	250	87	27	22	37	47	35
35	40	64	87	250	31	27	54	69	43
43	69	54	27	31	250	87	64	40	35
35	47	37	22	27	87	250	82	40	40
33	46	47	37	54	64	82	250	84	65
30	37	46	47	69	40	40	84	250	89
28	30	33	35	43	35	40	65	89	250

(d) Matérn, $\tau = 10^{-9}$

250	52	24	9	7	8	6	6	5	5
52	250	50	13	10	31	9	9	7	5
24	50	250	49	24	28	10	14	9	6
9	13	49	250	50	8	6	10	9	6
7	10	24	50	250	11	8	28	31	8
8	31	28	8	11	250	50	24	10	7
6	9	10	6	8	50	250	49	13	9
6	9	14	10	28	24	49	250	50	24
5	7	9	9	31	10	13	50	250	52
5	5	6	6	8	7	9	24	52	250

(e) $\frac{1}{r}$, $\tau = 10^{-4}$

250	44	21	9	8	10	8	8	7	6
44	250	42	12	9	26	10	10	8	7
21	42	250	41	21	23	9	13	10	8
9	12	41	250	41	7	6	9	10	8
8	9	21	41	250	10	7	23	26	10
10	26	23	7	10	250	41	21	9	8
8	10	9	6	7	41	250	41	12	9
8	10	13	9	23	21	41	250	42	21
7	8	10	10	26	9	12	42	250	44
6	7	8	8	10	8	9	21	44	250

(f) Exp., $\tau = 10^{-4}$

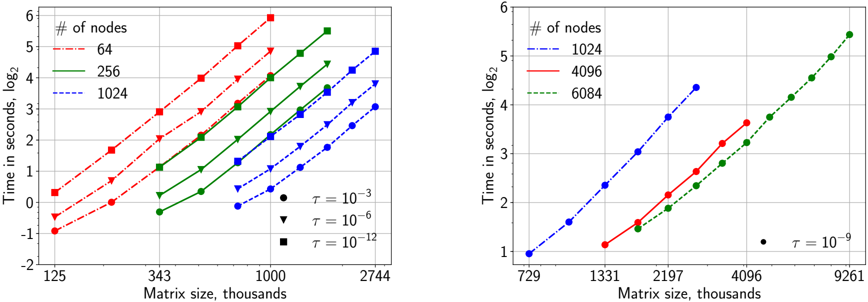
31	28	23	15	13	15	12	10	9	9
28	33	27	14	15	22	17	16	13	9
23	27	34	26	22	17	13	16	16	10
15	14	26	33	27	9	8	13	17	12
13	15	22	27	30	12	9	17	22	15
15	22	17	9	12	30	27	22	15	13
12	17	13	8	9	27	33	26	14	15
10	16	16	13	17	22	26	34	27	23
9	13	16	17	22	15	14	27	33	28
9	9	10	12	15	13	15	23	28	31

(g) Sqr.exp., $\tau = 10^{-4}$

39	29	21	13	11	14	11	9	9	9
29	41	28	13	13	21	15	15	13	9
21	28	43	27	21	18	11	14	15	9
13	13	27	40	28	8	7	11	15	11
11	13	21	28	39	10	8	18	21	14
14	21	18	8	10	39	28	21	13	11
11	15	11	7	8	28	40	27	13	13
9	15	14	11	18	21	27	43	28	21
9	13	15	15	21	13	13	28	41	29
9	9	9	11	14	11	13	21	29	39

(h) Matérn, $\tau = 10^{-4}$

TLR on a Distributed-Memory System



Dependence of time-to-approximate on matrix size. Up to **9 millions DOFs**, up to **6084** nodes ($\approx$ 200000 cores).

Thank you for attention!

STARS-H: a High-Performance $\mathcal{H}$-Matrix Market Library for Large-Scale Systems

A. Mikhalev, H. Ltaief, D. Keyes

KAUST

Extreme Computing Research Center   Computer, Electrical and Mathematical Sciences and Engineering Division   King Abdullah University of Science and Technology

STARS-H is

a high performance parallel open-source **S**oftware for **T**esting **A**ccuracy, **R**eliability and **S**calability of **H**ierarchical computations. STARS-H provides a hierarchical matrix market in order to benchmark performance of various libraries for hierarchical matrix compressions and computations. STARS-H introduces a standard for assessing accuracy and performance of hierarchical matrix libraries on a given hardware architecture environment. STARS-H package is available online at https://github.com/ecrc/stars-h.

Why $\mathcal{H}$-matrix market?

- $\mathcal{H}$-matrices appear in N-body problems and integral equations as a result of discretization process (e.g. astrophysics, aerodynamics, drug design).
- There are no standard tests to assess performance of a software for exploiting hierarchically low-rank structure.
- To meet our own needs and those of the community, we offer a parameterizable library.

Goals of STARS-H

- Provide a set of applications to standardize comparison of hierarchical libraries.
- Enable standardized input for computations of different H-matrix libraries with optimized performance on a range of hardware.

STARS-H 0.1.0

- Data formats: Tile Low-Rank (TLR) with more to come.
- Operations: approximation, matvec product, CG solve.
- Synthetic applications: random TLR, Cauchy.
- Real applications: electrostatics, electrodynamics, spatial statistics.
- Programming models: OpenMP, MPI and task-based (StarPU).
- Approximation techniques: SVD, Rank-Revealing QR, Randomized SVD.

Distribution of ranks ($\tau = 10^{-9}$)

Fig. 1: $\frac{\sin(kr)}{r}$, $k=1$ Fig. 2: $\frac{\sin(kr)}{r}$, $k=10$ Fig. 3: $\frac{\sin(kr)}{r}$, $k=50$

Fig. 4: $\frac{\sin(kr)}{r}$, $k=100$ Fig. 5: $\frac{\cos(kr)}{r}$, $k=1$ Fig. 6: $\frac{\cos(kr)}{r}$, $k=10$

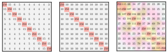


Fig. 7: Cauchy   Fig. 8: Random TLR   Fig. 9: $\frac{1}{r}$

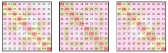


Fig. 10: $e^{-\frac{r}{\ell}}$ Fig. 11: $e^{-\frac{r^2}{2\ell^2}}$ Fig. 12: Matérn

Performance on Intel architectures

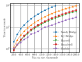


Fig. 13: 2D problem on different two-socket shared-memory Intel x86 architectures

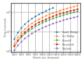


Fig. 14: 3D problem on different two-socket shared-memory Intel x86 architectures

Shaheen-II performance

STARS-H was tested on a Shaheen-II, a CRAY XC40 system. The system has 6,174 dual sockets compute nodes based on 16 core Intel Haswell processors running at 2.3GHz. Each node has 128GB of DDR4 memory running at 2300MHz. Overall the system has a total of 197,568 processor cores and 790TB of aggregate memory.

Single-node Shaheen-II

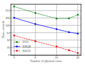


Fig. 15: TLR Approximation of 90k-by-90k matrix of ones

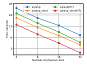


Fig. 16: Generation of 90k-by-90k matrix of 2D spatial statistics problem

Multi-node Shaheen-II

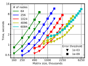


Fig. 17: Approximation of 2D spatial statistics problem (exponential kernel)

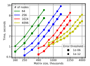


Fig. 18: Approximation of 2D spatial statistics problem (exponential kernel)

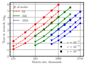


Fig. 19: Approximation of 3D spatial statistics problem (exponential kernel)

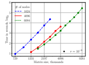


Fig. 20: Approximation of 3D spatial statistics problem (exponential kernel)

Summary

Matrices generated by STARS-H library have various rank distributions over tiles. It enables different inputs for further operations, like Cholesky factorization or matrix-matrix multiplication, making debugging and search for optimal implementation much easier. Although the TLR format itself is the simplest from a theoretical point of view, it scales nearly perfectly due to intrinsic load balance. The TLR could potentially be the best choice for emerging architectures, as a trend is to increase the number of computational cores, while sometimes decreasing power of a single core. As a popularity and number of compute nodes with several GPUs increases, TLR makes even more sense!

Future work

- Extend to other problems in a matrix-free form.
- Support **HODLR**, **HSS**, $\mathcal{H}$ and $\mathcal{H}^2$ data formats.
- Implement other approximation schemes (e.g., ACA).
- Port to GPU accelerators.
- Apply other dynamic runtime systems and programming models (e.g., PARSEC).

References

1. L. Greengard and V. Rokhlin. 1987. A fast algorithm for particle simulations. *Journal of computational physics* 73, 2 (1987), 325-348.
2. E. Tyrtyshnikov. Mosaic-skeleton approximations. *Calcolo* 33, 1 (1996) 47-57.
3. W. Hackbusch. 1999. A Sparse Matrix Arithmetic Based on $\mathcal{H}$-Matrices. Part I: Introduction to $\mathcal{H}$-Matrices. *Computing* 62, 2 (1999), 89-108.
4. W. Hackbusch, B. Khoromskij and S. Sauter. 2000. On H2-Matrices. *Lectures on Applied Mathematics* (2000), 9-29.
5. SMart 2018. Structured Matrix Market. (Accessed January 2018). http://smart.math.purdue.edu.
6. S. Ambikasaran, D. Foreman-Mackey, L. Greengard, D. Hogg and M. O'Neil. 2016. Fast Direct Methods for Gaussian Processes. *IEEE Trans. Pattern Anal. Mach. Intell* 38, 2(2016), 252-265.
7. Y. Sun, B. Li and M. Genton. 2012. Geostatistics for Large Datasets. *In Space-Time Processes and Challenges Related to Environmental Problems*, 55-77.
8. C. Augonnet, S. Thibault, R. Namyst and P. Wacrenier. 2011. StarPU: a unified platform for task scheduling on heterogeneous multicore architectures. *Concurrency and Computation: Practice and Experience*, 23(2), pp.187-198.

LaTeX TikZposter