USPARS direct linear solver

D. Bykov, M. Cherepanov, V. Kostin, A. Semenov, S. Solovyev

Unipro

5 октября 2021 г.

Оглавление

USPARS - прямой решатель разреженных СЛАУ

Поддерживаемые архитектуры и реализации BLAS:

- x86 - mkl, FLAME, OpenBlas
- arm - Apple Accelerate, OpenBlas
- e2k - EML (Elbrus Math Library)

Известные прямые решатели

Прямые решатели разреженных матриц (* - с открытым исходным кодом):

- *TAUCS https://www.tau.ac.il/~stoledo/taucs/ [v 2.2, 2003].
- *UMFPACK https://people.engr.tamu.edu/davis/suitesparse.html [v. 5.10.1, 2021];
- PARDISO https://www.pardiso-project.org/
- MKL PARDISO https://software.intel.com/content/www/us/en/develop/tools/oneapi/components/onemkl.html#gs.6l3p2m [v. 2021.3, 2021].
- *MUMPS http://mumps.enseeiht.fr/ [v. 5.4.0, 2021].
- *SuperLU https://portal.nersc.gov/project/sparse/superlu/

Предпосылки к разработке собственного прямого решателя

- ✓ Наличие заказчиков.
- ✓ Оперативная и надежная поддержка.
- ✓ Использование в собственных продуктах.
- ✓ Компетенции в разработке ПО.

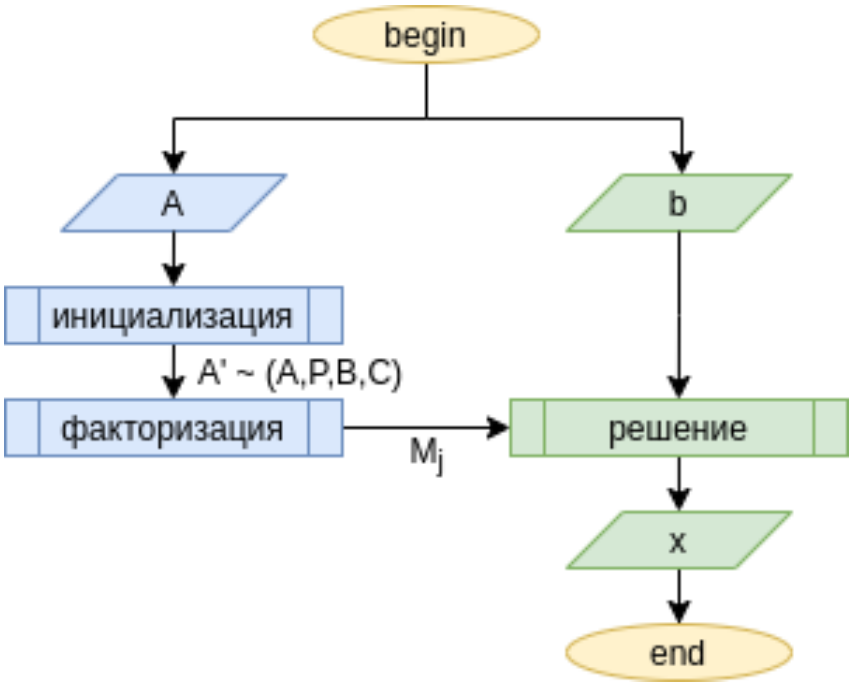
Схема вычислений

$$Ax = b$$

$$A^{'} = BPAP^{T}C$$

P – матрица перестановок, B, C – невырожденные

$$A^{'} = \prod_{j \in J} M_j, M_j - \text{треугольные}$$



Инициализация

$A' = BPAP^T C$, для положительно определенных $B = C = E$

- P перестановка, минимизирующая заполняемость ненулевыми элементами треугольных факторов (fill-in)
  Metis (`http://glaros.dtc.umn.edu/gkhome/metis/metis/download`).
- В случае неопределенных матриц $B = D_l * P_1, C = Dr$, где:
  - P_1: перемещение больших элементов A на главную диагональ (matching);
  - D_l, D_r - масштабирование элементов матрицы A (scaling)

Минимизация fill-in эффекта

Пример:

$$A = \begin{pmatrix} 9 & 1.5 & 6 & 0.75 & 3 \\ 1.5 & 0.5 & & & \\ 6 & & 12 & & \\ 0.75 & & & 0.625 & \\ 3 & & & & 16 \end{pmatrix}.$$

$A = MM^T$, где:

$$M = \begin{pmatrix} 3 & & & & \\ 0.5 & 0.5 & & & \\ 2 & -2 & 2 & & \\ 0.25 & -0.25 & -0.5 & 0.5 & \\ 1 & -1 & -2 & -3 & 1 \end{pmatrix}.$$

$$A^{'} = PAP^{T} = M^{'}M^{'T} = \begin{pmatrix} 16 & & & & 3 \\ & 0.5 & & & 0.75 \\ & & 12 & & 6 \\ & & & 0.625 & 1.5 \\ 3 & 1.5 & 6 & 0.75 & 9 \end{pmatrix}$$

где:

$$M^{'} = \begin{pmatrix} 4 & & & & \\ & \frac{1}{\sqrt{2}} & & & \\ & & 2\sqrt{3} & & \\ & & & \frac{\sqrt{10}}{4} & \\ \frac{3}{4} & \frac{3}{\sqrt{2}} & \sqrt{3} & \frac{3}{\sqrt{10}} & \sqrt{\frac{3}{80}} \end{pmatrix}.$$

Типы факторизаций

- Матрицы общего вида:

$$A = LU$$

- Положительно определенные матрицы:

$$A = \begin{cases} LL^T & \text{симметричная действительная} \\ LL^H & \text{комплексная Эрмитова} \end{cases}$$

- Неопределенные матрицы:

$$A = \begin{cases} LDL^T & \text{симметричная} \\ LDL^H & \text{комплексная Эрмитова} \end{cases}$$

L, U - нижне- и верхнетреугольная матрицы.
D - диагональная.

Стек технологий

c++, python, gtest, git, Teamcity

Тестовые данные

- Sparse Suite Matrix Collection (`https://sparse.tamu.edu`) - порядка 800 матриц.
- Автоматическая генерация тестов (уравнения Лапласа, Гельмгольца, Сильвестра)

Методика

- Тестирование по расписанию (Teamcity CI).
- Базы функциональных регрессий (x86, arm, e2k).
- Сравнение с референсным солвером (MKL Pardiso).
- Автоматизация отслеживания регрессий производительности.

Характеристики тестовых машин

Таблица: Computing system specifications: Haswell (hsw), Elbrus (e2k) and ARM.

	hsw	e2k	ARM (Apple M1)
HW	Server 2 × 12 cores Intel® Xeon® CPU E5-2680 v3 @2.50 GHz	Desktop E8C @1.3GHz	3.2 GHz (4xFirestorm) + 2.064 GHz (4xIcestorm) + undeclared AMX2 matrix coprocessor
GFlop/s	480	115	
Math libs	MKL 2019	EML:0901:20201019:msvs-p164	Open BLAS 0.3.15, Accelerated framework (Xcode 13.0 build v. 13A233)
Compiler	Intel® C Compiler V.19.1.0.166	lcc:1.25.10:Oct–1-2020:e2k-v4-linux	clang v. 11.1.0 (Target arm64-apple-darwin 20.5.0)
Open MP	v. 5.0	v. 3.1	v.5.0

USPARS (x86,e2k) vs MKL PARDISO (x86)

Matrix	Order	NNZ(A)	NNZ(L)	Computational time (s)		
				PARDISO	USPARS X86	USPARS E2K
nd24k	72 000	4 393.816	327 211 249	36.8	21.2	75.1
dc3	116 835	766 395	1 323 772	1.53	2.75	11.9
lung2	109 460	492 563	799 357	0.17	0.47	2.5
pwtk	217 918	5 926 170	51 138 816	1.09	3.76	17.4
torso3	259 156	4 429 041	228 195 289	6.02	6.79	24.5
af_0_k101	503 625	9 027 149	105 478 877	1.93	7.23	32.3
ecology1	1 000 000	2 998 000	39 917 032	3.85	6.64	39.6

8 OMP потоков для MKL PARDISO и USPARS.

- лучший результат

x86: USPARS vs MKL PARDISO vs MUMPS

Matrix	Order	NNZ(A)	NNZ(L)	Computational time (s)		
				PARDISO	USPARS	MUMPS
Ge99H100	112 985	4 282 190	607 149 662	76.7	35.7	39.86
Ga10As10H30	113 081	3 114 357	605 772 098	76.0	32.9	39.93
Ga19As19H42	133 123	4 508 981	801 244 103	125.9	48.1	52.01
dielFilterV3real	1 102 824	45 204 422	543 931 990	16.7	44.8	25.44
thermal2	1 228 045	4 904 179	54 112 868	7.1	11.0	13.56
atmosmodj	1 270 432	8 814 880	1 917 317 133	124.4	64.6	64.78
Serena	1 391 349	32 961 525	2 740 334 167	214.7	134.9	120.88
Geo_1438	1 437 960	32 297 325	2 541 908 275	102.6	125.2	82.7
StocF_1465	1 465 137	11 235 263	1 064 389 725	36.9	39.6	44.43
atmosmodl	1 489 752	10 319 760	1 961 496 567	97.6	59.2	66.34
Hook_1498	1 498 023	31 207 734	1 574 369 499	65.4	70.1	50.46
dielFilterV2real	1 557 456	24 848 204	554 736 755	17.4	33 6	31.3
Flan_1565	1 564 794	59 485 419	1 535 182 163	30.9	66.5	41.97
ss	1 652 680	34 753 577	2 845 017 068	387.3	189.7	166.66
memchip	2 707 524	14 810 202	63 920 499	9.9	24.8	24.96
Freescale1	3 428 755	18 920 347	56 718 368	13.4	29.0	32.39
Queen_4147	4 147 110	166 823 197	14 474 925 849	1544.9	1109.0	n/a

24 OMP потока для MKL PARDISO и USPARS, 24 MPI процесса для MUMPS.

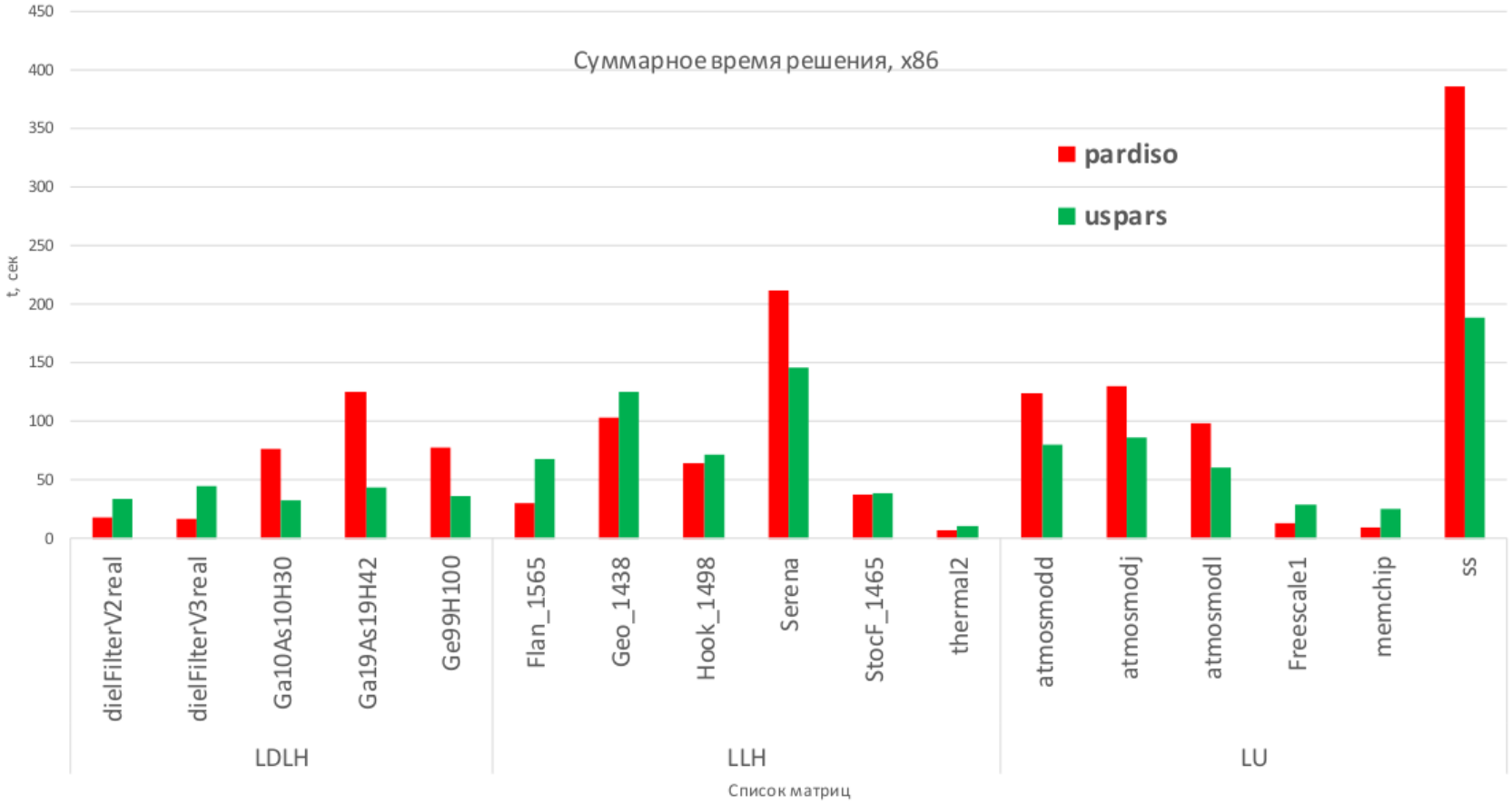
USPARS vs MUMPS vs Accelerate framework on ARM

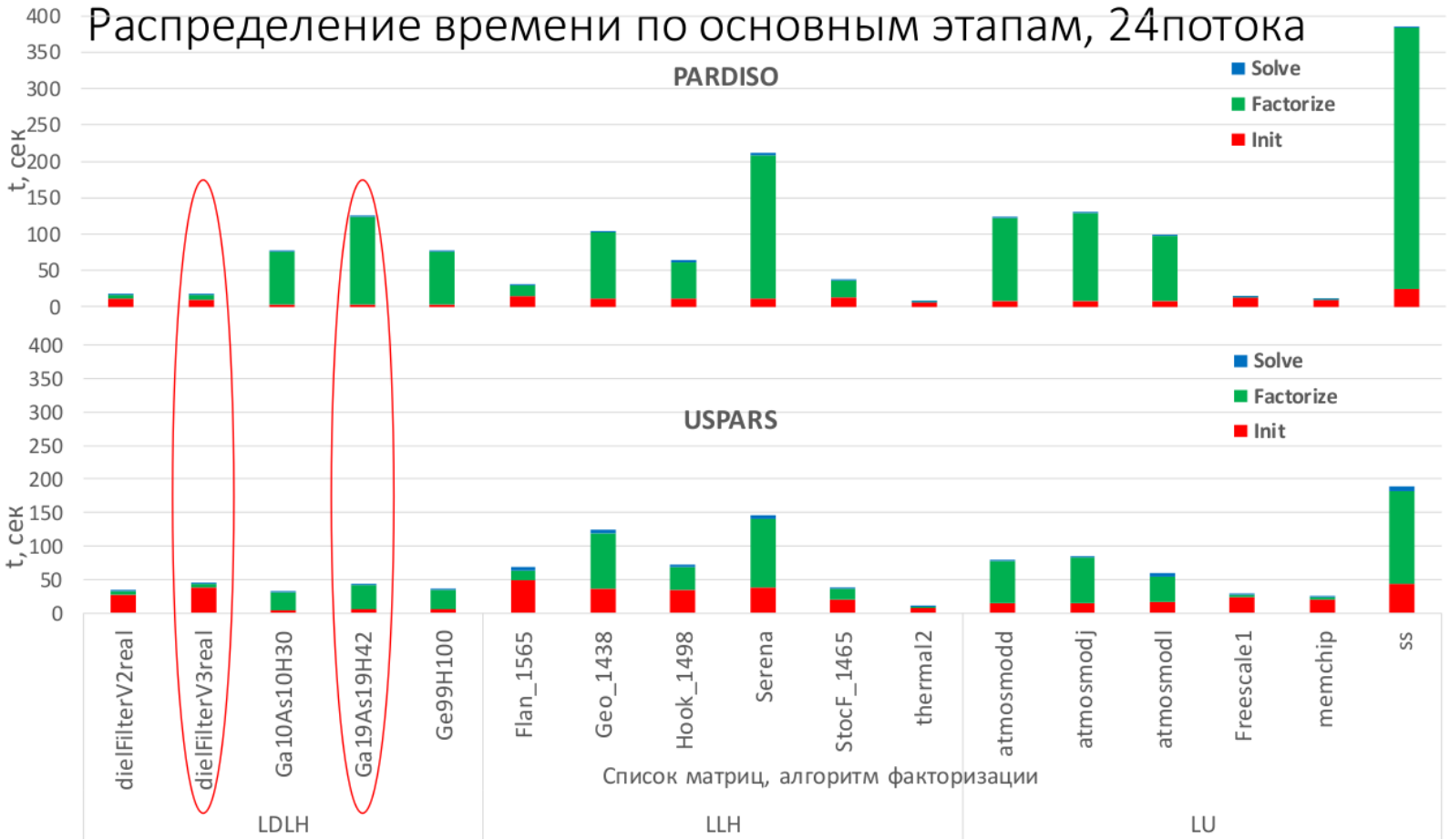
Matrix	Order	NNZ(A)	Computational time (s)		
			USPARS	MUMPS	APPLE
SiH4	5 041	88 472	0.26	0.42	1.05
nemeth24	9 506	758 028	0.37	0.12	0.17
Si10H16	17 077	446 500	2.14	2.21	17.42
crystm03	24 696	304 233	0.32	0.32	0.17
ct20stiff	52 329	1 375 396	1.05	0.6	0.27
t3dh_a	79 171	2 215 638	3.64	3.72	10.77
shipsec1	140 874	3 977 139	3.31	1.6	0.52
c_73	169 422	724 348	0.89	4.24	0.78
Lin	256 000	1 011 200	7.56	8.21	48.45
F1	343 791	13 590 452	16.65	13.1	27.28
c_big	345 241	1 343 126	6.20	6.98	21.36
af_shell1	504 855	9 046 865	4.75	6.68	21.14
af_shell3	504 855	9 046 865	8.64	4.16	6.83
nlpkkt80	1 062 400	14 883 536	313.44	3130	5847.58
thermal2	1 228 045	4 904 179	9.43	11	6.01
Geo_1438	1 437 960	32 297 325	465.33	281.8	246.62

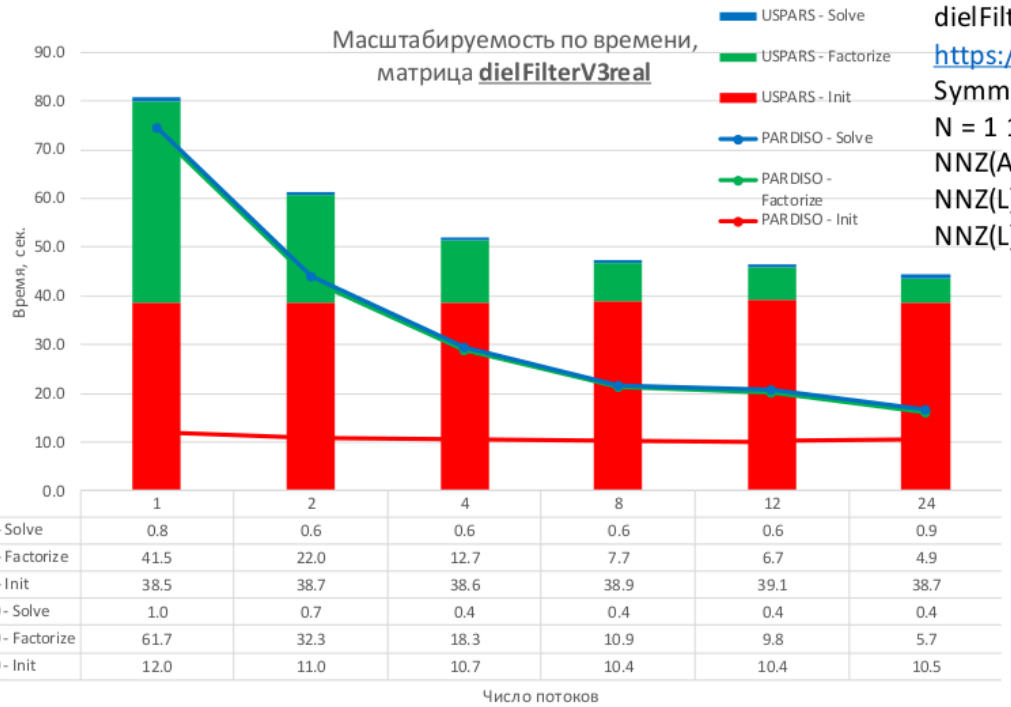
всюду BLAS из AMX2.
USPARS – 4 OMP потока.
MUMPS – 4 MPI процессов.
APPLE – разреженный решатель из Accelerate framework.

- Улучшения
  - Времени, масштабируемости, расходуемой памяти;
  - Блочные варианты факторизации
  - Использование сжатия промежуточных результатов;
- Расширение функциональности
  - Переопределенные и недоопределенные системы линейных уравнений
  - Проблема собственных значений
- Повышение надежности / улучшение точности
  - Итерационное уточнение
  - Стратегии выбора ведущего элемента

Спасибо за внимание

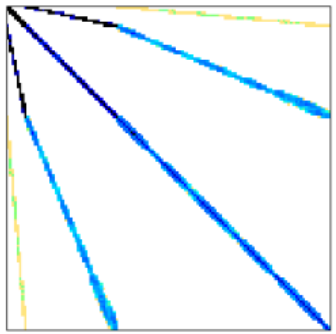

Суммарное время решения, x86
pardiso
uspars
t, сек
450
400
350
300
250
200
150
100
50
0
dielFilterV2real
dielFilterV3real
Ga10As10H30
Ga19As19H42
Ge99H100
Flan_1565
Geo_1438
Hook_1498
Serena
StocF_1465
thermal2
atmosmodd
atmosmodj
atmosmodl
Freescale1
memchip
ss
LDLH
LLH
LU
Список матриц

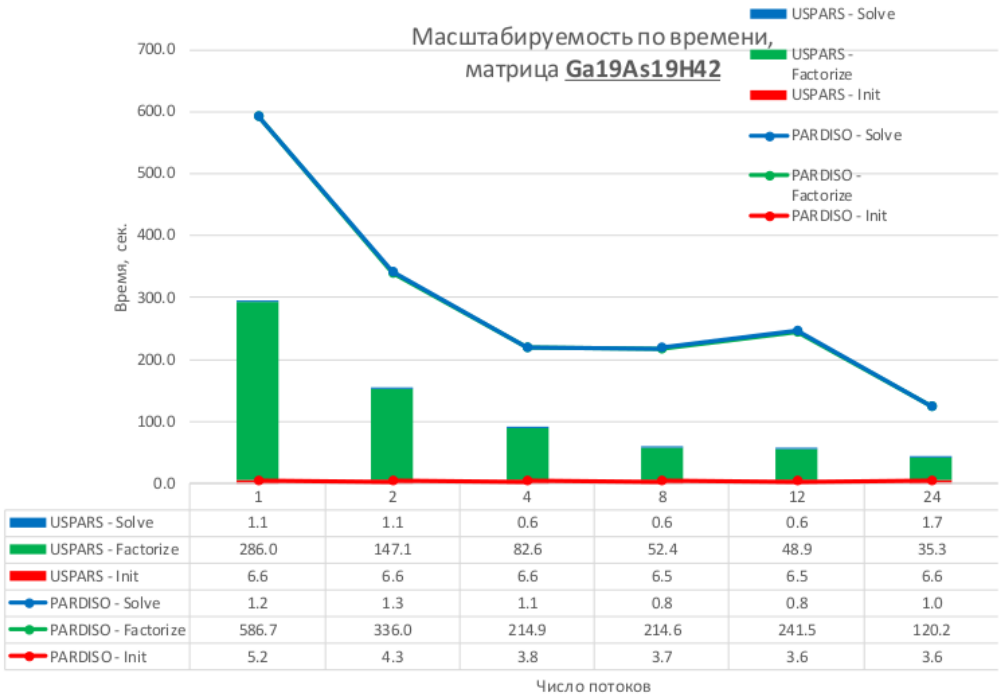

Распределение времени по основным этапам, 24потока
PARDISO
USPARS
Solve
Factorize
Init
t, сек
400
350
300
250
200
150
100
50
0
dielFilterV2real
dielFilterV3real
Ga10As10H30
Ga19As19H42
Ge99H100
Flan_1565
Geo_1438
Hook_1498
Serena
StocF_1465
thermal2
atmosmodd
atmosmodj
atmosmodl
Freescale1
memchip
ss
Список матриц, алгоритм факторизации
LDLH
LLH
LU



	1	2	4	8	12	24
USPARS - Solve	0.8	0.6	0.6	0.6	0.6	0.9
USPARS - Factorize	41.5	22.0	12.7	7.7	6.7	4.9
USPARS - Init	38.5	38.7	38.6	38.9	39.1	38.7
PARDISO - Solve	1.0	0.7	0.4	0.4	0.4	0.4
PARDISO - Factorize	61.7	32.3	18.3	10.9	9.8	5.7
PARDISO - Init	12.0	11.0	10.7	10.4	10.4	10.5

dielFilterV3real
https://sparse.tamu.edu/Dziekonski/dielFilterV3real
Symmetric (LDLH)
N = 1 102 824
NNZ(A) = 45 204 422 (Upper triangle)
NNZ(L) = 543 931 990 (PDS)
NNZ(L) = 602 494 155 (USP)





	1	2	4	8	12	24
USPARS - Solve	1.1	1.1	0.6	0.6	0.6	1.7
USPARS - Factorize	286.0	147.1	82.6	52.4	48.9	35.3
USPARS - Init	6.6	6.6	6.6	6.5	6.5	6.6
PARDISO - Solve	1.2	1.3	1.1	0.8	0.8	1.0
PARDISO - Factorize	586.7	336.0	214.9	214.6	241.5	120.2
PARDISO - Init	5.2	4.3	3.8	3.7	3.6	3.6

Ga19As19H42
https://sparse.tamu.edu/PARSEC/Ga19As19H42
Unsymmetric (LU)
N = 133 123
NNZ(A) = 4 508 981 (Upper triangle)
NNZ(L) = 825 501 440 (PDS)
NNZ(L) = 801 244 103 (USP)