



Table Ronde

Quels Outils pour classer les Machines ?

Participants (6):

- Michel Kern (Maison de la Simulation)
- Jérôme Pansanel (IPHC)
- Arnaud Renard (Mésocentre ROMEO)
- Frédéric Suter (CC IN2P3)
- Gunther Roth (Nvidia)
- Damien Declat (ATOS)

Animateur : Nicolas Renon (Mésocentre CALMIP)



Outils =

des Benchmarks, des métriques (flop/s ; Go/s ; d.o.f/s,)

Linpack (HPL), flop/s

HPCG, flop/s

Green500, Gflop/s/watts

IO-500, GB/s, kIOP/s

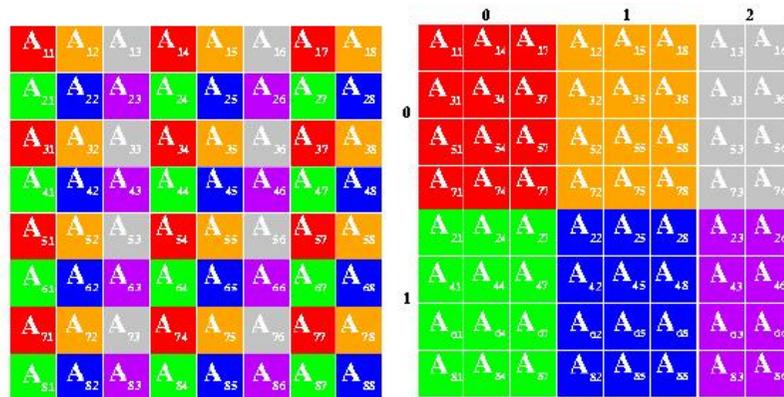
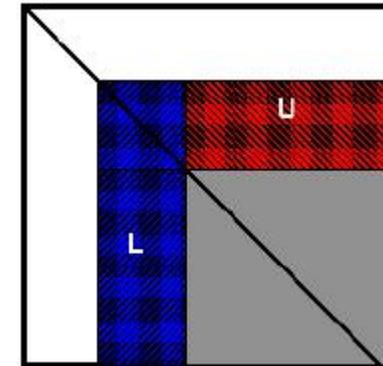
HPGMG, d.o.f./s

...

Top 500 / Linpack

Introduit par J. Dongarra (Linpack. 1979), Top500 en 1993 (H. Meurer + JD)

- Résolution d'un système linéaire
- Matrice pleine, méthode directe
- Caractérisation de cette méthode (et du Linpack) :
 - utilisation efficace des caches (haute intensité arithmétique)
 - peu d'impact des communications



- Factorisation LU (élimination de Gauss), avec pivot partiel sur architecture distribuée
- Vue logiquement comme une grille P x Q
- Algorithme par bloc, équivalent à la version par points
- Nombre d'opérations connu (et imposé)
- $(2N^3)/3$
- Noyau de calcul : produit matrice x matrice
- N^3 opérations, N^2 communications

<http://www.netlib.org/benchmark/hpl/algorithm.html>



Top 500 (juin 2018)

Rank	Nom	Pays	Type	Fournisseur	Total Cores [CPU / Accel.]	Rmax [PFlop/s]	Rpeak [PFlop/s]	Puissance (MW)
1	Summit	USA	Power / GPU	IBM	191 000 2 090 880	122	188	8,8
2	Sunway TaihuLight	China	Sunway	NRCPC	10 649 600	93	125	15
3	Sierra	USA	Power / GPU	IBM	190 000 1 382 400	71	119	
4	Tianhe-2A	China	Xeon / GPU	NUDT	427008 4 554 752	61	101	18
5	AI Bridging Cloud Infrastructure	Japan	Xeon / GPU	Fujitsu	43520 348 160	20	32	1,6
6	Piz Daint	Suisse	Xeon / GPU	Cray Inc.	63840 297 920	20	25	2,3
7	Titan	USA	Opteron / GPU	Cray Inc.	299 008 261 632	18	27	8
8	Sequoia	USA	BG / Q	IBM	1 572 864	17	20	7,9
9	Trinity	USA	Xeon Phi	Cray Inc.	979 968	14	44	3,8
10	Cori	USA	Xeon Phi	Cray Inc.	622 336	14	28	3,9



Top 500 en France

Rank	Name	Manufacturer	Site	Total Cores	Rmax [PFlop/s]	Rpeak [PFlop/s]	Power (MW)	
14	Tera-1000-2	Bull	CEA	561408		12,0	23,4	3,2
30	Pangea	HPE	Total EP	220800		5,3	6,7	4,1
34	JOLIOT-CURIE SKL	Bull	CEA/TGCC-GENCI	79488		4,1	6,6	0,9
70	occigen2	Bull	(GENCI-CINES)	85824		2,5	3,6	1,4
81	Prolix2	Bull	Meteo France	72000		2,1	2,5	0,8
82	Beaufix2	Bull	Meteo France	73440		2,1	2,6	0,8
90	Tera-1000-1	Bull	(CEA)	70272		1,9	2,6	1
144	Sid	Bull	Atos	49896		1,4	1,7	0,5
145	Curie thin nodes	Bull	CEA/TGCC-GENCI	77184		1,4	1,7	2
151	JOLIOT-CURIE KNL	Bull	CEA/TGCC-GENCI	56304		1,3	2,3	0,3
153	Cobalt	Bull	(CEA)/CCRT	38528		1,3	1,5	0,5
167	Diego	Bull	Atos	46800		1,2	1,6	0,5
222	Turing	IBM	CNRS/IDRIS-GENCI	98304		1,1	1,3	0,5
234	Tera-100	Bull	(CEA)	138368		1,1	1,3	4,5
249	Romeo	Bull	ROMEO	17640		1,0	1,5	0,1
297	Sequana BXI	Bull	Atos	35360		1,0	1,6	0,2
301	Eole	Bull	EDF	29568		0,9	1,1	0,4
498	Zumbrota	IBM	EDF R&D	65536		0,7	0,8	0,3



France, Allemagne, > 1Pflop

Rank	Name	Site	Rmax [PFlop/s]	Power (MW)
14	Tera-1000-2	CEA	12	3,2
30	Pangea	Total Exploration Production	5	4,1
34	JOLIOT-CURIE SKL	CEA/TGCC-GENCI	4	0,9
70	occigen2	(GENCI-CINES)	2,5	1,4
81	Prolix2	Meteo France	2,1	0,8
82	Beaufix2	Meteo France	2,1	0,8
90	Tera-1000-1	(CEA)	1,9	1
144	Sid	Atos	1,4	0,5
145	Curie thin nodes	CEA/TGCC-GENCI	1,4	2
151	JOLIOT-CURIE KNL	CEA/TGCC-GENCI	1,3	0,3
153	Cobalt	(CEA)/CCRT	1,3	0,5
167	Diego	Atos	1,2	0,5
222	Turing	CNRS/IDRIS-GENCI	1,1	0,5
234	Tera-100	(CEA)	1,1	4,5
249	Romeo	ROMEO	1	0,1
15			39,5	

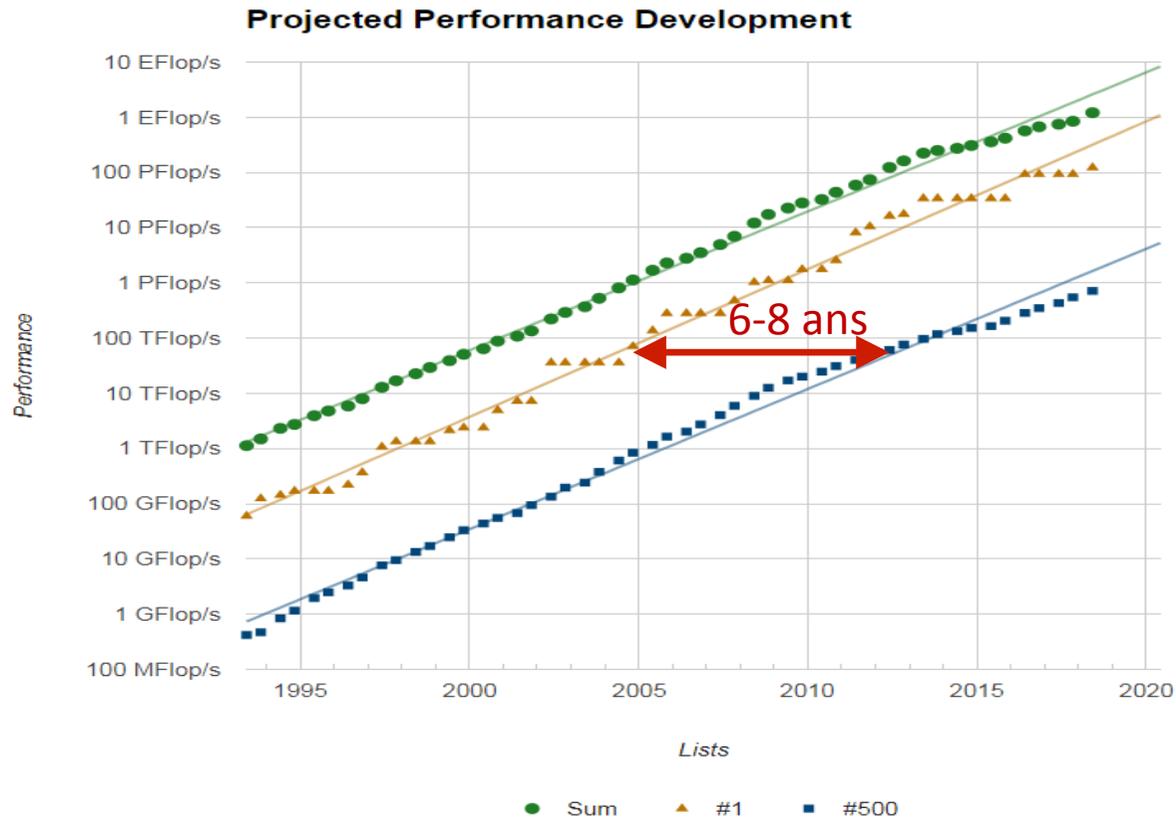
Rank	Name	Site	Rmax [PFlop/s]	Power (MW)
23	JUWELS	Jülich	6,2	1,4
27	Hazel Hen	HLRS Stuttgart	5,6	3,6
28	COBRA	Max-Planck-MPI/IPP	5,6	1,6
38	JURECA	Jülich	3,8	1,3
55	Mistral	DKRZ Klima Zentrum	3	1,1
57	SuperMUC	LRZ München	2,9	3,4
58	SuperMUC Phase 2	LRZ München	2,8	1,5
87	Mogon II	Mainz	2	0,7
94	QURIOSITY	BASF	1,8	
156		Max Planck MPI/IPP	1,2	1,3
170		Deutscher Wetterdienst	1,2	0,5
245	Taurus	TU Dresden	1	0,5
258		Deutscher Wetterdienst	1	0,4
273	Konrad	ZIB Berlin	1	0,9
14			39,3	

Évolution historique

(Beaucoup)
mieux que la loi
de Moore

“I would rather have
today’s algorithms on
yesterday’s computers
than vice versa.”

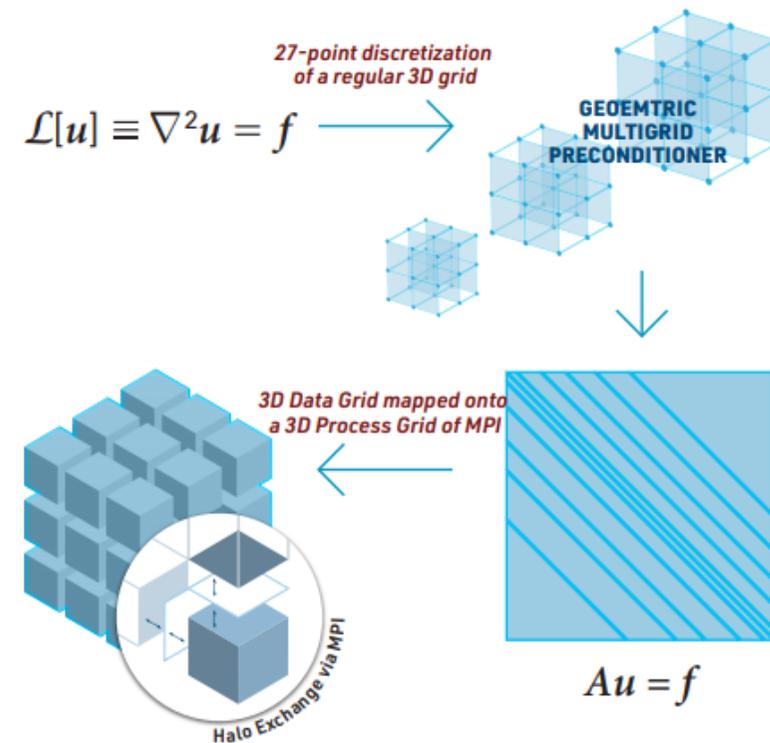
Philippe Toint



HPCG: High performance conjugate gradient

Nouveau benchmark (Dongarra, SC13) qui mesure les performances des opérations de base dans un code unifié

- Algorithme du gradient conjugué préconditionné par multigrille
- Active les noyaux sur une suite de grilles emboîtées
- Multiplication matrice creuse par un vecteur
- Opérations vectorielles
- Produit scalaire (distribué, global)
- Résolution systèmes triangulaires creux (Gauss-Seidel)
- Implémentation de référence en C++, support MPI ou OpenMP
- Caractérisation de cette méthode (et du CG) :
 - Très bande passante (**faible intensité arithmétique**)
 - Poids des communications relativement important



: <http://www.hpcg-benchmark.org/>



HPCG(juin 2018)

Rang	Top 500	Nom	Pays	Fournisseur	Top500 [PFlop/s]	HPCG [PFlop/s]
1	1	Summit	USA	IBM	122	2,93
2	3	Sierra	USA	IBM	71	1,80
3	16	K Computer	Japan	Fujitsu	10	0,60
4	9	Trinity	USA	Cray Inc.	14	0,55
5	6	Piz Daint	Switzerland	Cray Inc.	20	0,49
6	2	Sunway TaihuLight	China	NRCPC	93	0,48
7	12	Oakforest-PACS	Japan	Fujitsu	14	0,39
8	10	Cori	USA	Cray Inc.	14	0,36
9	14	Tera1000-2	France	Atos	12	0,33
10	8	Sequoia	USA	IBM	17	0,33
16	30	Pangea	France	Atos	5	0,16



HEP-SPEC06 (HS06)

- Proposé en janvier 2009 au CERN
- Mesure de performance des infrastructures de calcul pour WLCG
- Maintenant utilisé plus largement par d'autres communautés scientifiques
- Permet d'avoir des *pledges* basées sur un référentiel commun
- Basé sur SPEC CPU2006
- Procédure requière SL6 et compilation 32-bits
- Moyenne de 4 *benchmarks* sur les flottants et 3 sur les entiers
- Problème avec les systèmes virtualisés et les processeurs récents
- → groupe de travail HEPIX (*fast benchmark* & SPEC2017)

Questions :

- qu'est-ce que ces outils nous permettent de dire réellement sur la performance d'une machine ?
(flop/s, consommation kw, performance applicative du point de vue de l'utilisateur= temps de restitution)
- qu'est-ce que ces outils nous permettent de dire sur la performance de telle machine par rapport à telle machine ?
- sont-ils utilisés ou non, ou partiellement, dans les spécifications pour l'acquisition de machines ?
- est-ce que l'adhérence des machines à ces critères conduit à des machines utiles ?
- Horizon à 10 ans, Exascale ou Exaflop pour les Mésocentres ?