

EEA-Aware For Large-Scale Scientific Applications On Heterogeneous Architectures

John Anderson John Anderson Garcia Henao Garcia Henao, Esteban Hernandez, Philippe O.A. Navaux, Carlos Jaime Barrios-Hernandez

▶ To cite this version:

John Anderson John Anderson Garcia Henao Garcia Henao, Esteban Hernandez, Philippe O.A. Navaux, Carlos Jaime Barrios-Hernandez. EEA-Aware For Large-Scale Scientific Applications On Heterogeneous Architectures. ACM Student Research Competition at Supercomputing. The International Conference for High Performance Computing, Networking, Storage and Analysis, Nov 2016, Salt Lake City, Utah, United States. hal-01739585

HAL Id: hal-01739585 https://hal.science/hal-01739585

Submitted on 21 Mar 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



GIIRA



Energetically Efficient Acceleration EEA-Aware For Scientific Applications of Large-Scale On Heterogeneous Architectures John A. García. H.^{1,2,4}, Esteban Hernandez B.^{1,3}, Philippe O. A. Navaux², Carlos. J. Barrios H.¹

^{1.} High Performance and Scientific Computing Center, SC3

- Universidad Industrial de Santander, UIS Bucaramanga, Colombia
- ^{2.} Parallel and Distributed Processing Group, GPPD Universidade Federal do Rio Grande do Sul, UFRGS - Porto Alegre, Brasil

^{3.} Interoperability of Academic Networks Group - GIIRA,

- Universidad Distrital Francisco Jose de Caldas, UD Bogota, Colombia
- ⁴ Laboratoire d'Informatique, Signaux et Systèmes de Sophia Antipolis I3S, Université Côte d'Azur, UNA - CNRS - Nice, France

resources used to obtain high performance.

challenge for build the next generation of Exascale Systems"



 $\frac{1}{\partial t} =$

Planning workload experimental on Ondes3D© StarPU© task programming library for hybrid architectures 7.2 Millions Grid (300×300×80) - 625 MB 36 Task processing (20 Velocity; 16, Stress) 144 Task of communication Block Size 150x150 2D seismic model discretized in stencil Task dependency: Directed Acyclic Graph (DAG) blocks of primary grid. Ref. John A. García H. et al. Energetically Efficient Acceleration EEA-Aware. Degree work to obtain the title of Master of Science in Systems Engineering and Informatics at UIS 2016 Ref. Víctor Martinez et al. Task-based programming on Low-power Manycore Architectures: Seismic Simulation on Jetson TK1. CARLA 2015



FRANCISCO JOSÉ DE CALDA





^{1,2,4} henao@i3s.unice.fr, ^{1,3}ejhernandezb@udistrital.edu.co, ²navaux@inf.ufrgs.br, ¹cbarrios@uis.edu.co

EEA-Aware Structure



<u>vww.i3s.unice.fr/</u>



esting:

Step 2: Architectural Characterization

1	170.12	067	942 12	212 28	149.77	241.04	460.25	
1	441.92	1597.44	635.90	494.85	1871.22	783.43	569.37	
1	337.16	435.06	669.28	337.16	435 43	700.63	337.16	
1	149	695	425.6	170	814.8	548.9	198.8	
1	531.31	222.47	363.28	909.34	327.9	486.76	1344.09	
	5	6	6	6	6	6	6	
	537	228	369	915	333	493	1350	
1	GyPU I	Results		10	8		Î.C.	
1	1	4	8	1	4	8	1	
1	12	3	1	12	3	1	12	
	1	4	8	1	4	8	1	
2	arameter	rs for Nv	idia GP	U				
	24	80	30	30	72	48	48	
	768	360	960	960	576	864	864	
r	ident Pa	rameters	8					
)	2048	768	2048	2048	1024	1536	1536	
			61440		73728			
A	orkload	Paramet	ers		2			
	ET	ES	PC	ET	ES	PC	ET	

The second level uses the enerGyPU monitor in the post-processing for data visualization and statistical characterization of each architecture. In which the Figure below shows that the accuracy using the key factors is much more accurate than just using the number of GPUs. In addition, this monitor displays information via sequence data, statistics, histograms and tables showing results in terms of energy efficiency, for each experiment.

Minimization model with Task and Architectural Parameters

GPUPowerRuntime (GPU_worker[GPU_Factors], GPU_idle[GPU_Factors]) Accuracy of Linear and Polynomial Models Among Learning Curves Multivariable Model



The third level uses the projected multivariable regression model results to see metrics such as time, performance, power consumption, power consumption and performance per watt, which are used to execute the HPL for each combination of computational resources and and calculate the best combination of computational resources, as shown in Figure. Efficiently Energetic Acceleration | EE/

	+++++ FEA Prediction System of Computational Resources for Matrix 49152 to Execute HPL-2.0 on GUANE +++++							
00+[25].								
Uut[25].		Best Prediction	Time_to_Solution	Performance_Rmax	Power_Consumption	Energy_to_Solution	Performance_per	
	block_size	768.000000	1024.000000	768.000000	2048.000000	1024.000000	768.000000	
	Number_task	64.000000	48.000000	64.000000	24.000000	48.000000	64.000000	
	Task_slze	288.000000	384.000000	288.000000	768.000000	384.000000	288.000000	
	GPU	3.000000	5.000000	5.000000	1.000000	3.000000	3.000000	
	Cores	4.000000	2.000000	2.000000	12.000000	4.000000	4.000000	
	Time_to_Solution	93.730042	13.185518	56.792579	751.901756	50.122982	93.730042	
	Performance_Rmax	662.169353	640.145940	674.374812	118.395452	627.940482	662.169353	
	Power_Consumption	546.218050	609.774292	607.899581	356.528464	548.092760	546.218050	
	Energy_to_Solution	55.832045	41.343681	65.305453	241.895380	31.870272	55.832045	
	Performance_per_Watt	1.218755	1.080187	1.151112	0.406359	1.147830	1.218755	

EEA-Aware Scheme for Scalability and Portability of Large Scale Applications on Heterogeneous Architectures:



[1] John A. García H. et al. Energetically Efficient Acceleration EEA-Aware. Degree work to obtain the title of Master of Science in Systems Engineering and Informatics at UIS 2016. [2] John A. García H. et al. enerGyPU and enerGyPhi Monitor for Power Consumption and Performance Evaluation on Nvidia Tesla GPU and Intel Xeon Phi. CCGrid 2016 - IEEE/ACM. [3] John A. García H. et al. eGPU for Monitoring Performance and Power Consumption on Multi-GPUs. XIII Workshop de Processamento Paralelo e Distribuído, WSPPD 2015 - IEEE. [4] Raj jain. The art of computer systems performance analysis - techniques for experimental design, measurement, simulation, and modeling. Wiley professional computing. Wiley, 1991. [5] K. Bergman et al. Exascale Computing Study: Technology Challenges in Achieving Exascale Systems. Peter Kogge, Editor & Study Lead. Technical report, 2008.





Ref. john A. García H. et al. Energetically Efficient Acceleration EEA-Aware. Degree work to obtain the title of Master of Science in Systems Engineering and Informatics at UIS 2016. ef. Jack Dongarra. Algorithmic and Software Challenges For Numerical Libraries at Exascale, 2013.

Step 3: EEA Prediction System

A	Pred	iction	System