BOOK OF ABSTRACTS

cost IC1305

NESUS Winter School & PhD Symposium 2018

22nd - 25th January 2018, Zagreb, Croatia

Dimensionality Reduction in Large Scale Image

Mohammed Amin Belarbi, Sidi Ahmed Mahmoudi, Said Mahmoudi, Ghalem Belalem.
University of Mons, Faculty of Engineering, Computer science department, Belgium
Email : MohammedAmin.BELARBI@student.umons.ac.be

Abstract

Large-scale image retrieval is one of the critical technological fields using Big Data effectively. Content-Based Image Retrieval (CBIR) has become the popular method, which detects and extracts visual features of image (global and local features) automatically by means of image processing and computer vision algorithm. In most of cases, a retrieval system extracts visual features of a query image, and then compare to a set of image features stored in the database. As result, a list of images having similar features with the query are shown to the user [1]. CBIR methods mainly include two key parts: feature extraction and similarity measures. Considering the rapid development of digital image intake equipment, the diversity and complexity of high resolution image content increases the size of image files and images databases. High dimension features are usually extracted to describe image content accurately, especially for large scale image retrieval system [2], where the number of features is large-scale as well. If these high-dimensional data are processed directly, this may lead to the "Curse of Dimensionality" phenomenon which cannot improve search algorithms performances. Dimensionality reduction is one of the effective methods used to overcome these problems [3]. The idea behind these approaches is that Image features are pre-processed by projecting the original data form high dimensional space to a lower dimensional one [4]. Therefore, dimensionality reduction methods play important and significant role to overcome the "Curse of Dimensionality.

References:

[1] B. Cheng, L. Zhuo, and J. Zhang, "Comparative Study on Dimensionality Reduction in Large-Scale Image Retrieval," in 2013 IEEE International Symposium on Multimedia, 2013, pp. 445–450.
[2] H. Cai, X. Wang, and Y. Wang, "Compact and robust fisher descriptors for large-scale image retrieval," in 2011 IEEE International Workshop on Machine Learning for Signal Processing, 2011, pp. 1–6.
[3] E. Bingham and H. Mannila, "Random Projection in Dimensionality Reduction: Applications to Image and Text Data," pp. 245–250, 2001.
[4] M. Belarbi, S. Mahmoudi and G. Belalem, "Indexing Video by the Content", Advances in Intelligent Systems and Computing, pp. 21-33, 2016.

Proactive-reactive Project Scheduling with Flexibility and Quality requirements

Mario Brcic, Damir Kalpic, Nikica Hlupic
University of Zagreb, Faculty of Electrical Engineering and Computing, Department of Applied Computing, Croatia
Email : mario.brcic@fer.hr

Abstract

This research deals with new families of robustness measures, such as Cost-based exibility (CBF) which introduces dependency of rescheduling costs on the temporal distance of the baseline schedule change. A mathematical model describes the need for collaborators synchronization and aims to comply with the due dates. Quality requirements relate to complying with the due dates, while exibility requirements model the need for collaborators synchronization. As objective functions containing such elements are not necessarily regular and the existing theory does not cover that case, theoretical foundations are laid out. Exact solving procedure, as well as the two heuristic procedures are presented. The heuristic procedures are based on approximate dynamic programming and use simulation and statistical inference to make scheduling and synchronization decision. The results show that our algorithms substantially outperform the benchmark algorithm. In conclusion, CBF measure, unlike stability measure, enables proactive rescheduling.

Data Locality Models and Techniques for Extreme Scale Systems

Silvina Caíno-Lores
Carlos III University of Madrid, Computer Science and Engineering Department, Spain
Email : scaino@inf.uc3m.es

Abstract

Nowadays, new hybrid computing models --like high-performance analytics and data-intensive scientific computing-- are pushing forward the need to merge Big Data and traditional HPC. This talk introduces our work on the integration of Big Data platforms like Hadoop and Spark with HPC-oriented simulation libraries. First, we provide a meaningful contextualisation of this topic and discuss the desirable characteristics of future HPC and Big Data platforms in terms of performance, scalability, productivity and fault-tolerance. Then, we analyse how the former could be attained with current technologies. Finally, we describe our efforts to fill the gap between the data abstractions, programming and execution models of both approaches, and briefly introduce our future work.

Towards Smart and Automatic Optimization for Big Data: Real-time Application Monitoring and Proling

Jonatan Enes, Roberto R. Exposito, Juan Touriño
University of A Coruña, Spain
Email : (jonatan.enes,rreye,juan)@udc.es

Abstract

Massive data processing technologies, or Big Data, are being widely used in many elds where traditional approaches are not capable of handling datasets due to their large size and need to be distributed and processed in parallel across many nodes. However, such technologies tend to be highly resource-demanding, not only with commonly monitored resources like CPU and memory, but also with disk and network. To cater for such demands, commodity clusters are created using o-the-shelf hardware, virtual machines, the Cloud or even software containers thanks to the latest operating-system-level virtualization improvements.

Leaving aside the underlying infrastructure used, this PhD Thesis aims to come up with techniques for a more abstract and automatic optimization for the emerging frameworks such as Spark or Flink. However, before any technique is developed, it is vital to have an in-depth analysis available in order to identify both resource or program bottlenecks and overall characterize any application or framework behaviour. To do so, a monitoring and proling framework has been developed, BDWatchdog (http://bdwatchdog.dec.udc.es/), that takes a dierent approach from traditional resource monitoring systems and proling.

With BDWatchdog, the focus is put on the applications executed, rather than on the underlying hosts, in order to properly analyze an application that spans across a cluster. In addition, BDWatchdog has been developed to work in real-time and remain scalable. The available results are two-fold. For monitoring, a series of resource usage data that can be processed to create time series to plot specic applications usages (e.g., the overall disk write bandwidth of a MapReduce application). For proling, the more user-friendly ame graphs are plotted to show at any interval the amount of CPU time consumed by a process, program or subroutine, thanks to the use of stacked calls. Future work will be based upon the data produced to detect bottlenecks, in order to inform the user or even adapt the resources and infrastructure available to tackle such issues.

Performance Optimization of Mobile Applications using Paralldroid

Sergio Afonso[1], Alejandro Acosta[2] and Francisco Almeida[3]
Universidad de La Laguna, Spain
Email : safonsof@ull.es[1], yaacostad@ull.es[2], zfalmeida@ull.es[3]

Abstract

Modern hand-held devices have integrated technologies previously only available for desktop computers. They now integrate multi-core CPUs, GPUs and DSPs on a single chip. From the developer's point of view, the architecture is a traditional heterogeneous system where memory is shared between the CPU and accelerators. The exploitation of all of these heterogeneous processors is required in order to benefit from the computing capabilities of these devices.

Even though frameworks have been created to support software development for hand-held devices, they focus on fast application development. Tools for easily taking full advantage of the underlying architecture are required. Our goal is to provide one such tool without negatively impacting code's maintainability and portability.

Android allows the use of three programming models for writing applications. Java provides a comprehensive API which simplifies application development. Most Android applications are mainly written in Java, so Android developers are familiar with this language. Renderscript is designed for CPU-intensive tasks. It requires the developer to learn a new language. C allows the access to native libraries, and it can sometimes offer performance improvements over equivalent Java code.

Paralldroid is a development framework that allows the automatic generation of C, Renderscript and OpenCL code on the Android platform, starting from annotated sequential Java code. Using Paralldroid, developers write regular Java classes to which they add certain Java annotations. These annotations are used to produce a new class that incorporates the parallel code sections to run on the GPU, using the user-specified target language. This is achieved through Abstract Syntax Tree (AST) transformations.

A data environment is created for each @Target class, which handles communications between Java and the target environment. @Map and @Declare annotations are used in order to specify memory movement semantics. Paralldroid lets the user define data-parallel methods through the @Parallel annotation. All these annotations are applied to the main components of a class (fields, methods, parameters, . . .). This makes it easier to use and more suited to object-oriented programming than other annotation-based approaches like OpenMP or OpenACC.

A new extension developed for the Paralldroid programming model allows application programmers to tune the amount of computation carried out by each GPU thread by just specifying an expression defining a chunk size. The generated code will then implement tiling, a performance optimization technique also known as cache blocking that improves the memory access patterns and the granularity of algorithms. This way the performance can be significantly improved.

A set of benchmarks based on stencil codes show very substantial speedups over Java code when the right chunk sizes are used within Paralldroid. There is still potential for improvement, since it is still the application developer who has to determine the optimal chunk size in order to obtain the best performance. Our current focus is towards an automatic adaptive methodology for the discovery of optimal chunk sizes.

Adaptation of Selected Parallel Algorithms to the PGAS Programming Model and their Implementation in Java with the use of PCJ Library

Łukasz Górski
Faculty of Mathematics and Computer Science, Nicolaus Copernicus University in Toruń, Poland
Email : lgorski@mat.umk.pl

Abstract

PGAS (Partitioned Global Address Space) model is a new and promising concurrent and distributed programming paradigm. It assumes that - given a concurrent application - each of its threads of execution is affiliated with some private address space. Selected variables from this private memory area may become shared and available for addressing by other threads of execution. Many existing implementations are still experimental and feasibility analyses are necessitated.

The thesis aims to adapt selected parallel algorithms to the PGAS model using one-sided asynchronous communication and communication-computation overlap; developed implementations are subject to scalability verification. PCJ library was used in pursue of this aim. It is a Java-based solution facilitating development of PGAS applications under JVM (Java Virtual Machine).

The following kernels were chosen for paralelization: random memory access test, fast Fourier transform (1- and 3-dimensional), parallel version of LINPACK benchmark and parallel differential evolution metaheuristic. The latter algorithm was used in the scientific inquiry pertaining to the optimization of the nematode C. Elegans' connectome model, aiming to align model's predictions with the empirical data. Ultimately, the extensive parameter search allowed to find model's parameters that almost uniformly fall into standard deviation region of experimental data.

The research ended with the conclusion that Java and PCJ library allow for easy exploitation of distributed computing environments and development of complex scientific applications. Those traits were already acknowledged in high performance computing community, and two of developed implementations (random access and FFT kernel) were the basis for the reception of the prestigious HPC Challenge Award for productivity (Class 2 Award for Code Elegance). Highly performing and scalable code can be developed in Java-PGAS environment under the conditions that asynchronous communication and communicationcomputation overlap can be exploited. Such conditions allow to develop software that scales up to the thousands of threads of execution.

Energy-Aware Software for ManyCore Systems

Simon Holmbacka
Åbo Akademi University, Finland
Email : sholmbac@abo.fi

Abstract

Many-core systems provide a great potential in application performance with the massively parallel structure. Such systems are currently being integrated into most parts of daily life from high-end server farms to desktop systems, laptops and mobile devices. Yet, these systems are facing increasing chal-lenges such as high temperature causing physical damage, high electrical bills both for servers and individual users, unpleasant noise levels due to active cooling and unrealistic battery drainage in mobile devices; factors caused directly by poor energy efficiency.

Power management has traditionally been an area of research providing hardware solutions or runtime power management in the operating system in form of frequency governors. Energy awareness in application software is currently non-existent. This means that applications are not involved in the power management decisions, nor does any interface between the applications and the runtime system to provide such facilities exist. Power management in the operating system is therefore performed purely based on indirect implications of software execution, usually referred to as the workload. It often results in over-allocation of resources, hence power waste.

This thesis discusses power management strategies in many-core systems
in the form of increasing application software awareness of energy efficiency. The presented approach allows meta-data descriptions in the applications and is manifested in two design recommendations:

1) Energy-aware mapping
2) Energy-aware execution

which allow the applications to directly influence the power management decisions. The recommendations eliminate over-allocation of resources and increase the energy efficiency of the computing system. Both recommendations are fully supported in a provided interface in combination with a novel power management runtime system called Bricktop. The work presented in this thesis allows both new- and legacy software to execute with the most energy efficient mapping on a many-core CPU and with the most energy efficient performance level. A set of case study examples demonstrate real- world energy savings in a wide range of applications without performance degradation.

Physical Intrusion Detection Method based on Machine Learning and Context-aware Activity Recognition in Real-time

Nenad Katanić
Centre for Informatics and Computing, Ruđer Bošković Institute, Zagreb, Croatia
PhD student at Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia
Email : nkatanic@irb.hr

Abstract

With current maturity and wide accessibility of low-cost sensor technologies, sensor-based human activity recognition is increasingly popular in various domains and innovative applications. Huge amount of research in this area is driven by smart-home assistive living applications focused on development of methods and applications which can provide assistance with everyday human activities. On the other hand, personal security aspect is often neglected. Most of the related work on physical activity recognition assume that sensors are attached on human body. Sensor data is often pre-recorded instead of being processed in real-time and the activity to be recognized and classified is always performed within the same context. This research aims to develop a context-aware method for real-time physical activity recognition and classification based on vibration measurements from a single tri-axial accelerometer placed in the environment where the activity is performed. Proposed method combines data-driven and knowledge-driven approaches to human activity recognition and relies on real-time data stream processing and machine learning techniques. Validity of developed method will be evaluated on the prototype system for physical intrusion detection and intrusion type classification on home door in real-life environment.

Intelligent Information Processing Systems, Successful realization of tasks that require not just algorithmic solutions but an intelligent behavior of program or agent

Nelda Kote
Polytechnic University of Tirana, Albania
Email : nkote@fti.edu.al

Abstract

Nowadays, natural language processing and text mining are an ongoing and important research fields. There are too many research done in natural languages processing in different languages such as English, German, Spanish, Italian, Japanese, Arabic or Chinese but a few in Albanian language. The scope of our research work is to develop a tool for preprocessing the text in Albanian language.

At the first phase of our research work we were focused on identifying the research done in field of natural language processing and text mining in Albanian language. Some research work is done in experimental evaluation of machine learning algorithms for text classification. There are few attempts to develop preprocessing tools as rule-based stemmers or lemmatizers for Albanian language. Analyzing this done work we conclude that there are more rules that are not included in the stemmer or lemmatizer algorithms that could improve its capabilities.

We have made a thorough experimental evaluation of machine learning algorithms used for opinion mining in Albanian language in document level. The experimental results are interpreted with respect to various evaluation criteria for the different algorithms showing interesting features on the performance of each algorithm. We used single-domain and cross-domain opinions corpus categorized as positive and as negative opinions to test the performance of the algorithms. First, we have cleaned the dataset by preprocessing the text data with a stop-word removal and a stemmer. Then the clean dataset is used to train and test the classification algorithms. We use Weka software to evaluate the performance of 53 classification algorithms.

The use of preprocessing phase composed by a stop-word and a stemmer increases the performance of the classification algorithms. But there are too many missing rules and wrong used rules to find the stem of a word.

Based on the grammar and morphology of the Albanian language and in the reviewed research work during this time we have decided to focus in the development of a lemmatization tool using part-of- speech and rules to identify the lemma of a word.

Numerical Analysis and Optimization of Parallel Algorithms for Problems with Big Computational Costs

Rima Kriauzienė[1,2], Andrej Bugajev[1], Raimondas Čiegis[1]
1 Vilnius Gediminas Technical University, Sauletekio ave. 11, Vilnius, Lithuania
2 Institute of Data Science and Digital Technologies of Vilnius University, Akademijos str. 4, Vilnius, Lithuania
Email : rima.kriauziene@vgtu.lt

Abstract

We consider a general strategy of three-level parallelisation. Different levels can be characterised by different sets of properties that should be properly addressed. We consider a special case of these three levels. The first level of parallelisation can represented as a part of an algorithm, i.e. the degree of parallelism can be chosen during the computations. The second level generates tasks with different sizes, i.e. the work amount distribution between tasks is non-uniform – this makes the parallelisation challenging. The third level is dedicated to parallelise the solution of tasks from the second level. The third level can be used to perform load balancing at the second level and the first level is used to improve the scalability of the whole three-level parallel algorithm. The usage of the first level can lead to the efficiency loss. The third level can be used alone, however, it is limited due to Amdahl's law. The second level needs a workload distribution strategy, however, it is not trivial due to non-linearity of computational times curves. We investigate the workload distribution optimisation when on the first level a modified simplex downhill method is used, on the second level there are multiple differential equations being solved and at the third level the Wang's algorithm is used.

Efficient Data Management for Near Real-Time Edge Analytics

Ivan Lujic
Institute of Information Systems Engineering, Vienna University of Technology
Favoritenstrasse 9-11/194, A-1040 Vienna, Austria
Email : ivan@ec.tuwien.ac.at

Abstract

Internet of Things (IoT) devices are expected to surpass mobile phones as the largest category of connected devices in 2018 according to an Ericsson's report, increasing at a compounded annual growth rate of 23 percent between 2015 and 2021. Consequently, the exponential growth in data volume is caused by massive amount of data generated from growing number of IoT sensors. IoT sensors are used in a wide set of applications, such as smart cities, smart grids, eHealth and intelligent traffic management systems. Managing such systems generally requires three steps: (1) collecting data through sensors, (2) processing these data and (3) acting based on the obtained information. Traditionally, such processing is done in the cloud. However, performing data analytics in the cloud data centers brings several challenges: (i) transferring data from sensors to massive data centers may result in high latency; (ii) current bandwidth capabilities and network infrastructures cannot easily scale with the growing amount of data; (iii) critical applications such as traffic management or eHealth, require results of analytics with strict latency and accuracy.

Edge computing has been proposed as a solution to these issues. In edge analytics, analysis and collection of data coming from IoT devices is performed in edge nodes (for example, gateways and micro data centers) that are closer to them than massive data centers. Still, storing a certain amount of data and performing analytics on the edge can result in several problems due to the storage and computing limitations on the edge nodes. Furthermore, in highly distributed IoT systems, missing or invalid data may appear because of different reasons including, monitoring system failures and network failures. Thus, performing analytics on incomplete data can lead to inaccurate results and imprecise decisions, with negative effects on the target systems.

As accurate data analytics require a huge amount of complete historical data, storage efficient near real-time analytics becomes a key issue on the edge. We investigate approaches that dynamically find a trade-off between high forecast accuracy necessary for efficient near real-time decisions, and the amount of data stored in the space-limited storage. We target sensorbased data in IoT environments. By utilizing edge nodes, it is possible to reduce the amount of data traversing through the network, then perform more accurate near real-time decisions in proximity of the user, minimizing latency and overall costs for systems and applications deployed on the edge.

Towards a Smart Selection of Cloud Resources for Multimedia Big Data Computing

Sidi Ahmed Mahmoudi, Sidi Ahmed Mahmoudi, Mohammed Amin Belarbi
University of Mons, Computer Science Department, Belgium
Email : sidi.mahmoudi@umons.ac.be

Abstract

The last few years have been marked by the presence of multimedia data (images and videos) in our everyday lives. The latter are characterized by a fast frequency of creation and sharing since images and videos can come from different devices such as cameras, smartphones or drones. They are generally used to illustrate objects in different situations (airports, hospitals, public areas, sport games, etc.). As result, image and video processing algorithms have got increasing importance for several computer vision applications that should be adapted to the scale of Big Data. In this work, we propose a real time cloud-based toolbox (platform) for computer vision applications. This platform integrates a toolbox of image and video processing algorithms that can be run in real time and in a secure way. The related libraries and hardware drivers are automatically integrated and configured in order to offer to users an access to the different algorithms without the need to download, install and configure software or hardware.

Experimentations were conducted within three kinds of algorithms: 1. image processing toolbox. 2. Video processing toolbox. 3. 3D medical methods such as computer-aided diagnosis for scoliosis and osteoporosis. These experimentations demonstrated the interest of our platform for sharing our scientific contributions related to computer vision domain.

[1] S. A. Mahmoudi et al., “Real Time GPU-Based Segmentation and Tracking of the Left Ventricle on 2D Echocardiography”, International work-conference on bioinformatics and biomedical engineering, 2016.
[2] S. A. Mahmoudi, P. Manneback, “Multi-CPU/Multi-GPU Based Framework for Multimedia Processing”, IFIP Advances in Information and Communication Technology. 456, 2015, 54-65, 2015.
[3] P. D. C. Possa, S. A. Mahmoudi, N. Harb, C. Valderrama, “A New Self-Adapting Architecture for Feature Detection”, 22nd International Conference on Field Programmable Logic and Applications, pp. 643-646, 2012
[4] S. A. Mahmoudi et al., “Traitements d’images sur architectures parallèles et hétérogènes”, Technique et Science Informatiques, 31/8-10 - 2012, 8-9-10/2012, 1183-1203, 10.3166/tsi.31.1183-1203, 2012
[5] S. A. Mahmoudi et al., “Détection optimale des coins et contours dans des bases d’images volumineuses sur architectures multicoeurs hétérogènes”, RenPar’20, Saint-Malo, France, 2011
[6] S. A. Mahmoudi, P. Manneback, “Multi-GPU based Event Detection and Localization using High Definition Videos » “, 4th International Conference on Multimedia Computing and Systems (ICMCS’14), Morocco, 2014
[7] M. A. Larhmam et al., “A portable multi-cpu/multi-gpu based vertebra localization in sagittal MR images”, International Conference Image Analysis and Recognition, pp. 209-218, Portugal, 20141

PID-controlled drug delivery system subject to flow-through bioreactor

Anatolij Nečiporenko[1], Feliksas Ivanauskas[1], Valdas Laurinavičius[2], Mifodijus Sapagovas[3]
[1]Faculty of Mathematics and Informatics, Vilnius University, Lithuania,
[2]Institute of Biochemistry, Vilnius University, Lithuania
[3]Institute of Mathematics and Informatics, Vilnius University, Lithaunia
Email : anatolij.neciporenko@mif.vu.lt

Abstract

Our mathematical model belongs to an intensely studied class of problems, namely differential equations with nonlocal boundary conditions. The system of two parabolic nonlinear reaction--diffusion equations with nonlocal boundary condition is used to model the PID-controlled flow-through bioreactor. Nonlocal boundary conditions are commonly referred to as the boundary (or initial) conditions describing the relationship between the desired solution values on multiple points [1].

Integral-type nonlocal boundary condition links the solution on the system boundary to the integral of the solution within the system inner range. This integral plays an important role in the nonlocal boundary condition and in the general formulation of the boundary value problem [1].

Main peculiarity of the present work is a sufficiently detailed explanation of the physical principles that were the basis of the nonlocal boundary condition, which reflects the control (regulation) principle. This model not only provides a set of diffusion-reaction equations, but also describes the underlying physical process together with its possible applications [1].

The solution at boundary points is calculated using the integral combined with the proportional-integral-derivative (PID) controller algorithm. The mathematical modeling was applied to control the drug delivery systems where prodrug is converted into active form in the enzyme-containing bioreactor [1].

In the present work, we propose a new type of nonlocal boundary condition for the parabolic reaction--diffusion equation system applied to the bioreactor modeling. The condition is nonlocal w.r.t. the time and space domains [1].

[1] Feliksas Ivanauskas, Valdas Laurinavičius, Mifodijus Sapagovas, Anatolij Nečiporenko, Reaction–diffusion equation with nonlocal boundary condition subject to PID-controlled bioreactor, Nonlinear Analysis: Modelling and Control, Vol. 22, No. 2, pp. 261–272, doi:10.15388/NA.2017.2.8

Numerical study of enzyme immobilized microbioreactor systems

Linas Petkevičius, Romas Baronas1
Institute of Computer Science, Vilnius University, Lithuania
Email : linas.petkevicius@mif.vu.lt

Abstract

The investigation of chemical kinetics based models with various nonlinearities is usually investigated using numerical methods. Mathematical modeling and numerical simulation of non-linear mathematical model for digital simulation of an enzyme loaded porous microreactor is investigated. The model is based on a system of reaction-diffusion equations, containing a non-linear term related to the Michaelis-Menten kinetics, and involves three regions: the enzyme microreactor where the enzyme reaction as well as mass transport by diffusion take place, a diffusion limiting region (the Nernst layer), where only the mass transport by diffusion takes place, and a convective region, where the analyte concentration is maintained constant. The finite difference method was used to evaluate concentration at steady state. The influence of the thickness of the diffusion layer on the behaviour of the product emission, as well as the impact of the diffusion modulus and the Biot number has been numerically investigated. The MPI was used to parallelize calculations of numerical scheme. Simulations was also parallelized over independent characteristics using MPI.

Additivity: A Criteria to Select Performance Events for Reliable Energy Predictive Analytical Modeling

Arsalan Shahid, Muhammad Fahad, Ravi Reddy, Alexey Lastovetsky
Heterogeneous Computing Laboratory, School of Computer Science and informatics,, University College Dublin, Beleld, Dublin 4, Ireland
Email : arsalan.shahid@ucdconnect.ie

Abstract

Performance events or performance monitoring counters (PMCs) have been originally conceived, and widely used to aid low-level performance analysis and tuning. Nevertheless, they were opportunistically adopted for energy predictive modeling owing to lack of a precise energy measurement mechanism in processors, and to address the need of determining the energy consumption at a component-level granularity in a processor. Over the years, they have come to dominate research works in this area.

Modern hardware processors provide a large set of PMCs. Determination of the best subset of PMCs for energy predictive modeling is a non-trivial task given the fact that all the PMCs can not be determined using a single application run. Several techniques have been devised to address this challenge. While some techniques are based on a statistical methodology, some use expert advice to pick a subset (that may not necessarily be obtained in one application run) that, in experts' opinion, are signicant contributors to energy consumption. However, the existing techniques have not considered a fundamental property of predictor variables thatshould have been applied in the rst place to remove PMCs unt for modeling energy. We propose to address this oversight in this talk.

We present a novel selection criterion for PMCs called additivity [1], which can be used to determine the subset of PMCs that can potentially be used for reliable energy predictive modeling. It is based on the experimental observation that the energy consumption of a serial execution of two applications is the sum of energy consumptions observed for the individual execution of each application. A linear predictive energy model is consistent if and only if its predictor variables are additive in the sense that the vector of predictor variables for a serial execution of two applications is the sum of vectors for the individual execution of each application. The criterion, therefore, is based on a simple and intuitive rule that the value of a PMC for a serial execution of two applications is equal to the sum of its values obtained for the individual execution of each application.

The PMC is branded as non-additive on a platform if there exists an application for which the calculated value diers signicantly from the value observed for the application execution on the platform. The use of non-additive PMCs in a model renders it inconsistent. This study will further be used to improve energy modeling for modern complex architectures and to improve optimization techniques and design space exploration [2].

References

[1] A. Shahid, M. Fahad, R. Reddy, and A. Lastovetsky. Additivity: A selection criterion for performance events for reliable energy predictive modeling. Supercomputing Frontiers and Innovations, 4(4), 2017.

[2] A. Shahid, M. Y. Qadri, M. Fleury, H.Waris, A. Ahmad, and N. N. Qadri. Ac-dse: Approximate computing for the design space exploration of recongurable mpsocs. Journal of Circuits, Systems and Computers, page 1850145, 2018.