Google Apps

Amir Rahmani <amirr1@uci.edu>

251-DE31 - 55th DAC Submission - Accepted

1 message

papers@mpassociates.com <papers@mpassociates.com> Tue, Feb 20, 2018 at 10:57 PM
Reply-To: papers@mpassociates.com
To: spakan@utu.fi
Cc: cristiana.bolchini@polimi.it, dutt@uci.edu, pakrli@utu.fi, antonio.miele@polimi.it, amirr1@uci.edu

Dear Anil Kanduri,

Congratulations!

Your manuscript titled **"Approximation-Aware Coordinated Power/Performance Management for Heterogeneous Multi-cores"**, paper number **251-DE31**, was accepted for the 2018 DAC Technical Program.

You, as the submitter, are REQUIRED TO TAKE ACTION. As the submitter of the manuscript, you are the **only one that will have access to the Official Acceptance Letter and Acceptance Confirmation Form.**

Please log into your NAVIGATION CENTER at the URL listed below:
https://www.mpassociates.com/ml/account/login.aspx?redirectkey=42

Click "View" next to the submission to access the Official Acceptance Letter and the Acceptance Confirmation Form.

- Download the Official Acceptance Letter for more deadlines and details
- Complete the Acceptance Confirmation Form

Follow the instructions provided on the Acceptance Confirmation Form. The Acceptance Confirmation Form includes the title, list of authors, their affiliations, city, state and country information. You are required to check the title, list of authors, their affiliations, city, state and country to make sure all the information is correct. This information will be used in all print and electronic publications. Once you have checked the information, please submit the form electronically. You must have Adobe Reader to complete this form.

Your Acceptance Confirmation Form is due to the DAC office by February 28.

Best regards,
Anand Raghunathan
55th DAC Technical Program Chair

Google Apps

Amir Rahmani <amirr1@uci.edu>

Re: Paper Accepted - DAC 2018

1 message

Cristiana Bolchini <cristiana.bolchini@polimi.it> Wed, Feb 21, 2018 at 3:12 PM
To: Anil Kanduri <spakan@utu.fi>
Cc: Antonio Rosario Miele <antonio.miele@polimi.it>, Pasi Liljeberg <pasi.liljeberg@utu.fi>, Nikil Dutt <dutt@ics.uci.edu>, "amirr1@uci.edu (amirr1@uci.edu)" <amirr1@uci.edu>

Dear Anil,
thank you.
I received the reviews directly from the web system and I would say that some comments can provide useful hints for future work … and the paper was well received.
Again, nice work.
cristiana

On 21 Feb 2018, at 14:19, Anil Kanduri <spakan@utu.fi> wrote:

Hi,
I am not sure if everybody has access to the portal, below are the 6 reviewer's comments.

251-DE31
Approximation-Aware Coordinated Power/Performance Management for Heterogeneous Multi-cores

Reviewer ID: 51111
Overall Recommendation:
4 - Accept
Novelty
3
Technical depth and soundness
4
Experimental evaluation
4
Writing quality
3

COMMENTS TO THE AUTHOR
The paper presents a run-time resource management strategy for performance guar-
antees under power constraints that exploits accuracy-performance tradeoffs exploiting functionally approximate kernels. The proposed controller integrates approximation with the power knows.?

The paper addresses an interesting problem and reads well. This reviewer appreciated the fact that the approach is validated and evaluated on a real platform.?

==========

Reviewer ID: 37842
Overall Recommendation:
2 - Possible Reject
Novelty
2
Technical depth and soundness
2
Experimental evaluation
2
Writing quality
2

COMMENTS TO THE AUTHOR
This work tackles the problem of achieving performance requirements under power constraints. Traditional resource management strategies may sacrifice computational power (e.g., lowering operating frequency) in order to stay under a safe power limit. In addition to the traditional power/performance knobs of DVFS and task migration, this work uses approximation knobs that may be activated in order to maintain performance (expressed as a desired rate of progress/output in applications) when actions for power reduction are taken. When approximations are activated, there is a reduction in computational load, and therefore the system may be able to achieve the desired performance (albeit with reduced quality of results) even with reduced computational capabilities. - The work is compared with two previous papers, showing favorable results. - The metric used for performance, heartbeat/sec, is only useful for specific classes of long-running applications. It is unclear how this work would apply to applications that do not fit into the model of a long-running loop yielding results periodically. - There is no discussion on the impact of the approximations to quality of app results. In the extreme case, the approximation may yield unusable results. If the approximations always yield ?good enough? results, why not apply the approximation at all times, and save energy? - The model used to estimate power is very simplistic. The experimental section implies that on-board power sensors were used (rather than the model in 5.3). It is unclear how power is actually measured/estimated in the experiments. - The work in ref. [1] is more general (applicable to more types of approximations), and tackles essentially the same problem of maintaining required performance (formulated in terms of Service-Level Agreements) under resource constraints. Baek?s paper furthermore integrates app quality estimations in its optimization loop, and demonstrates the work with a broad range of apps. - The authors may want to consider integrating a general approximation framework such as Green or Petabricks [Ansel] into the task migration and DVFS optimization loop presented in this work.

==========

Reviewer ID: 62233
Overall Recommendation:
3 - Possible Accept
Novelty
3
Technical depth and soundness
3
Experimental evaluation
3
Writing quality
4

COMMENTS TO THE AUTHOR
The paper
presents an approach to combine run-time approximation to run-time management techniques like DVFS, task migration, power gating and reduction in CPU quota for specific tasks for an overall ?hybrid? run-time resource manager in the context of heterogeneous multi-cores platforms.

Overall,
the paper presents a nice idea and it is well written. The approach has been tested on an ARM big.LITTLE architecture, by adopting different benchmarks. The results show the effectiveness of the combined management technique and a comparison with some approaches at the state of the art. I?d suggest the authors to extend the analysis and description of the approximation concept in the decision strategy and experimental results. The authors state that the level of approximation is set proportional to the estimated performance loss. It would be interesting to know the approximation levels of application necessary to make the approach really efficient.

Minor
comments:

-) eqn (2)
pag 5. I guess \sum_{i=1}^N

==========

Reviewer ID: 44123
Overall Recommendation:
3 - Possible Accept
Novelty
4
Technical depth and soundness
3
Experimental evaluation
2
Writing quality
4

COMMENTS TO THE AUTHOR
The paper presents a run-time power management (RTM) scheme.
The authors propose to add adaptive imprecise computation as additional control parameter for RTM next to DVFS, task migration, power gating and CPU quota.

?

Strengths of the paper

- Low power computation of machine learning kernels is a great challenge. Hence, the topic addressed in this paper is highly relevant. Approximate computing is the direction many groups now focus on, yet the idea to manage the imprecision together with power at run-time is novel and interesting.

- The RTM heuristic in Algorithm 1 seems well designed, simple and straight-forward to apply.

- The paper is well written and the approach is clearly described.

Weaknesses of the paper

- The experimental results are not completely convincing to me. The authors claim in the conclusion that the loss of accuracy is acceptable. From the provided numbers in Table 3, this is hard to judge for the reader. No data on how the application is affected by this approx.. is provided.

- The paper focuses heavily on ARM?s BIG.little architecture. It remains unclear whether the solution is general, e.g., how the solution (algorithm 1) performs on other architectures.

- The authors present their results for very long software scenarios of up to 600s. The control period for the RTM method is in the seconds range. Existing approaches ? to my knowledge - update their strategy much more frequently to adapt fast to changing computing loads. It remains unclear what specifically motives these long control periods.

==========

Reviewer ID: 15986
Overall Recommendation:
3 - Possible Accept
Novelty
3
Technical depth and soundness
3
Experimental evaluation
3
Writing quality
4

COMMENTS TO THE AUTHOR
- Summary:

The paper describes run-time resource management strategy for performance guarantees under power constraints that exploit accuracy-performance trade-offs for error-resilient approximate applications. The described system controls the following knobs: DVFS, CPU quota, task migration and approximation levels. The strategy uses a coordinated approach to control power-approximation knobs in reactive and proactive manner. Authors described the resource allocation policy and made a direct evaluation with two state-of-the-art approaches.

Pros:
+ Good and easy-to-understand motivation to use coordinated strategy is provided.

+ RTM system works with multiple applications

+ RTM evaluates performance loss/gain in order to proactively tune the knobs

+ Experimental results show that the proposed strategy outperform other state-of-the-art approaches both in terms of power and performance violation.

+ Paper is easy-to-read and well-structured

Cons:
- The overhead of the described approach is not evaluated.

Minor details
* Algorithm 1 should be revised:?
? ?* Lines 7-13 are executed for each application on big cluster that is not migrated to the little cluster. In the text, these lines are executed if no application is migrated.?
? ?* Line 7. According to the algorithm only VF values of the little cluster are updated.??
? ?* Line 11. app -> app_i?
? ?* Line 12. app_i.perfEstimate is absolute value, while on Lines 11, 30 it is a relative value.?
? ?* Line 43. u -> U
?
* According to Figure 3, the scenarios with the performance violations in the idle phase could be also considered, though in the considered approximate applications this might never happen.

* Equation 3, F_i -> F_j where j is a cluster

==========

Reviewer ID: 51907
Overall Recommendation:
2 - Possible Reject
Novelty
3
Technical depth and soundness
2
Experimental evaluation
2
Writing quality
3

COMMENTS TO THE AUTHOR
Strength
- This paper proposes a power/performance management technique with application approximation on heterogeneous multicore systems.
- The authors discussed diverse existing techniques, and compared with them in the evaluation.

Weakness
As discussed in Introduction, some specific applications can be approximated. For the readers who are not familiar with those applications and this topic, it is good to provide a more detailed explanation about the scope of applications.
- In evaluation, this paper considers a limited set of applications, mainly oriented to small case studies. To show the practical value of the proposed technique, more applications should be considered.
There are a number of papers regarding approximation of diverse applications, e.g., image processing. They also use different approximation mechanisms from hardware to software.?Here?s a part of them:
i) Imani, Mohsen, et al. "ACAM: Approximate Computing Based on Adaptive Associative Memory with Online Learning." ISLPED. 2016.
ii) Sampson, Adrian, et al. "Accept: A programmer-guided compiler framework for practical approximate computing." University of Washington Technical Report UW-CSE-15-01 1 (2015).
iii) Rahimi, Abbas, et al. "Approximate associative memristive memory for energy-efficient GPUs." Proceedings of the 2015 Design, Automation & Test in Europe Conference & Exhibition. EDA Consortium, 2015.
- The experiment section only provides high-level evaluation of power and performance of the tested case studies. In-depth evaluation are highly required: e.g., controlling behaviors of diverse scenarios, and overhead of the proposed techniques.
- The efficiency gain of approximate computing stems from the quality loss of application outputs. Thus, this should be also thoroughly evaluated with the experimental results.
- The system may run a set of the applications, which can be approximated, with non-approximated applications. This situations would be usual in practice, e.g., running background tasks initiated by either user-level applications or OS. It would be good to discuss how the proposed technique works in those cases.

On Fri, Feb 9, 2018 at 10:11 PM, Antonio Rosario Miele <antonio.miele@polimi.it> wrote:

> Congratulations anil :)
>
> Inviato da iPhone

Il giorno 09 feb 2018, alle ore 20:25, Cristiana Bolchini <cristiana.bolchini@polimi.it> ha scritto:

very well, great news
cristiana

excuse my brevity

From: Amir Rahmani <amirr1@uci.edu>
Sent: Friday, February 9, 2018 6:29:12 PM
To: 'Anil Kanduri'; Antonio Rosario Miele; 'Pasi Liljeberg'; Cristiana Bolchini; 'Nikil Dutt'
Subject: RE: Paper Accepted - DAC 2018

Congrats Anil! Excellent news!

Best,

Amir

From: Anil Kanduri [mailto:spakan@utu.fi]
Sent: Friday, February 9, 2018 9:06 AM
To: Antonio Miele <antonio.miele@polimi.it>; Amir Rahmani <amirr1@uci.edu>; Pasi Liljeberg <pasi.liljeberg@utu.fi>; cristiana.bolchini@polimi.it; Nikil Dutt <dutt@ics.uci.edu>
Subject: Paper Accepted - DAC 2018

Hi,

Our paper "Approximation-Aware Coordinated Power/Performance Management for Heterogeneous Multi-cores" (251-DE31) has been accepted at DAC 2018, as per their website https://dac.com/2018/accepted-papers. *I haven't received any notification yet though.

Many thanks for your guidance and contributions.