Received: 15 February 2020 | *Revised: 10 March 2020* | *Accepted: 15 March 2020*
DOI: 10.37917/ijeee.00.0.0 *Vol. | Issue | Month Year*

Open Access

Iraqi Journal for Electrical and Electronic Engineering
Original Article or Review Article or Short Article

Understanding Power Gating Mechanism Based on Workload Classification of Modern Heterogeneous Many-Core Mobile Platform in the Dark Silicon Era

Haider Alrudainy*[1], **Muaad Hussein**[1], **Ali K. Marzook**[2], **Ali Aalsaud**[3], **Rishad Shafik**[4]
[1]Basra Technical Engineering College, Southern Technical University, Basra, Iraq
[2]School of Electrical Engineering, Basra University, Basra, Iraq
[3]School of Engineering, Al-Mustansiriyah University, Baghdad, Iraq
[4]School of EEE, University of Newcastle, Newcastle upon Tyne, NE1 7RU, England, UK

Correspondence
* Haider Alrudainy
Electronic Department, Basra Technical Institute, Basra, Iraq
Email: h.m.a.alrudainy@stu.edu.iq

Abstract
The rapid progress in mobile computing necessitates energy efficient solutions to support substantially diverse and complex workloads. Heterogeneous many-core platforms are progressively being adopted in contemporary embedded implementations for high performance at low energy cost estimations. These implementations experience diverse workloads that offer drastic opportunities to improve energy efficiency.
In this paper, we propose a novel per core power gating (PCPG) approach based on workload classifications (WLC) for drastic energy cost minimization in the dark silicon era. Core of our paradigm is to use an integrated sleep mode management based on workloads classification indicated by the performance counters. A number of real applications benchmark (PARSEC) are adopted as a practical example of diverse workloads, including memory- and CPU-intensive ones. In this paper, these applications are exercised on Samsung Exynos 5422 heterogeneous many-core system showing up to 37 to 110% energy efficient when compared with our most recent published work, and ondemand governor, respectively. Furthermore, we illustrate low-complexity and low-cost runtime per core power gating algorithm that consistently maximize IPS/Watt at all state space.

Keywords
Dark silicon, Energy-efficient, Multi-core mobile system, Per core power gating, Workload classification.

I. INTRODUCTION

In the recent times, the continuing demand of low energy cost at desirable throughput has led to the advent of heterogynous many core mobile systems. These platforms, characterized by an ever rising number of cores on a single chip, provide significant computational capability. However, this increasing number of core incorporates with a significant set of challenges, profoundly emphasized by the emergence of dark silicon [1]. In the same context, continuing scaling the technology node according to Moore's Law has led to reach to a point at which large portion of the chip has to be shut down to avoid significant power consumption. It is demonstrated that at 22 nm technology node 21% of a chip must be powered off. While at 8nm technology the percentage of the dark silicon portion increases drastically to more than 50% [1][2]. Other researchers show that 64% of the total 64-core chip has been observed as dark silicon [3][4][5]. Thus, it is predicted that the power consumption of many core platforms will be increased by a factor of 10 over the next decade due to the dark silicon phenomenon [6].

Unlike homogenous many-core systems, heterogeneous many-core platforms are widely being recently adopted in contemporary embedded mobile implementations. This is due to its superior energy efficiency at the desirable throughput compared to the homogenous cores counterparts. To alleviate the trade-offs between energy consumption and throughput a common approach is to assign heterogeneous computing resources (cores) on these platforms. Contemporary platform such as Samsung Exynos 5422 big.LITTLE octa-core system, which comprises 4 LITTLE (ARM A7) cores, and 4 big (ARM A15), is a reasonable choice of illustration in this work.

TABLE I.
LIMITATIONS AND FEATURES OF THE PRESENT APPROACHES

Reference	Architecture	Verification	Design abstraction	Platform	Key of Novelty
[11]	Homogeneous	Hardware	system	Not specified	power gating +Task mapping
[12]	Homogeneous	Hardware	Micro-architecture	ARM Cortex-M0	Sub-clock Power gating,
[13]	Homogeneous	Hardware	system	Intel Core I7	Power gating, (Turbo boost)
[14]	Homogeneous	Hardware	system	AMD Opteron 6168	Power gating, manually adjusted DVS
[15]	Heterogeneous	Hardware	system	Odroid-XU3	Low complexity+ Realtime+DVFS+TM
[16-20]	Heterogeneous	Hardware	system	Odroid-XU3	Realtime +DVFS+TM
[21]	Heterogeneous	Simulink (Gem5)	system	ARM Cortex-A15	Power modelling
Proposed	Heterogeneous	Hardware+ Simulink (Cadence)	system	Odroid-XU3	PCPG based on workload classifications +DVFS+TM

Over the past few years a substantial research has been conducted to meet energy cost reduction in heterogeneous embedded mobile platforms such as those from Arm and Intel [7-9] [15-18]. Such efforts normally manage dynamic voltage frequency scaling (DVFS) decisions, combined with the core allocation to threads to respond to workload variations. For instance, when a higher workload is experienced more number of cores are assigned with appropriately chosen DVFS combination. On the other hand, when a lower workload is encountered, fewer cores are allocated with decreased DVFS combination.

Although DVFS coupled with core allocation (TM) play a significant role in minimizing dynamic energy consumption in contemporary heterogynous many core systems, dark silicon contributes to significantly unuseful power consumption, principally decreasing the battery operating active time. To decrease the dark silicon energy consumption, power gating technique was adopted in many recent published research [10-14]. The fundamental key is to utilize a layer of sleep transistors to shut down the inactive cores by disconnecting the power supply voltage. In this paper, we propose a novel per power gating technique combined with DVFS and thread to core allocation in order to drastically decrease energy consumption. To the best of our knowledge, we consider that per core power gating (PCPG) based workload classification (WLC) performed on Samsung Exynos 5422 heterogeneous mobile system to be a novel effort. In our proposed approach, the following major contributions has been made:

- propose a per core power gating (PCPG) approach for contemporary heterogeneous mobile platform based on workload classification to effectively support various workloads,
- core of the approach is an integrated power saving management for dark silicon area based on workload classification metrics, modelled adopting the performance counters feedback,
- validate by means of various type of real application benchmarks to illustrate reasonable advantages and trade-offs.

The rest of the paper is structured as follows. Section 2 limitation and features of the present approaches are extensively explained. The system architecture and application is comprehensively described in Section 3. Workload classification metrics obtained from performance counter, and PCPG control decision based on workload classification details have been demonstrated in Section 4. The proposed approach is expansively discussed in Section 5, deals with per core power gating management and power switch network. Section 6 discuss the results of the experiments, and, lastly, Section 7 provides the conclusion the paper.

II. RELATED WORK

Energy efficiency of many-core mobile platforms has been investigated expansively in recent years. Table I outlines limitations and contributions of the most recent existing approaches. Over recent years significant research has been conducted addressing real-time energy reduction approaches. These techniques have taken into account single metric based optimization: mainly performance improvement within a particular power budget, or performance-constrained for power reduction [18]. For instance, real-time dynamic voltage frequency scaling (DVFS) control method for power reduction of many-core embedded platforms has been proposed in [19]. Their method utilizes performance and user experience constraints to obtain the minimum DVFS combinations by adopting reinforcement learning and transfer principles. Others illustrated another power reduction method that models real-time workload analysis to constantly maintain the core allocations and DVFS combination through predictive controls using multinomial logic regression [20]. A number of research papers have also demonstrated analytical investigations adopting simulation frameworks, including

McPAT, and gem5. These studies have utilized task mapping, DVFS, and offline optimization methods to significantly reduce the power dissipation under workloads variations [21]. A novel work in [15] presented low complexity runtime management approach based on workload classification for heterogeneous many core platforms. This approach addresses most configuration space of odroid-xu3 platform including core types, threads allocation, optimum dynamic voltage and frequency scaling.
A hardware based stateless load balancing scheme for homogeneous many-core system is assessed in aspect of power consumption and thermal behavior [11]. In this scheme, a power minimization is reached by powering off the dark silicon area. In [12], to minimize static power consumption during the sub-clock cycle, a power gating based sub-clock approach was implemented in ARM Cortex-M0 processor. In the same context, Charles et al. [13] performed per core power gating (PCPG) in contemporary homogeneous Intel Core i7 processor. It is illustrated that additional power headroom can be transferred to the active cores by power gating dark silicon area, idle cores, to boost their frequency and voltage without violating the thermal and power envelop. Likewise, transferring the saved power from dark silicon area into enabled cores was studied in [14] using a homogeneous many-core platforms named as AMD Opteron 6168. The practical outcomes of this work are relied on manually adjustment of dynamic voltage scaling (DVS) combination integrated with per core power gating approach.

III. SYSTEM ARCHITECTURE AND APPLICATIONS

The impetus of adopting heterogeneous architectures, comprising two or various types of CPUs, is recently increasing. Although these platforms provide superior performance, it is essential to ensure optimum energy consumption while exercising various types of workloads. The Odroid-XU3 board facilitates approaches including affinity, DVFS, and core manually disabling, normally utilized to enhance system operation in respect of energy consumption and performance. The Odroid-XU3 board is a small heterogeneous 8-cores computational platform. This board can run Android 4.4 or Ubuntu 14.04 operating systems. The primary element of Odroid-XU3 board is the 28 nm Application Processor Exynos 5422. The main processor architecture depicted in Fig. 1. This multiprocessor system on chip (MPSoC) is developed by ARM big.LITTLE heterogeneous architecture and comprises of a low power Cortex-A7 quad core block, a high performance Cortex-A15 quad core processor block, 2GB DRAM LPDDR3, and a Mali-T628 GPU. Further, this board comprises of 4 real time current sensors that provide the opportunity to measure power consumption on the 4 separated power blocks: little (A7) CPUs, big (A15) CPUs, DRAM, and GPU. In addition, there are also 1 temperature sensor for the GPU and 4 temperature sensor for each of the A15 CPUs. The clock frequency and supply voltage (Vdd) of the Odroid-XU3 board, for each power block, can be adjusted using a range of pre-defined range of values. For example, the low power Cortex-A7 quad core block has a set of frequencies ranged between 200 MHz and 1400 MHz with a step size of 100 MHz, while the performance Cortex-A15 quad core block features a set of frequencies ranged between 200 MHz and 2 GHz with a step size equal to 100MHz.
The PARSEC real application benchmark suite supports both emerging and current workloads for multiprocessing hardware [22]. It contains a various set of workloads from diverse domains including systems applications or interactive animation that mimic large-scale commercial workloads. In our paper, Therefore, PARSEC applications has been adopted and exercised on the Odroid-XU3 system on chip (SoC) whose heterogeneity can be illustrative of various design choices that can significantly impact workloads. PARSEC benchmark suite experience diverse: data sharing patterns, workload partitions, and memory behaviors from majority other benchmark suites in widespread use. Table II shows the characteristics of PARSEC benchmark suit which are adopted in our work. Three set of applications (ferret/cannel, fluidanimate/streamcluster, and bodytrack) are opted to illustrate CPU-intensive, memory-intensive, and mixed memory with CPU-intensive, respectively.
DVFS of odroid-xu3 platform is enabled by the power governors at the system software layer. For example, Linux incorporates various power governors that can be actuated based on the system demands. These comprise *powersave* for low performance and low power mode, *performance* for higher performance mode, *ondemand* for performance-sensitive DVFS level, and *userspace* for user-customized DVFS combination. These governors aim to appropriately adjust the voltage/frequency combination in compliance with energy and performance requirements of the exercised workload.
Core allocations to threads (TM) are typically governed by a scheduler [23]. A Linux scheduler typically allocates the overall workload across all available cores to attain substantial utilization. However, given a certain performance requirement, various types of threads must be processed differently for energy and performance optimization. For example, there is typically no differentiation about the specification of thread being allocated, including memory- or CPU-intensive. Tackling energy efficiency in heterogeneous many core systems exercising concurrent workload behaviors requires a great deal of effort. This is because the state space is significantly large and each workload demands different optimization. Therefore, the hardware state space of a many-core heterogeneous platform comprises all practicable DVFS combinations and threads to core allocations (TM).

TABLE II.
CHARACTERISTIC of PARSEC BENCHMARK [22]

Program	Application domain	Type
ferret	Similarity Search	CPU
cannel	Engineering	CPU
bodytrack	Computer Vision	CPU+mem
streamcluster	Data Mining	mem
fluidanimate	Animation	mem

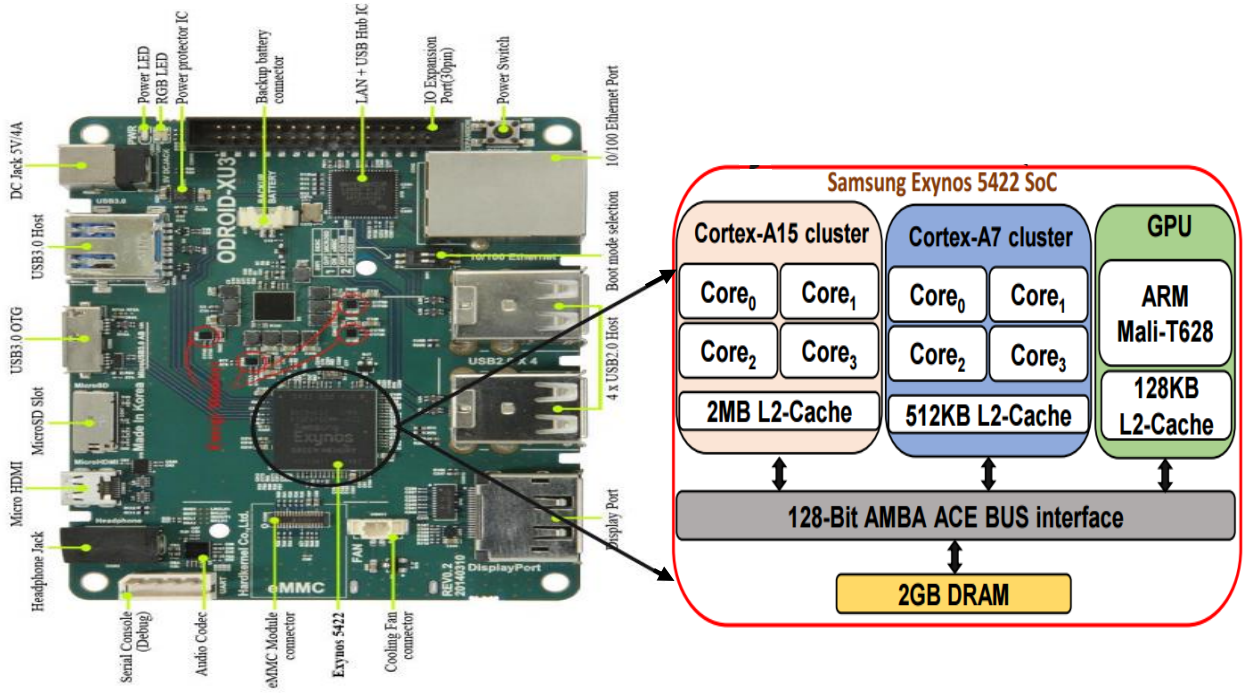

Fig. 1. Odroid-XU3 board comprising Samsung Exynos 5422 heterogeneous MPSoC.

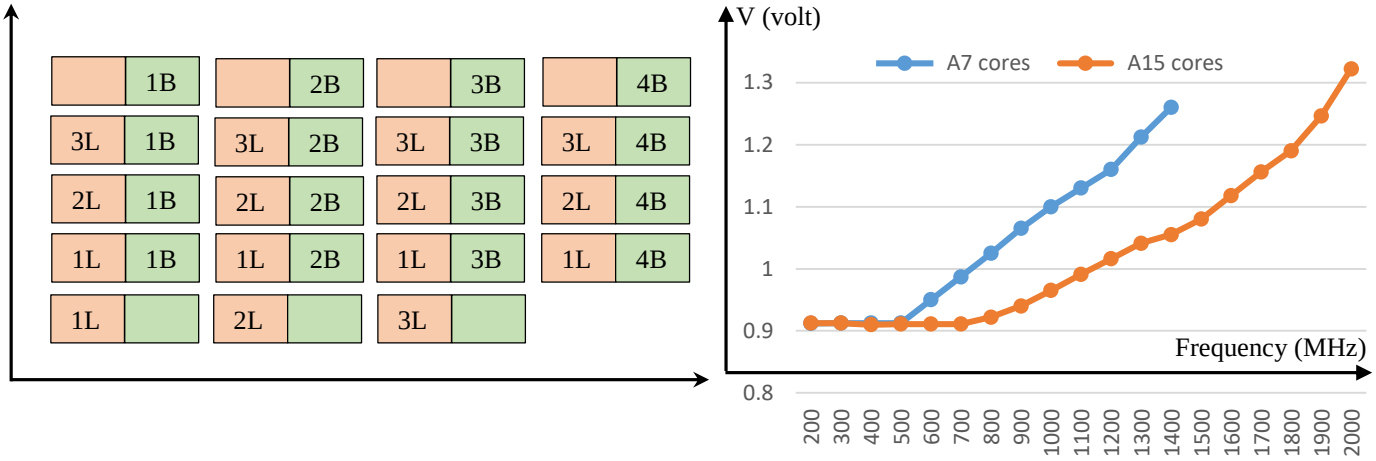

Fig. 2. a) number of potential core allocations of exercising single application; b) experimental data of cannel application demonstrating the number of DVFS combination for A7 and A15 cores.

For instance, the number of feasible big.LITTLE core allocations of exercising single application is 19 as can be illustrated in Fig. 2(a). Considering maximum of one thread per core is permitted, each application must have at least one thread, and one of the little cores is employed for running the operation system. These possible scenarios are latterly multiplied by the potential range of DVFS combinations as shown in Fig 2(b), which is computed as MA15. MA7, where MA15 is the potential of DVFS combination in the A15 block, and MA7 is the potential of DVFS combination in the A7 block. Exercising two applications concurrently requires 111 possible core allocation making energy efficiency optimization is extremely challenging. Therefore, exercising concurrent applications can make energy efficiency optimization a non- trivial task as the number of possible scenarios will exponentially increases.

IV. WORKLOAD CLASSIFICATIONS

The categorization of workload classes defined in the present work distinguishes between memory-intensive and CPU-intensive workloads, with low- or high-activity. Precisely, workloads are categorized into the following listed four classes:

- Class 0: low workloads activity;
- Class 1: intensive CPU workloads;
- Class 2: intensive memory and CPU workloads; and
- Class 3: intensive memory workloads.

Large-scale investigative experiments are exercised in our previous published work in [15] to examine the rationality of these inclusive concepts. These investigations demonstrate that optimum energy efficiency can be achieved if fewer little cores are used from memory-intensive application while it is advantageous to exercise more big core in parallel

for CPU-intensive applications. The classification is achieved by computing a range of metrics from performance counter readings, hence obtaining the classes relying on whether if these metrics have traversed a predetermined threshold as can be illustrated in Table III. For instance, the more nnmipc rises, the more CPU-intensive workload becomes. Therefore, for CPU-intensive workload all A7 cores must be powered off while A15 cores can be activated at higher DVFS level to maximize energy efficiency (IPS/Watt). Contrary, for memory-intensive workload, all A15 cores must be power gated while all A7 cores can be activated at maximum DVFS level. Power gating of big cores while exercising memory intensive workloads can drastically improve energy efficiency of heterogynous many core systems due to their high power consumption when they act as dark silicon.

V. PROPOSED APPROACH

Our proposed paradigm interacts with runtime performance counter to appropriately compute classification metrics of the exercised application, thereby application class can be determined. As a result, dark silicon cores can be power gated according to various workload scenarios. Power gating of unused cores can be achieved by using power switch network (PSN) based CMOS transistors. This PSN is controlled by adopting a PCPG management routine. In the following sections we briefly describe our approach, emphasizing the Exynos 5422 platform and PCPG based WLC interactions.

A. Performance counter

The Exynos 5422 big. LITTLE system on chip (SoC) board has been chosen to validate our proposed approach, as shown in Figure 3(a). To activate the observing of power-performance counter, we prepared a custom system routine compatible with ARM's technical specification document capable of record diverse performance counter readings at pre-determined periodic spans. This routine, along with its libraries, is presently being intended for open release. In this work, performance counter is adopted to report system performance events including, instruction retired cache misses, and cycles. Further, it is used to monitor temperature, current, voltage, and power from the sensors in the Odroid-XU3 board by adopting the approach provided by Walker et al. [24]. In this work, the low activity class which leads to the significant dark silicon power consumption, as the supply voltage and clock remain operational, has been captured for different PCPG scenarios and frequencies.

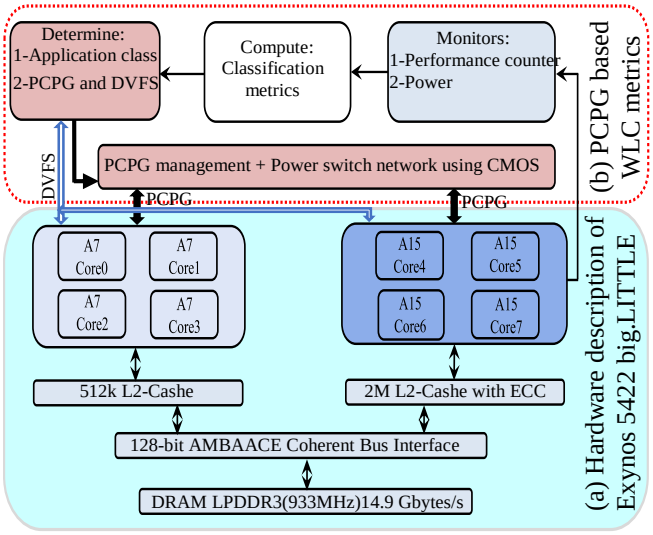

Fig. 3. (a) Hardware description of Samsung Odroid-XU3 platform; (b) Proposed PCPG based WLC metrics

The two observations below can be summarized as follows. Primarily, as the number of dark silicon related cores increase (big or LITTLE) the power consumption will drastically increase. For instance, the power consumption of 4 big idle cores at 2 GHz is about 1.5 Watt, which decreases to roughly 0.6 Watt when only one big core acts as dark silicon area. Secondly, the dark silicon related power consumption is also reliant on the clock frequency. As an example, when parallel threads of cannel application are assigned to LITTLE cores only, the dark silicon related power consumption of the idle four big cores increases from about 0.6 Watt at 1400 MHz to almost 1.5 Watt at 2 GHz.

B. Power gating management and PSN

The proposed runtime PCPG based WLC coupled with DVFS control is performed according to Algorithm I. This algorithm specifies the type of the workload, number of PCPG, and DVFS combinations for any exercised application. This can be achieved through comparing observed reading from performance counter to pre-set thresholds acquired from offline experiments at design time. Depending on extensive experiment from [8] [15], we categorize workloads by their processing demands (CPU) and communication intensive (memory). Therefore, workloads can be classified into three categories: memory-intensive (MI), CPU-intensive (CI), and mix of CPU and memory-intensive (MIX).

TABLE III.
Performance counter metrics, and PCPG control decision based on workload classification detail.

Metrics	Definitions	Metrics range	Classification	Freq.	A7	A15
nipc	*(InstRet/Cycles)(1/IPC$_{max}$)*	urr [0, 0.11]	0: low-activity	min	PCPG	PCPG
iprc	*InstRet/ClockRe f*	nnmipc [0.35, 1]	1: CPU-intensive	max	PCPG	max
nnmipc	*(1/IPC$_{max}$) (InstRet/Cycles − Mem/Cycles)*	nnmipc [0.25, 0.35)	2: CPU+memory	min	max	max
urr	*Cycles/ClockRe f*	nnmipc [0, 0.25)	3: mem-intensive	max	max	PCPG

ALGORITHM I: PCPG AND DVFS BASED WLC METRICS

Input: power, performance counter readings including (unhalted CPU cycles, memory access, instruction retired);
Constant: Parameters $urr_H = 0.11$, $nnmipc_H = 0.35$, $nnmipc_L = 0.25$
Output: WLC type, PCPG and DVFS
Compute: urr and nnmipc

If: $urr \leq urr_H$;
WL type = LA; ⇨ Class 0: Low-Activity
Allocated single LITTLE core;
DVFS f_{A7}=Min.;
Else if: $nnmipc > nnmipc_H$;
WLC type = CI; ⇨Class1: CPU-intensive
Allocated_cores A15 cores alone;
DVFS f_{A15}=Max.;
Else if: $nnmipc_L < nnmipc < nnmipc_H$;
WLC type = Mixed;⇨Class2: Combination
Allocated_cores big.LITTLE cores;
DVFS f_{A15}=Max. & f_{A7}= Max.;
Else if: $nnmipc < nnmipc_L$;
WLC type = MI; ⇨Class3: memory intensive
Allocated_cores A7 cores alone;
DVFS f_{A7}=Max.;
End if

A number of flag registers, on every time interval, are modified by the system software routine relying on the number of dark silicon cores. For instance, when two big cores (core 4 and core 5) are acted as dark silicon area the corresponding flag bits are assigned to 1 highlighting the advantages of power gating. These bits can then be adopted to activate the PSN based CMOS transistors, thereby shutting down those dark silicon related cores. In this work, PSN has been designed using Cadence Virtuoso tool box at 45nm technology node. Practically, the maximum current drawn when CPU-intensive workload is exercised on the A15 core has been reported to be Imax=1A per core at f=2000 MHz. Therefore, a number of switch transistor is connected in parallel at their maximum width to ensure providing maximum current of 1A to the active core [25]. As a result, the power consumption overhead causing by adopting the PSN has been computed using Cadence tool for fair comparison.

II. RESULTS

For verification purposes five different set of applications have been adopted in this work as a case study. As expected our PCPG based WLC achieves the highest energy efficiency (IPS/Watt) improvement when streamcluster application is exercised, compared to the ondemand governor and previous work published in [15]. This is because streamcluster application is a memory-intensive workload which prefer to be exercised at little cores with lower DVFS level. This indicating the advantages of switching off the big cores which causing the highest dark silicon related power consumption. These results illustrate about 110%

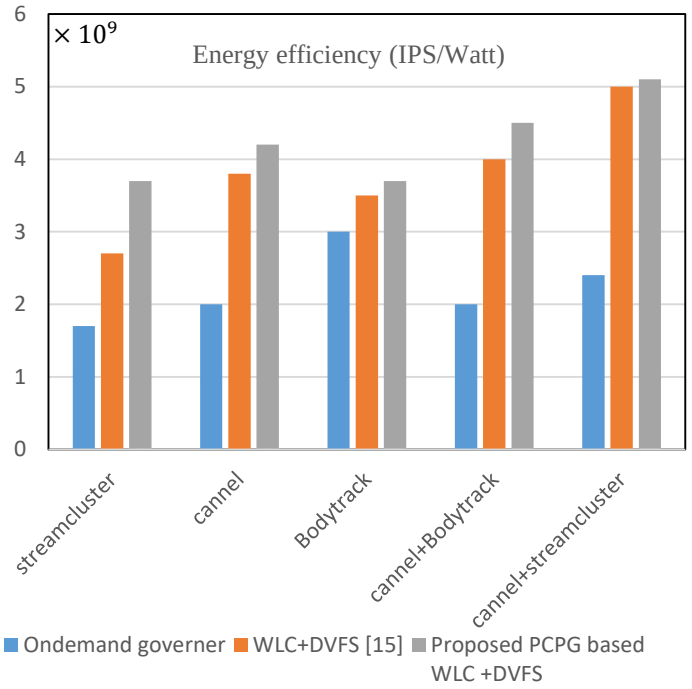

Fig. 4. IPS/Watt measured for the proposed PCPG based WLC, MLR+WLC, and only ondemand governor exercised on odroid XU3 platform.

improvements of energy-efficiency (IPS/Watt) over the ondemand governor, while improvement of 37% has been reported in comparison to WLC+MLR approach published in [15].
Exercising of cannel, CPU-intensive application, using our proposed approach can achieve 10 to 100% improvement over the WLC+MLR and ondemand governor, respectively. Mixed memory and CPU-intensive concurrent application shows slighter energy efficiency improvements of 12 to 2% for cannel+Bodytrack and cannel+streamcluster respectively. This is attributed to the fact that application behavior fluctuated from one class to another causing of high energy consumption in the power switch network which outweigh its energy saving at some interval of exercising the application. Therefore, our proposed approach can achieve significant energy saving on minor fluctuated application, which rarely move from one class to other.

VIII CONCLUSION

Emerging of heterogamous many-core platform offers promising solution for ever increasing demands of energy-efficient mobile computing system. Using power gating technique based workload classification coupled with core allocation and DVFS, these platforms can effectively minimize energy consumption and optimize resource utilization. Thereby, contributing to significantly mitigate dark silicon effect occurring due to an implementing of Moore's Law technology scaling. In this work, IPS/Watt has been improved 37 to 110% for memory intensive workload compared to published work in [15], and ondemand governor, respectively.
In conclusion, the combination of heterogeneous many-core platform and workload classification plays substantial role in

revolutionize the way we approach energy-efficient computing, making it a fundamental driver for a more sustainable and powerful future in the world of computing technology. For future work, PCPG based WLC can be implemented on GPU to drastically reduce energy consumption, hence significant energy saving can be improved.

REFERENCES

[1] H. Esmaeilzadeh, E. Blem, R. S. Amant, K. Sankaralingam and D. Burger, "Dark silicon and the end of multicore scaling," *2011 38th Annual International Symposium on Computer Architecture (ISCA)*, San Jose, CA, USA, 2011, pp. 365-376.

[2] Esmaeilzadeh, Hadi, et al. "Dark silicon and the end of multicore scaling." IEEE micro 32.3 (2012): 122-134.

[3] Shafaei Bejestan, Alireza, Yanzhi Wang, Srikanth Ramadurgam, Yuankun Xue, Paul Bogdan, and Massoud Pedram. "Analyzing the dark silicon phenomenon in a many-core chip multi-processor under deeply-scaled process technologies." *In Proceedings of the 25th edition on Great Lakes Symposium on VLSI, pp. 127-132. 2015.*

[4] Henkel, Jörg, et al. "Dark silicon: From computation to communication." *Proceedings of the 9th International Symposium on Networks*-on-Chip. 2015.

[5] X. Wang, A. K. Singh, B. Li, Y. Yang, H. Li and T. Mak, "Bubble Budgeting: Throughput Optimization for Dynamic Workloads by Exploiting Dark Cores in Many Core Systems," in *IEEE Transactions on Computers*, vol. 67, no. 2, pp. 178-192, 1 Feb. 2018.

[6] Wang, Xiaohang, et al. "A pareto-optimal runtime power budgeting scheme for many-core systems." *Microprocessors and Microsystems* 46 (2016): 136-148.

[7] Aalsaud, Ali, Ashur Rafiev, Fei Xia, Rishad Shafik, and Alex Yakovlev. "Model-free runtime management of concurrent workloads for energy-efficient many-core heterogeneous systems." *In 2018 28th International Symposium on Power and Timing Modeling, Optimization and Simulation (PATMOS)*, pp. 206-213. IEEE, 2018.

[8] Aalsaud, Ali, et al. "Power--aware performance adaptation of concurrent applications in heterogeneous many-core systems." *Proceedings of the 2016 International Symposium on Low Power Electronics and Design*. 2016.

[9] Mandal, Sumit K., et al. "An energy-aware online learning framework for resource management in heterogeneous platforms." *ACM Transactions on Design Automation of Electronic Systems* (TODAES) 25.3 (2020): 1-26.

[10] "Odroid XU3." http://www.hardkernel.com

[11] E. Musoll, "Hardware-based load balancing for massive multicore architectures implementing power gating," TCAD, vol. 29, no. 3, pp. 493–497, 2010

[12] J. N. Mistry, B. M. Al-Hashimi, D. Flynn, and S. Hill, "Sub-clock power gating technique for minimising leakage power during active mode," *in DATE*, pp. 1–6, March 2011

[13] J. Charles, P. Jassi, N. S. Ananth, A. Sadat, and A. Fedorova, "Evaluation of the intel R core i7 turbo boost feature," *in IISWC*, pp. 188–197, IEEE, 2009

[14] K. Ma and X. Wang, "Pgcapping: exploiting power gating for power capping and core lifetime balancing in cmps," *in PACT*, pp. 13–22, ACM, 2012.

[15] Aalsaud, A., Xia, F., Rafiev, A., Shafik, R., Romanovsky, A. and Yakovlev, A., 2020. Low-Complexity Run-time Management of Concurrent Workloads for Energy-Efficient Multi-Core Systems. *Journal of Low Power Electronics and Applications*, 10(3), p.25.

[16] Tzilis, S., Trancoso, P. and Sourdis, I., 2019. Energy-efficient runtime management of heterogeneous multicores using online projection. *ACM Transactions on Architecture and Code Optimization (TACO)*, 15(4), pp.1-26.

[17] Singh, A. K., Prakash, A., Basireddy, K. R., Merrett, G. V., & Al-Hashimi, B. M. (2017). Energy-efficient run-time mapping and thread partitioning of concurrent OpenCL applications on CPU-GPU MPSoCs. *ACM Transactions on Embedded Computing Systems (TECS)*, 16(5s), 1-22.

[18] Hankendi, Can, and Ayse K. Coskun. "Adaptive power and resource management techniques for multi-threaded workloads." *2013 IEEE International Symposium on Parallel & Distributed Processing*, Workshops and Phd Forum. IEEE, 2013.

[19] Shafik, R. A., Yang, S., Das, A., Maeda-Nunez, L. A., Merrett, G. V., & Al-Hashimi, B. M. (2015). Learning transfer-based adaptive energy minimization in embedded systems. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 35(6), 877-890.

[20] Das, A., Kumar, A., Veeravalli, B., Shafik, R., Merrett, G., & Al-Hashimi, B. (2015, March). Workload uncertainty characterization and adaptive frequency scaling for energy minimization of embedded systems. In 2015 *Design, Automation & Test in Europe Conference & Exhibition (DATE)* (pp. 43-48). IEEE.

[21] Reddy, B. K., Walker, M. J., Balsamo, D., Diestelhorst, S., Al-Hashimi, B. M., & Merrett, G. V. (2017, September). Empirical CPU power modelling and estimation in the gem5 simulator. *In 2017 27th International Symposium on Power and Timing Modeling, Optimization and Simulation (PATMOS)* (pp. 1-8). IEEE.

[22] Bienia, C., & Li, K. (2009, June). Parsec 2.0: A new benchmark suite for chip-multiprocessors. *In Proceedings of the 5th Annual Workshop on Modeling, Benchmarking and Simulation* (Vol. 2011, p. 37).

[23] Torrey, A.; Cleman, J.; Miller, P. Comparing interactive scheduling in Linux. Softw. Pract. Exp. 2007, 34, 347–364

[24] Walker, M.J.; Diestelhorst, S.; Hansson, A.; Das, A.K.; Yang, S.; Al-Hashimi, B.M.; Merrett, G.V. Accurateand stable run-time power modeling for mobile and embedded cpus. *IEEE Trans. Comput. Aided Des. Integr. Circuits Syst.* 2017, 36, 106–119

[25] B. Amelifard et al., "Optimal selection of voltage regulator modules in a power delivery network," *in DAC*, pp. 168–173, IEEE, 2007.

[26] H. A. Leftah, S. Boussakta, and S. Ikki, "Enhanced Alamouti space-time block-coding transmission based on a developed OFDM system," *in Proc. 9th Int. Symp. Commun. Syst., Netw. Digit. Sign*, 2014, pp. 591–595.