Meeting Thermal Safe Power in Fault-Tolerant Heterogeneous Embedded Systems

Mohsen Ansari, Mostafa Pasandideh, Javad Saber-Latibari, and Alireza Ejlali

***Abstract*— Due to the system-level power constraints, it is encountered that not all cores in a multicore chip can be simultaneously powered-on at the highest voltage/frequency levels. Also, in the future technology nodes, reliability issues due to the susceptibility of systems to transient faults should be considered in multicore platforms. Therefore, two major objectives in designing multicore embedded systems are low energy/power consumption and high reliability. This letter presents an energy management system that optimizes the energy consumption such that it satisfies reliability target and meets timing, Thermal Design Power (TDP) and Thermal Safe Power (TSP) constraints. Towards the TDP/TSP-constrained energy-reliability optimization, the proposed method schedules periodic real-time applications on different types of cores with voltage/frequency variations for heterogeneous multicore embedded systems. Experiments show that our proposed system provides up to 38.19% (in average by 29.66%) energy saving and up to 54.73% peak power reduction (in average by 24.55%) under different reliability targets and TDP/TSP constraints when compared to state-of-the-art techniques.**

***Index Terms* – Peak Power Consumption, Energy, Reliability, Thermal Design Power, Thermal Safe Power, Optimization.**

I. INTRODUCTION

On-chip systems due to continuing the scaling of feature size are thermally constrained [1][2][3][6][7]. Technology scaling allows more transistors to be integrated onto a multicore chip [2][4][5][12]. The chip-level power constraint, Thermal Design Power (TDP), is the highest sustainable power that a chip can dissipate to avoid performance throttling mechanisms [1][2]. However, TDP as the power constraint of a system can be very pessimistic, therefore, having better power budget is a major requirement towards dealing with performance losses [1][2]. A new power budget concept called Thermal Safe Power (TSP) provides safe and efficient power constraint. If the peak power consumption of each core violates its TSP, it automatically restarts or significantly reduces its performance to prevent permanent damage. TSP is computed in the offline phase for the worst-case scenarios, or unlike TDP in the online phase for a specific mapping of cores. When core heterogeneity or timing guarantees are involved, TSP can also guide task partitioning and mapping decisions. In order to meet the TDP/TSP constraints, some solutions like heat-sink and chip's cooling are proposed while due to their negative effects on the system reliability these solutions are not used in reliable embedded systems [2]. It should be noted that most of hard real-time systems are fan-less because fans are electro-mechanical components that in most cases have lower reliability characteristics compared to other components on a semiconductor-based system [16]. Therefore, peak power minimization is an efficient way to meet the power constraints and prevent the system from producing high temperature. Another limitation of embedded systems is that most of them are battery-based, and hence, the energy consumption of them should be reduced [4][13][14][15][19]. Energy is the integration of power consumption through time while the peak power consumption is instantaneous power consumption. Therefore, existing energy minimization schemes are unsuitable for peak power reduction and vice versa [8][9]. In order to prolong battery lifetime and meet the chip/core power constraints, energy minimization and peak power reduction are two major issues in modern embedded systems [2].

Meanwhile, in real-time embedded systems, reliability is another main design objective, and hence, the proposed system of this letter is subjected to different types of faults [2][4][14]. Multicore systems have an inherent redundancy that provides opportunities to implement various task replication techniques to tolerate transient and permanent faults [2]. In addition, violating the chip TDP and core TSP constraints degrades the system reliability because some cores may become reset or inactive [2]. Also, high temperatures may accelerate the occurrence of permanent faults in embedded systems [11]. Besides the temperature-dependent increase in soft errors [10], rapidly changing power levels may lead to transient faults due to the lower voltage level [11]. Recently, heterogeneous multicore systems provide an effective solution wherein every core can have an individual voltage but it is costly for implementation [3]. Due to the heterogeneity, the worst-case execution time and the energy/peak power consumption of tasks change according to the task-to-core mapping, presenting a new challenge for energy minimization and peak power reduction.

The purpose of this letter is to minimize energy consumption while keeping the peak power consumption below the power constraints and the system reliability at an acceptable level in heterogeneous multicore embedded systems without violating any timing constraints. In order to evaluate the effectiveness of the optimization method, we compared our scheme with three state-of-the-art techniques. The rest of this letter is formed as follows. In section II, we present our system model. In section III, we present the details of the problem and our solution. The experimental results are shown in section IV and we conclude the letter in section V.

Manuscript received May 23, 2019; accepted July 25, 2019. This work was supported by the Sharif University of Technology. This paper was recommended by Associate Editor D. Sciuto. (*Corresponding author: Alireza Ejlali*.)

The authors are with the Department of Computer Engineering, Sharif University of Technology, Tehran 14588, Iran (e-mails: {mansari; smpasandideh; jsaber}@ce.sharif.edu; ejlali@sharif.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier x.y/z

II. Models and Assumptions

A. Application, System, Power/Energy, and Fault Model

Task Model: In this letter, we consider a set of periodic hard real-time tasks $\psi=\{T_1,\ldots, T_n\}$, where each task T_i has a period π_i, a worst-case execution time wc_i. The j^{th} job of a task T_i (T_{ij}) arrives at time $t_{ij}=(j-1)\times\pi_i$ and must execute by its deadline $j\times\pi_i$. Also, the relative deadline D_{ij} of the job T_{ij} is equal to the period $j\times\pi_i$. Also, the worst-case execution time of the jobs for a task is equal to the worst-case execution time of that task. The utilization of the task T_i is defined as wc_i/π_i. Therefore, the sum of all tasks utilization is U_{tot}.

System, Power, and Energy Model: The system model is based on a heterogeneous multicores architecture with m cores consisting of two heterogeneous islands. These islands are: (*i*) High-Performance Island, (*ii*) Low Power Island, where each island has a number of homogeneous processing cores. Also, due to supporting Dynamic Voltage Scaling (DVS), each core may have a different voltage. The total power consumption of the system consists of static and dynamic power components [1][2][4][5]. Also, each core can operate in active and sleep modes. The core executes tasks in the active mode and in this mode Eq. 1 gives the power consumption of the system.

$$P_{total}(V_i, f_i) = P_{static} + P_{dynamic} = I_0 e^{\frac{-V_{th}}{\eta V_T}} V_i + \alpha C_L V_i^2 f_i \quad (1)$$

Under DVFS, let V_{max} be the maximum voltage corresponding to the maximum frequency f_{max}. Considering the almost linear relationship between voltage and frequency [2][4][20][25], we can write: $\rho_i=V_i/V_{max}=f_i/f_{max}$. Therefore, Eq. 1 can be rewritten as:

$$P_{total}(V, f) = \rho_i(I_0 e^{\frac{-V_{th}}{\eta V_T}} V_{\max}) + \rho_i^3(\alpha C_L V_{\max}^2 f_{\max}) = \rho_i.P_s^{\max} + \rho_i^3.P_d^{\max} \quad (2)$$

The energy consumption of the system is the sum of the energy consumption of all jobs of all tasks executed on different cores. The total energy consumption can be expressed as [4][5]:

$$E_{sys} = \sum_{k=1}^{m} \sum_{\forall T_{ij} \in core_m} \sum_{j=1}^{n_m} ((\rho_i.P_s^{\max} + \rho_i^3.P_d^{\max}).\frac{wc_{ij}}{\rho_i f_{\max}}) \quad (3)$$

Fault Model: We consider a transient fault model similar to [2][4][5]. The average fault rate λ is dependent on the core voltage so that decreasing core voltage, λ increases exponentially. The average fault rate on the voltage V can be expressed as:

$$\lambda(V) = \lambda_0 10^{\frac{V_{\max}-V}{d}} \quad (4)$$

where $\lambda_0=10^{-7}$ (faults per us) is the transient fault rate at V_{max} and d determines the sensitivity of the system to voltage scaling. Like the works [2][4][5], we consider d=2 in this letter. Therefore, the functional reliability of the job of a task can be written as [2][4]:

$$R(T_{ij}) = e^{-\lambda(V)\times wc_{ij}} \quad (5)$$

In the task replication technique, the execution of tasks will be unsuccessful only if all the replicas encounter transient faults during their executions. Therefore, the probability of failure φ and the reliability of a task T_i with k replicas is found as [5]:

$$\varphi(T_i) = (1-R_i)^k \quad (6)$$

$$R(T_i) = 1-(1-R_i)^k \quad (7)$$

III. Problem Definition and Our Solution

A. Concept Overview

In this letter, we consider a heterogeneous system that executes preemptive periodic hard real-time tasks. In this letter, we optimally minimize the system energy consumption in the offline phase that is subjected to reliability, timing, and the chip-level and core-level power constraints.

B. Problem Definition

Dertouzos in [26] has demonstrated that the EDF scheduling is the optimal solution in feasibility. However, EDF does not guarantee meeting TDP, TSP, reliability requirements and timing constraints simultaneously. In the heterogeneous multicore hard real-time systems, in addition to meeting all timing constraints, the system must satisfy the system reliability requirement and meet the chip-level and core-level power constraints [1][2]. Therefore, we use the following notation to present energy and peak power consumption, voltage and frequency level and task-to-core mapping. In this formulation, n is the number of tasks, s is the maximum number of jobs of tasks, m is the number of cores, and v is the number of available V-f levels for each core:

- The peak power consumption is represented by the matrix $P\epsilon\mathbb{R}^{n\times s\times m\times v}$, where each element $P_{i,j,k,l}$ denotes the power consumption for the job j of task i when the task is executed on the core k under the V-f level l.
- The task-to-core mapping and V-f level assignments are represented by the matrix $X\epsilon\{0,1\}^{n\times m\times v}$. The task i is mapped to the core k and is executed under the V-f level l if and only if $X_{i,k,l}$ = 1.

We formulate the above problem in the following.

Optimization Goal: Minimize the total energy consumption defined by the sum of the energy consumption of all jobs of all tasks.

$$\text{Minimize } E_{sys} = \sum_{k=1}^{m} \sum_{\forall T_{ij} \in core_m} \sum_{j=1}^{n_m} ((\rho_i.P_s^{\max} + \rho_i^3.P_d^{\max}).\frac{wc_{ij}}{\rho_i f_{\max}}) \quad (8)$$

Chip-Level Power Constraint: The instantaneous power consumption of the chip must be less than the chip TDP constraint. In the following equation, h is the least common multiple of all task periods called *hyperperiod*.

$$\begin{aligned} &\forall i,j,l: P_{total} = \sum_{k=1}^{m} P_{i,j,k,l} = \sum_{k=1}^{m} (\rho_i.P_s^{\max} + \rho_i^3.P_d^{\max}).X_{i,k,l}) \le P_{TDP,chip} \\ &\forall i,j,l: 1 \le i \le n + \sum_{a=1}^{n} nreplica(T_a), 1 \le j \le \frac{hyperperiod}{D_i}, 1 \le l \le m \end{aligned} \quad (9)$$

Core-Level Power Constraint: The peak power of each underlying core at each time slot t must be less than the core TSP constraint.

$$\begin{aligned} &\forall i,j,l: P_{i,j,k,l} = (\rho_i.P_s^{\max} + \rho_i^3.P_d^{\max}).X_{i,k,l} \le P_{TSP,core} \\ &\forall i,j,l: 1 \le i \le n + \sum_{a=1}^{n} nreplica(T_a), 1 \le j \le \frac{hyperperiod}{D_i}, 1 \le l \le m \end{aligned} \quad (10)$$

Tasks Timing Constraint: The worst-case execution time of each job $wc_{i,j}/f_{kl}$ on the core k and at the frequency level l should not exceed the task timing constraint (defined by the D_{ij}).

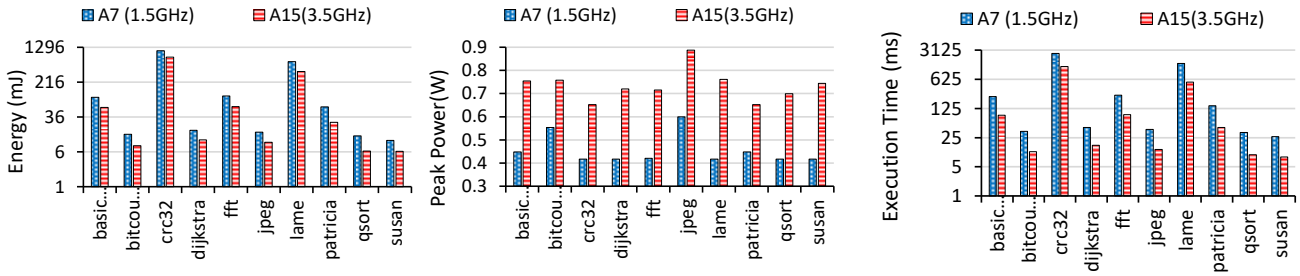

Fig. 1. Energy consumption, peak power consumption and execution time based on simulations in gem5 [23] and McPAT [24] and measured on ARM Cortex-A7 and Cortex-A15 [21] for applications from the MiBench benchmark suite [22].

$$\forall i,j,k,l:(j-1)\pi_i+\frac{wc_{i,j}}{\rho_i.f_{\max}}.X_{i,k,l}\le D_{ij}$$
$$\forall i,j:1\le i\le n+\sum_{a=1}^{n}nreplica(T_a),1\le j\le\frac{hyperperiod}{D_i} \quad (11)$$

Core Assignment Constraint: Each task can be only mapped to one core.

$$\forall l:\sum_i\sum_k X_{i,k,l}=1 \quad (12)$$

V-f Levels Assignment Constraint: Each task can be only executed under a single V-f level on a core (the V-f level does not change during the task execution).

$$\forall i,k:\sum_l X_{i,k,l}\le 1 \quad (13)$$

The number of replicas: In order to determine the number of replicas, there is a lower bound on the number of replicas required to achieve a certain reliability target. In other words, high-reliability levels necessitate the use of more replicas. Based on equations 6 and 7, we can determine the minimum number of replicas needed to achieve the reliability target at a given frequency level [5]:

$$\forall i:1\le i\le n:\varphi_{target}\ge(\varphi_{T_i})^{nreplica(T_i)}$$
$$\forall i:1\le i\le n:nreplica(T_i)\ge\left\lceil\frac{\log(\varphi_{target})}{\log(1-e^{wc_{ij}.(-\lambda_0).10^{\frac{V_{max}(1-\rho_i)}{d}}})}\right\rceil \quad (14)$$

In the task replication technique, it is sufficient to have at least one task copy execution that passes the acceptance test [17]. Also, the task replication technique requires a fault detection method. For this purpose, our processing cores typically employ a low-cost hardware checker like Argus [18].

The formulated problem is a convex problem that can be solved by the available convex solvers, and it is categorized as an NP-hard problem [1][2][4][5]. On the other hand, the complexity of such problems may increase exponentially with the increase of problem size, e.g., with the number of ready tasks, islands, cores, and V-f levels. In order to solve the problem, we use the Yalmip solver [27] in MATLAB. In the next section, we show the results of our simulations.

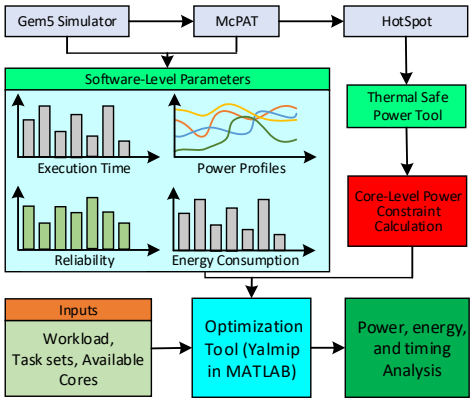

Fig. 2. Our tool flow for power, energy, and timing evaluation.

IV. Results and Discussion

In this section, we evaluate the effectiveness of our optimal solution via simulation with various task sets including real-life embedded applications of MiBench Benchmark suite [22] running on a target heterogeneous multicore chip. Fig. 2 shows our tool flow and simulation setup for power, energy, and timing evaluation. Our evaluation consists of the comparison between our optimal solution and three state-of-the-art schemes. We compared our optimal solution with [5]-EM, TMR [4], and [2]-PPM schemes. In order to evaluate our optimal solution, for each data point, we generated 100 task sets and the average results are reported. Each task set consists of 10 to 100 tasks based on different utilization targets. In our evaluations, the accuracy of the results is higher than 99.99%. The task sets are selected randomly from Fig. 1. We exploited gem5 full-system simulator [23] and McPAT [24] to conduct this figure.

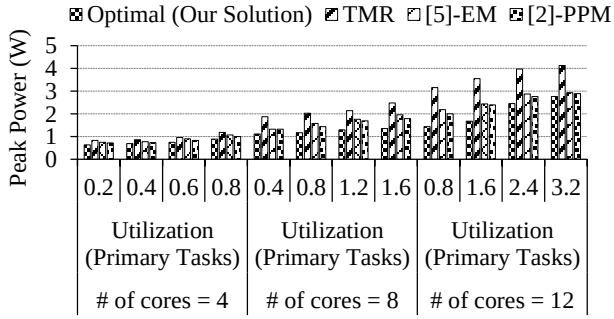

Fig. 3. Peak power consumption in different system utilizations.

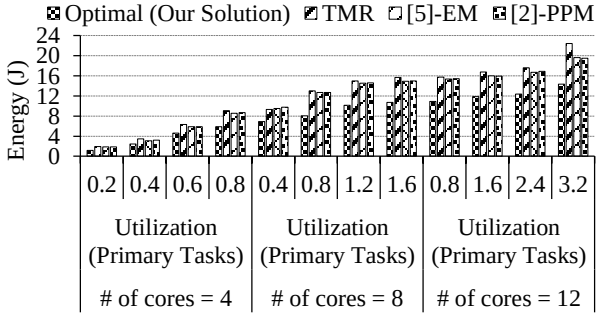

Fig. 4. Energy consumption in different system utilizations.

We evaluated the ratio of peak power reduction and energy saving of our optimal solution versus [5]-EM, TMR [4], and [2]-PPM for the different number of cores (M=4, 8 and 12) and different core workloads. Fig. 3 shows the results of peak power consumption for the cases when tasks are generated from applications of Fig. 1. What can be inferred from this figure is that our optimal solution completely outperforms the other three schemes for all systems configurations. Each case of Fig. 3 was simulated for 1000 times with different parameters of the applications and the average results are reported, i.e. accuracy is 99.95%. This figure shows that our scheme provides up to 54.73% (on average by 24.55%) peak power reduction compared to three state-of-the-art techniques. From Fig. 3 it can be concluded that in all utilization points, the peak power reduction of our optimal solution is higher than other schemes. Also, in Fig. 4, for all utilization, points our optimal solution can save more energy, compared [5]-EM, TMR [4], and [2]-PPM. The main reason is that another scheme is forced to employ a heuristic method to save more energy, while our optimal solution minimizes energy consumption. On the whole, by increasing the utilization, the energy saving of our optimal solution compared other scheme decreases because when utilization is high, less slack time can be achieved. Experiments show that our optimal solution provides up to 38.19% (on average by 29.66%) energy savings compared to three state-of-the-art techniques. Also, it can be seen from Fig. 4 that when the utilization of the cores increases, the energy saving decreases because the amount of static and dynamic slack times decreases, and hence we cannot achieve significant energy savings.

V. Conclusions

In this letter, we have addressed two main issues which are low power consumption and high reliability in heterogeneous multicore embedded systems. In order to achieve these objectives, we minimize energy consumption while keeping the peak power consumption below the chip-level and core-level power constraints and the system reliability at an acceptable level. Experiments show that our proposed system provides up to 38.19% (on average by 29.66%) energy saving when compared to three state-of-the-art techniques.

References

[1] S. Pagani, et. al, "Thermal Safe Power (TSP): Efficient Power Budgeting for Heterogeneous Manycore Systems in Dark Silicon," *IEEE Trans. on Comp.*, vol. 66, no. 1, pp. 147-162, 2017.

[2] M. Ansari, et. al, "Peak Power Management to Meet Thermal Design Power in Fault-Tolerant Embedded Systems," *IEEE Trans. on Par. and Dis. Sys.*, vol. 30, no. 1, 2019.

[3] J. Lee, B. Yun and K. G. Shin, "Reducing Peak Power Consumption in Multi-Core Systems without Violating Real-Time Constraints," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 4, pp. 1024-1033, 2014.

[4] M. Salehi, A. Ejlali, and B.M. Al-Hashimi, "Two-Phase Low-Energy N-Modular Redundancy for Hard Real-Time Multi-Core Systems," *IEEE Trans. on Parall. and Distr. Sys. (TPDS)*, vol. 25, no. 4, pp. 1024-1033, 2016.

[5] M. A. Haque, H. Aydin and D. Zhu, "On Reliability Management of Energy-Aware Real-Time Systems Through Task Replication*," IEEE Trans. on Parall. and Dis. Sys.*, vol. 28, no. 3, pp. 813-825, 2017.

[6] M. Shafique, D. Gnad, S. Garg, J. Henkel, "Variability-Aware Dark Silicon Management in On- Chip Many-Core Systems", *IEEE/ACM 18th Design, Automation and Test in Europe Conference (DATE)*, Mar. 2015.

[7] M. Ansari, S. Safari, F. R. Poursafaei, M. Salehi, A. Ejlali "AdDQ: Low-Energy Hardware Replication for Real-Time Systems through Adaptive Dual Queue Scheduling," *The CSI Journal on Computer Science and Engineering (JCSE)*, vol. 15, no. 1, pp. 31-38, 2017.

[8] H. Khdr, et. al, "Power Density-Aware Resource Management for Heterogeneous Tiled Multicores," *IEEE Transactions on Computers*, vol. 66, no. 3, pp. 488-501, 2017.

[9] S. Pagani, J. J. Chen and J. Henkel, "Energy and Peak Power Efficiency Analysis for the Single Voltage Approximation (SVA) Scheme," in *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 34, no. 9, pp. 1415-1428, 2015.

[10] M. Shafique, S. Garg, D. Marculescu, J. Henkel, "The EDA Challenges in the Dark Silicon Era: Temperature, Reliability, and Variability Perspectives", *ACM/EDAC/IEEE 51st Design Automation Conference (DAC)*, 2014.

[11] D. Brooks, R.P. Dick, R. Joseph, and L. Shang, "Power, Thermal, and Reliability Modeling in Nanometer-Scale Microprocessors," In *IEEE Micro*, pp.49-62, 2007.

[12] M. Shafique, S. Garg, "Computing in the Dark Silicon Era: Current Trends and Research Challenges", in *IEEE Design & Test (DnT)*, vol. 34, no. 2, pp. 5-7, 2017.

[13] B. Safaei, A. A. M. Salehi, A. M. H. Monazzah, A. Ejlali, "Effects of RPL Objective Functions on the Primitive Characteristics of Mobile and Static IoT Infrastructures," *Microprocessors and Microsystems* (MICPRO), 2019.

[14] M. Ansari, A. Yeganeh-Khaksar, S. Safari, and A. Ejlali, "Peak-Power-Aware Energy Management for Periodic Real-Time Applications," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 2019.

[15] S. Aminzadeh, and A. Ejlali, "A Comparative Study of System-Level Energy-Management Methods for Fault-Tolerant Hard Real-Time Systems," *IEEE Trans. on Comp.*, vol. 60, no. 9, pp. 1288-1299, 2011.

[16] M. Salehi, and A. Ejlali, "A Hardware Platform for Evaluating Low-Energy Multicore Embedded Systems Based on COTS Devices," *IEEE Trans. on Industrial Electronics*, vol. 62, no. 2, pp. 1262-1269, 2015.

[17] S. Safari, M. Ansari, G. Ershadi and S. Hessabi, "On the Scheduling of Energy-Aware Fault-Tolerant Mixed-Criticality Multicore Systems with Service Guarantee Exploration," in *IEEE Transactions on Parallel and Distributed Systems*, 2019.

[18] A. Meixner, M. E. Bauer and D. Sorin, "Argus: Low-Cost, Comprehensive Error Detection in Simple Cores," *40th Annual IEEE/ACM International Symposium on Microarchitecture (MICRO)*, Chicago, IL, pp. 210-222, 2007.

[19] M. Khavari Tavana, M. Salehi, and A. Ejlali, "Feedback-Based Energy Management in a Standby-Sparing Scheme for Hard Real-Time Systems," in *Proc. of the 32nd IEEE Real-Time Systems Symposium, RTSS*, Vienna, 2011.

[20] F. R. Poursafaei, et. al, "Offline Replication and Online Energy Management for Hard Real-Time Multicore Systems," *RTEST*, Tehran, Iran, October, 2015.

[21] P. Greenhalgh, "Big.LITTLE processing with ARM Cortex-A15 & Cortex-A7," ARM Limited, White Paper, September 2011.

[22] M.R. Guthaus, J.S. Ringenberg, D. Ernst, T.M. Austin, T. Mudge, and R.B. Brown, "MiBench: A Free, Commercially Representative Embedded Benchmark Suite," *Proc. Fourth IEEE Ann. Workshop on Workload Characterization*, pp. 3-14, 2001.

[23] N. Binkert, et. al, "The gem5 simulator," *ACM SIGARCH Computer Architecture News*,vol. 39, no. 2, pp. 1–7, May 2011.

[24] S. Li, at. al, "McPAT: An integrated power, area, and timing modeling framework for multicore and manycore architectures," in *MICRO*, pp. 469–480, 2009.

[25] S. Safari, M. Ansari, M. Salehi, and A. Ejlali, "Energy-Budget-Aware Reliability Management in Multi-Core Embedded Systems with Hybrid Energy Source," *The CSI Journal on Computer Science and Engineering (JCSE)*, vol. 15, no. 2, pp. 31-43, 2018.

[26] M. L. Dertouzos. "Control robotics: the procedural control of physical processes," *Information Processing*, 74, 1974.

[27] J. Lofberg, "YALMIP: a toolbox for modeling and optimization in MATLAB," *IEEE International Conference on Robotics and Automation (IEEE Cat. No.04CH37508)*, New Orleans, LA, pp. 284-289, 2004.