Reusable System-level Power-aware IP Modeling Approach *

Antonio Genov[1,2], François Verdier[1], and Loic Leconte[2]

[1] University of Cote d'Azur, CNRS, LEAT, 06903 Sophia-Antipolis, France
firstname.lastname@univ-cotedazur.fr
[2] NXP SEMICONDUCTORS, 06250 Mougins, France
firstname.lastname@nxp.com

Abstract. In many areas of the semiconductor industry, System-on-Chip (SoC) power analysis and management is a critical process for maintaining low-power profiles. Due to the complexity of modern SoC designs and the definition of their power management strategies, it is common to reuse available resources such as Intellectual Properties (IP) and power intent descriptions in the Unified Power Format (UPF). However, this is typically initiated at the Register Transfer Level (RTL) and beyond. In this study, we provide a new, earlier approach to enable system-level power intent definition and power management strategy description around IPs or entire SoCs. This approach follows UPF semantics and allows for the creation of reusable, power-aware IP models. In addition, it adds a description of clock intent into this model and simplifies the separation of designs into clock domains and the application of dynamic power reduction techniques. The approach is built around the PwClkARCH power modeling and estimation tool and its capabilities are demonstrated using an example of reusing power-aware SystemC-TLM IP models from NXP's i.MX8 SoC series.

Keywords - System-On-Chip (SoC); Power modeling; SystemC-TLM; Power estimation; Interoperability

1 Introduction

1.1 Context

The most modern areas of development, such as 5G, healthcare devices, and automotive, rely on advanced technologies, high performance and low power designs. We are increasingly reliant on battery-powered devices and reducing energy consumption in general. Designing energy-efficient devices is therefore becoming essential. In order to keep up with the pace of technological evolution and the timeframes set by customer needs, multiple important constraints and limitations must be reevaluated and resolved.

* University of Cote d'Azur, CNRS, LEAT, 06903 Sophia-Antipolis, France
NXP SEMICONDUCTORS, 06250 Mougins, France

Most major semiconductor manufacturers are targeting the latest technology nodes and design flow methodologies for their designs. As a result, with the steady decrease in transistor size, the number of Intellectual Properties (IP) in a System-on-Chip (SoC) continues to increase with each new technology node. This technological advancement improves chip performance, but exponentially increases the complexity of the designs and especially the management of their power consumption. This trend requires the development of new design methodologies to avoid severely affecting the energy efficiency, cost and development schedule of SoCs.

Several proven design techniques, such as IP reuse and Platform-based design [1], used from the Register Transfer Level (RTL), have been slightly extended to the Electronic System Level (ESL)[2] to simplify and accelerate testing and architecture definition. In addition, the IEEE 1801 Unified Power Format (UPF) reference standard [3] for specifying power intent has also been slightly extended to a higher level of abstraction (before RTL) to address the increasing complexity of power management in SoCs earlier in the flow. A new IEEE standard for power modeling, called IEEE 2416 or Unified Power Model (UPM) [4], was released to complement UPF and define a common methodology for defining and refining power during the design process with ESL as the starting point. The junction between the UPF and UPM standards provides power models with IP models that can be parameterized by the end user and estimate power consumption using event files or specific UPM APIs with specific UPF tools. However, these standards are aimed at power data representation (UPM) and dividing the SoC or IPs into power intents (UPF). The power management units (PMUs) or power reduction strategies must be developed by the SoC/IP user, which is also an effortful and not very efficient task at the system level.

We have created a realistic proof-of-concept example for the cross-platform electronic system level power-aware IP reuse methodology as well as a detailed power analysis. With this example, we demonstrate the portability and reusability of power models defined with PwClkARCH [8] tool. In addition, we highlight the importance of modeling power reduction techniques during early power analysis to obtain realistic power estimation results.

Our demonstration consists of the reuse and integration of a complex interconnect and traffic scheduling subsystem on three different generations of NXP i.MX8 SoC platforms. We reuse a previously verified IP performance model and its power model [9] to verify the accuracy of the power estimation after integration into another existing platform. Then, we reuse it on a next-generation NXP SoC and extract power estimates during the very early stages of SoC development.

In Section 1.2, we will examine a well-known power-related design constraint. In Sections 2 and 3, we will present our approach, its added value and some additional improvements. Finally, in Section 4, we will present our conclusion on this topic.

1.2 The Power in the Power, Performance, Area trade-off

Devices autonomy and power consumption are becoming more and more critical and are starting to really take their place in the PPA trade-off. Nowadays, this is a major design constraint that can cause a number of flaws due to limited power budget, glitches, or sub-optimal power intentions caught too late in the design flow. The new release of Wilson Research Group and Mentor Functional Verification Study [5] shows that it is still the third major flaw contributing to Application-Specific Integrated Circuits (ASIC) design respin, despite the positive trend compared to previous years (Fig. 1).

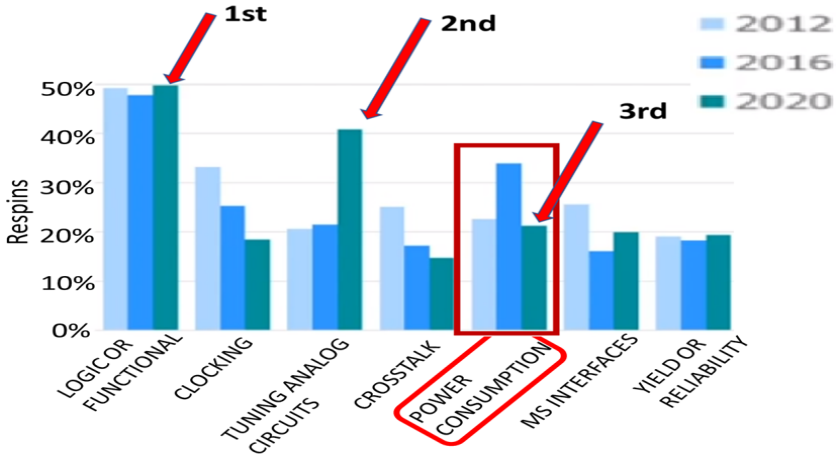

Fig. 1. Type of ASIC Flaws contributing to Respin [5]

This positive trend comes from the significant implication of IC engineers in the design and development of energy-efficient systems and the improved understanding and application of the existing power reduction techniques [6]. We can compare the percentage of power-aware design projects in the two most recent releases [5] [7] and conclude on the current involvement in the low-power design. From 2014 to 2018, there is no major changes in this percentage, but only from 2018 to 2020 we have a 10% increase in favor of power-aware designs.

However, due to the increasing complexity and performance of these power-aware ASIC and SoC designs, we are starting to hit the limits of the current design flow capabilities, especially on the power analysis side. Our believe is that there is a lot of room for improvement in studying power consumption during the design flow. Thus, following the shift-left tendency, we studied some additional hypothetical steps by raising the level of abstraction. The biggest gap here is at the very beginning of the flow. In the early stages of architecture definition, we only have a rough idea of what the energy consumption might be. At this stage, the energy consumption can be optimized through careful selection

of components and power management strategy, but this is not obvious without simulation-based exploration. In addition, due to the large number of IPs in modern SoCs, power-aware models reuse becomes an inevitable key element in the development of new architectures.

2 Methodology & propositions

2.1 PwClkARCH

PwClkARCH [12] [13] is a C++/SystemC-TLM library developed for ESL power estimation using UPF semantics abstracted to TLM level (Figure 2).

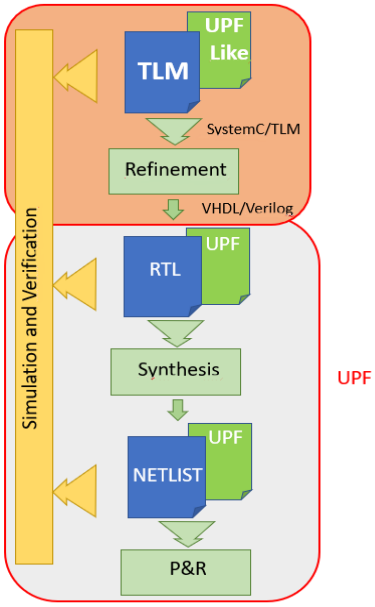

Fig. 2. Extending UPF to System-level

The details of the low-level power description are neglected in order to allow power modeling at a very early stage. An automated generic assertion verification mechanism is integrated in the library in order to ensure the consistency between the behavioral and power models. The library includes components such as Design Elements (DEs), Power State Table (PST), power switches and supply nets. Since there is no current standard for the description of the clock domains, PwClkARCH introduces a methodology that allows for the insertion of external clocks, generated clocks, DPLLs, Clock Managers (CMs) and a Clock State Table (CST) similar to the UPF PST. This clock domains description approach is inspired by the current most widely used clock distribution approaches.

This library is not IP-specific and has been tested on several industrialized IPs for different power observations. It has been used in an Intel's pre-silicon simulation environment for clock gating strategy definition on L2 Copro - hardware accelerators [8]. It has been used for DRAM energy consumption estimates with the DVFS application in a co-simulation with the DRAMPower tool [10] [8] [11]. It has also been used for power correlation between simulations and silicon measurements of an industrialized complex interconnect module [9] [11].

This study aims to further demonstrate the reusability of power models attached to reusable IPs and to clarify the implementation of this approach. We propose a solution for extending reusable transactional level IP models with an power intent and power management descriptions. In this approach, we attach to an IP a UPF-like power intent description, a similar clock intent controlled by a CM, and a power management strategy automatically and dynamically controlled by a PMU module (Figure 3). The power and clock states can be captured in PST and CST tables. Then, using an Operating Performance Points Table (OPPT), we described the legal combinations of power and clock states.

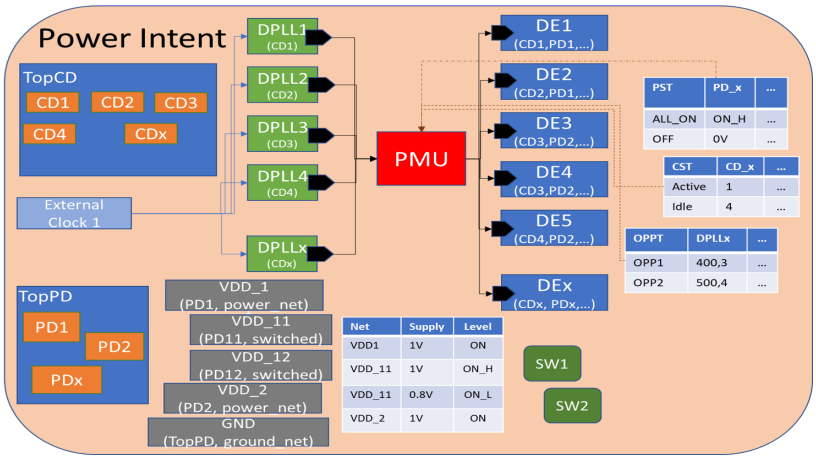

Fig. 3. PwClkARCH - Power/Clock intent structure

One or more master modules in the design can change the OPP points based on their functional state. These master modules are only responsible for initiating the OPP transitions and the PMU is responsible for automatically changing the states of all internal components based on the new OPPs [9].

As can be seen in the description of the PwClkARCH structure, much of the UPF semantics are reused here. Thus, PwClkARCH is also able to take a file in UPF format (including additional clock domain description commands like *create_clock_domain*, *create_supply_clock* and *create_dpll*) and automatically generate the C++/SystemC-TLM code for the power model.

2.2 Separation of concerns

The power model and the IP functional model can be considered as two independent SystemC models, which means that they can be easily reused, modified and tested with different power intents or power management stategies and options. In addition, they can be reused and easily re-adapted on multiple platforms with similar structures.

Similarly to UPF, PwClkARCH allows us to maintain the separation between the functional model and the power intent (Figure 4). This means that there is no intrusive power related code in the functional model, allowing it to be reused for performance estimation alone and making it more readable and efficient. This also allows parallel simulation between the functional/performance model and the power model and to easily study multiple power/clock intent structures and power management strategies for the same functional model. Thus *"what-if"* analyses are largely facilitated with this approach. We can observe the influence of each power management strategy on the performance of the functional model and easily detect bugs, such as poor power management and possible data loss. The PwClkARCH library is able to continuously monitor the power consumption of all IPs in any architecture running any application. The relevance of this approach can only be validated if the power consumption models are sufficiently accurate with respect to the real system and if the performance/power consumption relationship is also sufficiently accurate.

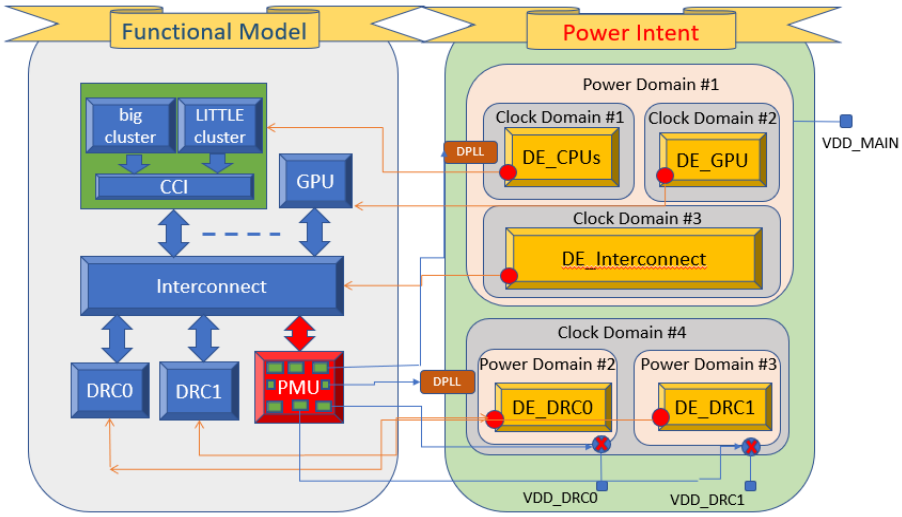

Fig. 4. PwClkARCH semantics [9]

The deliverable IP can be a configurable and pre-compiled combination of a functional model and several power intent models, defining possible power man-

agement strategies. The end user can choose the power intent from the available power descriptions to obtain the optimal solution for the given application. If we reuse a functional model (similar structure), but with different or additional features, we can again reuse the same power intents separately from the previous precompiled combination. We can also create configurable and reusable functional and power IP models that can be easily reprogrammed using a configuration file.

3 Experimental results

In this example, a power-aware Switch Matrix (SM) model is reused on three different NXP i.MX8 SoC platforms: i.MX8QM, i.MX8QXP and i.MX8nextGen.

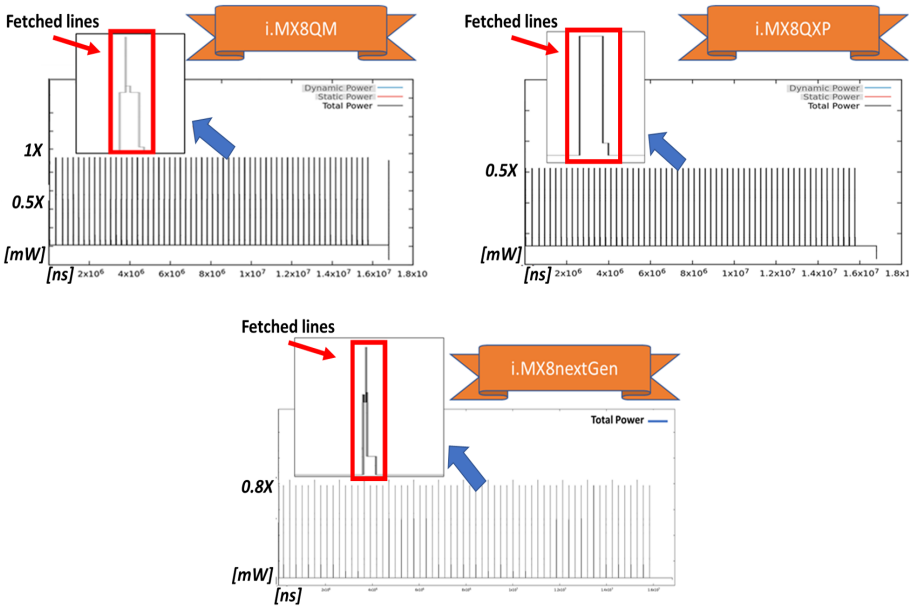

Fig. 5. i.MX8QM, i.MX8QXP and i.MX8nextGen - QVGA Display Refresh DRAM+Switch Matrix Total Power consumption. *All power-related axis scales had to be masked for confidentiality reasons. Let us consider the fixed value "X" [mW] as the most significant power consumption between the 3 platforms presented here.*

Once we had the power-aware model of the SM integrated in the i.MX8QM, it took less than 2 working days (one man effort) to complete the full integration (+ verification) into the i.MX8QXP platform and to simulate the QVGA, VGA and HD1080 display refresh and 256KB CPU memory copy use cases considered for silicon power consumption correlation. In the final step, we configured the reusable IP and integrated it on a higher performance and more complex next

generation i.MX8 platform. The integration and simulation was very intuitive and took less than a week (one-person effort), which is slightly longer than the first case due to the use and verification of additional logic and algorithms embodied in the IP model, but not used for the previous platforms.

In Fig. 5, we illustrate some of the results obtained after running a QVGA display refresh use case on the three platforms containing the reusable, power-aware SM subsystem model. The total power shown here contains the SM and 2 DRAM Controllers+LPDDR4 (DRC) in the i.MX8QM and i.MX8nextGen cases and only 1 DRC+LPDDR4 in the i.MX8QXP case. On the total power we have a periodic traffic generation corresponding to the different lines extracted from the memory.

To correlate the results of this simulation with silicon measurements, we extracted the power measurements using Baylibre ACME+Beaglebone and a maxTC temperature controller to maintain the ambient temperature (Figure 6).

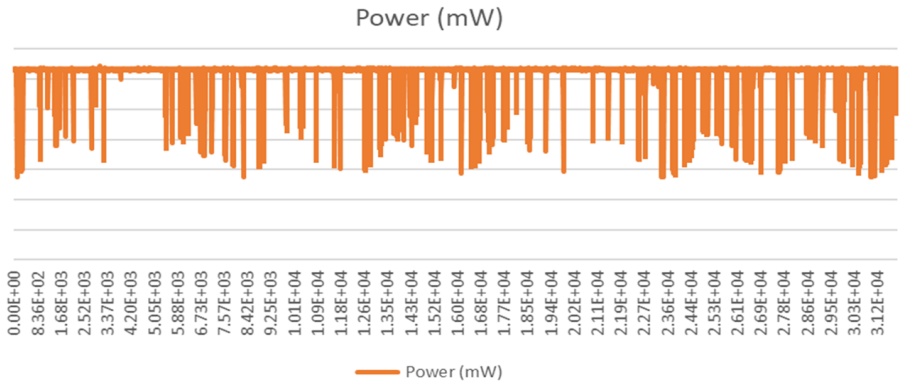

Fig. 6. i.MX8QM measurements - Display Refresh DRAM+Switch Matrix Total Power consumption. *All power-related axis scales had to be masked for confidentiality reasons.*

The power correlation results of i.MX8QM and i.MX8QXP show an accuracy of about 90-95%. In addition, at present, the silicon power measurement tools available without significant investment have relatively low resolution. Their results are usually an average value for a period of time/number of samples. For example, the resolution of our measurement tools is almost equal to the time needed to execute a display refresh. Thus, at the end of the measurement, the value we observe is the average value of all the power consumption variations during the retrieval of an image. With high-level simulation tools like PwClkARCH, we can extract a more detailed power consumption profile than with these measurement tools and observe its behavior closely. Therefore, due to the lower resolution, we compared the average values extracted from the HD display refresh simulations and the power measurements (for a short period of time).

Using this approach we were able to reuse an power-aware IP SystemC/TLM2.0 model on three different platforms, one of them is at its specification/product definition phase. Moreover, we were able to compare the power consumption with and without power management strategies (Figure 7) and to compare the energy consumed by each one of the three platforms (Figure 8).

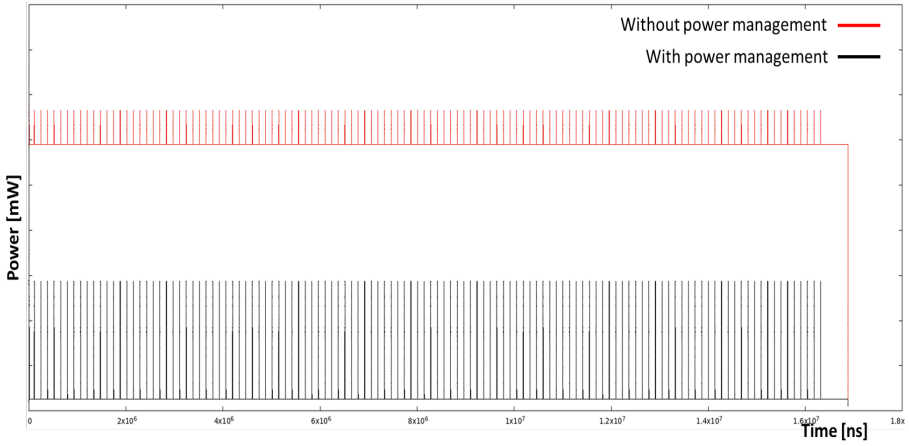

Fig. 7. i.MX8QM - Display refresh VGA use case: Total power with and without power management [mW]

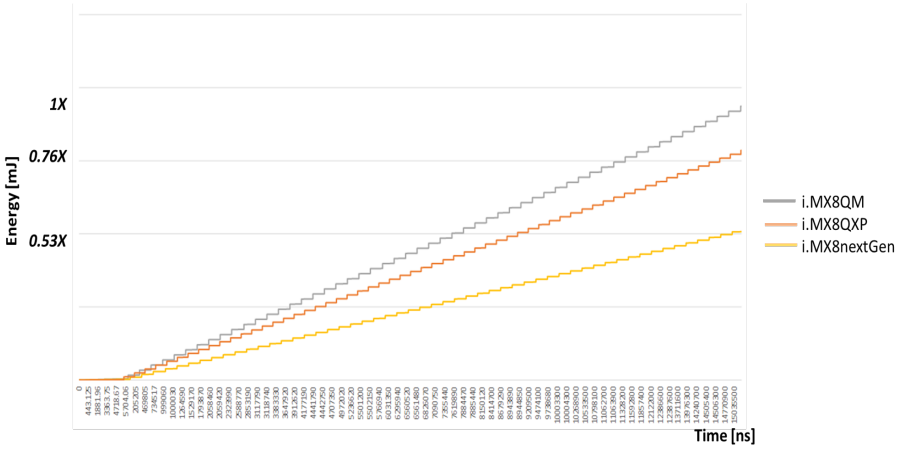

Fig. 8. Display refresh VGA use case: Energy comparison [mW]

4 Conclusion

This easy reuse of functional and/or power models demonstrated the usefulness of the separation of concerns semantics applied in our approach. We were able to simulate the models with and without power optimization, correlate the results, and observe the differences in power consumption between three generations of i.MX8 SoCs. All this at the system level in less than 2 weeks (after the first model was available). *The PwClkARCH library is not open source. If you are interested to use it, please contact one of the authors of the article.*

References

1. Keating M. and Bricaud P., *Reuse Methodology Manual: For System-On-Chip Designs*, Kluwer Academic Publisher, 2002.
2. Omer Cheema M., Lacassagne L. and Hammami O., *System-Platforms-Based SystemC TLM Design of Image Processing Chains for Embedded Applications*, EURASIP Journal on Embedded Systems, p. 14, 2007.
3. IEEE Std 1801-2018, *IEEE Standard for Design and Verification of Low-Power, Energy-Aware Electronic Systems*, no. 10.1109/IEEESTD.2019.8686430., 2019.
4. IEEE Std 2416-2019, *IEEE Standard for Power Modeling to Enable System-Level Analysis*, no. 10.1109/IEEESTD.2019.8782907., pp.1-63, 2019.
5. Foster H., *2020 Functional Verification Study, © Wilson Research Group and Mentor, A Siemens Business*, 2020.
6. dos Santos L., *Low Power Techniques for SoC Design: basic concepts and techniques*, Embedded Systems - Federal University of Santa Catarina, 2014.
7. Foster H., *2018 Functional Verification Study, © Wilson Research Group and Mentor, A Siemens Business*, 2018.
8. Ameur A., *Transactional simulation approach for modelling performance and energy of a heterogeneous SoC memory system*, Université Côte d'Azur, 2019. ⟨NNT : 2019AZUR4048⟩. ⟨tel-02570752⟩
9. Genov A., Leconte L. and Verdier F., *Mixed Electronic System-Level Power/Performance Estimation using SystemC/TLM2.0 Modeling and PwClkARCH Library*, in DVCon Europe, Virtual, 2020.
10. Chandrasekar K., Weis C., Li Y., Goossens S., Jung M., Naji O., Akesson B., Wehn N. and Goossens K., *DRAMPower: Open-source DRAM Power & Energy Estimation Tool*, [Online]. Available: http://www.drampower.info.
11. Genov A., Ameur A., Verdier F. and Leconte L., *Timing-Aware high level power estimation of industrial interconnect module*, in DVCon Europe, Virtual, 2020.
12. Affes H., Auguin M., Verdier F. and Pegatoquet A., *A methodology for inserting clock-management strategies in transaction-level models of system-on-chip*, in Forum on Specification and Design Languages (FDL), Spain, Sept. 2015.
13. Mbarek O., Pegatoquet A., Auguin M., *A Methodology for Power-Aware Transaction-Level Models of Systems-on-Chip Using UPF Standard Concepts*. In: Integrated Circuit and System Design. Power and Timing Modeling, Optimization, and Simulation (PATMOS) Lecture Notes in Computer Science, vol 6951. Springer, Berlin, 2011
14. Mbarek O., Pegatoquet A., Auguin M., *Using unified power format standard concepts for power-aware design and verification of systems-on-chip at transaction level*. IET Circuits, Devices & Systems, Institution of Engineering and Technology, pp.287-296, 2012