Reliability Optimization Under Severe Uncertainty for NoC Based Architectures Using an Info-Gap Decision Approach

Wenkai Guan, Jinhua Zhang and Cristinel Ababei
Department of Electrical and Computer Engineering
Marquette University
Email: {wenkai.guan, jinhua.zhang, cristinel.ababei}@marquette.edu

***Abstract*—Increased uncertainties in design parameters undermine the accuracy of the mapping of embedded applications to Network-on-Chip (NoC) based manycore architectures. In this paper, we attempt for the first time to apply the info-gap theory to uncertainty modeling in the context of embedded systems design. We first propose a novel info-gap based uncertainty-aware reliability model for NoC based manycore platforms. We then develop an uncertainty-aware solution to the problem of mapping in embedded systems. The solution is implemented as a computer program that can generate robust Pareto frontiers. Simulation results indicate that the proposed info-gap based uncertainty-aware mapping generates Pareto frontiers that have significant differences from the ones obtained with a traditional deterministic approach. Identifying and quantifying these differences is an important first step towards the development of better mapping optimization processes in order to arrive to optimal rather than suboptimal solutions.**

I. Introduction

The increasingly large process, voltage, and temperature variations negatively affect the design and optimization process of the embedded systems. In order to address these uncertainties in design parameters, previous studies [1], [2] attempted to capture uncertainty in the process of optimization of embedded systems. However, system reliability was considered as the only uncertain design parameter. The work in [3] is the first attempt to address the problem of multi-objective (reliability, performance, and energy) mapping for embedded systems under uncertainties. The study in [4] extends the work in [3] by modeling multiple correlated sources of uncertainty. However, these studies focused on hardware architectures formed by multiple processing elements (PEs) interconnected via a bus-based communication only, and Network-on-Chip (NoC) based architectures were not considered.

Our recent work [5] is the first attempt to quantify the impact of uncertainty in embedded systems mapping for NoC based architectures. The proposed design flow in [5] constructed of Monte Carlo simulation and evolutionary algorithms, which is an iterative process of exploration, learning, and evaluation. This decision process can handle uncertainty through *probabilistic* models that capture uncertainty in the form of mixtures of probability distribution functions, uniform value intervals, and fixed deterministic parameters. However, what if one did not have access to such probabilistic mixture models? This situation occurs when one does not have access to concrete RTL implementations of different cores, or their characterization is too difficult and time-consuming, or because one wants to explore versions of such cores implemented in future technology nodes whose parameters are largely unknown, notwithstanding efforts such as Predictive Technology Model (PTM) [6] to make predictions about many such parameters. These situations are cases of *severe* uncertainty rather than of known variability.

Therefore, in this paper, we investigate the use of *non-probabilistic* models of uncertainty and adopt the usage of different robustness metrics by casting the decision-making process of the multi-objective optimization mapping algorithm into a new formulation that is based on the *info-gap theory of uncertainty* [7]. The info-gap decision theory - which is an instance of the Wald's Maximin paradigm [8] - has been used in a variety of problem formulations, including water resources management, multi-agent search, project management, job scheduling, energy economics, and others [9], [10]. Here, we attempt for the first time to apply the info-gap decision theory in the context of embedded systems design. More specifically, the main novelty of this paper is the application of the info-gap decision theory to modeling uncertainty in parameters in embedded systems and the use of the model in developing a solution to the problem of mapping with optimization of energy and of uncertainty-aware reliability. The solution consists of a design flow approach based on evolutionary algorithms, similar to that from [3], [4].

II. Robust Mapping of Embedded Systems Under Severe Uncertainty

A. Uncertainty Modeling

In employing an approximate model, one acknowledges a large information gap between what *is known* and what needs *to be known* in order to make fully competent decisions. This information gap is a severe form of uncertainty and often occurs in practice. For example, let $\widetilde{d}(w)$ denote the delay of the critical path of a hardware core as a function of different quantities w that include process variations, temperature, number of logic levels, etc., based on the best available model. The actual delay of a given core realization $d(w)$ deviates in an unknown manner from the nominal model $\widetilde{d}(w)$. If one has no information with which to express the likelihood of various alternative delay functions - hence one is unable to specify a probabilistic model for the uncertainty in the function $d(w)$ - an *info-gap model of uncertainty* can be formulated as the set of all functions consistent with the nominal function, $\widetilde{d}(w)$, up to a given level of deviation [7].

$$U(\alpha, \widetilde{d}) = \{d(w) : |d(w) - \widetilde{d}(w)| \leq \alpha\}, \alpha \geq 0 \quad (1)$$

where $U(\alpha, \widetilde{d})$ is the set of all functions whose deviation from $\widetilde{d}(w)$ is nowhere greater than α, the *uncertainty parameter*. For a fixed α, this set represents uncertainty in the delay function by specifying a range of variation of $d(w)$ around the nominal delay $\widetilde{d}(w)$. Because in practice, the value of α itself is not known, eq. 1 is not a single set but rather a family of nested sets. Examples of info-gap models include energy-bound, envelope-bound, Minkowski-norm, and Ellipsoid-bound.

Let q be a decision vector of design variables. Then, the robustness $\alpha(q)$ of decision vector q is the greatest value of uncertainty parameter α for which specified minimal requirements are always satisfied. The degree of success is often assessed by a scalar reward function $R(q, u)$, which depends on the vector q of decisions and on an uncertain vector u whose variations are described by an info-gap model $U(\alpha, \widetilde{u})$, $\alpha \geq 0$. In this case, the minimal requirement is that the reward $R(q, u)$ is no less than a critical value of r_c.

$$\widetilde{\alpha}(q, r_c) = max\{\alpha : \min_{u \in U(\alpha, \widetilde{u})} R(q, u) \geq r_c\} \quad (2)$$

where $\widetilde{\alpha}(q, r_c)$ expresses robustness, the degree of resistance to uncertainty and immunity against failure, so a large value of $\widetilde{\alpha}(q, r_c)$ is desirable.

B. Info-gap Based Uncertainty-aware Reliability Model

The system reliability is defined as the correct functioning of each component of the system over a given time interval without failure [11]. The first of the design objectives to be optimized in this work is the reliability of the system that must be maximized. We attempt for the first time to apply the info-gap theory to build the uncertainty aware reliability model. We use the tile as the fundamental component in our info-gap based uncertainty-aware reliability model, and the symbol τ is used to stand for a tile in the NoC based architecture platform. The reliability of a tile at time t is defined assuming the model from [11]:

$$R_\tau(t) = e^{-\lambda_\tau t} \quad (3)$$

Similarly, the reliability of a link is described as:

$$R_l(t) = e^{-\lambda_l t} \quad (4)$$

With the definition of the reliability of a tile and of a link described above, the reliability of the NoC at time t is expressed as:

$$R_{NoC}(t) = \prod_{j=1}^{T} R_\tau^j(t) \prod_{k=1}^{L} R_l^k(t) \quad (5)$$

where L is the total number of links and T is the total number of tiles in the NoC.

When calculating the reliability R_{NoC} according to eq. 5, we assume the absorbing Discrete Time Markov Chain (DTMC) models described in [12]. DTMC models are graphical models consisting of finite state machine like state graphs [13]. Thus, the system reliability can be calculated as $R = S_{(1,n)} R_n$, where S is the fundamental matrix of the DTMC, $S_{i,j}$ is the expected number of visits to state j starting from state i before it is absorbed, and n is the number of states in the DTMC model. R_n is the reliability of the n_{th} component, which is estimated with eq. 3. Please refer to [12] for more details about the DTMC models.

If, however, one wanted to capture the influence of uncertainty in the design parameters on the reliability for the NoC based architectures, then, one needs to modify the above deterministic reliability model. Previous work used *probabilistic* approaches that captured the uncertainty in design parameters in the form of probability distributions [3], [4]. But, what if one did not have access to such probabilistic models? What if the design parameters (such as the failure rate) are under *severe* uncertainty rather than of known variability? In such situations, one can employ the info-gap theory based uncertainty modeling method presented in Section II-A to capture the uncertainty in design parameters. In this case, for instance, the failure rate design parameter that is assumed to be affected by severe uncertainty can be modeled as a set of failure rates:

$$U(\alpha, \widetilde{\lambda}) = \{\lambda : |\lambda - \widetilde{\lambda}| \leq \alpha\}, \alpha \geq 0 \quad (6)$$

where, the design parameter $\widetilde{\lambda}$ is the nominal failure rate that is used in the traditional deterministic reliability model, such as the reliability model presented above, while the design parameter λ forms a set of failure rates, specifying a range of variation around the nominal failure rate $\widetilde{\lambda}$.

Then, the robustness of the NoC architectures $\widetilde{\alpha}(q, r_{DTMC})$ can be modeled as follows:

$$\widetilde{\alpha}_{DTMC}(q, r_{DTMC}) = max\{\alpha : \min_{\lambda \in U(\alpha, \widetilde{\lambda})} R_{DTMC}(q, \lambda) \geq r_{DTMC}\} \quad (7)$$

where $R_{DTMC}(q, \lambda)$ is called the reliability reward function, which has been described in Section II-A, that can be calculated as follows:

$$R_{DTMC}(q, \lambda) = \prod_{j=1}^{T} e^{-\lambda_j t} \prod_{k=1}^{L} e^{-\lambda_k t} = S_{(1,n)} R_n \quad (8)$$

Therefore, for the objective of reliability, which needs to be maximized (or converted to a minimization of $1-Reliability$), we need to identify the solution that offers the highest robustness $\widetilde{\alpha}_{DTMC}(q, r_{DTMC})$ such that the minimal reliability reward function $R_{DTMC}(q, \lambda)$ is no less than a critical value r_{DTMC}.

C. Modified Mapping Optimization Approach

To provide a solution to the problem of mapping under severe uncertainty, we adopt the multi-objective mapping

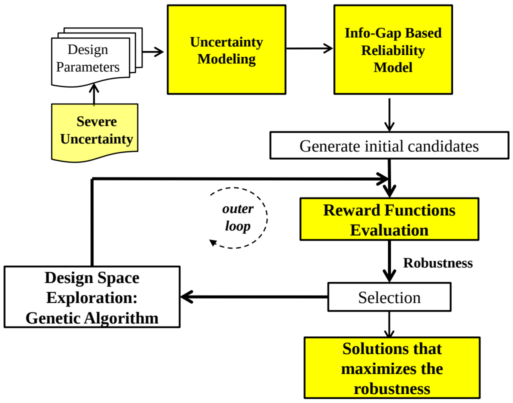

Fig. 1. Mapping optimization flow.

problem formulation from [3], [4] and modify the solution flow based on genetic algorithms illustrated in Fig. 1. We restrict ourselves to only two objectives: energy consumption and reliability. The diagram from Fig. 1 is modified to cast the problem of mapping into an info-gap decision theory based problem formulation. In this case, the Monte Carlo Simulation is replaced with the assessment or evaluation of the reliability reward function from in eq. 8. The goal is to transform the optimization process described by the outer-loop into a robust strategy, which maximizes the robustness or immunity to uncertainty. In other words, we want to seek the decision variable q (discussed in Section II-B), which guarantees the reliability reward function is of at least r_{DTMC} and which maximizes the immunity to the unknown variation of variables affected by uncertainty.

III. SIMULATIONS

The info-gap based uncertainty aware reliability model was implemented inside the proposed modified design flow. As an implementation of the genetic algorithm from Fig. 1, we use the implementation of the popular NSGA-II [15]. To improve the execution time of our implementation, we use OpenMP as a simple yet effective parallelization method, which helps to achieve an execution time linear with the number of iterations of the genetic algorithm. All simulations are conducted on a 64 bit Intel i7-7820 CPU, 2.90 GHz x8 running Ubuntu 16.04 LTS operation system with 40 GB memory.

We report simulations on the H.264 testcase available from [16] together with the ABS (anti-lock brake system), ACC (adaptive cruise control), and JPEG (picture compression) adopted from the study in [3]. This H.264 testcase is provided as a complete C/C++ description of each of the computational components which consist of the testcase. These components are integrated within an NoC simulator inside the same simulation framework named VNOC+H.264. Therefore, the testcase can be simulated for inputs that are real video streams, which represent workloads that exercise the NoC with traffic that is *realistic* and not synthetic. For example, we used the benchmark video file named *Plane* (available from [18]) as input to the H.264 testcase. The simulated platform is a 3x3

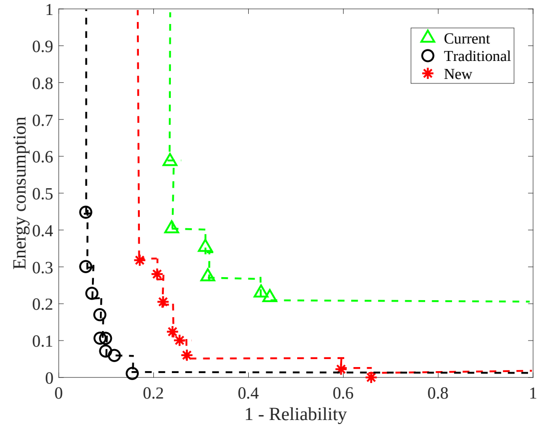

Fig. 2. The Pareto frontiers generated by the three tools investigated in this paper for the H.264 testcase. Axes values are normalized.

regular mesh NoC based architecture platform similar to the one used in [16]. This NoC is large enough to provide a router for each of the processing elements of the H.264 testcase, implemented in this work as *software* components connected to the NoC routers through network interfaces (NIs). The failure rates of the processing elements are adopted from [12]. The energy consumption of NoC routers and links are estimated using the open-source NoC simulator [17].

A. Results

In this section, we compare the proposed modified mapping design flow from Fig. 1, called *New* tool, with the *Traditional* tool from [11] and with the current state-of-the-art *Current* tool from [5]. All simulations generate 2D Pareto frontiers where the objectives are energy (we adopted the traditional state-of-the-art deterministic energy consumption model from [14]) and reliability (affected by severe uncertainty and modeled as described in this paper).

In the simulations, the *Current* tool is set up to inject a level of 5% uncertainty. The *New* tool is set up to use a reliability reward value of 0.525, which was empirically found to translate into a similar amount of injected uncertainty. It can be seen that there are significant differences between these frontiers. This is expected because the *Traditional* tool does not model uncertainties in design parameters on one hand, and because it is these differences in Pareto frontiers that are important to identify when considering the impact of uncertainty modeling because it helps to quantify how far solutions found by classic or traditional approaches can be from the truly optimal ones found by approaches that do model or capture uncertainty in design parameters as well.

To quantify the difference between the Pareto frontiers generated by the three tools, we use the concept of *hypervolume indicators* introduced in [4] with the reference point (1,1). The calculation results are reported in Table I. It can be observed that the difference between the *New* and the *Traditional* is $14.76\% - 31.6\%$ while the difference between the *New* and the *Current* is $6.8\% - 26.49\%$.

TABLE I

HYPERVOLUMES AND THEIR DIFFERENCE FOR THE PARETO FRONTIERS GENERATED BY THE NEW, TRADITIONAL AND CURRENT TOOLS.

Testcase	$H(New)$	$H(Traditional)$	$H(Current)$	New Vs. Traditional [11]	New Vs. Current [5]
H.264	0.7854	0.9214	0.5773	14.76%	26.49%
ABS	0.6676	0.9760	0.6222	31.6%	6.8%
ACC	0.6857	0.9316	0.6167	26.39%	10.06%
JPEG	0.7458	0.9127	0.6147	18.29%	17.57%

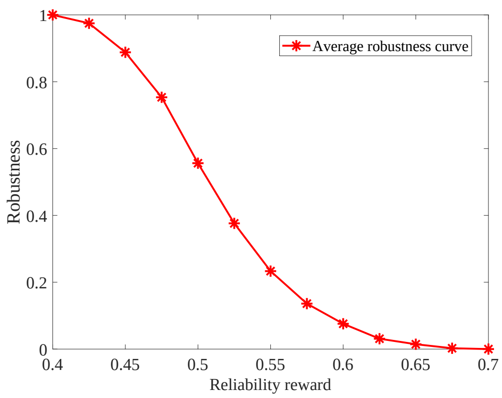

Fig. 3. The average robustness curve obtained by the *New* tool with different required reliability reward values for the H.264 testcase. Axes values are normalized.

Furthermore, we investigate the average robustness obtained by the *New* tool for different reliability reward values. The average robustness is calculated by all the mapping solutions generated by the *New* tool with a given reliability reward value. It can be used to measure the degree of resistance to uncertainty and immunity against failure for the mapping solutions generated by the *New* tool. The average robustness exhibits a variation as shown in Fig. 3. The average robustness vs. reliability reward curve highlights a *trade-off* between demanded-reward and immunity-to-uncertainty: if a large reliability reward is required, then only low immunity to uncertainty is possible. One can use the average robustness curve to decide the required reliability reward value in the *New* tool, so that it can generate mapping solutions that achieve good *trade-off* between demanded-reward and immunity-to-uncertainty. It should be noted that when applying the info-gap decision theory in the embedded systems design, the failure rate of the processing element has its natural range of $[0, 1]$, thus the robustness has also resulted in lower and upper range according to eq. 6 in Section II-B. This is the main reason why when a lower reliability reward is required, most of the solutions generated by the *New* tool have robustness close to the upper bound. Similarly, when a higher reliability reward is required, the average robustness is close to the lower bound.

IV. CONCLUSION

In this paper, we employ the info-gap theory of uncertainty to present an approach of modeling uncertainty in design parameters, when the amount of uncertainty is severe and unknown even in terms of probability distribution form. We use these modeling in the context of multi-objective mapping in embedded systems. Specifically, we consider failure rates of components in NoC based manycore architectures to be affected by uncertainty, and formulate the problem of mapping with two objectives: reliability, which is affected by the uncertainty and energy, which is modeled using a traditional deterministic approach. We develop a genetic algorithm based solution to this multi-objective mapping problem and discuss techniques to generate and quantify the difference between the robust Pareto frontiers generated by the proposed approach and those frontiers generated by previous approaches.

ACKNOWLEDGEMENT

The authors would like to acknowledge the time and considerations of our anonymous reviewers. This work was supported by the NSF, grant CCF-1524909. Any findings and conclusions expressed herein are those of the authors and do not necessarily reflect the views of the NSF.

REFERENCES

[1] I. Meedeniya, A. Aleti and L. Grunske, “Architecture-driven reliability optimization with uncertain model parameters,” *The Journal of Systems and Software*, vol. 85, no. 10, pp. 2340-2355, 2012.

[2] F. Khosravi, M. Muller and M. Glaß and J. Teich, “Uncertainty-aware reliability analysis and optimization,” *ACM/IEEE Design, Automation & Test in Europe Conference & Exhibition (DATE)*, 2015.

[3] W. Guan, M.G. Moghaddam, and C. Ababei, “Uncertainty aware mapping of embedded systems for reliability, performance, and energy," *IEEE ISQED*, 2018.

[4] W. Guan, M.G. Moghaddam, and C. Ababei, “Impact of uncertainty and correlations on mapping of embedded systems," *IEEE MWSCAS*, 2019.

[5] W. Guan, M.G. Moghaddam, and C. Ababei, “Quantifying the Impact of Uncertainty in Embedded Systems Mapping for NoC Based Architectures," *Elsevier Microprocessors and Microsystems (MICPRO)*, 2020 (under view).

[6] Predictive Technology Model (PTM), Arizona State University, 2014, http://ptm.asu.edu

[7] Yakov Ben-Haim, Information Gap Decision Theory - Decisions under severe uncertainty, Academic Press, 2001.

[8] M.Sniedovich, “The art and science of modeling decision-making under severe uncertainty," *Decision Making in Manufacturing and Services*, vol. 1, no. 1-2, pp. 111-136, 2007.

[9] I. Sisso, T. Shima, and Y. Ben-Haim, “Info-gap approach to multi agent search under severe uncertainty," *IEEE Trans. on Robotics*, vol. 26, no. 6, pp. 1032-1041, 2010.

[10] A. Soroudi and M. Ehsan, “IGDT based robust decision making tool for DNOs in load procurement under severe uncertainty," *IEEE Trans. on Smart Grid*, vol. 4, no. 2, pp. 886-895, 2013.

[11] N. Chatterjee, P. Mukherjee and S. Chattopadhyay, “Reliability-aware application mapping onto mesh based Network-on-Chip," *Integration, the VLSI Journal*, vol. 62, pp. 92-113, June 2018.

[12] Indika Meedeniya, “Architecture Optimisation of Embedded Systems under Uncertainty in Probabilistic Reliability Evaluation Model Parameters", *Ph.D. Thesis*, Faculty of Information and Communication Technologies, Swinburne University of Technology.

[13] R.C. Cheung, “A user-oriented software reliability model”, *IEEE Transactions on Software Engineering*, 1980.

[14] L. Liu, C. Wu, C. Deng, S. Yin, Q. Wu, J. Han, and S. Wei, “A flexible energy- and reliability-aware application mapping for NoC-based reconfigurable architectures," *IEEE Trans. on Very Large Scale Integration (VLSI) Systems*, vol. 23, no. 11, pp. 2566-2580, 2015.

[15] Kalyanmoy Deb, Multi-objective NSGA-II code in C. http://www.egr.msu.edu/ kdeb/codes.shtml

[16] M.G. Moghaddam and C. Ababei, “Performance evaluation of Network-on-Chip based H.264 video decoders via full system simulation," *IEEE Embedded Systems Letters*, vol. 9, no. 2, pp. 49-52, 2017.

[17] C. Ababei, ReliableNOC, 2010. [Online]. Available: http://www.dejazzer.com/

[18] FastVDO: H.264 Video Streams, 2016. [Online]. Available: http://www.fastvdo.com/H.264.html