Co-Design of Sustainable Embedded Systems-on-Chip

Jan Spieck, Dominik Walter, Jan Waschkeit, Jürgen Teich
{jan.spieck, dominik.l.walter, jan.waschkeit, juergen.teich}@fau.de
Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Germany

***Abstract*—This paper introduces a novel approach to the co-design of sustainable embedded systems through multi-objective design space exploration (DSE). We propose a two-phase methodology that optimizes both the multiprocessor system-on-chip (MPSoC) architecture and application mappings, considering sustainability, reliability, performance, and cost as optimization objectives. Our method thereby accounts for both operational and embodied emissions, providing a more comprehensive assessment of sustainability. First, an individual intra-application DSE is performed to explore Pareto-optimal constraint graphs for each application. The second phase, an inter-application DSE, combines these results to explore sustainable target architectures and corresponding application mappings. Our approach incorporates detailed models for embodied emissions (scope 1 and scope 2), operational emissions, reliability, performance, and cost. The evaluation demonstrates that our sustainability-aware DSE is able to explore design spaces, supported by superior results in four key objectives. This enables the development of sustainable embedded systems whilst achieving high performance and reliability.**

***Index Terms*—Sustainability, MPSoC, DSE, Emissions**

A. Introduction

In this work, we analyze and systematically explore the design space of tile-based multiprocessor system-on-chip (MPSoC) implementations of embedded systems with respect to sustainability by considering both their so-called *embodied* and *operational* greenhouse gas (GHG) emissions. *Operational* emissions are the emissions that occur during the operation of the system, while the *embodied* emissions refer to the emissions that occur during production of the chip. These can be differentiated into three *scopes* [12]. *Embodied Scope 1* emissions include chemicals and gases produced directly during chip manufacturing. The emissions are represented in CO_2 equivalents. *Embodied Scope 2* emissions represent the portion of emissions attributed to the energy required during the manufacturing process. Unlike scope 1 emissions which occur directly within the production environment, scope 2 emissions are dependent on the energy mix used to power the production facility. This energy mix varies by location based on the region's reliance on renewable or fossil fuel energy sources. *Embodied Scope 3* emissions include upstream and downstream supply-chain emissions and are omitted in this work due to their high variability and difficulty in estimation. While the operational emissions can be reduced by designing more energy-efficient computer architectures, embodied emissions can only be reduced—from a hardware design perspective—by either building smaller less powerful chips or extending the lifetime of the chip to amortize the fixed emissions during manufacturing.

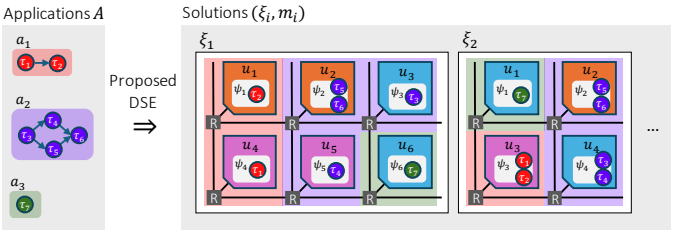


Fig. 1. Application graphs a_j mapped onto explored target architectures ξ_i.

Therefore, the lifetime, or more generally the reliability, of a system plays a critical role in achieving a sustainable design.

Current DSE methods [7, 9, 10, 1] neglect the embodied emissions of embedded systems, which are predicted to dominate future emissions [3]. Recent studies emphasize the need for a holistic approach to sustainability in computing, considering both operational and embodied emissions [3, 5, 2].

B. DSE for sustainable co-design

We consider a given set of applications A, each $a \in A$ to be mapped onto SoC architectures ξ of interconnected tiles $u \in U$, which contain processing resources $\psi \in \Psi$ (see Fig. 1), e.g., CPUs or accelerators [11]. We propose a two-phase DSE approach that explores architectures ξ and mappings m jointly for sustainability, performance (deadline miss rate dm) and operational energy consumption eng. First, for each application a, an *intra-application DSE* determining so-called *constraint graphs* $\eta_a \in H_a$ according to [9, 10] is performed. Let $R_v(\eta_a)$ denote the number of allocated tiles of each type $v \in V$ of constraint graph η_a:

$$\underset{\eta_a \in H_a}{\text{minimize}}(dm(\eta_a, \xi), eng(\eta_a, \xi), R_{v_1}(\eta_a), ..., R_{v_{|V|}}(\eta_a)) \quad (1)$$

Subsequently, a sustainability-aware *inter-application DSE* receives the non-dominated sets of constraint graphs H_a of each application $a \in A$ and then explores combinations of these graphs—and as a result mappings. Formally, it optimizes a selection of constraint graphs $\eta' = \{\eta_{a_1}, ..., \eta_{a_{|A|}}\}$, and a *production location* μ such that the global goals of sustainability $Su(\eta', \xi, \mu)$, reliability $Re(\xi)$ are maximized, and the cost $Co(\xi, \mu)$, and mean deadline miss rate $DM(\eta', \xi)$ are minimized:

$$\underset{\eta', \xi, \mu}{\text{maximize}}(Su(\eta', \xi, \mu), Re(\xi), -DM(\eta', \xi), -Co(\xi, \mu)) \quad (2)$$

The architecture ξ is given implicitly by the union of resources allocated for each application, exploiting spatial isolation and composability [10]. The sustainability $Su(\eta'_v, \xi, \mu)$ of the constraint graph vector η'_v on architecture ξ is estimated by the

978-3-9826741-0-0/DATE25/© 2025 EDAA

sum of the embodied emissions (scope 1 and scope 2) and the expected operational emissions over the lifetime of the MPSoC:

$$Su(\eta'_v, \xi, \mu) = -F_{\text{Scope1}}(\xi, \mu) - F_{\text{Scope2}}(\xi, \mu) - F_{\text{op}}(\eta'_v, \xi) \quad (3)$$

The *embodied scope 1* emissions F_{Scope1} can be estimated by the chip yield y, number of produced chips n_c, the number of wafers per chip $n_{\text{w/c}}$ for the target architecture ξ and the CO_2 emissions per wafer $E_{\text{CO}_2\text{/w}}$ at location μ (extended from [3]):

$$F_{\text{Scope1}}(\xi, \mu) = \frac{1}{y} \cdot \lceil n_c \cdot n_{\text{w/c}}(\xi) \rceil \cdot E_{\text{CO}_2\text{/w}}(\mu) \quad (4)$$

The *embodied scope 2* emissions F_{Scope2} can be calculated as follows where eng_{w} denotes the needed energy to produce one wafer (extended from [3]):

$$F_{\text{Scope2}}(\xi, \mu) = \frac{1}{y} \cdot \lceil n_c \cdot n_{\text{w/c}}(\xi) \rceil \cdot eng_{\text{w}} \cdot \rho_\mu \quad (5)$$

The factor ρ_μ expresses the CO_2 intensity of the energy mix at location μ during production. The *operational emissions* F_{op} can be computed as follows. Let $C \subseteq \mathbb{R} \times W$ be the set of tuples (r, λ) which describes that $r \cdot 100\,\%$ of the produced chips are expected to be operated in country λ. By considering both the ratio of chips $n_c \cdot r$ that are operated using a local energy mix ρ_λ and the energy consumption eng of the chip during its lifetime, the total operational emissions can be estimated as:

$$F_{\text{op}}(\eta'_v, \xi) = \sum_{(r,\lambda) \in C} (n_c \cdot r \cdot \rho_\lambda \cdot \sum_{\eta_a \in \eta'_v} eng(\eta_a, \xi)) \quad (6)$$

The reliability $Re(\xi)$ of the architecture ξ models the mean time to failure (MTTF) of the system.

$$Re(\xi) = \int_0^\infty \prod_{\psi \in \xi} \varsigma_\psi(t) dt = \frac{1}{\sum_{\psi \in \xi} \lambda_\psi} \quad (7)$$

ς_ψ denotes the reliability function of processing resource ψ and λ_ψ its failure rate. The performance $DM(\eta'_v, \xi)$ of the constraint graph vector η'_v on architecture ξ is estimated as follows. Given the set of periodic real-time applications where each application a has a priority p_a and a deadline θ_a, we want to minimize the average deadline miss rate $dm(\eta_a, \xi)$ of all mapped applications:

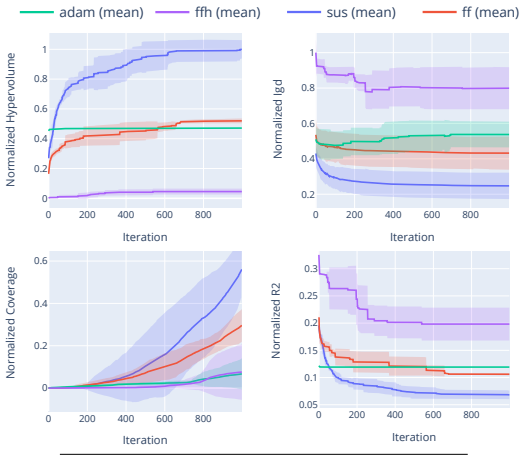
$$DM(\eta'_v, \xi) = \frac{1}{|A|} \cdot \sum_{\eta_a \in \eta'_v} dm(\eta_a, \xi) \cdot p_a \quad (8)$$

The cost $Co(\xi)$ of an architecture ξ is modeled by the wafer cost κ_{w}, and the electricity cost κ_{el} per wafer:

$$Co(\xi) = \frac{\kappa_{\text{w}} + eng_{\text{w}} \cdot \kappa_{\text{el}}}{n_{\text{w/c}} \cdot y} \quad (9)$$

C. Evaluation

We compare the proposed sustainability-aware DSE against various baselines regarding hypervolume, inverted generational distance (IGD), coverage, and r2 indicator [13]. This includes *FirstFit*, an extension of the multi-objective DSE from [10] with reliability and cost objectives, *AdaMD+*, an extension of the fast multi-application mapping optimization methodology by [8], and *Composition*, which uses *AdaMD* to determine the architecture as a composition of the mappings per application. This evaluation is based on the emission data from [5, 6, 5, 4, 2]. Acc. to Fig. 2, our proposed approach demonstrates superior performance across all metrics. The hypervolume metric rapidly converges to 1.0, while the best baseline, FirstFit, plateaus at 0.5. The IGD plot also shows that our solutions are consistently closer to the true Pareto front. Our method further excels in the coverage metric, finding nearly twice as many non-dominated solutions as FirstFit. Finally, the R2 indicator validates our method's superior convergence and diversity.



Metric on convergence	R2 (↓)	Hypervolume (↑)	IGD (↓)	Coverage (↑)
Composition	0.20	0.05	0.80	0.08
AdaMD+	0.12	0.47	0.53	0.07
FirstFit	0.11	0.52	0.43	0.30
Proposed	**0.07**	**1.0**	**0.25**	**0.56**

Fig. 2. Evolution of the normalized hypervolume, IGD, R2, and coverage metric for FirstFit (ff), AdaMD+ (adam), Composition (ffh), and our approach (sus). Shown are the average results across 5 runs (solid line) with the shaded area representing the ± 1 standard deviation from the mean.

References

[1] Nidhi Anantharajaiah et al. *Invasive Computing*. FAU University Press, 2022. ISBN: 978-3-96147-571-1.

[2] M Garcia Bardon et al. "DTCO including sustainability: Power-performance-area-cost-environmental score (PPACE) analysis for logic technologies". In: *2020 IEDM*. IEEE. 2020, pp. 41–4.

[3] Lieven Eeckhout. "Kaya for Computer Architects: Toward Sustainable Computer Systems". In: *IEEE Micro* 43.1 (2023), pp. 9–18.

[4] Robert Falkner. "The Paris Agreement and the new logic of international climate politics". In: *International Affairs* 92.5 (2016), pp. 1107–1125.

[5] Udit Gupta et al. "Chasing carbon: The elusive environmental footprint of computing". In: *HPCA*. IEEE. 2021, pp. 854–867.

[6] Nikhil Krishnan et al. "A hybrid life cycle inventory of nano-scale semiconductor manufacturing". In: *Environmental science & technology* 42.8 (2008), pp. 3069–3075.

[7] Behnaz Pourmohseni et al. "Hybrid Application Mapping for Composable Many-Core Systems: Overview and Future Perspective". In: *JLPEA* 10 (2020), pp. 1–37.

[8] Basireddy Karunakar Reddy et al. "AdaMD: Adaptive Mapping and DVFS for Energy-Efficient Heterogeneous Multicores". In: *TCAD* 39.10 (2020), pp. 2206–2217.

[9] Jan Spieck et al. "A Learning-based Methodology for Scenario-aware Mapping of Soft Real-time Applications onto Heterogeneous MPSoCs". In: *ACM TODAES* 28.1 (2023), 4:1–4:40.

[10] Jan Spieck et al. "A Scenario-Based DVFS-Aware Hybrid Application Mapping Methodology for MPSoCs". In: *TODAES* 29.4 (2024).

[11] Dominik Walter et al. "ALPACA: An Accelerator Chip for Nested Loop Programs". In: *ISCAS, Singapore, May 19-22, 2024*. IEEE, 2024, pp. 1–5.

[12] World Resources Institute et al. *Global Protocol for Community-Scale Greenhouse Gas Emission Inventories*. World Resources Institute, 2014.

[13] Eckart Zitzler et al. "Multiobjective evolutionary algorithms: a comparative case study and the strength Pareto approach". In: *TEC* 3.4 (1999), pp. 257–271.

Acknowledgment: This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project number 146371743 – TRR 89: Invasive Computing.