Abstract—The increased power densities of deep submicron process technologies have made on-chip temperature to become a critical design issue for high-performance integrated circuits. In this paper, we address the datapath merging problem faced during the design of coarse-grained reconfigurable processors from a thermal-aware perspective. Assuming a reconfigurable processor able to execute a sequence of datapath configurations, we formulate and efficiently solve the thermal-aware datapath merging problem as a minimum cost network flow. In addition, we integrate floorplan awareness of the underlying reconfigurable processor guiding the merging decision to account also for the effects of heat diffusion. Extensive experimentation regarding different configuration scenarios, technology nodes and clock frequencies showed that the adoption of the proposed thermal-aware methodology delivers up to 8.27K peak temperature reductions and achieves better temperature flattening in comparison to a low power but thermal-unaware approach.

I. INTRODUCTION

Coarse-grained reconfigurable processors (CGRPs) have been proposed as a new architectural paradigm which addresses the design requirements of hardware flexibility, high performance and fast reconfiguration time, found in several application domains, i.e. telecommunication, multimedia etc. Binding together the configurability of software solutions with the high performance of specialized hardware architectures, CGRPs are used to accelerate various computationally intensive kernels, customizing the datapath according to the application specific characteristics.

A specialized set of synthesis techniques, formally known as datapath merging [1]–[3], have been proposed in order to map datapath configurations onto the shared resources of a CGRP in an optimized manner. In particular, while the operation binding consists of the resource sharing decisions among the operations of a single DFG, the datapath merging refers to the resource sharing decisions among the functional units of multiple datapaths (see Figure 1), where each datapath is generated after performing the operation binding on its corresponding DFG. So far, datapath merging techniques proposed in literature focused on either the area or the interconnection optimization, neglecting power and/or temperature optimizations.

However, the aggressive transistor scaling of modern process technologies, is expected to lead to significant increase of the CGRP’s power density. These increased power density resulted in an increment of the IC’s temperature, which nowadays is considered as one of the dominating factors regarding the reliability of integrated circuits. High temperature impacts circuit reliability in many different ways, i.e. time to failure, timing and power violations. More specifically, increased on-chip temperature exponentially decreases the mean time to failure (MTTF), i.e. a 50% MTTF reduction is imposed for every 10-15 °C increase [4]. In addition, it decreases the switching speed of the transistors, thus negatively impacting the overall timing of the circuits, while also increasing – exposing cyclic dependency – leakage power, which forms a major source of power consumption in deep submicron technologies [5].

Several design-time temperature management techniques have been proposed targeting to architectural synthesis [6]–[9]. However, limited research results are available regarding thermal-aware design of CGRP architectures [10]–[12]. Inherently, the design of CGRPs adds a new degree of freedom to the design approaches focusing on architectural synthesis of Application Specific ICs (ASICs). Rather than optimizing the thermal profile of a single architecture instance, new synthesis methodologies and tools have to be developed taking into consideration the thermal features of the various architectural

Fig. 1. Datapath merging example. CW refers to the configuration word signals implementing the operation binding decisions. CID refers to the configuration ID signal implementing the datapath merging decisions.
In this paper, we address the aforementioned problem by introducing a thermal-aware design framework for CGRPs. We assume component-dominated CGRPs, in which the computing elements dominate the costs of multiplexers and wires added to introduce configurability [13]. Specifically, we propose a thermal-aware datapath merging (TADM) technique for peak temperature reduction and hotspot elimination for CGRPs designed in deep sub-micron process technologies. The paper introduces an innovative formulation of TADM as a min-cost network flow problem, which is efficiently solved by modern constraint/mathematical programming frameworks [14]. The proposed datapath merging technique works in an inter-configuration manner and it manages to intelligently redistribute the power densities (generated by intra-configuration synthesis decisions) of the CGRP’s functional units, thus reducing the peak temperature and flattening the CGRPs steady state thermal profile (hotspot elimination). In addition, the proposed technique is complementary to temperature reduction techniques proposed in [6], [9], since it is applied in an inter-configuration manner.

We experimentally evaluate the efficiency of the proposed TADM approach by using a set of representative multimedia configuration sequences as CGRP’s workloads. We show that the proposed datapath merging is able to deliver significant peak temperature reductions and eliminate thermal hotspots. We compared the proposed approach with a power-aware version of datapath merging based on [15]. The proposed approach can deliver up to 8.27K peak temperature reductions, showing that conventional power management techniques cannot guarantee also thermal optimization. Furthermore, we investigate the CGRP’s thermal behavior under differing clock frequencies for 90nm, 65nm and 45nm process technologies. The aforementioned exploration shows that scaling technology nodes, CGRP datapaths suffer from thermal emergency issues whenever clock frequency is getting closer to its upper limits. However, under the same target clock frequency, significant temperature stress relaxation can be achieved on the CGRP’s datapath when moving towards smaller technology nodes.

The paper is organized as follows. Section II introduces the state-of-the-art techniques on CGRP architectures and thermal-aware design methodologies. Section III introduces the proposed thermal-aware methodology, while Section IV describes the experimental evaluation of the proposed approach. Finally, Section V summarizes the relevant contributions of this work.

II. RELATED WORK

So far, several CGRP architectures [16]–[21] have been proposed exhibiting differing performance, area and power features. Datapath merging techniques can be used to efficiently share resources among multiple configurations mapped to the CGRP datapath. Existing datapath merging algorithms focus mainly on optimizing reconfiguration latency [3], interconnection overheads [1], [2] and routability [22]. However, the aforementioned approaches do not consider power and/or temperature optimization, which form major design objectives in sub-micron process technologies.

In the field of CGRP design, the authors in [10], proposed a row-wise activity migration scheme to prevent hotspot creation. The activity is migrated in an empirical manner interchanging operation executions between non adjacent rows. In [12], a random search algorithm is invoked in order to find an efficient activity migration following a row-based approach as in [10]. In [11], the authors presented a thermal-aware dynamic placement planner for optimizing peak temperature transients through dynamic reconfiguration.

Several thermal-aware synthesis techniques have been proposed in the literature, targeting to monolithic (non reconfigurable) circuit architectures. In [6], [9], iterative operation re-binding is proposed to reduce the peak temperature of the datapath. Simultaneous average and peak power minimization is proposed in [7], based on the usage of a genetic algorithm. In [8], the authors address temperature optimization at the physical level through voltage selection starting from a power optimized implementation. In [23], a compiler-directed power distribution is presented for thermal management of the functional units of VLIW processors, while in [24] the authors explore the performance, energy, and temperature trade-offs for application-specific instruction processors. We highly differentiate from the aforementioned approaches, since we are targeting to thermal-aware datapath merging for generating temperature optimized reconfigurable datapaths.

III. TADM METHODOLOGY

In this section, we formulate the TADM problem as a minimum-cost network flow problem. Network flow formulations have gained a lot of attention in the field of low power design, regarding either operation binding [15], [25], or memory [26] and register allocation [27]. The aforementioned techniques target only to the low power domain without considering thermal optimization. In [9], the authors iteratively modify and solve a network flow that corresponds to local operation binding decisions for minimizing the temperature of the hottest units. In this paper, we adopt a different approach by focusing on coarse-grained reconfigurable processors (CGRPs) rather than monolithic ICs and by developing a network flow formulation performed as a single pass rather than iteratively to reduce and flatten CGRP’s on-chip temperature. The target optimization problem can be stated as follows:

**Thermal-Aware Datapath Merging Problem:** Generate a global resource sharing plan that minimizes peak temperature across the overall configuration chain, once given (i) the number and the type of the shared resources available in a CGRP and (ii) the operation binding decisions for each datapath included in a configuration chain.

Next sections introduce the adopted thermal model (section III-A), the min-cost network flow formulation for the TADM (section III-B) and the restructuring procedure for including floorplan-aware constraints within the TADM network flow model (section III-C), respectively.

A. Thermal Model

We assume a CGRP that consists of \(Res_N\) functional units (FUs) shared among \(C_N\) configurations. In each configuration, the CGRP is able to execute a set of operation groups that are mapped to the available FUs. Let the index \(i \in Res_N\) refer to the \(i^{th}\) FU of the CGRP and index \(k \in C_N\) to the \(k^{th}\) configuration mapped onto the CGRP. The operation group mapped onto \(FU_i\), during configuration \(k\) is indexed using the vector of indexes \((i,k)\). Each \(FU_i\) is able to perform one or more types of operations, i.e. addition, multiplication, addition.
& multiplication etc. Let $OT$ be the set of available operation types that can be performed by the overall set of FUs of the CGRP. We define the function $M : ResSN \rightarrow OT$, where $M_i$ returns the set of operation types able to be executed by $FU_i$.

During configuration $k$, the functional unit $FU_i$ executes an operation group (set of sequenced operations), $(i,k)$, according to the binding decisions considered for the datapath of the specific configuration. If $c_i$ is the energy per operation averaged on the set of operation types $M_i$ on $FU_i$, then the overall power, $P_{i,k}$, dissipated by $FU_i$ during configuration $k$ is given by the following equation:

$$P_{i,k} = P_{i,k}^{Dyn} + P_{i,k}^{Static} = \#events_{i,k} \times \frac{c_i}{t_{k}^{DF}} + P_{i,k}^{Static}$$ (1)

For a single configuration $k \in C_N$, the temperature $T_{i,k}$ of the functional unit $FU_i$, can be calculated as follows [28]:

$$T_{i,k} = P_{i,k} \times R = P_{i,k} \times \left( \frac{t}{k_1 \times A_i} \right) = PD_{i,k} \times \frac{t}{k_1}$$ (2)

where $T_{i,k}$ is the temperature, $P_{i,k}$ is the dissipated power, $PD_{i,k}$ is the power density of the $FU_i$ during configuration $k$, $R$ is the thermal resistance between the block and the environment, $t$ is the thickness of the chip, $k_1$ is the thermal conductivity of the material and $A_i$ is the area of $FU_i$. Thus, for a single configuration the temperature of functional units strongly depends on the its power density, $PD_{i,k}$.

**B. Min-cost Network Flow for TADM**

The datapath elements of a CGRP are shared among the various mapped configurations, $C_N$, thus each $FU_i$ is shared among the mapped operation groups decided during the intra-configuration operation binding phase. According to [2], the inter-configuration temporal behavior of a CGRP can be abstracted as a sequence of datapath specifications each instantiating the architecture of its corresponding configuration. From this inter-configuration model, we define the Inter-Configuration Power Density (ICPD) table, defined as follows:

**Definition 1.** The Inter-Configuration Power Density (ICPD) table is defined as a table of $R_N$ rows and $C_N$ columns. Each row corresponds to a functional unit $FU_i$, $i \in R_N$ of the CGRP, while each column corresponds to a configuration $k$, $k \in C_N$. Each element, $FU_{i,k}$, reports the power density, $PD_{i,k}$ value of the $FU_i$ for configuration $k$.

The thermal-aware datapath merging problem can be mapped to that of re-assigning the operation groups bounded to $FU_{i,k}$ – decided during the intra-configuration operation binding – in a thermal optimized way along the overall configuration plan. Thus, the problem is equivalent to finding the optimal permutation of the column cells of ICPD table (without permuting the cells of the first column/configuration) in order to minimize the temperature in all the rows of ICPD.

For $FU_i$, the transition between two consecutive configurations, i.e. $k \rightarrow k+1$, implies the execution of two operation groups, i.e. $i$ and $i'$ that generate a temperature difference, $T_{i \rightarrow i',k \rightarrow k+1}$, which can be calculated according to the following formula:

$$T_{i \rightarrow i',k \rightarrow k+1} = T_{i',k+1} - T_{i,k} = (PD_{i',k+1} - PD_{i,k}) \times \frac{t}{k_1}$$ (3)

![Fig. 2. Network flow graph for a CGRP with 3 homogeneous FUs.](image)

It is important to notice that the temperature difference between two configurations can be either a positive or a negative value. A positive value of $T_{i \rightarrow i',k \rightarrow k+1}$ implies that the consecutive execution of the operation groups $i \rightarrow i'$ associated to the specified configuration scenario heats up the FU, while a negative value implies that the combination of the specific operation groups cools down the FU.

We cast the target optimization problem as a min-cost network flow problem. Min-cost flow problems can be solved in polynomial time using linear programming [29]. We first construct the network flow graph of the problem and then we formulate the corresponding min-cost network flow optimization problem.

Let us assume the acyclic graph $G = \{V,E\}$, where $V$: set of nodes and $E$: set of edges $e_{v_i,j,v_{i',j'}}$, $v_{i,j},v_{i',j'} \in V$. There is a vertex $v_{i,j} \in V$ for each entry of the ICPD table. Thus, each vertex $v_{i,j}$ models the operation group $(i,j)$ mapped onto $FU_i$, $i \in R_N$ under configuration $j$, $j \in C_N$. For each of distinct vertices, i.e. $v_{i,j}$ and $v_{i',j'}$, we add a directed connection edge $e_{v_{i,j},v_{i',j'}}$, iff $(j' = j + 1) \cap (M_i \subseteq M_{i'})$. Thus, we add $e_{v_{i,j},v_{i',j'}}$ if vertex $v_{i',j'}$ corresponds to the exact subsequent configuration of $v_{i,j}$ and iff the two FUs associated with the examined vertices are compatible in respect to their operation types. Also, we insert two special nodes to $G$, namely the source and the sink node. For each vertex $v_{i,j}$, $j = 1$ (vertex corresponding to configuration $j = 1$), we add a directed edge, $e_{\text{source},v_{i,j}}$. Accordingly, we add a directed edge between each vertex $v_{i,j}$, $j = |C_N|$ and the sink node. For each edge $e_{v_{i,j},v_{i',j'}}$, $v_{i,j},v_{i',j'} \in V$, we associate (i) an upper and lower bound of the capacity value, $u_{v_{i,j},v_{i',j'}}$, $l_{v_{i,j},v_{i',j'}}$, and (ii) a cost value, $c_{v_{i,j},v_{i',j'}}$, $v_{i,j},v_{i',j'} \in V$. For all edges, we set upper bound capacity equal to 1 and lower bound capacity equal to 0. For the edges initiated from vertex source $(e_{\text{source},v_{i,j}}, j = 1)$ and for the edges terminated to vertex sink $(e_{v_{i,j},\text{sink}}, j = |C_N|)$, we assign zero cost. For the rest of the $G$’s edges, the cost value $c_{v_{i,j},v_{i',j'}}$ models the thermal cost associated with each $FU_i$ when a configuration transition occurs.

For each transition, the $c_{v_{i,j},v_{i',j'}}$ is derived through a thermal cost function that tries to balance the power consumption across the overall configuration plan for all the FUs of the CGRP. The overall power mapped onto the CGRP during the configuration plan is $P_{tot} = \sum_{k \in C_N} \sum_{i \in Res_{SN}} P_{i,k}$. For a balanced thermal profile of the CGRP, each $FU_i$ in each configuration $k$ should dissipate an ideal amount of power given by $P_{ideal} = \left| Res_{SN} \times C_N \right|$. [23]. The cost, $c_{v_{i,j},v_{i',j'}}$, of each configuration transition is formulated as the difference between $P_{ideal}$ and a two coefficient term formed by (i) the power dissipated from $FU_i$ during the execution of configuration $j'$ and (ii) the temperature difference that $FU_i$ exhibits during the
transition between the two consecutive configurations \( j \rightarrow j' \).

The \( e_{v_{ij},v_{ij'}} \) is formed according to the following equation:

\[
e_{v_{ij},v_{ij'}} = |P_{ij'} + T_{ij'j}^{DF} - P_{ideal}| \tag{4}
\]

In the context of multiple configuration transitions, the formulated \( e_{v_{ij},v_{ij'}} \) tries to minimize the difference from \( P_{ideal} \) generated in each \( FU_i \) among the overall configuration span. The bigger the absolute difference from \( P_{ideal} \), the larger the non-uniformity of the power distribution. Thus, we use the absolute operator in order to manage situations in which the \( T_{ij'j}^{DF} \) has a negative value. As suggested in [30], [31], temperature flattening can be achieved through activity migration, i.e., mapping temperature intensive operation groups to "cold" functional units and vice versa. However, migration should be performed in a way that power is uniformly distributed across the CGRP, avoiding the formation of very "cold" and very "hot" regions. Thus, the cost function minimizing the absolute difference from \( P_{ideal} \), targets to the temperature flattening across configurations in order to prevent the generation of localized hotspots.

Finally, to ensure that the flow paths are node-disjoint, we apply a node splitting on the flow graph \( G = (V,E) \). Specifically, each node \( v_{ij} \in V - \{source, sink\} \) is split to two disjoint nodes \( v_{ij}^a, v_{ij}^b \) connected with an edge \( e_{v_{ij}^a, v_{ij}^b} \) with \( u_{v_{ij}^a,v_{ij}^b} = 1 \), \( l_{v_{ij}^a,v_{ij}^b} = 0 \) and \( c_{v_{ij}^a,v_{ij}^b} = 0 \). Figure 2 shows an exemplary network graph for the case of 3 configurations mapped onto an CGRP with 3 homogenous functional units.

We introduce the variable, \( x_{v_{ij}^a,v_{ij'}^a} \), \( x_{v_{ij}^b,v_{ij'}^b} \), \( v_{ij}, v_{ij'} \) \( V \), to represent the flow in edge \( e_{v_{ij},v_{ij'}} \) in \( G \). The flow along any directed edge of \( G \) must be positive valued, less than or equal to the upper capacity bound of the corresponding edge. The formulation of the network \( G \) imposes the flow variables to have a value either 1 or 0. Thus, the flow variables form the decision variables of our problem instantiation. Specifically, a decision \( x_{v_{ij}^a,v_{ij'}^a} = 1 \), indicates that the group of operations of \( FU_i \) during configuration \( j \) will be mapped together with the groups of operations of \( FU_i \) during configuration \( j' \). The FU that will execute the two mutual exclusive group of operations \((i,j)\) and \((i',j')\) is determined by the FU that initiates the corresponding flow path at configuration \( j = 1 \).

For each node \( v_{ij} \) in the network graph \( G \), the flow conservation constraints have to be satisfied, meaning that the flow inserting into the node is equal to the flow exiting out of the node. For the two special nodes of \( G \), source and sink, the flow out of source is equal to the flow into the sink. Edges incident to source only leave node source, while edges incident to node sink are only directed into node sink.

The minimum cost network flow instance for the thermal-aware datapath merging problem can be stated as follows: Given a fixed amount of flow, \( F = |Res_N| \) from source to sink, find the flow that minimizes the total cost, defined as the sum over all edges of the product of the flow, \( x_{v_{ij},v_{ij'}} \), and the cost, \( c_{v_{ij},v_{ij'}} \) in each edge of graph \( G \). Table I includes the formulation of the min-cost network flow problem for thermal-aware datapath merging.

\[
\begin{align*}
\text{TABLE I} \\
\text{TADM NETWORK FLOW FORMULATION.}
\end{align*}
\]

| Objective: | Minimize \( \sum_{(v_{ij},v_{ij'}) \in E} c_{v_{ij},v_{ij'}} \times x_{v_{ij},v_{ij'}} \) |
| subject to: | - Flow conservation constraints, \( \forall v_{ij} \in V, v_{ij} \neq \{source, sink\}: \sum_{(v_{ij},v_{ij'}) \in E} x_{v_{ij},v_{ij'}} - \sum_{(v_{ij},v_{ij'}) \in E} x_{v_{ij'},v_{ij}} = 0 \) |
| | - Maximum demand flow constraints: \( \sum_{(v_{ij},sink) \in E} x_{v_{ij},sink} = F, \forall v_{ij} \in V \) |
| | - Maximum supply flow constraints: \( \sum_{(source,v_{ij}) \in E} x_{source,v_{ij}} = F, \forall v_{ij} \in V \) |
| | - Flow range constraints: \( l_{v_{ij},v_{ij'}} \leq x_{v_{ij},v_{ij'}} \leq u_{v_{ij},v_{ij'}} \forall (v_{ij},v_{ij'}) \in E \) |

\[
\begin{align*}
\text{C. Floorplan-Aware Network Restructuring}
\end{align*}
\]

Given the floorplan of the CGRP datapath, the temperature of \( FU_i, T_{i,k} \), depends both (i) on the power, \( P_{i,k} \), consumed on \( FU_i \) during configuration \( k \) and also (ii) on the heat diffusion between the \( FU_i \) and all its adjacent functional units. The heat diffusion between two adjacent units is proportional to the product of their temperature difference and the shared length between them. Several previous works [6], [7], [9], considered heat diffusion based on iteratively performing temperature simulation of the floorplanned architecture for each mapping decision. However, temperature simulations are introducing additional runtime costs, which in the case of complex systems including several components and for high temperature resolutions are not negligible. In this paper, we adopt a different way to account for heat diffusion effects when floorplan is available. Rather than resorting to many temperature simulations for each possible merging decision, we restructure the TADM network flow model by inserting floorplan-aware constraints that reduce the heat flow warming.

These floorplan-aware constraints eliminate merging decisions (= edges in the TADM network flow model) that do not lead to significant temperature reduction on the specified FU (= node of the TADM network flow model). Merging the operation group of \( FU_i \) to one of its neighbouring FUs will have limited impact to the temperature reduction of \( FU_i \) due to the heat diffusion effect. The restructuring of the TADM network flow is based on eliminating network flow edges that invoke merging between neighbouring FU nodes. The proposed restructuring is heuristic in nature, since it silently assumes that the heat diffusion is uniform towards all the directions.

Specifically, given the floorplan of a CGRP and the thermal characteristics of the device, (such as the thermal conductivity of the material), we define a cyclic region with radius \( r_i \) for each \( FU_i \), where the heat can be diffused. The radius, \( r_i \) is specific for each FU, since the region of diffusion also depends on the size and the ratio of the FU. We associate \( r_i \) with the dimension of \( FU_i \) according to the following formula:

\[
r_i = a \times \max(L_i^x, L_i^y) \tag{5}
\]

where \( L_i^x/L_i^y \) is the length of \( FU_i \) on the x-axis/y-axis and \( a \) is a user defined weight variable, respectively. The \( a \) parameter affects the quality of the extracted merging decisions (in worst cases this could also impact the feasibility itself of the TADM network flow). The higher the value of \( a \) variable the more merging decisions are eliminated.
The restructuring of the TADM network flow model to eliminate non promising merging decisions is performed according to the following steps:

- For each $FU_i$ generate the set of its neighbouring nodes based on the floorplanned CGRP, $NG_i = \{ FU_{i,j} \in R_n - i \mid Distance(FU_{i},FU_{j}) \leq r_i \}$, where $Distance(FU_{i},FU_{j})$ refers to the euclidean distance between the center points of $FU_i$ and $FU_j$ on the floorplanned design.
- Propagate $NG_i$ as a constraint matrix to the TADM network flow generator.
- For each configuration $j$ and for each $FU_i$ eliminate the network flow edges, $(v_i,j, v_i,j+1)$ that satisfy the following condition: $j' \in NG_i$.

IV. EXPERIMENTAL RESULTS

In order to evaluate the proposed TADM methodology, we developed the automated software framework shown in Figure 3. The TADM framework accepts as inputs (i) the floorplan of the target CGRP, (ii) the power, $P_i,k$, dissipated by each $FU_i$ in configuration $k$ and it automatically constructs and solves the TADM network flow formulation. The corresponding TADM min-cost network flow instance is automatically generated in the AMPL modeling language and solved using the CPLEX network solver [14]. The decision variables (flow variables) of the TADM are propagated to (i) a re-writing procedure that structures the RTL code and the corresponding configuration instructions of the unified CGRP datapath and to (ii) a power trace generator, producing the power trace of the CGRP once the TADM methodology has been applied. Finally, the floorplanned CGRP and its corresponding power trace are propagated to the Hotspot [32] temperature simulator in order to generate the thermal map of the CGRP.

Additionally, we utilized the publicly available SPARK-HLS tool [33] to generate the RTL descriptions and Hot-Floorplan [32] for deriving the CGRP floorplan\(^1\). Moreover, the area and power libraries integrated into the McPAT [34] architectural simulator have been used to characterize the area and power values of each CGRP’s resource per configuration.

\(^1\)The floorplan can be either specified by the designer or extracted by other automatic floorplanning tools.

We considered a CGRP architecture consisting of 16 FUs, 12 ALU and 4 multipliers. We evaluated 8 reconfiguration scenarios in which either (i) the CGRP is reconfigured executing the same functionality (i.e., Matmul, Sobel, DCT, IDCT, SAD and YUV2RGB) or (ii) different kernels are mapped onto the CGRP within the configuration plan (see Table II).

Since there are not yet other thermal-aware approaches for datapath merging, we compare the proposed TADM approach with a (thermal unaware) low-power datapath merging approach (LPDM), minimizing the average power. We implement LPDM based on a low-power network flow formulation inspired by [15] for low power module sharing. Since [15] targets monolithic ICs and not CGRPs, the following adjustments have been considered for the LPDM, (i) use of a single commodity min-cost network flow formulation like the one proposed in TADM rather than a multi-commodity one [15], (ii) each configuration of the CGRP corresponds to a pipeline stage of [15], (iii) the cost in the LPDM formulation considers both the dynamic and static power rather than the switching activity used in [15]. We explored the efficiency of the proposed TADM with respect to the LPDM approach on three differing technology nodes (90nm, 65nm and 45nm) for two clock frequency configurations (2GHz and 3GHz).

Figure 4 reports the peak temperature reductions delivered by the proposed approach. For each combination of technology node, clock frequency and configuration scenario, a single customized CGRP floorplan has been considered. Thus, the peak temperature comparison between the two approaches, TADM and LPDM, is done on the same floorplanned CGRP. Specifically, the TADM methodology manages to deliver lower peak temperatures in all the analyzed cases, up to 8.27K, outperforming the LPDM approach\(^2\). As expected, the advantage in terms of peak temperature varies in respect to the mapped configuration scenario and technology node making the temperature reduction more evident where the thermal

\(^2\)Those values are in the same order of what have been reported in [23] and [24] for similar approaches in different areas.
stress is more intense, i.e. at higher frequency. Moreover, it is interesting to notice that for the examined configurations, the thermal stress is relaxed as we moving towards smaller technology nodes, mainly due to the \( V_{dd} \) scaling. Passing from 90nm to 65nm and 45nm the peak temperature reduction respectively scales from 3K to 1.6K and 0.9K for the 2GHz case, while for the 3GHz case from 4K, 2.3K and 1.7K.

Furthermore, we evaluated the TADM efficiency regarding temperature flattening, by considering for each examined scenario the difference between the peak and the average temperature, i.e. \( T_{\text{peak}} - T_{\text{average}} \). Small values of \( T_{\text{flat}} \) indicates that the temperature is distributed in a more uniform manner. We observed that the proposed TADM approach outperforms LPDM, delivering more flattened temperature profiles up to 8K. Large advantages in terms of temperature flattening are noticed when the working frequency is at 3GHz, implying higher thermal stressed scenarios.

V. CONCLUSION

In this paper, we addressed the problem of thermal-aware datapath merging for peak temperature reduction and flattening of coarse-grained reconfigurable processors. The target problem has been formulated as a minimum-cost network flow problem, thus being able to efficiently solved by state-of-art mathematical/constraint programming frameworks. In addition, floorplan awareness is introduced by a network flow restructuring technique to take into consideration the heat diffusion during the datapath merging decisions. Extensive experimentation regarding different configuration scenarios, technology nodes and clock frequencies showed that the proposed thermal-aware approach delivers higher quality solutions in respect to a low power oriented one, reaching peak temperature reductions up to 8.27K.

REFERENCES


