

Energy Minimization in DAG Scheduling on MPSoCs at Run-Time: Theory and Practice

Bertrand Simon

Universität Bremen, Germany
bsimon@uni-bremen.de

Joachim Falk

Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Germany
joachim.falk@fau.de

Nicole Megow

Universität Bremen, Germany
nicole.megow@uni-bremen.de

Jürgen Teich

Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Germany
juergen.teich@fau.de

Abstract

Static (offline) techniques for mapping applications given by task graphs to MPSoC systems often deliver overly pessimistic and thus suboptimal results w.r.t. exploiting time slack in order to minimize the energy consumption. This holds true in particular in case computation times of tasks may be workload-dependent and becoming known only at runtime or in case of conditionally executed tasks or scenarios. This paper studies and quantitatively evaluates different classes of algorithms for scheduling periodic applications given by task graphs (i.e., DAGs) with precedence constraints and a global deadline on homogeneous MPSoCs purely at runtime on a per-instance base. We present and analyze algorithms providing provably optimal results as well as approximation algorithms with proven guarantees on the achieved energy savings. For problem instances taken from realistic embedded system benchmarks as well as synthetic scalable problems, we provide results on the computation time and quality of each algorithm to perform a) scheduling and b) voltage/speed assignments for each task at runtime. In our portfolio, we distinguish as well continuous and discrete speed (e.g., DVFS-related) assignment problems. In summary, the presented ties between theory (algorithmic complexity and optimality) and execution time analysis deliver important insights on the practical usability of the presented algorithms for runtime optimization of task scheduling and speed assignment on MPSoCs.

2012 ACM Subject Classification Software and its engineering → Scheduling; Theory of computation → Scheduling algorithms

Keywords and phrases energy minimization, speed scaling, precedence graphs, scheduling, critical path, MPSoC

Digital Object Identifier 10.4230/OASICS.NG-RES.2020.2

1 Introduction

Dynamic voltage and frequency scaling (DVFS) on modern processors is a mean to actively control the power and energy consumption of an MPSoC (multi-processor system-on-chip). It is used for thermal chip management in combination with dynamic power management (DPM) [5]. But it can also be used in the context of dynamic energy minimization of programs executed on the MPSoC, e.g., for real-time applications. Here, a plethora of methods has been proposed to optimize the mapping (including task assignment and scheduling) of tasks of one or multiple applications to processor cores including the selection of processor speed(s) such that, given worst-case task execution times, a global deadline is met. While first investigations only considered uni-processor systems, a great number of approaches has emerged to apply DVFS optimization algorithms offline when targeting MPSoCs [7, 15, 18].



© Bertrand Simon, Joachim Falk, Nicole Megow, and Jürgen Teich;
licensed under Creative Commons License CC-BY

Workshop on Next Generation Real-Time Embedded Systems (NG-RES 2020).

Editors: Marko Bertogna and Federico Terraneo; Article No. 2; pp. 2:1–2:13



OpenAccess Series in Informatics

OASICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

These approaches, however, generally suffer from assuming fixed execution times of tasks given (e.g., WCETs). However, for most applications, the execution times of tasks may depend on the workload to be processed. Or, tasks may only be conditionally executed according to control flow information [22]. Hence, a static assignment of schedule and speeds for executing tasks might not be optimal. Choudhury et al. [6] proposed a combination of offline techniques to compute worst-case and average case execution times of tasks. At run-time, a computationally inexpensive method calculates observed slack, and adapts processor speeds for energy reduction, while still guaranteeing a global deadline not to be violated. Other approaches such as [23] exploit the knowledge of special models of computation such as synchronous dataflow (SDF) to apply a mixed offline and online DVFS optimization for MPSoCs. Still, the structure of the task graph and thus periodicity of executions is assumed static. In most general applications, however, both the execution times and the task graph structure may vary over time. Here, approaches using control/dataflow graphs (CDFGs) have been proposed such as in the work of Tariq et al. [25]. However, the presented computationally complex analysis and optimization is again purely static as task execution probabilities are used and thus only the *expected* energy consumption is targeted.

On the theoretical side, Yao et al. [27] initiated the algorithmic study of speed scaling in 1995. This area received a lot of attention since then; see the surveys [2, 12]. Most of these studies focus on scheduling *independent tasks* (without precedences) and a *single processor*. Regarding the speed choice model, only few theoretical works address the *discrete speed* model which is computationally much more complex but more realistic; see, e.g. [13, 16, 17]. Most related to our investigations is the work by Aupy et al. [3] that studies the problem of minimizing the energy consumption under a given mapping of tasks to cores, and where the power consumed by a core running at speed s is equal to s^α . They consider both the continuous and the discrete speed model. Pruhs et al. [21] focus on the problem of minimizing the makespan under an energy budget in the continuous speed model with the same power law. In this framework, they designed an approximation algorithm with a polylogarithmic ratio. Bampis et al. [4] later proposed a 2-approximation for the same problem, which matches the best known algorithm for makespan minimization without energy considerations [10]. Our contributions include to rephrase these results for our framework (energy minimization with a fixed deadline) and analyze the algorithms performance experimentally. We also add new results building on this earlier work.

A major goal of this paper is to analyze whether algorithms providing provably optimal results or at least approximation bounds on the quality of the results can be implemented and practically applied in a real MPSoC system to be executed at runtime. In this regard, at least to our knowledge, the following questions have so far not satisfactorily been answered for the problem of scheduling task graphs purely at runtime based on dynamically emerging task dependence structure and worst case execution times.

- **Are there in theory sound algorithms that also can be applied in practice on an MPSoC?** E.g., depending on the absolute time scale, many real-world applications do require solutions to be computed within a time scale of 1 to 10 ms in order to be of practical use.
- **How do these execution times scale with the problem size?** Problem instances ranging in size between 10 to 500 tasks should be handled in practice within such time scales. If not possible:
- **Are there fast and scalable algorithms with provable approximation bounds on the optimality of energy consumption?**

Our main contribution is to bring together theoretical and computational results for continuous and discrete speed scaling for precedence-constrained task systems with the goal to minimize the energy consumption. We distinguish two classes of problems: The first assuming

a processor assignment and schedule of tasks on cores given, and the second computing the full schedule including the task to processor mapping as well while minimizing energy. We present algorithms building on mathematical optimization techniques such as convex and integer linear programs as well as rounding solutions of relaxations. Previously known methods are adapted to our setting and we also provide new results. For the full portfolio of considered settings (continuous/discrete speed choice, unlimited/bounded number of processors) we give both, an exact algorithm and a computationally efficient (i.e., polynomial-time) algorithm. In cases where optimal polynomial-time algorithms are ruled out under standard complexity assumptions, we give a polynomial-time algorithm with an approximation guarantee, i.e., we guarantee for any input instance that the total energy needed by our algorithm to finish all tasks by the deadline is at most a certain factor away from a minimum energy needed by any algorithm. Such theoretical approximation results give very rigorous worst-case guarantees on the solution quality under any possible input. They are of high importance particularly for safety-critical real-time applications. In our experiments on real-world instances, it is shown that the solution quality is substantially better than the ones guaranteed in our theorems for the worst case.

Moreover, we rigorously analyze the applicability of all of our algorithms on problem instances taken from realistic embedded system benchmarks as well as synthetic scalable problems. As one result, it turns out that the mathematical optimization methods are applicable for MPSoC system applications despite their complexity. Running times between 1 to 10 *ms* for instances up to 100 tasks are in the acceptable range for many applications. If not, also a linear-time algorithm (previously used in similar settings [3, 11, 20]) that combines optimality with scalable performance for a majority of task graph instances exhibiting a series/parallel dependence structure is presented. Overall, our results include new and old algorithms with optimality/approximation guarantees while revealing their practicability for use in MPSoCs.

2 Formal problem definition and notation

We are given a set of tasks to be executed without preemption on m cores. Precedence relations between the tasks are given as a directed acyclic graph $G = (V, E)$, where each node in the graph is associated with a task. If there is an arc in E from task j to task k then task k cannot start before task j is completed. A task $j \in V$ has a nominal execution time, or weight, $w_j \geq 0$.

For comparability of the analyzed algorithms, we assume a homogeneous multi-processor architecture in the following with uniform cores. At any time, the speed s of a core can be set to any *eligible* value between $s_{min} > 0$ and $s_{max} \geq s_{min}$, and it is part of a scheduling algorithms decision to which speed to set the processor. It depends on the particular model, which values in $[s_{min}, s_{max}]$ are eligible; we consider the *continuous model*, in which any rational value is eligible, and the *discrete model*, which allows speeds only from a given finite set of speeds. A core that is set to speed s consumes power at the rate s^α , where $\alpha \geq 1$ is a small constant. The total energy consumed is the power consumption integrated over time.

In the continuous model, we may assume that a task is executed at a uniform speed. This follows directly from the convexity of the power function [27]. For discrete speeds, we add the restriction that a task has to run at a uniform speed. This is a reasonable assumption as in many processing environments it is not possible to change the processor speed during the execution of a task. If a task j of weight w_j is executed at speed $s_j \in [s_{min}, s_{max}]$, then the time to complete is $x_j = w_j/s_j$ and the energy consumed during the computation of j is

$$E_j = x_j \cdot s_j^\alpha = w_j \cdot s_j^{\alpha-1} = w_j^\alpha / x_j^{\alpha-1}. \quad (1)$$

We consider the following problem: given a deadline $D > 0$ and a node-weighted graph $G = (V, E, w)$, schedule all tasks in graph G and decide upon the processor speeds such that all tasks finish before the deadline D and the total energy consumption is minimized. If minimizing the energy consumption is intractable, we design approximation algorithms. An algorithm is called an r -approximation if it always computes a solution finishing before the deadline, with an energy consumption being at most r times the minimal energy consumption.

In our investigations we distinguish two problem classes of different complexity:

- SPEEDSCALING: we are given the mapping of each task to its core and the order in which each core executes the tasks mapped to it (encoded in G). The problem is then equivalent to minimizing the *critical path* of the graph G . That is, find speeds such that the total execution time of the longest path (w.r.t. execution times x_j) is minimized.
- SPEEDS&SCHEDULING: in addition to selecting the speeds at which each task should be executed, we provide a schedule for the tasks, i.e., we determine the core and the starting time for each task.

3 Continuous speeds

We consider the setting in which each core can be set to any rational speed value in the given interval $[s_{min}, s_{max}]$.

3.1 SpeedScaling Problem

As mentioned earlier, this problem is equivalent to determining the speeds such as to minimize the *critical path* of the graph G . This problem has been studied to some extent before. We summarize relevant known algorithms and provide new ones. We present two algorithms:

1. an optimal polynomial time algorithm CVX-SPEED which relies on a convex programming formulation inspired by the idea of Bampis et al. [4];
2. a linear-time algorithm SPG-SPEED for a special graph class, namely Series-Parallel Graphs, which are very common in practice. Our algorithm is a small modification of an algorithm in [3, 11, 20] and it computes an exact optimal solution when there is no limitation in the speeds. Our experiments show that this limitation is not prohibitive in our context.

Details on the algorithms follow below. The experimental evaluation is presented in Section 5.

3.1.1 CVX-speed

We provide a convex programming formulation with linear constraints that computes the exact solution for the energy minimization problem in the SPEEDSCALING setting. Such programs can be solved in polynomial time up to an arbitrary precision [19] with the Ellipsoid method. The formulation is inspired by a convex program for makespan minimization by Bampis et al. [4].

Each task j is associated to a constant speed s_j . The variable x_j represents the processing time of Task j in the solution, which is equal to w_j/s_j . The variable d_j represents the completion time of task j .

$$\min \sum_{j \in V} \frac{w_j^\alpha}{x_j^{\alpha-1}} \quad (2)$$

$$\text{s.t. } d_j \leq D, \quad \forall j \in V \quad (3)$$

$$x_j \leq d_j, \quad \forall j \in V \quad (4)$$

$$d_j + x_k \leq d_k, \quad \forall (j, k) \in E \quad (5)$$

$$w_j/s_{max} \leq x_j \leq w_j/s_{min}, \quad \forall j \in V. \quad (6)$$

The first three constraints ensure that tasks are executed one after the other, without preemption, respecting the precedence constraints and meeting the deadline D . Constraint 6 ensures that the speed limits are respected. Finally, the objective function computes the energy consumption for the schedule that is to be minimized. For a computed solution of the convex program, the speed s_j for task j is implied by w_j/x_j . We therefore have the following result.

► **Theorem 1.** *CVX-SPEED computes an optimal solution in polynomial time.*

3.1.2 SPG-speed

In the most general definition by Lawler [14], series-parallel graphs (or SP-graphs) are defined recursively as being either a single task, the series composition of two graphs (noted $(G_1; G_2)$), or the parallel composition of two graphs (noted $(G_1 || G_2)$). In $(G_1; G_2)$, the tasks of G_2 cannot start before all tasks of G_1 have terminated. In $(G_1 || G_2)$, there exist no precedence constraints between the tasks of G_1 and G_2 .

In the context of minimizing the makespan of malleable jobs, an algorithm has been proposed and studied in [11, 20], and a similar algorithm has been used in our context in [3]. The principle of the algorithm is to define an *equivalent* task of a series and a parallel composition of two graphs. Specifically, if \mathcal{L}_G represents the equivalent weight of G , we have:

- $\mathcal{L}_{T_i} = w_i$
- $\mathcal{L}_{(G_1; G_2)} = \mathcal{L}_{G_1} + \mathcal{L}_{G_2}$
- $\mathcal{L}_{(G_1 || G_2)}^\alpha = \mathcal{L}_{G_1}^\alpha + \mathcal{L}_{G_2}^\alpha$

The problem of selecting the speeds for a graph G in order to minimize the energy consumption is then equivalent to the problem of selecting the speed for a unique task of weight \mathcal{L}_G . The minimum energy necessary to schedule a graph G under a deadline D is therefore equal to $\mathcal{L}_G^\alpha/D^{\alpha-1}$, using the speed \mathcal{L}_G/D , see Equation (1). In order to compute the speed at which each task has to be scheduled in such a solution, the algorithm SPG-SPEED associates a speed s to each subgraph:

- $s(G) = \mathcal{L}_G/D$
- In $(G_1; G_2)$, $s(G_1) = s(G_2) = s(G_1; G_2)$.
- In $(G_1 || G_2)$, $s(G_1) = s(G_1 || G_2) \mathcal{L}_{G_1} / \mathcal{L}_{(G_1 || G_2)}$.

This result however requires to use speeds arbitrarily large, so the solution found may not respect the speed bounds, as specified in the following theorem.

► **Theorem 2** ([3, 11, 20]). *Given an SP-graph and ignoring the constraints s_{min} and s_{max} , SPG-SPEED computes an optimal solution in linear time.*

3.2 Speeds&Scheduling Problem

Consider the setting in which an algorithm determines both, the speed allocation and the actual schedule including the mapping of tasks to cores. If the optimal solution requires to use the speed s_{max} for each task, then computing a schedule meeting a given deadline is already an NP-hard problem, as it is reducible to the classic $P|prec|C_{max}$ problem in the Graham three-field notation. The SPEEDS&SCHEDULING problem can therefore not have an approximation algorithm unless $P = NP$, as this includes computing a schedule meeting the given deadline. The best known scheduling algorithm for $P|prec|C_{max}$ is a 2-approximation [10], and cannot be improved under some complexity assumptions [24]. We therefore assume that the optimal solution uses speeds at most $s_{max}/2$, in order to focus on the problem of minimizing the energy and not on meeting the deadline, which is not the core of this paper. We show the following result.

► **Theorem 3.** *APX-SCHED is a $2^{\alpha-1}$ -approximation if the optimal solution uses speeds at most $s_{max}/2$.*

The main idea of the algorithm builds on work in [4] for the related problem of minimizing the makespan under a fixed energy budget. The algorithm consists of two steps: firstly, a convex program is solved for computing the optimal speeds in a particular relaxation. Secondly, we fix these speeds and run a greedy heuristic for assigning the tasks to cores. The convex programming relaxation is as follows (recall that m is the number of cores).

$$\min \sum_{j \in V} \frac{w_j^\alpha}{x_j^{\alpha-1}} \quad (7)$$

$$\text{s.t. } \sum_{j \in V} x_j/m \leq D/2 \quad (8)$$

$$d_j \leq D/2, \quad \forall j \in V \quad (9)$$

$$x_j \leq d_j, \quad \forall j \in V \quad (10)$$

$$d_j + x_k \leq d_k, \quad \forall (j, k) \in E \quad (11)$$

$$w_j/s_{max} \leq x_j \leq w_j/s_{min}, \quad \forall j \in V. \quad (12)$$

Given an optimal solution for this program, we fix the speeds for the tasks. In the second step of the algorithm, we schedule the tasks using a list scheduling algorithm proposed by Graham [10]. That is, we consider tasks in any topological ordering (i.e., respecting the given precedence order) and assign a task to the core with currently smallest last completion time. If the makespan C obtained is smaller than D , the speeds are then lowered by a factor C/D .

Proof of Theorem 3. For a fixed speed assignment let $V := \sum_{i \in V} x_i/m$ denote the *volume* and let L denote the length of the critical path in G . Both, volume and critical path, are well known lower bounds on the makespan. Graham's list scheduling [10] yields a makespan of at most $V + L$. The convex program computes a speed assignment that minimizes the energy among all speed assignments for which both the volume and the critical path are not larger than $D/2$. Hence, Graham's list scheduling achieves a schedule where all tasks complete before $V + L \leq D$ and, thus, all tasks meet the deadline.

On the other hand, one can show that the energy consumed by this schedule is at most a factor $2^{\alpha-1}$ larger than the optimal. Indeed, consider an optimal schedule of makespan D using speeds at most $s_{max}/2$, and multiply every speed by 2. We obtain a speed assignment which is a solution to the convex program above, and whose energy cost is a factor $2^{\alpha-1}$ away from the optimal. As the speed assignment computed by the algorithm minimizes the objective function, its energy cost is not larger. ◀

In Section 5, we will show that on real-world instances, the solution quality is substantially better than the one guaranteed in Theorem 3 above. Finally, we remark that the problem is computationally highly intractable. Even for a given speed assignment, it is NP-complete to compute an optimal schedule even if all tasks have unit execution time [26] or if there are no precedence relations [9].

4 Discrete speeds

Consider the setting in which each core can run at $k \in \mathbb{N}$ possible speeds v_1, v_2, \dots, v_k with $v_i < v_{i+1}$. Let the maximum ratio of speeds be $r = \max_i v_{i+1}/v_i$. Note that the mapping problem in this setting is already NP-hard even with $k = 2$ [3]. However, the more general model in which speed modifications are allowed during the execution admits a polynomial exact algorithm [3]. We also underline that the approximation ratios given in this section still hold if the optimal solution is allowed to use any rational speed in the interval $[v_1; v_k]$.

4.1 SpeedScaling problem

Assume the task-to-core assignment is given and we need to determine the speeds such as to minimize the *critical path* of the graph G . We present two algorithms: (1) an optimal exponential time algorithm ILP-D-SPEED based on an integer linear programming (ILP) formulation, (2) a polynomial time algorithm APX-D-SPEED that solves a convex program within an approximation factor $r^{\alpha-1}$.

4.1.1 ILP-D-speed

We define nk boolean variables $y_{i,\ell}$ which are equal to 1 if task i runs at speed v_ℓ and to 0 otherwise, and consider the following program similar to the convex program (2)-(6). The main difference is that the execution time of a task i is now equal to $\sum_{\ell \leq k} \frac{w_i}{v_\ell} y_{i,\ell}$ and its energy consumption is equal to $\sum_{\ell \leq k} w_i v_\ell^{\alpha-1} y_{i,\ell}$.

$$\text{minimize } \sum_{i \in V} w_i \sum_{\ell \leq k} v_\ell^{\alpha-1} y_{i,\ell} \quad (13)$$

$$d_i \leq D \quad \forall i \in V \quad (14)$$

$$\sum_{\ell \leq k} \frac{w_i}{v_\ell} y_{i,\ell} \leq d_i \quad \forall i \in V \quad (15)$$

$$d_i + \sum_{\ell \leq k} \frac{w_j}{v_\ell} y_{j,\ell} \leq d_j \quad \forall (i, j) \in E \quad (16)$$

$$\sum_{\ell \leq k} y_{i,\ell} = 1 \quad \forall i \in V \quad (17)$$

$$\forall \ell \leq k \quad y_{i,\ell} \in \{0, 1\} \quad \forall i \in V. \quad (18)$$

The correctness of this ILP formulation therefore follows from the correctness of CVX-SPEED.

► **Theorem 4.** ILP-D-SPEED computes an optimal solution in exponential time.

In general, integer linear programs cannot be solved in polynomial time. However, our experiments show that on the datasets considered (up to 1000 tasks), this algorithm is at most 5 times slower than the polynomial-time algorithm CVX-SPEED.

4.1.2 APX-D-speed

The following algorithm is inspired by [3]. In a first step, we compute optimal *continuous* speeds \bar{s}_j for each task j . This is done by running the fast algorithm SPG-SPEED, and, in case this algorithm does not succeed (e.g., the SP-graph restriction is not met), we solve the convex program (2)-(6) (algorithm CVX-SPEED) with $s_{min} = v_1$ and $s_{max} = v_k$. Then, we run each task j at the speed s_j that is equal to the smallest speed v_i such that $v_i \geq \bar{s}_j$.

► **Theorem 5.** APX-D-SPEED computes an $r^{\alpha-1}$ -approximate solution in polynomial time.

Proof. Consider a speed setting computed by the algorithm. Observe that the tasks respect the deadlines as the speeds s_j are not smaller than the speeds \bar{s}_j that gave a valid solution. Let OPT be the energy consumed in an optimal solution. First, note that the energy consumed by executing each task at speed \bar{s}_j is not larger than OPT . The algorithm runs each task j at speed s_j , consuming an energy $w_j s_j^{\alpha-1}$. The total energy consumed is then:

$$E = \sum_j w_j s_j^{\alpha-1} \leq \left(\frac{s_j}{\bar{s}_j}\right)^{\alpha-1} \sum_j w_j \bar{s}_j^{\alpha-1} \leq r^{\alpha-1} OPT. \quad \blacktriangleleft$$

4.2 Speeds&Scheduling problem

In this setting, an algorithm determines both, the speed allocation and the actual schedule including the mapping of tasks to cores. We present two algorithms: (1) an optimal exponential time algorithm ILP-D-SCHED based on solving an ILP, (2) a polynomial time algorithm APX-D-SCHED that solves a convex program within approximation factor $(2r)^{\alpha-1}$.

4.2.1 ILP-D-sched

We extend the ILP (13)-(18) by adding nm boolean variables $z_{i,c}$ equal to 1 if task i is executed on core c and to 0 otherwise, as well as n^2 variables $e_{i,j}$ indicating if task i has to be scheduled before task j . In particular, if two tasks are executed on the same core, then either $e_{i,j}$ or $e_{j,i}$ equals 1.

$$\text{minimize } \sum_{i \in V} w_i \sum_{\ell \leq k} v_\ell^{\alpha-1} y_{i,\ell} \quad (19)$$

$$d_i \leq D \quad \forall i \in V \quad (20)$$

$$\sum_{\ell \leq k} \frac{w_i}{v_\ell} y_{i,\ell} \leq d_i \quad \forall i \in V \quad (21)$$

$$d_i + \sum_{\ell \leq k} \frac{w_j}{v_\ell} y_{j,\ell} \leq d_j + D(1 - e_{i,j}) \quad \forall i, j \in V \quad (22)$$

$$\sum_{\ell \leq k} v_{i,\ell} = 1 \quad \forall i \in V \quad (23)$$

$$v_{i,\ell} \in \{0, 1\} \quad \forall \ell \leq k, \forall i \in V \quad (24)$$

$$\sum_{c \leq m} z_{i,c} = 1 \quad \forall i \in V \quad (25)$$

$$z_{i,c} \in \{0, 1\} \quad \forall i \in V, \forall c \leq m \quad (26)$$

$$e_{i,j} \in \{0, 1\} \quad \forall i, j \in V \quad (27)$$

$$e_{i,j} = 1 \quad \forall (i, j) \in E \quad (28)$$

$$z_{i,c} + z_{j,c} - e_{i,j} - e_{j,i} \leq 1 \quad \forall i, j \in V, \forall c \leq m \quad (29)$$

Equation (22) ensures that task j is executed after task i if $e_{i,j} = 1$, and does not have any impact if $e_{i,j} = 0$, so the program returns the same result as ILP-D-SPEED on a graph where the edges are represented by the variables $e_{i,j}$. The second important constraint is Equation (29), which ensures that if two tasks belong to the same core, either $e_{i,j}$ or $e_{j,i}$ equals 1. Therefore, a valid valuation of the variables $e_{i,j}$ corresponds to a directed graph which contains the edges of E , and which contains an edge between any two tasks that are placed on the same core (by the variables $z_{i,c}$). This corresponds to a valid input to the ILP-D-SPEED programming, so we have the following result.

► **Theorem 6.** ILP-D-SCHED computes an optimal solution in exponential time.

4.2.2 APX-D-sched

This algorithm combines the ideas of APX-SCHED and APX-D-SPEED: assuming again that the optimal solution uses speeds at most $v_k/2$, solve the convex program of APX-SCHED in order to associate each task to a speed $\bar{s}_j \in [v_1; v_k]$. Then, the algorithm runs each task j to the speed s_j equal to the smallest speed v_i such that $v_i \geq \bar{s}_j$, and schedules the tasks using a list scheduling algorithm.

► **Theorem 7.** APX-D-SCHED computes a $(2r)^{\alpha-1}$ -approximate solution in polynomial time if the optimal solution uses speeds at most $v_k/2$.

Proof. We first note that, similarly to the APX-D-SPEED case, the energy used by the schedule obtained by APX-D-SCHED is at most a factor $r^{\alpha-1}$ away from the energy used by the APX-SCHED solution. Then, assuming that the optimal solution uses speeds at most $v_k/2$, we know that the energy used by the APX-SCHED solution is within a factor $2^{\alpha-1}$ of the optimal energy consumption. Combining these two results completes the proof. ◀

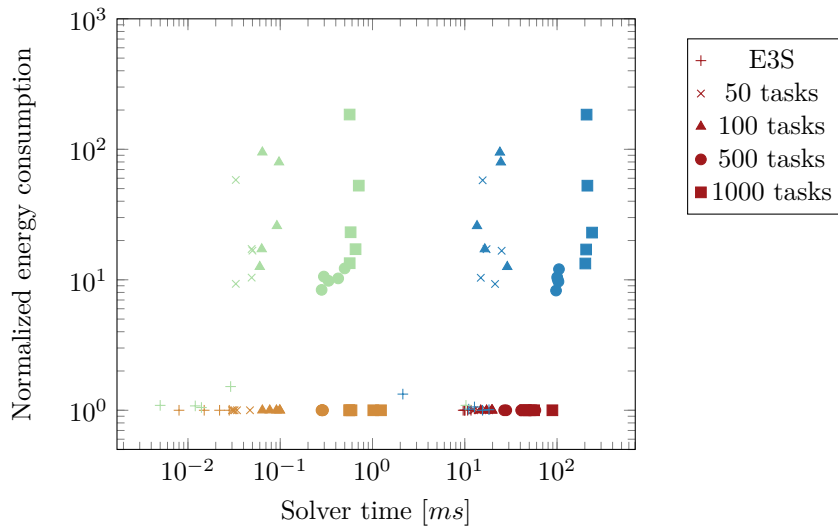
5 Experimental Results

In order to evaluate the quality of the presented approaches, we use a total of 5×5 benchmark graphs, i.e., five groups of five graphs of similar size. Our 5 smallest graphs are comprised of around 10 tasks and are derived from the Embedded System Synthesis Benchmarks Suite (E3S) [8]. These instances target processors of maximum frequency 250MHz, with a minimum frequency equal to 0.1MHz. 20 eligible speeds can be selected equally distributed between these limits. The deadlines associated to these graphs equal a few milliseconds, and are rather tight: several tasks need to be run at the maximum frequency. For larger graphs with 50, 100, 500, and 1000 tasks each, we selected graphs from the GENOME dataset of the Pegasus library [1]. The homogeneous processors used here were specified at a maximum frequency equal to 1.0GHz and again 20 equidistant speed setting, but assumed looser deadlines. All benchmarks belong to the class of SP-graphs, thus allowing the application of SPG-SPEED.

The benchmarks are executed on an Intel(R) Core(TM) i7-4770 CPU running at 3.40 GHz with 32 GiB of RAM using Ubuntu 18.04 LTS as underlying OS. To solve the ILPs for the ILP-D-SCHED and ILP-D-SPEED approaches, we use CPLEX 12.6 with a running time deadline of 5s. For the convex programs used by the CVX-SPEED and APX-D-SPEED approaches, we used MOSEK 8.1.

Figure 1 presents the results for the SPEEDSCALING problem, both in the continuous (CVX-SPEED and SPG-SPEED) and discrete speed (ILP-D-SPEED and APX-D-SPEED) settings.

Our first observation from the experiments is that the algorithm SPG-SPEED can be applied to all problem instances computing an optimal solution except for one single E3S graph instance where the prescribed speed limits were not respected. Moreover, it is really

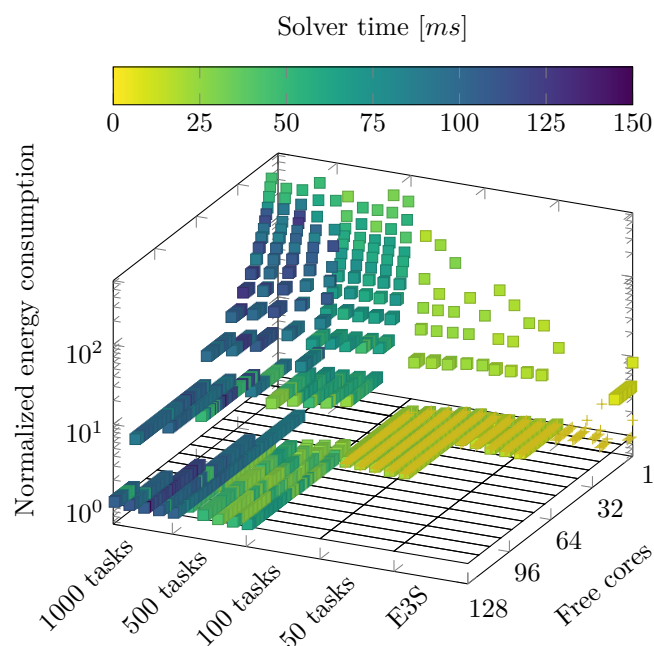


■ **Figure 1** Depicted above are the trade-offs between solver run-time and energy used by the solution for the four approaches – ■ CVX-SPEED, ■ SPG-SPEED, ■ ILP-D-SPEED, and ■ APX-D-SPEED – that assume that mapping and scheduling is already given (only speed assignment). These trade-offs have been determined for five classes of five benchmark graphs each from the E3S benchmark suite and the Pegasus library. The energy displayed by the minimal energy consumed with continuous speeds.

fast, needing at most 0.02ms for each of the five E3S graphs and 1ms only for the largest graphs with 1000 tasks. It can therefore be applied at runtime even for problems with very small and tight deadlines. As a consequence, the algorithm APX-D-SPEED runs at a comparable speed, except for the one instance which is not solved by SPG-SPEED. Even solving optimally the convex program (CVX-SPEED) is possible in less than 10 ms for the E3S benchmarks, 15ms for 100-tasks graphs, but may be unaffordable for very large graphs (in average 60ms for 1000 tasks). When solving the ILP for discrete speeds, the solver time can even increase to 200ms for the largest graphs, but we do not observe an exponential growth for this dataset, contrarily to the worst-case theoretical complexity. Surprisingly, the quality of the solution of APX-D-SPEED is only a few percents away from the optimal discrete solution (ILP-D-SPEED). Therefore, APX-D-SPEED can obtain near-optimal results two orders of magnitude faster than by solving the ILP, on SP-graph instances. The restriction to the discrete speed model implies a higher increase in energy consumption for the GENOME dataset. This can be explained by the fact that the deadlines are looser, so the optimal continuous speeds are lower, and being forced to select a discrete speed incurs higher losses.

Figure 2 presents the results of the APX-D-SCHED algorithm that performs also task-to-core assignment and scheduling apart from speed selection. From the color code, it can be seen that the solver times are roughly equal the ones of the APX-SCHED algorithm. For each of the 25 graphs, the number of cores has been varied between 1 and 128. In each design point, the energy of the found solution has been normalized to the optimal energy for the discrete speed case with no core constraints as determined by the ILP-D-SPEED approach.

It can be seen that APX-D-SCHED is able to solve many instances of graphs with 50 to 100 tasks in less than 25ms. However, it does not find a solution for 4 out of 5 E3S graphs because of the tightness of deadlines assumed in these benchmarks and the assumptions made in Theorem 7. Finally, we omit to present and compare the solver times of ILP-D-SCHED as these start in the range of minutes even for the smallest and easiest problem instances. Hence, we conclude this approach to be of no use to be applied on an MPSoC at run-time.



■ **Figure 2** Consumed energy of the 5×5 benchmark graphs for solutions found by the APX-D-SCHED approach (squares) subject to a fixed number of available (free) cores ranging from 1 to 128. The results are normalized: a value of 1 corresponds to the case with optimum discrete speeds and infinitely many cores. The required solver time to find these solutions ranges from 7 ms (■) to 150 ms (■) according to the given color key. The crosses denote optimal-energy solutions as determined by the ILP-D-SCHED approach.

6 Conclusions

We have shown that for many task graphs of real-world applications, the graph structure allows to determine energy-optimal speed assignments in the range of a ms given real-time constraints by applying an algorithm called SPG-SPEED in case tasks have been mapped already to cores. For the more complex problem of additionally determining the task-to-core assignment and schedule of tasks on these cores, even problem instances with few tasks cannot be practically solved optimally at runtime. Yet here, approximation algorithms have been analyzed and shown to offer affordable solving times to determine at least solutions with provable guarantees on the solution quality.

In the future, we would like to extend our analysis of the ties between theory and practice from homogeneous MPSoCs to systems with more diverse and complex communication architectures. Moreover, the presented set of algorithms shall be integrated into a framework for run-time resource management on many-core systems that are required to stay within given bounds on execution time, energy and also other user-specific requirement corridors.

References

- 1 Epigenomics dataset from the Pegasus library. <https://confluence.pegasus.isi.edu/display/pegasus/WorkflowGenerator>. [Online; accessed 02-September-2019].
- 2 Susanne Albers. Energy-efficient algorithms. *Communications of the ACM*, 53(5):86–96, 2010.
- 3 Guillaume Aupy, Anne Benoit, Fanny Dufossé, and Yves Robert. Reclaiming the energy of a schedule: models and algorithms. *Concurrency and Computation: Practice and Experience*, 25(11):1505–1523, 2013.

- 4 Evripidis Bampis, Dimitrios Letsios, and Giorgio Lucarelli. A note on multiprocessor speed scaling with precedence constraints. In *Proceedings of the 26th ACM Symposium on Parallelism in Algorithms and Architectures (SPAA)*, pages 138–142, 2014.
- 5 Gang Chen, Kai Huang, and Alois Knoll. Energy optimization for real-time multiprocessor system-on-chip with optimal DVFS and DPM combination. *ACM Trans. Embedded Comput. Syst.*, 13:111:1–111:21, 2014. doi:10.1145/2567935.
- 6 Pravanjan Choudhury, P. P. Chakrabarti, and Rajeev Kumar. Online Dynamic Voltage Scaling using Task Graph Mapping Analysis for Multiprocessors. In *20th International Conference on VLSI Design*, pages 89–94, 2007.
- 7 Pepijn J. de Langen and Ben H. H. Juurlink. Leakage-Aware Multiprocessor Scheduling. *Signal Processing Systems*, 57(1):73–88, 2009. doi:10.1007/s11265-008-0176-8.
- 8 R. Dick. Embedded System Synthesis Benchmarks Suite (E3S). <http://ziyang.eecs.umich.edu/~dickrp/e3s/>. [Online; accessed 02-September-2019].
- 9 M.R. Garey and D.S. Johnson. Strong NP-completeness results: motivation, examples, and implications. *J. Assoc. Comput. Mach.*, 25(3):499–508, 1978.
- 10 R. L. Graham. Bounds for certain multiprocessing anomalies. *The Bell System Technical Journal*, 45(9):1563–1581, November 1966. doi:10.1002/j.1538-7305.1966.tb01709.x.
- 11 Abdou Guermouche, Loris Marchal, Bertrand Simon, and Frédéric Vivien. Scheduling trees of malleable tasks for sparse linear algebra. In *European Conference on Parallel Processing*, pages 479–490. Springer, 2015.
- 12 Sandy Irani and Kirk Pruhs. Algorithmic problems in power management. *SIGACT News*, 36(2):63–76, 2005.
- 13 Woo-Cheol Kwon and Taewhan Kim. Optimal voltage allocation techniques for dynamically variable voltage processors. *ACM Trans. Embedded Comput. Syst.*, 4(1):211–230, 2005.
- 14 Eugene L Lawler. Sequencing jobs to minimize total weighted completion time subject to precedence constraints. In *Annals of Discrete Mathematics*, volume 2, pages 75–90. Elsevier, 1978.
- 15 Keqin Li. Performance Analysis of Power-Aware Task Scheduling Algorithms on Multiprocessor Computers with Dynamic Voltage and Speed. *IEEE Trans. Parallel Distrib. Syst.*, 19(11):1484–1497, 2008. doi:10.1109/TPDS.2008.122.
- 16 Minming Li and F. Frances Yao. An Efficient Algorithm for Computing Optimal Discrete Voltage Schedules. *SIAM J. Comput.*, 35(3):658–671, 2005.
- 17 Nicole Megow and José Verschae. Dual Techniques for Scheduling on a Machine with Varying Speed. *SIAM J. Discrete Math.*, 32(3):1541–1571, 2018.
- 18 Andrew Nelson, Orlando Moreira, Anca Mariana Molnos, Sander Stuijk, Ba Thang Nguyen, and Kees Goossens. Power Minimisation for Real-Time Dataflow Applications. In *14th Euromicro Conference on Digital System Design, Architectures, Methods and Tools (DSD 2011)*, pages 117–124, 2011.
- 19 Yurii Nesterov and Arkadii Nemirovskii. *Interior Point Polynomial Algorithms in Convex Programming*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 1994.
- 20 G. N. Srinivasa Prasanna and Bruce R. Musicus. Generalized Multiprocessor Scheduling and Applications to Matrix Computations. *IEEE TPDS*, 7(6):650–664, 1996. doi:10.1109/71.506703.
- 21 Kirk Pruhs, Rob van Stee, and Patchrawat Uthaisombut. Speed scaling of tasks with precedence constraints. *Theory of Computing Systems*, 43(1):67–80, 2008.
- 22 Dongkun Shin and Jihong Kim. Power-aware scheduling of conditional task graphs in real-time multiprocessor systems. In *Proceedings of the 2003 International Symposium on Low Power Electronics and Design, 2003, Seoul, Korea, August 25-27, 2003*, pages 408–413, 2003.
- 23 Amit Kumar Singh, Anup Das, and Akash Kumar. Energy optimization by exploiting execution slacks in streaming applications on multiprocessor systems. In *The 50th Annual Design Automation Conference 2013 (DAC 2013)*, pages 115:1–115:7, 2013.
- 24 Ola Svensson. Hardness of Precedence Constrained Scheduling on Identical Machines. *SIAM J. Comput.*, 40(5):1258–1274, 2011.

- 25 Umair Ullah Tariq and Hui Wu. Energy-Aware Scheduling of Periodic Conditional Task Graphs on MPSoCs. In *Proceedings of the 18th International Conference on Distributed Computing and Networking*, page 13, 2017.
- 26 J.D. Ullman. NP-complete scheduling problems. *J. Comput. System Sci.*, 10:384–393, 1975.
- 27 F. Frances Yao, Alan J. Demers, and Scott Shenker. A Scheduling Model for Reduced CPU Energy. In *Proc. of the 36th Annual Symposium on Foundations of Computer Science (FOCS 1995)*, pages 374–382, 1995.