Worst-case throughput analysis for parametric rate and parametric actor execution time scenario-aware dataflow graphs


DOI:
10.4204/EPTCS.145.7

Document status and date:
Published: 01/01/2014

Document Version:
Publisher’s PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:
• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:
www.tue.nl/taverne

Take down policy
If you believe that this document breaches copyright please contact us at:
openaccess@tue.nl
providing details and we will investigate your claim.

Download date: 13. Oct. 2023
Worst-case Throughput Analysis for Parametric Rate and Parametric Actor Execution Time Scenario-Aware Dataflow Graphs

Mladen Skelin  
Norwegian University of Science and Technology  
Trondheim, Norway

Marc Geilen  
Eindhoven University of Technology  
Eindhoven, The Netherlands

Francky Catthoor  
IMEC vzw.  
Leuven, Belgium

Sverre Hendseth  
Norwegian University of Science and Technology  
Trondheim, Norway

Scenario-aware dataflow (SADF) is a prominent tool for modeling and analysis of dynamic embedded dataflow applications. In SADF the application is represented as a finite collection of synchronous dataflow (SDF) graphs, each of which represents one possible application behaviour or scenario. A finite state machine (FSM) specifies the possible orders of scenario occurrences. The SADF model renders the tightest possible performance guarantees, but is limited by its finiteness. This means that from a practical point of view, it can only handle dynamic dataflow applications that are characterized by a reasonably sized set of possible behaviours or scenarios. In this paper we remove this limitation for a class of SADF graphs by means of SADF model parametrization in terms of graph port rates and actor execution times. First, we formally define the semantics of the model relevant for throughput analysis based on (max,+)-linear system theory and (max,+)-automata. Second, by generalizing some of the existing results, we give the algorithms for worst-case throughput analysis of parametric rate and parametric actor execution time acyclic SADF graphs with a fully connected, possibly infinite state transition system. Third, we demonstrate our approach on a few realistic applications from digital signal processing (DSP) domain mapped onto an embedded multi-processor architecture.

1 Introduction

Synchronous dataflow (SDF) [19] was introduced as a restriction of Kahn process networks (KPN) [18] to allow compile-time scheduling. The term synchronous means static or regular. Synchronous dataflow graphs (SDFGs) are directed graphs where nodes are called actors and edges are called channels. The numbers of data samples produced or consumed are known at compile time. We refer to these data samples as tokens and to the token production and consumption numbers as rates. Although SDF is very fitted to model regular streaming applications, it is due to its static nature, very lacking in its ability to capture the dynamic behaviour of modern streaming applications. Therefore, a notable number of SDF extensions has been proposed over the years. Cyclo-static dataflow (CSDF) [6] allows token production and consumption to vary between actor firings as long as the variation forms a certain type of a periodic pattern, while models such as parametrized synchronous dataflow (PSDF) [5], variable-rate dataflow (VRDF) [23], variable-rate phased dataflow (VPDF) [23] and schedulable parametric dataflow (SPDF) [10] introduce parametric rates. Scenario-aware dataflow (SADF) [22] encodes the dynamism of an application by identifying a finite number of different behaviours called modes or scenarios. Each of the modes is represented by a single synchronous dataflow graph. The modes or scenarios can occur in known or unknown sequences. A finite state machine (FSM) is used to encode occurrence patterns.
SADF is equipped with a technique that yields the tightest possible performance guarantees [12]. The power of this technique lies in its ability to consider transitions over all possible scenario sequences as given by the FSM. Considering only the worst-case scenario, i.e. the scenario with the lowest throughput, without considering scenario transitions could be too optimistic. On the other hand, merging all application SDFGs into one SDFG where an actor takes the worst-case execution time over all SDFGs in SADF would be too pessimistic. This is due to the fact that subsequent iterations belonging to different scenarios may overlap in time, i.e. execute in a pipelined fashion. However, SADF is limited by its finiteness. It can only handle a reasonably sized set of application scenarios.

To illustrate this, let us define an abstract parallel application consisting of a nested for loop with parametric affine loop bounds:

```csharp
ProcessData.A(out g, out h);

for (i=0; i<=g; i++){
    for (j=0; j<=h; j++){
        // Perform two tasks in parallel
        #region ParallelTasks
        // Perform two tasks in parallel
        Parallel.Invoke(() =>
            {
                ProcessData.B(i,j);
            }, // close first parallel action
        () =>
        {
            ProcessData.C(i,j);
        }) // close second parallel action
    // close Parallel.Invoke
    #endregion

    ProcessData.D(i,j);
}
```

The example application consists of 4 subtasks: ProcessData.A, ProcessData.B, ProcessData.C and ProcessData.D with known worst-case execution times. Data parallelism is elegantly specified using the Parallel.Invoke construct. Inside the Parallel.Invoke construct, an Action delegate is passed for each item of work. The application is mapped onto a multi-processor platform. The task assignment employed is purely static. In order to add complexity, we assume that the application executes in a pipelined fashion, i.e. more instances of the application can be active at the same time. Such an assumption introduces resource dependencies over subsequent activations of the application. In other words, a subtask of the \((i+1)\)th activation of the application might have to wait for a certain subtask of the \(i\)th activation to complete and release the corresponding processing element. As specified by the example code, \(g\) and \(h\) can take different values during each application execution, i.e. they are data-dependent and are the result of input data processing performed by the subtask ProcessData.A. Let us assume we know that \(g\) can take the value from the interval \([0, \frac{q}{2}]\) and \(h\) can take the value from the interval \([0, \frac{m}{2}]\). In that case, from a pure timing perspective, this application will exhibit as many behaviours as there are integer points in the rational 2-polytope \(P_{n,m}\) given by the set of constraints.
\{0 \leq \frac{1}{n}, 0 \leq \frac{1}{m}\}. For n = 4500 and m = 2001, to be able to use SADF to derive the tightest worst-case performance bounds, even for such a simple application executing in a pipelined fashion on a multi-processor platform, we would have to generate 2,252,126 SDFGs \[8\]. The situation gets even worse when dealing with platforms that support dynamic voltage and frequency scaling (DVFS), which is a commonly used technique that adapts both voltage and frequency of the system in respect to changing workloads \[20\]. In this case also the execution times of the application subtasks would vary depending on the current DVFS setting of the processing element they are mapped to.

In our work, we will remove these limitations which hamper the use of SADF in important application domains. For this purpose, we will add parametrization to the basic SADF modeling approach both in terms of parametric rates and parametric actor execution times given over a parameter space, which is a totally non-trivial extension because the current core of the SADF framework relies strongly on the constant nature of the rates and actor execution times. We raise the problem of SADF parametrization in the scope of existing parametric dataflow models. PSDF \[5\] and SPDF \[10\] are two semantically very similar models that provide a high level of generalization. We prefer SPDF due to syntactical convenience. By incorporating SPDF semantics into the definition of our parametric rate and parametric actor execution time SADF (PSADF), we show that the SPDF model can at run-time be treated as a special case of a SADF. We then derive a technique for worst-case throughput analysis for PSADF. We demonstrate our approach on a few realistic applications from the digital signal processing (DSP) domain.

2 Related Work

Throughput analysis of SDFGs is studied by many authors. Reference \[15\] gives a good overview of the existing methods. Due to the static nature of SDF, these methods cannot be applied to any form of parametric dataflow. \[14\] presents three methods for throughput computation for an SDFG where actor execution times can be parameters. However, the technique does not consider parametric rates and can only handle the static case, i.e. the graph cannot change parameter values during its execution. \[12\] introduces the (max,+) semantics for the SADF model relevant for worst-case performance analysis, but is, as previously mentioned, practically limited to a reasonably sized set of scenarios. The most closely related work to ours can be found in \[9\]. It combines the approaches presented in \[12\] and \[14\] and yields a technique that finds throughput expressions for an SADF where actors can have parameters as their execution times. However, the (max,+) semantics introduced in \[9\] can consider only parametric actor execution times and not parametric rates. A straightforward extension of \[9\] to cover the case of parametric rates is not possible because it is not clear how to symbolically execute the graph in the presence of parametric rates. In the scope of rate parametric dataflow models \[5\][10], little attention has been given to the aspect of time. Two examples of parametric models that explicitly deal with time are VRDF \[23\] and VPDF \[23\]. These address the problem of buffer capacity computation under a throughput constraint, but both have a structural constraint that each production of \( p \) tokens must be matched by exactly one consumption of \( p \) tokens. That drastically limits the scope of applications it can consider.

So, the current approaches in throughput analysis for dataflow MoCs either cannot consider parametric rates \[15\][14][12][9], or impose too hard structural constraints that severely limit the expressivity of the model \[23\]. In our work we will remove these limitations by embedding the SPDF model \[10\] which provides a high level of generalization into the SADF model \[22\][12].
3 Preliminaries

3.1 Synchronous Dataflow Graphs

SDFG is a directed graph \((\mathcal{A}, \mathcal{E})\) where nodes represent actors which in turn represent functions or tasks, while edges represent their dependencies. We also refer to edges as channels. Execution of an actor is denoted as firing and it is assigned with a time duration. In SDF, the number of tokens consumed and produced by an actor is constant for each firing. We refer to these numbers as rates. Actors communicate using tokens sent over channels from one actor to another. Fig. 1a shows an example of an SDFG with 5 actors \((\mathcal{A} = \{A, B, C, D, E\})\) and 9 channels \((\mathcal{E} = \{(A, B), (B, C), (C, C), \ldots\})\). Some channels might contain initial tokens, depicted with solid dots. The example graph contains 5 initial tokens that are labeled \(t_1, \ldots, t_5\). Each actor is assigned with a firing time duration, denoted in the actor node, below the actor name, e.g. actor A has a firing duration of 29 time-units. Each port is assigned with a rate. When the value is omitted, it means that the value equals to 1. As rates in SDF are constant for each firing, it is possible to construct a finite schedule (if it exists) that can be periodically repeated [19]. Such a schedule assures liveness and boundedness [19]. We call such minimal sequence of firings an iteration of the SDFG. This is a sequence of firings that has no net effect on the token distribution in the graph. The numbers of firings of each actor within an iteration constitute the repetition vector of an SDFG. We only consider dataflow graphs that are bounded and live. Throughput is considered in terms of the number of iterations per time-unit, i.e. the number of iterations executed in one period normalized by the repetition vector divided by the duration of the period [15]. It is natural to do so, because an iteration represents a coherent set of calculations, e.g. decoding of a video frame. For more details we refer to [19][15].

3.2 \((\text{max,})\) Algebra for SDFGs

Let \(a \oplus b = \max(a, b)\), \(a \odot b = a + b\) for \(a, b \in \mathbb{R} = \mathbb{R} \cup \{-\infty\}\). By max-algebra we understand the analogue of linear algebra developed for the pair of operations \((\oplus, \odot)\) extended to matrices and vectors [4]. Let \(\vec{\gamma}\) denote the vector of production times of tokens that exist in their different channels in between iterations, i.e. it has an entry for each initial token in the graph. Then \(\vec{\gamma}_k\) denotes the vector of production times of initial tokens after \(k\) iterations of the graph. These vectors then can be found using \((\text{max,})\) algebra [4]. The evolution of the graph is then given by the following equation: \(\vec{\gamma}_{k+1} = G \vec{\gamma}_k\), where \(G = \{g_{ij}\}\) is a \((\text{max,})\) characteristic matrix of the graph. Entry \(g_{ij}\) specifies the minimal elapsed time from the production time of the \(j^{\text{th}}\) token in the previous iteration to the production time of the \(i^{\text{th}}\) token in the current iteration. When the \(i^{\text{th}}\) token is not dependent on the \(j^{\text{th}}\) token, then \(g_{ij} = -\infty\). The specification of the algorithm for obtaining \(G\) can be found in [13]. The \((\text{max,})\) characteristic matrix for the example SDFG in Fig. 1a takes the form:

\[
G = \begin{bmatrix}
29 & -\infty & -\infty & 29 & -\infty \\
33 & 4 & -\infty & 33 & -\infty \\
63 & -\infty & 30 & 63 & -\infty \\
-\infty & -\infty & -\infty & -\infty & 0 \\
64 & 5 & 31 & 64 & -\infty 
\end{bmatrix}
\]
For example, $\vec{\gamma}_1$ can be calculated as below:

$$\vec{\gamma}_1 = \begin{bmatrix}
29 & -\infty & -\infty & 29 & -\infty \\
33 & 4 & -\infty & 33 & -\infty \\
63 & -\infty & 30 & 63 & -\infty \\
-\infty & -\infty & -\infty & -\infty & 0 \\
64 & 5 & 31 & 64 & -\infty \\
\end{bmatrix} = \begin{bmatrix}
\max(29 + 0, 29 + 0) \\
\max(33 + 0, 4 + 0, 33 + 0) \\
\max(63 + 0, 30 + 0, 63 + 0) \\
\max(0 + 0) \\
\max(64 + 0, 5 + 0, 31 + 0, 64 + 0) \\
\end{bmatrix} = \begin{bmatrix}
29 \\
33 \\
63 \\
0 \\
64 \\
\end{bmatrix}.$$  

Paper [12] explains how to obtain the throughput of an SDFG from the matrix $G$. Briefly, matrix $G$ defines a corresponding (max,+) automaton graph (MPAG) [11]. MPAG has as many nodes as there are initial tokens in the graph. An edge with the weight $g_{ij}$ is created from the $j$th node to the $i$th if $g_{ij} \neq -\infty$. The maximum cycle mean (MCM) $\lambda$ of the MPAG identifies the critical cycle of the SDFG. The critical cycle limits the throughput of the SDFG which takes the value $1/\lambda$. MPAG of the example SDFG graph is displayed in Fig. 1b. The cycle with weights $g_{14} - g_{51} - g_{45}$ (denoted with bold arrows) determines the throughput which takes the value of $1/31$ iterations per time-unit.

### 3.3 Scenario-Aware Dataflow Graphs (SADFG)

SADF models the dynamism of an application in terms of modes or scenarios. Every scenario is modeled by an SDFG, while the occurrence patterns of scenarios are given by an FSM. We give the following definition of an SADFG.

**Definition 1.** A Scenario-aware dataflow graph (SADFG) is a tuple $\text{SADFG} = (S, F)$, where:

- $S = \{s_i \mid s_i = (\text{scen}_i, G_i)\}$ is a set of ordered pairs of scenarios and their corresponding SDFGs;
- $F = (Q, q_0, \delta, \Sigma, E)$ is the scenario finite state machine consisting of a finite set $Q$ of states, an initial state $q_0 \in Q$, a transition relation $\delta \subseteq Q \times Q$, a scenario labelling $\Sigma : Q \rightarrow S$ and a set of final states $E$, where $E = Q$.

Fig. 2a shows an example SADFG with two scenarios, $a$ and $b$. In this example both scenarios use the same scenario graph, but the actor execution times differ. For example, actor $A$ has a firing duration of 29 time-units in scenario $a$ and 28 time units in scenario $b$. The scenario FSM is fully connected and thus allowing arbitrary scenario order.

Every finite path of arbitrary length $\overline{q}$ over the FSM corresponds to a sequence $\overline{s}$ with $s(k) = \Sigma(\overline{q}(k))$. When the FSM performs a transition, the SDFG graph associated with the destination state is executed for
Worst-case Throughput Analysis for PSADF Graphs

(a) An example SADFG.
(b) MPAG of the example SADFG.

Figure 2: Scenario-aware dataflow

exactly one iteration. Let $G(s_i)$ denote the $n \times n$ (max,+) characteristic matrix for the scenario $s_i$, where $n$ is the number of initial tokens in the SADFG. Then the completion time of a $k$-long sequence of scenarios can then be defined as a sequence of (max,+) matrix multiplications $S(s_1 \ldots s_k) = G(s_k) \ldots G(s_1)^\vec{i}$, where $\vec{i}$ specifies the initial enabling times of the graph’s initial tokens and usually $\vec{i} = \vec{0}$. The worst case increase of $S(\vec{x})$ for a growing length of $\vec{x}$ specifies the worst-case throughput for any sequence of scenarios [11][12]. Reference [12] explains how to build the MPAG of an SADFG. Again, the inverse of the MCM $(1/\lambda)$ of the obtained MPAG denotes the worst-case throughput of that particular SADFG. A special case that arises in practice, which will be of the utmost importance in our SADF parametrization, is when scenarios can occur in arbitrary order, yielding the SADF FSM to be fully connected and with a single state for each scenario. In that case, the throughput of an SADFG equals to the maximum cycle mean of the MPAG that corresponds to the (max,+) matrix $G = \max_{q \in Q}(G(\Sigma(q)))$ [12]. The operator $\max$ denotes taking the maximum of the elements of the individual scenario matrices. The corresponding scenario matrices for the example SADFG in Fig. 2a are:

$$G(a) = \begin{bmatrix} 29 & -\infty & -\infty & 29 & -\infty \\ 33 & 4 & -\infty & 33 & -\infty \\ 63 & -\infty & 30 & 63 & -\infty \\ -\infty & -\infty & -\infty & -\infty & 0 \\ 64 & 5 & 31 & 64 & -\infty \end{bmatrix} \quad G(b) = \begin{bmatrix} 28 & -\infty & -\infty & 28 & -\infty \\ 34 & 6 & -\infty & 34 & -\infty \\ 72 & -\infty & 24 & 72 & -\infty \\ -\infty & -\infty & -\infty & -\infty & 0 \\ 82 & 16 & 34 & 82 & -\infty \end{bmatrix}.$$ 

The critical cycle of the corresponding MPAG obtained from the maximized matrix $G = \max(G(a), G(b))$, is denoted by bold arrows in Fig. 2b. Throughput in this case equals 1/37 iterations per time-unit. This example also demonstrates that the worst-case throughput value cannot simply be obtained by only considering the ‘worst-case’ scenario, or by analysing the graph where each actor takes its worst-case execution time over all scenarios.

4 Parametric Rate and Actor Execution Time SADF Analysis

We start this section by formally defining the PSADF model and showing the (max,+) equivalence between SADF and PSADF. We use this result in defining the PSADF worst-case throughput calculation problem as a constrained optimization problem over the PSADF graph (PSADFG) parameter space,
where the objective functions are elements of the symbolic PSADFG (max,+) characteristic matrix. We conclude by giving the theoretical foundation and the algorithm for symbolic PSADFG (max,+) characteristic matrix extraction.

4.1 Motivation and Model Definition

SADF becomes impractical or even infeasible when it faces applications with a vast set of possible behaviours. We overcome this limitation by parametrization. The problem of parametrization of a dataflow model in terms of rates is not an easy task as it raises questions about properties like liveness, boundedness and schedulability. A naive approach in just declaring any rate of interest as parametric, could render the graph to deadlock, be unbounded or unschedulable. Therefore we start from SPDF \cite{10}. The liveness and boundedness properties for SPDF are decidable. SPDF extends SDF by allowing rates to be parametric while preserving static schedulability. Rates are products of static natural numbers and/or parameters that can change dynamically. The changes of each parameter $p$ are made by a single actor called its modifier each $\alpha$th time it fires using ‘set $p[\alpha]$’ annotation. We re-define SPDF \cite{10} by adding the notion of time of SDF/SADF to it.

**Definition 2.** A schedulable parametric dataflow graph (SPDFG) is a tuple SPDFG = $(G, PR, PD, i, r, e, M, \alpha)$, where:

- $G$ is a directed connected graph $(\mathcal{A}, \mathcal{E})$ with $\mathcal{A}$ set of actors and $\mathcal{E} \subseteq \mathcal{A} \times \mathcal{A}$ set of edges (channels);
- $PR$ is a set of rate parameters (symbolic variables) used to define SPDF rates by the grammar $FR := k \mid pr \mid FR_1 \cdot FR_2$, where $pr \in PR$, $k \in \mathbb{N}^+$;
- $PD$ is a set of actor execution time parameters (symbolic variables) used to define SPDF actor execution times by the grammar $FD := k \cdot pd \mid FD_1 + FD_2$, where $pd \in PD$, $k \in \mathbb{R}_0^+$;
- $i : \mathcal{E} \rightarrow \mathbb{N}_0$ returns for each edge channel its number of initial tokens;
- $r : \mathcal{A} \times \mathcal{E} \rightarrow FR$ returns for each port (represented by an actor and one of its edges) its rate;
- $e : \mathcal{A} \rightarrow FD$ returns for each actor its execution time;
- $M : PR \rightarrow A$ and $\alpha : PR \rightarrow FR$ returns for each rate parameter its modifier and its change period.

We consider only live SPDFGs as defined in \cite{10}. We allow parameters (rates and actor execution times) to change in between iterations. The introduction of parametric actor execution times to SPDF does not influence the liveness property. We define actor execution times as linear combinations of parameters. This gives us the ability to encode dependence, e.g. in case two actors are mapped onto the same processor, the ratio of their execution times will always be constant within an iteration.

Fig. 3a shows an example of a SPDF graph where actors have parametric $(p, q, s)$ or constant rates and parametric execution times $(a, b, c, d, e)$. Parametric rates $p$ and $s$ are modified by the actor $A$ every time it fires, while the parametric rate $q$ is modified by the actor $B$ every $p$th time it fires.

Now we can define our parametric SADF model, by subjecting SPDF to the operational semantics of SADF.

**Definition 3.** A parametric rate and parametric actor execution time SADFG (PSADFG) is a tuple PSADFG = $(G, \Omega, F)$, where:

- $G$ is a live SPDFG;
The worst-case throughput of a PSADFG for which Proposition 1.

Proposition 1. The worst-case throughput of a PSADFG for which \( \delta = Q \times Q \) and for which exists a bijective mapping \( z : Q \rightarrow \Omega \) equals to the inverse of the maximum cycle mean of the MPAG defined by the matrix \( G = \max_{q \in Q} (G(z(q))) \).

Proof. Given the operational semantics of PSADFG previously described and the fact that \( \Omega \) is bounded and closed, it follows straightforwardly from \([12][11]\).

4.2 Worst-Case Throughput Analysis

4.2.1 Problem Definition.

Given \( G(p) = \{ g_{ij}(p) \} \) as a matrix of continuous function over the closed and bounded parameter space \( \Omega \) that possesses an appropriate mathematical formulation, e.g. as equalities and inequalities over a certain \((|\mathcal{P}| + |\mathcal{R}|)\)-dimensional vector space, using Proposition 1 our worst-case throughput calculation problem becomes a set of maximally \((n \times n)\) constrained optimization problems with \( G(p) = \{ g_{ij}(p) \} \) as the objective function(s) and \( \Omega \) as the constraint set:
(a) An example SPDFG/PSADFG.

(b) PSADF actor model.

Figure 3: Parametric SADF

\[
\text{foreach } (i, j) \text{ s.t. } g_{ij}(\vec{p}) \neq -\infty \text{ do}
\]

\[
\begin{align*}
\text{maximize} & \quad g_{ij}(\vec{p}) \\
\text{subject to} & \quad \vec{p} \in \Omega.
\end{align*}
\]

A continuous function over a bounded and closed set admits a maximum. Of course, the term continuous includes also discrete functions that are continuous in the Heine sense. After maximizing all the element functions of \( G(\vec{p}) \), the worst-case throughput will equal to the MCM of the MPAG given by the maximized PSADFG \((\max,+)^+\) characteristic matrix. Our main challenge is thus to derive a technique for the analytical formulation of the symbolic PSADFG \((\max,+)^+\) characteristic matrix \( G(\vec{p}) \). \( G(\vec{p}) \) is a matrix of functions that in the \((\max,+)^+\) sense encodes the time distances between initial tokens in adjacent iterations of a PSADFG. We will show that this is a matrix of polynomial functions of \( \vec{p} \). Polynomial functions are continuous. Then the problem can be solved as a polynomial programming problem over \( \Omega \). There exists a variety of techniques for solving such problems depending on the ‘shape’ of \( \Omega \). Do note here that these optimization problems are solved independently as we are interested in the worst-case increase of \( A(\vec{p}) \) for a growing length of \( \vec{p} \) (over a growing number of iterations).

4.2.2 \((\max,+)^+\) Algebra for PSADF.

In PSADF we only allow parameters to change between graph iterations, i.e. \( \#M(pr_j) = 1 \) for parametric rates in the context of SPD. The same goes for parametric actor execution times. Currently, our \( G(\vec{p}) \) extraction technique requires that the considered PSADFG is ‘acyclic within an iteration’. If we take a PSADFG and convert it to a directed acyclic graph (PSADFG-DAG) by removing the edges with initial tokens, we require that only the PSADFG-DAG sink actors can produce tokens on the removed edges, and only the PSADFG-DAG source actors can consume from those edges. We do not include self-edges in this restriction. That is to say that we only allow cyclic dependencies tied to one actor. However, we can still consider PSADFGs that are serial compositions of subgraphs that are ‘acyclic within an iteration’ if the subgraph performs only one iteration during an iteration of the composite PSADFG. Our \( G(\vec{p}) \) extraction process will depend on the PSADFG quasi-static schedule which can be obtained using the procedure from [10]. Basically, the PSADFG-DAG is sorted topologically. Result of the topological sorting is a string of actors. For PSADFG in Fig. 3a this string equals to \( ABCDE \). Now we replace every actor \( X \) with \( X^\#X \), where \( \#X \) is the PSADFG repetition vector entry for actor \( X \). For PSADFG in Fig. 3a the final quasi-static schedule takes the form \( AB^pC^qD^rE \).

We continue by giving an appropriate \((\max,+)^+\) model of the PSADF actor as displayed in Fig. 3b.
First let us briefly explain the (max,+) semantics of a dataflow actor firing. If $T$ is the set of tokens needed by an actor to perform its firing and for every $\tau \in T$, $t_\tau$ is the time that token becomes available, then the starting time of the actor firing is given by $\bigoplus_{\tau \in T} t_\tau$. If $d$ is the execution time of that actor then the tokens produced by the actor firing become available at $\bigoplus_{\tau \in T} t_\tau + d$. Now, let $\gamma(A_i,k)$ be the completion time of the $k^{th}$ firing of actor $A_i$. This annotation is present in Fig. 3a for each of the actors. In order for an actor to fire, it must have all its input dependencies satisfied. We can now derive the expression for $\gamma$:

$$\gamma(A_i,k) = \left( \bigoplus_{(A_h,A_i) \in \mathcal{E}} \gamma\left( A_b, \left\lfloor \frac{r(A_i,(A_h,A_i))k - i(A_h,A_i)}{r(A_h,(A_h,A_i))} \right\rfloor \right) \right) \otimes e(A_i).$$  \hfill (1)

The completion time of the $k^{th}$ firing of actor $A_i$ corresponds to the maximal completion times of appropriately indexed firings of actors that feed its input edges $A_h \mid (A_h,A_i) \in \mathcal{E}$ increased by its own execution time $e(A_i)$. The quotient $\left\lfloor \frac{r(A_i,(A_h,A_i))k - i(A_h,A_i)}{r(A_h,(A_h,A_i))} \right\rfloor$ is used to index the appropriate firing of the actors that feed its input edges. The $i(A_h,A_i)$ member in the nominator of the fraction accounts for initial tokens. Initial tokens have the semantics of the initial delay and form the initial conditions used to solve (max,+) difference equations, analogue to the initial conditions in classical linear difference (recurrence) equations. We comply with the liveness criteria from [10] which among others requires that all SPDGF cycles are live, i.e. within a cycle there is an edge with initial tokens to fire the actor the needed number of times to complete an iteration, either a global one or a local one. Liveness and the ‘acyclic within an iteration’ restriction render (1) solvable and we can always obtain a solution for (1) in terms of initial conditions. The analytical solution of a system of such (max,+) linear difference equations evaluated at the iteration boundary for every actor of the graph will exactly give us the needed symbolic PSADF characteristic (max,+) matrix. We follow the order of actors from the quasi-static schedule. This guarantees that we respect data/resource dependencies. Element $X^{\#X}$ tells us that we have to solve (1) for actor $X$ at $k = \#X$. The obtained solution is propagated to the next iteration of the algorithm. We continue until we reach the end of the quasi-static schedule. At this point we will obtain solutions for all actors in terms of dependence of their completion times at the iteration boundary on initial conditions. From these solutions we can then easily construct the symbolic PSADF characteristic (max,+) matrix.

Let us consider the PSADFG example in Fig. 3a. We write down (max,+) equations for each actor (we omit the sign $\otimes$, i.e. $a \otimes b$ will be denoted as $ab$):

$$\gamma(A,k) = (\gamma(A,k-1) \oplus \gamma(E,k-2)) a = a\gamma(A,k-1) \oplus a\gamma(E,k-2),$$  \hfill (2)

$$\gamma(B,k) = b\gamma(A,\left\lceil \frac{k}{p} \right\rceil),$$  \hfill (3)

$$\gamma(C,k) = \left( \gamma(B,\left\lceil \frac{k}{q} \right\rceil) \oplus \gamma(C,k-1) \right) c = c\gamma(B,\left\lceil \frac{k}{q} \right\rceil) \oplus c\gamma(C,k-1),$$  \hfill (4)

$$\gamma(D,k) = \left( \gamma(A,\left\lceil \frac{k}{s} \right\rceil) \oplus \gamma(D,k-1) \right) d = d\gamma(A,\left\lceil \frac{k}{s} \right\rceil) \oplus d\gamma(D,k-1),$$  \hfill (5)

$$\gamma(E,k) = (\gamma(C,pqk) \oplus \gamma(D,sk)) e = e\gamma(C,pqk) \oplus e\gamma(D,sk).$$  \hfill (6)
The initial conditions are:

\[
\gamma(A, 0) = t_1, \quad \gamma(D, 0) = t_2, \quad \gamma(C, 0) = t_3, \quad \gamma(E, -1) = t_4, \quad \gamma(E, 0) = t_5.
\]

We can now evaluate and solve them at an iteration boundary given by the sequential schedule \(AB^pC^{pq}D^sE\).

Firing actor \(A\) using (2) with \(k = 1\) we obtain:

\[
\gamma(A, 1) = a\gamma(A, 0) \oplus a\gamma(E, -1) = at_1 \oplus at_4.
\]

Firing \(B^p\) using (3) with \(k = p\) and using (8) we obtain:

\[
\gamma(B, p) = abt_1 \oplus abt_4.
\]

Firing \(C^{pq}\) using (4) with \(k = pq\) and (9) we obtain (backward substitution):

\[
\gamma(C, pq) = abct_1 \oplus abct_4 \oplus c\gamma(C, pq - 1) = abc^{pq}t_1 \oplus c^{pq}t_3 \oplus abc^{pq}t_4.
\]

Firing \(D^s\) using (5) with \(k = s\) similarly evaluates to:

\[
\gamma(D, s) = ad^st_1 \oplus d^st_2 \oplus ad^st_4.
\]

Firing \(E\) using (6) with \(k = 1\) and (10) (11) we obtain:

\[
\gamma(E, 1) = aet_1 (bc^{pq} \oplus d^s) \oplus d^se t_2 \oplus c^{pq}et_3 \oplus aet_4 (bc^{pq} \oplus d^s).
\]

In (12) initial conditions \(t_1\) and \(t_4\) are \((\text{max}, +)\) multiplied by a symbolic \((\text{max}, +)\) summation term \((bc^{pq} \oplus d^s)\). We refer to this situation as a conflict. The production time of the tokens generated by actor \(E\) will depend on the relationship between \((b + pqc)\) and \(sd\). Before proceeding, we have to consider two cases. One given by \((b + pqc \geq sd)\) and the other by \((b + pqc \leq sd)\). We must check the intersection of newly added constraints and the already existing ones to reason against feasibility. If there are no feasible points in one of the subregions, we drop the further evaluation within the same subregion. In this example let us assume that both subregions contain feasible points. We easily construct the symbolic matrices from the solutions that are all expressed in terms of their dependence on initial conditions at an iteration boundary. We write down once more the solutions of the equations at the iteration boundary for actors that reproduce the initial tokens. Those are actors \((A, C, D, E)\). We will change the notation from \(\gamma(A_i, k)\) to \(t'_j\) depending on the indexes of initial conditions (tokens) and the producing actor. We obtain for \((b + pqc \geq sd)\):

\[
\begin{align*}
t'_1 &= at_1 \oplus at_4, \quad (13) \\
t'_2 &= ad^st_1 \oplus d^st_2 \oplus ad^st_4, \quad (14) \\
t'_3 &= abc^{pq}t_1 \oplus c^{pq}t_3 \oplus abc^{pq}t_4, \quad (15) \\
t'_4 &= t_5, \quad (16) \\
t'_5 &= abc^{pq}et_1 \oplus d^se t_2 \oplus c^{pq}et_3 \oplus abc^{pq}et_4. \quad (17)
\end{align*}
\]

From (13)-(17) we then easily obtain the rows of the symbolic \((\text{max}, +)\) matrix:

\[
G_{(b + pqc \geq sd)} = \begin{bmatrix}
-a & -\infty & -\infty & a & -\infty \\
-a + sd & sd & -\infty & a + sd & -\infty \\
-a + b + pqc & -\infty & pqc & a + b + pqc & -\infty \\
-\infty & -\infty & -\infty & -\infty & 0 \\
-a + b + pqc + e & sd + e & pqc + e & a + b + pqc + e & -\infty
\end{bmatrix}.
\]
The same procedure is used for the \((b + pqc \leq sd)\) case. The evolution of the PSADF graph is then governed by the following equations over the parameter space \(\Omega\): 

\[
\vec{\gamma}_{k+1} = G_{(b + pqc \geq sd)} \vec{\gamma}_k
\]

\[
\vec{\gamma}_{k+1} = G_{(b + pqc \leq sd)} \vec{\gamma}_k,
\]

depending in which region of \(\Omega\) is the \((k+1)\)th iteration scheduled. If \((b + pqc = sd)\), any of the two can be chosen. In the definition of both regions we use the \(\leq\) and \(\geq\) operators to have them remain closed. The functions that constitute the symbolic \((\max,+)\) matrices are polynomial functions of \(\vec{p}\).

In order to obtain the worst case throughput we will have to solve a mixed-integer polynomial programming problem for \(G_{(b + pqc \geq sd)}\) and \(G_{(b + pqc < sd)}\) over \((\Omega \cap (b + pqc \geq sd))\) and \((\Omega \cap (b + pqc \leq sd))\), respectively. A collection of techniques that solve such problems for a variety of definitions of \(\Omega\), e.g. convex, non-convex or restricted to take only a few discrete values, can be found in [21]. The matrix \(\max(G_{(b + pqc \geq sd)}, G_{(b + pqc \leq sd)})\) will define the MPAG of the example PSADFG. The inverse of the MCM of this MPAG equals to the worst-case throughput.

At this point we present our recursive algorithm for symbolic PSADF \((\max,+)\) characteristic matrix extraction (Algorithm 1). The inputs to the algorithm are the pre-computed sequential quasi-static schedules \(Qss\), the set of PSADF \((\max,+)\) difference equations \(MpEqSet\), the initial parameter space \(\Phi = \Omega\) and the initial solution set \(Ss = \emptyset\). The solution set \(Ss\) is a set of ordered pairs \(Ss = \{(G_{\Phi_i}, \Phi_i)\}\), where \(G_{\Phi_i}\) is the symbolic \((\max,+)\) matrix that governs the evolution of the PSADF in the region \(\Phi_i \subseteq \Omega\) generated by

---

**Algorithm 1** Symbolic PSADFG \((\max,+)\) characteristic matrix extraction

1: function SYMBOLIC_EXTRACT(Qss, MpEqSet, \(\Phi\), Ss)
2: fBranchingNode \(\leftarrow\) false
3: while not Qss.isFinished() do
4: currQssElem \(\leftarrow\) Qss.popNextElem()
5: currSol \(\leftarrow\) SOLVE(MpEqSet, currQssElem)
6: if currSol.Conflicted() then
7: fBranchingNode \(\leftarrow\) true
8: while new\(\Phi\) \(\leftarrow\) currSol.getNextConflict() do
9: if FEASIBILITY_CHECK(new\(\Phi\), \(\Phi\)) then
10: curr\(\Phi\) \(\leftarrow\) \(\Phi\)
11: curr\(\Phi\).Add(new\(\Phi\))
12: currMpEqSet \(\leftarrow\) MpEqSet
13: currMpEqSet.ResolveC(new\(\Phi\))
14: Ss.Add(SYMBOLIC_EXTRACT(Qss, currMpEqSet, curr\(\Phi\), Ss))
15: end if
16: end while
17: else
18: MpEqSet.Update(currSol)
19: end if
20: end while
21: if not fBranchingNode then
22: return (mpEqSet, \(\Phi\))
23: else
24: return \(\emptyset\)
25: end if
26: end function
adding conflict resolving constraints to $\Omega$ during the execution of the algorithm. Algorithm traverses the sequential schedule taking one actor with its repetition count at a time (Line 18). Function SOLVE (Line 5) solves Equation (11) for the considered actor. If there are no conflicts in the solution, the algorithm updates the equation set with the current solution that can be used in later iterations (Line 13). If there are conflicts, i.e., there are $\bigoplus_i y_i$ terms multiplying the initial conditions, we have to split the parameter space (Line 8). For example, if the term $y_1 \oplus y_2 \oplus y_3$ is multiplying an initial condition, we have to consider three cases: $(y_1 > y_2, y_1 > y_3)$, $(y_2 > y_1, y_2 > y_3)$, and $(y_3 > y_1, y_3 > y_2)$. Function FEASIBILITYCHECK (Line 9) checks the emptiness of the intersection of the current constraint set $\Phi$ and the new constraints. If the intersection is non-empty, new constraints are added to the current set for this branch of exploration (Line 11), conflicts are resolved (Line 13) and $\text{SYMBOLICEXTRACT}$ is recursively called again (Line 14). If the intersection is non-feasible, this branch is dropped. If we continue in this fashion we will eventually reach a non-branching node (Line 22).

We demonstrate our approach on the example PSADF graph in Fig. 3a. The example models a dynamic streaming application consisting of loops with interdependent parametric affine loop bounds. We define the ranges for parametric loop bounds (PSADF rates) as: $p \in \{10, 2000\}$, $q \in \{10, 15\}$, and $s \in \{100, 1500\}$. We also define linear dependencies between them: $p + s \leq 1400$ and $q \leq p$. Our application is run on a multi-processor platform where each loop body (actor) is mapped onto a different processor. Let PSADF actor execution times take the values of their nominal execution times multiplied by the parameter $c_i \in [1, 5]$ to account for six different possible platform dynamic voltage and frequency scaling (DVFS) settings. We obtain: $a = 30c_i$, $b = 20c_i$, $c = 4c_i$, $d = 3c_i$, $e = c_i$. These constraints define $\Omega$ for our example. To obtain the worst-case throughput value we must maximize the matrices $G_{(b+pqc\geq sd)}$ and $G_{(b+pqc\leq sd)}$ over $\Omega$ as given by the previously listed constraints. These become two mixed integer polynomial programming problems over $\Omega \cap (b + pqc \geq sd)$ and $\Omega \cap (b + pqc \leq sd)$ and can be solved using the technique from [21]. Throughput is given by the inverse of the MCM of the MPAG defined by the matrix $\max \{ G_{(b+pqc\leq sd)}, G_{(b+pqc\geq sd)} \}$ and equals to $1/390000$ iterations per time-unit.

5 Experimental results

We demonstrate our throughput analysis technique on five representative DSP applications with parametric interdependent affine loop bounds listed in Table 1. The first column shows the number of PSADFG actors, the second denotes the number of initial tokens, the third shows the number of parametric rates, the fourth gives the number of parametric actor execution times and the last shows the number of scenarios as the number of points in the PSADFG parameter space $\Omega$. All applications, except the bounded block parallel lattice reduction algorithm for MIMO-OFDM [3], are mapped onto a two-processor scalar architecture. The latter is mapped onto a vector/SIMD architecture. To obtain the nominal actor execution times for our benchmark set, we used the AVR32 [1] simulator under a reference frequency of 32 MHz. For bounded block parallel lattice reduction algorithm [3] we used random numbers for nominal actor execution times, as the source code of the algorithm is not publicly available. We assume that the frequency of each processor can be placed inside the range from 32 to 64 MHz, with the step of 1 MHz. For a 2 processor platform this will give 32 possible combinations. In contrast to the conventional SADF approach from [12] which would have to generate $|\Omega|$ SDFGs, our approach in each of these cases will solve maximally $(n \times n)$ polynomial programming problems without the need for the enumeration of $\Omega$ which is a difficulty by itself. Actually, in practice this number is usually less than $(n \times n)$, because not all initial tokens depend on all other initial tokens in the graph rendering the matrices to be quite sparse. Moreover, sometimes the entries in the symbolic PSADF $(\max, +)$ characteristic
Table 1: Experimental results

| Benchmark | $|A|$ | $|PR|$ | $|PD|$ | $|\Omega|$ |
|-----------|-----|------|------|-------|
| Fundam. freq. detector based on norm. autocorr. [2] | 12 | 6 | 2 | 2 | 16,687,681 · 32 |
| Normalized LMS alg. [2] | 9 | 6 | 2 | 2 | 385 · 32 |
| High resolution spectral analysis [2] | 9 | 6 | 2 | 2 | 385 · 32 |
| Adaptive predictor program [7] | 6 | 4 | 2 | 2 | 400 · 32 |
| Bound. block parallel latt. reduct. alg. [3] | 12 | 5 | 3 | 1 | 300 · 16 |

Matrix are repetitive, so we only have to solve the corresponding problem once. The symbolic PSADF (max,+) characteristic matrices of the benchmark applications were extracted manually using Algorithm 1, while the corresponding optimization problems were solved using CVX, a package for specifying and solving convex programs [17][16].

6 Conclusion

In this paper we have presented an extension to SADF that allows to model applications with vast or infinite sets of behaviours. We refer to our model as PSADF. We have proven the semantical equivalence of the two models and used that result in the formulation of worst-case throughput calculation problem for PSADF graphs with a fully connected state transition system within a generic optimization framework. The objective functions are functionals that represent the elements of the symbolic PSADF (max,+) characteristic matrices. Furthermore, we have derived a (max,+) linear theory based algorithm that is able to generate these matrices by combining a (max,+) difference equation solver and a recursive parameter space exploration for a subclass of PSADF graphs that are ‘acyclic within an iteration’. As future work, we want to fully automate our technique and investigate the problem of parametric throughput analysis of PSADF graphs.

References


