Requirements for Hybrid Cosimulation

David Broman
Lev Greenberg
Edward A. Lee
Michael Massin
Stavros Tripakis
Michael Wetter

Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2014-157
http://www.eecs.berkeley.edu/Pubs/TechRpts/2014/EECS-2014-157.html

August 16, 2014

Acknowledgement

This work is supported in part by the iCyPhy Research Center (Industrial Cyber-Physical Systems, supported by IBM and United Technologies), and the Center for Hybrid and Embedded Software Systems (CHESS) at UC Berkeley (supported by NSF award #0931843 (ActionWebs), NRL award #N0013-12-1-G015, the Assistant Secretary for Energy Efficiency and Renewable Energy, Office of Building Technologies of the U.S. DoE, under Contract No. DE-AC02-05CH11231, and the following companies: Denso, National Instruments, and Toyota). This work was also supported by the Academy of Finland and by the NSF via projects COSMOI: Compositional System Modeling with Interfaces and ExCAPE: Expeditions in Computer Augmented Program Engineering. This work was also supported by the Swedish Research Council #623-2013-8591.

Requirements for Hybrid Cosimulation *

David Broman[1,2], Lev Greenberg[3], Edward A. Lee[1],
Michael Masin[3], Stavros Tripakis[1,4], Michael Wetter[5]
[1]{broman,eal,stavros}@eecs.berkeley.edu, [3]{levg,michaelm}@il.ibm.com, [5]mwetter@lbl.gov
[1]University of California, Berkeley, [2]KTH Royal Institute of Technology,
[3]IBM Research – Haifa, Israel, [4]Aalto University, [5]Lawrence Berkeley National Laboratory

August 16, 2014

Abstract

This paper defines a suite of requirements for future hybrid cosimulation standards, and specifically provides guidance for development of a hybrid cosimulation version of the Functional Mockup Interface (FMI) standard. A cosimulation standard defines interfaces that enable diverse simulation tools to interoperate. Specifically, one tool defines a component that forms part of a simulation model in another tool. We focus on components with inputs and outputs that are functions of time, and specifically on inputs and outputs that are mixtures of discrete events and continuous time signals. This hybrid mixture is not well supported by existing cosimulation standards, and specifically not by FMI 2.0, for reasons that are explained in this paper. The paper defines a suite of test components, giving a mathematical model of an ideal behavior, plus a discussion of practical implementation considerations. The discussion includes acceptance criteria by which we can determine whether a standard supports definition of each component. In addition, the paper defines a set of test compositions of components. These compositions define requirements for coordination between components, including consistent handling of timed events.

1 Introduction

FMI (Functional Mock-up Interface) is an evolving standard for composing model components designed using distinct modeling and simulation tools [16]. The standard consists of a C API for simulation components and an XML schema for describing components. An **FMU** (Functional Mock-up Unit) is a component, typically exported from a modeling and simulation tool, that can be instantiated and used as part of a simulation in another modeling tool. To date, the emphasis of the standard has been on components that model the dynamics of physical systems, such as mechanical and electrical components. This emphasis reflects the origins of FMI as a way to achieve interoperability of simulators for models of automotive suppliers and OEMs [4].

FMI provides two distinct mechanisms for interaction between an FMU and a host simulator. The first mechanism is **model exchange**, where the host simulator is responsible for all numerical integration methods. The FMU provides procedures representing model equations, for example to compute the derivative of state variables of the component given current values for those state variables. The host simulator

*Acknowledgments: This work was supported in part by the iCyPhy Research Center (Industrial Cyber-Physical Systems, supported by IBM and United Technologies), and the Center for Hybrid and Embedded Software Systems (CHESS) at UC Berkeley (supported by the National Science Foundation, NSF award #0931843 (ActionWebs), the Naval Research Laboratory (NRL #N0013-12-1-G015), the Assistant Secretary for Energy Efficiency and Renewable Energy, Office of Building Technologies of the U.S. Department of Energy, under Contract No. DE-AC02-05CH11231, and the following companies: Denso, National Instruments, and Toyota). This work was also supported by the Academy of Finland and by the NSF via projects *COSMOI: Compositional System Modeling with Interfaces* and *ExCAPE: Expeditions in Computer Augmented Program Engineering*. This work was also supported by the Swedish Research Council #623-2013-8591.

is responsible for advancing time and updating the values of state variables, for example by performing numerical integration.

The second mechanism is **cosimulation**, where the FMU implements its own mechanisms for advancing the values of state variables, for example by internally implementing its own numerical integration method. The host simulator provides input values to the FMU, requests that the FMU advance its state variables and output values in time, and then queries for the updated output values. Cosimulation provides a looser coupling of tools than model exchange, and in principle can have a smaller, simpler interface with fewer assumptions and constraints. In the FMI 2.0 standard [16], both of these mechanisms, but most particularly cosimulation, are optimized for simulating continuous dynamics. The emphasis in model exchange is on providing enough information about the FMU component so that the host simulator can effectively use sophisticated integration algorithms. The emphasis in cosimulation is on loose coupling between continuous integration algorithms, where time can advance somewhat independently within the FMU and the host simulator. For example, the current version of the standard provides no mechanism for a cosimulation FMU to output an instantaneous reaction to a changed input value (such reactions are central to **reactive systems** [7]). The assumption is that inputs and outputs change reasonably smoothly, so delayed reactions have small effects on the overall simulation behavior.

In view of these assumptions, the community-driven standardization process is considering a third mechanism called **hybrid cosimulation** that strives for the loose coupling of cosimulation, but with support for discrete and discontinuous signals and instantaneous events. The intent of this mechanism is to support hybrid systems [13, 1, 6, 19, 17], where continuous dynamics are combined with discrete mode changes and discrete events. Any mechanism that supports hybrid systems can also support purely discrete systems and purely continuous systems, since those are both special cases.

In this paper, we provide a series of examples of simulation components (test components) that must be supported by any comprehensive mechanism that supports hybrid systems. These examples are chosen to comprehensively cover the space of functionality needed to effectively model mixed continuous and discrete systems, and to be as simple as possible. In addition, we provide a small set of simulation models (test models) that compose test components. For the test components and models, we mathematically define an ideal behavior and discuss practical acceptance criteria by which we can determine whether a practical implementation conforms with the ideal behavior.

Note that although FMI 2.0 does not currently support hybrid cosimulation, small extensions to the standard are sufficient to enable such support. Some such extensions are described and analyzed in [5, 21], where it is shown for example how to support reactive systems. We believe, therefore, that a hybrid cosimulation FMI standard that conforms with the requirements in this paper and yet is similar in spirit and style to the current standard is practical.

Together with the FMI 2.0 standards document, the FMI Development Group has developed an FMI Test Package. This package provides a number of models given in the Modelica language [15, 20], together with a specification of submodels that should be exported by a tool as FMUs, and re-imported by a tool as FMU. A tool passes the test if the behaviors all match the Modelica behavior. Compatibility between tools can be checked by having one tool import FMUs exported by another. If all tests match the Modelica behavior, the tools are compatible.

Although we could find no explicit statement to this effect, the tests specified in this package appear to be tests of the Model Exchange mechanism in FMI 2.0, since many of the tests are not implementable using Co-Simulation. Nevertheless, many of these tests are just as relevant to a future Hybrid Cosimulation standard as they are to the existing Model Exchange standard. Those tests include simple connections that test for proper FMU sorting, connections with cycles (that can be resolved without algebraic loop solving), linear and non-linear systems, systems with mixed data types, systems with event propagation, and models that test initialization of connected FMUs. Most (and possibly all) of these test cases could also form useful test cases for a Hybrid Cosimulation standard. Hence, we mostly avoid in our test cases in this paper situations that are already covered by those tests.

In this paper, we also take a different approach towards specifying test cases. Instead of providing a Modelica program as a reference, we give a mathematical ideal. The ideal correct behavior of a model is

unambiguous and language and tool independent.

In some cases, the mathematical ideal is not exactly realizable by any computational system. Real numbers, for example, are always approximated in a computational system, typically by floating point numbers. Such approximations are necessarily part of the test criteria, with specified tolerances.

Also, in some cases, the tests given here may not be directly realizable in Modelica, which, for example, lacks the notion of an absent value on a signal. These tests reflect discrete-event behaviors that are found more commonly in other types of simulators, such as network simulators (e.g. ns-3, `http://www.nsnam.org/`, and OPNET, from Riverbed)), hardware description languages (such as VHLD and SystemC), DEVS-based system simulators [22], and synchronous-reactive languages [3]. Although these tests can be emulated in Modelica through suitable encoding of signals, and Modelica tools should be able to import FMUs using this notion through such suitable encodings, it would be awkward and indirect to specify the tests in this way.

2 Principles

2.1 Definitions

We adopt the following definitions:

- A **host simulator** is a tool that imports a modeling component (an FMU) that is either written by hand or exported from another tool (or possibly even the same tool).
- The **master algorithm** is the execution procedure and policy by which the host simulator invokes the interface procedures of a component (an FMU).
- A **communication point** is the simulation time at which the master algorithm invokes an interface procedure of an FMU.
- A **deterministic FMU** is one where the output values and states are uniquely defined given initial conditions, input values, and communication points.
- A **deterministic composition of deterministic FMUs** is one where for a valid sequence of communication points, given initial conditions and inputs from outside the composition, the values of outputs of the deterministic FMUs are uniquely defined. Note that practical implementations may approximate those values using imprecise numerical methods, but the composition is still deterministic if the ideal correct values are uniquely defined.

The definition of determinism is a bit subtle. See [10] for a rigorous definition.

2.2 Assumptions

The FMI specification for hybrid cosimulation defines a contract between FMUs and master algorithms. This paper assumes the following principles to be followed in defining the specification:

- We prefer a weaker contract over a stronger contract. That is, we prefer fewer constraints on the design of FMUs and master algorithms. Constraints that can be eliminated without violating the requirements of this document should be eliminated. This will maximize interoperability of FMUs and master algorithms.
- We assume superdense time and piecewise continuous signals (see below and [11]). Specifically, two distinct communication points can share the same notion of real simulation time and yet be ordered in time. This is necessary for rigorous modeling of discontinuities and discrete signals and is already included in FMI 2.0 for model exchange.
- Where possible, the specification should use mechanisms already present in FMI 1.0 and 2.0, rather than replacing them with new mechanisms.
- The specification should enable, but not require, efficient execution.

2.3 Requirements

The FMI specification extension for hybrid cosimulation shall provide the following:

- It shall be able to handle all test cases that are described in this document, with approximations where necessary, as indicated in the discussion of the test cases below.
- It shall provide an unambiguous way of cosimulating FMUs.
- It shall enable (but not necessarily mandate) master algorithms that assure deterministic composition of deterministic FMUs.
- It shall be backwards compatible, meaning that current version 2.0 FMUs for cosimulation shall be possible to simulate, at least with continuous inputs, together with new FMUs designed specifically for hybrid cosimulation.

2.4 Notation

We use a particular mathematical notation to define the test cases in this paper. This notation is not intended to be used explicitly in the FMI specification for hybrid cosimulation. It is a mathematical idealization of what would be realized in an FMU and the host simulator.

The set $T = \mathbb{R}_+ \times \mathbb{N}$ represents time, where $\mathbb{R}_+$ is the set of non-negative real numbers, and $\mathbb{N} = \{0, 1, 2, \cdots\}$ is the set of natural numbers. A **superdense time** $\tau \in T$ is two-tuple, $\tau = (t, n)$, where the real number t represents a time in the usual Newtonian sense, and n is the **microstep**, which indexes sequences of values at Newtonian time t. Every **communication point** is a member of the set T.

T is a totally ordered set, where for any $\tau_1, \tau_2 \in T$ where $\tau_1 = (t_1, n_1)$ and $\tau_2 = (t_2, n_2)$, then $\tau_1 > \tau_2$ if either $t_1 > t_2$, or $t_1 = t_2$ and $n_1 > n_2$. Otherwise, $\tau_1 \leq \tau_2$.

A **signal** x is a function of the form

$$x \colon T \to \mathbb{R} \cup \{\varepsilon\}, \tag{1}$$

where ε represents the absence of a value. In the ideal, the signal is total, defined at all T, but in a simulation, signal values will be computed only at a finite subset of values of T. Note that the FMI specification will need to deal with data types other than reals as well, but we assume here that those data types simply match what is provided by FMI 2.0. There is no need for a hybrid cosimulation standard to deviate from the existing standard in this regard. Also note that nothing in this paper requires that FMI include any explicit representation of ε. It is a semantical concept.

A **continuous-time** (**CT**) signal is one that has a non-absent value for all $\tau \in T$. A **discrete-event** (**DE**) signal is one that has a non-absent value at only some $\tau \in T$. Specifically, following [11], a DE signal x has a non-absent value $x(\tau)$ only for $\tau \in D \subset T$, where D is a **discrete set**.[1]

A signal is **discontinuous** at any time $t \in \mathbb{R}$ if there exist $n, m \in \mathbb{N}$ such that $x(t, n) \neq x(t, m)$.

The **initial-value signal** x_i for a signal x is a function of the form $x_i \colon \mathbb{R}_+ \to \mathbb{R} \cup \{\varepsilon\}$ given by

$$x_i(t) = x(t, 0)$$

for all $t \in \mathbb{R}_+$.

At any $t \in \mathbb{R}_+$, the **final microstep** m_t of a signal x is a number $m_t \in \mathbb{N}$ such that for all $m > m_t$, $x(t, m) = x(t, m_t)$. The **final value** at time t is $x(t, m_t)$. If for any $t \in \mathbb{R}_+$, x has no final microstep, then x is said to be a **chattering Zeno** signal. It has a Zeno condition at time t, where it has an infinite sequence of changing values. Put another way, for a chattering Zeno signal, there exists a time where the signal does not settle to a final value.

The **final-value signal** x_f for a non-chattering Zeno signal x is a function of the form $x_f \colon \mathbb{R}_+ \to \mathbb{R} \cup \{\varepsilon\}$ given by

$$x_f(t) = x(t, m_t)$$

for all $t \in \mathbb{R}_+$, where m_t is the final microstep at time t.

A **continuous signal** is a CT signal where for all $t \in \mathbb{R}_+$, $m_t = 0$ and x_i is continuous at t (in the usual sense for functions of reals).

A **piecewise-continuous CT signal** is one where

[1] A **discrete set** is an ordered set that is order isomorphic with a subset of the natural numbers [11].

1. $m_t = 0$ for all $t \in \mathbb{R}_+$, except $t \in D$, where $D \subset \mathbb{R}_+$ is a discrete set.
2. x_i is left-continuous for all t (in the usual sense for functions of reals).
3. x_f is right-continuous for all t.

We can extend the notion of a piecewise-continuous signal to include DE signals and signals that are neither CT nor DE (they are absent over some intervals and present over others). A **piecewise-continuous signal** is one where

1. $m_t = 0$ for all $t \in \mathbb{R}_+$, except $t \in D$, where $D \subset \mathbb{R}_+$ is a discrete set.
2. If $x_i(t) \neq \varepsilon$, then x_i is left-continuous at t (in the usual sense for functions of reals).
3. If $x_i(t) = \varepsilon$, then there exists a $\delta > 0$ such that for all $0 \leq \epsilon < \delta$, $x_i(t - \epsilon) = \varepsilon$.
4. If $x_f(t) \neq \varepsilon$, then x_f is right-continuous at t (in the usual sense for functions of reals).
5. If $x_f(t) = \varepsilon$, then there exists a $\delta > 0$ such that for all $0 \leq \epsilon < \delta$, $x_f(t + \epsilon) = \varepsilon$.

This simply extends the usual notion of left and right continuity to absent values.

A **well formed simulation** models a system where all signals are piecewise continuous. Every piecewise-continuous signal has a well-defined (possibly empty) sequence of times $d_0, d_1, d_2, \cdots$, ordered in time, at which it is discontinuous. An interval between these times, $t \in (d_i, d_{i+1})$, is called a **continuous interval** of the signal.

Note that FMI 2.0 for model exchange already requires piecewise-continuous CT signals, and also defines these in terms of superdense time, so this formulation is not fundamentally new to FMI. Our only change is to extend the formulation to signals that may be absent at some times, and hence to discrete event and mixed signals.

Note that we use the notation $t \in (d_i, d_{i+1})$ for an interval of real numbers. In this paper, we use the following notation conventions:

- (a, b) represents a set of real values x where $a < x < b$;
- $[a, b]$ represents a set of real values x where $a \leq x \leq b$; and
- $[a, b)$ represents a set of real values x where $a \leq x < b$.

This notation should not be confused with $(t, n) \in T$, where (t, n) is a **tuple**, not an interval.

If a signal is not piecewise continuous, then it is not possible for samples of the signal to unambiguously distinguish a discontinuous signal from a rapidly varying continuous signal. Consider the following non-piecewise-continuous signal,

$$x(t, n) = \begin{cases} 0 & \text{if } t < 1 \\ 1 & \text{otherwise} \end{cases}$$

Any discrete set of samples of this signal is also a set of samples of a continuous signal. In contrast, consider the piecewise-continuous signal,

$$y(t, n) = \begin{cases} 0 & \text{if } t < 1 \\ 0 & \text{if } t = 1 \text{ and } n = 0 \\ 1 & \text{otherwise} \end{cases}$$

A discrete set of samples that includes $y(1, 0)$ and $y(1, 1)$ cannot be a set of samples of any continuous signal. The samples unambiguously represent a discontinuity.

For numerical solvers with variable step sizes, piecewise continuity is a very valuable property. For the signal x above, a solver may attempt to reduce the step size in the vicinity of $t = 1$ because of the rapid variation of a possibly continuous signal. But no reduction in step size will be sufficient to make the signal smooth, and hence the solver may fail with an error, indicating that it cannot ensure accuracy. With the piecewise-continuous signal y, however, the solver will not see any rapid variation in the interval up to time $t = 1$, and hence will be able to preserve high accuracy with large step sizes. The solver needs to be "aware" of superdense time only to the extent that it must include time $t = 1$ as a boundary between integration intervals, and it should use the final value of the signal y as the initial value for the next integration interval.

Because of the desirable properties of piecewise-continuous signals, the components in this paper all have the property that if all their inputs are piecewise continuous, then all their outputs are piecewise continuous. If they have no inputs, then the outputs are by construction piecewise continuous.

Note that for a DE signal x to be piecewise continuous, we need

$$x_i(t) = x_f(t) = \varepsilon$$

for all $t \in \mathbb{R}_+$. This is an important property that enables numerical integrators to interoperate cleanly with discrete events. Numerical integrators will only deal with initial and final-value functions. For a DE signal, these signals are absent everywhere, and therefore such signals have no effect on numerical integration. Hence, such signals can be used to represent cyber events in cyber-physical systems. For such signals to affect the physical dynamics, they can be first converted to CT signals, for example using the Zero-Order Hold component described below.

Note that CT signals that are piecewise continuous can have discontinuities. At a point in time t where there is a discontinuity, $m_t \neq 0$, and the initial and final-value signals may differ. At this point in time, the signal $x(t, n)$ has a sequence of values $x(t, 0), x(t, 1), \cdots, x(t, m_t)$. A numerical integrator deals only with the values $x(t, m_t)$ (which is a boundary value, specifically an initial value in an initial-value problem), and $x(t, 0)$ (which is the final value in a final-value problem). Note the somewhat confusing reversal of terminology, where an initial value of a signal at time t is the final value over a time interval that ends at t, and the final value of a signal at time t is the initial value over a time interval that begins at t.

In all of the test cases below, $m_t = 0, 1$, or 2 at all t. I.e., the discontinuities are simple ones, where a signal either changes from one value to a single other value at time t, or progresses from absent to present and back to absent (in the case of discrete-event signals). However, in general, it is useful to support arbitrary $m_t \in \mathbb{N}$. A simple physical example is given by Newton's cradle, where five steel balls suspended by threads swing and collide. When one of the end balls collides with the remaining four, which are stationary, the momentum of that ball is successively transferred from one ball to the next down the chain until the fifth ball, which has nothing further to collide with, lifts up and swings. The chain of instantaneous transfers of momentum are naturally and cleanly modeled in sequences of microsteps. Details are given in [9].

More generally, complex hybrid system models will likely include software subsystems that lack any temporal semantics and are most cleanly modeled as sequences of instantaneous events, as done in the synchronous-reactive languages [3]. Such sequences of instantaneous events also benefit from arbitrary final microsteps m_t (and even from the possibility of chattering Zeno conditions, which model non-terminating computations).

Notice that a progression through a sequence of microsteps is *not the same* as the iteration commonly used in simulators to solve algebraic loops. The purpose of those iterations is to find the value of signals *at a single point in superdense time*. The sequence of intermediate values in such iterations is an accident of the algebraic loop solving strategy, and is not an intrinsic part of the model.

Note that FMI 2.0 for model exchange [16] provides an **event mode** for handling discrete events and discontinuities. No such mechanism is provided in FMI for cosimulation. However, FMI for model exchange lacks the notion of an absent value. To see why this notion is important, consider a networked system, where components communicate over, say, a packet-switched network. There is an important semantic distinction between receipt of a packet that happens to carry the same payload as the previous packet and the failure to receive a packet. Any simulation of such a system necessarily has to make this distinction. Another example where the notion of an absent signal is important is with voting schemes (e.g. for fault tolerance). The absence of a vote is distinctly different from, say, a nay vote.

One way to make this distinction is to explicitly encode the absence of a value as a particular datatype value. Tools providing hybrid cosimulation will be free to handle absent values this way, but it is more efficient to handle absence of a value by doing nothing. This is how discrete-event simulators generally work. Components react only when there are input events present (or when a certain time has elapsed). To ensure that such situations can be handled efficiently, in this paper, we treat absence of a value as a first-class semantic concept. It does not require any explicit encoding, though it can be implemented with explicit encoding if necessary to support it in any particular tool.

2.5 Time

The specifications in this paper use real numbers to represent time. Computer programs typically approximate real numbers using floating-point numbers, which can create problems. Specifically, real numbers can be compared for equality, but it rarely makes sense to do so for floating point numbers.

Several of the components described below specify output events that occur at specific points in time. When composing components, it will be beneficial if the view of time across components is consistent. For example, if two components produce periodic events with the same period, then it should be assured that those events will appear simultaneously at any other component that observes them. Periods that are simple multiples of one another should also yield simultaneous events. Quantization errors should not be permitted to weaken this property. In short, a useful hybrid cosimulation standard should provide a model of time with a sound notion of simultaneity.

The following properties are sufficient for a sound notion of simultaneity:

1. The precision with which time is represented should be finite and should be the same for all observers in a model. Infinite precision (as provided by real numbers) is not practically realizable in computers, and if precisions differ to different observers, then the different observers will not agree on which events are simultaneous.
2. The precision with which time is represented should be independent of the absolute magnitude of the time. In other words, the time origin (the choice for the meaning of time zero) should not affect the precision.
3. Addition of time should be associative. That is, for any three time intervals t_1, t_2, and t_3,

$$(t_1 + t_2) + t_3 = t_1 + (t_2 + t_3).$$

The last two of these three properties are not satisfied by floating-point numbers, so floating-point numbers should not be used as the primary representation for time. The first property implies that the precision of the representation of time should be a global property of a composition of components, not a property of individual components. For a practical implementation of a model of time that satisfies all three properties, see [18].

3 Test Components

An FMI hybrid cosimulation standard should at least be capable of supporting the suite of components defined in this section. These vary wildly in sophistication, with some being quite trivial and some quite subtle. In each case, we give a Platonic ideal, a mathematical description of idealized behavior, which will be approximated by any real implementation (such as an FMU). The ideals are expressed as functions from superdense-time signals to superdense-time signals, and include operations on real numbers, such as testing for equality of reals, that are not precisely realizable in computational systems. Nevertheless, these ideals define useful test cases, in the same sense that the mathematics of real numbers provides useful test cases for floating-point arithmetic in digital hardware. Regression tests using these ideals will need to be parameterized with specified precision requirements.

We have chosen the suite of components to be a (nearly) minimal set that represents hybrid (mixed discrete and continuous) behaviors, plus a small set of components required to create useful test cases for composition of components, described in Section 4 below. This set of components is not meant to be comprehensive. Many more components would be needed for a reasonable simulation library. And each component in this set is not designed to be general and reusable. Library components similar to these would likely have more parameters and a richer set of capabilities. This set is intended instead to represent the capabilities that are needed for hybrid cosimulation, and to be barely sufficient to construct test cases that test these capabilities.

3.1 Constant Signal Generator

- CT output signal y;
- Real parameter c.

For all $\tau \in T$,

$$y(\tau) = c \tag{2}$$

Discussion. As with many test cases here, this component provides an output value at all values τ of superdense time. Since there are an uncountably infinite number of such values, no computer execution of this component can actually provide output values at all such times. An FMU and host simulator for this test case is deemed correct if at every point τ in superdense time where the host simulator chooses to observe the output of this FMU (a communication point), then the value of that output will be c. This test case imposes no constraints on when such observations are made.

3.2 Gain

- Input signal x;
- Output signal y;
- Real parameter a.

For all $\tau \in T$,

$$y(\tau) = \begin{cases} ax(\tau) & \text{if } x(\tau) \neq \varepsilon \\ \varepsilon & \text{otherwise} \end{cases} \tag{3}$$

Discussion. How this component handles absent inputs is an essential part of the definition. As defined, given a DE input, the output will be DE. An alternative definition would use the most recently seen input value if the input is absent. The Zero-Order Hold component below can provide such behavior, so for this case, we stick to the definition in equation (3).

Note that if the input is piecewise continuous, the output will also be piecewise continuous.

With the definition in (3), it may be reasonable for an FMU to require that at every communication point τ for this FMU, $x(\tau) \neq \varepsilon$. That is, the master algorithm should not invoke interface procedures of this FMU except at times when the input is present. This would make handling absent inputs extremely efficient. The FMU would need to do nothing at all to handle them. The output will be absent if the input is absent because the FMU will not be invoked to make the output non-absent. To support this, the hybrid cosimulation standard would need to provide a way for an FMU to declare that it requires all input to be present when invoked. Or conversely, an FMU may indicate, as part of its interface definition, that it can react even if some (or all) inputs are absent. We do not include in this paper any such requirements, however, because we prefer weaker contracts over stronger ones, and we are focused on a minimal set of requirements, and not on making cosimulation more efficient.

Note that there is no requirement that the FMU be invoked at all times when x is present, except at points of discontinuity, where any reasonable master algorithm will invoke it to ensure that a discontinuous input results in a discontinuous output. For CT signals, it is not in fact possible to invoke an FMU at all times when x is present, because x is present at all times, which would imply an uncountably infinite number of communication points.

Note that the 2.0 FMI standard for cosimulation cannot realize this component, even for CT inputs only. On page 95 of the standard, in the "mathematical description," the document states "During the interrupted simulation the slave [FMU] (and its individual solver) can receive values for inputs ... and send values of outputs ..." This encouragingly suggests that a master algorithm can "interrupt" the slave (the FMU) to provide a new input value $x(\tau)$ at a communication point τ, and then retrieve output values that depend on that input value according to the above equation. However, such a sequence of operations is not permitted in the standard. On page 104 the document states "There is the additional restriction in 'slaveInitialized' state that it is not allowed to call `fmi2GetXXX` functions after `fmi2SetXXX` functions without an `fmi2DoStep` call in between." This requires that after an input value is provided, time must advance before the resulting output can be retrieved. But, as stated on page 95 of the FMI 2.0 specification, "The communication step

size has to be greater than zero." The amount of the time advance is up to the host simulator, not up to the FMU. Hence, this FMU cannot implement the specified function once it enters the 'slaveInitialized' state.

This implies that the calculation performed by this component can only be performed during 'initialization mode' of an FMU, and that it would make no sense for the master algorithm to call `fmi2GetXXX` during 'slaveInitialized' mode. We believe that there is no implementation of this component consistent with this constraint that supports the calling sequences defined in Figure 11 on page 103 of the standard document. The only way we see to implement this component in the 2.0 standard requires that the FMU never enter the 'slaveInitialized' state. But this seems to imply that time cannot advance.

3.3 Triggered Constant

- Input signal x;
- Output signal y;
- Real parameter c.

For all $\tau \in T$,

$$y(\tau) = \begin{cases} c & \text{if } x(\tau) \neq \varepsilon \\ \varepsilon & \text{otherwise} \end{cases} \tag{4}$$

Discussion. This component extends the trivial Constant component with a trigger input that regulates whether an output is produced or not. That is, an output is produced only if the input is present. If the input is piecewise continuous, then the output will be piecewise continuous. For same reason as the Gain component, this component is not implementable with the FMI 2.0 standard for cosimulation. The communication points should include at least every $\tau \in T$ where $x(\tau) \neq \varepsilon$.

3.4 Adder

- Input signals x_1 and x_2;
- Output signal y;

For all $\tau \in T$,

$$y(\tau) = \begin{cases} x_1(\tau) + x_2(\tau) & \text{if } x_1(\tau) \neq \varepsilon \text{ and } x_2(\tau) \neq \varepsilon \\ x_1(\tau) & \text{if } x_1(\tau) \neq \varepsilon \text{ and } x_2(\tau) = \varepsilon \\ x_2(\tau) & \text{if } x_1(\tau) = \varepsilon \text{ and } x_2(\tau) \neq \varepsilon \\ \varepsilon & \text{otherwise} \end{cases} \tag{5}$$

Discussion. This component illustrates that an FMU may be presented at a communication point with some inputs that are absent and some that are not, and that its behavior may depend on which inputs are present. Of course, a simpler Adder component would require all inputs to be present simultaneously, or would use previous input values if an input is not present. We can certainly design such Adder components, and indeed a library for simulation might prefer those semantics. But our goal here is to test capabilities that may be required in hybrid cosimulation, and reacting differently to different patterns of presence of inputs most certainly will be required.

This component imposes no constraints on the communication points, but as usual, points of discontinuity of the inputs must be presented as communication points in order for the output to reflect the ensuing discontinuity.

For same reason as the Gain component, this component is not implementable with the FMI 2.0 standard for cosimulation (and neither are the simpler Adder variants).

3.5 Periodic Piecewise Constant Signal Generator

- CT output signal y;
- Real parameters a, b, p.

Informally, this component outputs the constant a from time zero to p, b from time p to $2p$, a from $2p$ to $3p$, etc., alternating between a and b, as illustrated in Figure 1.

We require that the output be piecewise continuous. Specifically, for all $\tau = (t, n) \in T$,

$$y(t,n) = \begin{cases} a & \text{if } kp < t < (k+1)p \text{ and } k \in \mathbb{N} \text{ is even;} \\ b & \text{if } kp < t < (k+1)p \text{ and } k \in \mathbb{N} \text{ is odd;} \\ b & \text{if } t \text{ is an odd multiple of } p \text{ and } n \geq 1; \\ b & \text{if } t \text{ is an even multiple of } p,\ t > 0, \text{ and } n = 0; \\ a & \text{otherwise.} \end{cases}$$

Discussion. A correct implementation of this component and host simulator will produce at least the output values shown in Figure 1 as filled and unfilled dots. Hence, in any correct implementation, there must be at least two communication points at each multiple of p, one at microstep zero and one at microstep one. A host simulator may choose to invoke the FMU implementation at additional communication points, but this is not required. Typically a host simulator will use a step-size adjustment algorithm to choose communication points.

Note that this component is also not implementable using FMI 2.0 for cosimulation. First, the 2.0 standard requires strictly positive time advances when time is advanced, so the progression from microstep zero to microstep one at multiples of p is not realizable.

In addition, the 2.0 standard provides only an awkward mechanism for an FMU implementation of this component to ensure that a communication point occurs at multiples of p. The FMU can reject proposed time increments by returning `fmi2Discard` when its `fmi2DoStep` procedure is called. After the FMU rejects a proposed step, the host simulator can enquire the end time of the last successfully completed communication step.

However, for this component, the time of the next interesting event is knowable in advance. There is a **predictable breakpoint**, a required communication point that is knowable arbitrarily far in advance. Specifically, at any communication point $(t, n) \in T$, there exists a $k \in \mathbb{N}$ such that $kp \leq t < (k+1)p$. The next required communication point is

$$(t', n') = \begin{cases} (kp, 1) & \text{if } kp = t \text{ and } n = 0 \\ ((k+1)p, 0) & \text{otherwise} \end{cases} \tag{6}$$

So with better coordination between the FMU and the host simulator, there is no need for the host to propose a step that will be rejected. A specific extension to FMI that supports such predictable breakpoints is proposed in [5].

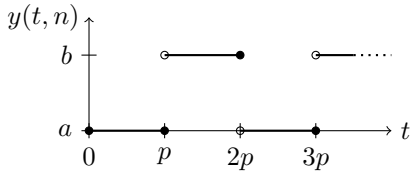

Figure 1: Example output from the Periodic Piecewise Constant component. The unfilled dots show values that occur only at microsteps $n \geq 1$, whereas the filled dots and lines show values at $n = 0$.

3.6 Periodic Discrete Signal Generator

- DE output signal y;
- Real parameters a, p.

This component outputs the constant a at integer multiples of p, and otherwise its output is absent. To be piecewise continuous, the output signal should be absent for all $(t, n) \in T$ where $n = 0$ or $n > 1$. Specifically, for all $\tau = (t, n) \in T$,

$$y(t, n) = \begin{cases} a & \text{if } t = kp \text{ and } n = 1, \text{ where } k \in \mathbb{N}; \\ \varepsilon & \text{otherwise.} \end{cases}$$

This is illustrated in Figure 2.

Discussion. This component provides a canonical source for a DE signal. It can be used to build regression tests to verify, for example, that the Gain component above behaves correctly with DE inputs. It can also be used to provide discrete inputs to any of the test cases below that require discrete inputs.

This component requires that there be a communication point at all $t = (kp, 1)$, $k \in \mathbb{N}$. Communication points at other times are not required, but if the host simulator provides them, then this component will produce no output (its output will be absent).

Just like the Periodic Piecewise Constant Signal Generator, this component has predictable breakpoints similar to (6), but slightly different. At any communication point $(t, n) \in T$, where $k \in \mathbb{N}$ is such that $kp \leq t < (k + 1)p$, the next required communication point is

$$(t', n') = \begin{cases} (kp, 1) & \text{if } kp = t \text{ and } n = 0 \\ ((k + 1)p, 1) & \text{otherwise} \end{cases} \tag{7}$$

This component is not realizable using the FMI 2.0 standard, which lacks any notion of absent values.

3.7 Modal Model with Discrete Control

- DE input signal x;
- CT output signal y;
- Real parameters a, b.

This component initially outputs the constant a. When the first input event arrives, it switches to producing output b. When the second input event arrives, it switches back to a. Etc. Formally,

$$y(t, n) = \begin{cases} a & \text{if } s(t, n) = 0 \\ b & \text{otherwise.} \end{cases} \tag{8}$$

where s is a CT signal such that $s(t, n)$ is the **state** of the component at time (t, n), defined as follows. Let $d_0, d_1, d_2, \cdots$ denote the superdense times τ, in temporal order, at which $x(\tau) \neq \varepsilon$. We are assured that such a sequence, which may be empty, finite, or infinite, exists, because x is a DE signal. At any superdense time $(t, n) \in T$, there may exist a maximum $i \in \mathbb{N}$ such that $(t, n) > d_i$. If no such i exists, then either there are

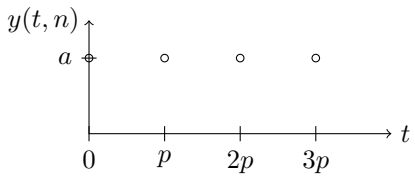

Figure 2: Example output from the Periodic Discrete Signal Generator. The unfilled dots are the only non-absent values, and they occur only at microstep $n = 1$.

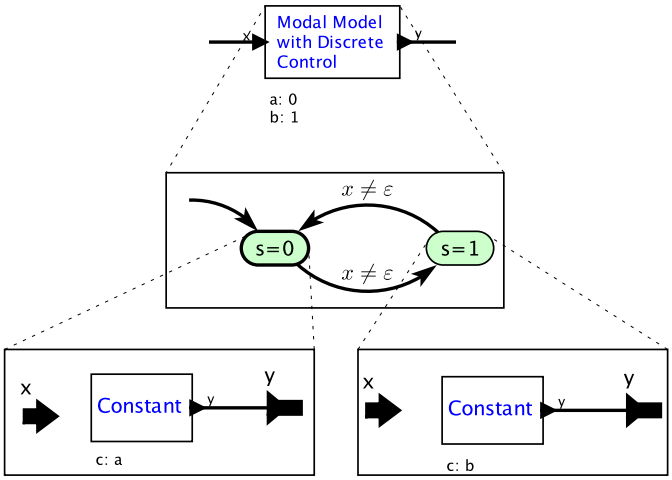

Figure 3: Modal Model with Discrete Control.

no events at all in x or $(t, n) \leq d_0$. Hence, at time (t, n), the state s is given by

$$s(t,n) = \begin{cases} 0 & \text{if no such } i \text{ exists} \\ 1 & \text{if } s(d_i) = 0 \\ 0 & \text{if } s(d_i) = 1 \end{cases} \tag{9}$$

In this definition, at time (t, n), d_i, if it exists, is the time of the most recent but strictly earlier input event. Hence, at time (t, n), $s(d_i)$ is the *previous* state of the component. The *current* state is always the opposite of the previous state.

In words, the state s is initially 0. Each time an input event arrives, the state toggles from 0 to 1 or from 1 to 0. The toggle occurs *strictly after* the input event arrives. So when an input event arrives at time (t, n), the output depends on the current state $s(t, n)$ as given by (8), and then, strictly later, the state is updated as in (9).

An illustration of such a component is given in Figure 3 using a hierarchical modal model in the style of [11]. The logic of the component is given as a **finite state machine** (FSM) (middle level) with two states, each with a mode refinement that outputs a constant.

Discussion. The state of this component is changed after each non-absent input. Since the input is required to be a DE signal, the state changes occur only at a discrete subset of T. Hence, the state updates are enumerable in temporal order, and hence computable.

The state of this component is changed *after* producing an output value. This ensures that the output is piecewise continuous, assuming the input is piecewise continuous.

Notice that if we combine this component with a Periodic Discrete Signal Generator to drive its input, then we can implement the Periodic Piecewise Constant Signal Generator. Nevertheless, we keep the Periodic Piecewise Constant Signal Generator in the suite of test cases to make it easier for readers to understand the progression of capabilities.

The only constraint that this component imposes on communication points is that there be a communication point at each superdense time $d_i = (t, n)$ where the input is not absent and one superdense time later $(t, n + 1)$. At (t, n), the output will be determined by the current state, whereas at time $(t, n + 1)$, the output will be determined by the next state. If $a \neq b$, then this creates a discontinuity, but the output remains piecewise continuous.

Notice that once we have an FMI standard that supports this component and the other components defined here, then it is easy to generalize this so that the output is not piecewise constant, but rather varies during the time the component resides in each state, $s = 0$ and $s = 1$. Each state could have dynamics,

including perhaps an internally implemented numerical integration algorithm, so that instead of producing the constant a while in state $s = 0$, it produces some dynamically changing output. An FMU realizing this component could influence step sizes taken by the master algorithm in a manner similar to the Integrator component below. It is further easy to generalize this component to have more than two modes (values of s). Hence, once we can support this component and the others included here, we can implement a broad class of **hybrid systems** [13], including **timed automata** [2] and many others.

3.8 Integrator

The output is the integral of the input.

- CT input signal x;
- CT output signal y;

For all $(t, n) \in T$,

$$y(t, n) = \int_0^t x(\alpha, 0) d\alpha. \tag{10}$$

We propose three variants of this component, which are meant to capture key properties of the most commonly used numerical integration algorithms. All three variants require a communication point (t, m_t) at every t where x is discontinuous, where m_t is the final microstep at t (note that this requires that the input not have a chattering Zeno condition for time to advance).

The first two variants do not require that the input $x(t, n)$ be set at a communication point (t, n) before the output $y(t, n)$ is retrieved (it must be eventually provided, but it can be provided after the output is retrieved). By definition (10), the output $y(t, n)$ does not depend on the input $x(t, k)$ for any $k \in \mathbb{N}$ at any time t. Hence, there is no **direct dependence** between the input x and the output y (i.e., no instantaneous dependence, unlike most of the components above). A component without such a direct dependence is called a **non-strict** component, meaning that the input need not be known at a particular time for the output at that time to be produced. The third variant below, however, reflects that some integration algorithms *are* strict, despite the definition (10) (namely, those using implicit integration methods). The variants are:

1. Variant 1 imposes no additional constraints on the communication points.
2. Variant 2 requires communication points at (t, m_t) for $t \in D$, where $D \subset \mathbb{R}_+$ is an arbitrary discrete set, chosen by the component.
3. Variant 3 is strict, and requires a communication points at $(t, 0)$ for any t where there is a communication point.

In all cases, any additional communication points are optional, up the host simulator.

Discussion. Note that by definition (10), the output of this component does not depend on input values at non-zero microsteps, but a realization will need to see the final values at discontinuities of the input to maintain accuracy. For this reason, all variants require that inputs be provided at *final* microsteps. The final value of the input x at time t provides the boundary value for an initial value problem to be solved by the numerical integration algorithm. The algorithm needs this value, even if the mathematical ideal (10) does not. The only variant that requires the initial value $x(t, 0)$ to be also provided is variant 3, which reflects the requirements of an implicit integration algorithm. Such an algorithm requires input values at both the start and end of an integration interval. At the start, it needs the final value, and at the end, it needs the initial value (confusingly).

Suppose that the input is discontinuous at times t_1 and t_2, where $t_1 < t_2$, and is continuous in between. In each such an interval, an FMU realizing any variant of this component needs to solve an initial-value problem, where the initial value of the integrator state y is equal to the final value of the signal y at time t_1, and the initial value of x (the derivative of y) is the final value of x at time t_1. The integration algorithm calculates the integral up to time t_2 using standard numerical integration techniques, and the value of y that it determines at t_2 will define the initial value of the signal y at time t_2.

During an interval (t_1, t_2) between discontinuities, the FMU may or may not be provided with additional input values x by the host simulator. I.e., there may or may not be communication points in between. In any

case, it cannot be provided with *all* such input values, because in general there are an uncountably infinite number of them. Numerical analysts have developed various techniques for approximating such integrals given partial information about the input x. For example, some more sophisticated solvers assume that the input signal x is not only continuous, but also *smooth*, meaning that all its derivatives exist at all points in the interval. In this case, given only the final value of x at time t_1 and all its derivatives, then the value of x at any point in the interval is fully defined and can, in principle, be calculated. In other words, given the value of the input x at time t_1 and its derivatives, **extrapolation** to any point in the interval (t_1, t_2) is possible.

Even with the smooth assumption, providing *all* the derivatives at time t_1 is not practical. Only a finite number will be provided by a master algorithm (providing none or one derivative are the most common). To be interoperable with a wide range of simulation hosts, therefore, the FMU will have to adapt its integration algorithm to the information provided. For example, if no derivatives are provided, the FMU might extrapolate using **zero-order hold**, which assumes that the input is constant over the interval with the value given by the final value of x at time t_1. Or it might use the recent history of the input to approximate the derivatives.

No matter what integration algorithm is chosen, and what information is provided by the host simulator, at least for some inputs, there will be errors compared to the ideal value given by (10). The three variants above are intended to reflect key properties of the most commonly used algorithms.

Variant 1 integrates between points in time chosen by the host simulator (which are required to include points of discontinuity of the input). Hence, if the FMU internally takes multiple steps between points provided by the host, it will need to extrapolate the input.

To prevent large errors due to extrapolation, it may be better for an Integrator component to influence the step sizes used by its host simulator, specifically requiring that the host simulator provide more input values x within an interval in order to maintain accuracy. To support such a scenario, it is essential for a hybrid cosimulation FMI standard to provide a mechanism for an FMU to influence the step size choices of the master algorithm. Variant 2 constrains the host simulator so that it at least provides input values at communication points of the FMU's choosing. This can model the most commonly used fixed-step-size algorithms, using for example the forward Euler method, with the caveat that additional communication points may be provided by the host (it is required to provide additional communication points if the input has a discontinuity at some t that is not in D). In our test cases, therefore, even a fixed-step-size algorithm must tolerate additional communication points, and the FMU must provide reasonable outputs at those points.

Variant 2 can also model variable-step-size algorithms. For a fixed-step-size algorithm, the set D is known ahead of time, so the breakpoints are predictable, like those of the Periodic Piecewise Constant Signal Generator and Periodic Discrete Signal Generator. For a variable-step-size algorithm, the set D is not known ahead of time. It depends on the input, and the breakpoints are not predictable. **Unpredictable breakpoints** require a different mechanism for the FMU and the master algorithm to coordinate step sizes. The technique that currently exists in FMI 2.0, where an FMU can reject a proposed step, can work in certain circumstances, but as explained in [5], this only works in general if all components support getting and setting their state. That is, in general, a mechanism for backtracking an FMU is required if unpredictable breakpoints are to be supported. See [5] for suggested mechanisms.

Considering a variable-step-size Integrator can help us understand why we require piecewise-continuous signals. Such an FMU will adjust the step size of its integration algorithm based on an estimate of the integration error. Typically, this estimate depends on how quickly the input x is varying. Small step sizes are required when x is varying rapidly. If we attempt to use such an algorithm to integrate over a region that includes a discontinuity in x, the step-size-adjustment algorithm is likely to fail, because at the point of the discontinuity, the input is changing infinitely fast, and no step size will be small enough. This is the key reason for requiring piecewise continuity in signals. It ensures that standard numerical integration can be used during the intervals when the input is continuous.

Some numerical integration algorithms are **implicit**, meaning that for each time interval (t_1, t_2) over which they integrate a function x, they use both the value of x at time t_1 (confusingly, the final value

at that time) and the value of x at time t_2 (confusingly, the initial value at that time). In this case, instead of extrapolating the input value, the numerical integrator **interpolates** the input. For example, the **trapezoidal method** assumes that the input x varies linearly between t_1 and t_2, given only input values at t_1 and t_2.

Variant 3 supports implicit integration methods. However, as pointed out above, this variant makes the FMU strict, meaning that the output at any time t depends on the input at time t. Such a dependence is not present in the ideal definition of integration, given in (10). Use of such algorithms makes it more challenging to use such a component in a feedback loop, because such use could yield a causality loop.

3.9 Integrator with Reset

The output is the integral of the input, but an additional discrete input can reset the state of the integrator.

- CT input signal x;
- DE input signal r;
- CT output signal y;

Let $D \subset T$ denote the set of time stamps $(t, n) \in D$ where $r(t, n) \neq \varepsilon$. I.e., for all $(t, n) \in T$ where $(t, n) \notin D$, $r(t, n) = \varepsilon$. Since this set is order isomorphic with a subset of the natural numbers, we can list its elements in order, d_0, d_1, d_2, $\cdots$, where $d_n < d_{n+1}$. For all $(t, n) \in T$, the output is defined to be

$$y(t,n) = \begin{cases} \int_0^t x(\alpha, 0) d\alpha & \text{if } (t,n) < d_0 \text{ or } D = \emptyset \\ r(d) + \int_u^t x(\alpha, 0) d\alpha & \text{otherwise} \end{cases}$$

where $d = (u, m)$ is the largest element in D s.t. $(u, m) \leq (t, n)$. We assume the same three variants as the Integrator component, with the additional constraint that for all variants, there must be a communication point at every $(t, n) \in D$. In addition, to ensure a piecewise-continuous output, for every $(t, n) \in D$, we require a communication point at $(t, 0)$.

Discussion. Absent any reset events r, this is identical to the Integrator above. At the time of reset events, the state of the integrator (its output y) gets reset to the value of the reset event. Like the Integrator, the output y has no immediate dependence on the input x, and it does not depend on the value of the input at non-zero microsteps (in the ideal). But the output y does immediately depend on the input r.

Assuming r is piecewise continuous, it is absent at all $(t, 0)$. Reset events can only occur at microsteps greater than zero. This ensures that the output y is piecewise continuous. Suppose a reset event occurs at superdense time $(t_1, 1)$. The output $y(t_1, 0)$ is unaffected by it, and hence is part of the preceding continuous segment. At the next microstep, $y(t_1, 1)$ takes on the value given by $r(t_1, 1)$. If there are no further reset events at time t_1, then this will be the final value of the output, which provides the initial integrator state for the next integration interval.

3.10 Zero-Crossing Detector

The output is a discrete event when the input hits or crosses zero.

- CT input signal x;
- DE output signal y;

For all $(t, n) \in T$, the output is

$$y(t,n) = \begin{cases} 0 & \text{if } x(t,0) = 0, n = 1, \text{ and there exists a } \delta > 0 \text{ s.t.} \\ & \text{for all } 0 < \epsilon < \delta, x(t-\epsilon, 0) \neq 0 \\ 0 & \text{if } n > 0 \text{ and } x(t, n-1) \text{ and } x(t, n) \text{ have opposite signs,} \\ 0 & \text{if } n > 0 \text{ and } x(t, n-1) \neq 0 \text{ and } x(t, n) = 0, \\ \varepsilon & \text{otherwise} \end{cases} \tag{11}$$

To remove any ambiguity, "opposite signs" means that one value is negative, and the other value is positive.

Discussion. This is the most subtle of the components considered in this paper, but it represents widely used functionality in numerical simulation. Every widely used simulator provides mechanisms for monitoring signals for zero crossings, which may represent, for example, collisions of physical objects or other discrete physical phenomena (see [9] for a few examples). Note that "crossings" is a bit of a misnomer here, because our zero-crossing detector will output an event if the input merely "hits" but does not "cross" zero. This simplifies the test case considerably, because to detect a legitimate "crossing" would require non-causal behavior. The component would need to see the future of the input in order to output an event. Non-causal I/O relationships are problematic. Semantics of discrete-event, continuous-time, and hybrid systems models all rely on causal I/O relationships [22, 8, 11, 12, 14].

Existing simulators implement variations of this functionality. This particular definition is carefully constructed to ensure that if the input is piecewise continuous, then the output is a piecewise-continuous DE signal. In particular, if an input hits zero and then stays there, this component produces an output event only when it first hits zero. Simulink's "Hit Crossing" block, in contrast, would continue to produce an event indicator as long as the input signal remains at zero. Our test case is simpler because the Simulink behavior would yield an output signal that is neither a CT signal nor a DE signal. While such signals are useful and legitimate, the additional complexity of handling them does not add anything fundamental to our test suite.

Here, we assume that the input x is piecewise continuous. Otherwise, it is not clear that there is a reasonable definition of a zero crossing detector. Piecewise continuity ensures that if there is a zero crossing within a continuous interval, then the first condition in (11) will be satisfied within this interval at the point of the zero crossing. Furthermore, the second and third conditions in (11) ensure that if a discontinuous signal crosses or hits zero during a discontinuity, the event is detected.

If the input is not piecewise continuous, then this definition of a zero-crossing detector is not assured of detecting zero crossings. Consider a non-piecewise-continuous input given by

$$x(t, n) = \begin{cases} -1 & \text{if } t \leq 1 \\ 1 & \text{otherwise} \end{cases}$$

Our zero-crossing detector produces a constant absent output, failing to detect this zero crossing. But when does this zero crossing occur? There is no earliest time t at which $x(t, n) = 1$, so there is no clean definition of time of the zero crossing. We could, of course, invoke contortions using limits from the right, but the definition will be non-causal. An event produced at time $t = 1$ would depend on *future* inputs. No such difficulties arise if the input is piecewise continuous, so we simply define a test case that yields expected behavior when the input is piecewise continuous. We do not care what the behavior is when the input is not piecewise continuous, because the behavior yielded by our component is as defensible as any other.

The output of this component is a piecewise-continuous DE signal. Note that it is always absent at microstep zero, and that as long as the input has no chattering Zeno condition, the final value of the output is always absent.

The I/O dependencies of this component are interesting. When the input is continuous, there is arguably a microstep delay, because the input is zero at microstep zero, but the output event is not produced until microstep one. However, at discontinuities, the second and third conditions of equation (11) have no such microstep delay. So in general there is no microstep delay from input to output.

The first condition in (11) compares a real-valued input x to a fixed constant zero. Such a comparison is not exactly realizable in a computer when the input is a computed value of a continuous signal defined on a time continuum; hence an approximation is needed. A typical approximation allows $x(t, 0)$ to be non-zero as long as it is small and has a sign that is opposite to that of $x(t - \epsilon, 0)$ over an interval $\epsilon \in (0, \delta)$. That is, a zero-crossing is allowed to overshoot by a small amount. Undershoot is typically not allowed, because then there is no assurance that either a crossing or a "hit" has occurred.

With such an approximation, the time at which the crossing is detected will be slightly late. But any digital representation of time must be quantized in any case (see Section 2.5 below), so some approximation is always necessary. A useful regression test will specify a precision requirement, so a useful implementation of this ideal component should provide a mechanism for controlling the precision (a parameter, for example).

This component requires a communication point at or near the point of a zero crossing. Like the Integrator above, this introduces an **unpredictable breakpoint**, because the time of this zero crossing cannot be

known in advance, in general. A typical realization of such a Zero-Crossing Detector cooperates with the master algorithm to regulate the step size taken by a simulator so that the detection delay is less than some specified time resolution. As a consequence, a hybrid cosimulation FMI standard should include a mechanism for an FMU to cooperate with the host simulator, regulating time steps taken by the simulator. One mechanism would be for the FMU to be able to reject step sizes proposed by the host simulator until the step size is small enough to detect the zero crossing within the specified resolution. Again, see [5] for a discussion of such mechanisms.

When an FMU rejects a step size because a zero crossing has occurred, it may be able to estimate the time at which the zero crossing actually occurs by interpolating between the specified inputs. Hence, for efficiency, a hybrid cosimulation standard should also include a mechanism for an FMU to be able to suggest a better step size when it rejects one.

Note that it appears that a reasonable approximation of a zero-crossing detector is difficult using FMI 2.0 for cosimulation as defined in [16]. In particular, when advancing time using `fmi2DoStep`, the host simulator provides no information about the value of the inputs at the *end* of the step. Hence, the FMU cannot reject a step size proposed by `fmi2DoStep`. The API does provide a way to reject a setting of input values using `fmi2SetReal` (which can return `fmi2Discard`). But almost certainly this is not intended to be used to reject a previously accepted step size and trigger a redoing of the step.

This component can only be approximated in FMI 2.0 for cosimulation by extrapolating the input, using the derivatives of the input, if provided, or an extrapolation based on the history of the input.

3.11 Zero-Order Hold

- DE input signal x;
- CT output signal y;

The output is defined to be

$$y(\tau) = \begin{cases} x(d) & \text{if } d \text{ exists,} \\ \varepsilon & \text{otherwise} \end{cases}$$

where d is the largest element in T s.t. $d \leq \tau$ and $x(d) \neq \varepsilon$.

Discussion. Recall that a DE signal is non-absent at a discrete subset of times. This property ensures that if there is any input event at all, then all outputs at the time of that event and beyond are not absent. Specifically, if there is any time $\tau_0 \in T$ where $x(\tau_0) \neq \varepsilon$, then for all $\tau \in T$ where $\tau \geq \tau_0$, there exists a largest $d \leq \tau$ such that $x(d) \neq \varepsilon$.

If the input is piecewise continuous, then the output will also be piecewise continuous. This fact is not entirely obvious from the definition. Note that for x to be a piecewise-continuous DE signal, it must be true that both its initial and final value functions are absent. Hence, if an event arrives on x at time $t_0 \in \mathbb{R}_+$, then the event must occur at a microstep strictly greater than 0. That is, $x(t_0, 0) = \varepsilon$. Hence, $y(t_0, 0)$ will have the value specified by *previous* event (or be absent if there is no previous event). Suppose that $x(t_0, 1) \neq \varepsilon$. Then $y(t_0, 1) = x(t_0, 1)$. If final value of x at t_0 is $x(t_0, 2) = \varepsilon$, then the final value of y at t_0 will be $y(t_0, 1) = x(t_0, 1)$, and this will be the value of y until (and including) $(t_1, 0)$, where t_1 is the time of the arrival of the next event.

The output of this component has a discontinuity at each time t where $x(t, n) \neq \varepsilon$. In order for this discontinuity to manifest correctly as a sequence of two distinct values at the same time t with distinct microsteps, there will need to be two distinct communication points at the time t of the input event. Suppose that an input event occurs at (t, n) for some $n > 0$. That is, $x(t, n) \neq \varepsilon$. It is sufficient for the host simulator to provide a communication point at both $(t, 0)$ and (t, n). But to provide a communication point at $(t, 0)$, the host simulator needs to "anticipate" the event at (t, n). This seems to imply non-causal behavior.

Fortunately, we can meet this requirement without any non-causal mechanisms. The simplest mechanism is already realized in FMI 2.0, where whenever a non-zero step size is taken, the next communication point automatically occurs at microstep zero. There is no mechanism in FMI 2.0 to skip microstep zero. The Ptolemy II Continuous director [18], which implements hybrid system simulation, also always provides a

communication point $(t, 0)$ for every communication point (t, n). So a Zero-Order Hold FMU will never be forced to reject a time step if Ptolemy II is the host simulator.

But even if the future hybrid cosimulation standard does provide a mechanism to advance from communication point (t_0, n_0) to (t_1, n_1), where $t_1 > t_0$ and $n_1 > 0$, then an FMU implementing this Zero-Order Hold component can reject the proposed step. It thereby ensures that for every communication point (t, n), there will also be a communication point $(t, 0)$. This mechanism does not require anticipating future events.

3.12 Sampler

The output is a discrete sampling of a continuous input.

- Input signal x;
- DE input signal s;
- DE output signal y;

For all $\tau \in T$, the output is

$$y(\tau) = \begin{cases} x(\tau) & \text{if } s(\tau) \neq \varepsilon \\ \varepsilon & \text{otherwise} \end{cases}$$

Discussion. As long as the input s is a piecewise continuous DE signal, the output will be a piecewise continuous DE signal. Since the output is a DE signal, it will be absent at all $(t, 0) \in T$. If the input s is free of chattering Zeno conditions, then the output will also be free of chattering Zeno conditions.

The communication points should include at least every $\tau \in T$ where $s(\tau) \neq \varepsilon$.

3.13 Discrete Time Delay

The output is a time-delayed input.

- DE input signal x;
- DE output signal y;
- Real parameter d, where $d > 0$.

For all $(t, n) \in T$, the output is

$$y(t, n) = \begin{cases} x(t - d, n) & \text{if } t \geq d \\ \varepsilon & \text{otherwise} \end{cases}$$

Discussion. The communication points should include every (t, n) where $x(t, n) \neq \varepsilon$, and also every $(t + d, n)$.

Notice that we do not attempt to define this for CT inputs because such time delays are extremely difficult to realize in simulation. In theory, they have an uncountably infinite state space. And they wreak havoc with step-size control mechanisms in variable-step solvers.

3.14 Discrete Microstep Delay

The output is a microstep-delayed input.

- DE input signal x;
- DE output signal y;

For all $(t, n) \in T$, the output is

$$y(t, n) = \begin{cases} x(t, n - 1) & \text{if } n \geq 1 \\ \varepsilon & \text{if } n = 0 \end{cases}$$

Discussion. The communication points should include at least every (t, n) where $x(t, n) \neq \varepsilon$ and also $(t, n+1)$. This component, therefore, requires a mechanism for ensuring zero step advancement of superdense time (which the microstep only advances).

Notice that we do not attempt to define this for CT inputs. Indeed, if presented with a CT input, this component will produce a rather odd output signal, one whose initial value is always absent, and subsequent values are present. If the input is a piecewise continuous DE signal, then the input is always absent at microstep zero, and the output will also be piecewise continuous. If the input is free of chattering Zeno conditions, then the output will be free of chattering Zeno conditions. In this case, the final value of the input is also ε, so the final value of the output will be ε.

Notice further that we do not generalize this component to have a parameter m to delay by m microsteps. See the discussion in Section 4.4 below for the reason for this.

4 Composition Test Cases

A hybrid cosimulation FMI standard that enables definition of the above components provides a rich framework for composition of discrete and continuous simulation tools. Any such standard should be able to unambiguously define FMUs that realize such components and should ensure that host simulators are capable of executing these FMUs. Such capabilities can be verified using unit tests that check each of the above components individually by providing a range of inputs and verifying that the outputs match the ideal (up to some precision, where appropriate). But such unit tests are not quite sufficient. We also need to ensure that interactions between multiple components behave correctly.

In this section, we discuss some test cases that combine a few of the above components, and give acceptance criteria that define correct behavior. These test cases are, in effect, constraints on master algorithms. Host simulators that conform with the standard must implement master algorithms that satisfy these acceptance criteria.

4.1 Synchronous Events

This test case checks that multiple components with discrete timed behavior coordinate their representations of time. Consider the composition shown in Figure 4. This has three components:

1. A Periodic Discrete Signal Generator with period p=1/3 and a=1.
2. A Periodic Discrete Signal Generator with period p=2/3 and a=1.
3. A Sampler with DE input x

The test criterion is that the output of the Sampler should equal the output of second Periodic Discrete Signal Generator at all superdense times. More generally, we would like the periods to be $p = q$ and $2q$, where q is a representable time, given as 1/3 in our test case.

Discussion. FMUs may internally use representations of time that are different from that of the host simulator. This test criterion is intended to ensure that no matter how the FMU and host simulator internally represent time, the Sampler and Periodic Discrete Signal Generator semantics are respected. This test case also checks for a well-defined notion of simultaneity. In particular, the periods chosen are not

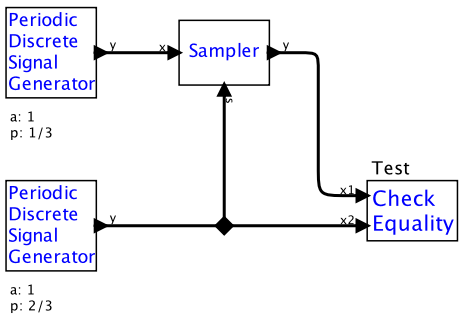

Figure 4: Test case for sampling of discrete event signals.

exactly representable with double precision floating point numbers, and the test is intended to ensure that despite any roundoff errors, every second output of the first Periodic Discrete Signal Generator is exactly synchronous with every output of the second one. The Test component must see the events at the same communication point in superdense time.

4.2 Integrating Discontinuous Signals

Figure 5 shows a test case that integrates a discontinuous input. The output is continuous, and should match Figure 6. Every acceptable test result will produce at least one output sample at the times of the discontinuities of the output of the Periodic Piecewise Constant Signal Generator. Samples in between these points in time are optional, and may depend on the step-size control algorithm used by the host simulator.

Discussion. The key feature being tested here is that a host simulator does not get confused by a signal that has two distinct values at the same (real) time, at distinct microsteps, and that a sequence of values at a (real) time does not affect the output of an Integrator, except that the final value at a discontinuity becomes the initial value for the next integration interval.

4.3 Integrating Glitches

Figure 7 shows a test case that verifies that the Integrator output is unaffected by input values whose duration is zero. In this test case, a constant-valued signal is modified using an Adder so that its value at integer-valued times sequences from 1 to 2 and back to 1, without time elapsing. These **glitches** have zero width, and hence should not affect the output of the Integrator.

4.4 Zero-Delay Feedback

Figure 8 shows a test model using a Zero-Crossing Detector in a feedback loop. The Integrator with Reset is integrating a constant 1, and hence will produce a line with unit slope. When that line crosses 1, the Zero-Crossing Detector is triggered. The event it produces is fed back through a Discrete Microstep Delay and through a Triggered Constant, which provides a discrete event with value 0 to the reset input of the Integrator with Reset.

The expected output is shown in Figure 9. Because of the approximate nature of the Zero-Crossing Detector (see Section 3.10), the times at which the reset occurs and the value at which it is triggered are approximate, so a regression test needs to specify a tolerance.

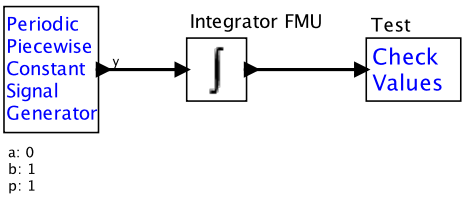

Figure 5: Test case for integrating discontinuous inputs.

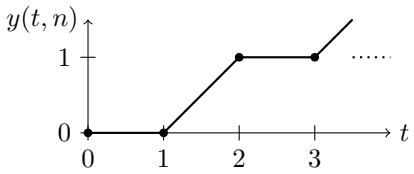

Figure 6: Result of the composition of components shown in Figure 5. The dots show required samples.

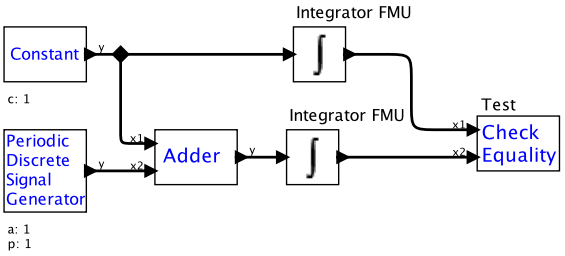

Figure 7: Test case for integrating glitches.

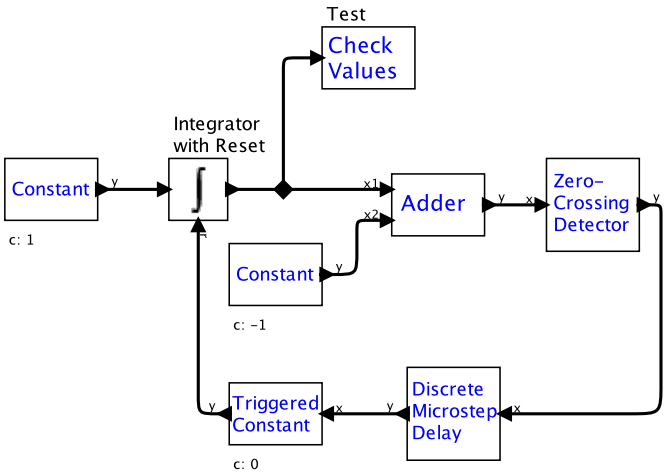

Figure 8: Test case for zero-delay feedback. The required output is shown in Figure 9.

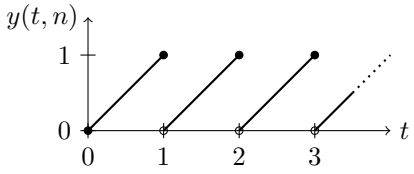

Figure 9: Required output from the test model in Figure 8. The unfilled dots show values that occur only at microsteps $n \geq 2$, whereas the filled dots and lines show values at $n = 0$ and $n = 1$.

In the plot in Figure 9, the filled and unfilled dots are required samples, occurring at microsteps 0 and 2 respectively. Samples in between are optional and may depend on the step-size control algorithm of the host simulator. Specifically, at time $t = 1$, the output of the Integrator With Reset should be

$$\begin{aligned} y(1,0) &= a \\ y(1,1) &= a \\ y(1,2) &= 0 \end{aligned}$$

where $a \approx 1$.

Notice that the reset actually occurs in microstep 2, because the event at the output of the Zero-Crossing Detector occurs at microstep 1, and it is then delayed by one additional microstep. In this particular instance, the Discrete Microstep Delay in the feedback loop might not seem to be required because the input to the Zero-Crossing Detector is continuous, and by the definition of the Zero-Crossing Detector, it introduces a microstep delay when the zero crossing occurs in a continuous region of the input. Nevertheless, our test case includes a Discrete Microstep Delay for two reasons. First, it provides a test where microsteps explicitly go beyond 1. Second, the Zero-Crossing Detector introduces a microstep delay only for some inputs. So the presence of a microstep delay in the loop is not a static property, which complicates scheduling of the components. Specifically, the Discrete Microstep Delay is non-strict, meaning that its input at superdense time τ need not be known to retrieve its output at τ. A scheduler can take this into account to break the apparent dependency loop created by the feedback. The Zero-Crossing Detector, however, is only non-strict at microstep zero (because its output is always absent at microstep zero). Hence, without the Discrete Microstep Delay, we would have a causality loop at all microsteps but zero. A master algorithm would have to ensure that at the input to the Zero-Crossing Detector, $m_t = 0$ for all t. In general, this is difficult to ensure.

Discussion. A subtle point raised by this composition is that the master algorithm needs to "know" at any time t when all signals have reached their final microstep. Specifically, it is not sufficient to stop incrementing microsteps at time t when all signals become absent at time t. First, CT signals never become absent at time t, so the mere presence of a CT signal will foil this strategy. Second, the Discrete Microstep Delay may have an absent input, and yet, in the next microstep, produce a non-absent output.

Our assumption is that each FMU constrains step sizes so that prior to reaching the final microstep of its outputs, it prevents the master algorithm from advancing time. It only permits advances in microstep. When no component does this, the master algorithm can assume that all signals have reached their final microstep.

For simplicity, we have constrained Discrete Microstep Delay so that it can only produce a non-absent output at the very next microstep. I.e., it implements a unit microstep delay, not an m-step microstep delay. It is certainly possible to generalize this mechanism to allow an m-step microstep delay, but we have not encountered any need for such generality, so we have not included this in the test cases for a hybrid cosimulation standard.

The Discrete Microstep Delay, the Discrete Time Delay, and both Integrator components have the property that they prevent causality loops when forming feedback compositions.[2] Specifically, these components have the unique property that they are able to produce outputs at a superdense time τ without any knowledge of their input values at τ. Such components are said to be **non-strict**. The Integrator with Reset is non-strict only with respect to the x input. It is strict with respect to the r input. Hence, strictness is a relation between input and output ports, and not a property of a component. This relationship needs to be declared as a static interface property of a component, so that a master algorithm can analyze causality properties, check models for constructiveness (see [9]), and optimize the schedule of invocation of FMUs. The FMI 2.0 specification supports an optional element for specifying such dependencies between input and output ports. For a detailed discussion of such causality relations, see [23].

[2]This assumes the Integrators implement variants 1 or 2, using an explicit integration method.

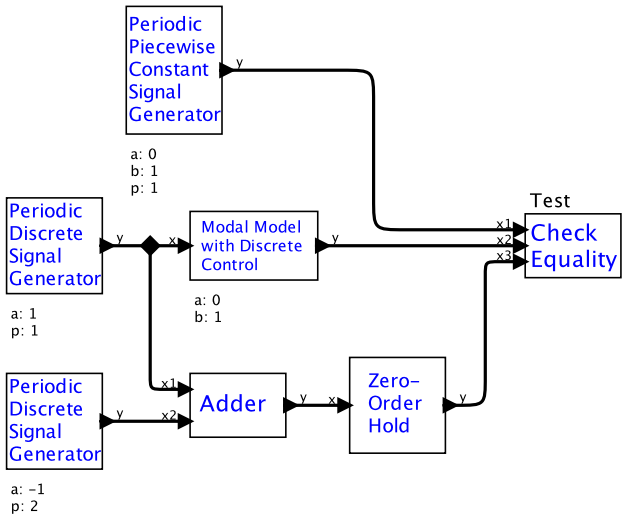

Figure 10: Test case verifying that various ways of producing piecewise-constant signals produce identical signals. The required output is shown in Figure 1.

4.5 Piecewise Constant Signals

The test components given in Section 3 offer a number of ways to construct the piecewise-constant signal in Figure 1. The model in Figure 10 tests that these various ways produce identical signals.

5 Conclusion

Hybrid cosimulation is a far from trivial goal. A standard that enables composition of simulation tools has two conflicting objectives. It needs to be sufficiently rigorous to define the meaning of a composition of components. And it needs to be flexible enough to embrace industry-standard and established simulators. The former demands a rigorous semantics, but the later creates pressure for less well-defined semantics.

This paper does not define such a standard, but instead enumerates a set of capabilities that such a standard must support to be useful. In particular, the paper enumerates a suite of components and test cases that cover a wide range of hybrid system behaviors. We have attempted to keep the components as simple as possible, but some cases, "as simple as possible" is not very simple. The most complex of the components are the Integrators, the Zero-Crossing Detector, and the Modal Models, which, although they represent widely used functionality in hybrid system simulation, have subtle properties.

The components defined in this paper are meant to be test cases, not library components. In most cases, to be useful in a library of components, they would need to be further generalized and parameterized. For example, the Zero-Crossing Detector would probably have a parameter that would specify whether upwards, downwards, or both directions of crossing are of interest. It could also have another parameter for a level crossing other than zero to detect. While such enhancements would increase the utility of the components, they do not enhance their value as test cases.

The following features, at least, are captured by the test cases:

1. A component may have zero latency from an input to an output.
2. A component can output discontinuous signals, and so that those signals are compatible with integration algorithms, they are piecewise continuous.
3. A component can reject step sizes based on internal accuracy estimates of numerical integration.
4. A component can reject step sizes in order to ensure that the simulation does not step over or skip a time at which a discrete event occurs.

5. A component can reject step sizes and suggest better step sizes in order to approximate event detection (such as zero crossings).
6. A component can react to DE inputs and produce DE outputs.
7. A component that reacts to DE inputs and produces continuous outputs produces piecewise-continuous outputs.
8. Discontinuities in signals can be exactly represented by samples of the signal, thanks to superdense time.
9. A component can monitor a continuous input and produce discrete events when some condition is detected, such as a zero-crossing.

References

[1] R. Alur, C. Courcoubetis, N. Halbwachs, T. Henzinger, P. Ho, X. Nicollin, A. Olivero, J. Sifakis, and S. Yovine. The algorithmic analysis of hybrid systems. *Theoretical Computer Science*, 138:3–34, 1995.

[2] R. Alur and D. L. Dill. A theory of timed automata. *Theoretical Computer Science*, 126(2):183–235, 1994.

[3] A. Benveniste and G. Berry. The synchronous approach to reactive and real-time systems. *Proceedings of the IEEE*, 79(9):1270–1282, 1991.

[4] T. Blochwitz, M. Otter, M. Arnold, C. Bausch, C. Clauß, H. Elmqvist, A. Junghanns, J. Mauss, M. Monteiro, T. Neidhold, D. Neumerkel, H. Olsson, J.-V. Peetz, and S. Wolf. The functional mockup interface for tool independent exchange of simulation models. In *Proc. of the 8-th International Modelica Conference*, Dresden, Germany, Mar. 2011. Modelica Association.

[5] D. Broman, C. Brooks, L. Greenberg, E. A. Lee, M. Masin, S. Tripakis, and M. Wetter. Determinate Composition of FMUs for Co-Simulation. In *Proceedings of the International Conference on Embedded Software (EMSOFT 2013)*. IEEE, 2013.

[6] L. P. Carloni, R. Passerone, A. Pinto, and A. Sangiovanni-Vincentelli. Languages and tools for hybrid systems design. *Foundations and Trends in Electronic Design Automation*, 1(1/2), 2006.

[7] D. Harel and A. Pnueli. On the development of reactive systems. In K. R. Apt, editor, *Logic and Models for Verification and Specification of Concurrent Systems*, volume F13 of *NATO ASI Series*, pages 477–498. Springer-Verlag, 1985.

[8] E. A. Lee. Modeling concurrent real-time processes using discrete events. *Annals of Software Engineering*, 7:25–45, 1999.

[9] E. A. Lee. Constructive models of discrete and continuous physical phenomena. *IEEE Access*, 2(1):1–25, 2014.

[10] E. A. Lee and A. Sangiovanni-Vincentelli. A framework for comparing models of computation. *IEEE Transactions on Computer-Aided Design of Circuits and Systems*, 17(12):1217–1229, 1998.

[11] E. A. Lee and H. Zheng. Operational semantics of hybrid systems. In M. Morari and L. Thiele, editors, *Hybrid Systems: Computation and Control (HSCC)*, volume LNCS 3414, pages 25–53, Zurich, Switzerland, 2005. Springer-Verlag.

[12] X. Liu, E. Matsikoudis, and E. A. Lee. Modeling timed concurrent systems. In *CONCUR 2006 - Concurrency Theory*, volume LNCS 4137, pages 1–15, Bonn, Germany, 2006. Springer.

[13] O. Maler, Z. Manna, and A. Pnueli. From timed to hybrid systems. In *Real-Time: Theory and Practice, REX Workshop*, pages 447–484. Springer-Verlag, 1992.

[14] E. Matsikoudis and E. A. Lee. On fixed points of strictly causal functions. In *International Conference on Formal Modeling and Analysis of Timed Systems (FORMATS)*, volume LNCS 8053, pages 183–197. Springer-Verlag, 2013.

[15] Modelica Association. Modelica textregistered - a unified object-oriented language for physical systems modeling: Language specification version 3.1. Report, May 27 2009.

[16] Modelica Association. Functional mock-up interface for model exchange and co-simulation. Report 2.0, July 25 2014.

[17] M. Otter, M. Malmheden, H. Elmqvist, S. E. Mattsson, and C. Johnsson. A new formalism for modeling of reactive and hybrid systems. In *Modelica Conference*. The Modelica Association, 2009.

[18] C. Ptolemaeus, editor. *System Design, Modeling, and Simulation using Ptolemy II.* Ptolemy.org, Berkeley, CA, 2014.

[19] P. Tabuada. *Verification and Control of Hybrid Systems: A Symbolic Approach.* Springer, 2009.

[20] M. M. Tiller. *Introduction to Physical Modeling with Modelica.* Kluwer Academic Publishers, 2001.

[21] S. Tripakis and D. Broman. Bridging the Semantic Gap Between Heterogeneous Modeling Formalisms and FMI. Technical Report UCB/EECS-2014-30, EECS Department, University of California, Berkeley, Apr 2014.

[22] B. P. Zeigler, H. Praehofer, and T. G. Kim. *Theory of Modeling and Simulation.* Academic Press, 2nd edition, 2000.

[23] Y. Zhou and E. A. Lee. Causality interfaces for actor networks. *ACM Transactions on Embedded Computing Systems (TECS)*, 7(3):1–35, 2008.