



HAL
open science

Bidirectional Reactive Programming for Machine Learning

Dumitru Potop-Butucaru, Albert Cohen, Gordon Plotkin, Hugo Pompougnac

► **To cite this version:**

Dumitru Potop-Butucaru, Albert Cohen, Gordon Plotkin, Hugo Pompougnac. Bidirectional Reactive Programming for Machine Learning. 2023. hal-04354071

HAL Id: hal-04354071

<https://inria.hal.science/hal-04354071v1>

Preprint submitted on 19 Dec 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Bidirectional Reactive Programming for Machine Learning

DUMITRU POTOP BUTUCARU, Inria, France

ALBERT COHEN, Google DeepMind, France

GORDON PLOTKIN, Google DeepMind, USA

HUGO POMPOUGNAC, Inria, France

Reactive languages are dedicated to the programming of systems which interact continuously and concurrently with their environment. Values take the form of unbounded streams modeling the (discrete) passing of time or the sequence of concurrent interactions. While conventional reactivity models recurrences forward in time, we introduce a symmetric reactive construct enabling backward recurrences. Constraints on the latter allow to make the implementation practical. Machine Learning (ML) systems provide numerous motivations for all of this: we demonstrate that reverse-mode automatic differentiation, backpropagation, batch normalization, bidirectional recurrent neural networks, training and reinforcement learning algorithms, are all naturally captured as bidirectional reactive programs.

1 INTRODUCTION

Conventional machine learning (ML) frameworks offer a tensor-centric view of the design and implementation of deep neural networks. But ML models do not stand by themselves as pure tensor functions. ML applications typically *interact with an environment and often operate on streams of data* collected and processed over time. For instance, the reactive control of a self-driving car [27, 40] operates on streams of data coming from sensors and controlling actuators. Training algorithms themselves embed a model into a reactive loop: it is decomposed into epochs and (mini-)batches allowing the efficient scheduling of computations and I/O, parameter updates, etc. The same applies to reinforcement learning (RL) agents [13]. Back to the automated driving example, *stateful* behavior is essential to taking into account previously-inferred facts such as speed limits, whether the current lane is a left turn etc., long after the acquisition of sensor inputs. Other examples of ML components embedded into stateful reactive feedback loops include model-predictive maintenance [37], control [8], and digital twins [21]. ML models themselves involve stateful constructs in the form of recurrent neural network (RNN) layers. When generating optimized code, even matrix products and convolutions in feedforward networks can be folded over time, using (stateful) buffering to reduce memory footprint [35]. In distributed settings, the efficient implementation of large models involves pipelined communications and computations [7, 17, 22], which amounts to locally recovering a streaming execution pattern.

Considering this broad range of scenarios, we observe that existing ML frameworks inadequately capture reactive aspects, raising barriers between differentiable models and the associated control, optimization, and input/output code. These barriers worsen the gap between ML research and system capabilities [3], particularly in the area of control automation where embedded ML engineering relies on undisclosed, ad-hoc implementations [40]. Our contributions are threefold:

Dataflow reactive machine learning. We first observe that conventional (dataflow or `fold_left`) reactive programming allows the natural specification of a variety of ML inference models, optimization and RL algorithms. By indexing data and parameter streams along the logical time steps of computation and interaction, one may integrate the internal processing of ML models with data pre-/post-processing stages, and embed everything into a reactive control loop.

Bidirectional reactive programming. We extend such a core reactive language with recurrences backward in time (`fold_right`). While classical reactive programs compute the values of cycle

$n + 1$ from those of cycle n , forming recurrences forward in time, bidirectional reactive programs allow a value of cycle n to depend on one of cycle $n + 1$.¹ This extension captures reverse-mode automatic differentiation, backpropagation through recurrent layers, training parameter update, distributed and pipelined training, bidirectional recurrent networks, and layers crossing data and time boundaries such as batch normalization; all of this without the need for an explicit gradient tape [34] or caching mechanism [24]. This extension preserves the bounded time and memory guarantees of reactive languages whenever recurrences backwards in time have a bounded horizon.

Formal semantics. The semantics of this bidirectional reactive language is defined from purely functional stream denotations. We provide a denotational semantics reminiscent of synchronous Kahn networks [10]. To ensure productivity (i.e. the absence of deadlocks), chains of backward dependences must *always eventually* terminate, so that execution can be implemented as *globally forwards and locally backwards* with finite buffering.

We unify independently developed ML models and algorithms on top of reactive constructs with deterministic formal semantics. This allows homogeneous and unambiguous specification, spanning from individual layers to RL algorithms. It also paves the way for the combination of efficient ML processing with the scheduling and resource allocation expressiveness of reactive languages.

2 RELATED WORK

This paper contributes to programming languages for both machine learning and reactive control systems. We address them separately.

Machine learning. While dataflow modeling has long been the norm in deep learning frameworks, the focus has typically been on efficient data processing, one input (batch) and one version of the trained parameters at a time. The design of ML compilers has been primarily driven by this common-case scenario, where both inference and training involve computations over data sets that have been already processed and stored. This contrasts with the needs of control applications that involve reactive and stateful interactions with an environment. Such interactions are represented outside of the ML model, which makes it difficult to provide implementations that satisfy stringent real-time and memory requirements [16, 27, 35, 40, 44].

| Network Type | Application scenario | | |
|---------------------------|-------------------------|-------------------------|-------------------------|
| | Inference | Training | Reinforcement Learning |
| Feedforward w/o BatchNorm | Stateless | Stateful, Forward | Stateful, Forward |
| Feedforward w/ BatchNorm | Stateless | Stateful, Bidirectional | Stateful, Bidirectional |
| Recurrent | Stateful, Forward | Stateful, Bidirectional | Stateful, Bidirectional |
| Bidirectional | Stateful, Bidirectional | Stateful, Bidirectional | Stateful, Bidirectional |

Fig. 1. Expressiveness requirements for neural networks operating on streams w.r.t. network type. The core language covers the pale and bright yellow cells. Bidirectional reactivity is required for the orange cells.

By comparison, we provide a language allowing the natural representation of reactive programs on streams, to express transformations over these such as automatic differentiation. It has been shown recently [31] that dataflow reactive programming [4, 5, 9, 32] captures a large class of ML applications—the light and yellow cells of Figure 1. In particular, the variety of ML constructs in these applications can be reduced to just four dataflow constructs. Bidirectional reactivity—the orange cells of Figure 1—improves expressiveness further, with support for the automatic differentiation of arbitrary networks and context-dependent operations such as batch normalization

¹“Bidirectional” borrows here from machine learning terminology. It is not about computing from updated outputs [29].

(whose semantics differ in inference and training modes). One no longer needs a different programming model for training in the data center vs. inference on the edge, for latency- vs. throughput-optimization, or a specific “out-of-the-language” toolkit for RL.

Reactive languages. We aim to facilitate the design and implementation of ML applications by exploiting abstractions and methods originating from reactive systems. Our main influence is the dataflow synchronous language Lustre, and its commercial version Scade, introduced in the 1980s [4] to allow the formal specification, the efficient execution and the automatic verification of deterministic concurrent systems with feedback and conditional control. Lustre, as well as the broader class of functional reactive languages [14],² do not allow the representation of bidirectional recurrences.

Lucid [2], one of the inspiration for Lustre, did introduce a `next` construct with semantics similar to the `post` construct we introduce for bidirectional recurrences. The designers of Lucid explicitly disallowed recurrences into the future however, limiting the usage of `next` in right-hand side expressions to non-recursive definitions.

3 LANGUAGE SYNTAX AND OVERVIEW

Figure 2 introduces the syntax of $R(ML)^2$ —for Reactive Machine Learning Meta Language. It is based on that of the classical dataflow reactive language Lustre [4, 5] from which it borrows its core set of dataflow primitives (in black) [6]. The Kleene star stands for comma-separated lists when applied to expressions and variables, for lists of equations in nodes, and for list of nodes at the top level of the program. The main node serves as the entry point to the computation; its inputs and outputs characterize the semantics of the whole program.

| | | | |
|--------|-------|---|--|
| e | $::=$ | $x \mid c$ | variable, constant stream |
| | | $f(e^*)$ | node instantiation and function application |
| | | $e \text{ when } e$ | sub-sampling (for conditional execution) |
| | | merge $e \ e \ e$ | merging (combining, up-sampling) |
| | | $e \text{ fby } e$ | recurrence forward in time (initialized delay) |
| | | post e | recurrence backward in time |
| eq | $::=$ | $x^* = e ;$ | stream equation |
| $node$ | $::=$ | node $f(x^*) \rightarrow (x^*) \ eq^*$ | node |
| $prog$ | $::=$ | $node^*$ | program |

Fig. 2. Syntax of the $R(ML)^2$ reactive language.

To this classical reactive core, $R(ML)^2$ adds the single new primitive `post`. It brings a fundamental extension over classical reactive languages: recurrences backwards in time, i.e. the ability to traverse time backwards. In a ML context, this is required to represent bidirectional networks [26], the training or reinforcement learning of stateful networks by means of backpropagation through time [25, 42], and batch normalization [18]. The expressive power of this construct will be discussed in Section 5 and its formal semantics will be defined in Section 6.

Note that $R(ML)^2$ features neither general recursion nor higher order functions. This choice, inherited from the synchronous reactive language Lustre, allows implementation in bounded time and memory.³ More expressive variants of the language could be considered in the future, in the spirit of FRP [14] or Reactive ML [23], but this is left for future work.

For brevity, we did not include tensor notation and tensor algebra operations in this syntax. A few of the examples below use conventional Python notations and Numpy or Keras operations, with square brackets for functional arrays.

²See also https://en.wikipedia.org/wiki/Functional_reactive_programming.

³E.g. for embedded and real-time applications.

As noted in Figure 1 and justified in the next section, the core language (i.e. without `post`) allows the natural representation of inference and reinforcement learning, for feedforward networks and one-directional RNNs without batch normalization. This includes models with complex control such as gated mixture of experts [38]. In the case of feedforward networks, it can also represent forward- and reverse-mode differentiation, backpropagation, a wide range of (non-pipelined) training algorithms, and mixing ML models with more general reactive control. We dedicate Section 4 to presenting its expressive power from an ML perspective.

Note that synchronous reactive languages like Lustre have *clocks*, a dependent type system making explicit the otherwise implicit wellformedness of sampled stream expressions, and the associated conditional execution [12]. The inability to infer the clock of a variable results in the program being rejected as incorrect, while properly clocked reactive programs are called *synchronous* [4]. We leave clocks out of this paper for the sake of conciseness, since for the most part, applications to ML and bidirectional extensions are orthogonal to clock types and inference.

4 DATAFLOW REACTIVE MACHINE LEARNING

Informal semantic principles. The underlying reactive language without `post` allows the definition of (forward) dataflow *nodes* that compute *streams* of outputs from streams of inputs. It has a concurrent semantics that parallels that of the classical `fold_left` iterator in functional languages. While classical functional languages evaluate recurrences *in space*, left-to-right along lists or vectors, the iterations of a reactive node unfold *in time* along potentially infinite input streams of data. At each cycle, the inputs of one recurrence step are acquired along with the current value of the `fold_left` accumulator, known as the *program state* in reactive programming; then computations for this cycle take place, yield the associated outputs and the new program state (the value of the accumulator), the latter are propagated to the next cycle before its evaluation may start.

The motivations for choosing a concurrent dataflow semantics are similar in both reactive control [4] and ML applications [1]. It captures the intrinsic concurrency of sensors and actuators in control automation, and the intrinsic concurrency of compute graphs in ML; it natively models parallel and distributed execution, including pipelining [43], and it provides ample freedom to the compiler to implement graph-level optimizations such as different forms of fusion [39].

Furthermore, a reactive language generalizes the `fold_left` iterator by allowing conditional input/output. One cycle does not need to acquire a value on every input stream and produce a value on every output stream. Instead, any variable can be marked as *absent* in a cycle. Within each cycle, execution is data-driven: computations take place as soon as enough input variables have acquired a value. As a result, *variable uses need not be dominated by their definition*: the syntactic ordering of equations in a node is semantically irrelevant.

Core constructs. The reactive core of $R(ML)^2$ uses only four dataflow primitives to implement these semantic principles: function application, initialized delay, conditional (sub-)sampling, and conditional up-sampling a.k.a. merging. All other constructs of Lustre can be reduced to the four primitives by structural expansion, and thus can be implemented on top of core $R(ML)^2$. The usual scalar operators ($+$, $*$, etc.), lifted to tensors through point-wise application, and other pure functions imported from a classical (non-reactive) language can be used in $R(ML)^2$ programs. The function application primitive of $R(ML)^2$ lifts them to streams by pointwise application at each time step. Among these pure functions we mention the (eager) conditional expression on streams `if e_1 then e_2 else e_3` , which does not need special handling as a primitive.

The remainder of this section explores these constructs and their intuitive semantics in greater detail. Much of the discussion is based on the example of Figure 3, which implements the gradient descent of a two-layer feedforward network. To simplify the presentation, we apply the layers

to scalars instead of higher-dimensional tensors, and we omit the batch dimension. The back-propagation control structure is not simplified however. The model has two parameters (in layer dense); it exercises all $R(ML)^2$ primitives, including forward recurrences—in blue—and backward recurrences—in red—and as we shall see, it already goes beyond the ML state of the art.

```

1  node param (init)->(o)
2  o = init;
3
4  node dense(i)->(o)
5  b = param(0.0);
6  k = param(1.0);
7  o = k * i + b;
8
9
10
11 node multiply(i)->(o)
12 o = i * i;
13
14 node app(i)->(o)
15 x = dense(i);
16 o = multiply(x);
17
18 node app_pipe(i)->(o)
19 x = dense(i);
20 y = 0.0 fby x;
21 o = multiply(y);
22
23
24
25 node mse_grad_i(i,gt)->(o)
26 o = 2.0 * (i-gt);
27
28 node infer(i)->(o)
29 o = diff_app_pipe(i,false,0.0,0.0);

node diff_param(init,bp,do)->(o)
o = init fby o1;
o1 = o + merge bp do 0.0;
node diff_dense(i,bp,do)->(o,di)
b = diff_param(0.0,bp,db);
k = diff_param(1.0,bp,dk);
o = k * i + b;
di = k * do; (* ∂o/∂i = k *)
dk = i * do; (* ∂o/∂k = i *)
db = do; (* ∂o/∂b = 1 *)
node diff_multiply(i,bp,do)->(o,di)
o = i * i;
di = do * 2.0 * i when bp; (* ∂o/∂i = 2i *)
node diff_app(i,bp,gt,learn_rate)->(o)
x,di = diff_dense(i,bp,dx);
o,dx = diff_multiply(x,bp,do);
do = -learn_rate * mse_grad_i(o when bp,gt);
node diff_app_pipe(i,bp,gt,learn_rate)->(o)
x,di = diff_dense(i,bpi,dx);
y,bpi,dx = diff_fby1(0.0,x,bp,dy);
o,dy = diff_multiply(y,bp,do);
do = -learn_rate * mse_grad_i(o when bp,gt);
node diff_fby1(init,i,bpo,do)->(o,bpi,di)
o = init fby i;
bpi = post bpo;
s = merge bpo do ((post s) when not bpo);
di = s when bpi;
node train(i,gt)->(o)
_ = diff_app_pipe(i,true,gt,0.01);

```

Fig. 3. Complete gradient descent implementation in $R(ML)^2$. On the left (lines 1-21) the specification (node app) and a pipelined version of it (node app_pipe). On the right, the gradient descent implementation for every node, obtained by structural translation not detailed in this paper. At the bottom, inference and training nodes obtained by specialization of the gradient descent node diff_app_pipe.

The two-layer feedforward network implements the function $\text{app}(i) = \text{multiply}(\text{dense}(i)) = (k \times i + b)^2$, where k and b are parameters initialized with values 1.0 and 0.0, respectively. During training, the objective is to perform gradient descent on the parameters in order to reduce the Mean Square Error (MSE) with respect to a ground truth gt . If the ground truth provided for input i is $gt(i) = (2.0 \times i - 3.0)^2$, the parameters will converge towards 2.0 and -3.0, respectively.

4.1 Hierarchy, dataflow equations, normalization

An $R(ML)^2$ program consists of a list of dataflow *nodes*. Each node comprises an *interface* and zero or more *equations* connected into a dataflow graph by means of variables. In an ML context, $R(ML)^2$ dataflow nodes represent elementary computations, layers, inference, training, or the full application. A dataflow node such as app can *instantiate* another. Semantically, dataflow instantiation (which we inherit from Lustre) has the same effect as inlining the code of the instantiated node.⁴ Due to this inlining semantics, while code may be reused across multiple instances of the same node, the storage of stateful nodes is *not* shared between instances. In our example,

⁴The param node deserves a specific note: it has dedicated, non-standard translation during the synthesis of backpropagation code, as discussed in Section 4.5. Yet in inference mode param translates to the identity function. For consistency throughout this paper, we assume that the synthesis of backpropagation code is performed prior to any inlining of param.

node `diff_dense` contains two instances of node `diff_param`, each one independently storing the value of one parameter during training (using the `fbv` primitive at Line 2).

Dataflow dependences between primitives of a node are either explicitly represented using named variables, or implicitly within expressions. For instance, in node `dense`, in Line 8, the operator “+” takes as input the output of operator “*”. It is always possible to transform equations with complex expressions into multiple single-primitive equations by exposing implicit dataflow dependences with fresh variables. This process is traditionally referred to as *normalization* and is analogous to the administrative normal form in continuation-passing style. The result of this process for node `app`, combined with the inlining of the instantiation of `dense` and `multiply` (for easier reference in simulations), is provided in Figure 4. Note in lines 4-5 the single-primitive equations produced as part of the normalization of the dataflow expression in Line 8 of node `dense`.

4.2 Function application

As explained earlier, at each execution cycle, an equation performs its computation as soon as its variables have acquired a value are are known as absent.

```

1  node app(i)->(o)
2    b = 0.0;
3    k = 1.0;
4    t = k * i;
5    x = t + b;
6    o = x * x;
```

Fig. 4. Inlining and normalization of the `app` node

| cycle | 0 | 1 | 2 | 3 | 4 |
|----------|-----|-----|--------|---|-----|
| t | 1 | 2 | 1.5 | □ | 1 |
| b | 0 | 0 | 0 | □ | 0 |
| x | ↓ 1 | ↓ 2 | ↓ 1.5 | □ | ↓ 1 |
| o | ↓ 1 | ↓ 4 | ↓ 2.25 | □ | ↓ 1 |

Fig. 5. Sample chronogram of the equations in lines 5-6 of Figure 4. Brown arrows represent data dependences.

The function application primitive of $R(ML)^2$ incorporates classical (non-reactive, non-streaming) functions into our reactive language by means of pointwise application. At each cycle, the function application primitive will wait for one value on each input variable before calling the classical function and then assigning its results to the output variables. If one input or output is absent during the cycle, then all inputs and outputs must be absent and the function is not called. If we denote with $t_0 < t_1 < t_2 < \dots$ the indices of the cycles where the inputs and outputs of the function application “ $o_1, \dots, o_m = f(i_1, \dots, i_n)$ ”, are present and with y_n the value of a variable y at cycle n , then we will have, for all $n \geq 0$: $o_{1_{t_n}}, \dots, o_{m_{t_n}} = f(i_{1_{t_n}}, \dots, i_{n_{t_n}})$. We depict in Figure 5 a possible execution trace—known as a *chronogram*⁵—of the equations in lines 5-6 of Figure 4. The trace is 5 cycles long (indices 0 to 4). In cycle 3 all variables are absent—the □ symbol.

Notice how, inside each cycle, the computation is *causal*, with the producer of a variable value being executed before all consumers. Function applications can be composed into arbitrarily complex (hierarchical) stateless computations, which can represent the computation of layers (such as `dense` in Figure 3), or full feedforward models, such as `app`. The main constraint here is that *the resulting dataflow graphs are acyclic*.⁶

Most functions used in ML applications are pure functions such as additions, convolutions, `tanh`, `sigmoid`. However, ML practice also makes significant use of functions that are not pure. These are random functions used in initializers (e.g. `glorot_uniform`, `orthogonal`) or in dropout layers.

4.3 Stateful computations

We have seen in the previous section how function applications can be composed into arbitrarily complex stateless computations represented with acyclic dataflow graphs. However, *ML practice*

⁵A chronogram displays the successive values of streams of interest at every cycle index in a given range. Cyclic execution allows us to graphically align multiple streams, since indexes coincide with computation cycles.

⁶In a cyclic graph formed only of function applications, a computation would depend on its own input, resulting in deadlock.

always involves stateful behaviors implementing forward recurrences along the time direction. As we shall see, these recurrences are mandatory in the training process (to store the parameter values), but they may already appear at specification level, as in recurrent neural networks (RNNs). Beyond the ML code itself, ML-based systems design also involves significant stateful components, such as data pre-processing involving sliding windows⁷ or efficiently-scheduled pipelined implementations [17, 43]. The R(ML)² language construct and a few applications are discussed in this section, while more advanced uses of stateful behaviors in ML are discussed later, in Section 4.5.

4.3.1 Initialized delay. The R(ML)² and Lustre statement allowing to cover all these needs is `fbv`. Its general form is $y = i \text{ fby } e$; where i and e are expressions of the same type and y is a stream variable. Its semantics is that of a loop-carried dependence, and it requires that i , e and y are all present at the same cycles. We will denote with $t_0 < t_1 < t_2 < \dots$ the indices where i , x and y are all present. Then, $y_{t_0} = i_{t_0}$ and $y_{t_n} = e_{t_{n-1}}$ if $n > 0$. Figure 6 shows one possible execution trace for a `fbv` equation.

| cycle | 0 | 1 | 2 | 3 | 4 | 5 |
|----------|-----|-----|---|-----|-----|-----|
| i | 3.0 | 5.1 | □ | 6.1 | 3.0 | 2.2 |
| x | 4.3 | 0.8 | □ | 3.3 | 1.9 | 7.7 |
| y | 3.0 | 4.3 | □ | 0.8 | 3.3 | 1.9 |

Fig. 6. Chronogram of $y = i \text{ fby } x$. Forward dependences in blue. In brown, the dependence of y on i in the first cycle

Notice that, unlike function application, the output of `fbv` does not depend *within the current cycle* on its second input. This allows the construction of cyclic dataflow graphs, such as the one in

```

1  node counter()->(o)
2    o = 0 fby u;
3    u = o + 1;

```

Fig. 7. Counter node in MLR

Figure 7. In the first cycle (of index 0) its output (the value of variable o) is the value of the first input, which is 0. In cycle 1, it is the value of its second input (variable u) in cycle 0, which is 1. Inductively, we can prove that in every cycle n the value of o is n .

Note that each `fbv` equation of a node defines one element of the recurrence accumulator which defines the program state. Thus, the counter node above has a state formed of one integer, whereas the state of node `app` in Figure 4 is formed of two floating-point values. For a node n that has not yet been inlined, the state is formed of the `fbv` equations defined directly in n , plus the state of all the node instances of n .

4.3.2 Modeling recurrent neural networks. We start to showcase the expressive power of `fbv` by modeling recurrent neural networks (RNNs), taking as example the classical Long Short-Term Memory (LSTM) [15]. RNNs are stateful networks, where one cycle passes a state that is used in the next. We provide in Figure 8 an implementation of the LSTM later.⁸

The presentation of the LSTM layer is meant to facilitate its comparison with classical ML frameworks such as Keras, preserving the separation between the `cell` node `lstm_cell` representing the function iterated within the recurrence from the `lstm` node which implements the reduction over time. Parameter initialization follows the Keras default: `glorot_uniform` for the kernel, zeros for the bias, or `thogonal` for the recurrent state. Notice how the two LSTM-specific recurrence state elements (c and h) are encoded in node `lstm` using two `fbv` equations.

Pairing the `fbv` operators within the top-level non-recurrent node is mainly due to the limitations of existing ML frameworks. Using the node hierarchy of R(ML)², we could restructure the LSTM cell to emphasize its internal organization into *gates*, as defined in the literature [15]. Collecting `fbv` operators for efficient iteration may be deferred to the compiler (see Pompougnac et al. [31], Section 4).

⁷Also known as rotating buffers.

⁸Node `dense1` provides here a more faithful implementation of the Keras layer `Dense` than the node `dense` of Figure 3.


```

1 node dense1(units, size_x, x) -> (o)
2   kernel = param(glorot_uniform([units, size_x]));
3   bias = param(zeros([units]));
4   o = matmul(kernel, x) + bias;
5 node lstm_cell(units, size_x, h, c, x) -> (h_next, c_next)
6   zr = reshape([4, units], dense1(4 * units,
7     units + size_x, concat(x, h)));
8   c_next = sigmoid(zr[0]) * c + sigmoid(zr[1]) * tanh(zr[3]);
9   h_next = sigmoid(zr[2]) * tanh(c_next);
10 node lstm(units, size_x, x) -> (h_next)
11   h_next, c_next = lstm_cell(units, size_x, h, c, x);
12   h = param(orthogonal([units])) fby h_next;
13   c = param(orthogonal([units])) fby c_next;

```

Fig. 8. LSTM layer (square brackets index functional arrays).

4.3.3 Pipelined inference. Given the complexity of mapping ML applications onto heterogeneous hardware, efficient designs may involve source-level mapping and scheduling [17, 43]. This is cumbersome at best with existing ML frameworks. Figure 3 shows how to split node app in two pipeline stages by inserting a fby operator between the two layers. The result is shown in node app_pipe, and the corresponding backpropagation code in node diff_app_pipe.

4.3.4 Data preprocessing. Finally, using only function application and fby, one may implement non-trivial data pre-processing typically occurring in conjunction with ML models.

```

1 node mlp(shape, x) -> (o)
2   y = dense1(100, shape, x);
3   z = relu(x);
4   o = dense1(1, 100, y);
5 node window(x) -> (w)
6   x1 = 0.0 fby x;
7   x2 = 0.0 fby x1;
8   x3 = 0.0 fby x2;
9   w = [x, x1, x2, x3];
10 node timeseries(x) -> (o)
11   b = window(x);
12   o = mlp(4, b);

```

We showcase this in Figure 9 with a time series application where inputs are buffered into a sliding window (also known as a rotating buffer) of size 4 before being fed to a Multi-Level Perceptron (MLP). The sliding window node uses a delay line where x_1 is x delayed by 1 cycle, x_2 is further delayed by 1 cycle, and finally x_3 is x delayed by 3 cycles. The sliding window output w is obtained by grouping all four variables into a vector.⁹

Beyond this small example, data pre-processing and post-processing can be arbitrarily complex, from signal processing to extract a spectral representation to using the output of the ML algorithm for model-predictive control.

Fig. 9. MLP with embedded data pre-processing.

4.4 Conditional execution and sampling

Piecewise functions (e.g. ReLU) and data sampling have long been part of ML practice. Yet general primitives such as the `tf.where` choice operator or `tf.cond`/`jax.lax.cond` for conditional execution¹⁰ are rarely used directly by programmers. However, the pressure to reduce the computational requirements of ML applications naturally results in rapidly growing popularity of conditional execution, such as the sparsely-gated mixture-of-experts layer [38].

By comparison, reactive programming exposes powerful conditional execution primitives. The danger associated with conditional execution in a concurrent system is non-determinism: when the producer and the consumer of a variable are executed on different conditions, the consumer may read an undefined value. This is why dataflow reactive formalisms propose correct-by-construction semantics ensuring that all values used in computations are well-defined. This is the approach we follow in $R(ML)^2$, borrowing the `when` and `merge` conditional execution primitives of Lustre [4, 5].

⁹Specific language constructs, like TensorFlow’s `stop_gradient`, are needed to avoid backpropagation through the data processing code, but these mechanisms are outside the scope of our paper.

¹⁰https://jax.readthedocs.io/en/latest/_autosummary/jax.lax.cond.html

| cycle | 0 | 1 | 2 | 3 | 4 |
|-------|-------|-------|-------|---|-------|
| c | true | false | true | □ | false |
| y | 2 | 7 | 5 | □ | 1 |
| x | ↓ 2 ↓ | □ | ↓ 5 ↓ | □ | □ |

 Fig. 10. Sample chronogram of $x = y$ when c .

In the dataflow setting of Lustre and $R(ML)^2$, where execution is driven by the arrival of data, conditioning the execution of a node in a cycle is done by making its input variables present or absent in that cycle. This is done using the (sub-)sampling operator when. In equation “ $x = y$ when c ”, it is required that variable c is of Boolean type and that c is present in a cycle if and only if y is present. Then, in cycles where c is present with value true, x takes the value of y . In cycles where c is absent or present with value false, x is absent and computations that depend on it are not executed. If c is absent, then x is absent. We provide in Figure 10 a possible execution trace. In cycles 0 and 2, c is present with value true, so the value of y is copied onto x . In cycles 1 and 4, both inputs are present but c is false so x is absent. Finally, all variables are absent in cycle 3.

Primitive “ $x = \text{merge } c \ y \ z$ ” defines a deterministic multiplexer. Its inputs must satisfy the following conditions:

- c is known as the *control stream* of the merge operator and must be of Boolean type;
- y is be present in a cycle *iff* c is present with value true;
- z is be present in a cycle *iff* c is present with value false;
- y and z have the same data type.

Under these conditions, x is a variable that is present *iff* c is present, which takes the value of y if c is true and the value of z if c is false. Figure 11 shows a possible execution trace. In cycles 0 and 2, c is present with value true, so the value of y is copied onto x and z must be absent. In cycles 1 and 4, c is present with value false, so the value of z is copied onto x and y must be absent. In cycle 3, all variables are absent.

| cycle | 0 | 1 | 2 | 3 | 4 |
|-------|-------|-------|-------|---|-------|
| c | true | false | true | □ | false |
| y | 2 | □ | 5 | □ | □ |
| z | □ | 3 | □ | □ | 1 |
| x | ↓ 2 ↓ | ↓ 3 ↓ | ↓ 5 ↓ | □ | ↓ 1 ↓ |

 Fig. 11. Sample chronogram of $x = \text{merge } c \ y \ z$.

As a first ML application of when and merge, Figure 12 describes a sparsely-gated mixture-of-experts layer [38]. We instantiate it for 3 experts, assuming that they are provided as nodes `expert1`, `expert2`, and `expert3`, which produce tensors of the same shape.

```

1 node gating_network(x)->(g1,g2,g3)
2 node expert1(x)->(o)
3 node expert2(x)->(o)
4 node expert3(x)->(o)
5
6 node mixture_of_experts(shape,x)->(o)
7   g1,g2,g3 = gating_network(x);
8   o1 = expert1(x when c1); c1 = (g1 != 0);
9   o2 = expert2(x when c2); c2 = (g2 != 0);
10  o3 = expert3(x when c3); c3 = (g3 != 0);
11  o = g1 * (merge c1 o1 zeros(shape)) +
12     g2 * (merge c2 o2 zeros(shape)) +
13     g3 * (merge c3 o3 zeros(shape));
    
```

Fig. 12. Sparsely-gated Mixture of Experts layer.

We assume that expert control is performed by node `gating_network`, which outputs at each cycle one floating point value per expert, serving as weights in the computation of the output (in lines 11-13). In a cycle where the weight corresponding to an expert is 0, the expert must not be computed. This behavior is obtained by sampling the network input x . For `expert1` this is done in line 8. Boolean variable $c1$ stores the Boolean activation condition. To compute the weighted sum output, $o1$, $o2$, and $o3$ cannot be added because they are not present at the same cycle. Instead, the merge operators of lines 11-13 extend the outputs $o1$, $o2$, and $o3$ to variables that are present at all cycles, padding with 0 in cycles where the experts do not produce a value.

Notice that Lustre and $R(ML)^2$ offer a modular form of logical time, by allowing subsets of equations, including `fbv` operators, to be executed under specific Boolean conditions. For instance, the experts or the gating network of Figure 12 can be themselves stateful/recurrent [11] involving

fby operators. For every such expert, the recurrences relate only the cycles where the expert is active. In the context of control applications—typically multi-periodic and involving multiple execution modes—this facilitates interfacing the ML code with the reactive system-level code driving it [28, 41].

4.5 Backpropagation training of feedforward networks

We are now ready to discuss the reactive modeling of gradient backpropagation when training neural networks. In the example of Figure 3, node `diff_app` implements backpropagation for the feedforward network `app`. Note that `diff_app` preserves all the inputs, outputs, and equations of `app`, structurally adding to them the error computation and backward path variables and equations and the control governing them. The new Boolean input `bp`, known as the *backpropagation signal*, is present at all cycles. Backpropagation is performed in cycles where `bp` is true. In these cycles, the node also receives a ground truth value `gt`. Based on the value of `gt`, the error is computed (on the current sample only), it is multiplied by the learning rate, and then backpropagated. Parameters are eventually updated, with the objective of reducing the error in future cycles.

Note that an execution of node `diff_app` can mix cycles of training with cycles of pure inference, by choosing the sequence of values of `bp`. This feature is needed in Reinforcement Learning [13], and our modeling approach supports it out-of-the-box. Pure inference and pure training implementations of our specification (like those produced by TensorFlow) can then be obtained by setting `bp` to constant false or to constant true, respectively, as we do in nodes `infer` and `train`.

The synthesis of the backpropagation and parameter update code from the specification is not detailed in this paper.¹¹ We only provide the result of this synthesis process to show that $R(\text{ML})^2$ allows its representation. Backpropagation consists in computing, for each value v computed on the forward path, an error estimation dv . In our backpropagation approach based on static program transformations, for each specification-level differentiable variable v we create in the backpropagating implementation another variable dv . The computation of the dv variables reverses all the causalities of the forward path. For instance, in node `dense` of Figure 3 variable `o` is computed from `k`, `i`, and `b`. Then, in node `diff_dense`, variables `dk`, `di`, and `db` all depend on `do`. Furthermore, backward path computations may depend on forward path values (when the forward path function is not affine). For instance, `dk` depends on `i`. The forward path depends on the backward path through the update of parameters only. This explains the form of the `diff_param` node, which accumulates the corrections `do` at every cycle where `bp` is present. The updated parameter value is fed to the next cycle regardless of the value of `bp`.

From a language design perspective, `param` is an ML-specific primitive (the only one used for the scope of this paper), which is expanded during synthesis of training code (not detailed in this paper) into `diff_param`. Further coverage of ML parameters, from a recurrent reactive specification perspective, follows in Section 5.3.

5 RECURRENCES BACKWARD IN TIME

The reactive core of $R(\text{ML})^2$, along with the time/space and state manipulation extensions of the previous section, already covers the expressiveness requirements of ML applications in the yellow cells of Figure 1. The particularity of these ML applications is that they only define data dependences (recurrences) forward in time—a value of a cycle cannot depend on values of future cycles.

But to cover the needs of modern Machine Learning, we also need to cover the applications in the orange cells of Figure 1. These applications feature dependences backwards in time coming

¹¹And neither are, for of the sake of the presentation presentation, the issues of selectively saving or loading the state of fby primitives.

| cycle | 0 | 1 | 2 | 3 | 4 | 5 |
|----------|-----|-----|---|-----|-----|-----|
| x | 4.3 | 3.0 | □ | 3.3 | 1.9 | 7.7 |
| y | 3.0 | 3.3 | □ | 1.9 | 7.7 | 1.9 |

Fig. 13. Sample chronogram of $x = \text{post } y$. In red, data dependences.

from 3 main sources: bidirectional networks [26], batch normalization [18], and backpropagation through time used in the training and reinforcement learning of recurrent networks [25, 42].

To this end, we extend reactive dataflow programming with a new primitive, with the form “ $y = \text{post } x$;”. This equation specifies that x and y are present at the same cycles and that, if we denote with $t_0 < t_1 < t_2 < \dots$ the indices of the cycles where x and y are both present, then for all n we have $y_{t_n} = x_{t_{n+1}}$. Figure 13 shows one possible execution trace for this equation. We start our exploration of the new primitive by noticing that “ $\text{post } (y \text{ fby } x)$ ” is always equal to x , whereas “ $y \text{ fby } (\text{post } x)$ ” is equal to x for all cycles save the first, where it is equal to y (since post discards the first value of x which cannot be recovered).

5.1 Backpropagation on pipelined applications

The simplest use of post is in extending the backpropagation method of Section 4.5 to pipelined applications. Recall from Section 4.3.3 that fby primitives can be introduced in feedforward specifications to mark the frontier between pipeline stages of the inference algorithm.

These fby primitives do not create cycles in the dataflow of the node, and backpropagation code can be synthesized for such specifications without considering the full complexity of time recurrences, which is covered in the next section. More precisely, backpropagation in pipelined feedforward applications simply requires the realignment of backward propagation with the forward one by means of post primitives inserted on the backwards path in places where on the forward path a fby primitive exists.

We illustrate this on Figure 3: `app_pipe` is a pipelined version of `app`, and `diff_app_pipe` shows the effect of injecting backpropagation code. In the process, the unique fby primitive is replaced with an instance of the `diff_fby1` node which not only includes the forward-path fby that separates the pipeline stages, but also includes the backward-path post primitives that realign backpropagated values with the forward path computations.¹²

In this chronogram, the value i_1 is produced by the first pipeline stage (node `dense` in our example) in cycle 1, but is delayed to be processed by the second pipeline stage in cycle 2. In cycle 2, the backpropagation indicator is true in the second stage (`bpo = true`) and value do_2 must be backpropagated to the cycle 1 of the first pipeline stage. This is realized by the two post primitives in node `diff_fby1`. The first one simply anticipates the values of `bpo` to produce `bpi`. The second is part of a more complex protocol that produces an output `di` at a different rate from the input `do`.

| cycle | 0 | 1 | 2 | 3 | 4 | 5 |
|------------|-------|--------|--------|---|--------|--------|
| i | i_0 | i_1 | i_2 | □ | i_4 | i_5 |
| o | 0.0 | i_0 | i_1 | □ | i_2 | i_4 |
| bpo | false | false | true | □ | false | true |
| do | □ | □ | do_2 | □ | □ | do_5 |
| bpi | false | true | false | □ | true | false |
| di | □ | do_2 | □ | □ | do_5 | □ |

Fig. 14. Backpropagation in pipelined applications: a possible trace of node `diff_fby` of Figure 3. In blue, dependences forward in time. In red, dependences backward in time.

5.2 Globally-forward, locally-backward propagation

Note that, while the fby and post primitives of $R(\text{ML})^2$ both embody recurrences over time, they are not symmetrical, as classical functional programming iterators `fold_left` and `fold_right` are.

¹²To simplify the code in Figure 3, we assume in the definition of `diff_fby1` that its first argument is a constant, requiring no backpropagation. A full-fledged version of fby backpropagation is provided in Section 5.3.

As any execution has a start (in cycle 0) forward recurrences always have a well-defined initialization cycle. By comparison, reactive programming does not identify an end cycle and executions are *a priori* infinite. For this reason, no initial value is specified by the post primitive.

Like for fby, the output of a post does not depend on its input inside a cycle. This allows the creation of cyclic dataflow graphs, like the counter node of Figure 7, which represent recurrences in time. Replacing in the counter node the fby by a post produces the node of Figure 15(left). We have pictured in Figure 15(right) the data dependences specified by its equations.¹³ Clearly, the program can compute no output, as all values depend on the infinite future.

Whenever a program includes a post primitive in a dataflow cycle, we must ensure that, at every execution cycle in the program execution, the dependence on the future is eventually broken. In formal terms, the predicate p identifying the cycles where the dependence on the future is broken must satisfy the temporal logic [30] formula *always eventually p*.

```

1  node pcounter()->(o)
2  o = post u;
3  u = o + 1;

```

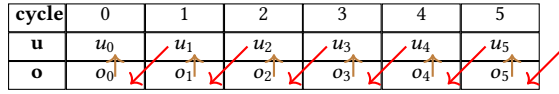


Fig. 15. Dependences on the infinite future; backward dependences in red, instantaneous ones in brown.

```

1  node backfill(i,bp)->(o)
2  o = merge bp
3    (i when bp)
4    ((post o) when not bp);

```

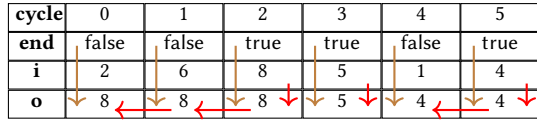


Fig. 16. Bounded broadcast towards the past.

This requirement only applies to post, not to fby. Thus, the execution of the program can proceed *globally forwards, locally backwards*.

In Figure 16(left), the condition breaking the dependence on the future is provided by the bp input. When bp is true, the value of the input is sampled at the current cycle and broadcasted to all previous cycles, until reaching an earlier cycle where bp was true. One execution trace of backfill is pictured in Figure 16(right). Notice how the cycles where bp is true cut the execution into intervals, with the output remaining constant over each interval.

5.2.1 Batch normalization. We are now ready to study the *batch normalization* layer [19], which is one of the most complex from a reactive control perspective. In training mode, the forward path is provided in Figure 17 by node batch_norm.

We operate on batches of samples of the input x . Batch termination is identified by the Boolean input batch_end. Node mean_var computes the mean and variance, propagating data both forward and backward in time. The number of samples in a batch and the partial sum of its elements are accumulated in the forward direction in variables cnt and sum, respectively. Node fby_end applies forward in time recurrences piecewise: whenever end is true, the input is dropped and in the next cycle the output is set to init. In other cycles, the output is set to the delayed input. An execution trace of fby_end is provided in Figure 18.

In node mean_var, instances of fby_end are used to reset the accumulation of partial sums at batch ends, so that for each batch starts with an accumulator set to 0. At batch ends, the value of expression “sum/cnt”—the batch mean—is sampled and broadcast to all batch cycles by the first instance of backfill as described above. Using the batch mean, one may again accumulate in forward direction, in variable var_sum, the partial sums needed to compute the variance. At batch end, the variance is broadcast to the whole batch by the second instance of backfill.

¹³Brown dependences inside a cycle are due to the + operator. Red dependences between cycles are associated with post.

```

1  node fby_end(shape, end, init, i)->(o)
2    i1 = if end then zeros(shape) else i;
3    o = if (true fby end) then init else (init fby i1);
4  node mean_var(shape, x, batch_end)->(mean, var)
5    (* fw computation of batch size *)
6    cnt    = 1 + fby_end([], batch_end, 0, cnt);
7    (* fw computation of batch sum *)
8    sum    = x + fby_end(shape, batch_end, zeros(shape), sum);
9    (* backfill batch mean *)
10   mean   = backfill(sum / cnt, batch_end);
11   var_sum = (x - mean) * (x - mean) +
12             fby_end(shape, batch_end, zeros(shape), var_sum);
13   (* backfill batch variance *)
14   var    = backfill(var_sum / cnt, batch_end);
15  node norm(shape, epsilon, mean, var, x)->(y)
16   gamma = fby_end(shape, batch_end, param(ones(shape)), gamma);
17   beta  = fby_end(shape, batch_end, param(zeros(shape)), beta);
18   y     = gamma * (x - mean) / (var + epsilon) + beta;
19  node batch_norm(shape, epsilon, x, batch_end)->(y)
20   mean, var = mean_var(shape, x, batch_end);
21   y        = norm(shape, epsilon, mean, var, x);

```

Fig. 17. Batch normalization.

| | | | | | | |
|-------|-------|-------|------|------|-------|------|
| cycle | 0 | 1 | 2 | 3 | 4 | 5 |
| end | false | false | true | true | false | true |
| init | 1 | 0 | 9 | 3 | 5 | 8 |
| i | 2 | 6 | 8 | 5 | 1 | 4 |
| o | 1 | 2 | 6 | 3 | 5 | 1 |

Fig. 18. Trace of node fby_end of Figure 17.

The computed batch mean and variance are then used to normalize the input, also using two trainable parameters γ and β . The use of `fby_end` in the context of trainable parameters will be covered in detail in the next section.

5.3 Backpropagation through time for RNN training

Consider the `lstm` node of Figure 8, which represents the recurrent layer LSTM. In its first execution cycle, the state accumulator variables h and c are initialized with the value of the two parameters. Then, at every cycle (including the first), the non-recurrent `lstm_cell` node is executed once, taking as input the current accumulator and the current value of the `lstm` input x , and producing the updated accumulators (of which one is also output). If we denote with h_n the value of h at cycle n , the computation of h_{n+1} causally depends on that of h_n for all n .

Recall that in a backpropagation training context, for every value v computed on the forward path, an error estimation dv is computed on the backward path, and the computation of the latter reverses all the causalities of the forward path. Thus, if the difference associated with h_n is denoted dh_n , we know that dh_n depends on dh_{n+1} for all n , and so the computation of dh_0 (to update parameters) depends on the infinite future.

As explained in Section 5.2, such a dependence on the infinite future cannot be effectively resolved: backpropagation must be reduced to a globally forward, locally-backward propagation, where the predicate determining when to restart backpropagation is *always eventually* true. Such a predicate is readily available when training with mini-batches. Indeed, RNN training samples consist of finite input sequences accompanied by ground truth values. In a training or RL, sequences are provided cyclically, possibly separated by samples meant for inference only. The result can be seen as a single stream of inputs, where the end of training sequences is marked by the Boolean input signal `end`.

```

1  node param_end(shape,end,p)->(o)
2    o = o1; o2 = o1;
3    o1 = fby_end(shape,end,p,o2);
4  node dense1_end(units,sex,x,end)->(o)
5    shape = [units,sex];
6    kernel = param_end(shape,end,
7      param(glorot_uniform(shape)));
8    bias = param_end([units],end,
9      param(zeros([units])));
10   o = matmul(kernel,x) + bias;
11  node lstm_cell_end(units,sex,h,c,x,end)->(h_next,c_next)
12    zr = reshape([4,units],dense1_end(4 * units,
13      sex + units,concat(x,h),end));
14    c_next = sigmoid(zr[0]) * c + sigmoid(zr[1]) * tanh(zr[3]);
15    h_next = sigmoid(zr[2]) * tanh(c_next);
16  node lstm_end(units,sex,x,end)->(h_next)
17    h_next,c_next = lstm_cell_end(units,sex,h,c,x,end);
18    h = fby_end(end,param(orthogonal([units])),h_next);
19    c = fby_end(end,param(orthogonal([units])),c_next);

```

Fig. 19. LSTM of Figure 8 extended with backpropagation reset.

```

1  node diff_fby_end(shape,end,init,i,bp,do)->(o,dinit,di)
2    o = fby_end(end,init,i);
3    rst1 = rst when bp; end1 = end when bp;
4    dinit = merge rst1 (do when rst1) zeros(shape);
5    di = merge end1 zeros(shape) (di1 when not end1);
6    di1 = post (merge rst1 zeros(shape)
7      (do when not rst1));
8  node diff_param_end(shape,end,p,bp,do)->(o,dp)
9    o = o1; o2 = o1; (* dup *)
10   do1 = do + do2; (* backpropagation for dup *)
11   o1,_,do2 = diff_fby_end(shape,end,p,o2,bp,do1);
12  node diff_lstm_end(units,sex,x,end,bp,dh_next)->(h_next,dx)
13    h_next,c_next,dh,dc,dx = lstm_cell_end(units,sex,h,
14      c,x,end,bp,dh_next,dc_next);
15    hp = diff_param(orthogonal([units]),bp,dhp);
16    cp = diff_param(orthogonal([units]),bp,dcp);
17    h,dhp,dh_next = diff_fby_end(end,hp,h_next,bp,dh);
18    c,dc,dc_next = diff_fby_end(end,cp,c_next,bp,dc);

```

Fig. 20. LSTM backpropagation training code.

agation code. Indeed, it is expected that RNN parameters are only updated once per training sequence. During backpropagation over a sequence, parameter updates arriving at each cycle must be accumulated before applying the sum when backpropagation reaches the first cycle of the sequence. This is the exact behavior of the backpropagation code for node `param_end` in Figure 20. Notice in line 12 the accumulation of differences. The behavior of `diff_fby_end` ensures that the accumulated update is only applied to parameters on sequence start cycles. The actual parameter update is performed by `diff_param` (see Figure 3). With these definitions, the backpropagation code of the LSTM may be obtained by structural translation, as pictured in Figure 20.

5.4 Bidirectional RNNs

We have seen that the 5 reactive primitives of $R(ML)^2$ —the 4 classical ones and `post`—allow to model batch normalization and backpropagation with reset, as well as the alternation of training and inference within RL algorithms. We complete our presentation of the expressive power of $R(ML)^2$ with the encoding of a bidirectional recurrent network [36].

Bidirectional networks feature recurrences in time in both directions. Thus, `post` operators are *needed in the model itself*. For such a specification to be productive, it must include the conditions allowing a globally forward, locally backward execution, as discussed in previous sections.

One may finally weave backpropagation with reset into the LSTM of Figure 8. The result of this source-to-source transformation is provided in Figure 19, where `fby` and `post` operators are replaced with the resettable versions `fby_end` in Figures 17 and 18 and `post_end` in Figure 21, respectively. Parameters also need to be wrapped into separate `param_end` nodes, which are particular instantiations of `fby_end` with the output fed back into their delayed input.

The `param_end` wrappers are required during synthesis of backpropagation code.

```

1  node post_end(shape,end,init,i)->(o)
2    i1 = if (true fby end) then zeros(shape) else i;
3    o = if end then init else (post i1);
4
5  node lstm_bi(units,shape,x,end)->(o)
6    h_next = lstm_end(units,shape,x,end);
7    h_pred,c_pred = lstm_cell_end(units,shape,h,c,x,end);
8    h = post_end(shape,end,param(orthogonal()),h_pred);
9    c = post_end(shape,end,param(orthogonal()),c_pred);
10   o = h_next + h_pred;

```

Fig. 21. Bidirectional LSTM layer.

We assume that the Boolean input variable `end` of the node identifies the cycles where backward recurrences begin.

The bidirectional LSTM layer is described in Figure 21. It builds upon the extension of the LSTM with backpropagation reset capability (nodes `lstm_end` and `lstm_cell_end` of Figure 19). Between every two cycles where `end` is true, the layer implements two independent recurrences in time, one forward (implemented by the `lstm_end` node instance) and one backward. The outputs of the two recurrences are summed at each cycle to produce the output of the layer.

6 SEMANTICS

As illustrated in the earlier examples, $R(\text{ML})^2$ significantly extends the expressiveness of reactive (synchronous) languages such as Lustre. In particular, the execution of a program involving `post` (recurrences backwards in time) may not compute all the stream variables at a cycle n before moving on to the next cycle $n + 1$.

Comparison with Kahn networks. As a first attempt at formalizing the semantics of $R(\text{ML})^2$, it is tempting to build upon the dataflow framework of Kahn [20]: concurrent programs whose asynchronous semantics can be defined by means of compositions of *monotonous stream functions*. Unfortunately this is not so simple. Consider, for instance, in Figure 16 the node `backfill` and its execution trace. Here, the values of `o` are not computed as a stream, where new values are always added at the end: the value of `o` at cycle 2 is known before those of cycles 1 and then 0.

Notations and conventions. If T is a set of values, T^* denotes the set of *finite streams of elements of T* , T^ω is the set of *infinite streams*, and $T^\infty = T^* + T^\omega$. To ease the presentation, we will consider a single set T encompassing all values occurring in $R(\text{ML})^2$ examples.

The empty stream is noted ε and \cdot is the stream concatenation operator; lifted to infinite streams inductively and allowing the concatenation of a finite stream (left-hand side operand) with a finite or infinite stream (right-hand side operand). The length of a stream s is noted $|s|$ —it is a natural number or ∞ . The n -th value of stream s is noted s_n , the sub-stream starting from index n and ending immediately before index m (i.e., m excluded) is noted $s_{n:m} = s_{\{i \in \mathbb{N} | n \leq i < m\}}$, and the suffix starting from index n is noted $s_n = s_{\{i \in \mathbb{N} | i \geq n\}}$. For example, $s_0 \cdot s_1$ denotes the stream whose head is s_0 and tail is s_1 .

Considering a program P in the syntax of Figure 2, we assume P has been normalized to an administrative form where every equation consists of a single primitive.

V denotes an environment of stream variables. In such an environment, $x := v$ denotes the binding of variable x to the value v , and $V(x)$ denotes the value of variable x in the environment V . $\llbracket e \rrbracket$ is a function from environments to environments: $\llbracket e \rrbracket(V)$ denotes the interpretation of expression e applied to the environment V . Let $\text{FV}(e)$ be the set of free variables of an expression e and $\text{FV}(x = e) = \text{FV}(e) \setminus \{x\}$. Let $\text{DV}(x = e) = \{x\}$ denote the (left-hand side) variables bound by an equation eq . These notations are lifted to tuples of equations, defining $\text{FV}(x^1, \dots, x^d = e^1, \dots, e^d) = \text{FV}(e^1, \dots, e^d) \setminus \{x^1, \dots, x^d\}$ and $\text{DV}(x^1, \dots, x^d = e^1, \dots, e^d) = \{x^1, \dots, x^d\}$. For the

sake of conciseness, when clear from the context, we overload x to denote a tuple of variables, e a tuple of expressions, and $x = e$ a tuple of equations.

Variables and semantic interpretations are defined within a directed-complete partial order (dCPO or simply CPO). We will later (somewhat loosely) refer to any CPO as a *domain* in the following. Given two CPOs D and D' , their product $D \times D'$ and the set of Scott-continuous functions $D \rightarrow D'$ are also CPOs by considering the pointwise order. If φ is a continuous function from D to D where \perp is the infimum (a.k.a. greatest lower bound) of D , we shall write $\text{FIX}(\varphi) = \lim_{n \rightarrow \infty} \varphi^n(\perp)$ for the smallest fix point of φ by the Kleene fixed point theorem.

6.1 Denotational synchronous semantics

We are interested in soundness guarantees about the termination of a single reaction, about the overall productivity (deadlock freedom), and on bounded memory execution. These motivations call for a semantics suitable to offer such guarantees “by construction” or “by design” as opposed to relying on best-effort static or dynamic analyses after the fact. This is the purpose of our denotational synchronous semantics. Note that these guarantees are typically associated with the ability to compile the program to a statically scheduled form with in-place updates [4], but such concerns are outside the scope of this paper.

Domain of sparsely indexed streams. One may relax the prefix ordering of the Kahn domain [10] with explicitly indexed streams, making all streams effectively infinite and randomly indexable. Given a natural number, one may reason about the value of an infinite stream at this specific index. This value may be an element of T , or it may be the special *unknown* value \perp capturing the fact that the actual value in T (if any) is yet to be determined in the process of a fixed point computation. Note that finite streams can be extended into infinite ones by concatenating \perp^ω ; one may thus model finite computations while considering infinite streams only. We also introduce an explicit *absent* signal \square and an *error* value \varkappa , and define $T_{\perp\square\varkappa} = T \cup \{\perp, \square, \varkappa\}$. \square effectively sparsifies streams and computations, modeling the absence of a value in T at a given index, while \varkappa is associated with incorrectly synchronized programs.¹⁴

One may define a simple ordering relation \sqsubseteq over $T_{\perp\square\varkappa}^\omega$ by lifting the flat CPO on individual values $\forall x \in T_{\perp\square\varkappa}, \perp \leq_S x \wedge x \leq_S \varkappa$ (a.k.a. the Scott domain) to infinite streams.

This lets us define the *sparsely indexed stream domain* $D = (T_{\perp\square\varkappa}^\omega, \sqsubseteq, \perp^\omega)$, a CPO whose join operator is denoted by \sqcup , with infimum \perp^ω and supremum \varkappa^ω .

Synchronous semantics on sparsely indexed streams. Figure 22 provides the denotations for the synchronous semantics. c is a constant and x is a stream variable; s, b, t and f are streams in D ; i and e are $\text{R}(\text{ML})^2$ expressions; n is a natural integer. The mark $\#$ distinguishes the syntactic construct from its interpretation as a stream transformer. Scalar operations are lifted to streams of values of the same type; this includes the usual unary and binary operations $-$, $+$, or , etc. The interpretation of a node overloads the environment V to hold the bidding a node name f its implementation; conversely, the node instantiation $f(e)$ looks up the environment V to apply the corresponding function to the interpretation of e . The projection $\pi_o(eq)$ retains the tuple of streams associated with the set of variables o from the tuple of equations eq . The fixed point operator FIX is applied to every system of equations in a node: considering a tuple of equations $x = e$, it is applied to the function $\varphi = \lambda d. \llbracket e \rrbracket(V, x := d)$ mapping the tuple of expressions e to its tuple of bindings. The environment V provides a value for all free variables in e except for recursive definitions depending on x ; the latter are provided with the tuple d of arguments of the λ -term. Fixed point iterations

¹⁴The notation is meant as a symbol for stream misalignment.

start with values in V already defined while x takes the successive iterates $\varphi^n(\perp)$; we establish the existence and unicity of a least fixed point $\text{FIX}(\varphi)$ by proving the Scott continuity of φ .¹⁵

The synchronous stream transformers are provided in Figure 23; u and v denote individual values in T . This semantics insists on providing a cycle-by-cycle interpretation amenable to bounded-memory execution. The subscripted semantics notations $\llbracket e \rrbracket_n$ and $\llbracket e \rrbracket_n^\#$ denote the scalar value of expression e at cycle n and the suffix stream value of expression e at all cycles n and above, respectively. The absence of a value is made explicit with \square . We lift built-in functions, unary and binary operations to $T_{\perp\square\#}$, yielding \perp when one of more operands are \perp , and yielding $\#$ when one or more operands are \square or $\#$. We decompose the fby operator to isolate an “initial state” $\text{fby}^\#$ from the “steady state” $\text{pre}^\#$ triggered by the recurrent equations in Figure 23.

- The right-hand side of equations in Figure 23 has been aligned on the \cdot operator, concatenating the value at index n with the rest of the stream. The value at index n is referred to as *instantaneous*. It is available for all computations at index n and is obtained by restricting both left- and right-hand-sides of the equations to this specific index. This instantaneous interpretation is a direct benefit of synchrony: it will form the basis for *current cycle* effects when deriving an operational semantics. In the conventional forward-reactive case, it also makes efficient in-place evaluation possible with bounded memory limited to the values live within a given cycle—all stream values at index n as well as the “state” s_{n-1} associated with fby operators. In the bidirectional case, multi-cycle dependences arise from the evaluation of s_{n+1} , challenging any in-place evaluation strategy; this is the topic of Section 9.
- Equations in Figure 23 provide the full definitions for infinite streams, concatenating the former with recurrences on a stream suffix such as s_{n+1} . The value s_n of the stream s at index n only depends on the left hand side of \cdot in the right-hand-side of the equations, while the stream suffix s_n depends on the whole right-hand-side beyond the \cdot operator.

All cases not matched in Figure 23 yield error values. Formally, this is enforced by defining any remaining non-matched case at a given index n to the $\#$ value at index n ; e.g.

$$\begin{aligned} \text{fby}_n^\#(\square, v, i, e) &\stackrel{\text{def}}{=} \# \cdot \text{fby}_{n+1}^\#(v, e) \\ \text{merge}_n^\#(\text{true} \cdot b_{n+1}, \square \cdot t_{n+1}, \square \cdot f_{n+1}) &\stackrel{\text{def}}{=} \# \cdot \text{merge}_{n+1}^\#(b_{n+1}, t_{n+1}, f_{n+1}) \end{aligned}$$

As a corollary, \square never occurs together with present values in pointwise/scalar computations. In particular, the merge operator enforces one of its branches to be present while the other is absent (or both are unknown).

We may now define a notion of synchrony: a program is *synchronous* if the $\#$ value never occurs in the least fixed point solution of the semantic equations in Figure 22; formally, let $T_{\perp\square} = T \cup \{\perp, \square\} = T_{\perp\square\#} \setminus \{\#\}$ represent non-erroneous, i.e. well-synchronized values, and consider a program P from d -tuples to d' -tuples of streams, i.e. $\llbracket P \rrbracket : V^d \rightarrow V^{d'}$,

$$P \text{ is synchronous} \iff \forall i \in (T_{\perp\square}^\omega)^d, \llbracket P \rrbracket(i) \in (T_{\perp\square}^\omega)^{d'}$$

This is actually a problem when operating on constants that (by definition) are always present, preventing then for occurring as arguments of the merging operator. The happens with recursive definitions involving post , which are also always present. To avoid these issues, we enforce that every constant and post expression is sampled with an “appropriate” when b expression before being operated upon; here “appropriate” calls for inferring the proper condition b to enforce synchrony, which is the purpose of Lustre’s clock calculus.

¹⁵As a corollary of continuity, the iterates $\varphi^n(\perp)$ form an ascending chain, i.e. successive iterates of φ propagate information monotonically, structurally from sub-expressions of e and from input arguments bound in V .

Finally, we define a bounded horizon version of the synchronous semantics of a program P denoted as $\llbracket P \rrbracket^h$, where the recursive evaluation of $\text{post}^\#$ stops at a given index h : the rules are identical except that of Figure 23 are used at index $n = h$. We will use this bounded horizon semantics to establish forward-progress and bounded memory execution guarantees.

$$\begin{aligned}
\llbracket c \rrbracket(V) &\stackrel{\text{def}}{=} c^\omega \\
\llbracket x \rrbracket(V) &\stackrel{\text{def}}{=} V(x) \\
\llbracket i \text{ fby } e \rrbracket(V) &\stackrel{\text{def}}{=} \text{fby}^\#(V(\llbracket i \rrbracket(V)), V(\llbracket e \rrbracket(V))) \\
\llbracket \text{post } e \rrbracket(V) &\stackrel{\text{def}}{=} \text{post}^\#(V(\llbracket e \rrbracket(V))) \\
\llbracket e \text{ when } b \rrbracket(V) &\stackrel{\text{def}}{=} \text{when}^\#(V(\llbracket e \rrbracket(V)), \llbracket b \rrbracket(V)) \\
\llbracket \text{merge } b \text{ t } f \rrbracket(V) &\stackrel{\text{def}}{=} \text{merge}^\#(V(\llbracket b \rrbracket(V)), \llbracket t \rrbracket(V), \llbracket f \rrbracket(V)) \\
\llbracket f(e) \rrbracket(V) &\stackrel{\text{def}}{=} V(f)(\llbracket e \rrbracket(V)) \\
\llbracket \text{node } f(i) \rightarrow (o) \text{ eq} \rrbracket(V) &\stackrel{\text{def}}{=} f := \lambda d. \llbracket \pi_o(\text{eq}) \rrbracket(V, i := d) \text{ where } i \subseteq \text{FV}(\text{eq}) \wedge o \subseteq \text{DV}(\text{eq}) \\
\llbracket x = e \rrbracket(V) &\stackrel{\text{def}}{=} \llbracket e \rrbracket(V, x := x_{\text{FIX}}) \text{ where } x_{\text{FIX}} = \text{FIX}(\lambda d. \llbracket e \rrbracket(V, x := d))
\end{aligned}$$

Fig. 22. Synchronous semantics.

$$\begin{array}{lll}
n \cdot \dots & n \cdot n+1 \cdot \dots & \\
\text{fby}_n^\#(i, s) &\stackrel{\text{def}}{=} \square \cdot \text{fby}_{n+1}^\#(i, s) & \text{if } i_n = s_n = \square \\
\text{fby}_n^\#(i, s) &\stackrel{\text{def}}{=} i_n \cdot \text{pre}_{n+1}^\#(s) & \text{if } i_n, s_n \in T \cup \{\perp\} \\
\text{pre}_n^\#(s) &\stackrel{\text{def}}{=} \square \cdot \text{pre}_{n+1}^\#(s) & \text{if } s_{n-1} = \square \\
\text{pre}_n^\#(s) &\stackrel{\text{def}}{=} s_{n-1} \cdot \text{pre}_{n+1}^\#(s) & \text{if } s_{n-1} \in T \cup \{\perp\} \\
\text{post}_n^\#(s) &\stackrel{\text{def}}{=} \square \cdot \text{post}_{n+1}^\#(s) & \text{if } s_{n+1} = \square \\
\text{post}_n^\#(s) &\stackrel{\text{def}}{=} s_{n+1} \cdot \text{post}_{n+1}^\#(s) & \text{if } s_{n+1} \in T \cup \{\perp\} \\
\text{when}_n^\#(\square \cdot s_{n+1}, \square \cdot b_{n+1}) &\stackrel{\text{def}}{=} \square \cdot \text{when}_{n+1}^\#(s_{n+1}, b_{n+1}) & \\
\text{when}_n^\#(s_n \cdot s_{n+1}, \text{true} \cdot b_{n+1}) &\stackrel{\text{def}}{=} s_n \cdot \text{when}_{n+1}^\#(s_{n+1}, b_{n+1}) & \\
\text{when}_n^\#(s_n \cdot s_{n+1}, \text{false} \cdot b_{n+1}) &\stackrel{\text{def}}{=} \square \cdot \text{when}_{n+1}^\#(s_{n+1}, b_{n+1}) & \\
\text{when}_n^\#(s_n \cdot s_{n+1}, \perp \cdot b_{n+1}) &\stackrel{\text{def}}{=} \perp \cdot \text{when}_{n+1}^\#(s_{n+1}, b_{n+1}) & \\
\text{merge}_n^\#(\square \cdot b_{n+1}, \square \cdot t_{n+1}, \square \cdot f_{n+1}) &\stackrel{\text{def}}{=} \square \cdot \text{merge}_{n+1}^\#(b_{n+1}, t_{n+1}, f_{n+1}) & \\
\text{merge}_n^\#(\text{true} \cdot b_{n+1}, t_0 \cdot t_{n+1}, \square \cdot f_{n+1}) &\stackrel{\text{def}}{=} t_0 \cdot \text{merge}_{n+1}^\#(b_{n+1}, t_{n+1}, f_{n+1}) & \\
\text{merge}_n^\#(\text{false} \cdot b_{n+1}, \square \cdot t_{n+1}, f_0 \cdot f_{n+1}) &\stackrel{\text{def}}{=} f_0 \cdot \text{merge}_{n+1}^\#(b_{n+1}, t_{n+1}, f_{n+1}) & \\
\text{merge}_n^\#(\perp \cdot b_{n+1}, t_n, f_n) &\stackrel{\text{def}}{=} \perp \cdot \text{merge}_{n+1}^\#(b_{n+1}, t_{n+1}, f_{n+1}) &
\end{array}$$

Fig. 23. Synchronous stream transformers.

Figure 24 illustrates the synchronous semantics on the `backfill` example. The fixed point computation is best decomposed over two dimensions: for a given stream index m , rows correspond to successive iterations of the fixed point computation defining values at index m , while the index may be stepped to $m + 1$ as soon as a fixed point is reached at index m .

6.2 Properties

In the following we refer to streams with no unknown (\perp) values as *fully defined*. One essential property of practically useful reactive programs is *productivity*, a notion associated with (1) the ability to compute the least fixed point of a system of equation as the result of a finite computation, and (2) the resulting streams being *fully defined*. In the traditional Lustre setting, productivity

| | | | |
|---|-----------------------|---|-------------------------------------|
| $\llbracket i \rrbracket_{0:3}$ | 0 1 2 | $\llbracket i \rrbracket_{0:6}$ | 0 1 2 3 4 5 |
| $\llbracket bp \rrbracket_{0:3}$ | false false true | $\llbracket bp \rrbracket_{0:6}$ | false false true true false true |
| $\llbracket i \text{ when } bp \rrbracket_{0:1}$ | \square | $\llbracket i \text{ when } bp \rrbracket_{0:4}$ | $\square \square 2 3$ |
| $\llbracket \text{post } o \rrbracket_{0:1}$ | \perp | $\llbracket \text{post } o \rrbracket_{0:4}$ | 2 2 $\perp \perp$ |
| $\llbracket (\text{post } o) \text{ when not } bp \rrbracket_{0:1}$ | \perp | $\llbracket (\text{post } o) \text{ when not } bp \rrbracket_{0:4}$ | 2 2 $\square \square$ |
| $\llbracket o \rrbracket_{0:1}$ | \perp | $\llbracket o \rrbracket_{0:4}$ | 2 2 2 3 |
| $\llbracket i \text{ when } bp \rrbracket_{0:2}$ | $\square \square$ | $\llbracket i \text{ when } bp \rrbracket_{0:5}$ | $\square \square 2 3 \square$ |
| $\llbracket \text{post } o \rrbracket_{0:2}$ | $\perp \perp$ | $\llbracket \text{post } o \rrbracket_{0:5}$ | 2 2 $\perp \perp \perp$ |
| $\llbracket (\text{post } o) \text{ when not } bp \rrbracket_{0:2}$ | $\perp \perp$ | $\llbracket (\text{post } o) \text{ when not } bp \rrbracket_{0:5}$ | 2 2 $\square \square \perp$ |
| $\llbracket o \rrbracket_{0:2}$ | $\perp \perp$ | $\llbracket o \rrbracket_{0:5}$ | 2 2 2 3 \perp |
| $\llbracket i \text{ when } bp \rrbracket_{0:3}$ | $\square \square 2$ | $\llbracket i \text{ when } bp \rrbracket_{0:6}$ | $\square \square 2 3 \square 5$ |
| $\llbracket \text{post } o \rrbracket_{0:3}$ | $\perp \perp \perp$ | $\llbracket \text{post } o \rrbracket_{0:6}$ | 2 2 $\perp \perp \perp \perp$ |
| $\llbracket (\text{post } o) \text{ when not } bp \rrbracket_{0:3}$ | $\perp \perp \square$ | $\llbracket (\text{post } o) \text{ when not } bp \rrbracket_{0:6}$ | 2 2 $\square \square \perp \square$ |
| $\llbracket o \rrbracket_{0:3}$ | $\perp \perp 2$ | $\llbracket o \rrbracket_{0:6}$ | 2 2 2 3 $\perp 5$ |
| $\llbracket \text{post } o \rrbracket_{0:3}$ | $\perp 2 \perp$ | $\llbracket \text{post } o \rrbracket_{0:6}$ | 2 2 3 $\perp 5 \perp$ |
| $\llbracket (\text{post } o) \text{ when not } bp \rrbracket_{0:3}$ | $\perp 2 \square$ | $\llbracket (\text{post } o) \text{ when not } bp \rrbracket_{0:6}$ | 2 2 $\square \square 5 \square$ |
| $\llbracket o \rrbracket_{0:3}$ | $\perp 2 2$ | $\llbracket o \rrbracket_{0:6}$ | 2 2 2 3 5 5 |
| $\llbracket \text{post } o \rrbracket_{0:3}$ | 2 2 \perp | $\llbracket \text{post } o \rrbracket_{0:6}$ | 2 2 3 5 5 \perp |
| $\llbracket (\text{post } o) \text{ when not } bp \rrbracket_{0:3}$ | 2 2 \square | | |
| $\llbracket o \rrbracket_{0:3}$ | 2 2 2 | $\llbracket \text{backfill}(i, bp) \rrbracket_{0:6}$ | 2 2 2 3 5 5 |

Fig. 24. Illustration of the synchronous semantics.

(i.e. deadlock-freedom) stems from a conservative and modular interpretation of *causality*: no *instantaneous cycle* may exist when stepping through the computation of fixed points, i.e. when any recurrence on a given stream variable at index n can be traced back to the previous index $n - 1$ through the use of the fby operator. The presence of a fby operator on all cycles in the dataflow graph defined by every node of the program is a sufficient condition for the absence of instantaneous cycles. Weaker conditions and static analyses exist but none of these involve dependences on future indices, deemed non-reactive: the interested reader may refer to [33] for a survey and for extensions improving modularity and efficient compilation.

This pushes us to research an extended notion of productivity compatible with bidirectional reactions. Intuitively, lookahead dependences may be allowed as long as they are eventually broken, i.e. they do not form an infinite recurrence into the future. The problem is that allowing dependences into the future, even bounded ones, automatically gives rise to instantaneous cycles by composition with conventional dependences on the past. Reasoning on dependences only is unlikely to lead to solution. This intuition leads us to define the extended notion of *bidirectional causality*.

Definition 1 (Bidirectional causality). A program P is bidirectionally causal if and only if the following condition hold on the synchronous interpretation $\llbracket P \rrbracket : V^d \rightarrow V^{d'}$:

$$\forall m \in \mathbb{N}, \exists z \in \mathbb{N}, \llbracket P \rrbracket_m = \llbracket P \rrbracket_{0:m}^{m+z}$$

We are also interested in a stronger notion of *bidirectional bounded reactivity*, analogous to the bounded memory, instantaneous termination property of synchronous reactive languages [4]. To this end, we combine bidirectional causality with a notion of termination suitable to stream functions. Let $\varphi : D \rightarrow D$ be a Scott-continuous function; φ *terminates up to index* $n \in \mathbb{N}$ if and only if $\exists m \in \mathbb{N}, \text{FIX}(\varphi)_{0:n} = \varphi^m(\perp^\omega)_{0:n}$; and φ *terminates* if and only if it terminates up to index n for all $n \in \mathbb{N}$. A synchronous program P terminates (resp. terminates up to index n) if and only if all fixed point computations terminate (resp. terminate up to index n).

Definition 2 (Bidirectional bounded reactivity). A program P is bidirectionally bounded reactive if and only if the two following conditions hold on the synchronous interpretation $\llbracket P \rrbracket : V^d \rightarrow V^{d'}$: (i) $\llbracket P \rrbracket$ terminates, and (ii) $\exists z \in \mathbb{N}, \forall m \in \mathbb{N}, \llbracket P \rrbracket_m = \llbracket P \rrbracket_{0:m}^{m+z}$.

Let us now state two theorems that are essential to making the interpretation of the semantics effective and practical.

THEOREM 3 (FINITE AND BOUNDED MEMORY INTERPRETATION).

- *Finite memory (possibly unbounded): considering a bidirectionally causal program P and the look-ahead offset z guaranteed by bidirectional causality at a given index m , only needs to store windows of stream values of size $z + 1$ to effectively interpret the semantics.*
- *Bounded memory: considering a bounded reactive program P and the look-ahead offset \bar{z} guaranteed by bounded reactivity for all indices, only needs to store windows of stream values of size $\bar{z} + 1$ to effectively interpret the semantics.*

The proof stems from the observation that incremental left-to-right computation is possible for fby operators, holding the value at the previous index only.

In the following, *causal* is synonymous with *bidirectionally causal*, and *bounded reactive* is synonymous with *bidirectionally bounded reactive*, unless stated otherwise. It is immediate that a bounded reactive program is causal.

THEOREM 4 (EVENTUAL PROGRESS). *For a causal program P , the synchronous interpretation up to a certain index coincides with the synchronous semantics with a finite look-ahead horizon applied to a bounded future of the program's inputs. For a bounded reactive program, a global bound exists for both look-ahead horizon and inputs.*

Formally, given program P with $\llbracket P \rrbracket : V^d \rightarrow V^{d'}$, if P is causal then

$$\forall i \in D^d, \forall m \in \mathbb{N}, \exists z \in \mathbb{N}, \llbracket P(i) \rrbracket_{0:m} = \llbracket P(i_{0:m+z} \cdot \perp^\omega) \rrbracket_{0:m}^{\bar{m}+z}$$

and if P is bounded reactive then

$$\forall i \in D^d, \exists \bar{z} \in \mathbb{N}, \forall m \in \mathbb{N}, \llbracket P(i) \rrbracket_{0:m} = \llbracket P(i_{0:m+\bar{z}} \cdot \perp^\omega) \rrbracket_{0:m}^{\bar{m}+\bar{z}}$$

The proof is inductive on the structure of semantic equations.

7 FROM DENOTATIONAL SEMANTICS TO OPERATIONAL SIMULATION

The semantics of the previous section has the advantage introducing backwards recurrences in a conventional Lustre-like setting with minimal changes to the semantic domain. To this end, it makes the simplifying assumption that programs are normalized, i.e. that constants are present at every cycle, that the fby operator of forward recurrences is activated according to its first input, that the post operator of backward recurrences is present at every cycle, and that the execution environment provides the status of every input variable at every cycle. From these primary sources of control information, the status of every other variable can be incrementally constructed at every cycle by traversing every program equation to compute the status of the equation outputs based on that of the equation inputs. In this process, every equation is traversed exactly once per cycle.

Formalizations of classical reactive languages rely on similar normalization. In most cases, it is assumed that the activation of every variable and equation is described using a predicate, called a *clock*, which can be provided by the programmer or computed by means of static analysis—the clock calculus. The main downside of this classical approach is that the clocks need to be available and the normalization to take place even *before one is able to define the program semantics*. Indeed, the natural programming style implemented in classical compilers of dataflow reactive languages,¹⁶ does not require the definition of clocks nor does it expose the user with the need to normalize programs. An additional drawback is that such normalization may actually be wasteful

¹⁶Such as [Lustre](#) or [Heptagon](#).

of memory resources, by forcing the allocation of memory for all inputs of a node at every cycle irrespectively of their downstream sampling through normalization.

In this section, we solve this conundrum by providing a semantics for programs that are not heavily normalized. In doing this, we actually solve an old problem of reactive programming: introducing a semantic reference to establish the correctness of the static analysis algorithms producing clock annotations.

The semantics defined in this section has three more advantages:

- It provides support for the efficient simulation of bidirectionally causal programs (see Definition 1). This is actually the main reason to present this new normalization-free semantics in the context of bidirectional reactivity: bidirectionality within a given cycle is exactly what is needed to eliminate the need for normalization in the first place; while more general bidirectional reactivity amounts to extending the semantics and simulation to multi-cycle windows.
- Specifically, one may simulate finite executions of bidirectionally bounded reactive programs (see Definition 2) in a finite number of steps and with a finite memory footprint.
- From an system and input/output standpoint, the semantics directly captures the behavior of programs that drive their inputs, possibly driven by their outputs, an important demand-driven pattern in both embedded and datacenter applications.

We formally prove that the simulator provided in Section 7.7 implements the semantics of this section. We also conjecture that our semantics is a strict generalization of that of Section 6 and of that of classical reactive languages such as Lustre. Proving these results is work in progress.

7.1 Domain definition

Single-variable domain. Given a variable v , we denote with T_v the set of values of its data type. We also overload T_v to stand the data type itself. The semantics may also assign to a variable, during a cycle, another 4 values with specific meaning:

- \perp represents the absence of any information on the status of the variable (at the current point in execution).
- \square represents the absent status of a variable during a cycle. It is needed to ensure by semantic means, without requiring preliminary static analysis, that a variable value produced in one execution cycle is always consumed within the same cycle.
- $?$ is the *synchronization status* of a variable for which we determined that it is present, but have not yet computed the value. It allows the on-demand triggering of the computation of constants and of recurrence outputs. It also allows triggering input acquisition actions, as opposed to always assuming that the status is provided by the environment. This value is a new addition compared to the sparsely indexed stream domain of the previous section.
- \varkappa signifies that the variable is involved in a *synchronization error* in the current cycle – its present/absent status is inconsistent with that of other variables. It is required to ensure confluence when control can propagate backwards.¹⁷

The domain of v is therefore $D_v = T_v \cup \{\perp, \square, ?, \varkappa\}$, endowed with the partial order \leq defined by: \perp and e respectively are the infimum and supremum of D_v , and $?$ is lower than the values $x \in T_v$. We denote the (always defined) supremum operator of D_v with \vee .

¹⁷Which means that, for the same input, the synchronization error can be first detected, depending on the equation evaluation order, in different equations. The use of \varkappa allows propagating the error towards a unique end-state.

We define $nv : D_v \rightarrow \{\perp, \square, ?, \varkappa\}$ as the function that strips the value of a variable, preserving only its synchronization status:

$$nv(x) = \begin{cases} ? & \text{if } x \in T_v \\ x & \text{otherwise} \end{cases}$$

Multi-variable domain. Given a set of variables V , we represent their semantic status at a given execution step with elements of $D_V = \prod_{v \in V} D_v$. Given $v \in V$ and $c \in D_V$, we denote with $c[v]$ the value of v in c . We overload the \perp notation to (also) denote the element of D_V with $c[v] = \perp$ for all $v \in V$. We similarly overload in D_V the notations \square , $?$, and \varkappa . D_V is naturally endowed with the product domain structure, where \perp and \varkappa are respectively least and greatest elements, \leq is the partial order over D_V , and \vee is the (completely-defined) supremum operator.

If $s \in D_V$, $v_1, \dots, v_n \in V$ and $x_i \in D_{v_i}$, then $s[v_1, \dots, v_n \leftarrow x_1, \dots, x_n]$ is the element of D_V that equals x_i on v_i and $s[v]$ for all $v \notin V$. We also denote $s[v_1, \dots, v_n \leftarrow x_1, \dots, x_n] = s \vee s[v_1, \dots, v_n \leftarrow x_1, \dots, x_n]$ the update operator that never loses information.

We denote with $cerr$ the endomorphism $cerr : D_V \rightarrow D_V$ that returns \varkappa if at least one component of its input is \varkappa , and returns its input otherwise.

Domain of streams. If V is the set of variables of a program or set of equations p , then we will use streams of D_V^∞ to represent both (finite or infinite) execution traces and the semantic states used in the construction of the traces. Recall that if $s \in D_V^\infty$, then $s_n \in D_V$ is the element of s of index $n \geq 0$. In a trace/semantic state $s \in D_V^\infty$, we will say that execution has reached cycle $n \geq 0$ if there exists $m \geq n$ with $s_m \neq \perp$. For $s \in D_V^\infty$ we denote with $len(s) \in \mathbb{N} \cup \{\infty\}$ its length.

The partial order \leq on D_V induces a partial order \leq on D_V^∞ as follows: $s^1 \leq s^2$ if, by definition, $len(s^1) \leq len(s^2)$ and $s_i^1 \leq s_i^2$ for all $0 \leq i < len(s^1)$. Endowed with \leq , D_V^∞ is a complete lattice with infimum ϵ and supremum \varkappa^∞ . The supremum operator is denoted \vee .

7.2 Semantic functions

Given a set X , we denote with id_X the identity function over X . For instance, $id_{D_V^\infty}$ is the identity function over D_V^∞ .

Given a domain (D, \leq) , and another set A , we also denote with \leq the pointwise partial order induced on D^A (the set of functions from A to D) by the partial order of D . We respectively denote with \vee and \wedge the supremum and infimum operators induced by \leq .

Assume V is the set of variables used by a program or set of equations p . We will define the semantics of p as a function $\llbracket p \rrbracket \in (D_V^\infty)^{D_V^\infty}$. We must always have:

- $id_{D_V^\infty} \leq \llbracket p \rrbracket$, so that applying the semantics does not lose information present in the input.
- $\llbracket p \rrbracket$ is idempotent¹⁸ so that all the values that can be computed by p given the input s will be present in $\llbracket p \rrbracket(s)$, and applying $\llbracket p \rrbracket$ a second time builds no new values.

7.2.1 Fixpoint iteration function. Recall that given a directed-complete partial order (CPO) (D, \leq) , a continuous function $f : D \rightarrow D$, and $x \in D$, the least fixed point of f greater than x , denoted $\text{FIX}(f, x)$, exists and can be computed as:

$$\text{FIX}(f, x) = \lim_{n \rightarrow \infty} f^n(x)$$

To ensure that $\llbracket p \rrbracket(s)$ can be incrementally computed for every s , we will require that it is defined as the least fixed point greater than s of another function $\{\!\{ p \}\!\} \in (D_V^\infty)^{D_V^\infty}$ called the *fixpoint*

¹⁸Or, equivalently, $\llbracket p \rrbracket(s)$ is a fixed point of $\llbracket p \rrbracket$ for all s .

iteration function of p . We require that $\llbracket p \rrbracket$ is continuous and that $\llbracket p \rrbracket \geq id_{D_V^\infty}$. Then, for all s :

$$\llbracket p \rrbracket(s) = \text{FIX}(\llbracket p \rrbracket, s) = \lim_{n \rightarrow \infty} \llbracket p \rrbracket^n(s)$$

7.2.2 Semantics of stateless equations/programs. If p involves no recurrences (no fby or post equation), then its semantics can be described using a *stateless iteration function* $\llbracket p \rrbracket_\# : D_V \rightarrow D_V$ that must be continuous and have the property $\llbracket p \rrbracket_\# \geq id_{D_V}$. The iteration function $\llbracket p \rrbracket$ is then defined by $\llbracket p \rrbracket(s)_n = \llbracket p \rrbracket_\#(s_n)$ for all $0 \leq n < \text{len}(s)$. This function is greater than $id_{D_V^\infty}$ and is continuous, allowing the definition of $\llbracket p \rrbracket$.

7.3 Composing semantic functions

The semantics of R(ML)^2 primitives will be provided directly in Section 7.6. The semantics of program fragments including multiple statements/equations, and that of nodes, is provided in this section. It is obtained by composition from that of the primitives.

7.3.1 Notations. Given two sets of variables $V \subseteq W$ and $x \in D_W$, we denote with $x \upharpoonright_V$ the restriction of x to V , which is an element of D_V . In the other sense, given $y \in D_V$, we denote with $y \upharpoonright_W$ its extension to W , which equals y over V and \perp elsewhere. The restriction and extension operators are both extended pointwise to (respectively) elements of D_W^∞ and D_V^∞ . Thus, given $s \in D_W^\infty$, $s \upharpoonright_V$ is an element of D_V^∞ . Similarly, if $s \in D_V^\infty$ then $s \upharpoonright_W$ belongs to D_W^∞ .

The extension operator is generalized to allow the embedding of $(D_V^\infty)^{D_V^\infty}$ into $(D_W^\infty)^{D_W^\infty}$. If $f : D_V^\infty \rightarrow D_V^\infty$, then $f \upharpoonright_W : D_W^\infty \rightarrow D_W^\infty$ is defined by $f \upharpoonright_W(s) = f(s \upharpoonright_V) \upharpoonright_W$.

7.3.2 Semantics of composition. We can now define the semantics of composition. The semantics of each equation or set of equations is defined over the variables it involves. For instance, the semantics of “ $x = \text{merge } c \ y \ z$ ” is defined over domain $D_{\{x,c,y,z\}}$. When building the semantics of a code fragment (or node) containing an equation, we first use the extension operator to lift the semantics of the individual equation onto the variable set of the code fragment or node, which allows us to define composition by means of upper bound:

$$\llbracket eq_1; \dots; eq_n \rrbracket(s) = \text{FIX}(\llbracket eq_1; \dots; eq_n \rrbracket, s) \quad \llbracket eq_1; \dots; eq_n \rrbracket = \bigvee_{i=1}^n (\llbracket eq_i \rrbracket \upharpoonright_W)$$

where W is the set of variables taken as input or produced as output by equations eq_1, \dots, eq_n . Note that by applying the Knaster-Tarski theorem we also obtain $\llbracket eq_1; \dots; eq_n \rrbracket(s) = \text{FIX}(\bigvee_{i=1}^n (\llbracket eq_i \rrbracket \upharpoonright_W), s)$.

7.3.3 Semantics of nodes. Consider an R(ML)^2 node n . We assume that its equations are eq_1, \dots, eq_n , and that I, O , and L are respectively the (mutually disjoint) sets of input variables, output variables, and local variables that are neither inputs nor outputs. Then, $\llbracket eq_1; \dots; eq_n \rrbracket : D_{I \cup O \cup L}^\infty \rightarrow D_{I \cup O \cup L}^\infty$ and we define the semantics of node n with functions $\llbracket n \rrbracket, \llbracket n \rrbracket_\# : D_{I \cup O}^\infty \rightarrow D_{I \cup O}^\infty$:

$$\llbracket n \rrbracket(s) = \text{FIX}(\llbracket n \rrbracket, s) \quad \llbracket n \rrbracket_\#(s) = \llbracket eq_1; \dots; eq_n \rrbracket(s \upharpoonright_{I \cup O \cup L}) \upharpoonright_{I \cup O}$$

7.4 The “fby back” primitive

The semantics of this section will cover not only infinite executions/traces of R(ML)^2 programs,¹⁹ but also *finite* ones. In this case, backwards recurrences fully mirror, in time, the behavior of forward recurrences, including their initialization, which for backwards recurrences will happen in the *last* execution cycle of the recurrence.

As the post primitive does not allow providing an initialization value (instead producing an output that is absent in the last execution cycle), and for symmetry w.r.t. the forward recurrences, we replace post with a new primitive whose general form is “ $x = y \text{ fby back } z$,” where y and z are expressions of the same type and x is a stream variable. This primitive requires that x, y , and

¹⁹Like the semantics in Section 6 does.

z are present in the same cycles. We will denote with $t_0 < t_1 < t_2 < \dots$ the indices where x, y and z are all present, and let $l \in \mathbb{N} \cup \{\infty\}$ be the number of such cycles. Then, the recurrence relation defined by the primitive is:

$$x_{t_n} = \begin{cases} y_{t_n} & \text{if } n = l - 1 \quad \text{-- initialization} \\ z_{t_{n+1}} & \text{otherwise} \quad \text{-- recurrence backwards in time} \end{cases}$$

Notice that the initialization case of the definition does not apply for infinite executions/traces.

The post operator can always be expressed based on `fbym` back:

$$x = \text{post } y; \quad \text{becomes} \quad x = (y \text{ fbym back } y) \text{ when } (\text{false fbym back true});$$

The converse is not true, as it would require a means to identify the last execution cycle of the equation, which `post` and the other primitives do not provide.

7.5 Normalization

In the general form of the recurrence primitives `fbym` and “`fbym` back”, the initialization can be performed using any expression. However, this general form can always be reduced to the case where the initializer is a constant:

$$\begin{aligned} y = i \text{ fbym } x; & \quad \text{becomes} \quad x = \text{if } (\text{true fbym } \text{false}) \text{ then } i \text{ else } (k \text{ fbym } x); \\ y = i \text{ fbym back } x; & \quad \text{becomes} \quad x = \text{if } (\text{true fbym back } \text{false}) \text{ then } i \text{ else } (k \text{ fbym back } x); \end{aligned}$$

where the constant k is any value of the type of x .

This transformation improves the orthogonality of our set of primitives, by letting conditional control *inside the cycle* be always handled by stateless conditional execution primitives, whereas the recurrence primitives only focus on the recurrence itself, k being the recurrence initializer, which corresponds to the initial value of the accumulator in a `fold_left` or `fold_right` iterator, which is not computed during iteration.

The definition of the semantics of the primitives will assume that this transformation has been performed, and that furthermore complex expressions have been decomposed, so that:

- Every equation contains a single dataflow operator.
- No variable appears twice in an equation.
- Constants are only used as:
 - Right-hand side of *constant equations* with a single variable as output.
 - State initializers in `fbym` and “`fbym` back” equations.

7.6 Semantics of primitives

Definition 5 (Function application semantics). Consider a function $f : \prod_{i=1}^n T_{x_i} \rightarrow \prod_{j=1}^m T_{y_j}$, with $n \geq 1$ and the equation $eq = “(y_1, \dots, y_m) = f(x_1, \dots, x_n)”$. Its stateless iterator function is $\{\{eq\}\}_\# : \prod_{i=1}^n D_{x_i} \times \prod_{j=1}^m D_{y_j} \rightarrow \prod_{i=1}^n D_{x_i} \times \prod_{j=1}^m D_{y_j}$ defined for any $v_i \in D_{x_i}$ and $w_j \in D_{y_j}$ by:

$$\{\{eq\}\}_\#(v_1, \dots, v_n, w_1, \dots, w_m) = \begin{cases} cerr((v_1, \dots, v_n, u_1 \vee w_1, \dots, u_m \vee w_m)) & \text{if } s = ? \text{ and } \forall i : v_i \in T_{x_i} \\ (s \vee v_1, \dots, s \vee v_n, s \vee w_1, \dots, s \vee w_m) & \text{otherwise} \end{cases}$$

where $s = nv(v_1) \vee \dots \vee nv(v_n) \vee nv(w_1) \vee \dots \vee nv(w_m)$, and (when $\forall i : v_i \in T_{x_i}$) $f(v_1, \dots, v_n) = (u_1, \dots, u_m)$.

Notice that computation values are only propagated from inputs to the outputs (in the first case of the definition, through the application of f), whereas synchronization information will traverse the equation in both senses (the second case of the definition, and the application of $cerr$).

Whenever a synchronization error is identified on one of the variables, it is propagated to all variables, either explicitly through the application of $cerr$, or implicitly through the computation of s .

The semantics of constant equations (with a constant right-hand side) can be provided as a sub-case of function application. However, for clarity, we prefer defining it separately.

Definition 6 (Constant semantics). Consider equation $eq = "y = k"$, where $k \in T_y$. Its stateless iterator function is $\{\{eq\}\}_\# : D_y \rightarrow D_y$ defined by:

$$\{\{eq\}\}_\#(w) = \begin{cases} k \vee w & \text{if } w \geq ? \\ w & \text{otherwise} \end{cases}$$

Notice that the value of the constant is produced when the status of the output variable is "?". The same demand-driven protocol also allows inputs to be read on-demand.

Definition 7 (Semantics of when). Consider the equation $eq = "o = y \text{ when } c"$. Its stateless iterator function is $\{\{eq\}\}_\# : D_{\{o,y,c\}} \rightarrow D_{\{o,y,c\}}$ defined as:

$$\{\{eq\}\}_\# = cerr \circ when^o \circ when^y \circ when^c$$

where:

$$when^c(s) = \begin{cases} s \vee \square & \text{if } s[c] = \square \\ s[y \leftarrow ?] & \text{if } s[c] = ? \\ s[o, y \leftarrow s[y] \vee ?, ?] & \text{if } s[c] = \text{true} \\ s[o, y \leftarrow \square, ?] & \text{if } s[c] = \text{false} \\ s & \text{otherwise} \end{cases}$$

$$when^y(s) = \begin{cases} s \vee \square & \text{if } s[y] = \square \\ s[c \leftarrow ?] & \text{if } s[y] \in T_y \cup \{?\} \\ s & \text{otherwise} \end{cases} \quad when^o(s) = \begin{cases} s \vee ? & \text{if } s[o] \in T_y \cup \{?\} \\ s & \text{otherwise} \end{cases}$$

Notice how the definition of $\{\{eq\}\}_\#$ has been decomposed into multiple semantic functions, each one representing the semantic consequences of setting a particular input or output variable to a particular value. For instance, in the definition of function $when^c$, having $c = \text{true}$ in the input state results in y and o being set to present, and if y already has a value of T_y , it propagates towards o . This propagation action can result in o taking the value or, if o already has a value of \square or μ , its value becomes μ .

To obtain the semantics of the primitive, the per-variable functions are composed, and $cerr$ is applied so that an error detected in one variable is propagated to all other variables.

Definition 8 (Semantics of merge). Consider the equation $eq = "o = \text{merge } c \ y \ z"$. Its stateless iteration function is $\{\{eq\}\}_\# : D_{\{o,c,y,z\}} \rightarrow D_{\{o,c,y,z\}}$ defined as:

$$\{\{eq\}\}_\# = cerr \circ merce^o \circ merge^{y,z} \circ merge^c$$

where:

$$merge^c(s) = \begin{cases} s \vee \square & \text{if } s[c] = \square \\ s[o \leftarrow ?] & \text{if } s[c] = ? \\ s[o, y, z \leftarrow s[y] \vee ?, ?, \square] & \text{if } s[c] = \text{true} \\ s[o, y, z \leftarrow s[z] \vee ?, \square, ?] & \text{if } s[c] = \text{false} \\ s & \text{otherwise} \end{cases}$$

$$merge^{y,z}(s) = \begin{cases} s[c \leftarrow ?] & \text{if } \{s[y], s[z]\} \cap (T_y \cup \{?\}) \neq \emptyset \\ s & \text{otherwise} \end{cases} \quad merge^o(s) = \begin{cases} s \vee \square & \text{if } s[o] = \square \\ s[c \leftarrow ?] & \text{if } s[o] \in T_y \cup \{?\} \\ s & \text{otherwise} \end{cases}$$

Notice that the semantics of merge requires that the Boolean condition value is received before the data input can be propagated towards the output.

Definition 9 (Semantics of fby). Consider the equation $eq = "o = k \text{ fby } y"$, where $k \in T_o = T_y$. Its iteration function is $\{\{eq\}\}_\# : D_{\{o,y\}}^\infty \rightarrow D_{\{o,y\}}^\infty$ defined by $\{\{eq\}\}_\# = fby^{val} \circ fby^{sync}$, where:

$$\begin{aligned} fby^{sync}(s)_n &= s_n[o, y \leftarrow nv(s_n[y]), nv(s_n[o])] \text{—synchronization, involving no recurrence} \\ fby^{val}(s)_n &= s_n[o \leftarrow curr^{fw}(s, n)] \text{—forward recurrence} \end{aligned}$$

with:

$$curr^{fw}(s, n) = \begin{cases} s_n[o] & \text{if } s_n[o] \notin T_o \cup \{?\} \\ k \vee s_n[o] & \text{if } s_n[o] \in T_o \cup \{?\} \text{ and } \forall k < n : s_k[o] = \square \\ s_m[y] \vee s_n[o] & \text{if } s_n[o], s_m[y] \in T_o \cup \{?\} \text{ and } \forall m < k < n : s_k[o] = \square \\ \perp & \text{otherwise} \end{cases}$$

Definition 10 (Semantics of “fby back”). Consider the equation $eq = “o = k \text{ fby back } y”$, where $k \in T_o = T_y$. Its iteration function is $\llbracket eq \rrbracket : D_{\{o,y\}}^\infty \rightarrow D_{\{o,y\}}^\infty$ defined by $\llbracket eq \rrbracket = fbyback^{val} \circ fby^{sync}$, where:

$$fbyback^{val}(s)_n = s_n[o \leftarrow curr^{bw}(s, n)] \text{—backward recurrence}$$

with:

$$curr^{bw}(s, n) = \begin{cases} s_n[o] & \text{if } s_n[o] \notin T_o \cup \{?\} \\ k \vee s_n[o] & \text{if } s_n[o] \in T_o \cup \{?\} \text{ and } \forall n < k \leq len(s) - 1 : s_k[o] = \square \\ s_m[y] \vee s_n[o] & \text{if } s_n[o], s_m[y] \in T_o \cup \{?\} \text{ and } \forall n < k < m : s_k[o] = \square \\ \perp & \text{otherwise} \end{cases}$$

Note that, if the equation $eq = “o = x \text{ fby back } y”$ is activated (inputs or outputs present) only in a finite number of cycles of an infinite trace, then in the last activation cycle its output remains at value \perp . Indeed, determining that the equation is inactive in all subsequent cycles cannot be done algorithmically over an infinite number of cycles. Such \perp values can be filtered out using when equations.

The last primitive we need to consider is node instantiation.

Definition 11 (Node instantiation semantics). Consider equation $eq = “(y_1, \dots, y_m) = n(x_1, \dots, x_n)”$, where n is a node (not a function). Then $\llbracket eq \rrbracket = \llbracket n \rrbracket$ (as defined in Section 7.3.3).

Note that this definition makes the semantics of instantiation equations depend on that of the instantiated nodes, whereas Section 7.3.3 makes the semantics of nodes depend on that of equations. The correction of this recursive definition is guaranteed here by the rule governing module instantiation: no node can directly or indirectly instantiate itself. This rule implies that the nodes instantiated (directly or indirectly) by a node n form a *finite* tree with n as root, each node having as children the node instances it has as equations.

7.7 Fixed point computation

Consider a node p . Assume its set of variables is W and its set of inputs is $I \subseteq W$. Consider a valuation of its inputs $i \in D_I^\infty$. Then, the semantics of p for input i is $\llbracket p \rrbracket(\bar{i})$, where $\bar{i} = i \upharpoonright_V$.

Under the definition of Section 7.7, $\llbracket p \rrbracket(\bar{i})$ is obtained by cyclically applying $\llbracket p \rrbracket$ to compute $\llbracket p \rrbracket^n(\bar{i})$. But one application of $\llbracket p \rrbracket$ is already an infinite process, as it may build values at an infinity of cycles of \bar{i} . Therefore, this organization of the fixpoint definition cannot be used as a pattern for operational implementation.

To allow implementation, we need to choose a different order for the computation of variable values. We choose to always compute new values at the lowest cycle where it is possible. This is also the implementation strategy of classical reactive languages, where values are built cycle by cycle. In these languages, when all values of cycle n are constructed, execution can move to cycle $n + 1$. But when recurrences backwards in time are allowed, this is not always possible. For instance, in the training of RNNs by means of BPTT (as in Figure 19) a full recurrent input sample is processed by the forward path before back-propagation can start. For this reason, multiple cycles can have values under construction.

```

1  s :=  $\perp[0 \leftarrow \bar{i}_0]$  (* initial state *)
2  lst := 0 (* initial lst *)
3  while(true) (* iteration, a priori infinite *)
4      n = 0 (* start computation of the first cycle with actions to execute *)
5      while todon(s) =  $\emptyset$  do (* in cycle n no equation can be executed *)
6          if n < len( $\bar{i}$ ) then n := n + 1 (* try the next cycle *) else return
7          if n > lst then (* if needed open cycles to simulation *)
8              lst := n (* in cycle n I can execute no equation, try the next cycle *)
8              s := s[lst  $\leftarrow \bar{i}_n$ ] (* include the inputs of the newly open cycle in the state *)
10         end
11     end
12     eq := choose(todon) (* choose equation to execute from non-void set *)
13     act :=  $\{\{eq\}\}(s \upharpoonright_{V_{eq}})_n$  (* compute update action *)
14     s := s[n  $\leftarrow s_n \vee (act \upharpoonright_W)$ ] (* change simulation state *)
15 end

```

Fig. 25. Fixed point iteration algorithm. It only terminates (in line 6) if the input is finite.

At every step of the fixpoint computation, we denote with lst the last cycle under construction. Whenever no new values can be built between cycles 0 and lst without using information of cycles of index $n > lst$, we increment lst .

We denote with $todo_n(s)$ the set of all equations that, in fixed point simulation state s , would change the value of a variable in cycle n . It is formally defined as:

$$todo_n(s) = \{eq \mid \{\{eq\}\}(s \upharpoonright_{V_{eq}})_n > (s \upharpoonright_{V_{eq}})_n\}$$

where V_{eq} denotes the set of variables of equation eq . Given $s \in D_V^\infty$ and $x \in D_V$, we denote $s[n \leftarrow x]$ the element of D_V^∞ that equals x in cycle n and s_m in every cycle $m \neq n$.

With these notations, the computation of the fixed point is realized using the algorithm in Figure 25. To simplify presentation, the algorithm assumes that node p instantiates no other nodes. To cover node instantiation, the simulation state must be generalized to include not only the valuation of variables, but also the simulation state of every instantiated node, hierarchically under the form of a finite tree, and the state computation and update (lines 13-14) must be updated accordingly.

Notice that the fixed point iteration algorithm considers inputs incrementally. This means that state representations used during iteration are always finite, allowing implementation.

7.7.1 Equivalence with denotational semantics. Along a given execution of the algorithm in Figure 25, we shall denote with $\{\{p\}\}_m(\bar{i})$ the value of s at the end of the m^{th} iteration of the main iteration loop. Then, the fixed point iteration algorithm and the semantics are related by the following result:

THEOREM 12. *Under the previous definitions, for every $n \geq 0$, if $\llbracket p \rrbracket(\bar{i})_n \neq i_n$, then there exists $m \geq 0$ such that $\llbracket p \rrbracket(\bar{i})_{0:n} = \{\{p\}\}_m(\bar{i})_{0:n}$*

Assuming in the theorem hypothesis that $\llbracket p \rrbracket(\bar{i})_n \neq i_n$ amounts to assuming that the semantics assigns at least one new value to at least one variable in a cycle of index $k \geq n$. Without this assumption, the fixed point iteration algorithm would remain stuck in the internal while loop that computes n (lines 7-13), as the semantics has no values to compute.

PROOF OF THEOREM 12. We first prove that $\{\{p\}\}_m \llbracket p \rrbracket(\bar{i}) \leq \llbracket p \rrbracket(\bar{i})$ for all m . This is done by induction. For $m = 0$, $\{\{p\}\}_0(\bar{i}) = \perp[0 \leftarrow \bar{i}_0] \leq \bar{i} \leq \llbracket p \rrbracket(\bar{i})$. For the induction step, we denote with

f^{m+1} the function used to construct $\{\!\{p}\!\}_{m+1}(\bar{t})$ from $\{\!\{p}\!\}_m(\bar{t})$, as defined in lines 15-16 of the algorithm. As f^{m+1} is obtained from $\{eq^{m+1}\}$ for some equation eq^{m+1} , it is by construction smaller than $\{\!\{p}\!\}$ (itself the union of $\{eq\}$ for all equations eq). Then: $\{\!\{p}\!\}_{m+1}(\bar{t}) = f^{m+1}(\{\!\{p}\!\}_m(\bar{t})) \leq \{\!\{p}\!\}(\{\!\{p}\!\}_m(\bar{t})) \leq \llbracket p \rrbracket(\{\!\{p}\!\}_m(\bar{t})) \leq \llbracket p \rrbracket(\llbracket p \rrbracket(\bar{t}))$. And since $\llbracket p \rrbracket(\bar{t})$ is a fixed point of $\llbracket p \rrbracket$, the induction step is complete.

From $\{\!\{p}\!\}_m(\bar{t}) \leq \llbracket p \rrbracket(\bar{t})$ for all m we can then deduce $\{\!\{p}\!\}_m(\bar{t})_{0:n} \leq \llbracket p \rrbracket(\bar{t})_{0:n}$ for all m, n .

To complete the proof of our theorem, we need to find a value of m for which the reverse inequality holds. Since $\llbracket p \rrbracket(\bar{t}) = \lim_{r \rightarrow \infty} \{\!\{p}\!\}^r(\bar{t})$, we obtain that there exists r_0 such that $\llbracket p \rrbracket(\bar{t})_{0:n} = \{\!\{p}\!\}^r(\bar{t})_{0:n}$ for all $r \geq r_0$. Then, proving our theorem can be done by proving that for every $r \geq 0$ and $n \geq -1$ there exists $m_{r,n}$ such that

$$\{\!\{p}\!\}_{m_{r,n}}(\bar{t})_{0:n} \geq \{\!\{p}\!\}^r(\bar{t})_{0:n} \quad (1)$$

We prove this by outer induction over r and inner induction over n . Recall that $x_{0,-1}$ is the empty stream, so that property (1) is trivially satisfied for $n = -1$ for any r and $m_{r,-1}$. We choose $m_{r,-1} = 0$ for all r , and thus have solved the base case for the inner induction for all r .

Induction over r , base case. For $r = 0$ and $n \geq -1$, we need to prove that for every n there exists $m_{0,n}$ such that $\{\!\{p}\!\}_{m_{0,n}}(\bar{t})_{0:n} \geq \bar{t}_{0:n}$. For $n = -1$ the property has been established above. For $n \geq 0$, from the theorem hypothesis, $\llbracket p \rrbracket(\bar{t})_n \neq \bar{t}_n$, which implies that the fixed point computation proceeds to at least cycle n , and each time lst is incremented the values of \bar{t} for cycle lst are included in s (in line 11). Thus, for $m_{0,n} = n$, $\{\!\{p}\!\}_{m_{0,n}}(\bar{t})_{0:n} \geq \bar{t}_{0:n}$.

Induction step over r . As induction hypothesis we assume that for all n we have constructed $m_{r,n}$ such that $\{\!\{p}\!\}_{m_{r,n}}(\bar{t})_{0:n} \geq \{\!\{p}\!\}^r(\bar{t})_{0:n}$. From the monotony of $\{\!\{p}\!\}$ we have $\{\!\{p}\!\}^r(\bar{t}) \leq \{\!\{p}\!\}^{r+1}(\bar{t})$.

If $\{\!\{p}\!\}^r(\bar{t})_{0:n} = \{\!\{p}\!\}^{r+1}(\bar{t})_{0:n}$, then from the induction hypothesis we have $\{\!\{p}\!\}_{m_r}(\bar{t})_{0:n} \geq \{\!\{p}\!\}^{r+1}(\bar{t})_{0:n}$, so the induction step is verified for $m_{r+1,n} = m_{r,n}$.

If $\{\!\{p}\!\}^r(\bar{t})_{0:n} \neq \{\!\{p}\!\}^{r+1}(\bar{t})_{0:n}$, then $\{\!\{p}\!\}^r(\bar{t})_{0:n} < \{\!\{p}\!\}^{r+1}(\bar{t})_{0:n}$. This is equivalent to there existing $u \geq 1$ distinct variables $v_1, \dots, v_u \in W$ and $0 \leq k_1, \dots, k_u \leq n$ such that $\{\!\{p}\!\}^{r+1}(\bar{t})_{k_l}[v_l] > \{\!\{p}\!\}^r(\bar{t})_{k_l}[v_l]$. We will determine that for every $1 \leq l \leq u$ there exists $m^l > m_{r,n}$ such that $\{\!\{p}\!\}^{r+1}(\bar{t})_{k_l}[v_l] = \{\!\{p}\!\}_{m^l}(\bar{t})_{k_l}[v_l]$. Once this is proven, property (1) is satisfied for $m_{r+1,n} = \max_{l=1}^u m^l$.

Consider (v_l, k_l) one of the u pairs with $\{\!\{p}\!\}^{r+1}(\bar{t})_{k_l}[v_l] > \{\!\{p}\!\}^r(\bar{t})_{k_l}[v_l]$. By construction, v_l cannot be an input of p , but an output of a unique equation of p . Let eq be this equation. From the definition of $\{\!\{p}\!\}$, we know that the value is produced by $\{eq\}$ in the context provided by $\{\!\{p}\!\}^r(\bar{t})$. If eq is not a “fby back” equation, then this context can be restrained to $\{\!\{p}\!\}^r(\bar{t})_{0:n}$, which is included in $\{\!\{p}\!\}_{m_{r,n}}(\bar{t})_{0:n}$. Equation eq is therefore part of $todo_{k_l}$ in state $\{\!\{p}\!\}_{m_{r,n}}(\bar{t})$. As such, it will be executed in a finite number of steps of the fixed point iteration (in the worst case, after all variables are assigned a new value in all cycles $c \leq k_l$). Then, we can set $m^l = m_{r,n} + |W| * k_l$, where $|W|$ is the cardinal of W .

Finally, the case where $eq = “v_l = x \text{ fby back } y”$. From $\{\!\{p}\!\}^{r+1}(\bar{t})_{k_l}[v_l] > \{\!\{p}\!\}^r(\bar{t})_{k_l}[v_l]$ we determine that $\{\!\{p}\!\}^{r+1}(\bar{t})_{k_l}[v_l]$ is constructed during the $(r+1)^{th}$ application of $\{\!\{p}\!\}$. Constructing this value is done according to Definition 10. Then, the construction of $\{\!\{p}\!\}^{r+1}(\bar{t})_{k_l}[v_l]$ depends on a finite number of absent or present variable values of $\{\!\{p}\!\}^r(\bar{t})$. Let N be the maximum cycle index of such a variable value. Then, we know that all variable values needed to construct $\{\!\{p}\!\}^{r+1}(\bar{t})_{k_l}[v_l]$ are present in $\{\!\{p}\!\}_{m_{r,N}}(\bar{t})$. Then, we can set $m^l = m_{r,n} + |W| * N$. The proof is completed. \square

8 IMPLEMENTATION

To allow the evaluation of $R(ML)^2$, we have implemented a full front-end and simulator named `m1rc`. It covers the language introduced in the paper, extended with NumPy-like functional arrays and a (data) type system featuring 3 scalar types (`bool`, `int`, and a type floating point type `num`

that must be used on all variables subject to back-propagation) and a classical notation for tensor shapes. Thus, “`i:num [2,3]`” defines variable `i` as a 2-dimensional floating point tensor with 2 lines and 3 columns. To avoid ambiguity, scalars are represented as tensors with 0 dimensions. Thus, “`c:bool []`” defines a Boolean scalar. Our `mlrc` compiler relies on [XLA](#) to allow efficient execution of functions such as `Conv2D`.

The front-end includes type and shape inference phases that will assign a fixed type and shape to all variables, starting from those of the inputs. This is required to allow simulation.

Prior to simulation, the front-end will perform full type/shape inference, inlining of all nodes in the main node, and a pass of advanced constant folding (which itself requires clock inference) to reduce the state size.

The algorithm in [Figure 25](#) provides a first method to execute $R(ML)^2$ programs, one that was designed for generality (equivalence with denotational semantics) and simplicity. To attain efficiency, it must be further refined. The overall event-driven simulation method can be largely improved by optimizing the computation of the $todo_n$ sets and the search for an equation to activate, at each step of the algorithm. In `mlrc` we do both. Each time a new value is computed for a variable v in cycle n , we determine which equations having v as input can compute a new value. These equations are then added to the corresponding $todo_k$ set, to be removed when the corresponding value is computed. We also exclude from the search in lines 4-11 the *completed* cycles where all values have already been computed. Furthermore, the valuations of these cycles can be deleted from the simulation state provided the cycle is not adjacent to a non-completed one. This allows reducing to a minimum the number of currently open cycles of the simulation state.

Of course, in an ML context most performance gains require replacing the event-driven simulation with statically-scheduled code optimizable by a compiler such as XLA. We currently only do it for individual functions.

9 IMPLEMENTATION

Considering a *synchronous, bidirectionally causal* program P , the properties established in the previous section provide a concrete path to its interpretation. For the computation up to a given index m , [Theorem 4](#) states that it is sufficient to evaluate the semantic equations up to some bounded index $m + z$, and [Theorem 3](#) states that $z + 1$ live indices are sufficient to interpret the semantics equations of P up to index m . The value of z is a function of m in general, unless the program is actually bounded reactivewhere one may further establish an upper bound \bar{z} such that a maximal amount of $\bar{z} + 1$ live indices are sufficient.

This is not yet sufficient to allow a practical implementation: one still needs to compute z for any m . To that end, let $worklist_n$ denote the set of equations whose right-hand-side lead to a new (higher) value of a variable at cycle n : $worklist_m = \{(x = e) \in P \mid \llbracket x = e \rrbracket_m(V) \sqsupset V(x)_m\}$. This predicate may be evaluated recursively from the synchronous transformers in [Figure 23](#), considering only the instantaneous left-hand side of the \cdot operator, and memoizing the *window of live indices* lower or higher than the cycle m of interest. In particular, one may choose to always compute new values at the lowest index possible, matching the implementation of classical reactive languages where values are computed cycle by cycle. When all values at index m have reached a fixed point, execution can move to cycle $m + 1$, and memory for cycle m can be reclaimed except for values held by `fbv` operators. When the program is bounded reactive, the window of live indices remains bounded, making the implementation practical.

We implemented a front-end and interpreter for $R(ML)^2$ called `mlrc`. The syntax presented in the paper is extended with NumPy-like functional arrays and a (data) type system featuring 3 scalar types (`bool`, `int` and a floating point type `num` that must be used on all variables subject to back-propagation). Functional arrays feature a classical notation for tensor shapes: e.g. “`i:num [2,3]`”

defines variable `i` as a 2-dimensional floating point tensor with 2 rows and 3 columns. Scalars are represented as tensors with 0 dimensions: e.g. “`c : bool []`” defines a Boolean scalar. The front-end includes type and shape inference passes to assign a type and shape to all variables, starting from the inputs. The version `m1rc` provided as supplementary material only supports primitive Ocaml functions and operators, while the full version relies on XLA [39] to support realistic neural networks.

While the algorithm we just sketched interprets any $R(ML)^2$ program, there is a lot of room for performance improvement. In particular, the `m1rc` implementation specializes the computation of $worklist_n$ sets and the search for the next equation to evaluate. Each time a new value is computed for a variable v at cycle n , one may determine which equations having v as input can compute a new value. These equations are then added to the corresponding $worklist_k$ set, and retired when the value is eventually computed. One may also exclude from the search the completed cycles where all values have already been computed, and one may then recollect the storage locations associated with these unless the cycle is adjacent to a non-completed one.

10 CONCLUSION

We introduced $R(ML)^2$, a bidirectional reactive programming language on unbounded streams with both forward and backward dependences. We illustrated the modularity and expressiveness of the language on a variety of ML models, algorithms and usage scenarios, with unprecedented parsimony—four traditional reactive constructs plus one original backward recurrence primitive. We provided an denotational synchronous semantics with higher expressiveness than traditional synchronous Kahn networks preserving its main productivity and bounded memory execution properties. Starting from a single reactive program, one may transform it structurally to compute both forward and reverse-mode gradients, leveraging bidirectional reactive semantics on streams instead of referring to an external data structure such as a gradient tape or cache. From this single reactive program, one may also derive backpropagation and (batched) training algorithms by inserting dedicated control signals around initialization and sampling primitives. The language also unifies the expression of models interacting with their environment, from reinforcement learning to front-end/back-end data processing and real-time control. A prototype is provided as supplementary material, as well as detailed simulations and discussion of the denotational semantics.

REFERENCES

- [1] Martin Abadi, Michael Isard, and Derek G. Murray. A computational model for tensorflow. an introduction. In *Proceedings MAPL*, 2017.
- [2] E. A. Ashcroft and W. W. Wadge. Lucid, a nonprocedural language with iteration. *Commun. ACM*, 20(7):519–526, jul 1977.
- [3] Paul Barham and Michael Isard. Machine learning systems are stuck in a rut. In *Proceedings of the Workshop on Hot Topics in Operating Systems*, HotOS '19, page 177–183, New York, NY, USA, 2019. Association for Computing Machinery.
- [4] J.L. Bergerand, P. Caspi, D. Pilaud, N. Halbwachs, and E. Pilaud. Outline of a real time data flow language. In *Proceedings RTSS*, San Diego, CA, USA, December 1985.
- [5] D. Biernacki, co J.-L. Cola G. Hamon, and M. Pouzet. Clock-directed modular code generation for synchronous data-flow languages. In *Proceedings LCTES*, 2008.
- [6] T. Bourke, L. Brun, P.-E. Dagand, X. Leroy, M. Pouzet, and L. Rieg. A formally verified compiler for lustre. In *PLDI 2017: Proceedings of the 38th ACM SIGPLAN Conference on Programming Language Design and Implementation*, pages 586–601, Jun 2017.
- [7] Zixian Cai, Zhengyang Liu, S. Maleki, M. Musuvathi, T. Mytkowicz, J. Nelson, and O. Saarikivi. Synthesizing optimal collective algorithms. In *Proceedings PPOPP*, 2021.
- [8] Eduardo F Camacho and Carlos Bordons Alba. *Model predictive control*. Springer science & business media, 2013.
- [9] Paul Caspi, Adrian Curic, Aude Maignan, Christos Sofronis, Stavros Tripakis, and Peter Niebert. From simulink to scade/lustre to tta: A layered approach for distributed embedded applications. In *Proceedings of the 2003 ACM SIGPLAN Conference on Language, Compiler, and Tool for Embedded Systems*, LCTES '03, page 153–162, New York, NY, USA, 2003. Association for Computing Machinery.
- [10] Paul Caspi and Marc Pouzet. Synchronous kahn networks. *SIGPLAN Not.*, 31(6):226–238, jun 1996.
- [11] Shlomo E. Chazan, Jacob Goldberger, and Sharon Gannot. Deep recurrent mixture of experts for speech enhancement. In *2017 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 359–363, 2017.
- [12] J-L. Colaço and M. Pouzet. Clock as first class abstract types. In *EMSOFT 2003*, pages 134–135, 2003.
- [13] Gabriel Dulac-Arnold, Nir Levine, Daniel J. Mankowitz, Jerry Li, Cosmin Paduraru, Sven Gowal, and Todd Hester. Challenges of real-world reinforcement learning: definitions, benchmarks and analysis. *Machine Learning*, 110:2419–2468, 2021. <https://doi.org/10.1007/s10994-021-05961-4>.
- [14] Conal Elliott and Paul Hudak. Functional reactive animation. In *International Conference on Functional Programming*, 1997.
- [15] F. Gers, J. Schmidhuber, and F. Cummins. Learning to forget: Continual prediction with lstm. *Neural computation*, 12:2451–71, 10 2000.
- [16] Heitor Murilo Gomes, Jesse Read, Albert Bifet, Jean Paul Barddal, and João Gama. Machine learning for streaming data: State of the art, challenges, and opportunities. *SIGKDD Explor. Newsl.*, 21(2):6–22, nov 2019.
- [17] Yanping Huang, Youlong Cheng, Ankur Bapna, Orhan Firat, Mia Xu Chen, Dehao Chen, HyoukJoong Lee, Jiquan Ngiam, Quoc V. Le, Yonghui Wu, and Zhifeng Chen. Gpipe: Efficient training of giant neural networks using pipeline parallelism, 2019.
- [18] S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *ICML'15: Proceedings of the 32nd International Conference on International Conference on Machine Learnin*, volume 37, pages 448–456, Jul 2015.
- [19] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *CoRR*, abs/1502.03167, 2015.
- [20] Gilles Kahn. The semantics of a simple language for parallel programming. In Jack L. Rosenfeld, editor, *Information Processing, Proceedings of the 6th IFIP Congress 1974, Stockholm, Sweden*, pages 471–475. North-Holland, August 1974.
- [21] Michael G. Kapteyn, David J. Knezevic, and Karen Willcox. Toward predictive digital twins via component-based reduced-order models and interpretable machine learning. In *AIAA Scitech Forum*, 2020.
- [22] Qinghua Liu and Yuxiang Jiang. Dive into big model training, 2022.
- [23] Louis Mandel and Marc Pouzet. Reactiveml: a reactive extension to ML. In Pedro Barahona and Amy P. Felty, editors, *Proceedings of the 7th International ACM SIGPLAN Conference on Principles and Practice of Declarative Programming, July 11-13 2005, Lisbon, Portugal*, pages 82–93. ACM, 2005.
- [24] William Moses and Valentin Churavy. Instead of rewriting foreign code for machine learning, automatically synthesize fast gradients. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 12472–12485. Curran Associates, Inc., 2020.
- [25] M.C. Mozer. *Backpropagation: theory, architectures, and applications*, chapter A focused backpropagation algorithm for temporal pattern recognition. L. Erlbaum Associates Inc, 1995.

- [26] M.Schuster and K.K. Paliwal. Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, 45(11), Dec 1997.
- [27] Kaleb Ben Naveed, Zhiqian Qiao, and John M. Dolan. Trajectory planning for autonomous vehicles using hierarchical reinforcement learning. In *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*, pages 601–606, 2021.
- [28] V Peddinti, D. Povey, and S. Khudanpur. A time delay neural network architecture for efficient modeling of long temporal contexts. In *Proceedings INTERSPEECH*, Dresden, Germany, 2015.
- [29] Benjamin C. Pierce. Foundations for bidirectional programming. In *Proceedings of the 2nd International Conference on Theory and Practice of Model Transformations, ICMT '09*, page 1–3, Berlin, Heidelberg, 2009. Springer-Verlag.
- [30] Amir Pnueli. The temporal logic of programs. In *FOCS*, 1977.
- [31] H. Pompougnac, U. Beaugnon, A. Cohen, and D. Potop Butucaru. Weaving synchronous reactions into the fabric of ssa-form compilers. *ACM Trans. Archit. Code Optim.*, 19(2), mar 2022.
- [32] D. Potop-Butucaru and B. Caillaud. Correct-by-construction asynchronous implementation of modular synchronous specifications. *Fundamenta Informaticae*, CXXVII:1–27, 2006.
- [33] Marc Pouzet and Pascal Raymond. Modular static scheduling of synchronous data-flow networks - an efficient symbolic representation. *Des. Autom. Embed. Syst.*, 14(3):165–192, 2010.
- [34] Alexey Radul, Adam Paszke, Roy Frostig, Matthew J. Johnson, and Dougal Maclaurin. You only linearize once: Tangents transpose to gradients. *Proc. ACM Program. Lang.*, 7(POPL), jan 2023.
- [35] Oleg Rybakov, Natasha Kononenko, Niranjana Subrahmanya, Mirkó Visontai, and Stella Laurenzo. Streaming Keyword Spotting on Mobile Devices. In *Proc. Interspeech 2020*, 2020. <https://arxiv.org/abs/2005.06720>.
- [36] Mike Schuster and Kuldip Paliwal. Bidirectional recurrent neural networks. *Signal Processing, IEEE Transactions on*, 45, 12 1997.
- [37] Oscar Serradilla, Ekhi Zugasti, and Urko Zurutuza. Deep learning models for predictive maintenance: a survey, comparison, challenges and prospect, 2020.
- [38] Noam Shazeer, Azalia Mirhoseini, Krzysztof Maziarz, Andy Davis, Quoc Le, Geoffrey Hinton, and Jeff Dean. Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. Technical report, University of Toronto, 2017.
- [39] The XLA team. XLA. <https://github.com/openxla/xla>.
- [40] Towards AI editorial team. tesla’s self-driving algorithm explained. <https://towardsai.net/p/l/teslas-self-driving-algorithm-explained>. Published online on 27/05/2022.
- [41] T. Vlaar and B. Leimkuhler. Multirate training of neural networks. Technical report, University of Edinburgh, 2021.
- [42] P.J. Werbos. Backpropagation through time: what it does and how to do it. *Proceedings of the IEEE*, 78(10), Oct 1990.
- [43] Sam Whitlock, James R. Larus, and Edouard Bugnion. Extending tensorflow’s semantics with pipelined execution. *CoRR*, abs/1908.09291, 2019.
- [44] Tianfei Zhou, Fatih Porikli, David J. Crandall, Luc Van Gool, and Wenguan Wang. A survey on deep learning technique for video segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1–20, 2022.

A KAHN SEMANTICS

For reference and comparison, we recall the definition of the Kahn domain then tentatively derive an asynchronous semantics [20]. We show by example that such a semantics, being limited to prefix-closed stream values, leads to non-interesting least fixed points in the context of bidirectional reactive programming.

Kahn domain. As announced earlier, we first experiment with providing $R(ML)^2$ with an asynchronous dataflow semantics over finite and infinite streams by extending the Kahn network formalism.

Let \leq_K be the prefix order over streams: $x \leq_K y \iff \exists z, y = x \cdot z$. The ordered set $D_K = (T^\infty, \leq_K, \varepsilon)$ —known as the *Kahn domain* [20]—is a CPO with infimum ε and no supremum.

Asynchronous semantics on the Kahn domain. A tentative Kahn semantics extended to bidirectional reactivity is given in Figure 26. The interpretation of the associated stream transformers is given in Figure 27. The mark $\#_K$ distinguishes the syntactic construct from its interpretation as a stream transformer.

Notice the when false case may shrink an infinite stream into a finite (possibly empty) one. Notice also $\llbracket \text{post } e \rrbracket = \llbracket e \text{ when (false fby true)} \rrbracket$: in this asynchronous semantics, post happens to be syntactic sugar for sampling with the stream $\text{false} \cdot \text{true}^\omega$.

$$\begin{aligned}
\llbracket c \rrbracket(V) &\stackrel{\text{def}}{=} c^\omega \\
\llbracket x \rrbracket(V) &\stackrel{\text{def}}{=} V(x) \\
\llbracket i \text{ fby } e \rrbracket(V) &\stackrel{\text{def}}{=} \text{fby}^\#(V(\llbracket i \rrbracket(V)), V(\llbracket e \rrbracket(V))) \\
\llbracket \text{post } e \rrbracket(V) &\stackrel{\text{def}}{=} \text{post}^\#(V(\llbracket e \rrbracket(V))) \\
\llbracket e \text{ when } b \rrbracket(V) &\stackrel{\text{def}}{=} \text{when}^\#(V(\llbracket e \rrbracket(V)), \llbracket b \rrbracket(V)) \\
\llbracket \text{merge } b \text{ t } f \rrbracket(V) &\stackrel{\text{def}}{=} \text{merge}^\#(V(\llbracket b \rrbracket(V)), \llbracket t \rrbracket(V), \llbracket f \rrbracket(V)) \\
\llbracket f(e) \rrbracket(V) &\stackrel{\text{def}}{=} V(f)(\llbracket e \rrbracket) \\
\llbracket \text{node } f(i) \rightarrow (o) \text{ eq} \rrbracket(V) &\stackrel{\text{def}}{=} f := \lambda d. \llbracket \pi_o(\text{eq}) \rrbracket(V, i := d) \text{ where } i \subseteq \text{FV}(\text{eq}) \wedge o \subseteq \text{DV}(\text{eq}) \\
\llbracket x = e \rrbracket(V) &\stackrel{\text{def}}{=} \llbracket e \rrbracket(V, x := x_{\text{FIX}}) \text{ where } x_{\text{FIX}} = \text{FIX}(\lambda d. \llbracket e \rrbracket(V, x := d))
\end{aligned}$$

Fig. 26. Kahn semantics.

$$\begin{aligned}
\text{fby}_K^\#(i, \varepsilon) &\stackrel{\text{def}}{=} \varepsilon \\
\text{fby}_K^\#(\varepsilon, s) &\stackrel{\text{def}}{=} \varepsilon \\
\text{fby}_K^\#(i, s) &\stackrel{\text{def}}{=} i_0 \cdot s \\
\text{post}_K^\#(e) &\stackrel{\text{def}}{=} \varepsilon \\
\text{post}_K^\#(s_0 \cdot s_1) &\stackrel{\text{def}}{=} s_1 \\
\text{when}_K^\#(s, \varepsilon) &\stackrel{\text{def}}{=} \varepsilon \\
\text{when}_K^\#(\varepsilon, b) &\stackrel{\text{def}}{=} \varepsilon \\
\text{when}_K^\#(s_0 \cdot s_1, \text{true} \cdot b_1) &\stackrel{\text{def}}{=} s_0 \cdot \text{when}_K^\#(s_1, b_1) \\
\text{when}_K^\#(s_0 \cdot s_1, \text{false} \cdot b_1) &\stackrel{\text{def}}{=} \text{when}_K^\#(s_1, b_1) \\
\text{merge}_K^\#(\varepsilon, t, f) &\stackrel{\text{def}}{=} \varepsilon \\
\text{merge}_K^\#(\text{true} \cdot b_1, \varepsilon, f) &\stackrel{\text{def}}{=} \varepsilon \\
\text{merge}_K^\#(\text{false} \cdot b_1, t, \varepsilon) &\stackrel{\text{def}}{=} \varepsilon \\
\text{merge}_K^\#(\text{true} \cdot b_1, t_0 \cdot t_1, f) &\stackrel{\text{def}}{=} t_0 \cdot \text{merge}_K^\#(b_1, t_1, f) \\
\text{merge}_K^\#(\text{false} \cdot b_1, t, f_0 \cdot f_1) &\stackrel{\text{def}}{=} f_0 \cdot \text{merge}_K^\#(b_1, t, f_1)
\end{aligned}$$

Fig. 27. Stream transformers for the primitives of the Kahn semantics.

Figure 28 illustrates these definitions on the backfill example. Each row corresponds to the evaluation of one rule in the fixed point computation, effectively computing a new finite or infinite stream one row at a time. Blank cells are meant for visual alignment only, they are not meaningful in the D_K domain. From the $\text{merge}_K^\#(\text{false} \cdot b_1, t, \varepsilon)$ case of the $\text{merge}_K^\#$ stream transformer, the Kahn semantics computes an empty stream for o at the first iteration. As a result, it immediately reaches a fixed point, making no progress at all in the computation of o . This deadlocking behavior is the direct result of the domain operating on prefixed-closed streams and is common to reactive languages on streams [4, 14].

```

1 node backfill(i,bp)->(o)
2   o = merge bp (i when bp) ((post o) when not bp);
3

```

| | | | | | | | |
|--|-------|-------|------|------|-------|------|-----|
| $\kappa \llbracket i \rrbracket$ | 0 | 1 | 2 | 3 | 4 | 5 | ... |
| $\kappa \llbracket bp \rrbracket$ | false | false | true | true | false | true | ... |
| $\kappa \llbracket i \text{ when } bp \rrbracket$ | | | 2 | 3 | | 5 | ... |
| $\kappa \llbracket \text{post } o \rrbracket$ | | | | | | | |
| $\kappa \llbracket (\text{post } o) \text{ when not } bp \rrbracket$ | | | | | | | |
| $\kappa \llbracket o \rrbracket$ | | | | | | | |
| $\kappa \llbracket \text{backfill } (i, bp) \rrbracket$ | | | | | | | |

Fig. 28. Illustration of the Kahn semantics deadlocking on a backward recurrence.

B CYCLE-BY-CYCLE SEMANTIC SIMULATIONS

Beyond the backfill example discussed in the main paper, Figure 29 illustrates a more complex case of sampling over values computed with the lookahead operator. Notice how \perp values in o stick in at cycles 0 and 1, and eventually disappear at cycle 2 ($\llbracket o \rrbracket_{0:3}$).

Figure 30 explores the semantics of merge (starting at cycle 2 for readability): the interpretation switches from \perp to known (blue) values much earlier, without waiting for the prefix of c to be known in full. Figure 31 goes one step further, illustrating the expressiveness of the synchronous semantics by revisiting the recurrent merge case (starting at cycle 2).

```

1 node sample_backfill(i,bp)->(o)
2   nh = false fby (not nh);
3   c = backfill(nh,bp);
4   o = i when c;

```

| | | | | | | |
|--|---|---|---|---|---|---|
| $\llbracket i \rrbracket_{0:6}$ $\llbracket bp \rrbracket_{0:6}$ $\llbracket nh \rrbracket_{0:6}$ | 0 false | 1 false | 2 true | 3 true | 4 false | 5 true |
| $\llbracket nh \text{ when } bp \rrbracket_{0:1}$ $\llbracket \text{post } c \rrbracket_{0:1}$ $\llbracket (\text{post } c) \text{ when not } bp \rrbracket_{0:1}$ $\llbracket c \rrbracket_{0:1}$ $\llbracket o \rrbracket_{0:1}$ | \square \perp \perp \perp \perp | | | | | |
| $\llbracket nh \text{ when } bp \rrbracket_{0:2}$ $\llbracket \text{post } c \rrbracket_{0:2}$ $\llbracket (\text{post } c) \text{ when not } bp \rrbracket_{0:2}$ $\llbracket c \rrbracket_{0:2}$ $\llbracket o \rrbracket_{0:2}$ | \square \perp \perp \perp \perp | \square \perp \perp \perp \perp | | | | |
| $\llbracket nh \text{ when } bp \rrbracket_{0:3}$ $\llbracket \text{post } c \rrbracket_{0:3}$ $\llbracket (\text{post } c) \text{ when not } bp \rrbracket_{0:3}$ $\llbracket c \rrbracket_{0:3}$ $\llbracket o \rrbracket_{0:3}$ | \square \perp \perp \perp \perp | \square \perp \perp \perp \perp | false \perp \square false \perp | | | |
| $\llbracket \text{post } c \rrbracket_{0:3}$ $\llbracket (\text{post } c) \text{ when not } bp \rrbracket_{0:3}$ $\llbracket c \rrbracket_{0:3}$ $\llbracket o \rrbracket_{0:3}$ | \perp \perp \perp \perp | false false false \square | \perp \square false \square | | | |
| $\llbracket \text{post } c \rrbracket_{0:3}$ $\llbracket (\text{post } c) \text{ when not } bp \rrbracket_{0:3}$ $\llbracket c \rrbracket_{0:3}$ $\llbracket o \rrbracket_{0:3}$ | false false false \square | false false false \square | \perp \square false \square | | | |
| $\llbracket nh \text{ when } bp \rrbracket_{0:4}$ $\llbracket \text{post } c \rrbracket_{0:4}$ $\llbracket (\text{post } c) \text{ when not } bp \rrbracket_{0:4}$ $\llbracket c \rrbracket_{0:4}$ $\llbracket o \rrbracket_{0:4}$ | \square false false false \square | \square false false false \square | false false false false \square | true \perp \square true 3 | | |
| $\llbracket nh \text{ when } bp \rrbracket_{0:5}$ $\llbracket \text{post } c \rrbracket_{0:5}$ $\llbracket (\text{post } c) \text{ when not } bp \rrbracket_{0:5}$ $\llbracket c \rrbracket_{0:5}$ $\llbracket o \rrbracket_{0:5}$ | \square false false false \square | \square false false false \square | false false false false \square | true \perp \square true 3 | \perp \perp \perp \perp \perp | |
| $\llbracket nh \text{ when } bp \rrbracket_{0:6}$ $\llbracket \text{post } c \rrbracket_{0:6}$ $\llbracket (\text{post } c) \text{ when not } bp \rrbracket_{0:6}$ $\llbracket c \rrbracket_{0:6}$ $\llbracket o \rrbracket_{0:6}$ | \square false false false \square | \square false false false \square | false false false false \square | true \perp \square true 3 | \square \perp \perp \perp 3 | true \perp \square true 5 |
| $\llbracket \text{post } c \rrbracket_{0:6}$ $\llbracket (\text{post } c) \text{ when not } bp \rrbracket_{0:6}$ $\llbracket c \rrbracket_{0:6}$ $\llbracket o \rrbracket_{0:6}$ | false false false \square | false false false \square | true true true \square | \perp true true 3 | true \perp true 4 | \perp \square true 5 |
| $\llbracket \text{post } c \rrbracket_{0:6}$ $\llbracket (\text{post } c) \text{ when not } bp \rrbracket_{0:6}$ $\llbracket c \rrbracket_{0:6}$ $\llbracket o \rrbracket_{0:6}$ | false false false \square | false false false \square | true true true \square | \perp true true 3 | true true true 4 | \perp \square true 5 |
| $\llbracket \text{sample_backfill}(i,bp) \rrbracket_{0:6}$ | \square | \square | \square | 3 | 4 | 5 |

Fig. 29. Illustration of the synchronous semantics, sampling on a lookahead condition.

```

1 node merge_backfill(i,bp)->(o)
2   nh = false fby (not nh);
3   c = backfill(nh,bp);
4   o = merge c (i when c) (-i when not c);
5

```

| | | | | | | |
|---|---|---|---|---|---|--|
| $\llbracket i \rrbracket_{0:6}$ $\llbracket bp \rrbracket_{0:6}$ $\llbracket nh \rrbracket_{0:6}$ | 0 false false | 1 false true | 2 true false | 3 true true | 4 false false | 5 true true |
| $\llbracket nh \text{ when } bp \rrbracket_{0:3}$ $\llbracket \text{post } c \rrbracket_{0:3}$ $\llbracket (\text{post } c) \text{ when not } bp \rrbracket_{0:3}$ $\llbracket c \rrbracket_{0:3}$ $\llbracket i \text{ when } c \rrbracket_{0:3}$ $\llbracket \neg i \text{ when not } c \rrbracket_{0:3}$ $\llbracket o \rrbracket_{0:3}$ | \square \perp \perp \perp \perp \perp \perp | \square \perp \perp \perp \perp \perp \perp | false \perp \square false \square -2 -2 | | | |
| $\llbracket \text{post } c \rrbracket_{0:3}$ $\llbracket (\text{post } c) \text{ when not } bp \rrbracket_{0:3}$ $\llbracket c \rrbracket_{0:3}$ $\llbracket i \text{ when } c \rrbracket_{0:3}$ $\llbracket \neg i \text{ when not } c \rrbracket_{0:3}$ $\llbracket o \rrbracket_{0:3}$ | \perp \perp \perp \perp \perp \perp | false false false \square -1 -1 | \perp \square false \square -2 -2 | | | |
| $\llbracket \text{post } c \rrbracket_{0:3}$ $\llbracket (\text{post } c) \text{ when not } bp \rrbracket_{0:3}$ $\llbracket c \rrbracket_{0:3}$ $\llbracket i \text{ when } c \rrbracket_{0:3}$ $\llbracket \neg i \text{ when not } c \rrbracket_{0:3}$ $\llbracket o \rrbracket_{0:3}$ | false false false \square 0 0 | false false false \square -1 -1 | \perp \square false \square -2 -2 | | | |
| $\llbracket nh \text{ when } bp \rrbracket_{0:4}$ $\llbracket \text{post } c \rrbracket_{0:4}$ $\llbracket (\text{post } c) \text{ when not } bp \rrbracket_{0:4}$ $\llbracket c \rrbracket_{0:4}$ $\llbracket i \text{ when } c \rrbracket_{0:4}$ $\llbracket \neg i \text{ when not } c \rrbracket_{0:4}$ $\llbracket o \rrbracket_{0:4}$ | \square false false false \square 0 0 | \square false false false \square -1 -1 | false \perp \square false \square -2 -2 | true \perp \square true 3 \square 3 | | |
| $\llbracket \text{post } c \rrbracket_{0:4}$ | false | false | true | \perp | | |
| $\llbracket nh \text{ when } bp \rrbracket_{0:5}$ $\llbracket \text{post } c \rrbracket_{0:5}$ $\llbracket (\text{post } c) \text{ when not } bp \rrbracket_{0:5}$ $\llbracket c \rrbracket_{0:5}$ $\llbracket i \text{ when } c \rrbracket_{0:5}$ $\llbracket \neg i \text{ when not } c \rrbracket_{0:5}$ $\llbracket o \rrbracket_{0:5}$ | \square false false false \square 0 0 | \square false false false \square -1 -1 | false true \square false \square -2 -2 | true \perp \square true 3 \square 3 | \square \perp \perp \perp \perp \perp \perp | |
| $\llbracket nh \text{ when } bp \rrbracket_{0:6}$ $\llbracket \text{post } c \rrbracket_{0:6}$ $\llbracket (\text{post } c) \text{ when not } bp \rrbracket_{0:6}$ $\llbracket c \rrbracket_{0:6}$ $\llbracket i \text{ when } c \rrbracket_{0:6}$ $\llbracket \neg i \text{ when not } c \rrbracket_{0:6}$ $\llbracket o \rrbracket_{0:6}$ | \square false false false \square 0 0 | \square false false false \square -1 -1 | false true \square false \square -2 -2 | true \perp \square true 3 \perp 3 | \square \perp \perp true \perp \perp \perp | true \perp \square true 5 \square 5 |
| $\llbracket \text{post } c \rrbracket_{0:6}$ $\llbracket (\text{post } c) \text{ when not } bp \rrbracket_{0:6}$ $\llbracket c \rrbracket_{0:6}$ $\llbracket i \text{ when } c \rrbracket_{0:6}$ $\llbracket \neg i \text{ when not } c \rrbracket_{0:6}$ $\llbracket o \rrbracket_{0:6}$ | false false false \square 0 0 | false false false \square -1 -1 | true \square false \square 3 3 | \perp \square true 3 4 4 | true true true 4 5 5 | \perp \square \square \square \square \perp |
| $\llbracket \text{post } c \rrbracket_{0:6}$ | false | false | true | true | true | \perp |
| $\llbracket \text{merge_backfill}(i,bp) \rrbracket_{0:6}$ | 0 | -1 | -2 | 3 | 4 | 5 |

Fig. 30. Illustration of the synchronous semantics, merging on a lookahead condition.

```

1 node rec_merge_backfill(bp)->(o)
2   nh = false fby (not nh);
3   c = merge bp (nh when bp) ((post o) when not bp);
4   o = backfill(nh, c);
5

```

| | | | | | | |
|---|--|---|---|---|--|--------------|
| [[nh]] _{0:6} [[bp]] _{0:6} | false false | true false | false true | true true | false false | true true |
| [[nh when bp]] _{0:3} [[post o]] _{0:3} [[post o) when not bp]] _{0:3} [[c]] _{0:3} [[nh when c]] _{0:3} [[post o) when not c]] _{0:3} [[o]] _{0:3} | ⊠ ⊠ ⊠ ⊠ ⊠ ⊠ ⊠ | ⊠ ⊠ ⊠ ⊠ ⊠ ⊠ ⊠ | false ⊠ ⊠ false ⊠ ⊠ ⊠ | | | |
| [[nh when bp]] _{0:4} [[post o]] _{0:4} [[post o) when not bp]] _{0:4} [[c]] _{0:4} [[nh when c]] _{0:4} [[post o) when not c]] _{0:4} [[o]] _{0:4} | ⊠ ⊠ ⊠ ⊠ ⊠ ⊠ ⊠ | ⊠ ⊠ ⊠ ⊠ ⊠ ⊠ ⊠ | false ⊠ ⊠ false ⊠ ⊠ ⊠ | true ⊠ ⊠ true true ⊠ true | | |
| [[post o]] _{0:4} [[post o) when not bp]] _{0:4} [[c]] _{0:4} [[nh when c]] _{0:4} [[post o) when not c]] _{0:4} [[o]] _{0:4} | ⊠ ⊠ ⊠ ⊠ ⊠ ⊠ | ⊠ ⊠ ⊠ ⊠ ⊠ ⊠ | true ⊠ false ⊠ true true | ⊠ ⊠ true ⊠ ⊠ true | | |
| [[post o]] _{0:4} [[post o) when not bp]] _{0:4} [[c]] _{0:4} [[nh when c]] _{0:4} [[post o) when not c]] _{0:4} [[o]] _{0:4} | ⊠ ⊠ ⊠ ⊠ ⊠ ⊠ | true true true ⊠ ⊠ true | true ⊠ false ⊠ ⊠ true | ⊠ ⊠ true ⊠ ⊠ true | | |
| [[post o]] _{0:4} [[post o) when not bp]] _{0:4} [[c]] _{0:4} [[nh when c]] _{0:4} [[post o) when not c]] _{0:4} [[o]] _{0:4} | true true true false ⊠ false | true true true true ⊠ true | true ⊠ false ⊠ ⊠ true | ⊠ ⊠ true ⊠ ⊠ true | | |
| [[nh when bp]] _{0:5} [[post o]] _{0:5} [[post o) when not bp]] _{0:5} [[c]] _{0:5} [[nh when c]] _{0:5} [[post o) when not c]] _{0:5} [[o]] _{0:5} | ⊠ true true false ⊠ false | ⊠ true true true true true | false true false ⊠ true true | true ⊠ ⊠ true ⊠ true | ⊠ ⊠ ⊠ ⊠ ⊠ ⊠ | |
| [[nh when bp]] _{0:6} [[post o]] _{0:6} [[post o) when not bp]] _{0:6} [[c]] _{0:6} [[nh when c]] _{0:6} [[post o) when not c]] _{0:6} [[o]] _{0:6} | ⊠ true true true false ⊠ false | ⊠ true true true true true true | false true false ⊠ true true | true ⊠ ⊠ true ⊠ ⊠ true | true ⊠ ⊠ true ⊠ ⊠ true | |
| [[post o]] _{0:6} [[post o) when not bp]] _{0:6} [[c]] _{0:6} [[nh when c]] _{0:6} [[post o) when not c]] _{0:6} [[o]] _{0:6} | true true true false ⊠ false | true true true true ⊠ true | ⊠ ⊠ false ⊠ ⊠ true | true ⊠ true ⊠ ⊠ true | ⊠ ⊠ true ⊠ ⊠ true | |
| [[post o]] _{0:6} | true | true | true | false | true | ⊠ |
| [[rec_merge_backfill(i, bp)]] _{0:6} | false | true | true | true | false | true |

Fig. 31. Illustration of the synchronous semantics, merging on a recurrently defined lookahead condition.