Nondeterministic Asynchronous Dataflow in Isabelle/HOL

Rafael Castro G. Silva, Laouen Fernet and Dmitriy Traytel

Department of Computer Science University of Copenhagen

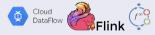
14/05/2025

Motivation

Motivation

Context:

- Stream Processing: programs that compute (possibly) unbounded sequences of data (streams)
- A common problem in the industry
- Frameworks:
 Apache Flink, Apache Samza, Apache Spark, Google Cloud Dataflow, and Timely Dataflow



- Why use frameworks?
 - Highly Parallel
 - Low latency (output as soon as possible)
 - Incremental computing (re-uses previous computations)

Motivation

Context:

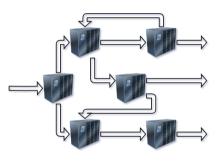
- Stream Processing: programs that compute (possibly) unbounded sequences of data (streams)
- A common problem in the industry
- Frameworks:
 Apache Flink, Apache Samza, Apache Spark, Google Cloud Dataflow, and Timely Dataflow



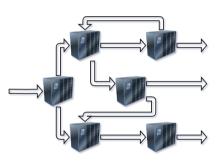
- Why use frameworks?
 - Highly Parallel
 - Low latency (output as soon as possible)
 - Incremental computing (re-uses previous computations)

Our goal:

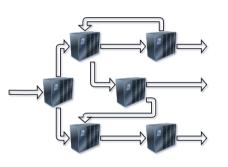
Mechanically Verify Timely Dataflow algorithms



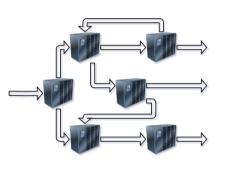
• Nondeterministic Asynchronous Dataflow



- Nondeterministic Asynchronous Dataflow
 - Dataflow: Directed graph of interconnected operators



- Nondeterministic Asynchronous Dataflow
 - Dataflow: Directed graph of interconnected operators
 - Asynchronous:
 - Operators execute independently: processes without an orchestrator
 - Operators can freely communicate with the network (read/write); do silent computation steps
 - Networks are unbounded FIFO queues



- Nondeterministic Asynchronous Dataflow
 - Dataflow: Directed graph of interconnected operators
 - Asynchronous:
 - Operators execute independently: processes without an orchestrator
 - Operators can freely communicate with the network (read/write); do silent computation steps
 - Networks are unbounded FIFO queues
 - Nondeterministic:
 - Operators can make nondeterministic choices
 - Operators are relations between inputs and outputs sequences

The Algebra for Nondeterministic Asynchronous Dataflow

- Bergstra et al. presents an algebra for Nondeterministic Asynchronous Dataflow
- Primitives: sequential and parallel composition; feedback loop...
- The 52 axioms
- An process calculus instance

Network Algebra for Asynchronous Dataflow*

J.A. Bergstra^{1,2,†} C.A. Middelburg^{2,3,§} Gh. Ştefănescu⁴

¹Programming Research Group, University of Amsterdam P.O. Box 41882, 1009 DB Amsterdam, The Netherlands

²Department of Philosophy, Utrecht University P.O. Box 80126, 3508 TC Utrecht, The Netherlands

³Department of Network & Service Control, KPN Research P.O. Box 421, 2260 AK Leidschendam, The Netherlands

⁴Institute of Mathematics of the Romanian Academy P.O. Box 1-764, 70700 Bucharest, Romania

E-mail: janb@fwi.uva.nl - keesm@phil.ruu.nl - ghstef@inar.ro

Isabelle/HOL Preliminaries

Isabelle/HOL

Classical higher-order logic (HOL):
 Simple Typed Lambda Calculus + axiom of choice + axiom of infinity + rank-1 polymorphism

Isabelle/HOL

- Classical higher-order logic (HOL):
 Simple Typed Lambda Calculus + axiom of choice + axiom of infinity + rank-1 polymorphism
- Isabelle: A generic proof assistant



Isabelle/HOL: Isabelle's flavor of HOL

Isabelle/HOL

- Classical higher-order logic (HOL):
 Simple Typed Lambda Calculus + axiom of choice + axiom of infinity + rank-1 polymorphism
- Isabelle: A generic proof assistant



• Isabelle/HOL: Isabelle's flavor of HOL

Why Isabelle/HOL?

- Codatatypes: (possibly) infinite data structures (e.g., lazy lists, streams)
- Corecursion: always eventually produces some codatatype constructor
- Coinductive predicate: infinite number of introduction rule applications
- Coinduction: reason about coinductive predicates

Operators as a Codatatype

Operators

Operators in Isabelle/HOL

```
codatatype (inputs: 'i, outputs: 'o, 'd) op =
Read 'i ('d \Rightarrow ('i, 'o, 'd) op) | Write (('i, 'o, 'd) op) 'o 'd
Silent ('i, 'o, 'd) op | Choice (('i, 'o, 'd) op) cset
```

- Type parameters: inputs/output ports; data
- Operator's actions
- Possibly infinite trees
- inputs/outputs: Sets of used ports

Examples 1

Uncommunicative operators

abbreviation

$$\emptyset \equiv \mathsf{Choice}\ \{\}_c$$

$$\otimes = \mathsf{Choice}\; ((\lambda_{-}.\otimes)_{c} \; \{()\}_{c})$$

corec silent_op (⊙) where

$$\odot = \mathsf{Silent} \ \odot$$

lemma spin_op_code:

$$\otimes = \mathsf{Choice}\ \{\otimes\}_{c}$$

- Quirk: any corecursive call guarded by the Choice must be applied using the proper map function
- They have the same meaning.

Examples 1

Uncommunicative operators

abbreviation

$$\oslash \equiv \mathsf{Choice}\ \{\}_c$$

corec spin op
$$(\otimes)$$
 where

$$\otimes = \mathsf{Choice}\; ((\lambda_{-}.\otimes)_{c} \; \{()\}_{c})$$

corec silent_op (⊙) where

 $\odot = \mathsf{Silent} \,\, \odot$

lemma spin_op_code:

 $\otimes = \mathsf{Choice}\ \{\otimes\}_{\mathit{c}}$

- Quirk: any corecursive call guarded by the Choice must be applied using the proper map function
- They have the same meaning. But are syntactically different ©

Operators Equivalences: Motivation

An ideal equivalence relation

- Only equate operators that **behave** the same
- Useful reasoning principle

Operators Equivalences: Motivation

An ideal equivalence relation

- Only equate operators that **behave** the same
- Useful reasoning principle

Milner's approach: Bisimilarity



- Based on labeled transition systems (LTS)
- Labels (actions): Read, Write, Silent (τ)

Intuition

Two operators are bisimilar if their corresponding transition systems can mutually simulate each other's transitions.

14/05/2025

Operators Equivalences: Strong Bisimilarity

Label Transition System

```
datatype ('i, 'o, 'd) IO = Inp 'i 'd | Out 'o 'd | Tau
inductive step where
step (Inp p x) (Read p f) (f x) | step (Out q x) (Write op q x) op
| step Tau (Silent op) op
| op \in_c ops \implies step io op op' \implies step io (Choice ops) op'
```

Operators Equivalences: Weak Bisimilarity

• foo

Asynchronous Dataflow Operators

Buffer Infrastructure

• foo

Asynchronous Dataflow Properties

```
\begin{aligned} & \text{B1: } op_1 \parallel (op_2 \parallel op_3) \approx \text{map\_op} & & & & & & & & & \\ & \text{B2\_1: } op \parallel (\mathcal{I} :: (\theta, 0, 'd) \ op) \approx \text{map\_op} \ \text{Inl Inl } op \\ & \text{B2\_2: } (\mathcal{I} :: (\theta, 0, 'd) \ op) \parallel op \approx \text{map\_op} \ \text{Inl Inl } op \\ & \text{B3: } (op_1 \bullet op_2) \bullet op_3 \approx op_1 \bullet (op_2 \bullet op_3) \\ & \text{B4\_1: } op \bullet \mathcal{I} \approx op \bullet \mathcal{I} \approx op \bullet \\ & \text{B5: } (op_1 \parallel op_2) \bullet (op_3 \parallel op_4) \approx (op_1 \bullet op_3) \parallel (op_2 \bullet op_4) \\ & \text{B6: } \mathcal{I} \parallel \mathcal{I} \approx \mathcal{I} \\ & \text{B7: } \mathcal{X} \bullet \mathcal{X} \approx \mathcal{I} \\ & \text{B8: } (\mathcal{X} :: ('i + \theta, \theta + 'i, 'd) \ op) \approx \text{map\_op} \ \text{id} \ \text{(case\_sum Inr Inl)} \ \mathcal{I} \\ & \text{B9: } \mathcal{X} \approx \text{map\_op} \ \cap \ \cap \ (\mathcal{X} \parallel \mathcal{I}) \bullet \text{map\_op} \ \text{id} \ \cap \ \mathcal{I} \parallel \mathcal{X} \\ & \text{B10: } (op_1 \parallel op_2) \bullet \mathcal{X} \approx \mathcal{X} \bullet (op_2 \parallel \theta op_1 \parallel op_1) \\ & \text{F1: } \mathcal{I} \uparrow \approx \langle \mathcal{I} :: (\theta, \theta, 'd) \ op) \end{aligned}
```

```
R1: \operatorname{Inr} : \operatorname{inputs} op_1 \cap \operatorname{defaults} = \{\} \Longrightarrow \operatorname{Inr} : \operatorname{outputs} op_1 \cap \operatorname{defaults} = \{\} \Longrightarrow op_2 \bullet (op_1 \uparrow) \approx ((op_2 \parallel \mathcal{I}) \bullet op_1) \uparrow
R2: \operatorname{Inr} : \operatorname{inputs} op_1 \cap \operatorname{defaults} = \{\} \Longrightarrow \operatorname{Inr} : \operatorname{outputs} op_1 \cap \operatorname{defaults} = \{\} \Longrightarrow (op_1 \uparrow) \bullet op_2 \approx (op_1 \bullet (op_2 \parallel \mathcal{I})) \uparrow
R3: \operatorname{Inr} : \operatorname{inputs} op_2 \cap \operatorname{defaults} = \{\} \Longrightarrow \operatorname{Inr} : \operatorname{outputs} op_2 \cap \operatorname{defaults} = \{\} \Longrightarrow \operatorname{op}_1 \parallel (op_2 \uparrow) \approx (\operatorname{map\_op} \land \land (op_1 \parallel op_2)) \uparrow
R4: \operatorname{Inr} : \operatorname{inputs} op_1 \cap \operatorname{defaults} = \{\} \Longrightarrow \operatorname{inr} : \operatorname{outputs} op_1 \cap \operatorname{defaults} = \{\} \Longrightarrow \operatorname{inputs} op_2 \cap \operatorname{defaults} = \{\} \Longrightarrow (\exists op_1 \bullet (\mathcal{I} \parallel op_2)) \uparrow \approx ((\mathcal{I} \parallel op_2) \bullet op_1 \vdash) \uparrow
R5: \operatorname{Inr} : \operatorname{inputs} op_2 = \{\} \Longrightarrow \operatorname{Inr} : \operatorname{outputs} op_2 \in \{\} \Longrightarrow \operatorname{map\_op} \operatorname{Inl} \operatorname{ln} ((op: :: ('i + \theta, 'o + \theta, 'd) op) \uparrow) \approx op
R6: \operatorname{Inr} : \operatorname{inputs} op_2 = \{\} \Longrightarrow \operatorname{Inr} : \operatorname{outputs} op_2 = \{\} \Longrightarrow \operatorname{Inr} : \operatorname{inputs} op_2 = \{\} \Longrightarrow \operatorname{Inr} : \operatorname{outputs} op_2 = \{\} \Longrightarrow \operatorname{outputs} op_
```

```
\begin{array}{lll} \mathrm{R1:} & \operatorname{Inr} \stackrel{>}{\sim} & \operatorname{inputs} \ op_1 \cap \operatorname{defaults} = \{\} \Longrightarrow \operatorname{Inr} \stackrel{>}{\sim} & \operatorname{outputs} \ op_1 \cap \operatorname{defaults} = \{\} \Longrightarrow \operatorname{op_2} \bullet (\operatorname{op_1}) \uparrow \approx ((\operatorname{op_2} \parallel \mathcal{I}) \bullet \operatorname{op_1}) \uparrow \\ \mathrm{R2:} & \operatorname{Inr} \stackrel{>}{\sim} & \operatorname{inputs} \ op_1 \cap \operatorname{defaults} = \{\} \Longrightarrow \operatorname{Inr} \stackrel{>}{\sim} & \operatorname{outputs} \ op_1 \cap \operatorname{defaults} = \{\} \Longrightarrow \operatorname{Inr} \stackrel{>}{\sim} & \operatorname{outputs} \ op_2 \cap \operatorname{defaults} = \{\} \Longrightarrow \operatorname{Inr} \stackrel{>}{\sim} & \operatorname{outputs} \ op_2 \cap \operatorname{defaults} = \{\} \Longrightarrow \operatorname{Inr} \stackrel{>}{\sim} & \operatorname{outputs} \ op_2 \cap \operatorname{defaults} = \{\} \Longrightarrow \operatorname{Inr} \stackrel{>}{\sim} & \operatorname{outputs} \ op_1 \cap \operatorname{defaults} = \{\} \Longrightarrow \operatorname{inr} \implies op_2 \cap \operatorname{defaults} = \{\} \Longrightarrow \operatorname{inr} \implies op_1 \cap \operatorname{defaults} = \{\} \Longrightarrow \operatorname{inr} \implies op_2 \cap \operatorname{defaults} = \{\} \Longrightarrow \operatorname{inr} \implies op_1 \cap \operatorname{defaults} = \{\} \Longrightarrow \operatorname{inr} \implies op_2 \cap \operatorname{op_2} \cap \cap \operatorname{op_2} \cap \operatorname{op_2} \cap \operatorname{op_2} \cap \operatorname{op_2} \cap \operatorname{op_2} \cap \operatorname{op_2} \cap \operatorname{o
```

```
\begin{aligned} & \text{B1: } op_1 \parallel (op_2 \parallel op_3) \approx \text{map\_op} & & & & & & & & & & \\ & \text{B2\_1: } op \parallel \|\mathcal{I} :: (\theta, 0, |d) \ op) \approx \text{map\_op} \ \text{Inl Inl } op \\ & \text{B2\_2: } (\mathcal{I} :: (\theta, 0, |d) \ op) \parallel op \approx \text{map\_op} \ \text{Inl Inl } op \\ & \text{B3: } (op_1 \bullet op_2) \bullet op_3 \approx op_1 \bullet (op_2 \bullet op_3) \\ & \text{B4\_1: } op \models \bullet \mathcal{I} \approx op \models \\ & \text{B4\_2: } \mathcal{I} = \text{B4\_2: } \mathcal{I} \bullet \dashv op \approx \dashv op \\ & \text{B5: } (op_1 \parallel op_2) \bullet (op_3 \parallel op_4) \approx (op_1 \bullet op_3) \parallel (op_2 \bullet op_4) \\ & \text{B6: } \mathcal{I} \parallel \mathcal{I} \approx \mathcal{I} \\ & \text{B8: } (\mathcal{X} :: (|i + \theta, \theta + |i, |d) \ op) \approx \text{map\_op} \ \text{id} \ \text{(case\_sum Inr Inl) } \mathcal{I} \\ & \text{B9: } \mathcal{X} \approx \text{map\_op} \ \cap \ \cap \ (\mathcal{X} \parallel \mathcal{I}) \bullet \text{map\_op} \ \text{id} \ \cap \ (\mathcal{I} \parallel \mathcal{X}) \\ & \text{B10: } (\neg op_1 \parallel \neg op_2) \bullet \mathcal{X} \approx \mathcal{X} \bullet (op_2 \vdash \parallel op_1 \Vdash) \\ & \text{F1: } \mathcal{I} \uparrow \approx \langle \mathcal{I} :: (\theta, \theta, |d) \ op) \end{aligned} \end{aligned}
```

```
R1: \operatorname{Inr} : \operatorname{inputs} op_1 \cap \operatorname{defaults} = \{\} \Longrightarrow \operatorname{Inr} : \operatorname{outputs} op_1 \cap \operatorname{defaults} = \{\} \Longrightarrow op_2 \bullet (op_1 \uparrow) \approx ((op_2 \parallel \mathcal{I}) \bullet op_1) \uparrow
R2: \operatorname{Inr} : \operatorname{inputs} op_1 \cap \operatorname{defaults} = \{\} \Longrightarrow \operatorname{Inr} : \operatorname{outputs} op_1 \cap \operatorname{defaults} = \{\} \Longrightarrow (op_1 \uparrow) \bullet op_2 \approx (op_1 \bullet (op_2 \parallel \mathcal{I})) \uparrow
R3: \operatorname{Inr} : \operatorname{inputs} op_2 \cap \operatorname{defaults} = \{\} \Longrightarrow \operatorname{Inr} : \operatorname{outputs} op_2 \cap \operatorname{defaults} = \{\} \Longrightarrow \operatorname{op}_1 \parallel (op_2 \uparrow) \approx (\operatorname{map\_op} \land \land (op_1 \parallel op_2)) \uparrow
R4: \operatorname{Inr} : \operatorname{inputs} op_1 \cap \operatorname{defaults} = \{\} \Longrightarrow \operatorname{inr} : \operatorname{outputs} op_1 \cap \operatorname{defaults} = \{\} \Longrightarrow \operatorname{inputs} op_2 \cap \operatorname{defaults} = \{\} \Longrightarrow (\exists op_1 \bullet (\mathcal{I} \parallel op_2)) \uparrow \approx ((\mathcal{I} \parallel op_2) \bullet op_1 \vdash) \uparrow
R5: \operatorname{Inr} : \operatorname{inputs} op_2 = \{\} \Longrightarrow \operatorname{Inr} : \operatorname{outputs} op_2 \in \{\} \Longrightarrow \operatorname{map\_op} \operatorname{Inl} \operatorname{ln} ((op: :: ('i + \theta, 'o + \theta, 'd) op) \uparrow) \approx op
R6: \operatorname{Inr} : \operatorname{inputs} op_2 = \{\} \Longrightarrow \operatorname{Inr} : \operatorname{outputs} op_2 = \{\} \Longrightarrow \operatorname{Inr} : \operatorname{inputs} op_2 = \{\} \Longrightarrow \operatorname{Inr} : \operatorname{outputs} op_2 = \{\} \Longrightarrow \operatorname{outputs} op_
```

```
B1: op_1 \parallel (op_2 \parallel op_3) \approx \mathsf{map\_op} \land \land (op_1 \parallel op_2) \parallel op_3

B2_1: op \parallel (\mathcal{I} :: (\theta, 0, 'd) op) \approx \mathsf{map\_op} Ind Ind op

B2_2: (\mathcal{I} :: (\theta, 0, 'd) op) \parallel op \approx \mathsf{map\_op} Ind Ind op

B3: (op_1 \bullet op_2) \bullet op_3 \approx op_1 \bullet (op_2 \bullet op_3)

B4_1: op \models \bullet \mathcal{I} \approx op \models

B5: (op_1 \parallel op_2) \bullet (op_3 \parallel op_4) \approx (op_1 \bullet op_3) \parallel (op_2 \bullet op_4)

B6: \mathcal{I} \parallel \mathcal{I} \approx \mathcal{I}

B7: \mathcal{X} \bullet \mathcal{X} \approx \mathcal{I}

B8: (\mathcal{X} :: ('i + \theta, \theta + 'i, 'd) op) \approx \mathsf{map\_op} id (case_sum Inr Inl) \mathcal{I}

B9: \mathcal{X} \approx \mathsf{map\_op} \land (\mathcal{X} \parallel \mathcal{I}) \bullet \mathsf{map\_op} id \land (\mathcal{I} \parallel \mathcal{X})

B10: (-op_1 \parallel -op_2) \bullet \mathcal{X} \approx \mathcal{X} \bullet (op_2 \models \parallel op_1 \models)

F1: \mathcal{I} \uparrow \approx (\mathcal{I} :: (\theta, \theta, 'd) op)
```

Conclusion

Conclusion

- Isabelle/HOL has a good tool set to formalize and reason about stream processing:
 - Codatatypes, coinductive predicates, corecursion with friends, reasoning up to friends (congruence),
 - Coinduction up to congruence principle is automatically derived for codatatypes (but not for coinductive principles)
- Next step: Feedback loop

Questions, comments and suggestions