

Nondeterministic Asynchronous Dataflow in Isabelle/HOL

Rafael Castro G. Silva, Laouen Fernet and Dmitriy Traytel

`razi@di.ku.dk`

Department of Computer Science
University of Copenhagen

14/05/2025

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



Cloud
DataFlow



Flink



- 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)

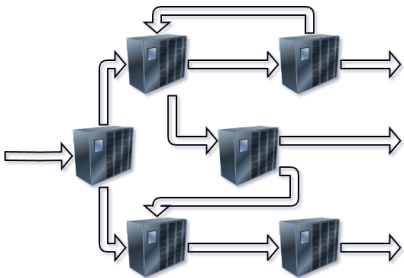
Our goal:

Mechanically Verify Timely Dataflow algorithms

A Good Foundation

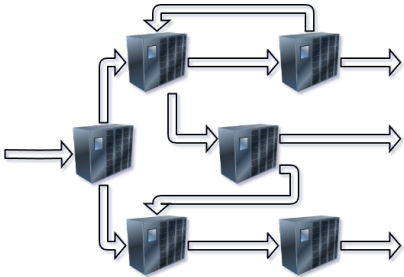
A Good Foundation

- Nondeterministic Asynchronous Dataflow

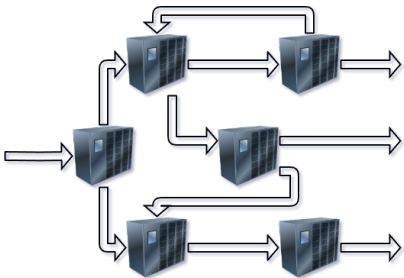


A Good Foundation

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

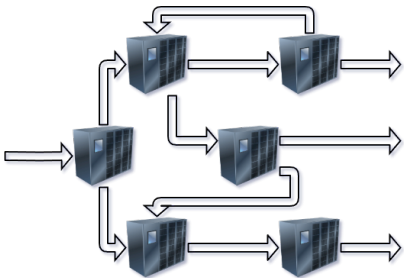


A Good Foundation



- 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

A Good Foundation



- 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. Ştefănescu^{4,‡}

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

- 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

- 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 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

$$\odot \equiv \text{Choice } \{\}_c$$

corec spin_op (\otimes) where

$$\otimes = \text{Choice } ((\lambda_ . \otimes)`_c \{()\}_c)$$

corec silent_op (\odot) where

$$\odot = \text{Silent } \odot$$

lemma spin_op_code:

$$\otimes = \text{Choice } \{\otimes\}_c$$

- They have the same meaning!
- A small quirk: any corecursive call guarded by the Choice constructor must be applied using the map function on *cset*

Operators Equivalences: Motivation

- foo

Operators Equivalences: Strong Bisimilarity

- foo

Operators Equivalences: Weak Bisimilarity

- foo

Asynchronous Dataflow Operators

- foo

Asynchronous Dataflow Properties

Basic network algebra properties

B1: $op_1 \parallel (op_2 \parallel op_3) \approx \text{map_op } \curvearrowright \curvearrowright (op_1 \parallel op_2) \parallel op_3$
 B2_1: $op \parallel (\mathcal{I} :: (0, 0, 'd) \text{ } op) \approx \text{map_op } \text{Inl } \text{Inl } op$
 B2_2: $(\mathcal{I} :: (0, 0, 'd) \text{ } op) \parallel op \approx \text{map_op } \text{Inr } \text{Inr } op$
 B3: $(op_1 \bullet op_2) \bullet op_3 \approx op_1 \bullet (op_2 \bullet op_3)$
 B4_1: $op \sqcap \mathcal{I} \approx op \sqcap$ B4_2: $\mathcal{I} \bullet \sqcap op \approx \sqcap op$
 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: $(\mathcal{X} :: ('i + 0, 0 + 'i, 'd) \text{ } op) \approx \text{map_op } \text{id } (\text{case_sum } \text{Inr } \text{Inl}) \mathcal{I}$
 B9: $\mathcal{X} \approx \text{map_op } \curvearrowright \curvearrowright (\mathcal{X} \parallel \mathcal{I}) \bullet \text{map_op } \text{id } \curvearrowright (\mathcal{I} \parallel \mathcal{X})$
 B10: $(\sqcap op_1 \parallel \sqcap op_2) \bullet \mathcal{X} \approx \mathcal{X} \bullet (op_2 \sqcap \parallel op_1 \sqcap)$
 F1: $\mathcal{I} \uparrow \approx (\mathcal{I} :: (0, 0, 'd) \text{ } op)$ F2: $\mathcal{X} \uparrow \approx \mathcal{I}$

R1: $\text{Inr } \vdash \text{inputs } op_1 \cap \text{defaults} = \{\} \implies \text{Inr } \vdash \text{outputs } op_1 \cap \text{defaults} = \{\} \implies$
 $op_2 \bullet (op_1 \uparrow) \approx ((op_2 \parallel \mathcal{I}) \bullet op_1) \uparrow$
 R2: $\text{Inr } \vdash \text{inputs } op_1 \cap \text{defaults} = \{\} \implies \text{Inr } \vdash \text{outputs } op_1 \cap \text{defaults} = \{\} \implies$
 $(op_1 \uparrow) \bullet op_2 \approx (op_1 \bullet (op_2 \parallel \mathcal{I})) \uparrow$
 R3: $\text{Inr } \vdash \text{inputs } op_2 \cap \text{defaults} = \{\} \implies \text{Inr } \vdash \text{outputs } op_2 \cap \text{defaults} = \{\} \implies$
 $op_1 \parallel (op_2 \uparrow) \approx (\text{map_op } \curvearrowright \curvearrowright (op_1 \parallel op_2)) \uparrow$
 R4: $\text{Inr } \vdash \text{inputs } op_1 \cap \text{defaults} = \{\} \implies \text{Inr } \vdash \text{outputs } op_1 \cap \text{defaults} = \{\} \implies$
 $\text{inputs } op_2 \cap \text{defaults} = \{\} \implies \text{outputs } op_2 \cap \text{defaults} = \{\} \implies$
 $(\sqcap op_1 \bullet (\mathcal{I} \parallel op_2)) \uparrow \approx ((\mathcal{I} \parallel op_2) \bullet op_1 \sqcap) \uparrow$
 R5: $\text{Inr } \vdash \text{inputs } op = \{\} \implies \text{Inr } \vdash \text{outputs } op = \{\} \implies$
 $\text{map_op } \text{Inl } \text{Inl } ((op :: ('i + 0, 'o + 0, 'd) \text{ } op) \uparrow) \approx op$
 R6: $\text{Inr } \vdash \text{inputs } op = \{\} \implies \text{Inr } \vdash \text{outputs } op = \{\} \implies$
 $\text{Inr } \vdash \text{Inl } \vdash \text{inputs } op = \{\} \implies \text{Inr } \vdash \text{Inl } \vdash \text{outputs } op = \{\} \implies$
 $(op \uparrow) \uparrow \approx (\text{map_op } \curvearrowright \curvearrowright op) \uparrow$

■ **Table 1** Basic network algebra properties

Basic network algebra properties

$$B1: op_1 \parallel (op_2 \parallel op_3) \approx \text{map_op } \curvearrowright \curvearrowright (op_1 \parallel op_2) \parallel op_3$$

$$B2_1: op \parallel (\mathcal{I} :: (0, 0, 'd) op) \approx \text{map_op } \text{Inl } \text{Inl } op$$

$$B2_2: (\mathcal{I} :: (0, 0, 'd) op) \parallel op \approx \text{map_op } \text{Inr } \text{Inr } op$$

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

$$B4_1: op \sqcap \bullet \mathcal{I} \approx op \sqcap$$

$$B4_2: \mathcal{I} \bullet \sqcap op \approx \sqcap op$$

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

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

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

$$B8: (\mathcal{X} :: ('i + 0, 0 + 'i, 'd) op) \approx \text{map_op } \text{id } (\text{case_sum } \text{Inr } \text{Inl}) \mathcal{I}$$

$$B9: \mathcal{X} \approx \text{map_op } \curvearrowright \curvearrowright (\mathcal{X} \parallel \mathcal{I}) \bullet \text{map_op } \text{id } \curvearrowright (\mathcal{I} \parallel \mathcal{X})$$

$$B10: (\sqcap op_1 \parallel \sqcap op_2) \bullet \mathcal{X} \approx \mathcal{X} \bullet (op_2 \sqcap \parallel op_1 \sqcap)$$

$$F1: \mathcal{I} \uparrow \approx (\mathcal{I} :: (0, 0, 'd) op)$$

$$F2: \mathcal{X} \uparrow \approx \mathcal{I}$$

$$R1: \text{Inr } \dot{\vdash} \text{ inputs } op_1 \cap \text{ defaults } = \{\} \implies \text{Inr } \dot{\vdash} \text{ outputs } op_1 \cap \text{ defaults } = \{\} \implies op_2 \bullet (op_1 \uparrow) \approx ((op_2 \parallel \mathcal{I}) \bullet op_1) \uparrow$$

$$R2: \text{Inr } \dot{\vdash} \text{ inputs } op_1 \cap \text{ defaults } = \{\} \implies \text{Inr } \dot{\vdash} \text{ outputs } op_1 \cap \text{ defaults } = \{\} \implies (op_1 \uparrow) \bullet op_2 \approx (op_1 \bullet (op_2 \parallel \mathcal{I})) \uparrow$$

$$R3: \text{Inr } \dot{\vdash} \text{ inputs } op_2 \cap \text{ defaults } = \{\} \implies \text{Inr } \dot{\vdash} \text{ outputs } op_2 \cap \text{ defaults } = \{\} \implies op_1 \parallel (op_2 \uparrow) \approx (\text{map_op } \curvearrowright \curvearrowright (op_1 \parallel op_2)) \uparrow$$

$$R4: \text{Inr } \dot{\vdash} \text{ inputs } op_1 \cap \text{ defaults } = \{\} \implies \text{Inr } \dot{\vdash} \text{ outputs } op_1 \cap \text{ defaults } = \{\} \implies \text{inputs } op_2 \cap \text{ defaults } = \{\} \implies \text{outputs } op_2 \cap \text{ defaults } = \{\} \implies (\sqcap op_1 \bullet (\mathcal{I} \parallel op_2)) \uparrow \approx ((\mathcal{I} \parallel op_2) \bullet op_1 \sqcap) \uparrow$$

$$R5: \text{Inr } \dot{\vdash} \text{ inputs } op = \{\} \implies \text{Inr } \dot{\vdash} \text{ outputs } op = \{\} \implies \text{map_op } \text{Inl } \text{Inl } ((op :: ('i + 0, 'o + 0, 'd) op) \uparrow) \approx op$$

$$R6: \text{Inr } \dot{\vdash} \text{ inputs } op = \{\} \implies \text{Inr } \dot{\vdash} \text{ outputs } op = \{\} \implies \text{Inr } \dot{\vdash} \text{Inl } \dot{\vdash} \text{ inputs } op = \{\} \implies \text{Inr } \dot{\vdash} \text{Inl } \dot{\vdash} \text{ outputs } op = \{\} \implies (op \uparrow) \uparrow \approx (\text{map_op } \curvearrowright \curvearrowright op) \uparrow$$

Table 1 Basic network algebra properties

Basic network algebra properties

$$B1: op_1 \parallel (op_2 \parallel op_3) \approx \text{map_op } \curvearrowright \curvearrowright (op_1 \parallel op_2) \parallel op_3$$

$$B2_1: op \parallel (\mathcal{I} :: (0, 0, 'd) op) \approx \text{map_op } \text{Inl } \text{Inl } op$$

$$B2_2: (\mathcal{I} :: (0, 0, 'd) op) \parallel op \approx \text{map_op } \text{Inr } \text{Inr } op$$

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

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

$$B4_2: \mathcal{I} \bullet op \approx op$$

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

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

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

$$B8: (\mathcal{X} :: ('i + 0, 0 + 'i, 'd) op) \approx \text{map_op } \text{id } (\text{case_sum } \text{Inr } \text{Inl}) \mathcal{I}$$

$$B9: \mathcal{X} \approx \text{map_op } \curvearrowright \curvearrowright (\mathcal{X} \parallel \mathcal{I}) \bullet \text{map_op } \text{id } \curvearrowright (\mathcal{I} \parallel \mathcal{X})$$

$$B10: (\dashv op_1 \parallel \dashv op_2) \bullet \mathcal{X} \approx \mathcal{X} \bullet (op_2 \dashv op_1)$$

$$F1: \mathcal{I} \uparrow \approx (\mathcal{I} :: (0, 0, 'd) op)$$

$$F2: \mathcal{X} \uparrow \approx \mathcal{I}$$

$$R1: \text{Inr } \vdash \text{inputs } op_1 \cap \text{defaults} = \{\} \implies \text{Inr } \vdash \text{outputs } op_1 \cap \text{defaults} = \{\} \implies op_2 \bullet (op_1 \uparrow) \approx ((op_2 \parallel \mathcal{I}) \bullet op_1) \uparrow$$

$$R2: \text{Inr } \vdash \text{inputs } op_1 \cap \text{defaults} = \{\} \implies \text{Inr } \vdash \text{outputs } op_1 \cap \text{defaults} = \{\} \implies (op_1 \uparrow) \bullet op_2 \approx (op_1 \bullet (op_2 \parallel \mathcal{I})) \uparrow$$

$$R3: \text{Inr } \vdash \text{inputs } op_2 \cap \text{defaults} = \{\} \implies \text{Inr } \vdash \text{outputs } op_2 \cap \text{defaults} = \{\} \implies op_1 \parallel (op_2 \uparrow) \approx (\text{map_op } \curvearrowright \curvearrowright (op_1 \parallel op_2)) \uparrow$$

$$R4: \text{Inr } \vdash \text{inputs } op_1 \cap \text{defaults} = \{\} \implies \text{Inr } \vdash \text{outputs } op_1 \cap \text{defaults} = \{\} \implies \text{inputs } op_2 \cap \text{defaults} = \{\} \implies \text{outputs } op_2 \cap \text{defaults} = \{\} \implies (\dashv op_1 \bullet (\mathcal{I} \parallel op_2)) \uparrow \approx ((\mathcal{I} \parallel op_2) \bullet op_1 \dashv) \uparrow$$

$$R5: \text{Inr } \vdash \text{inputs } op = \{\} \implies \text{Inr } \vdash \text{outputs } op = \{\} \implies \text{map_op } \text{Inl } \text{Inl } ((op :: ('i + 0, 'o + 0, 'd) op) \uparrow) \approx op$$

$$R6: \text{Inr } \vdash \text{inputs } op = \{\} \implies \text{Inr } \vdash \text{outputs } op = \{\} \implies \text{Inr } \vdash \text{Inl } \vdash \text{inputs } op = \{\} \implies \text{Inr } \vdash \text{Inl } \vdash \text{outputs } op = \{\} \implies (op \uparrow) \uparrow \approx (\text{map_op } \curvearrowright \curvearrowright op) \uparrow$$

■ **Table 1** Basic network algebra properties

Basic network algebra properties

$$\text{B1: } op_1 \parallel (op_2 \parallel op_3) \approx \text{map_op} \curvearrowright \curvearrowright (op_1 \parallel op_2) \parallel op_3$$
$$\text{B2_1: } op \parallel (\mathcal{I} :: (\theta, \theta, 'd) \text{ } op) \approx \text{map_op } \text{Inl } \text{Inl } op$$
$$\text{B2_2: } (\mathcal{I} :: (\theta, \theta, 'd) \text{ op}) \parallel \text{op} \approx \text{map_op lnr lnr op}$$
$$\text{B3: } (op_1 \bullet op_2) \bullet op_3 \approx op_1 \bullet (op_2 \bullet op_3)$$
$$\text{B4_1: } op \vdash \bullet \mathcal{I} \approx op \vdash \quad \text{B4_2: } \mathcal{I} \bullet \dashv op \approx \dashv op$$
$$\text{B5: } (op_1 \parallel op_2) \bullet (op_3 \parallel op_4) \approx (op_1 \bullet op_3) \parallel (op_2 \bullet op_4)$$

B6: $\mathcal{I} \parallel \mathcal{I} \approx \mathcal{I}$
B7: $\mathcal{X} \bullet \mathcal{X} \approx \mathcal{I}$

$$\text{B8: } (\mathcal{X} :: ('i + \theta, \theta + 'i, 'd) \text{ op}) \approx \text{map_op id (case_sum lnr lnl)} \mathcal{I}$$
$$\text{B9: } \mathcal{X} \approx \text{map_op} \circ \circ (\mathcal{X} \parallel \mathcal{I}) \bullet \text{map_op id} \circ (\mathcal{I} \parallel \mathcal{X})$$
$$\text{B10: } (\dashv op_1 \parallel \dashv op_2) \bullet \mathcal{X} \approx \mathcal{X} \bullet (op_2 \vdash \parallel op_1 \vdash)$$
$$\text{F1: } \mathcal{I} \uparrow \approx (\mathcal{I} :: (\theta, \theta, 'd) \text{ op}) \qquad \text{F2: } \mathcal{X} \uparrow \approx \mathcal{I}$$
$$\text{R1: } \text{Inr } \dot{\vdash} \text{ inputs } op_1 \cap \text{ defaults} = \{\} \implies \text{Inr } \dot{\vdash} \text{ outputs } op_1 \cap \text{ defaults} = \{\} \implies op_2 \bullet (op_1 \uparrow) \approx ((op_2 \parallel \mathcal{I}) \bullet op_1) \uparrow$$
$$\text{R2: } \text{Inr } \dot{\vdash} \text{ inputs } op_1 \cap \text{defaults} = \{\} \implies \text{Inr } \dot{\vdash} \text{ outputs } op_1 \cap \text{defaults} = \{\} \implies (op_1 \uparrow) \bullet op_2 \approx (op_1 \bullet (op_2 \parallel \mathcal{I})) \uparrow$$
$$\text{R3: } \text{Inr} \vdash \text{inputs } op_2 \cap \text{defaults} = \{\} \implies \text{Inr} \vdash \text{outputs } op_2 \cap \text{defaults} = \{\} \implies op_1 \parallel (op_2 \uparrow) \approx (\text{map_op } \hookrightarrow \hookrightarrow (op_1 \parallel op_2)) \uparrow$$
$$\begin{aligned} \text{R4: } \text{Inr} \vdash \text{inputs } op_1 \cap \text{defaults} = \{\} &\implies \text{Inr} \vdash \text{outputs } op_1 \cap \text{defaults} = \{\} \implies \\ \text{inputs } op_2 \cap \text{defaults} = \{\} &\implies \text{outputs } op_2 \cap \text{defaults} = \{\} \implies \\ (\neg op_1 \bullet (\mathcal{I} \parallel op_2)) \uparrow \approx ((\mathcal{I} \parallel op_2) \bullet op_1) \uparrow & \end{aligned}$$
$$\text{R5: } \text{Inr } \dot{\vdash} \text{ inputs } op = \{\} \implies \text{Inr } \dot{\vdash} \text{ outputs } op = \{\} \implies \text{map_op } \text{Inl } \text{Inl } ((op :: ('i + \theta, 'o + \theta, 'd) \text{ op}) \uparrow) \approx op$$
$$\begin{aligned} \text{R6: } \text{Inr } \dot{\vdash} \text{ inputs } op = \{\} &\implies \text{Inr } \dot{\vdash} \text{ outputs } op = \{\} \implies \\ \text{Inr } \dot{\vdash} \text{Inl } \dot{\vdash} \text{ inputs } op = \{\} &\implies \text{Inr } \dot{\vdash} \text{Inl } \dot{\vdash} \text{ outputs } op = \{\} \implies \\ (op \uparrow) \uparrow \approx (\text{map_op } \circ \circ \text{ op}) \uparrow \end{aligned}$$

■ **Table 1** Basic network algebra properties

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