
Upper Bounds on Stream I/O Using Semantic Interpretations

Marco Gaboardi^{*} and Romain Péchoux[†]

^{*} Dipartimento di Informatica - Università di Torino - gaboardi@di.unito.it

[†] Université Nancy 2 - Nancy Université - LORIA - romain.pechoux@loria.fr

Work partially supported by the French COMPLICE project
and by the Italian CONCERTO project.

▷ [Introduction](#)

[Preliminaries](#)

[Length Based I/O
Upper Bound](#)

[Size Based I/O Upper
Bound](#)

[Synchrony Upper
Bound](#)

Introduction

Setting and Motivations

- We deal with the problems of program **resource control** and **resource usage certification**.
- We are interested into studying such properties about programs working on **infinite data structures**, notably **streams**.
- **Lazy Functional Programming Languages** enable an elegant treatment of stream programs.
- Here, in order to prove properties of **Lazy Functional programs** dealing with **streams**, we apply **static analysis** methods inspired by Quasi-interpretation by Bonfante et al[2001] and Sup-interpretation by Marion and Péchoux[2006].
- In particular we give some **criteria** ensuring three distinct **Input-Output upper bound properties**.

Streams

- Streams are **infinite lists**

$0 : 1 : 2 : 3 : 4 : 5 : 6 : \dots : 65512 : 65513 \dots$

$1 : 1 : 2 : 3 : 5 : 8 : 13 : \dots : 46368 : 75025 : \dots$

$(0, 0) : (0, 1) : (1, 0) : (2, 0) : (1, 1) : (0, 2) : \dots : (64, 0) : (63, 1) : \dots$

- **Lazy functional programming languages** like Haskell allow us to easily define streams by **expressions**:

```
from x = x : (from (x + 1))
```

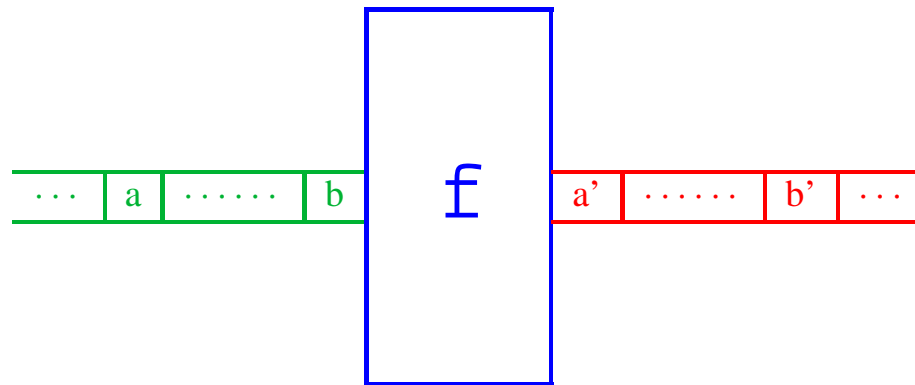
```
fibonacci = 1 : 1 : (sadd fibonacci (tail fibonacci))
```

Stream properties

- In the last years, many efforts have been made in the study of:
 - **stream programs equivalence** by means of denotational and operational semantics
 - **stream productivity** by means of several techniques: syntactical restrictions, type systems, data flow analysis
- Unfortunately only little efforts have been made in order to study **other stream properties**.
- Here we deal with some **complexity properties** of stream programs:
 - **relating** the **number of output writings** to the **number of input readings**.
 - **relating** the **number of output writings** to the **size of input reading elements**.
 - **relating synchronously** the **size of output writing elements** to the **size of input reading elements**.

Stream programs

- It is convenient to distinguish between different kinds of programs dealing with stream:
 - programs **generating** streams
 - programs **working on input** streams
- In this work we focus on **programs working on input streams**, we leave the study of **programs generating streams** for future developments.
- In particular, we look at programs as



Introduction

[▶ Preliminaries](#)

Length Based I/O
Upper Bound

Size Based I/O Upper
Bound

Synchrony Upper
Bound

Preliminaries

The sHask language

- A program P is a set of definitions described by the grammar:

p	$::= x \mid \mathbf{c} p_1 \cdots p_n$	(Patterns)
e	$::= x \mid \mathbf{t} e_1 \cdots e_n \mid \text{Case } \bar{e} \text{ of } \bar{p}_1 \rightarrow e_1, \dots, \bar{p}_m \rightarrow e_m$	(Expressions)
v	$::= \mathbf{c} e_1 \cdots e_n$	(Values)
d	$::= \mathbf{f} x_1 \cdots x_n = e$	(Definitions)

- There is a special **error constructor symbol \mathbf{Err}** of arity 0 which corresponds to pattern matching failure. We also allow **pattern matching on it**.
- As syntactic sugar we write a definition of the shape
 $\mathbf{f} \vec{x} = \text{Case } \bar{x} \text{ of } \bar{p}_1 \rightarrow e_1, \dots, \bar{p}_k \rightarrow e_k$ as a set of definitions

$$\begin{array}{lcl}
 \mathbf{f} \vec{p}_1 & = & e_1 \\
 & \vdots & \\
 \mathbf{f} \vec{p}_k & = & e_k
 \end{array}$$

Type system

- The set of **types** contains elements of the shape

$$A_1 \rightarrow (\dots \rightarrow (A_n \rightarrow A))$$

where **basic** and **value** types are defined as

$$\begin{aligned} \sigma &::= \alpha \mid \text{Nat} \mid \sigma \times \sigma \\ A &::= a \mid \sigma \mid A \times A \mid [\sigma] \end{aligned}$$

- The typing rules are

$$\frac{}{x :: A} \text{ (Var)} \qquad \frac{\bar{e} :: \bar{A} \quad \bar{p}_1 :: \bar{A} \quad \dots \quad \bar{p}_m :: \bar{A} \quad e_1 :: A \quad \dots \quad e_m :: A}{\text{Case } \bar{e} \text{ of } \bar{p}_1 \rightarrow e_1, \dots, \bar{p}_m \rightarrow e_m :: A} \text{ (Case)}$$

$$\frac{}{t :: A_1 \rightarrow \dots \rightarrow A_n \rightarrow A} \text{ (Tb)} \qquad \frac{t :: A_1 \rightarrow \dots \rightarrow A_n \rightarrow A \quad e_1 :: A_1 \quad \dots \quad e_n :: A_n}{t \ e_1 \ \dots \ e_n :: A} \text{ (Ts)}$$

- **Remarks:**

- **Err** :: A for every A.
- **Streams of streams** are forbidden.

Lazy Operational Semantics

- A natural way to deal with infinite data structure in a finitary way is by using a **lazy evaluation** \Downarrow .

$$\frac{c \in \mathcal{C}}{c e_1 \cdots e_n \Downarrow c e_1 \cdots e_n} \text{ (val)} \quad \frac{e\{e_1/x_1, \dots, e_n/x_n\} \Downarrow v \quad f x_1 \cdots x_n = e}{f e_1 \cdots e_n \Downarrow v} \text{ (fun)}$$

$$\frac{\text{Case } e^1 \text{ of } p_1^1 \rightarrow \dots \rightarrow \text{Case } e^m \text{ of } p_1^m \rightarrow d_1 \Downarrow v \quad v \neq \mathbf{Err}}{\text{Case } \bar{e} \text{ of } \bar{p}_1 \rightarrow d_1, \dots, \bar{p}_n \rightarrow d_n \Downarrow v} \text{ (c}_b\text{)}$$

$$\frac{\text{Case } e^1 \text{ of } p_1^1 \rightarrow \dots \rightarrow \text{Case } e^m \text{ of } p_1^m \rightarrow d_1 \Downarrow \mathbf{Err} \quad \text{Case } \bar{e} \text{ of } \bar{p}_2 \rightarrow d_2, \dots, \bar{p}_n \rightarrow d_n \Downarrow v}{\text{Case } \bar{e} \text{ of } \bar{p}_1 \rightarrow d_1, \dots, \bar{p}_n \rightarrow d_n \Downarrow v} \text{ (c)}$$

$$\frac{e \Downarrow c e_1 \cdots e_n \quad \text{Case } e_1 \text{ of } p_1 \rightarrow \dots \rightarrow \text{Case } e_n \text{ of } p_n \rightarrow d \Downarrow v}{\text{Case } e \text{ of } c p_1 \cdots p_n \rightarrow d \Downarrow v} \text{ (pm)}$$

$$\frac{e \Downarrow v \quad v \neq c e_1 \cdots e_n}{\text{Case } e \text{ of } c p_1 \cdots p_n \rightarrow d \Downarrow \mathbf{Err}} \text{ (pm}_e\text{)} \quad \frac{e'\{e/x\} \Downarrow v}{\text{Case } e \text{ of } x \rightarrow e' \Downarrow v} \text{ (pm}_b\text{)}$$

- For simplicity we do not consider **sharing**.

Some useful programs

In the sequel we need some **programs**:

$$\begin{array}{ll} \text{take} :: \text{Nat} \rightarrow [\alpha] \rightarrow [\alpha] & \text{lg} :: [\alpha] \rightarrow \text{Nat} \\ \text{take } 0 \quad s & = \quad \text{nil} & \text{lg } \text{nil} & = & \underline{0} \\ \text{take } (x + 1) \quad \text{nil} & = & \text{nil} & \text{lg } \mathbf{Err} & = & \underline{0} \\ \text{take } (x + 1) \quad (y : \text{ys}) & = & y : (\text{take } x \text{ ys}) & \text{lg } (x : \text{xs}) & = & (\text{lg } \text{xs}) + 1 \end{array}$$
$$\begin{array}{ll} \text{!!} :: [\alpha] \rightarrow \text{Nat} \rightarrow \alpha & \text{lInd} :: \text{Nat} \rightarrow [\alpha] \rightarrow [\alpha] \\ (x : \text{xs}) \quad \text{!!} \quad 0 & = \quad x & \text{lInd } x \quad \text{xs} & = & (\text{xs} \text{ !! } x) : \text{nil} \\ (x : \text{xs}) \quad \text{!!} \quad (y + 1) & = & \text{xs} \text{ !! } y \end{array}$$

Moreover we need a program

$$\begin{array}{l} \text{eval} :: A \rightarrow A \\ \text{eval } (\mathbf{c} \ e_1 \ \cdots \ e_n) = \hat{\mathbf{C}} (\text{eval } e_1) \ \cdots \ (\text{eval } e_n) \end{array}$$

where $\hat{\mathbf{C}}$ is a program representing the *strict version* of the primitive constructor \mathbf{c} .

Interpretation

Definition [Interpretation] A program P admits an **interpretation** if there is an **assignment** $\langle - \rangle$:

- total: $\forall \mathbf{t}, \langle \mathbf{t} \rangle : (\mathbb{R}^+)^n \rightarrow \mathbb{R}^+$
- monotonic: $X_i \geq Y_i \Rightarrow \langle \mathbf{t} \rangle(\dots, X_i, \dots) \geq \langle \mathbf{t} \rangle(\dots, Y_i, \dots)$,

such that for each definition in P of the shape $\mathbf{f} \overrightarrow{\mathbf{p}} = \mathbf{e}$:

$$\langle \mathbf{f} \overrightarrow{\mathbf{p}} \rangle \geq \langle \mathbf{e} \rangle$$

A program P admits an **additive interpretation** if $\langle - \rangle$ is such that for every symbol \mathbf{c} of arity n :

- $\langle \mathbf{c} \rangle = 0$ if $n = 0$
- $\langle \mathbf{c} \rangle(X_1, \dots, X_n) = \sum_{i=1}^n X_i + \alpha_{\mathbf{c}}$, with $\alpha_{\mathbf{c}} \geq 1$ otherwise

Properties

Lemma [Interpretation respects evaluation] Given a program P admitting the interpretation $\llbracket - \rrbracket$, for every closed expression e we have

- if $e \Downarrow v$ then $\llbracket e \rrbracket \geq \llbracket v \rrbracket$
- if $\text{eval } e \Downarrow v$ then $\llbracket e \rrbracket \geq \llbracket v \rrbracket$

Lemma [Interpretation vs. expression size] Given a program P admitting the interpretation $\llbracket - \rrbracket$, there is a function $G : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ such that for each expression e :

$$\llbracket e \rrbracket \leq G(|e|)$$

Introduction

Preliminaries

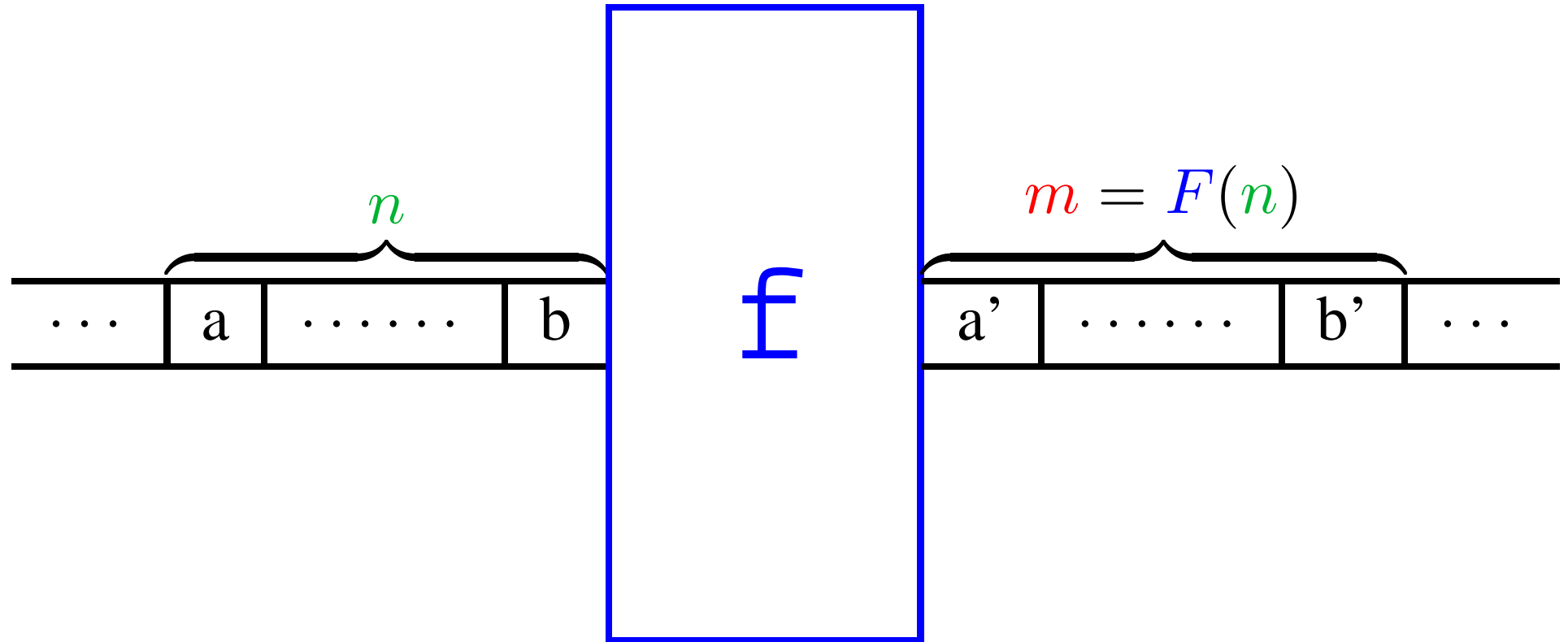
Length Based I/O
▷ Upper Bound

Size Based I/O Upper
Bound

Synchrony Upper
Bound

Length Based I/O Upper Bound

Length Based I/O Upper Bound - Intuition



Length Based I/O Upper Bound - Formally

□ Definition

A stream function $f :: [\sigma'] \rightarrow \tau \rightarrow [\sigma]$ has a **length based I/O upper bound** if there is $F : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ such that $\forall s :: [\sigma'], \forall e :: \tau$:

$$\forall \underline{n} \in \mathbb{N}, \text{ s.t. } \text{eval}(\text{lg}(f(\text{take } \underline{n} \text{ s}) e)) \Downarrow \underline{m}, F(\max(|\underline{n}|, |e|)) \geq |\underline{m}|$$

□ Examples

$\text{double} :: [\text{Nat}] \rightarrow [\text{Nat}]$

$\text{double } (x : xs) = x : (\text{double } xs)$

$\text{merge} :: [\alpha] \rightarrow [\alpha] \rightarrow [\alpha \times \alpha]$

$\text{merge } (x : xs) (y : ys) = (x, y) : (\text{merge } xs \text{ } ys)$

LBUB Criterion

□ Definition

A program is **LBUB** if it admits an **interpretation** $(|-)$ which satisfies $(+1)(X) = X + 1$ and which is **additive** but on the constructor symbol : where $(:)$ is defined as

$$(:)(X, Y) = Y + 1$$

□ Lemma

Given a **LBUB program** wrt the interpretation $(|-)$:

- for every $\underline{n} :: \text{Nat}$ we have $(\underline{n}) = |\underline{n}|$.
- the interpretation can be extended to the program `!g` by $(!g)(X) = X$ and to the program `take` by $(\text{take})(N, L) = N$

□ Theorem

If a program is **LBUB** then each stream function in it **has a length based I/O upper bound**.

LBUB Examples

The program `double` is LBUB, it admits the additive interpretation

$$\langle \text{double} \rangle (X) = 2X \quad \langle \cdot \rangle (X, Y) = Y + 1$$

We have:

$$\begin{aligned} \langle \text{double } (x : xs) \rangle &= \langle \text{double} \rangle \langle (x : xs) \rangle \\ &= 2 \langle (x : xs) \rangle \\ &= 2(\langle xs \rangle + 1) \\ &= \langle \text{double } xs \rangle + 2 \\ &= \langle x : (x : (\text{double } xs)) \rangle \end{aligned}$$

By defining $F(X) = \langle \text{double} \rangle (X)$ we have that if

$$\text{eval}(\text{lg}(\text{double } (\text{take } \underline{n} s_1))) \Downarrow \underline{m}$$

then $F(\underline{n}) = 2n \geq \underline{m}$.

Introduction

Preliminaries

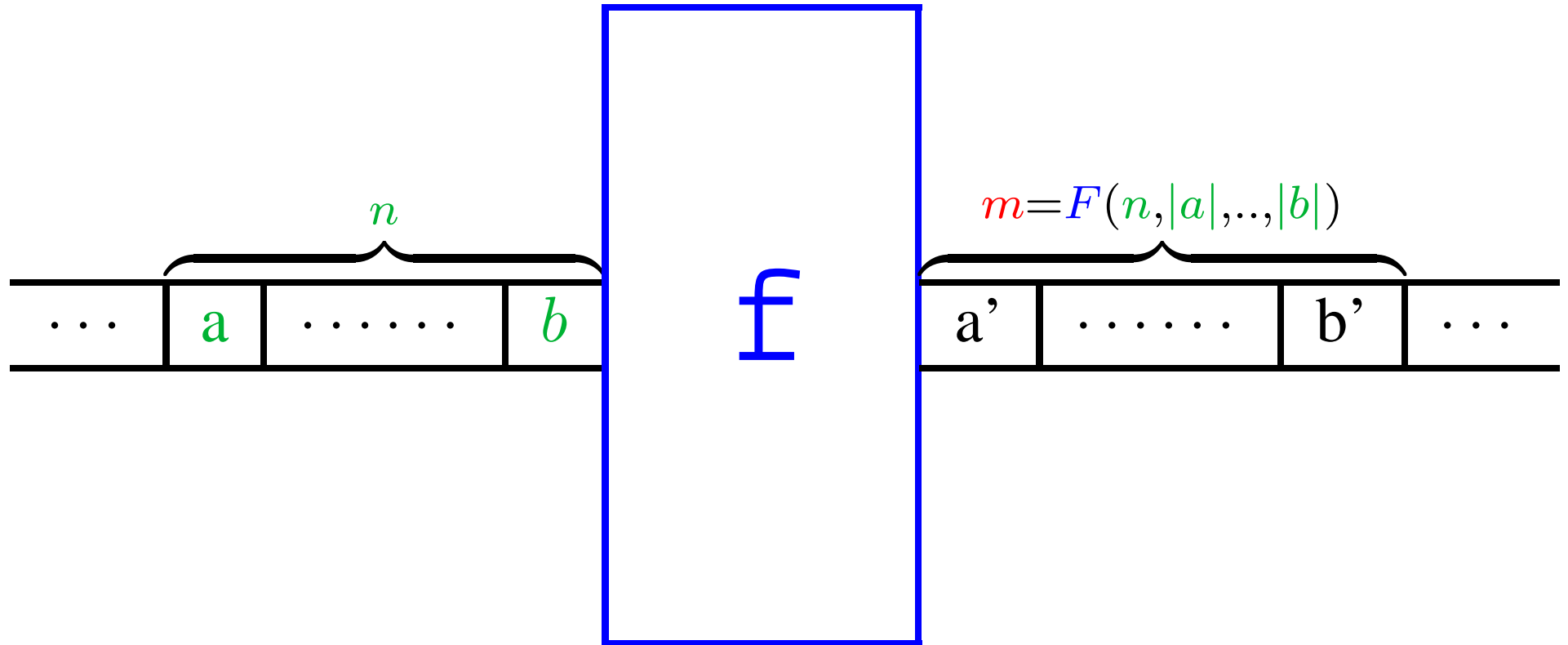
Length Based I/O
Upper Bound

Size Based I/O
▶ Upper Bound

Synchrony Upper
Bound

Size Based I/O Upper Bound

Size Based I/O Upper Bound - Intuition



Size Based I/O Upper Bound - Formally

- **Definition** A stream function $f :: [\sigma'] \rightarrow \tau \rightarrow [\sigma]$ has a **size based I/O upper bound** if there is $F : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ such that $\forall s :: [\sigma'], \forall e :: \tau$:

$$\forall \underline{n}_i \in \mathbb{N}, \text{ s.t. } \text{eval}(\text{lg}(f(\text{take } \underline{n} \ s) \ e)) \Downarrow \underline{m}, F(\max(|\underline{n}|, |s|, |e|)) \geq |\underline{m}|$$

- **Example**

$\text{app} :: [\alpha] \rightarrow [\alpha] \rightarrow [\alpha]$	$\text{upto} :: \text{Nat} \rightarrow [\text{Nat}]$
$\text{app } (x : xs) \ ys = x : (\text{app } xs \ ys)$	$\text{upto } 0 = \text{nil}$
$\text{app } \text{nil} \ ys = ys$	$\text{upto } (x + 1) = (x + 1) : (\text{upto } x)$

$\text{extendupto} :: [\text{Nat}] \rightarrow [\text{Nat}]$
 $\text{extendupto } (x : xs) = \text{app } (\text{upto } x) (\text{extendupto } xs)$

Note that `extendupto` has **not a length based I/O upper bound**, nevertheless it has a **size based I/O upper bound**.

SBUB Criterion

□ Definition

A program is **SBUB** if it admits an **additive interpretation** $(|-|)$ satisfying

$$(|+1|)(X) = X + 1 \text{ and } (|:|)(X, Y) = X + Y + 1$$

□ Lemma

Given a **SBUB program** wrt the interpretation $(|-|)$:

- for every $\underline{n} :: \text{Nat}$ we have $(|\underline{n}|) = |\underline{n}|$.
- the interpretation can be extended to the program !g by $(|\text{!g}|)(X) = X$ and to the program take by $(|\text{take}|)(N, L) = L$

□ Theorem

If a program is **SBUB** then each stream function in it **has a size based I/O upper bound**.

SBUB Example

The program `extendupto` is SBUB, it admits the additive interpretation

$$\begin{aligned}(\text{upto})(X) &= (\text{extendupto})(X) = 2X^2 & (\text{app})(X, Y) &= X + Y \\ (\text{:})(X, Y) &= X + Y + 1\end{aligned}$$

We have:

$$\begin{aligned}(\text{extendupto } (x : xs)) &= (\text{extendupto}) (\text{|x : xs|}) \\ &= 2(\text{|x : xs|})^2 \\ &= 2(\text{|x|} + \text{|xs|} + 1)^2 \\ &= 2(\text{|x|})^2 + 2(\text{|xs|})^2 \\ &= (\text{app } (\text{upto } x) (\text{extendupto } xs))\end{aligned}$$

By defining $F(X) = (\text{extendupto})(X)$ we have that if

$$\text{eval}(\text{lg}(\text{extendupto } (\text{take } \underline{n} \text{ } \underline{s}))) \Downarrow \underline{m}$$

then $F(\max(\underline{n}, \underline{s})) \geq \underline{m}$.

Introduction

Preliminaries

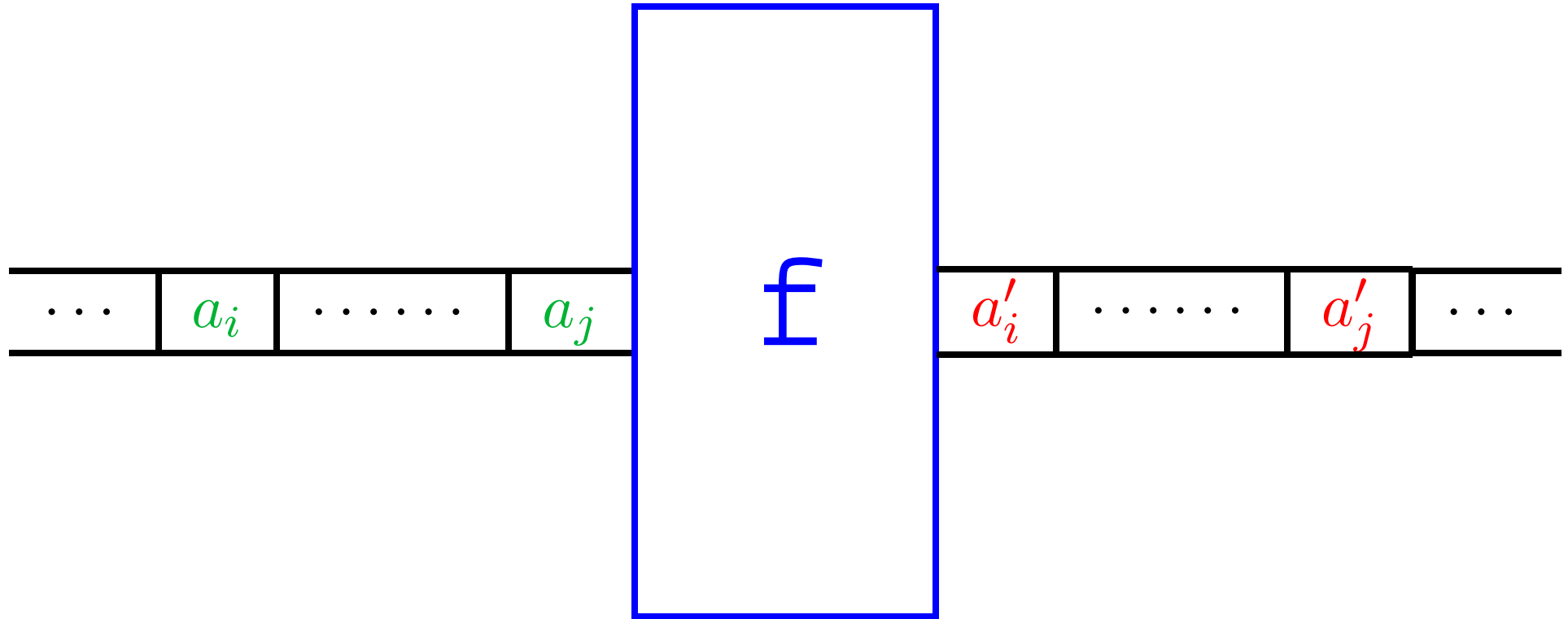
Length Based I/O
Upper Bound

Size Based I/O Upper
Bound

Synchrony Upper
▷ Bound

Synchrony Upper Bound

Synchrony Upper Bound - Intuition



where $|a'_i| = F(|a_i|), \dots, |a'_j| = F(|a_j|)$.

Synchrony Upper Bound - Formally

- **Definition** A stream function $f :: [\sigma'] \rightarrow \tau \rightarrow [\sigma]$ is said to be “Read, Write” if $\forall s :: [\sigma'], \forall e :: \tau$ and $\forall \underline{n} \in \mathbb{N}$:

If $\text{eval}(\text{lInd } \underline{n} (f \ s \ e)) \Downarrow v$ and $\text{eval}(f(\text{lInd } \underline{n} \ s) \ e) \Downarrow v'$ then $v = v'$

- **Definition** A “Read, Write” function $f :: [\sigma'] \rightarrow \tau \rightarrow [\sigma]$ has a **synchrony upper bound** if there is $F : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ such that $\forall s :: [\sigma'], \forall e :: \tau$ and $\underline{n} \in \mathbb{N}$:

If $\text{eval}(s \ !! \ \underline{n}) \Downarrow w$ and $\text{eval}((f \ s \ e) \ !! \ \underline{n}) \Downarrow v$ then $F(\max(|w|, |e|)) \geq |v|$

- **Example**

$\text{sadd} :: [\text{Nat}] \rightarrow [\text{Nat}] \rightarrow [\text{Nat}]$

$\text{sadd} \ (x : xs) \ (y : ys) = (\text{add } x \ y) : (\text{sadd } xs \ ys)$

$\text{add} :: \text{Nat} \rightarrow \text{Nat} \rightarrow \text{Nat}$

$\text{add} \ (x + 1) \ (y + 1) = ((\text{add } x \ y) + 1) + 1$

$\text{add} \ (x + 1) \ 0 = x + 1$

$\text{add} \ 0 \ (y + 1) = y + 1$

SUB Criterion

- **Definition** A stream function $f :: [\sigma'] \rightarrow \tau \rightarrow [\sigma]$ is **synchronously restricted** if it can be written (and maybe extended) by definitions of the shape:

$$\begin{aligned} f \quad (x : \mathbf{xs}) \quad \overrightarrow{p} &= \mathbf{hd} : (f \ \mathbf{xs} \ \overrightarrow{p}) \\ f \quad \mathbf{nil} \quad \overrightarrow{p} &= \mathbf{nil} \end{aligned}$$

where **xs** do not appear in the expression **hd**.

- **Lemma**

Every **synchronously restricted** function is “**Read, Write**”.

- **Definition** A program is **SUB** if it is **synchronously restricted** and it admits an **additive interpretation** $\langle - \rangle$ but on $:$ where

$$\langle : \rangle (X, Y) = X$$

- **Theorem**

If a program is **SUB** then each stream function in it **admits a synchrony upper bound**.

SUB Example

The program `sadd` is SUB because it is **synchronously restricted** and it admits the additive interpretation

$$\langle \text{add} \rangle (X, Y) = X + Y \quad \langle \text{sadd} \rangle (X, Y) = X + Y \quad \langle \cdot \rangle (X, Y) = X$$

We have:

$$\begin{aligned} \langle \text{sadd } (x : \text{xs}) (y : \text{ys}) \rangle &= \langle \text{sadd} \rangle \langle x : \text{xs} \rangle \langle y : \text{ys} \rangle \\ &= \langle x : \text{xs} \rangle + \langle y : \text{ys} \rangle \\ &= \langle x \rangle + \langle y \rangle \\ &= \langle \text{add } x \ y \rangle \\ &= \langle (\text{add } x \ y) : (\text{sadd } \text{xs} \ \text{ys}) \rangle \end{aligned}$$

By defining $F(X) = \langle \text{sadd} \rangle (X, X)$ we have that if

$\text{eval } \langle (x : \text{xs})!!\underline{n} \rangle \Downarrow w$, $\text{eval } \langle (y : \text{ys})!!\underline{n} \rangle \Downarrow w'$ and $\text{eval } \langle (\text{sadd } (x : \text{xs}) (y : \text{ys}))!!\underline{n} \rangle \Downarrow v$

then $F(\max(|w|, |w'|)) \geq |v|$.