
An Introduction to Data Refinement

Formal Methods II, 2002/03

J.N. Oliveira

FM software design process

- **Formal specification** — “what” the intended software system should do
- **Implementation** — machine code produced instructing the hardware about “how” to do it

In general, there is more than one way in which a particular machine can accomplish “what” the specifier bore in mind:

- Relationship between specifications and implementations is **one-to-many**
- Specifications are more **abstract** than implementations.

Data refinement

Principle of **data abstraction**: A abstracts B wherever

- A surjective abstraction function $A \xleftarrow{F} B$ can be found:

$$\text{img } F = \text{id} \quad (1)$$

F is thus **simple** but possibly partial.

- Any **entire** subrelation R of F° is said to be a **representation** for F . So $R \subseteq F^\circ$.

Representation relations

- It follows that R is **injective**, since $\ker R \subseteq \ker F^\circ$ and $\ker F^\circ = \text{img } F = id$.
- So, no two different abstract values $a, a' \in A$ get mixed up along the representation process.
- Altogether, $\ker R = id$ because $id \subseteq \ker R$ (R is entire).
- It also follows that R is a **right-inverse** of F , that is

$$F \cdot R = id \quad (2)$$

This is proved by circular inclusion

$$F \cdot R \subseteq id \subseteq F \cdot R$$

in the next slide.

Right invertibility

$$\begin{aligned} & F \cdot R \subseteq id \wedge id \subseteq F \cdot R \\ \equiv & \quad \{ \text{img } F = id \text{ and converses} \} \\ & F \cdot R \subseteq F \cdot F^\circ \wedge id \subseteq R^\circ \cdot F^\circ \\ \equiv & \quad \{ \text{ker } R = id \} \\ & F \cdot R \subseteq F \cdot F^\circ \wedge R^\circ \cdot R \subseteq R^\circ \cdot F^\circ \\ \Leftarrow & \quad \{ (F \cdot) \text{ and } (R^\circ \cdot) \text{ are monotone} \} \\ & R \subseteq F^\circ \wedge R \subseteq F^\circ \\ \equiv & \quad \{ R \subseteq F^\circ \text{ is assumed} \} \\ & \text{TRUE} \end{aligned}$$

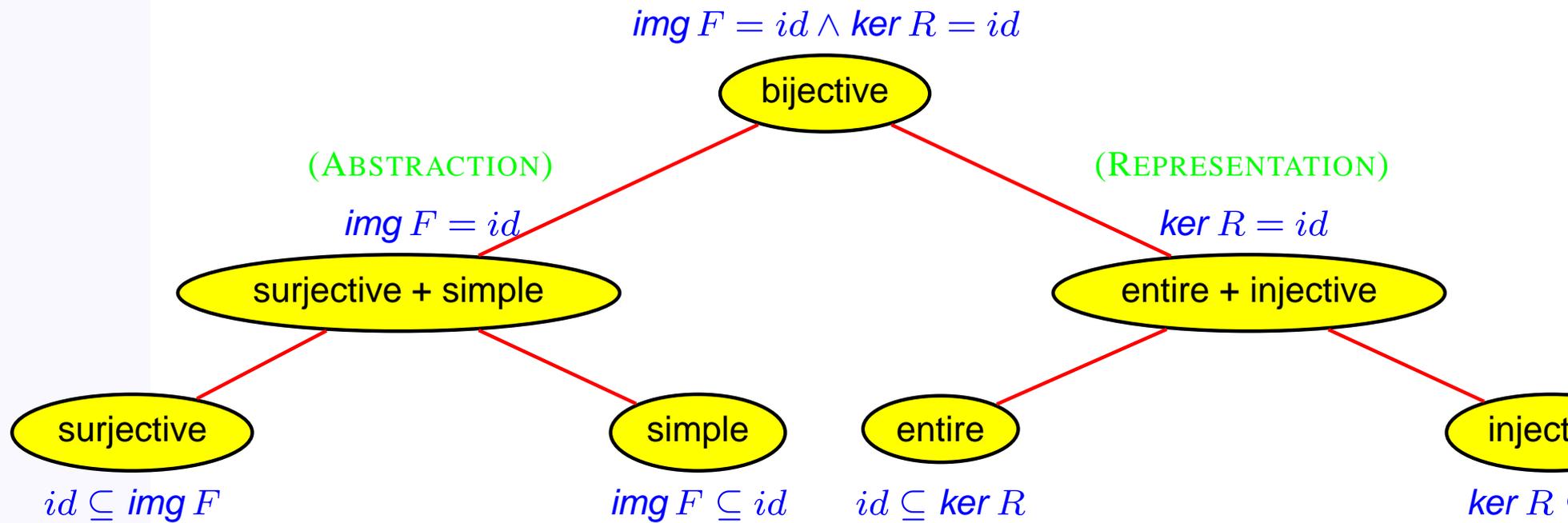
Refinement inequations

$$A \begin{array}{c} \xrightarrow{R} \\ \leq \\ \xleftarrow{F} \end{array} B \quad \text{such that} \quad F \cdot R = id_A$$

This inequation has several informal interpretations:

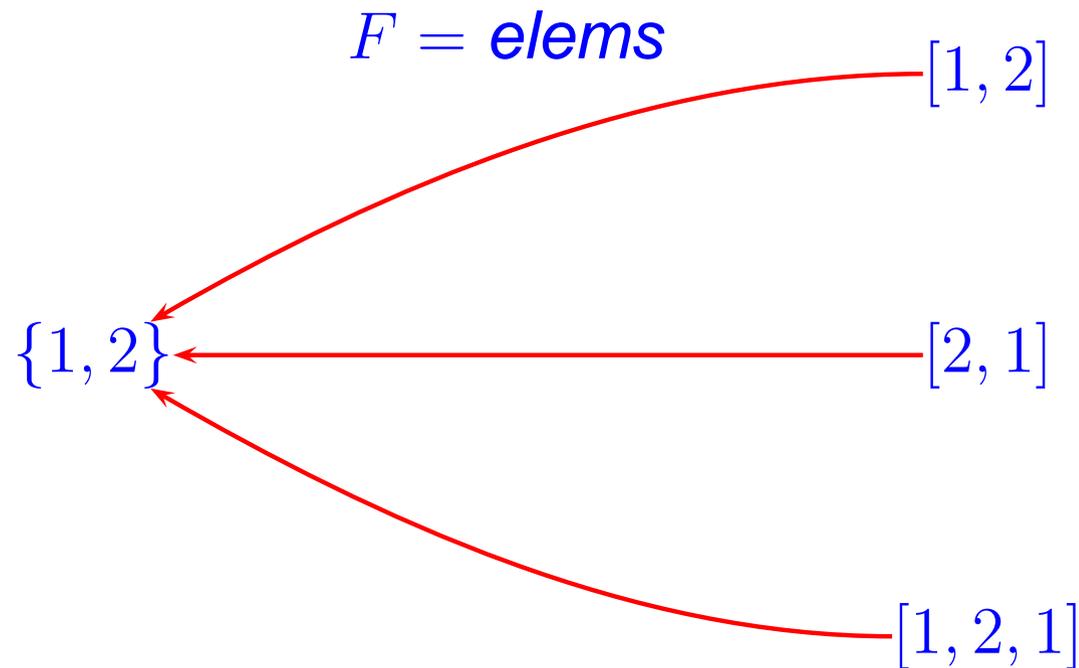
- A is “smaller” than B
- B is able to “represent” A
- B is “abstracted” by A
- A is “implemented” by B
- B is a refinement (“refines”) A

In a diagram



Example

Representing finite **sets** by finite **lists**:



Among the many $R \subseteq F^\circ$, we may choose the following:

Relational representation

```
Listify : set of nat -> seq of nat
Listify(s) ==
  if s = {} then []
  else let e in set s
       in [e] ^ Listify(s \ {e});
```

Intuitively,

$$\text{rng Listify} = \llbracket \text{noRepeats} \rrbracket$$

where

```
noRepeats(s) == card elems s = len s
```

Functional representation

```
listify : set of nat -> seq of nat
listify(s) ==
  if s = {} then []
  else let e = minset(s)
        in [e] ^ listify(s \ {e});
```

Intuitively,

$$\text{rng listify} = \llbracket \text{IsOrdered} \rrbracket \cdot \llbracket \text{noRepeats} \rrbracket$$

Concrete invariants

- Wherever

$$A \begin{array}{c} \xrightarrow{R} \\ \leq \\ \xleftarrow{F} \end{array} B \quad \text{such that } R \subseteq F^\circ \text{ and } \text{rng } R = \llbracket \phi \rrbracket$$

we say that ϕ is the **concrete invariant** induced by R .

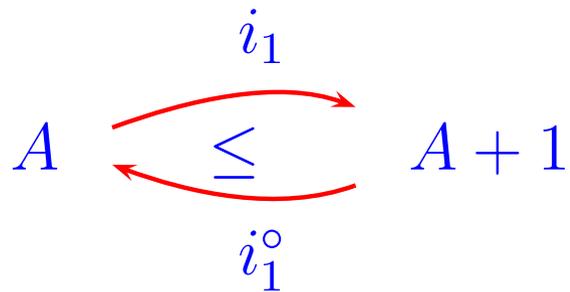
- In case R is a function, and because it always is injective, one has

$$A \cong B_\phi$$

where B_ϕ denotes the subset of B which satisfies concrete-invariant ϕ .

Example of a partial abstraction

Every element of datatype A can be represented by a “pointer”:



- **Simplicity** of the abstraction is ensured by a known fact: the converse of an injective relation is simple.
- **Concrete invariant**: $\phi = [\underline{\text{TRUE}}, \underline{\text{FALSE}}]$

Another partial abstraction

Finite mappings “are” (simple) finite relations:

$$\begin{array}{ccc} & \text{mkr} & \\ & \curvearrowright & \\ \text{map } A \text{ to } B & \leq & \text{set of } (A * B) \\ & \curvearrowleft & \\ & \text{mkf} = \text{mkr}^\circ & \end{array}$$

VDM-SL:

```
mkr : map A to B -> set of (A * B)
mkr(f) == { mk_(a,f(a)) | a in set dom f };

mkf : set of (A * B) -> map A to B
mkf(r) == { p.#1 |-> p.#2 | p in set r }
pre isSimple(r);
```

(Guess the concrete invariant.)

A fundamental iso abstraction

$$A \rightarrow B \begin{array}{c} \xrightarrow{\text{tot}} \\ \simeq \\ \xleftarrow{\text{untot}} \end{array} (B + 1)^A \quad (3)$$

where, for types A , B and $\text{Just}B :: \text{value} : B$,

```
tot: map A to B -> A -> [JustB]
tot(sigma)(a) ==
  if a in set dom(sigma) then mk_JustB(sigma(a)) else nil;

untot: (A -> [JustB]) -> map A to B
untot(f) == { a |-> b | a: A, b: B & f(a) = mk_JustB(b) };
```

Pointfree $untot = (i_1^\circ \cdot)$

As checked next:

$$untot\ f = i_1^\circ \cdot f$$

$$\equiv \{ \text{relations as set comprehensions} \}$$

$$untot\ f = \{(b, a) \mid a \in A, b \in B : b(i_1^\circ \cdot f)a\}$$

$$\equiv \{ \text{using rule } (f\ b)Ra \equiv b(f^\circ \cdot R)a \}$$

$$untot\ f = \{(b, a) \mid a \in A, b \in B : i_1\ b = f\ a\}$$

$$\equiv \{ \text{VDM-SL notation} \}$$

$$untot\ f = \{a \mid -> b \mid a:A, b:B \ \& \ f(a) = mk_JustB(b)\}$$

Easy consequence of *tot/untot*:

$$A \begin{array}{c} \xrightarrow{\cong} \\ \xleftarrow{\cong} \end{array} A^1$$

extends to partial functions as follows:

$$A + 1 \begin{array}{c} \xrightarrow{r} \\ \xleftarrow{f} \end{array} 1 \multimap A \quad (\text{guess } f \text{ and } r).$$

That is, the “singleton” finite map is a disguise of a “pointer” structure.

Properties of \leq :

Reflexivity

$$A \begin{array}{c} \xrightarrow{id} \\ \leq \\ \xleftarrow{id} \end{array} A \quad \text{cf. } id \cdot id = id$$

Transitivity

$$A \begin{array}{c} \xrightarrow{R} \\ \leq \\ \xleftarrow{F} \end{array} B \wedge B \begin{array}{c} \xrightarrow{S} \\ \leq \\ \xleftarrow{G} \end{array} C \Rightarrow A \begin{array}{c} \xrightarrow{S \cdot R} \\ \leq \\ \xleftarrow{F \cdot G} \end{array} C$$

Proof of transitivity

It is enough to show that composition preserves simplicity and surjectiveness:

$$\begin{aligned} & \text{img}(F \cdot G) = id \\ \equiv & \quad \{ \text{expanding and converses} \} \\ & F \cdot (\text{img } G) \cdot F^\circ = id \\ \equiv & \quad \{ G \text{ is simple and surjective} \} \\ & \text{img } F = id \\ \equiv & \quad \{ F \text{ is simple and surjective} \} \\ & id = id \end{aligned}$$

Also note that $S \cdot R \subseteq (F \cdot G)^\circ$ by monotonicity.

Structural data refinement

$$A \begin{array}{c} \xrightarrow{R} \\ \leq \\ \xleftarrow{F} \end{array} B \quad \Rightarrow \quad FA \begin{array}{c} \xrightarrow{FR} \\ \leq \\ \xleftarrow{FF} \end{array} FB$$

where F is an arbitrary relator (functor):

$$\begin{aligned} & id \\ = & \{ \text{functors commute with } id \} \\ & F id \\ = & \{ R \text{ is right-inverse of } F \} \\ & F(F \cdot R) \\ = & \{ \text{functors commute with composition} \} \\ & (FF) \cdot (FR) \end{aligned}$$

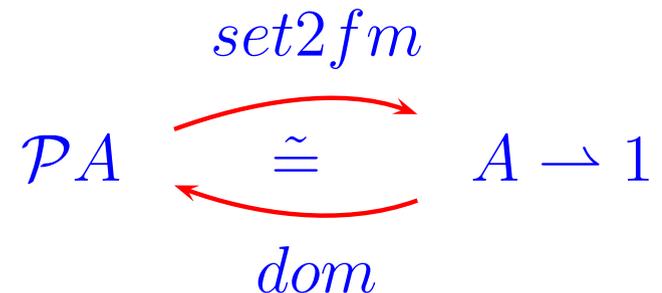
Refining finite sets (I)

$$\mathcal{P}A \overset{\cong}{\rightleftarrows} A \multimap 1$$

Calculation:

$$\begin{aligned} & A \multimap 1 \\ \cong & \quad \{ \text{tot representation} \} \\ & (1 + 1)^A \\ \cong & \quad \{ \text{basic} \} \\ & 2^A \\ \cong & \quad \{ 2^A \text{ is isomorphic to } \mathcal{P}A \} \\ & \mathcal{P}A \end{aligned}$$

Refining finite sets (1a)



VDM-SL

```
set2fm : set of A -> map A to Nil
set2fm(s) == { a |-> nil | a in set s };
```

Pointfree

$$\text{set2fm} \stackrel{\text{def}}{=} (!\cdot)$$

Right-invertibility

Calculation:

$$\mathit{dom} \cdot \mathit{set2fm} = \mathit{id}$$

$$\equiv \{ \}$$

$$\mathit{dom} (\mathit{set2fm} \ s) = s$$

$$\equiv \{ \}$$

$$\mathit{dom} (! \cdot s) = s$$

$$\equiv \{ ! \text{ is a function, } \mathit{dom} (f \cdot R) = \mathit{dom} R \}$$

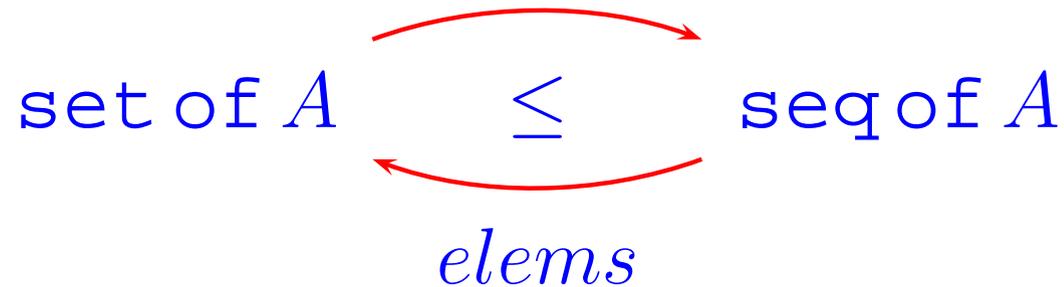
$$\mathit{dom} \ s = s$$

$$\equiv \{ s \text{ is coreflexive} \}$$

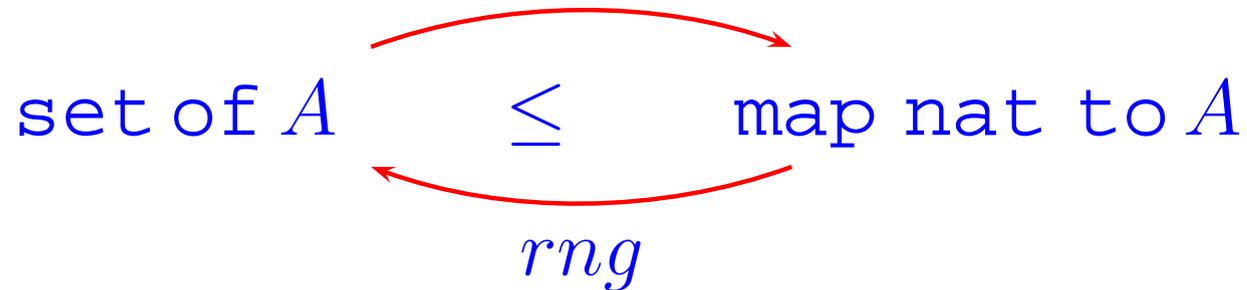
$$s = s$$

Refining finite sets (II)

List (cf. example before):

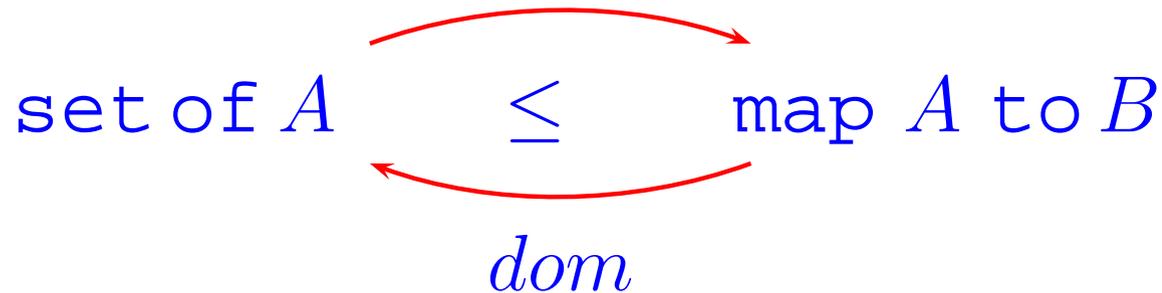


Index A :

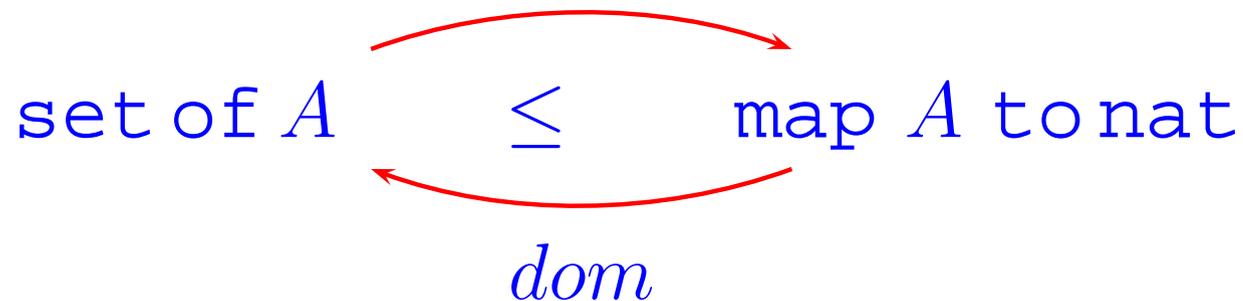


Refining finite sets (III)

Classify A by B ($B \supset \{\}$):



Quantify A ("multisets"):



Refining finite maps (I)

```
JustB::value:B;  
JustC::value:C;  
BorC = JustB | JustC ;
```

$$\text{map } (BorC) \text{ to } A \quad \cong \quad (\text{map } B \text{ to } A) \times (\text{map } C \text{ to } A)$$

peither

```
peither: (map B to A) * (map C to A) -> map BorC to A  
peither(m,n) == { mk_JustB(b) |-> m(b) | b in set dom m } union  
                { mk_JustC(c) |-> n(c) | c in set dom n };
```

Refining finite maps (Ia)

$$(B + C) \rightarrow A \quad \cong \quad (B \rightarrow A) \times (C \rightarrow A)$$

unpeither

peither

where

$$\text{peither}(\sigma, \tau) = [\sigma, \tau]$$

for $[R, S] = (R \cdot i_1^\circ) \cup (S \cdot i_2^\circ)$, that is

$$\text{peither} = \cup \cdot ((\cdot i_1^\circ) \times (\cdot i_2^\circ))$$

Refining finite maps (II)

$$A \rightarrow (B + C) \leq (A \rightarrow B) \times (A \rightarrow C)$$

uncojoin

cojoin

where

$$cojoin = \cup \cdot ((i_1 \cdot) \times (i_2 \cdot))$$

NB: *cojoin* is partial since the union of two partial functions not always is a partial function.

Refining finite maps (IIa)

Note the representation function:

```
uncojoin : map A to BorC -> (map A to B) * (map A to C)
uncojoin(f) ==
  mk_( { a |-> f(a).value
        | a in set dom f & is_JustB(f(a)) },
        { a |-> f(a).value
        | a in set dom f & is_JustC(f(a)) }
  );
```

The finite map bifunctor

- Note the $(\cdot i_1^\circ)$ s, $(i_1 \cdot)$ s, etc
- In general, for an injective f and any g , define bifunctor

$$f \rightarrow g \stackrel{\text{def}}{=} (g \cdot) \cdot (\cdot f^\circ)$$

that is

$$(f \rightarrow g)\sigma = g \cdot \sigma \cdot f^\circ$$

- So, we could have written e.g.

$$peither = \cup \cdot ((i_1 \rightarrow id) \times (i_2 \rightarrow id))$$

Refining finite maps (III)

$$\text{map } A \text{ to } B * C \quad \leq \quad (\text{map } A \text{ to } B) \times (\text{map } A \text{ to } C)$$

\bowtie

where (writing `join` for \bowtie)

```
join : (map A to B) * (map A to C) -> map A to (B * C)
join(m,n) == { a |-> mk_(m(a),n(a))
              | a in set dom m inter dom n };
```

Refining finite maps (IIIa)

$$\begin{array}{ccc} & \textit{unjoin} & \\ & \curvearrowright & \\ A \multimap B \times C & \leq & (A \multimap B) \times (A \multimap C) \\ & \curvearrowleft & \\ & \boxtimes & \end{array}$$

where

$$\sigma \boxtimes \tau \stackrel{\text{def}}{=} \langle \sigma, \tau \rangle$$

where $\langle R, S \rangle \stackrel{\text{def}}{=} (\pi_1^\circ \cdot R) \cap (\pi_2^\circ \cdot S)$. A right-inverse of *join* is

$$\textit{unjoin} \stackrel{\text{def}}{=} \langle id \multimap \pi_1, id \multimap \pi_2 \rangle$$

Refining finite maps (IV)

How do we extend

$$B^{C \times A} \begin{array}{c} \xrightarrow{\text{curry}} \\ \cong \\ \xleftarrow{\text{uncurry}} \end{array} (B^A)^C$$

to partial functions? Case $B := B + 1$

$$\begin{aligned} & (B + 1)^{C \times A} \cong ((B + 1)^A)^C \\ \equiv & \quad \{ \text{that is} \} \\ & (C \times A) \multimap B \cong (A \multimap B)^C \end{aligned}$$

Refining finite maps (IVa)

In general:

$$(C \times A) \multimap B \quad \leq \quad C \multimap (A \multimap B)$$

pcurry (top arrow) and *unpcurry* (bottom arrow)

```
unpcurry : map C to (map A to B) -> map (C * A) to B
unpcurry(f) ==
  merge { let g=f(a)
          in { mk_(a,b) |-> g(b) | b in set dom g }
        | a in set dom f };
```

Refining finite maps (IVb)

```
pcurry : map (C * A) to B -> map C to (map A to B)
pcurry(f) ==
  let y = { x.#1 | x in set dom f }
  in { a |-> { p.#2 |-> f(p)
             | p in set dom f & p.#1=a }
      | a in set y };
```

Transposing relations

Let $B := 2$ in the *curry/uncurry* isomorphism and obtain

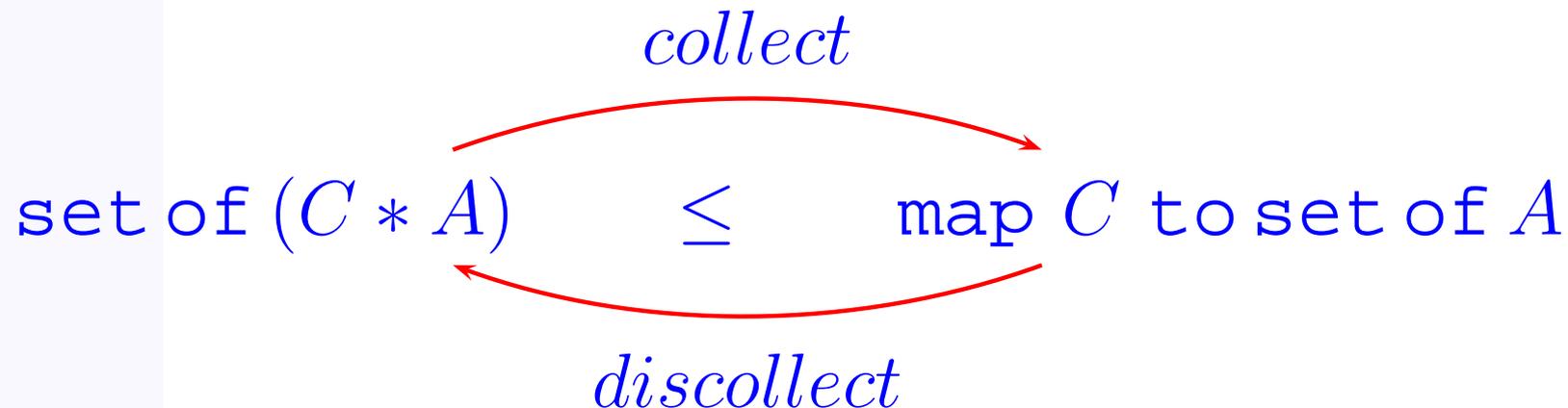
$$\mathcal{P}(A \times C) \begin{array}{c} \xrightarrow{\Lambda} \\ \cong \\ \xleftarrow{\Lambda^\circ} \end{array} (\mathcal{P}A)^C$$

where

$$f = \Lambda R \quad \equiv \quad R = \in \cdot f \quad (4)$$

and $A \xleftarrow{\in} \mathcal{P}A$ is the membership relation.

Transposing finite relations



```
collect : set of (C * A) -> map C to set of A
collect(r) == { c |-> { q.#2 | q in set r & c=q.#1 }
                 | c in set { p.#1 | p in set r } };

discollect : map C to set of A -> set of (C * A)
discollect(f) == dunion { { mk_(c,a) | a in set f(c) }
                         | c in set dom f };
```

Refining finite maps (V)

Last but not least

$$A \multimap D \times (B \multimap C) \leq (A \multimap D) \times ((A \times B) \multimap C) \quad (5)$$

unnjoin

\bowtie_n

where

$$\bowtie_n \stackrel{\text{def}}{=} \bowtie \cdot \langle \pi_1, \dagger \cdot ((id \multimap \underline{\emptyset}) \times pcurry) \rangle \quad (6)$$

and

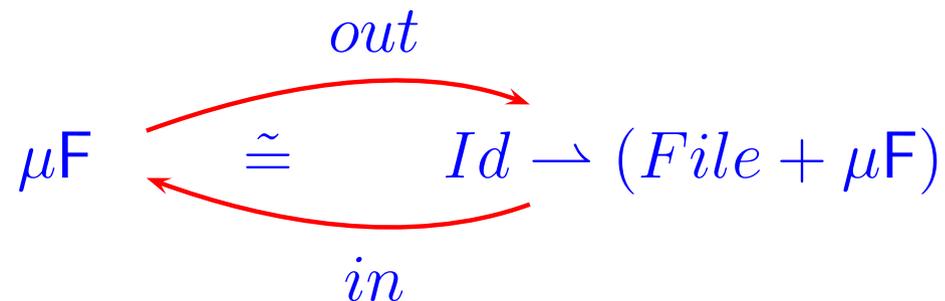
$$unnjoin \stackrel{\text{def}}{=} (id \times unpcurry) \cdot unjoin$$

Recursive data refinement

How does one refine recursive VDM-SL models such as e.g.

```
FS :: D: map Id to Node; -- FS means file system
Node = File | FS;       -- a Node is either a file
                        -- or a directory
Id = seq of char;      -- node identifiers
File :: F: seq of token -- sequential files
```

that is, $FS = \mu F$ for $F X = Id \rightarrow (File + X)$:



Recursive data refinement

or ...

```
DecTree :: question: What
          answers: map Answer to DecTree
What = seq of char;
Answer = seq of char;
```

that is, $DecTree = \mu F$ in

$$DecTree \cong What \times (Answer \rightarrow DecTree)$$

for $F X = What \times (Answer \rightarrow X)$

Recursion “removal”

Given

$$\mu F \begin{array}{c} \xrightarrow{\text{out}} \\ \cong \\ \xleftarrow{\text{in}} \end{array} F \mu F$$

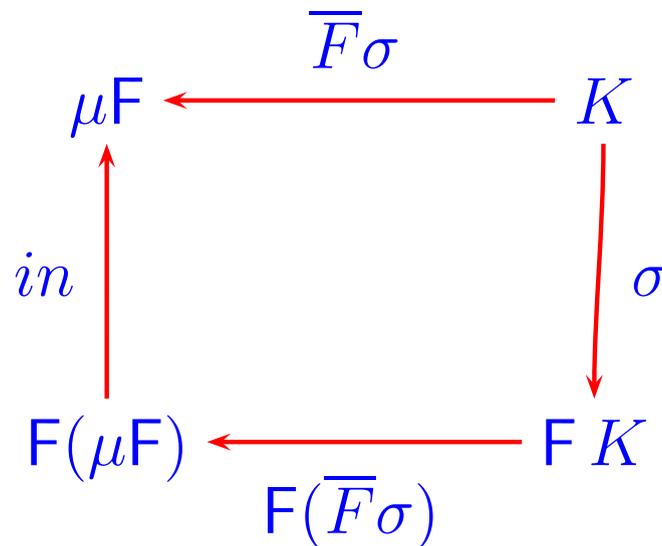
one has

$$\mu F \begin{array}{c} \xrightarrow{\quad} \\ \leq \\ \xleftarrow{F} \end{array} (K \multimap F K) \times K \quad (7)$$

for K a domain of “pointers” such that $K \cong \mathbb{N}$.

Abstraction function

- Main rôle in representation is played by a (partial) F -coalgebra $F K \xleftarrow{\sigma} K$, assumed as a (finite) piece of “linear storage”, a “heap” or a “database”.
- \overline{F} (the transpose of abstraction F) is of type $(K \rightarrow F K) \rightarrow K \rightarrow \mu F$ and one can build hylomorphism



$$\overline{F}\sigma = in \cdot F(\overline{F}\sigma) \cdot \sigma$$

Partiality of implementation

$F(\sigma, k) = (\overline{F}\sigma)k$ will be undefined wherever

- $k \notin \text{dom } \sigma$
- σ is not “closed” over itself (see below)
- σ is non-well-founded (see below)

Thus concrete invariant

$$\phi(\sigma, k) \stackrel{\text{def}}{=} k \in \text{dom } \sigma \wedge (\text{closed } \sigma) \wedge (\text{wellf } \sigma)$$

In order to define $\text{closed } \sigma$ and $\text{wellf } \sigma$ we need σ 's **accessibility** relation \prec_σ (next slide).

Accessibility and membership

Accessibility relation for σ :

$$K \xleftarrow{\lambda_\sigma} K$$
$$\lambda_\sigma \stackrel{\text{def}}{=} \in_F \cdot \sigma$$

where $K \xleftarrow{\in_F} F K$ extends $K \xleftarrow{\in} \mathcal{P}K$ inductively over polynomial functors, as follows:

$$\begin{aligned} \in_{\mathcal{P}} &\stackrel{\text{def}}{=} \in \\ \in_C &\stackrel{\text{def}}{=} \perp \\ \in_{\lambda X.X} &\stackrel{\text{def}}{=} id \\ \in_{F \times G} &\stackrel{\text{def}}{=} (\in_F \cdot \pi_1) \cup (\in_G \cdot \pi_2) \\ \in_{F+G} &\stackrel{\text{def}}{=} [\in_F, \in_G] \end{aligned}$$

Example

Let $F X = 1 + A \times X$. Then,

$$\begin{aligned} & \in_{1+A \times X} \\ = & \{ \in \text{ for coproduct bifunctor } \} \\ & [\in_1, \in_{A \times X}] \\ = & \{ \in \text{ for constant and product (bi)functors } \} \\ & [\perp, (\in_A \cdot \pi_1) \cup (\in_{\lambda X.X} \cdot \pi_2)] \\ = & \{ \in \text{ for constant and identity functor } \} \\ & [\perp, (\perp \cdot \pi_1) \cup (id \cdot \pi_2)] \\ = & \{ \perp \text{ and } [R, S] = (R \cdot i_1^\circ) \cup (S \cdot i_2^\circ) \} \\ & \pi_2 \cdot i_2^\circ \end{aligned}$$

Example (pointfree)

$$k \in_{1+A \times X} x$$

$$\equiv \{ \text{calculation above} \}$$

$$k(\pi_2 \cdot i_2^\circ)x$$

$$\equiv \{ \text{relational composition} \}$$

$$k(\pi_2)(a, k') \wedge x = i_2(a, k')$$

$$\equiv \{ \text{trivia} \}$$

$$x = i_2(a, k') \wedge k = k'$$

$$\equiv \{ \text{trivia} \}$$

$$x = i_2(a, k)$$

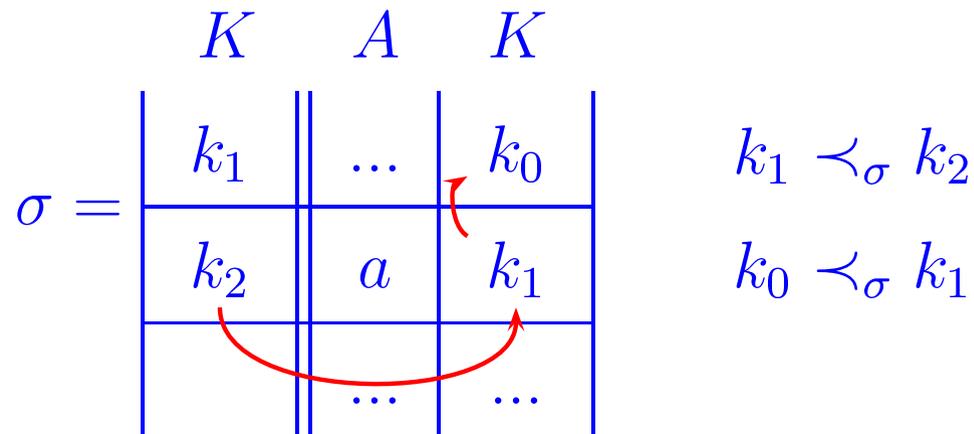
Accessibility (example)

Pointer reachability in case of a “linear” heap

$$(1 + A \times K) \xrightarrow{\sigma} K:$$

$$k_1 \prec_{\sigma} k_2 \equiv k_2 \in \text{dom } \sigma \wedge (\sigma k_2) = i_2(a, k_1)$$

In a drawing:



Closure and wellfoundedness

Let \prec_{σ}^{+} denote the transitive closure of \prec_{σ} . Then,

- *closed* $\sigma = \text{rng } \prec_{\sigma}^{+} \subseteq \text{dom } \sigma$ that is, all reachable k are defined.
- *wellf* $\sigma = (\prec_{\sigma}^{+}) \cap \text{id} = \perp$, that is, \prec_{σ}^{+} is irreflexive (no cycles)