

Making Program Logics Intelligible

John C. Reynolds
Carnegie Mellon University

Lovelace Lecture — June 8, 2011

Research partially supported by National Science Foundation Grant CCF-0916808

Dedication

To the British computer scientists who taught me so much when I was young.

Those who are gone:

Christopher Strachey Peter Landin Robin Milner

and those who continue to instruct me:

Tony Hoare Rod Burstall Alan Robinson

Introduction

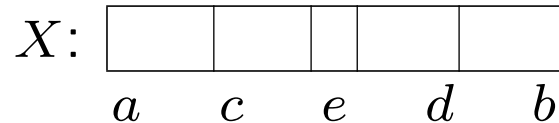
To verify program specifications, rather than generic safety properties, it will be necessary to integrate verification into the process of programming.

Program proving is unlike theorem proving in mathematics - mathematical conjectures may give no hint as to how they could be proved, but programs are written by programmers, who must understand informally why their programs work. The job of verification is not to explore some immense search space, but to formalize the programmer's intuitions until any faults are revealed.

This requires specifications and proofs that are succinct and intelligible - which in turn require logics that go beyond predicate calculus (the assembly language of program proving). In this talk, I will recount and illustrate several steps, old and new, towards this goal.

The Distance between Traditional Logic and Programming

Particularly when describing arrays, program specifications are typically full of inequalities and set definitions using inequalities. For example, a diagram that a programmer might write, e.g.,

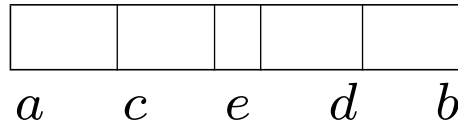


might be expressed by:

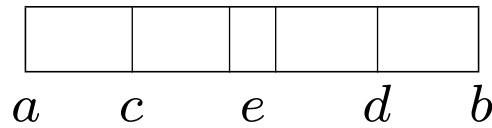
$$a \leq c \leq e \leq d \leq b \quad \text{and} \quad \text{dom } X = \{i \mid a < i < b\}.$$

Partition Diagrams

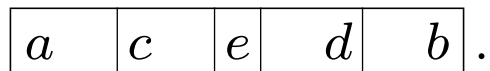
The obvious difficulty with diagrams such as



is that expressions tend to migrate slightly:



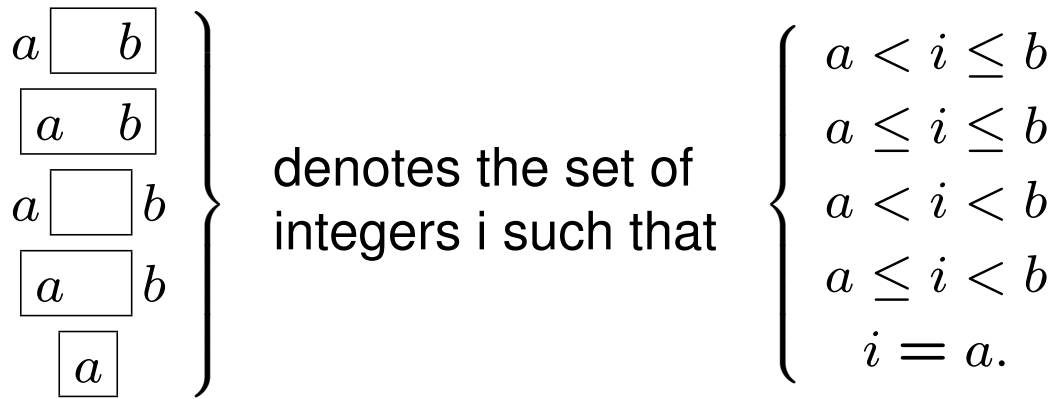
This fertile source of program errors can be avoided by capturing the expressions inside the boxes:



We call such a diagram a *partition diagram*.

Interval Diagrams

An *interval diagram* is an annotated box denoting a finite consecutive set of integers. In particular



Partition Diagrams

A diagram of the form

$$a_0 \boxed{a_1} \boxed{a_2} \cdots \boxed{a_n}$$

is called a *partition diagram*. The intervals denoted by

$$a_0 \boxed{a_1}, a_1 \boxed{a_2}, \dots, a_{n-1} \boxed{a_n}$$

are called the *component intervals* of the partition diagram, and the interval denoted by $a_0 \boxed{a_n}$ is called the *total interval* of the partition diagram.

A partition diagram is an assertion that its component intervals are a partition of its total interval, i.e., that the component intervals are disjoint and their union is the total interval.

Some Abbreviations

Just as with interval diagrams, we may write $|a$ instead of $a-1|$ and \boxed{a} instead of $\boxed{a \ a}$ within partition diagrams. For example,

$$\boxed{a \ k \ b}$$

has the same meaning as

$$\boxed{a \ k \ k \ b},$$

which has the same meaning as

$$a-1 \boxed{k-1 \ k \ b}.$$

Summing an Array (Annotated Specification)

$\{ \boxed{a \quad b} \subseteq \text{dom } X \}$

newvar k **in** ($k := a - 1 ; s := 0 ;$

$\{ \text{inv: } \boxed{a \quad k \quad b} \wedge s = \sum_{i \in \boxed{a \quad k}} X(i) \}$

while $k < b$ **do** (

$k := k + 1 ;$

$\{ \boxed{a \quad k \quad b} \wedge s = \sum_{i \in \boxed{a \quad k}} X(i) \}$

$s := s + X(k) \}$

$\{ s = \sum_{i \in \boxed{a \quad b}} X(i) \}$

Array Values as Functions

We regard an array as a variable whose value is a function, whose domain is an interval. For example,

real array $X(1 : 10) \dots$

declares X to be an array whose values have the domain $1 \ 10$.

When $S \subseteq \text{dom } X$, we write $X \upharpoonright S$ for the *restriction* of X to S , which is the function such that

$$\text{dom}(X \upharpoonright S) = S \quad \forall i \in S. (X \upharpoonright S) i = X i.$$

We also write $\{X\}$ for the *image* of X , i.e., the set

$$X = \{ X i \mid i \in \text{dom } X \}.$$

For example, suppose sq is the function with domain $\boxed{0 \ 5}$ that maps each number between 0 and 5 into its square. Then

$$\{sq\} = \{0, 1, 4, 9, 16, 25\} \quad \{sq\} \boxed{2 \ 4} = \{4, 9, 16\}.$$

Pointwise Extension of a Relation

Suppose ρ is a binary relation between values. Then the pointwise extension of ρ , written ρ^* , is the relation between sets of such values such that

$$S \rho^* T \text{ iff } \forall x \in S, y \in T. x \rho y.$$

For example,

$$\{2, 3\} \leq^* \{3, 4\} \quad \{2, 3\} \neq^* \{4, 5\}$$

are both true, but

$$\{2, 3\} <^* \{3, 4\} \quad \{2, 3\} =^* \{2, 3\} \quad \{2, 3\} \neq^* \{2, 3\}$$

are all false.

We also abbreviate

$\{x\} \rho^* T$ by $x \rho^* T$

$S \rho^* \{y\}$ by $S \rho^* y$.

Ordered Arrays

We define the assertion

$\text{ord}_\rho X$ iff $\forall i, j \in \text{dom } X. i < j$ implies $X(i) \rho X(j)$.

For example, $\text{ord}_{\leq} X$ asserts that X is ordered in nonstrict increasing order.

Binary Search

The following is the invariant of a simple while-loop program for binary search:

$$\begin{aligned} & \boxed{a \quad b} \subseteq \text{dom } X \wedge \text{ord}_{\leq}(X \upharpoonright \boxed{a \quad b}) \wedge \\ & \boxed{a \quad c \quad d \quad b} \wedge \\ & \{X \upharpoonright \boxed{a \quad c}\} \leq^* y \wedge y <^* \{X \upharpoonright d \boxed{b}\}. \end{aligned}$$

Compare this with

$$\begin{aligned} & (\forall i. a \leq i \leq b \text{ implies } i \in \text{dom } X) \wedge \\ & (\forall i, j. a \leq i < j \leq b \text{ implies } X(i) \leq X(j)) \wedge \\ & a - 1 \leq c - 1 \leq d \leq b \wedge \\ & (\forall i. a \leq i < c \text{ implies } X(i) \leq y) \wedge \\ & (\forall i. d < i \leq b \text{ implies } y < X(i)). \end{aligned}$$

Partition: The Invariants

$$I = \boxed{m \quad i \quad n} \wedge \{A\} \boxed{m \quad i} \leq^* r \wedge$$

$$\boxed{m \quad j \quad n} \wedge r \leq^* \{A\} \boxed{j \quad n} \wedge$$

$$(i \leq j \text{ implies } (\exists p. \boxed{i \quad p \quad n} \wedge r \leq A(p)) \wedge$$

$$(\exists q. \boxed{m \quad q \quad j} \wedge A(q) \leq r))$$

$$I_1 = \boxed{m \quad i \quad n} \wedge \{A\} \boxed{m \quad i} \leq^* r \wedge$$

$$(\exists p. \boxed{i \quad p \quad n} \wedge r \leq A(p))$$

$$I_2 = \boxed{m \quad j \quad n} \wedge r \leq^* \{A\} \boxed{j \quad n} \wedge$$

$$(\exists q. \boxed{m \quad q \quad j} \wedge A(q) \leq r)$$

$$I' = \boxed{m \quad i \quad n} \wedge \{A\} \boxed{m \quad i} \leq^* r \wedge$$

$$\boxed{m \quad j \quad n} \wedge r \leq^* \{A\} \boxed{j \quad n} \wedge$$

$$(i+1 \leq j-1 \text{ implies } (\exists p. i \boxed{\quad} p \boxed{\quad} n \wedge r \leq A(p)) \wedge$$

$$(\exists q. \boxed{m \quad q \quad \quad} j \wedge A(q) \leq r))$$

Partition: The Procedure

let Partition($A, i, j ; m, n$) =

$\{ \boxed{m \quad n} \subseteq \text{dom } A \wedge \boxed{m \quad n} \}$

newvar r in (

$r := A((m + n) \div 2) ; i := m ; j := n ;$

$\{\text{inv: } I\}$

while $i \leq j$ do (

$\{I_1 \wedge I_2\}$

while $A(i) < r$ do $i := i + 1 ;$

while $r < A(j)$ do $j := j - 1 ;$

$\{I_1 \wedge r \leq A(i) \wedge I_2 \wedge A(j) \leq r\}$

if $i \leq j$ then (

newvar t in ($t := A(i) ; A(i) := A(j) ; A(j) := t$) ;

$\{I_1 \wedge A(i) \leq r \wedge I_2 \wedge r \leq A(j) \wedge i \leq j\}$

$\{I'\}$

$i := i + 1 ; j := j - 1$)))

$\{ \boxed{m \quad j} \mid \boxed{i \quad n} \wedge \{A \mid \boxed{m} \quad i\} \leq^* \{A \mid j \quad \boxed{n}\} \}$

Rearrangement

We write $X \sim Y$, and say that X is a *rearrangement* of Y , when there is a bijection (i.e., a one-to-one correspondence) B from $\text{dom } X$ to $\text{dom } Y$ such that

$$\forall i \in \text{dom } X. X(i) = Y(B(i)).$$

Suppose $S \subseteq \text{dom } X = \text{dom } Y$. We write $X \sim_S Y$, and say that X is a *rearrangement restricted to S* of Y , when X is a rearrangement of Y and, for all $i \in (\text{dom } X) - S$, $X(i) = Y(i)$.

Partition Revisited

let Partition($A, i, j ; m, n$) =

$\{ \boxed{m \ n} \subseteq \text{dom } A \wedge \boxed{m} \boxed{n} \wedge A = A_0 \}$

newvar r **in** (

$r := A((m + n) \div 2) ; i := m ; j := n ;$

$\{\text{inv}: I \wedge A \sim_{\boxed{m \ n}} A_0\}$

while $i \leq j$ **do** (

$\{I_1 \wedge I_2 \wedge A \sim_{\boxed{m \ n}} A_0\}$

while $A(i) < r$ **do** $i := i + 1 ;$

while $r < A(j)$ **do** $j := j - 1 ;$

$\{I_1 \wedge r \leq A(i) \wedge I_2 \wedge A(j) \leq r \wedge A \sim_{\boxed{m \ n}} A_0\}$

if $i \leq j$ **then** (

newvar t **in** ($t := A(i) ; A(i) := A(j) ; A(j) := t$) ;

$\{I_1 \wedge A(i) \leq r \wedge I_2 \wedge r \leq A(j) \wedge i \leq j \wedge A \sim_{\boxed{m \ n}} A_0\}$

$\{I' \wedge A \sim_{\boxed{m \ n}} A_0\}$

$i := i + 1 ; j := j - 1$))

$\{ \boxed{m \ j} \boxed{i \ n} \wedge \{A \mid \boxed{m} \ i\} \leq^* \{A \mid j \ \boxed{n}\} \wedge A \sim_{\boxed{m \ n}} A_0 \}$

Quicksort

letrec Quicksort($A ; m, n$) =

$\{ \boxed{m \ n} \subseteq \text{dom } A \wedge A = A_0 \}$

if $m < n$ then newvar i, j in (

$\{ \boxed{m \ n} \subseteq \text{dom } A \wedge \boxed{m \ \ \ n} \}$

Partition($A, i, j ; m, n$) ;

$\{ \boxed{m \ j \ \ \ i \ n} \wedge \{A \mid \boxed{m \ \ \ i}\} \leq^* \{A \mid j \ \ \ n\} \wedge A \sim_{\boxed{m \ n}} A_0 \}$

Quicksort($A ; m, j$) ;

$\{ \boxed{m \ j \ \ \ i \ n} \wedge \text{ord}_{\leq A} \mid \boxed{m \ j} \wedge$

$\{A \mid \boxed{m \ \ \ i}\} \leq^* \{A \mid j \ \ \ n\} \wedge A \sim_{\boxed{m \ n}} A_0 \}$

Quicksort($A ; i, n$)

$\{ \boxed{m \ j \ \ \ i \ n} \wedge \text{ord}_{\leq A} \mid \boxed{m \ j} \wedge \text{ord}_{\leq A} \mid \boxed{i \ n} \wedge$

$\{A \mid \boxed{m \ \ \ i}\} \leq^* \{A \mid j \ \ \ n\} \wedge A \sim_{\boxed{m \ n}} A_0 \}$)

$\{ \text{ord}_{\leq A} \mid \boxed{m \ n} \wedge A \sim_{\boxed{m \ n}} A_0 \}$

Realignment

We write $X \simeq Y$, and say that X is a *realignment* of Y , when there is a *monotonic* bijection B from $\text{dom } X$ to $\text{dom } Y$ such that

$$\forall i \in \text{dom } X. X(i) = Y(B(i)).$$

When X and Y are array values, i.e., functions on intervals, $X \simeq Y$ holds only when X is a “shift” of Y .

But when X and Y are functions whose domains can be arbitrary finite sets of integers — which we might call “lacy array values” — things become more interesting.

Eliminating Zeroes

$$\{\boxed{a \quad b} \subseteq X \wedge X = X_0\}$$

newvar d in $(c := a ; d := a ;$

$$\{\text{inv: } \boxed{a \quad c \quad d \quad b} \wedge$$

$$X \upharpoonright \boxed{a \quad c} \simeq X_0 \upharpoonright (\boxed{a \quad d} \cap \{i \mid X(i) \neq 0\}) \wedge$$

$$X \upharpoonright \boxed{d \quad b} = X_0 \upharpoonright \boxed{d \quad b}\}$$

while $d \leq b$ do

if $X(d) = 0$ then $d := d + 1$

else $(X(c) := X(d) ; c := c + 1 ; d := d + 1)$

$$\{\boxed{a \quad c \quad b} \wedge X \upharpoonright \boxed{a \quad c} \simeq X_0 \upharpoonright (\boxed{a \quad b} \cap \{i \mid X(i) \neq 0\})\}$$

Separation Logic: An Example

Suppose

$$\text{list } \epsilon \ i \stackrel{\text{def}}{=} i = \text{nil}$$

$$\text{list}(a \cdot \alpha) \ i \stackrel{\text{def}}{=} \exists j. i \hookrightarrow a, j \wedge \text{list } \alpha \ j$$

and consider the program

$$LREV \stackrel{\text{def}}{=} j := \text{nil};$$

while $i \neq \text{nil}$ do $(k := [i + 1] ; [i + 1] := j ; j := i ; i := k)$.

To prove $\{\text{list } \alpha \ i\} LREV \{\text{list } \alpha^\dagger \ j\}$, the invariant

$$\exists \alpha, \beta. \text{list } \alpha \ i \wedge \text{list } \beta \ j \wedge \alpha_0^\dagger = \alpha^\dagger \cdot \beta$$

is inadequate.

An adequate invariant (in Hoare logic):

$$(\exists \alpha, \beta. \text{list } \alpha \text{ } i \wedge \text{list } \beta \text{ } j \wedge \alpha_0^\dagger = \alpha^\dagger \cdot \beta) \\ \wedge (\forall k. \text{reachable}(i, k) \wedge \text{reachable}(j, k) \Rightarrow k = \text{nil}).$$

An adequate invariant (in separation logic):

$$\exists \alpha, \beta. (\text{list } \alpha \text{ } i * \text{list } \beta \text{ } j) \wedge \alpha_0^\dagger = \alpha^\dagger \cdot \beta.$$

where $*$ is the *separating conjunction*.

Enriching the Concept of State

In separation logic, the *state* consists of two components:

- A *store*, which maps variables into their values,
- A *heap*, which maps addresses (which are values) into their values.

The Separating Conjunction

The assertion $p_1 * p_2$ holds for a heap h when h can be partitioned into two disjoint subheaps h_1 and h_2 such that p_1 holds in h_1 and p_2 holds in h_2 .

The Points-To Relation

The assertion $\ell \mapsto e$ holds when the heap consists of a single cell mapping the address ℓ into the value of e .

Some Inference Rules

- Frame Rule (O'Hearn)

$$\frac{\{p\} \ c \ \{q\}}{\{p * r\} \ c \ \{q * r\}},$$

where the free variables of r are not modified by c .

- Existential Quantification

$$\frac{\{p\} \ c \ \{q\}}{\{\exists v. p\} \ c \ \{\exists v. q\}},$$

where v does not occur free in c .

- Concurrent Composition (O'Hearn)

$$\frac{\{p_1\} c_1 \{q_1\} \quad \{p_2\} c_2 \{q_2\}}{\{p_1 * p_2\} c_1 \parallel c_2 \{q_1 * q_2\}},$$

where the free variables of p_1 , c_1 , and q_1 are not modified by c_2 , and vice-versa.

The Iterated Separating Conjunction

Suppose S is a finite set. Then the assertion $\odot_{i \in S} P(i)$ holds when the heap can be partitioned into a family h_i of heaplets indexed by i , such that, for all $i \in S$, $P(i)$ holds in h_i .

For example, $\odot_{i \in S} \ell + i \mapsto A(i)$ describes an array with base address ℓ , whose subscripts range over S , and whose value is the function A .

More Definitions

When S is a finite set of integers,

$$\ell \mapsto^S A \stackrel{\text{def}}{=} S \subseteq \text{dom } A \wedge \bigodot_{i \in S} \ell + i \mapsto A(i)$$

$$A \sim^S A' \stackrel{\text{def}}{=} (A \upharpoonright S) \sim (A' \upharpoonright S).$$

Quicksort Revisited: Some Assumptions

$$\{\ell \mapsto \boxed{m \ n} A_0\}$$

Quicksort(; ℓ, m, n)

$$\{\exists A. \ell \mapsto \boxed{m \ n} A \wedge A \sim \boxed{m \ n} A_0 \wedge \mathbf{ord}_{\leq}(A \upharpoonright \boxed{m \ n})\}$$

$$\{\ell \mapsto \boxed{m \ n} A_0 \wedge \boxed{m} \ \boxed{n}\}$$

Partition(i, j ; ℓ, m, n)

$$\{\exists A. \ell \mapsto \boxed{m \ n} A \wedge A \sim \boxed{m \ n} A_0 \wedge \boxed{m \ j} \ \boxed{i \ n} \wedge \\ \{A \upharpoonright \boxed{m} \ i\} \leq^* \{A \upharpoonright j \ \boxed{n}\}\}$$

Quicksort Revisited: The Procedure Body

$\{\ell \mapsto \boxed{m\ n} A_0\}$

if $m < n$ then newvar i, j in (

$\{\ell \mapsto \boxed{m\ n} A_0 \wedge \boxed{m\ \ \ n}\}$

Partition($i, j ; \ell, m, n$)

$\{\exists A_1. \ell \mapsto \boxed{m\ n} A_1 \wedge A_1 \sim \boxed{m\ n} A_0 \wedge \boxed{m\ j\ \ \ i\ n} \wedge$
 $\{A_1 \upharpoonright \boxed{m\ \ \ i}\} \leq^* \{A_1 \upharpoonright \boxed{j\ \ \ n}\}\}$

$\left\{ \begin{array}{l} \{\ell \mapsto \boxed{m\ j} A_1\} \\ \text{Quicksort}(; \ell, m, j) \\ \{\exists A_2. \ell \mapsto \boxed{m\ j} A_2 \wedge \\ A_2 \sim \boxed{m\ j} A_1 \wedge \\ \text{ord}_{\leq}(A_2 \upharpoonright \boxed{m\ j})\} \end{array} \right\} \parallel \left\{ \begin{array}{l} \{\ell \mapsto \boxed{i\ n} A_1\} \\ \text{Quicksort}(; \ell, i, n) \\ \{\exists A_3. \ell \mapsto \boxed{i\ n} A_3 \wedge \\ A_3 \sim \boxed{i\ n} A_1 \wedge \\ \text{ord}_{\leq}(A_3 \upharpoonright \boxed{i\ n})\} \end{array} \right\}$

$* \left(\begin{array}{l} \ell \mapsto \boxed{j\ \ \ i} A_1 \wedge A_1 \sim \boxed{j\ \ \ i} A_1 \wedge \\ A_1 \sim \boxed{m\ n} A_0 \wedge \boxed{m\ j\ \ \ i\ n} \wedge \\ \{A_1 \upharpoonright \boxed{m\ \ \ i}\} \leq^* \{A_1 \upharpoonright \boxed{j\ \ \ n}\} \end{array} \right) \} \exists A_1$

$\{\exists A'. \ell \mapsto \boxed{m\ n} A' \wedge A' \sim \boxed{m\ n} A_1 \wedge A_1 \sim \boxed{m\ n} A_0 \wedge$
 $\boxed{m\ j\ \ \ i\ n} \wedge \{A_1 \upharpoonright \boxed{m\ \ \ i}\} \leq^* \{A_1 \upharpoonright \boxed{j\ \ \ n}\} \wedge$
 $\text{ord}_{\leq}(A' \upharpoonright \boxed{m\ j}) \wedge \text{ord}_{\leq}(A' \upharpoonright \boxed{i\ n})\}$

$\{\exists A'. \ell \mapsto \boxed{m\ n} A' \wedge A' \sim \boxed{m\ n} A_0 \wedge \text{ord}_{\leq}(A' \upharpoonright \boxed{m\ n})\}$

where $A' = (A_2 \upharpoonright \boxed{m\ j}) \cup (A_1 \upharpoonright \boxed{j\ \ \ i}) \cup (A_3 \upharpoonright \boxed{i\ n})$.

What's Next

— Permissions

— Lacy Arrays