Reflections on, and predictions for, support systems for the development of programs

Cliff B. Jones

Computing Science Newcastle University

ASE 2008-09-19

Instead of conventional "Thank you for invite"

Huge Thank you!

My first time at an ASE: see real conversation between different approaches!

Contents

Background

Arguments

An example: ACMs

- Where to start a specification
- Splitting atoms (gently) in abstract state
- Retaining less history
- The four-slot representation
- Conclusions

Overall conclusions/summary

Me + FM support tools

- contributed to transition from VDL to VDM (language description)
 - ▶ we wrote large (including PL/I) definitions with minimal tooling
 - experience: problems mainly hit us when changes made!
 - originated much of "VDM" as for program development
- used support systems since Jim King's "Effigy"
 - I worry they lock user in to one method
 - suspect they constrain thought
 - (but used Effigy for top-down design!)
- "Formal Development Support System" IBM Hursley (1970s)
 - it was so rigid, even I couldn't use it!
- more public is the mural system (below) Manchester (1980s)
- observed use of many TP systems
 - have seen people "hack" without understanding what they are proving
- Rodin Toolset (below)

- 本部 とくき とくき とうき

mural [JJLM91]

- order of proof steps was very flexible
- a "logical frame" (e.g. used for LPF)
- focus on building theories
- but only minimal automatic support
- best seen as a UI experiment
- built from VDM spec
- implemented in SmallTalk'80 (turned out to be a mistake)
- kept multiple proof attempts difficult to delete!
- book now on the web
- yes, it contains the VDM spec (evolved)

Rodin ToolSet

- (EU) project developed tools "Rodin ToolSet"
- open source available from SourceForge
- kernel + plugins
- Eclipse based
- one key advantage: background proving
- also: nice work on computing impact of changes (minimise re-proof)
- now being used in the (EU) DEPLOY IP project
- "road map" discusses plans; invites input
- tool engenders an approach: everything in Contexts/Machines

Contents

Background

2 Arguments

An example: ACMs

- Where to start a specification
- Splitting atoms (gently) in abstract state
- Retaining less history
- The four-slot representation
- Conclusions

Overall conclusions/summary

Claim: (software) design is hard

- (Yes, I know this is stating the obvious!)
- requires a strange mixture of important (big) insights and detailed symbol pushing
- layers of abstraction (backed up by formal rules) are all we've got!
- (for many reasons) we must take our own medicine
 - reluctance to so do: Effigy, ...
- we must be seen to take our own medicine

Belief: there is a long way to go

- no current "formal methods support system" gives software engineers anything like the support given to hardware designers by their CAD systems
- destruction of design history is intellectual vandalism
- current programming languages are ill-suited to documenting design
- have to stop trying to build "complete" support systems
- build/link components
- care! there are pitfalls here (e.g. different logics)
- "whole" system includes (IMHO) tracking all versions
- ... and all tests on all versions

Thesis: level of generality

- there are all sorts of things I'd like to prove
- mistake to fix on one method (example below)
- but want more than a general purpose TP system
- there is no point in proving all of the verification conditions for one version of a program and then running a different (buggy) version so systems have to control all versions, tests, verifications, changes etc.
- might call it a "method frame"
- this can present problems diagnostics (and performance)

My hope for AI contribution

- (discussions with Ireland/Bundy)
- they planning to "mine proofs"
- loses info on how created (order) cf. mural view
- info on failed attempts long discarded!
- at detailed level, can't trace what "copied" from where
- system learns high-level strategy (not tactics?)

Argument: don't get locked into "legacy code" corner

- (I'm aware there are a lot of "testing" folk at ASE)
- BTW: I started out in IBM's Product Test division
- ideas like "abstract interpretation" ("symbolic execution"): making real progress for non-trivial systems
- handling "legacy" systems presents another set of challenges here the aim is to accumulate information such as avoidance of certain sorts of bad behaviour; again, such hard won information should not be discarded
- even if can't work on "green fields" projects, look at rational reconstructions

Theses

- model checking not only needs "abstraction" but it should be equipped to use ones that are available from design
- there are enough common problems between the various sorts of tool that interfacing components is imperative — apart from simple syntactic interfaces
- ... much in the style of the EPFL paper (Beyer?) Wednesday morning ("predicate abstraction" + "explicit analysis")
- such integration can pose hard semantic challenges

Idea: direct support for SOS was almost subject of ASE conf paper (now [HJ08])

- do not have complete axiom systems for any widely used programming language (by a big margin)
- might therefore have to reason from, say, an operational semantics
- our paper builds on mural approach
- obviously use Floyd/Hoare-like rules is applicable
- in fact, would be nice if this system supports justification of such rules

Broader worries: industrial perspective

- getting the "right" specification [JHJ07] for a non-trivial system is at least as much an issue as showing that a design matches its specification
- even during design, everything will change (in fact, designing for flexibility is often more important than aiming for efficiency) systems must maximise what is preserved over such changes
- we have to build our tools so that they can interface with whatever in-house engineering systems are being used by organisations we expect to adopt our formal tools

• . . .

Contents

Background

Arguments

3

An example: ACMs

- Where to start a specification
- Splitting atoms (gently) in abstract state
- Retaining less history
- The four-slot representation
- Conclusions

Overall conclusions/summary

Key abstractions

Argument: flexibility on methods

- \bullet Pre/post-conditions (as in VDM/B/...)
 - design by sequential "operation decomposition rules"
 - Floyd/Hoare-like rules (coping with relational post-conditions)
- Rely/Guarantee "thinking"
 - not (just) a specific set of rules
 - show importance of "frames" (cf. Separation Logic)
 - using "auxiliary variables"
- Abstract objects
 - choice of abstract data objects key for specifications
 - data "reification" (classic-VDM / Nipkow's rule)
 - link with R/G development
- "fiction of atomicity"
 - "splitting (software) atoms safely" [Jon07]
 - ▶ cf. database transactions [JLRW05], ...
 - cf. POBL [Jon96]

・ 同 ト ・ ヨ ト ・ ヨ ト

While (operation decomposition) rule

$$S \text{ sat } (P \land b, P \land W)$$

$$\underline{P \Rightarrow \delta_l(b)}$$

$$\underline{While\text{-}I} \xrightarrow{P \Rightarrow \delta_l(b)} \text{sat } (P, P \land \neg b \land W^*)$$

There's no reason why a system should hardwire the (standard) Hoare rule the rules should be data (to a method frame)

"posit and prove" is one way of supporting design; "Verified by Construction" has been shown to be viable for large systems

One R/G rule

$$\begin{array}{c} \{P, R \lor Gr\} \vdash sl \text{ sat } (Gl, Ql) \\ \{P, R \lor Gl\} \vdash sr \text{ sat } (Gr, Qr) \\ Gl \lor Gr \Rightarrow G \\ \hline \hline \hline P \land Ql \land Qr \land (R \lor Gl \lor Gr)^* \Rightarrow Q \\ \hline \left[Par\text{-}I\right] \hline \{P, R\} \vdash mk\text{-}Par(sl, sr) \text{ sat } (G, Q) \end{array}$$

... and there are lots more where this one came from

Image: A match a ma

ラト

Subtle link between R/G and data reification cf. [Jon07]

• in *FINDP*

- we have $t \leftarrow min(t, local)$ in n parallel processes
- assuming we don't want to "lock" t
- need a representation that preserves R/G conditions
- ▶ simple to represent as t as min(et, ot)

SIEVE

- \blacktriangleright we have to remove an element from a set s
- assuming we don't want to "lock" s (big!)
- need a representation that preserves R/G conditions $s \subseteq \overleftarrow{s}$
- (less obvious) represent s as a bit vector

Simpson

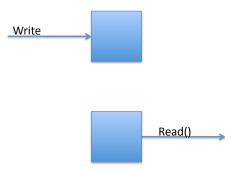
- extremely interesting
- my claim: this is the essence of Simpson's contribution

ACMs: [JP08] Communication (Atomic?)



ACMs

Atomic and (trying for) Asynchronous



Simpson's algorithm

- Simpson's algorithm
- several other folk still working on this
- run through my "rational reconstruction"
 - "explanation" via layers of abstraction
- essential to get the big steps right before detailed proof
- apologies for so much argument about eight lines of code ...

formulae in small fount not meant to be read!

Specification

 $\Sigma^a :: data-w: Value^*$ fresh-w: \mathbb{N} hold-r: \mathbb{N}

 $\begin{array}{l} {\rm inv} \ (mk \cdot \Sigma^a(data \cdot w, fresh \cdot w, hold \cdot r)) \triangleq \\ fresh \cdot w, hold \cdot r \in \{1.. {\rm len} \ data \cdot w\} \wedge hold \cdot r \leq fresh \cdot w \end{array}$

$$\sigma_0^a = mk\text{-}\Sigma^a([\mathbf{x}], 1, 1)$$

```
while true do

start-Write(v: Value): data-w \leftarrow data-w \frown [v];

commit-Write(): fresh-w \leftarrow len data-w

od

while true do

start-Read(): hold-r \leftarrow fresh-w;

end-Read()r: Value: r \leftarrow data-w(i) for some i \in \{hold-r..fresh-w\}

od
```

通 ト イヨ ト イヨト

Example 3

start-Read() start-Write(y) commit-Write() start-Write(z) commit-Write() end-Read() start-Read() end-Read()

- - $\dots \quad mk \cdot \Sigma^a([\mathbf{x}, \mathbf{y}, \mathbf{z}], 2, 1)$
 - .. $mk \Sigma^{a}([x, y, z], 3, 1)$
 - $r \in \{x, y, z\}$
 - .. $mk \Sigma^a([x, y, z], 3, 3)$

 $\dots r = z$

Specification in terms of four sub-operations (Write)

Atomic operations — therefore pure pre/post specification

while true do

```
start-Write(v: Value): data-w \leftarrow data-w \cap [v];
commit-Write(): fresh-w \leftarrow len data-w
od
```

A D A D A D A

Specification in terms of four sub-operations (Read)

while true do

 $start-Read(): hold-r \leftarrow fresh-w;$ $end-Read()r: Value: r \leftarrow data-w(i)$ for some $i \in \{hold-r..fresh-w\}$ od

 $\begin{array}{l} Read()r: Value\\ \textbf{local} \ hold-r: \mathbb{N}\\ start-Read()\\ \textbf{wr} \ hold-r\\ \textbf{rd} \ fresh-w\\ \textbf{post} \ hold-r = fresh-w\\ end-Read()r: Value\\ \textbf{rd} \ data-w, fresh-w\\ \textbf{post} \ \exists i \in \{hold-r.., fresh-w\} \cdot r = data-w(i) \end{array}$

Image: A mathematical states and a mathem

General messages

- note "algorithmic" specification
- "fiction of atomicity"
 - but single "atomic" variable does not cover all behaviour
- "frames" (for rd/wr access)
 - plus "local"
- data abstraction

Splitting atoms in Σ^a (*Write*)

Accept overlap (only read/write) — therefore rely/guarantee

Write(v: Value)start-Write(v: Value) **rd** fresh-w wr data-w $\textbf{rely } \textit{fresh-w} = \overleftarrow{\textit{fresh-w}} \land \textit{data-w} = \overrightarrow{\textit{data-w}}$ guar $\{1..fresh-w\} \triangleleft data-w = \{1..fresh-w\} \triangleleft data-w$ post $data-w = \overline{data-w} \frown [v]$ *commit-Write*(v: Value) rd data-w wr fresh-w pre data-w(len data-w) = vrely $fresh-w = fresh-w \wedge data-w = data-w$ **post** fresh-w = len data-w

Splitting atoms in Σ^a (*Read*)

$$\begin{aligned} & Read()r: Value \\ & start-Read() \\ & \mathbf{rd} \ fresh-w \\ & \mathbf{wr} \ hold-r \\ & \mathbf{rely} \ hold-r = \overleftarrow{hold-r} \\ & \mathbf{post} \ hold-r \in \{\overleftarrow{fresh-w}, fresh-w\} \\ & end-Read()r: Value \\ & \mathbf{rd} \ data-w, fresh-w, hold-r \\ & \mathbf{rely} \ hold-r = \overleftarrow{hold-r} \land \forall i \in \{hold-r..\overbrace{fresh-w}\} \cdot data-w(i) = \overleftarrow{data-w}(i) \\ & \mathbf{post} \ \exists i \in \{hold-r..\overbrace{fresh-w}\} \cdot r = \overleftarrow{data-w}(i) \end{aligned}$$

3

→ Ξ →

A = A + A = A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A

General messages

- phasing
 - ▶ makes clear *start-Write* cannot interfere with *commit-Write*
 - avoids implications in rely conditions
- $\bullet\,$ frames plus phasing significantly simplify R/G assertions
- cf. *rely-start-Write* on Σ^a above

Retaining less history

A data reification exercise — still very general

$$\begin{array}{rcl} \Sigma^i & :: & data{-}w{:} X & \stackrel{m}{\longrightarrow} Value \\ & fresh{-}w{:} X \\ & hold{-}r{:} X \\ & hold{-}w{:} X \end{array}$$

 $\sigma_0^i = mk\text{-}\Sigma^i(\{\alpha \mapsto \mathbf{x}\}, \alpha, \alpha, \alpha)$

< A > < 3

Relating Σ^i to Σ^a Using Nipkow's rule

 $r(\sigma_1^a,\sigma_1^i) \wedge \textit{post}^i(\sigma_1^i,\sigma_2^i) \ \Rightarrow \ \exists \sigma_2^a \in \Sigma^a \cdot \textit{post}^a(\sigma_1^a,\sigma_2^a) \wedge r(\sigma_2^a,\sigma_2^i)$

 $\begin{aligned} r: \Sigma^{a} \times \Sigma^{i} &\to \mathbb{B} \\ r(mk - \Sigma^{a}(data - w^{a}, fresh - w^{a}, hold - r^{a}), \\ mk - \Sigma^{i}(data - w^{i}, fresh - w^{i}, hold - r^{i}, hold - w^{i})) & \triangleq \\ \mathbf{rng} \ data - w^{i} &\subseteq \mathbf{elems} \ data - w^{a} \wedge \\ data - w^{a}(fresh - w^{a}) &= data - w^{i}(fresh - w^{i}) \wedge \\ data - w^{a}(hold - r^{a}) &= data - w^{i}(hold - r^{i}) \end{aligned}$

Specifications of the sub-operations on Σ^i

Still overlapped — still rely/guarantee

 $\begin{array}{l} \text{Write}(v: Value) \\ \text{local } hold-v: X \\ start-Write(v: Value) \\ \text{rd } hold-r, fresh-w \\ \text{wr } data-w, hold-w \\ \\ \text{guar } \{hold-r, hold-r\} \triangleleft data-w = \{hold-r, hold-r\} \triangleleft data-w \\ \text{guar } \{hold-r, hold-r\} \triangleleft data-w = \{hold-r, hold-r\} \triangleleft data-w \\ \text{post } hold-w \in (X - \{fresh-w, hold-r, hold-r\}) \land data-w = data-w \uparrow \{hold-w \mapsto v\} \\ commit-Write(v: Value) \\ \text{rd } data-w, hold-w \\ \text{wr } fresh-w \\ \text{pre } data-w(hold-w) = v \\ \text{rely } fresh-w = fresh-w \land data-w = data-w \\ \text{post } fresh-w = hold-w \\ \end{array}$

Specifications of the sub-operations on Σ^i

 $\begin{aligned} Read()r: Value \\ start-Read() \\ rd fresh-w \\ wr hold-r \\ rely hold-r = hold-r \\ post hold-r \in \{fresh-w, fresh-w\} \\ end-Read()r: Value \\ rd hold-r, data-w \\ rely hold-r = hold-r \land data-w(hold-r) = data-w(hold-r) \\ post r = data-w(hold-r) \end{aligned}$

(日) (同) (三) (三)

General messages

- simpler R/G because of read/write frames
- data reification
 - (potentially) reducing non-determinism
 - use of VDM's other reification rule
- still have "bold" atomicity assumptions
 - \blacktriangleright couldn't update data-w atomically on any reasonable machine
- still work to be done
- role of data reification in achieving rely conditions
- Simpson's representation crucial

The four-slot representation

Focus on Simpson's inspiration

$$\Sigma^{r} :: data-w: P \times S \xrightarrow{m} Value$$

$$pair-w: P$$

$$pair-r: P$$

$$slot-w: P \xrightarrow{m} S$$

$$wp-w: P$$

$$ws-w: S$$

$$rs-r: S$$

where (key assumptions about granularity (ρ)):

P, S = Token-set

$$P = S$$

card $P = 2$
 $\rho(i) \neq i$

< A D > < D >

Specifications of the sub-operations on Σ^r

Satisfies guarantee conditions (as well as post)

 $\begin{aligned} & Write(v: Value) \\ & \textbf{local } wp\text{-}w: P \\ & \textbf{local } ws\text{-}w: S \\ & wp\text{-}w \leftarrow \rho(pair\text{-}r); \\ & ws\text{-}w \leftarrow \rho(slot\text{-}w(wp\text{-}w)); \\ & data\text{-}w(wp\text{-}w, ws\text{-}w) \leftarrow v; \\ & slot\text{-}w(wp\text{-}w) \leftarrow ws\text{-}w; \\ & pair\text{-}w \leftarrow wp\text{-}w \end{aligned}$

 $\begin{aligned} Read()r: Value\\ \textbf{local } rs\text{-}r: S\\ pair\text{-}r \leftarrow pair\text{-}w;\\ rs\text{-}r \leftarrow slot\text{-}w(pair\text{-}r);\\ r \leftarrow data\text{-}w(pair\text{-}r, rs\text{-}r) \end{aligned}$

Conclusions (on example)

- all at ASE probably accept "refinement from abstractions"
- "splitting atoms" a new/old formal addition
- subsidiary points
 - rely/guarantee "thinking"
 - remember frame descriptions
 - combination with data reification
 - link with "phasing"
 - "auxiliary variables" + Nipkow's rule
 - tool support
 - try to avoid "coding logic into values"
- these ideas are not (all) in any single "method"

Contents

Background

2 Arguments

An example: ACMs

- Where to start a specification
- Splitting atoms (gently) in abstract state
- Retaining less history
- The four-slot representation
- Conclusions

Overall conclusions/summary

(My) general conclusions

- OK, tools *do* matter
- no one method solves covers all problems
- must design interworking components
 - XML is not the answer!
 - generality: diagnostics via rules?
 - ... performance via abstract interpretation?
- I hope to explore "method frame"
 - flexibility
 - way to combine
- GUI does matter
 - view onto huge data structure
 - much of which generated
 - quick/easy check to avoid wasting time trying to prove non-theorems?
- Programming Languages part of the problem (not solution)
 - "must try harder"
 - old TOPD question

Personal preferences

- (post this meeting) I still plan to work on Verification!
- a couple more examples
 - ▶ I got into "formal methods" (1969) because of PL/I-F compiler mess
 - my attempts to prove extant programs always failed
 - concede: I didn't have the good tools available here at ASE 2008
 - but: finding errors late still leaves "scrap and rework" issue
- real message: continue to search for synergy
 - I happen to be on Verification side of the fence
 - I do see the payoff with model checking etc.
 - Confess: VxC suites my personal research tastes!
- ... and I hope to work with AI people!

References



John R. D. Hughes and C. B. Jones.

Reasoning about programs via operational semantics: Requirements for a support system. Automated Software Engineering, 10.1007/s10515-008-0036-6, 2008.



Cliff B. Jones, Ian J. Hayes, and Michael A. Jackson.

Deriving specifications for systems that are connected to the physical world.

In Cliff B. Jones, Zhiming Liu, and Jim Woodcock, editors, Formal Methods and Hybrid Real-Time Systems: Essays in Honour of Dines Bjørner and Zhou Chaochen on the Occassion of Their 70th Birthdays, volume 4700 of Lecture Notes in Computer Science, pages 364–390. Springer Verlag, 2007.



C. B. Jones, K. D. Jones, P. A. Lindsay, and R. Moore.

mural: A Formal Development Support System. Springer-Verlag, 1991.



C. B. Jones, D. Lomet, A. Romanovsky, and G. Weikum.

The atomicity manifesto.

Journal of Universal Computer Science, 11(5):636-650, 2005.



C. B. Jones.

Accommodating interference in the formal design of concurrent object-based programs. Formal Methods in System Design, 8(2):105–122, March 1996.



C. B. Jones.

Splitting atoms safely. Theoretical Computer Science, 357:109–119, 2007.



Cliff B. Jones and Ken G. Pierce.

Splitting atoms with rely/guarantee conditions coupled with data reification. In *ABZ2008*, volume LNCS 5238, pages 360–377, 2008.

< 🗇 🕨

3 ×

ASE community might also like VSTTE (because the "E" is for "experiments") Toronto, October $6^{th}-9^{th}$

http://qpq.csl.sri.com/vsr/vstte-08