Computational Modeling and Analysis of Complex Systems (CMACS)

"The Reduced Product of Abstract Domains and the Combination of Decision Procedures" (1.) and "Termination: Foundations using Abstract Interpretation" (2.) Patrick Cousot

pcousot@cs.nyu.edu <u>http://cs.nyu.edu/~pcousot/</u> Joint ongoing work with

- I. Radhia Cousot and Laurent Mauborgne
- 2. Radhia Cousot and Andreas Podelski

**CMACS NYU** meeting

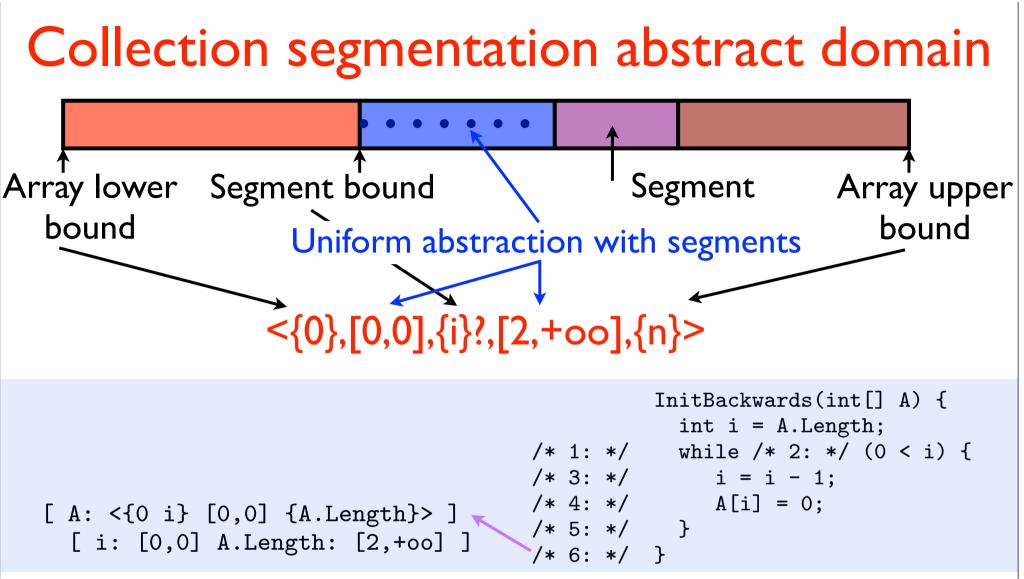
October 29, 2010

#### What has been done since Pittsburgh meeting

#### • Work since the Pittsburgh meeting:

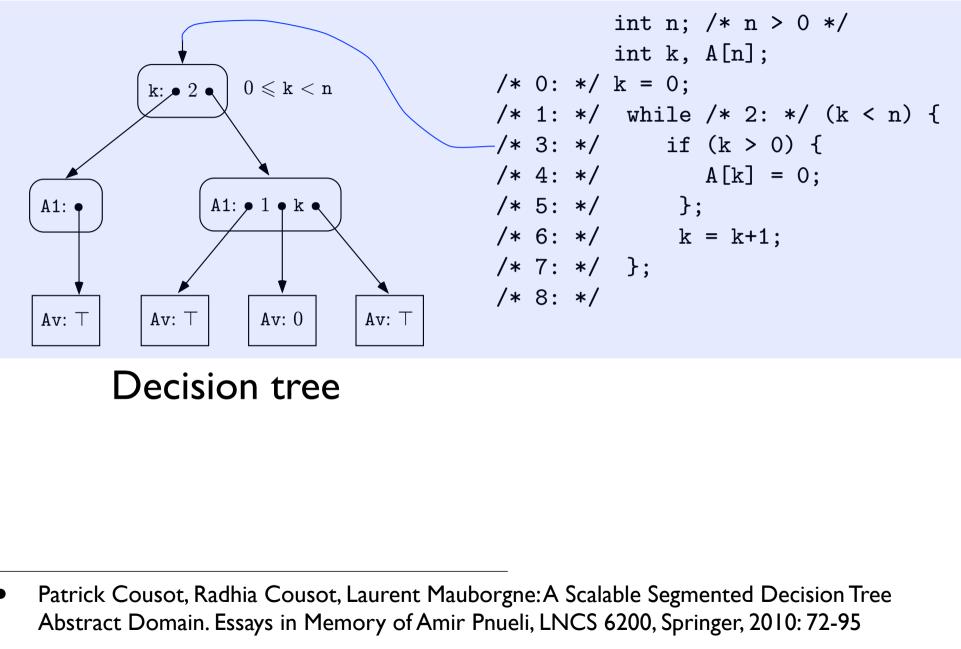
- Array content analysis (joint work with R. Cousot and F. Logozzo)
- Segmented decision tree abstract domain (joint work with R. Cousot and L. Mauborgne)
- Precondition inference from runtime-checked assertions (joint work with R. Cousot and F. Logozzo)
- Work in progress (today's presentations):
  - Probabilistic abstract interpretation: see talk by Michael Monerau
  - Logical abstract domains
  - Termination/Guarantee semantics, proof, and static analysis

# Work on Al since the Pittsburgh meeting



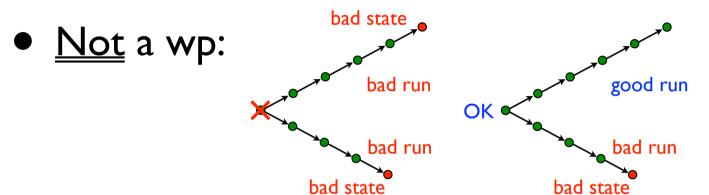
- Included by F. Logozzo in MSR Clousot (distributed with MS Visual Studio under Windows 7 pro)
- To appear in POPL'2011

#### Segmented decision tree abstract domain

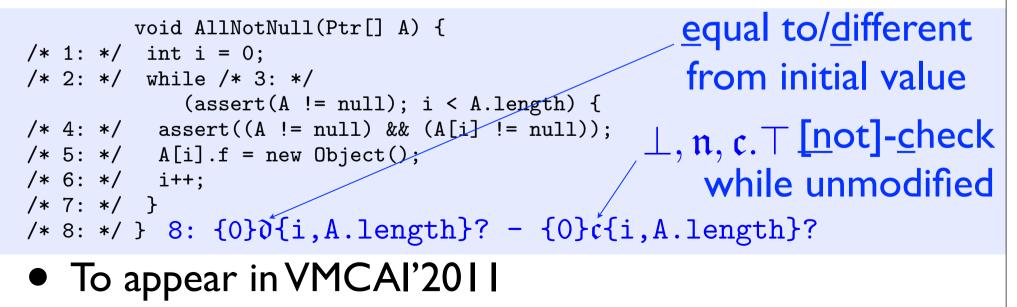


#### Preconditionerinference from asserts

• Derive a static precondition from programmers and languages runtime-checked assertions in the code



• Symbolic under-approximation:



# Ongoing work (1) Logical abstract domains

#### Combining Algebraic and Logical Abstractions (I)

- Model-checking is "logical" (temporal logic, BDDs, SMT solvers,...)
- Abstract interpretation is "algebraic" (orders, lattices, linear algebra, categories, reduced product, ...)
- MC & Al can be combined within "set theory", e.g.
  - Patrick Cousot, Radhia Cousot: Temporal Abstract Interpretation. POPL 2000, 12-25.
  - Patrick Cousot, Radhia Cousot: Refining Model Checking by Abstract Interpretation. Autom. Softw. Eng. 6(1): 69-95 (1999).

#### Combining Algebraic and Logical Abstractions (II)

 We propose a new MC & AI combination as "logical (i.e. SMT solvers) + algebraic (i.e. reduced product of abstract domains)" Logical abstract domains: an instance of algebraic abstract domains

- Abstract properties: a theory (set of logical formulae)
  - Order =>, join: \/, meet: /\, ...
  - Concretization γ: interpretation
  - Abstraction α: in general does not exist (no best abstraction e.g. in absence of infinite conjunctions)
- Transformers:
  - Forward: Floyd/sp
  - Backward: Hoare/wp/wlp

$$(\mathbf{x} = 0) \Rightarrow (\mathbf{x} = 0 \lor \mathbf{x} = 1) \Rightarrow \dots \Rightarrow \bigvee_{i=1}^{n} \mathbf{x} = i \Rightarrow \dots$$
$$(\mathbf{x} \neq -1) \Leftarrow (\mathbf{x} \neq -1 \land \mathbf{x} \neq -2) \Leftarrow \dots \Leftarrow \bigwedge_{i=1}^{n} \mathbf{x} \neq -i \Leftarrow \dots$$

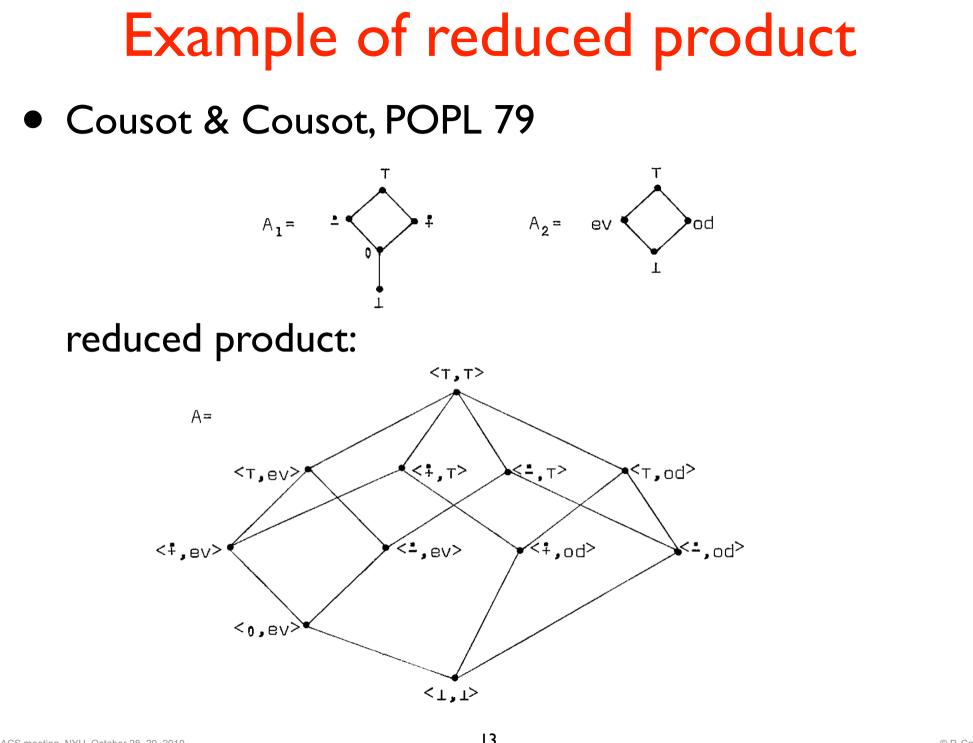
# The (iterated) reduced product in Al

#### **Reduced product**

- A Cartesian product of abstract domains:  $\prod_{i=1,...,n} A_i$
- Understood as a conjunction:

 $\gamma(a_1,...,a_n) = / \sum_{i=1,...,n} \gamma(a_i)$ 

- Implemented as a Cartesian product plus a reduction
   p to propagate shared information from one
   component to another
- Sound and complete



#### Iterated pairwise reduction

- pij : reduction between abstract domains Ai and Aj
- $\vec{\rho}_{ij}$  : extension to the Cartesian product

 $\vec{\rho}_{ij}(a_1,...,a_n) = (a_1,...,a'_{i},...,a'_{j},...,a_n)$ 

where  $(a'_i, a'_j) = \rho_{ij}(a_i, a_j)$ 

- Iterated pairwise reduction
  - $\vec{\rho}^*$  = iterate the  $\vec{\rho}_{ij}$ , i,j=1,...,n, i≠j until convergence
- Sound (but in general incomplete)

#### Example of pairwise iterated reduction

- Concrete domain:  $L = \mathcal{O}(\{a, b, c\})$
- Abstract domains:  $A_1 = \{\emptyset, \{a\}, \top\}$ 
  - $T = \{a, b, c\} \qquad A_2 = \{\emptyset, \{a, b\}, T\} \\ A_3 = \{\emptyset, \{a, c\}, T\}$
- Reduction of  $\langle \top, \{a, b\}, \{a, c\} \rangle$ :
- Global:  $\langle \{a\}, \{a, b\}, \{a, c\} \rangle$
- **Pairwise reductions:**  $\vec{\rho}_{ij}(\langle \top, \{a, b\}, \{a, c\} \rangle) = \langle \top, \{a, b\}, \{a, c\} \rangle$  for  $\Delta = \{1, 2, 3\}, i, j \in \Delta, i \neq j$
- Iterated reduction:

 $\vec{\rho}^*(\langle \top, \{a, b\}, \{a, c\}\rangle) = \langle \top, \{a, b\}, \{a, c\}\rangle$ 

# The Nelson-Oppen combination procedure

#### **Objective of the Nelson-Oppen procedure**

- Given deductive theories *T<sub>i</sub>* in F(Σ<sub>i</sub>), Σ<sub>i</sub> ⊆ Σ
   with equality and decision procedures sat<sub>i</sub> for satisfiability of quantifier free conjunctive formulæ *φ<sub>i</sub>* ∈ C(Σ<sub>i</sub>), *i* = 1, ..., *n*,
- Decide the satisfiability of a quantifier free conjunctive formula  $\varphi \in \mathbb{C}(\bigcup_{i=1}^{n} \Sigma_{i})$  in theory  $\mathcal{T} = \bigcup_{i=1}^{n} \mathcal{T}_{i}$  such that  $\mathfrak{M}(\mathcal{T}) = \bigcap_{i=1}^{n} \mathfrak{M}(\mathcal{T}_{i})$ .

[23] G. Nelson and D. Oppen. Simplification by cooperating decision procedures. *TOPLAS*, 1(2):245–257, 1979.

#### The Nelson-Oppen combination procedure

I. Purification: project the quantifier-free conjunctive formula  $\varphi$  as an equi-satisfiable conjonction of component formulæ in each theory by introducing fresh variables for alien terms

 $\varphi' = \exists \vec{x}_1, \ldots, \vec{x}_n : \bigwedge_{i=1}^n \varphi_i \text{ where } \varphi_i = \varphi'_i \wedge \bigwedge_{x_i \in \vec{x}_i} x_i = t_{x_i},$ 

- 2. Repeat the equality reduction: propagate [dis] equalities deduced from each component formula  $\varphi_i$ to the other components formulæ) until no new [dis] equality can be added
- 3. Test satisfiability of component formulæ, unsatifiable iff one is unsatisfiable else unknown (originally, satisfiable if all component formula are satisfiable)

# The Nelson-Oppen procedure is an iteratively reduced observation product

- The purification is a projection of the formula to an observation product (with auxiliary variables observing alien subterms)
- The reduction is iterative but only for [dis]equalities
- The unsatisfiability check is a reduction to (false)

#### Soundness of the procedure ?

- The unsatisfiability is sound
- More conditions for satisfiability soundness to ensure that all theories have isomorphic models such as
  - stably-infinity, politeness, etc ... so as to ensure that the models of the theories  $\mathcal{T}_i$  have the same cardinalities
  - shared symbols (e.g. equality) have isomorphic interpretations in all theories sharing them or theories are disjoint which avoids the problem

#### Completeness of the procedure ?

• The procedure is incomplete

so there exists formulæ satisfiable in two theories but not in their combination (e.g. integer arithmetics and bit vectors)

- Additional restrictions are necessary to ensure completeness
  - convexity (to avoid to have to reduce by disjunctions of [dis]equalities)
  - disjointness of the theories (but constants, to avoid to have to reduce on other properties than [dis] equality such as <)</li>

#### Who cares about completeness in static analysis?

- We care about soundness but not on completeness (since we always get a sound overapproximation)
- Abandoning completeness, we can
  - combine theories sharing symbols other than = (as signs and parity)
  - perform reduction (even for non-convex theories) that are simply not optimal

#### Combining logical and algebraic abstractions

We use an iteratively reduced observation product with:

- logical components in logical abstract domains sharing symbols and handled by SMT solvers
- algebraic components in algebraic abstract domains
- the reduction propagates
  - [dis]equalities of logical components to all other components
  - pairwise algebraic reductions (equalities and others) to all other components

#### Perspectives

- A new perspective to combine
  - SMT solvers based model-checking understood as logical abstract domains (with logical widenings)
  - abstract interpretation-based static analysis using classical abstract domains (with algebraic widenings)
- This might avoid costly iterative refinement methods thanks to the expressivity of first-order logic

# Ongoing work (2) Termination

#### Basic idea:

### Apply the abstract interpretation framework to termination

#### Abstract Interpretation framework

- Define the standard semantics:  $\langle \Sigma, \tau \rangle$
- Define the collecting semantics (most general property of interest):  $C \in \wp(C)$
- Express the collecting semantics in fixpoint form:  $\mathcal{C} = \mathsf{lfp}^{\subseteq} F \in \wp(C)$
- Finite (MC) : compute C iteratively;
- Infinite (AI) : define an abstraction:

$$\langle \wp(C), \subseteq \rangle \xleftarrow{\gamma} \langle A, \sqsubseteq \rangle$$

Abstract Interpretation framework (cont'd)

• Define an abstract transformer:

 $\alpha \circ F \circ \gamma \sqsubseteq \overline{F}$ 

- The fixpoint abstract semantics is sound:  $\alpha(\mathsf{lfp}^{\subseteq} F) \sqsubseteq \mathsf{lfp}^{\sqsubseteq} \overline{F}$
- Compute the abstract iterates iteratively:

$$\overline{F}^0 \triangleq \bot, \dots, \overline{F}^{n+1} \triangleq \overline{F}(\overline{F}^n), \dots$$

 Accelerating the convergence by widening narrowing (when necessary)

#### Termination analysis:

Applying the abstract interpretation framework to a termination collecting semantics

#### Standard semantics

• Traces on the set of states  $\Sigma$  :

- Traces of length n:  $\vec{s} = \vec{s}_0 \vec{s}_1 \dots \vec{s}_{n-1} \in \vec{\Sigma}^n$
- Finite traces:
- Infinite traces:

$$\vec{\Sigma}^{+} \triangleq \bigcup_{n \ge 1} \vec{\Sigma}^{n}$$
$$\vec{s} = \vec{s}_0 \vec{s}_1 \dots \vec{s}_i \vec{s}_{i+1} \dots \in \vec{\Sigma}^{\omega}$$

• Trace semantics:  $S_T = \langle \Sigma, \text{ init}, \text{ final}, \vec{T} \rangle$ 

• finite runs:

 $\forall n \ge 1 : \forall \vec{s} \in \vec{T} \cap \vec{\Sigma}^n : \vec{s}_0 \in \mathsf{init} \land \forall i \in [0, n-1(: \vec{s}_i \notin \mathsf{final} \land \vec{s}_{n-1} \in \mathsf{final})$ 

#### • Infinite runs:

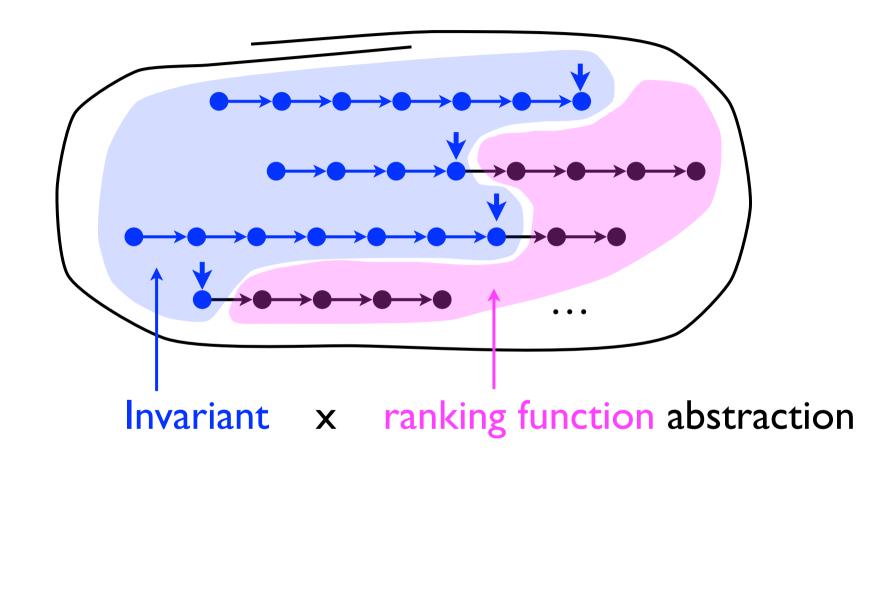
 $\forall \vec{s} \in \vec{T} \cap \vec{\Sigma}^{\,\omega} : \vec{s}_0 \in \text{init} \land \forall i \ge 0 : \vec{s}_i \notin \text{final}$ 

Example: traces generated by a transition system

- Transition system:  $\langle \Sigma, \tau \rangle$
- Trace semantics:  $S_{T}[\tau] = \langle \Sigma, \text{ init}, \text{ final}, \vec{T} \rangle$
- Generated by the transition system:

$$\forall \vec{s} \, ss' \vec{s}' \in \vec{T} : \tau(s, s')$$

#### À la Floyd/Turing invariant/ranking function abstraction



#### Past/future abstraction

- Past:  $\alpha_{\leftarrow}(\vec{T}) \triangleq \{\vec{s} \in \vec{\Sigma}^+ \mid \exists \vec{s}' \in \vec{\Sigma}^{\infty} : \vec{s} \vec{s}' \in \vec{T}\}$  $\mathcal{S}_{\leftarrow} \triangleq \langle \Sigma, \text{ init}, \text{ final}, \alpha_{\leftarrow}(\vec{T}) \rangle$
- Future:  $\alpha_{\rightarrow}(\vec{T}) \triangleq \{\vec{s}' \in \vec{\Sigma}^{\infty} \mid \exists \vec{s} \in \vec{\Sigma}^* : \vec{s} \, \vec{s}' \in \vec{T}\}$  $\mathcal{S}_{\rightarrow} \triangleq \langle \Sigma, \text{ init, final, } \alpha_{\rightarrow}(\vec{T}) \rangle$

#### Past fixpoint semantics

• Past fixpoint semantics:

$$\begin{aligned}
\mathcal{B}_{\leftarrow} \llbracket \tau \rrbracket &\in \wp(\vec{\Sigma}^{+}) \longrightarrow \wp(\vec{\Sigma}^{+}) \\
\mathcal{B}_{\leftarrow} \llbracket \tau \rrbracket(\vec{X}) &\triangleq \operatorname{init}^{1} \cup \vec{X} \circ \vec{\tau} \\
\mathcal{S}_{\leftarrow} \llbracket \tau \rrbracket &= \langle \Sigma, \operatorname{init}, \operatorname{final}, \operatorname{lfp}^{\subseteq} \mathcal{B}_{\leftarrow} \llbracket \tau \rrbracket \rangle
\end{aligned}$$

• Further abstractions yield invariants (in fixpoint form):

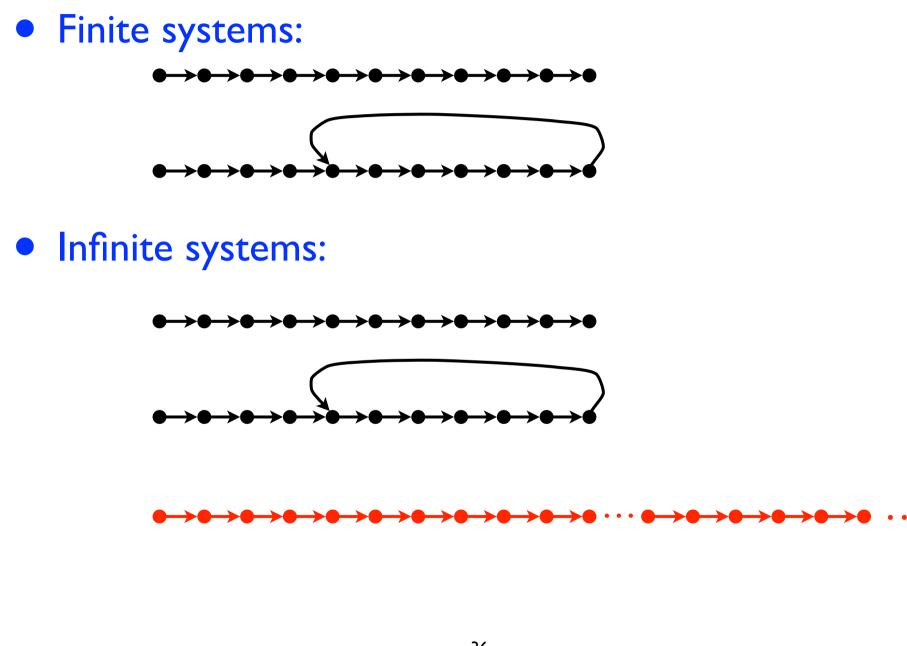
$$\begin{array}{ll} \alpha_i(\vec{T}) & \triangleq & \{\vec{s}_{n-1} \mid n \geqslant 1 \land \vec{s} \in \vec{T} \cap \vec{\Sigma}^n\} \\ \mathcal{S}_i & \triangleq & \langle \Sigma, \text{ init}, \text{ final}, \, \alpha_i \circ \alpha_\leftarrow(\vec{T}) \rangle \end{array}$$

 and automatic static analysis (iterative fixpoint computation with convergence acceleration by widening/narrowing)

#### Future fixpoint semantics Computational ordering $\vec{X} \ \vec{\Box} \ \vec{Y} \ \triangleq \ (\vec{X} \cap \vec{\Sigma}^+) \subset (\vec{Y} \cap \vec{\Sigma}^+) \land (\vec{X} \cap \vec{\Sigma}^\omega) \supset (\vec{Y} \cap \vec{\Sigma}^\omega)$ $\langle \wp(\vec{\Sigma}^{\infty}), \vec{\Box}, \vec{\Sigma}^{\omega}, \vec{\Sigma}^{+}, \vec{\sqcup}, \vec{\sqcap} \rangle$ is a complete lattice • Future fixpoint (termination collecting) semantics: $\mathcal{B}_{\rightarrow}\llbracket \tau \rrbracket \in \wp(\vec{\Sigma}^{\infty}) \longrightarrow \wp(\vec{\Sigma}^{\infty})$ $\mathcal{B}_{\rightarrow} \llbracket \tau \rrbracket (\vec{X}) \triangleq \text{final}^1 \cup (\vec{\tau} \circ \vec{X})$ $\mathcal{S}_{\rightarrow} \llbracket \tau \rrbracket \triangleq \langle \Sigma, \text{ init, final, } \mathsf{lfp}^{\overrightarrow{\sqsubseteq}} \mathcal{B}_{\rightarrow} \llbracket \tau \rrbracket \rangle$

Patrick Cousot: Constructive design of a hierarchy of semantics of a transition system by abstract interpretation. <u>Theor. Comput. Sci. 277(1-2)</u>: 47-103 (2002)

#### Future of finite versus infinite systems



#### Future approximation strategies

• Under-approximation of the termination domain:  $\mathrm{dmn}[\neg(\vec{T} \cap \vec{\Sigma}^{\,\omega})]$ 

 $\mathsf{dmn}[\vec{T}] \triangleq \{s \in S \mid \exists \vec{s} : s\vec{s} \in \vec{T}\}$ 

- Dual abstract interpretation (over-approximation of the complement)
- Extremely difficult
- Few known solutions (testing, bounded model-checking, symbolic execution, etc.), mostly ineffective
- Over-approximation of the termination argument:
  - Follow Lyapunov (stability), Turing, Floyd (ranking functions), Burstall, Ramsey, ...

#### The ranking abstraction

#### • Ordinals:

 $0 \triangleq \emptyset, 1 \triangleq \{0\}, 2 \triangleq \{0, 1\}, \dots, n+1 \triangleq \{0, \dots, n\}, \dots, \omega \triangleq \bigcup_{\delta < \omega} \delta, \omega + 1, \dots$ 

#### • Ranking abstraction:

$$\begin{array}{ll} \alpha_r(\vec{T}) & \triangleq & \{ \langle \vec{s}_0, \, 0 \rangle \mid \vec{s} \in \vec{T} \cap \vec{\Sigma}^1 \} \\ & \cup \{ \langle s, & \bigcup_{ss'\vec{s} \in \vec{T} \land \langle s', \, \delta \rangle \in \alpha_r(\vec{T}) } \delta + 1 \rangle \mid \exists \vec{s}' \in \vec{\Sigma}^{\, \alpha} : s\vec{s}' \in \vec{T} \} \\ & ss'\vec{s} \in \vec{T} \land \langle s', \, \delta \rangle \in \alpha_r(\vec{T}) \\ \mathcal{S}_r & \triangleq & \langle \Sigma, \text{ init, final, } \alpha_r \circ \alpha_{\rightarrow}(\vec{T}) \rangle \end{array}$$

#### Ancestors abstraction

- Abstract a partial function by its domain of definition:
  - $\begin{array}{lll} \alpha_{a}(f) & \triangleq & \mathsf{dmn}[f] \\ \mathcal{S}_{a} & \triangleq & \langle \Sigma, \text{ init}, \text{ final}, \ \alpha_{a} \circ \alpha_{r} \circ \alpha_{\rightarrow}(\vec{T}) \rangle \end{array}$

• We get pre[t\*] (final) ()

#### (1) P. Cousot, Thesis, Grenoble, March 1978

#### Fixpoint ranking semantics

$$\begin{split} \mathcal{B}_{r}\llbracket\tau\rrbracket &\in (\Sigma \not\rightarrow \mathbb{O}) \longrightarrow (\Sigma \not\rightarrow \mathbb{O}) \\ \mathcal{B}_{r}\llbracket\tau\rrbracket(X) &\triangleq \{\langle s, 0 \rangle \mid s \in \mathsf{final}\} \\ & \cup \{\langle s, \bigcup_{\tau(s,s') \land \langle s', \delta \rangle \in X} \delta + 1 \rangle \mid s \in \mathsf{pre}\llbracket\tau\rrbracket(\mathsf{dmn}[X])\} \\ \mathcal{S}_{r}\llbracket\tau\rrbracket &= \langle \Sigma, \mathsf{init}, \mathsf{final}, \mathsf{lfp}^{\subseteq} \mathcal{B}_{r}\llbracket\tau\rrbracket \rangle \end{split}$$

Proof

```
\alpha_r(\mathcal{B}_{\to}\llbracket\tau\rrbracket(X))
```

=...

 $= \mathcal{B}_r[\![\tau]\!](\alpha_r(X))$ 

 $\operatorname{def.} \mathcal{B}_r[\tau]$ 

#### Example

Consider the following program on  $\mathbb{N}$ .

```
while (i <> 1) {
    if even(i) { i = i div 2}
}
```

understood as defining the transition relation on  $\mathbb N$ 

$$\tau(i,i') \triangleq i \neq 1 \land (odd(i) \Rightarrow i' = i) \land (even(i) \Rightarrow i' = i/2)$$

- Let us prove by fixpoint computation that the ranking semantics is:
  - Termination domain:  $dom[f] = \{2^n \mid n \in \mathbb{N}\}$
  - Ranking function:  $f(n) = \log_2 n$ .

#### Iterates

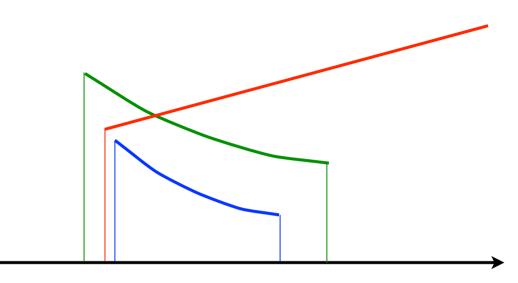
we calculate the iterates of

$$\begin{split} \mathcal{B}_r[\![\tau]\!](f) &\triangleq \{\langle s, 0 \rangle \mid s \in \mathsf{final}\} \cup \{\langle s, f(\tau(s)) + 1 \rangle \mid \tau(s) \in \mathsf{dmn}[f]\} \\ &= \{\langle 1, 0 \rangle\} \cup \{\langle i, f(i') + 1 \rangle \mid i \neq 1 \land (odd(i) \Rightarrow i' = i) \land (even(i) \Rightarrow i' = i/2) \land \\ &\quad i' \in \mathsf{dmn}[f]\} \end{split}$$

$$\begin{split} f^{0} &\triangleq \emptyset \\ f^{1} &\triangleq \mathcal{B}_{r}\llbracket\tau\rrbracket(f^{0}) &= \{\langle 1, 0 \rangle\} & \text{(since dmn}[f^{0}] = \emptyset\} \\ f^{2} &\triangleq \mathcal{B}_{r}\llbracket\tau\rrbracket(f^{1}) &= \{\langle 2, 1 \rangle, \langle 1, 0 \rangle\} \\ & \text{(since dmn}[f^{0}] = \{1\}, \text{ and } \operatorname{pre}\llbracket\tau\rrbracket(\operatorname{dmn}[f^{0}]) = \{2\} \text{ and } \tau(2, 1)\} \\ \cdots \\ f^{n} &= \{\langle 2^{i}, i \rangle \mid 0 \leqslant i < n\} & \text{(induction hypothesis of the recurrence)} \\ f^{n+1} &\triangleq \mathcal{B}_{r}\llbracket\tau\rrbracket(f^{n}) &= \{\langle 1, 0 \rangle\} \cup \{\langle 2^{i+1}, i+1 \rangle \mid 0 \leqslant i < n\} \\ &= \{\langle 2^{i}, i \rangle \mid 0 \leqslant i < n+1\} \\ \text{(since dmn}[f^{n}] &= \{2^{i} \mid 0 \leqslant i < n\}, \text{ and } \operatorname{pre}\llbracket\tau\rrbracket(\operatorname{dmn}[f^{n}]) = \{2^{i+1} \mid 0 \leqslant i < n\} \text{ and} \\ \tau(2^{i+1}, 2^{i}) \end{bmatrix} \\ \cdots \\ f^{\omega} &= \bigcup_{n \geqslant 0} f^{n} = \bigcup_{n \geqslant 0} \{\langle 2^{i}, i \rangle \mid 0 \leqslant i \leqslant n\} = \{\langle 2^{i}, i \rangle \mid 0 \leqslant i\} \\ f^{\omega+1} &= \mathcal{B}_{r}\llbracket\tau\rrbracket(f^{\omega}) = f^{\omega} = \operatorname{lfp}_{0}^{\subseteq} \mathcal{B}_{r}\llbracket\tau\rrbracket = \lambda n \in 2^{\mathbb{N}} \cdot \log_{2} n \\ \Box \end{split}$$

#### **Computable abstractions**

• Approximation:



- Abstraction by a reduced product of standard abstractions e.g.:
  - Linear equalities <sup>(I)</sup> (with negative slopes and minimum or positive slopes and maximum)
  - Powers <sup>(II)</sup>
  - ...

 (I) Michael Karr: Affine Relationships Among Variables of a Program. <u>Acta Inf. 6</u>: 133-151 (1976)
 (II) Isabella Mastroeni: Algebraic Power Analysis by Abstract Interpretation. <u>Higher-Order and</u> <u>Symbolic Computation 17(4)</u>: 297-345 (2004)

#### On going work ...

- Currently working on the formalization in AI terms
- and on abstractions for further methods:
  - Burstall <sup>(I), (II)</sup>
  - Ramsey (III)
  - ...
  - Checking temporal specifications of infinite systems (e.g. temporal logics)
- Rod M. Burstall: Program Proving as Hand Simulation with a Little Induction. <u>IFIP Congress 1974</u>: 308-312
- (II) Patrick Cousot, <u>Radhia Cousot</u>: Sometime = Always + Recursion = Always on the Equivalence of the Intermittent and Invariant Assertions Methods for Proving Inevitability Properties of Programs.<u>Acta Inf. 24(1)</u>: I-31 (1987)
- (III) Andreas Podelski, Andrey Rybalchenko: Transition Invariants. LICS 2004: 32-41

### Conclusion

#### Conclusion

 This foundational preliminary work is the first step towards methods and inference algorithms for proving liveness by over-approximation