Language-Parametric Formal Methods in the Field

Thesis Proposal

Daejun Park
April 20, 2018

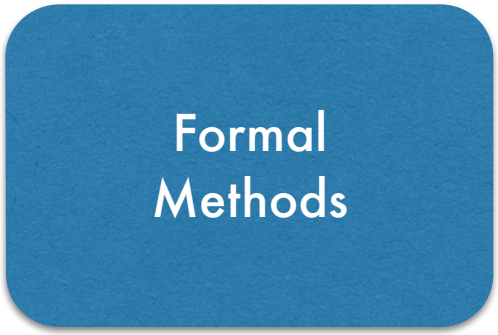
Reusability problem of formal tools

- Fragmentation: crafted for a fixed language
  - Implemented similar heuristics/optimizations: duplicating efforts
- Hard to re-target
  - Replace hardcoded semantics, or
  - Implement a translation to the original language

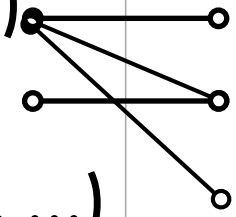
Language-parametric formal methods

- Mitigate reusability issue
- Develop universal formal methods
  - Parameterized by language semantics
- Instantiate formal tools
  - By plugging-in language semantics
- Admit operational semantics

Language-parametric formal methods



- C (c11, gcc, clang, ...)
- Java (6, 7, 8, ...)
- JavaScript (ES5, ES6, ...)
- ...

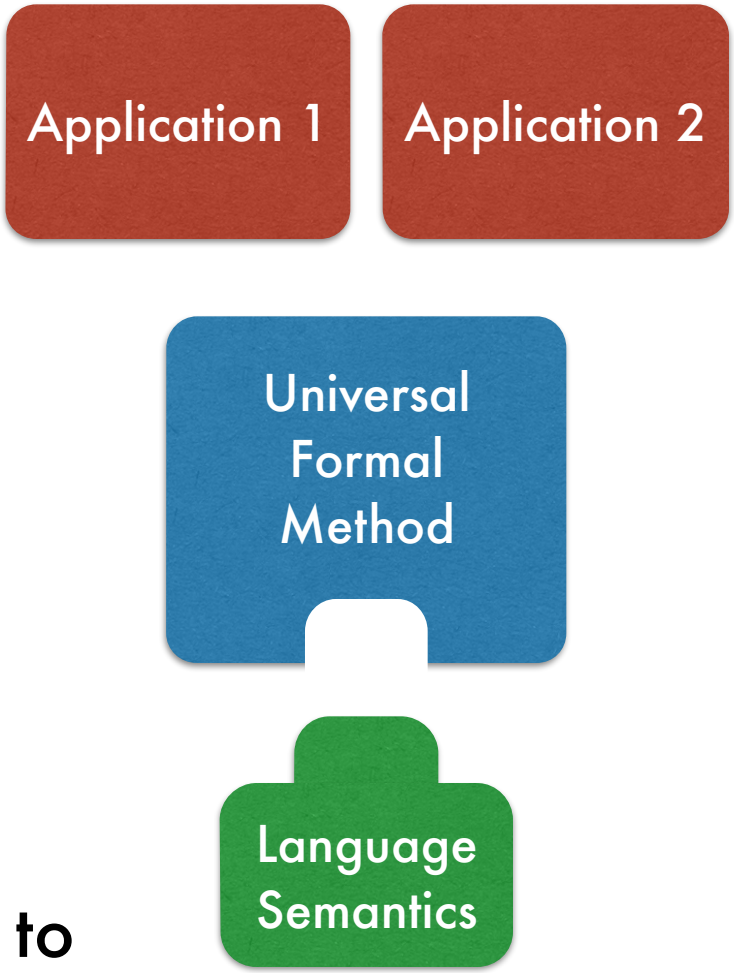


- Deductive verification
- Model checking
- Program equivalence checking
- ...

Defined/implemented once, and reused for all others

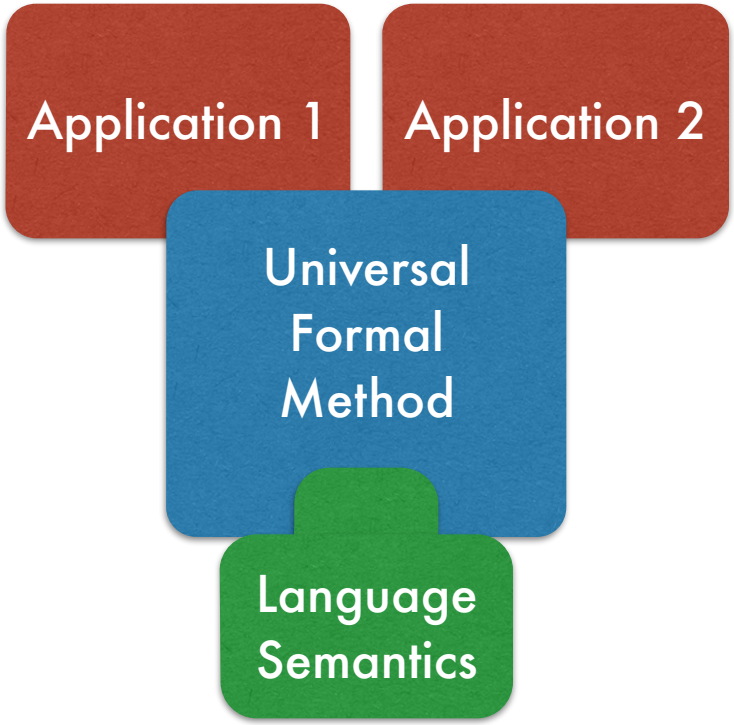
Thesis work

- Specifying real-world language semantics and measuring the specification effort
- Developing language-parametric formal methods
- Instantiating them by plugging-in various language semantics
- Applying the derived formal tools to real-world applications, demonstrating their practical feasibility

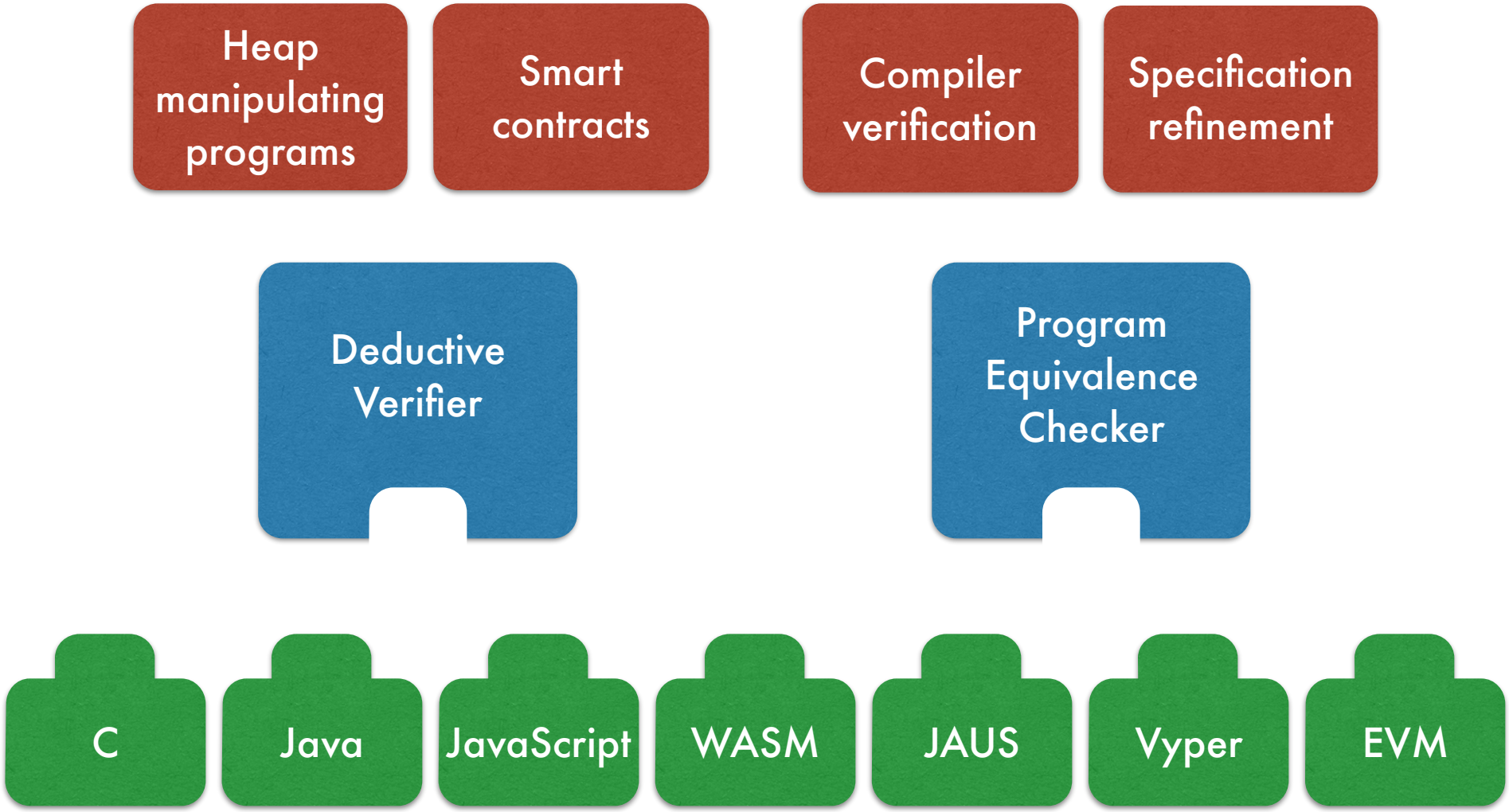


Thesis work

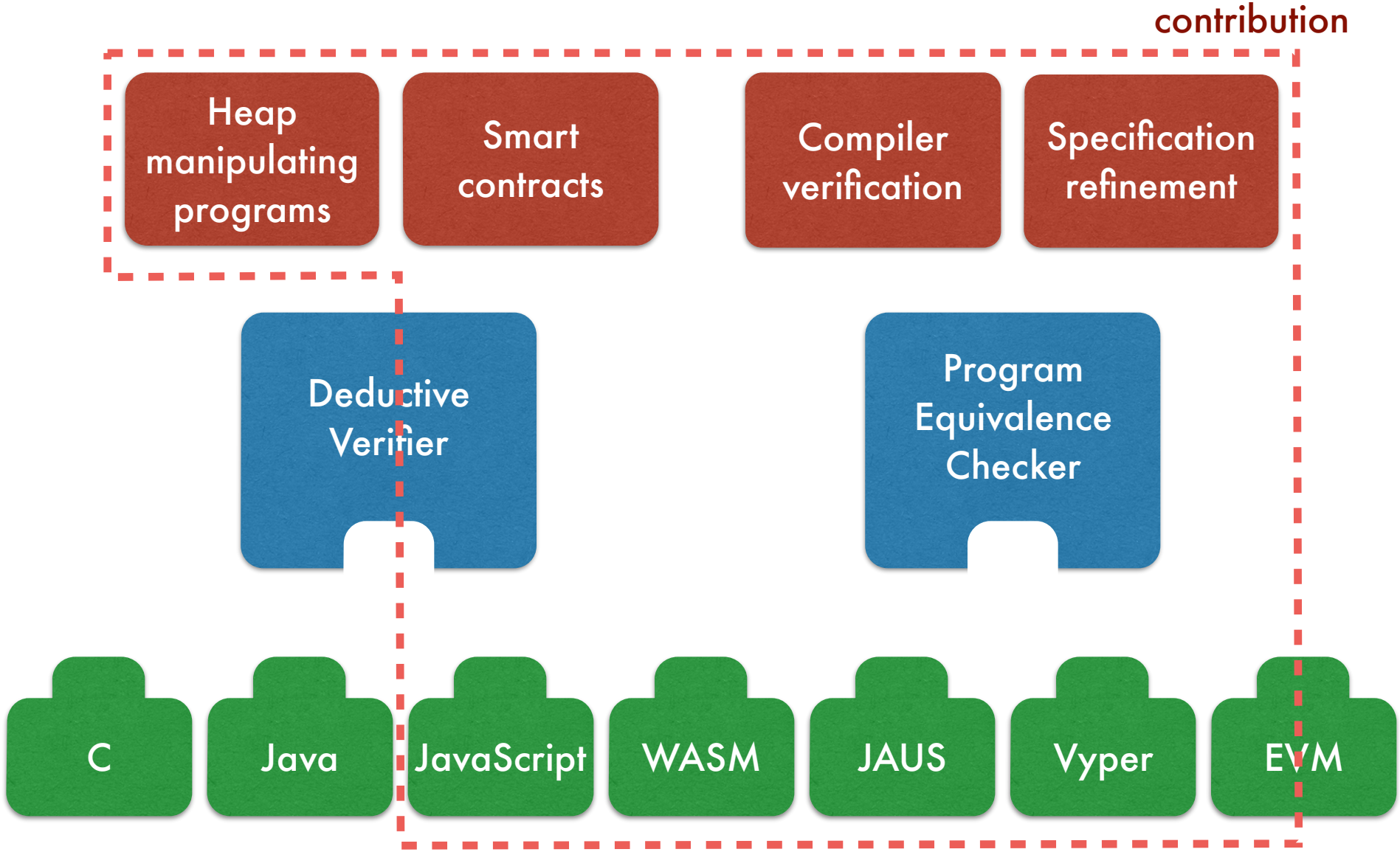
- Specifying real-world language semantics and measuring the specification effort
- Developing language-parametric formal methods
- Instantiating them by plugging-in various language semantics
- Applying the derived formal tools to real-world applications, demonstrating their practical feasibility



Thesis work

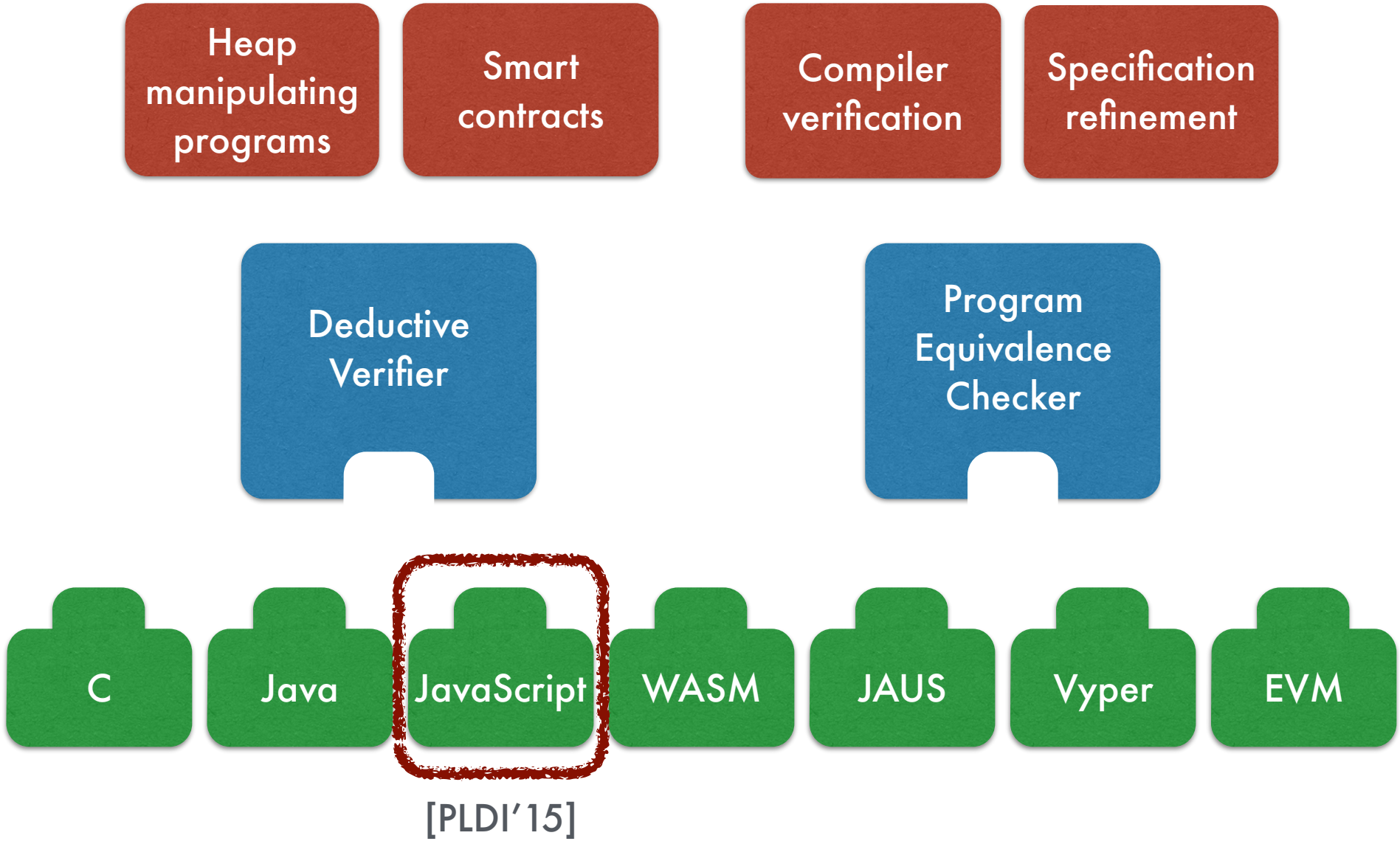

Heap manipulating programs
Smart contracts
Compiler verification
Specification refinement
Deductive Verifier
Program Equivalence Checker
C
Java
JavaScript
WASM
JAUS
Vyper
EVM

Thesis work


contribution
Heap manipulating programs
Smart contracts
Compiler verification
Specification refinement
Deductive Verifier
Program Equivalence Checker
C
Java
JavaScript
WASM
JAUS
Vyper
EVM

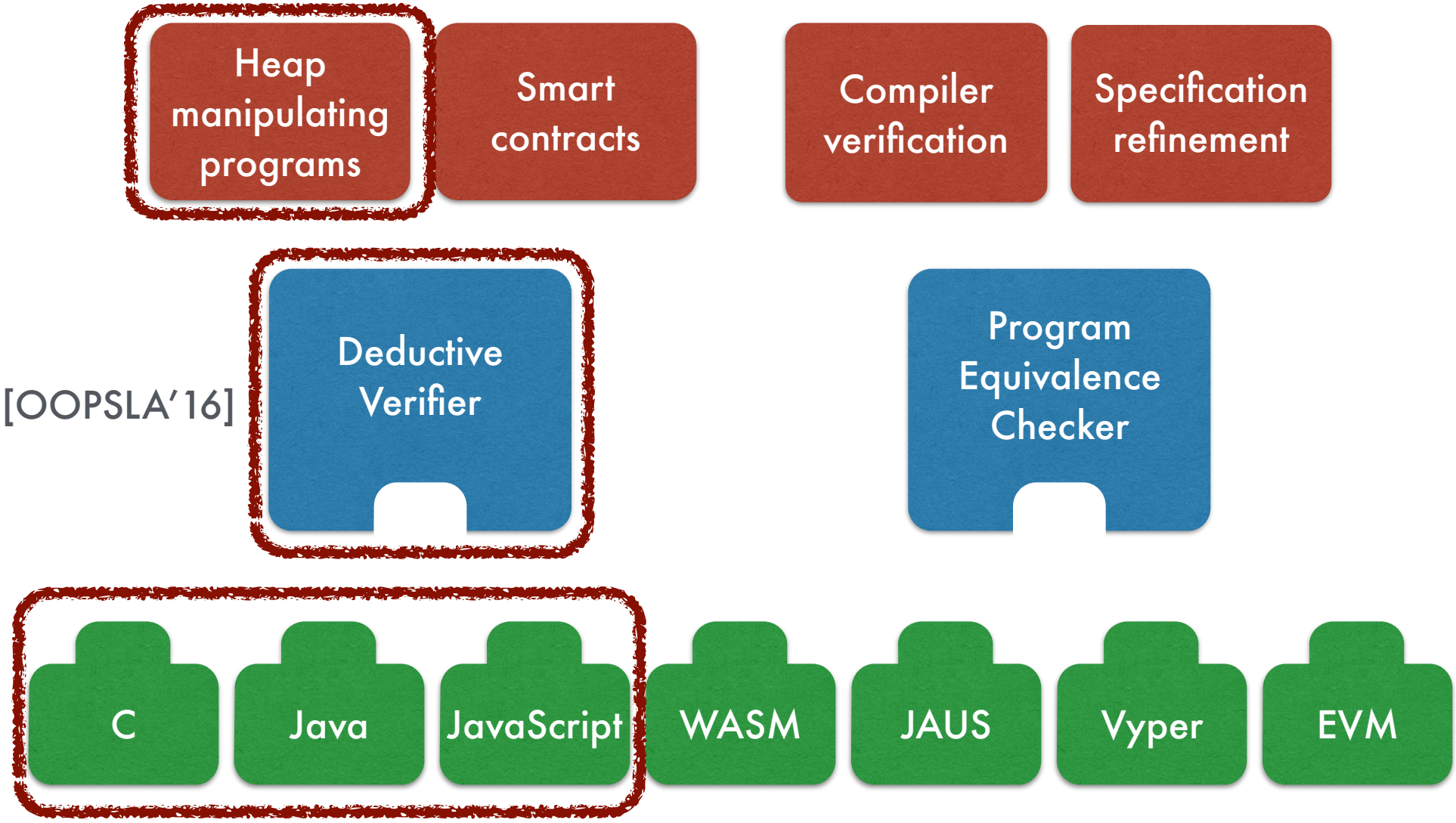
Current Results

Specifying language semantics



[PLDI'15]

Instantiating with various languages



Applying to real-world systems

[TR'17] [TR'18]

Heap manipulating programs

Smart contracts

Compiler verification

Specification refinement

Deductive Verifier

Program Equivalence Checker

C

Java

JavaScript

WASM

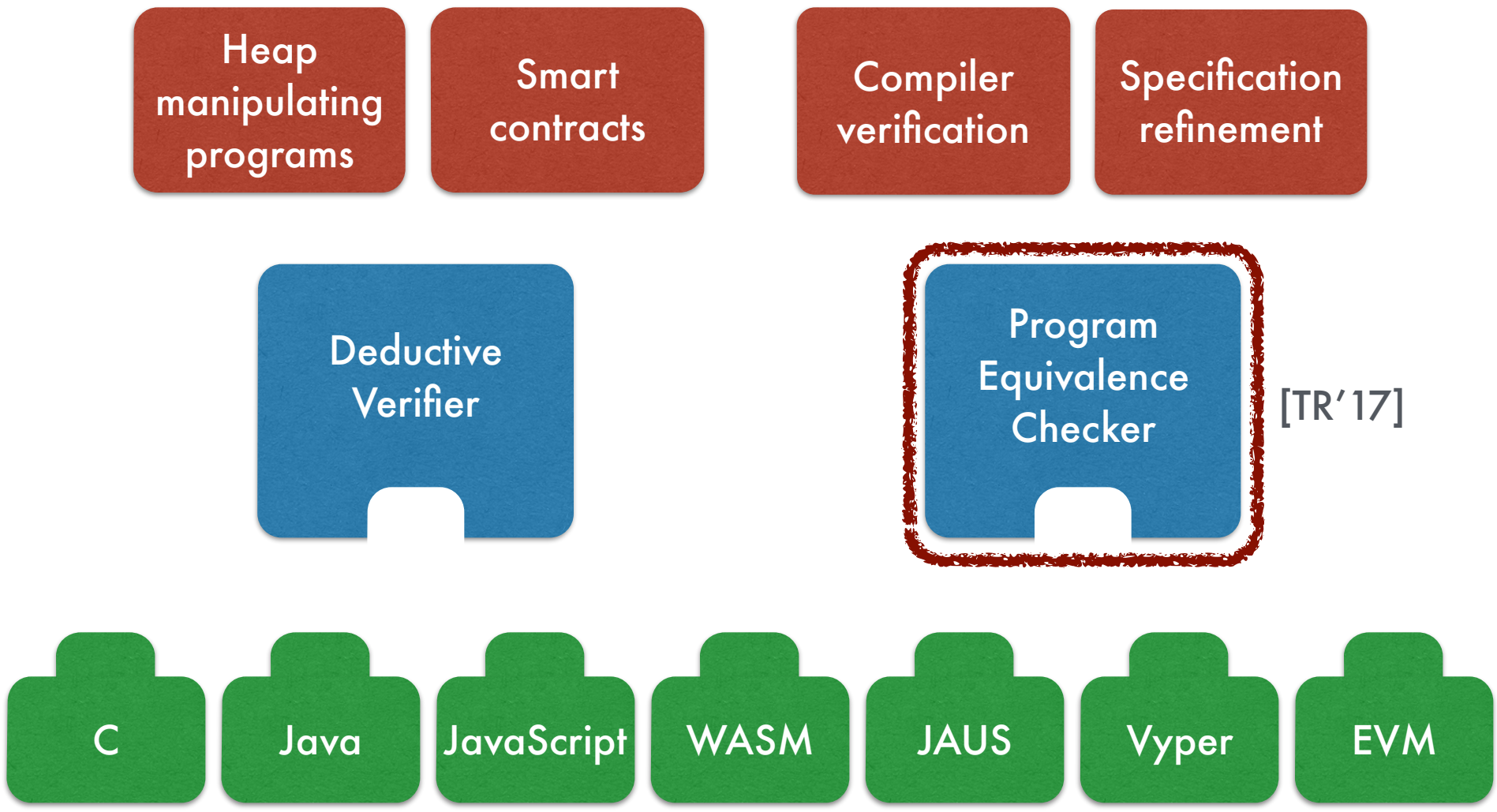
JAUS

Vyper

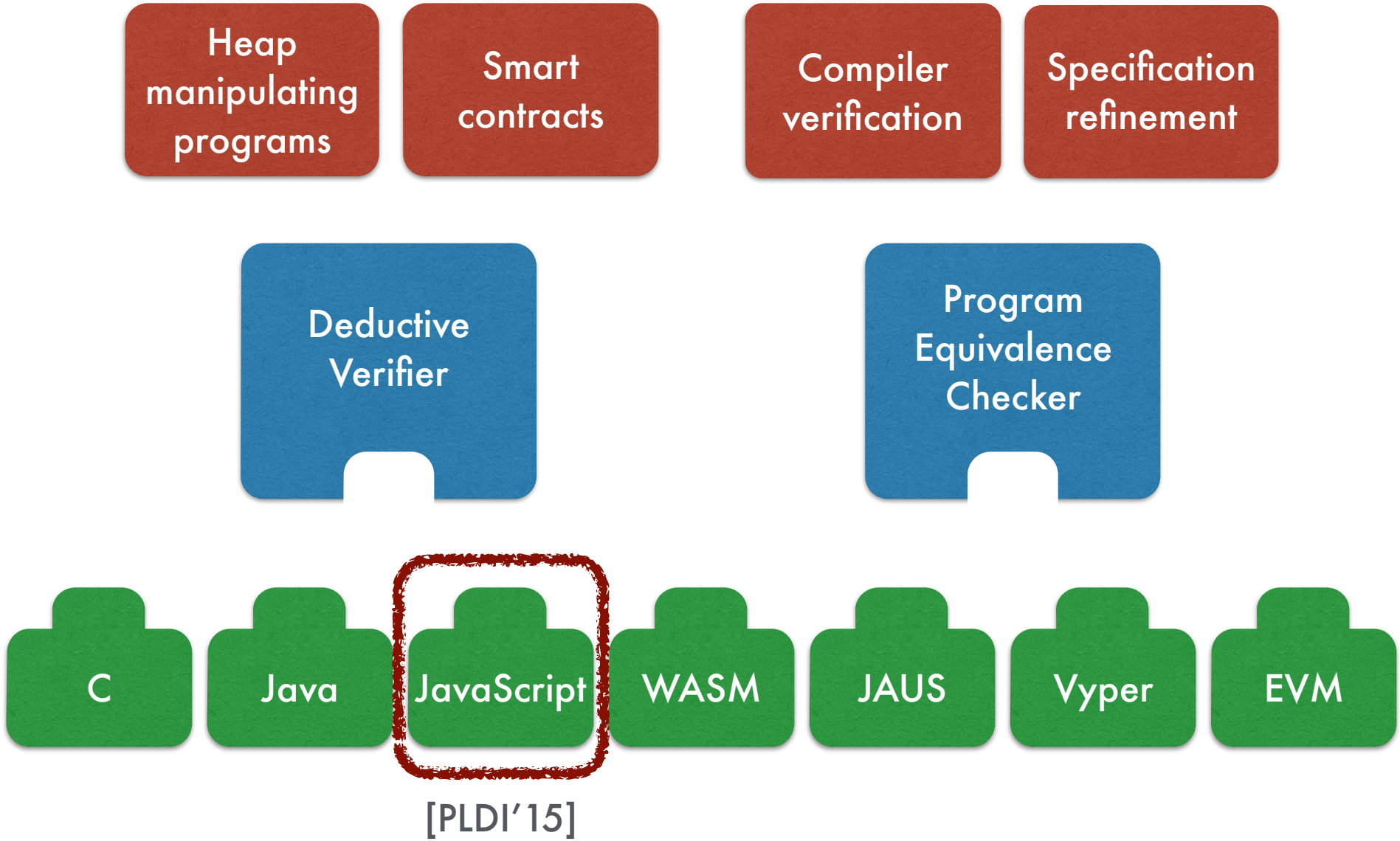
EVM

[CSF'18]

Developing new formal method



Specifying language semantics



[PLDI'15]

KJS: complete formal semantics of JavaScript

KJS *faithfully* formalizes ECMAScript 5.1 (informal) standard.

each step of informal description ⟷ (1-to-1 mapping, systematic translation) formal pseudo-code statement

ECMAScript 5.1 standard	KJS
The expression "++ *Expression*" is evaluated as follows:	`rule ++ Expression =>`
1. Let *expr* be the result of evaluating *Expression*.	`Let $expr = @GetReference(Expression);`
2. Let *oldValue* be ToNumber(GetValue(*expr*)).	`Let $oldValue = ToNumber(GetValue($expr));`
3. Let *newValue* be the result of adding the value 1 to *oldValue*.	`Let $newValue = @Addition($oldValue,1);`
4. Call PutValue(*expr*, *newValue*).	`Call PutValue($expr,$newValue);`
5. Return *newValue*.	`Return $newValue;`

manual inspection

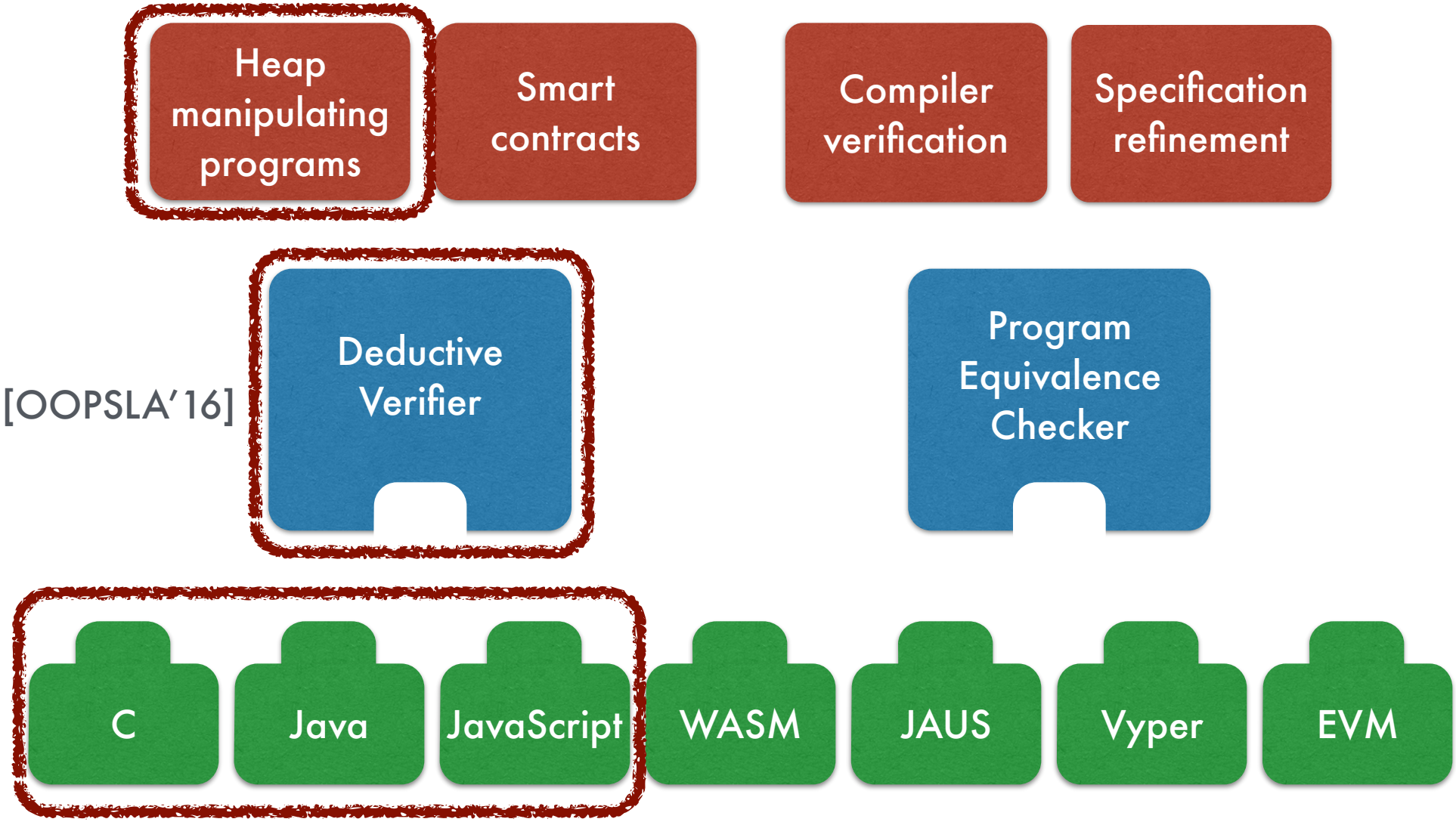
Completeness

Tested against ECMAScript conformance test suite.

Formal Semantics	Passed	Failed	% passed
KJS	2,782	0	100.0%
[Politz et al. 2012][5]	2,470	345	87.7%
[Bodin et al. 2014]	1,796	986	64.6%

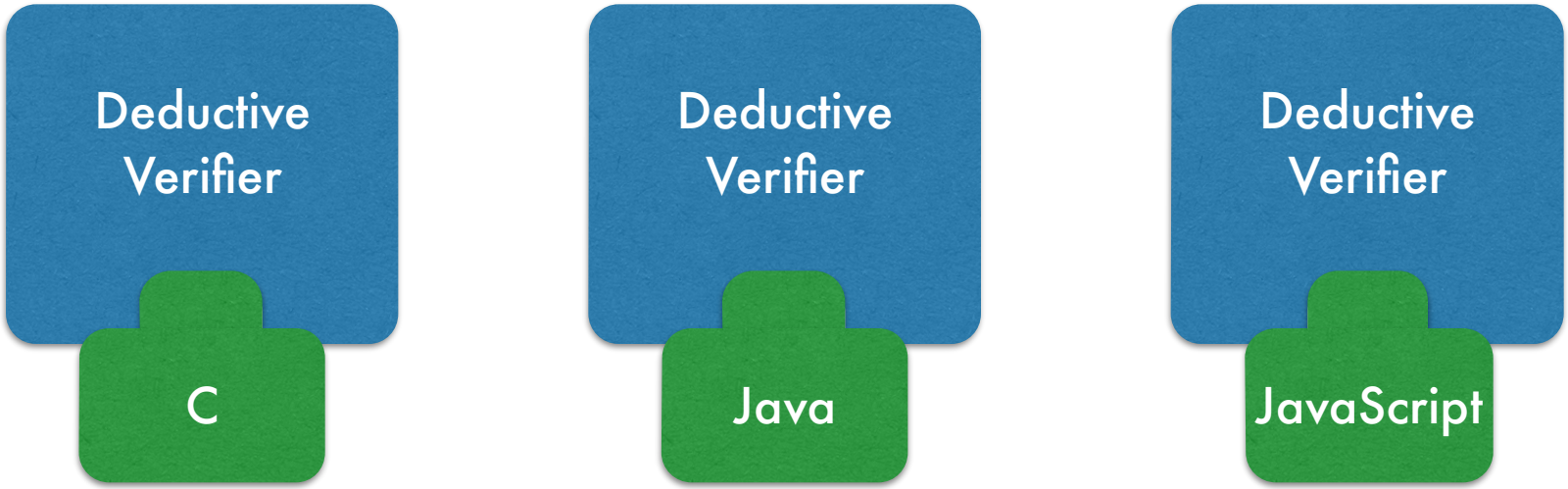
Took me only four months.

Instantiating with various languages



Deductive program verifiers

- Instantiated framework by plugging-in three language semantics.



- Verified challenging heap-manipulating programs implementing the same algorithms in all three languages.

Experiments [OOPSLA'16]

Time (secs)

Programs	C	Java	JS
BST find	14.0	4.7	6.3
BST insert	30.2	8.6	8.2
BST delete	71.7	24.9	21.2
AVL find	13.0	4.9	6.4
AVL insert	281.3	105.2	135.0
AVL delete	633.7	271.6	239.6
RBT find	14.5	5.0	6.8
RBT insert	903.8	115.6	114.5
RBT delete	1,902.1	171.2	183.6

Programs	C	Java	JS
Treap find	14.4	4.9	6.5
Treap insert	67.7	23.1	18.9
Treap delete	90.4	28.4	33.2
List reverse	11.4	4.1	5.5
List append	14.8	7.3	5.3
Bubble sort	66.4	38.8	31.3
Insertion sort	61.9	31.1	44.8
Quick sort	79.2	47.1	48.1
Merge sort	170.6	87.0	66.0
Total	4,441.1	983.5	981.2
Average	246.7	54.6	54.5

Experiments [OOPSLA'16]

Time (secs)

Programs	C	Java	JS
BST find	14.0	4.7	6.3
BST insert	30.2	8.6	8.2

Programs	C	Java	JS
Treap find	14.4	4.9	6.5
Treap insert	67.7	23.1	18.9

Full functional correctness:

```
/*
 * @pre  bst(t)
 * @post bst(t')
 * @post keys(t') == keys(t) \union { v }
 */
function insert(t, v) {
  ...
}
```

Total	4,441.1	983.5	981.2
Average	246.7	54.6	54.5

Experiments [OOPSLA'16]

Time (secs)

Programs	C	Java	JS
BST find	14.0	4.7	6.3
BST insert	30.2	8.6	8.2
BST delete	71.7	24.9	21.2
AVL find	13.0	4.9	6.4
AVL insert	281.3	105.2	135.0

Programs	C	Java	JS
Treap find	14.4	4.9	6.5
Treap insert	67.7	23.1	18.9
Treap delete	90.4	28.4	33.2
List reverse	11.4	4.1	5.5
List append	14.8	7.3	5.3

Performance is comparable to a state-of-the-art verifier for C, VCDryad, based on a separation logic extension of VCC:
e.g., AVL insert : 260s vs 280s (ours)

Total	4,441.1	983.5	981.2
Average	246.7	54.6	54.5

Applying to real-world systems

[TR'17] [TR'18]

Heap manipulating programs

Smart contracts

Compiler verification

Specification refinement

Deductive Verifier

Program Equivalence Checker

C

Java

JavaScript

WASM

JAUS

Vyper

EVM

[CSF'18]

Smart contracts

- Programs that run on blockchain
- Usually written in a high-level language
  - Solidity (JavaScript-like), Vyper (Python-like), …
- Compiled down to VM bytecode
  - EVM (Ethereum VM), IELE (LLVM-like VM), …

  - Runs on VM of blockchain nodes

Challenges for EVM bytecode verification

- Byte-twiddling operations
  - Non-linear integer arithmetic (e.g., modulo reduction)
- Arithmetic overflow detection
- Gas limit
  - Variable gas cost depending on contexts
- Hash collision

Byte-twiddling operations

Given:

$$x[n] \stackrel{\texttt{def}}{=} (x/256^n) \bmod 256$$

$$merge(x[i..j]) \stackrel{\texttt{def}}{=} merge(x[i..j+1]) \mathbin{\underline{*}} 256 \mathbin{\underline{+}} x[j] \quad \text{when } i > j$$

$$merge(x[i..i]) \stackrel{\texttt{def}}{=} x[i]$$

Prove:

$$\text{“}x = merge(x[31..0])\text{”}.$$

Abstractions

```
syntax Int ::= nthByte(Int, Int, Int) [function]
```

```
rule merge(nthByte(V, 0, N) ... nthByte(V, N-1, N))
  => V
  requires 0 <= V <  2 ^ (N * 8)
       and 1 <= N <= 32
```

Challenges for EVM bytecode verification

- Byte-twiddling operations
  - Non-linear integer arithmetic (e.g., modulo reduction)
- Arithmetic overflow detection
- Gas limit
  - Variable gas cost depending on contexts
- Hash collision

Verified smart contracts

- High-profile ERC20 token contracts
  - OpenZeppelin's
  - ConsenSys's
  - Hacker Gold (HKG) token
  - Vyper ERC20 token

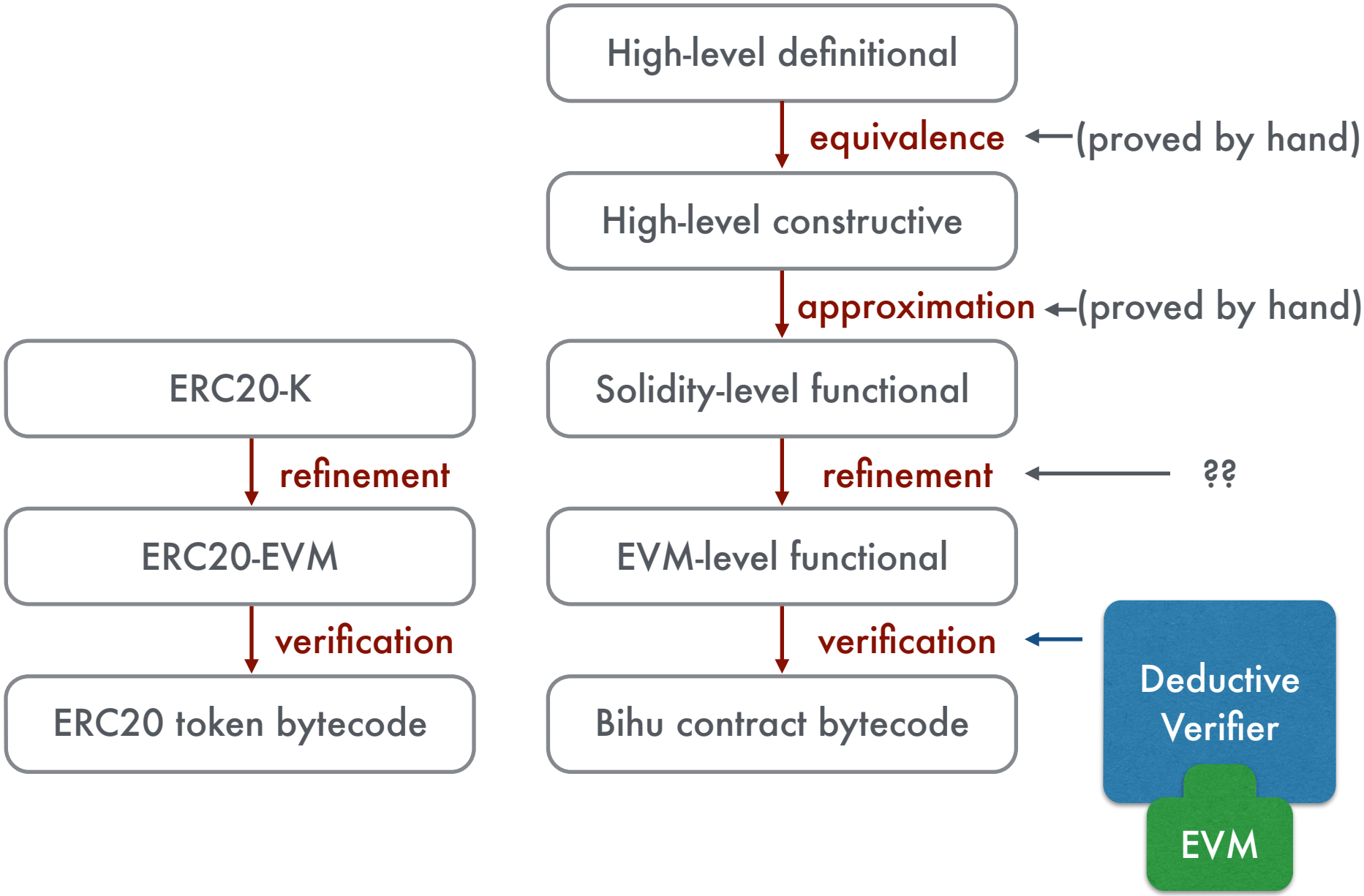
found divergent behaviors

- Commercial Bihu KEY operation contracts
  - KeyRewardPool
  - WarmWallet

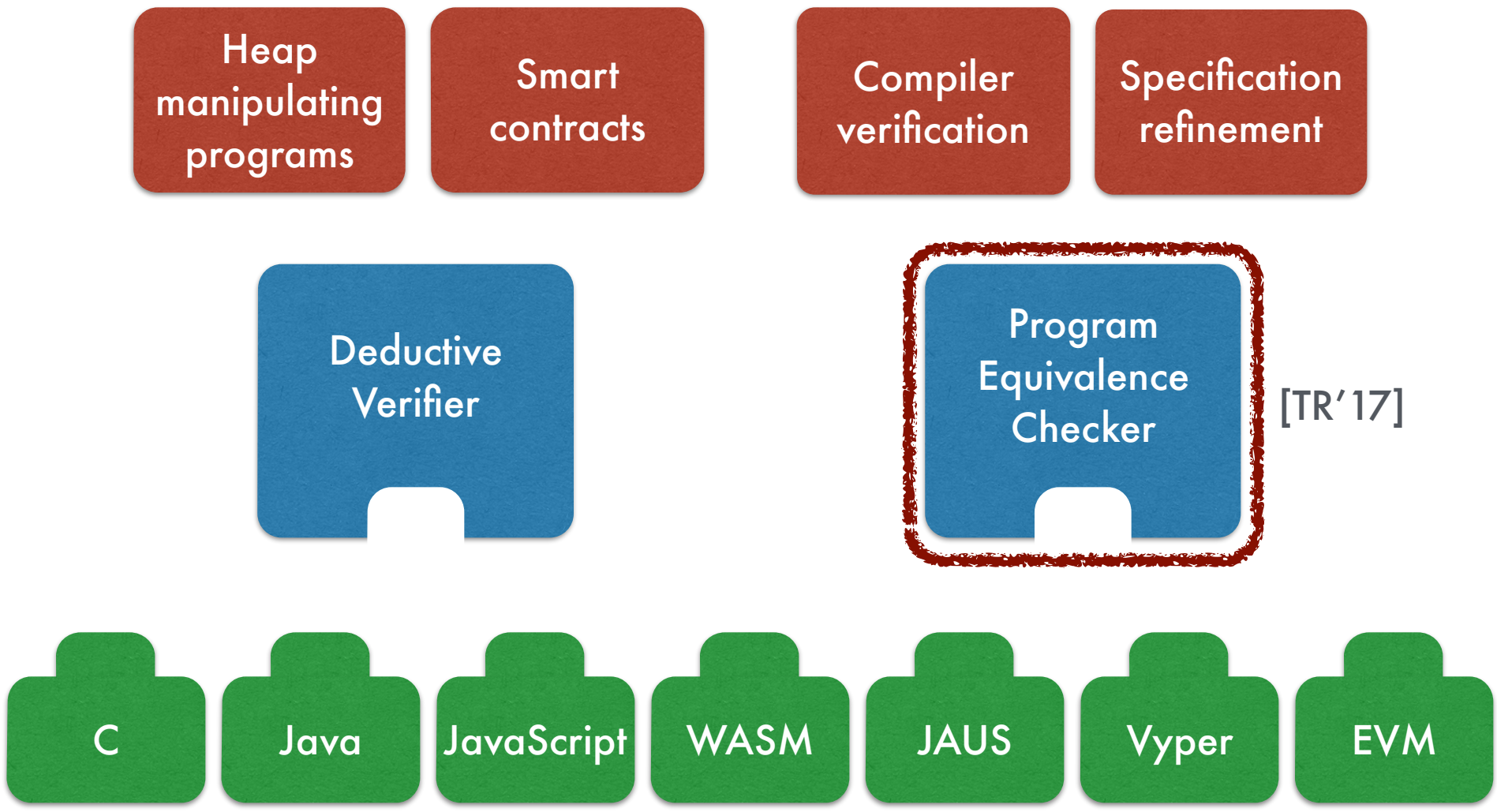
End-to-end verification

- Formalize high-level business logic
  - Confirmed by contract developers
- Refine it to EVM level
  - Capturing EVM-specific details
- Verify EVM bytecode using derived EVM verifier
  - No need to trust Solidity/Vyper compilers

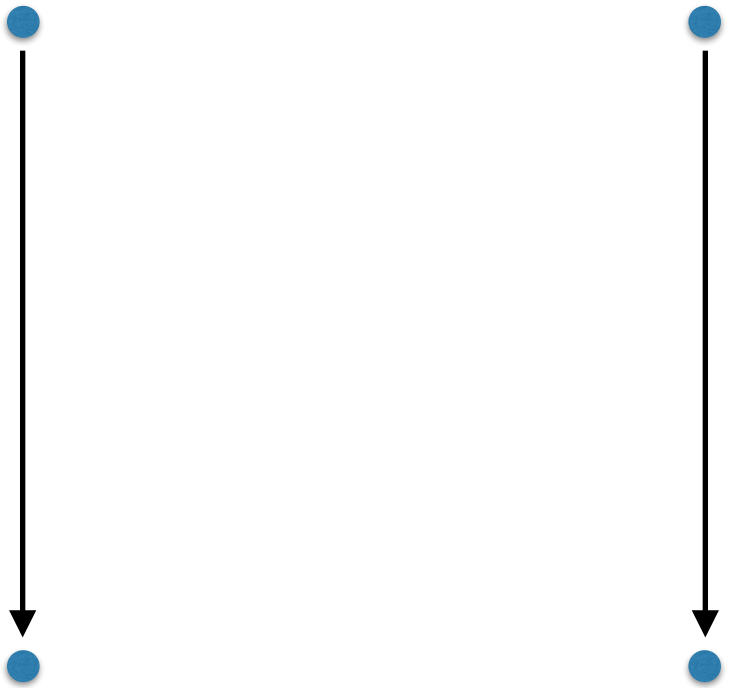
End-to-end verification



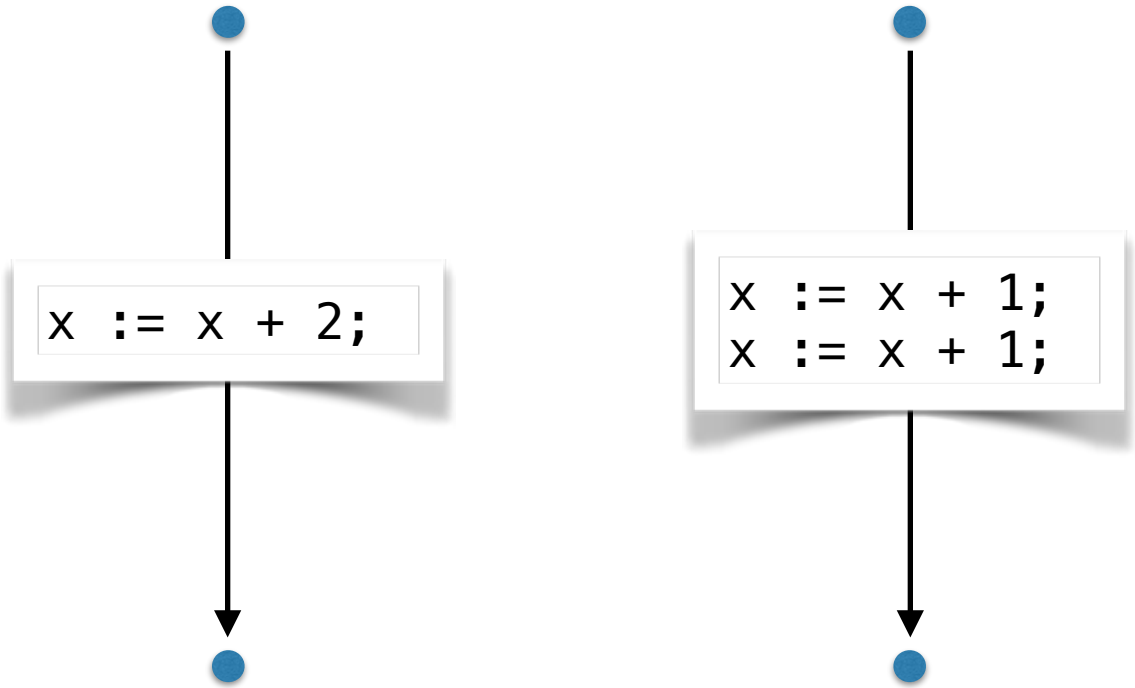
Developing new formal method



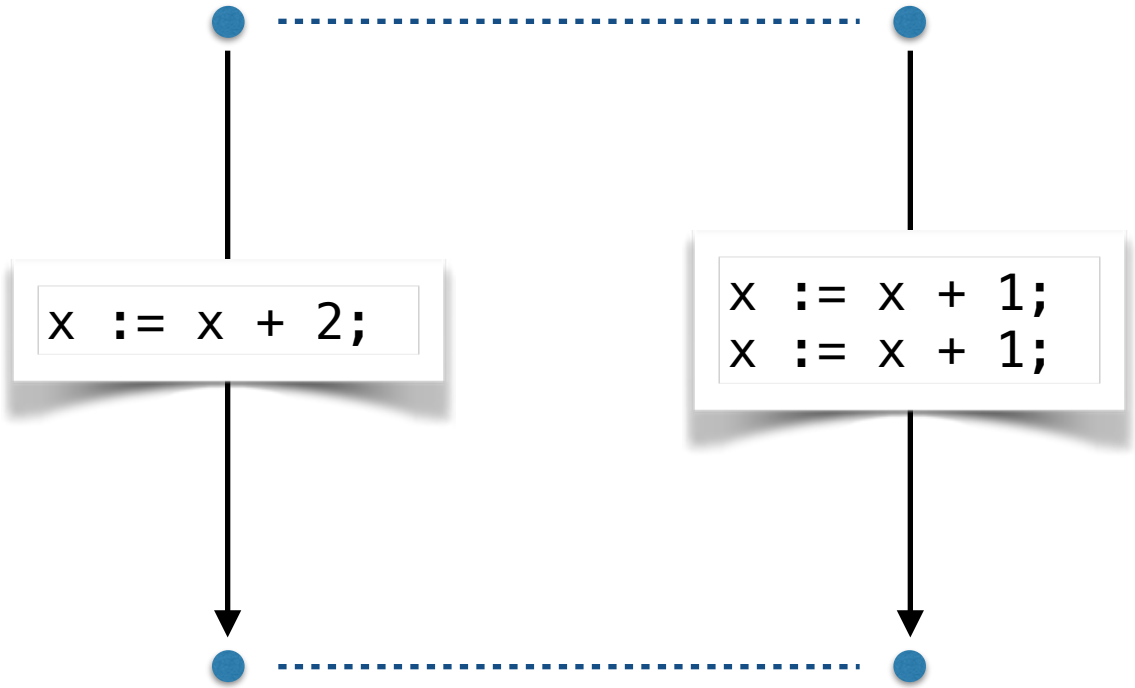
Deterministic, terminating programs



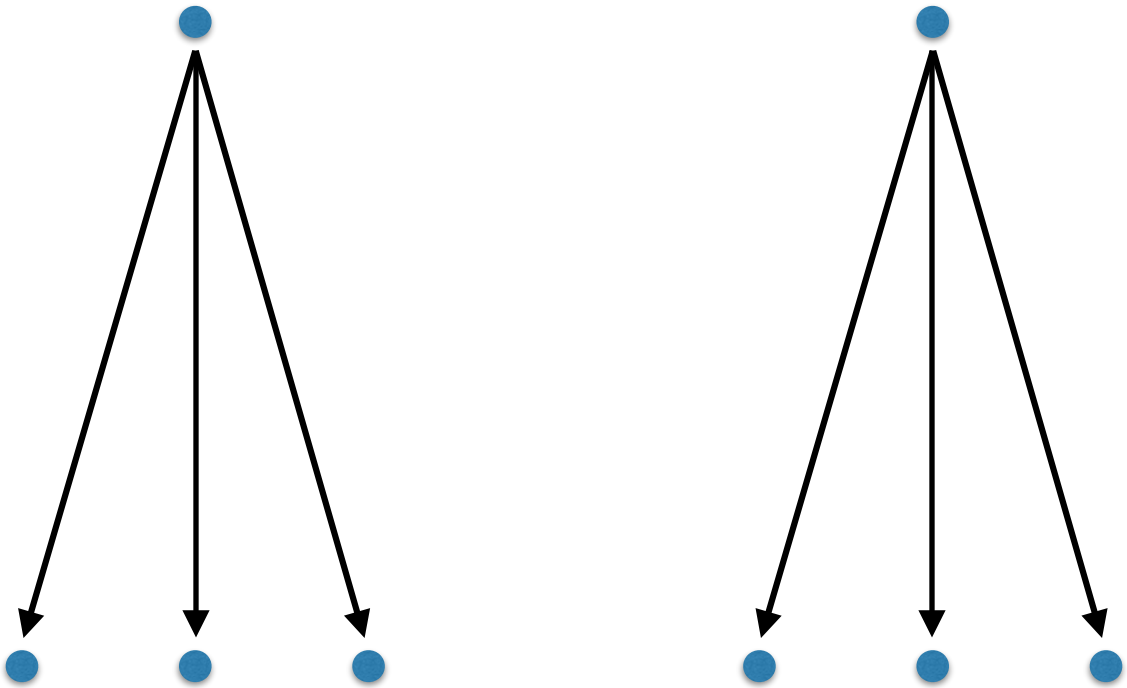
Deterministic, terminating programs



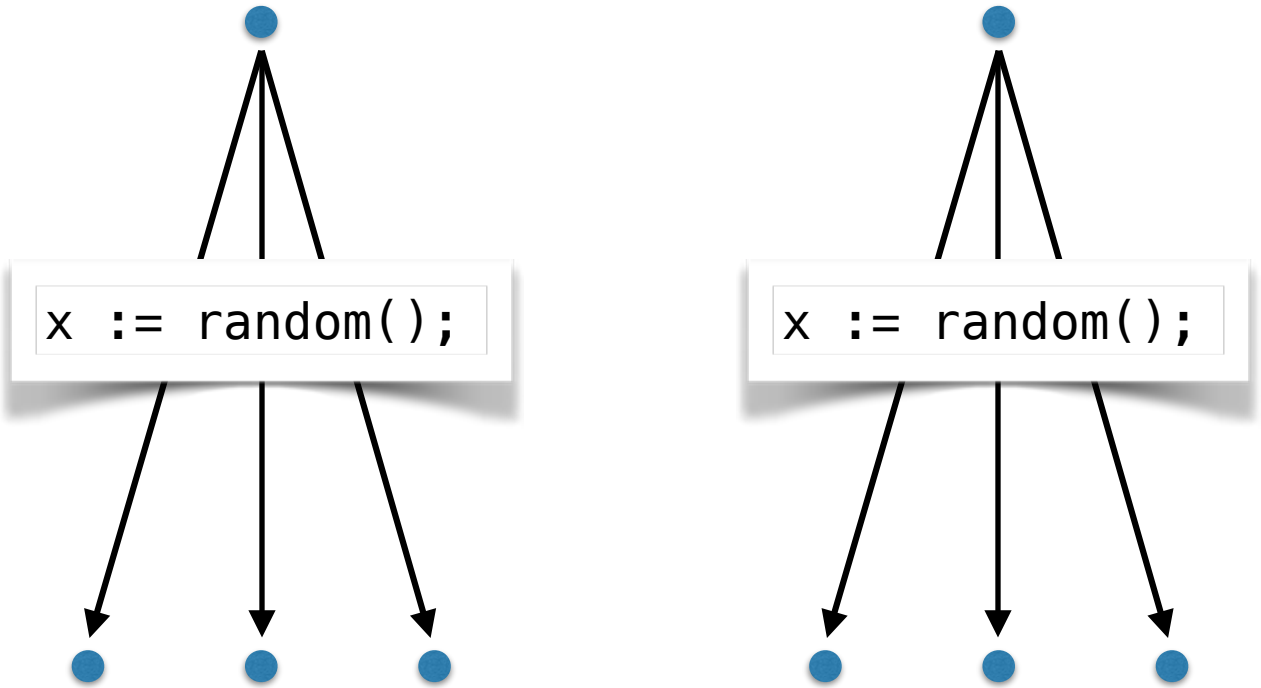
Deterministic, terminating programs



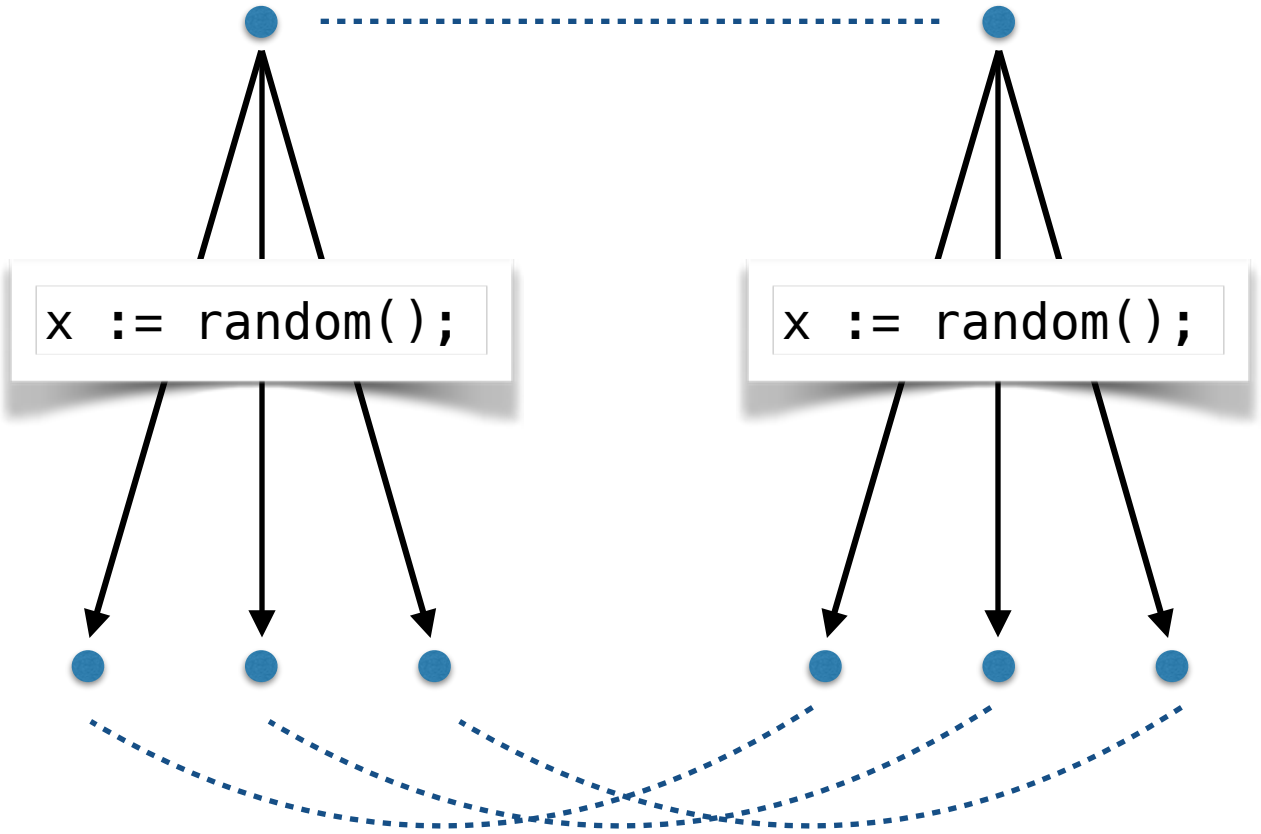
Nondeterministic programs



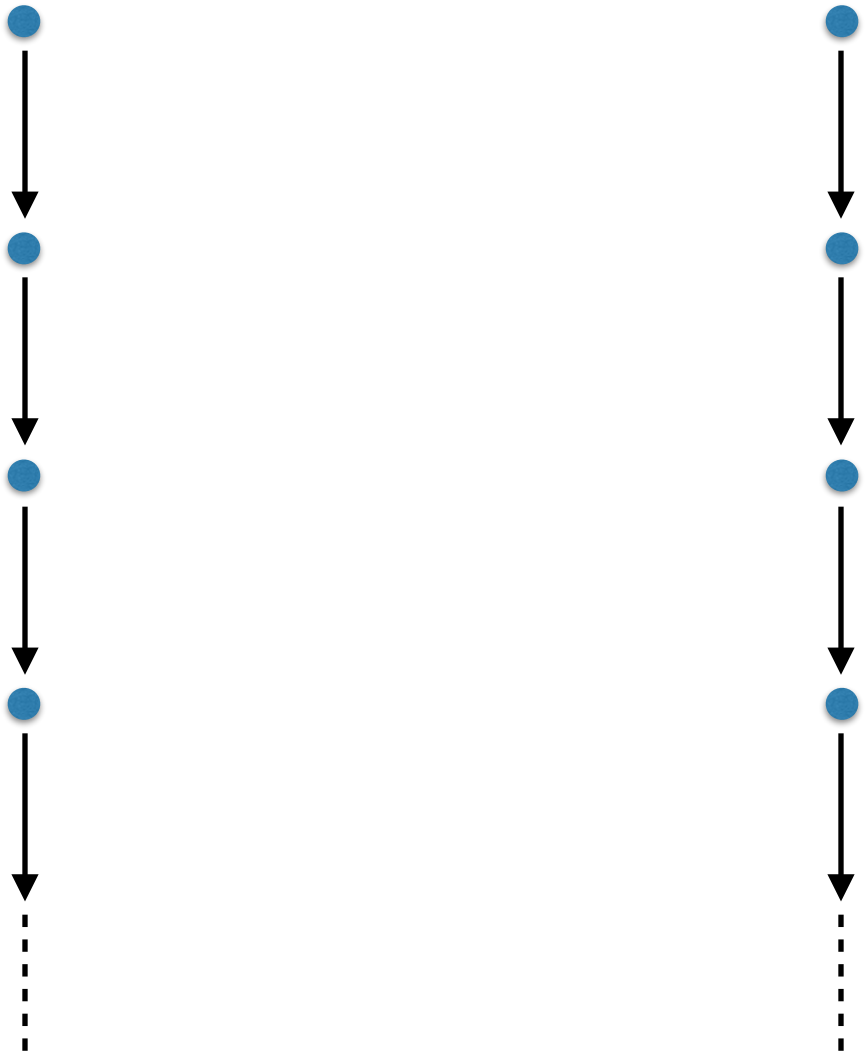
Nondeterministic programs



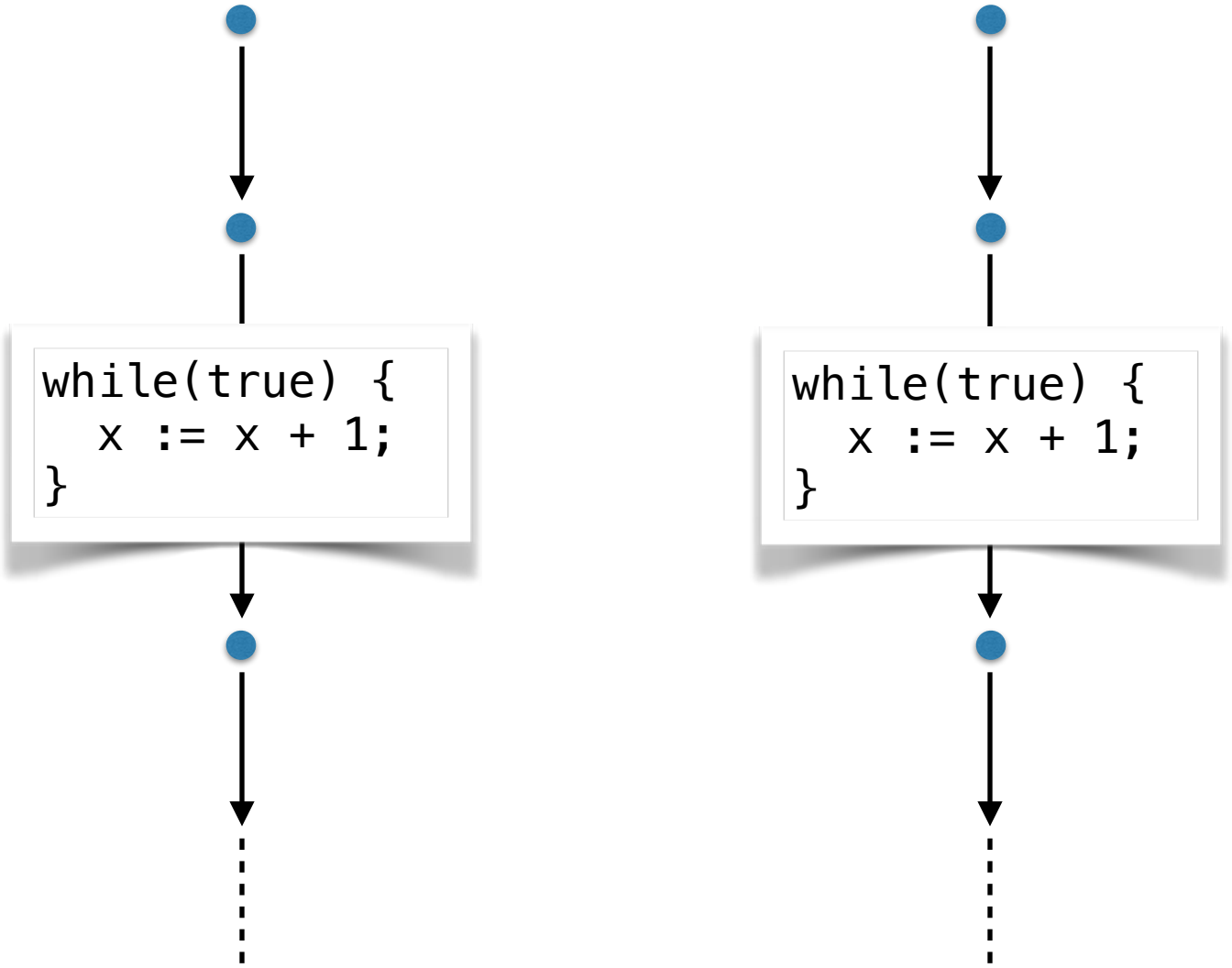
Nondeterministic programs



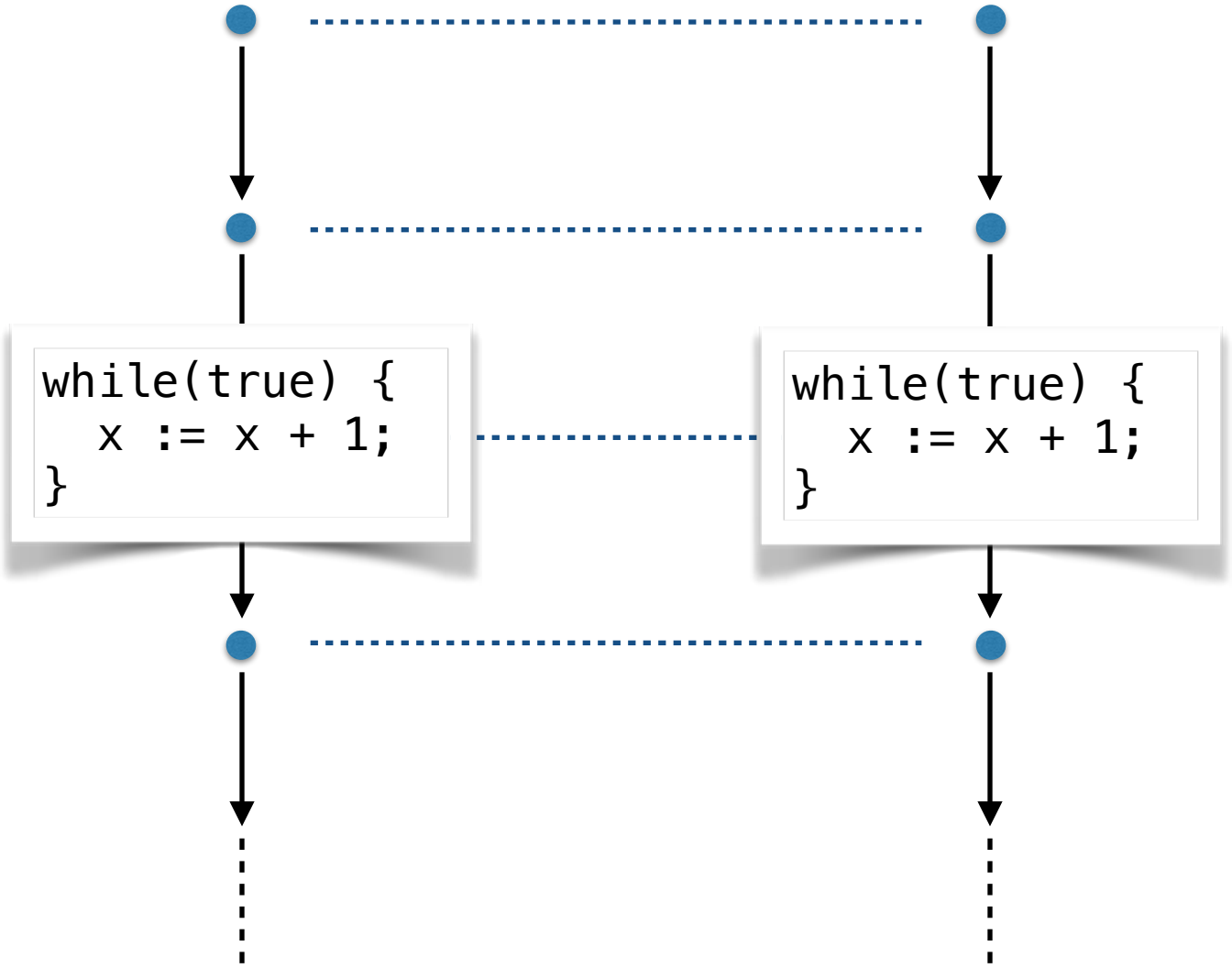
Non-terminating programs



Non-terminating programs



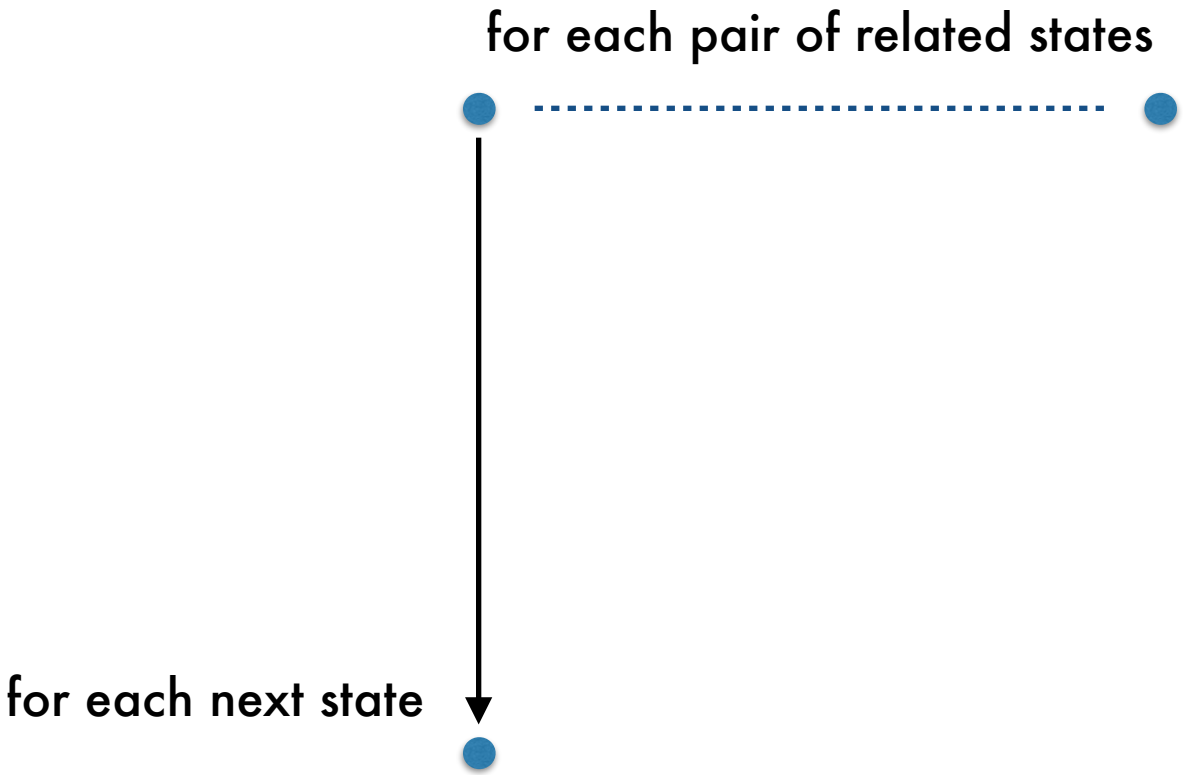
Non-terminating programs



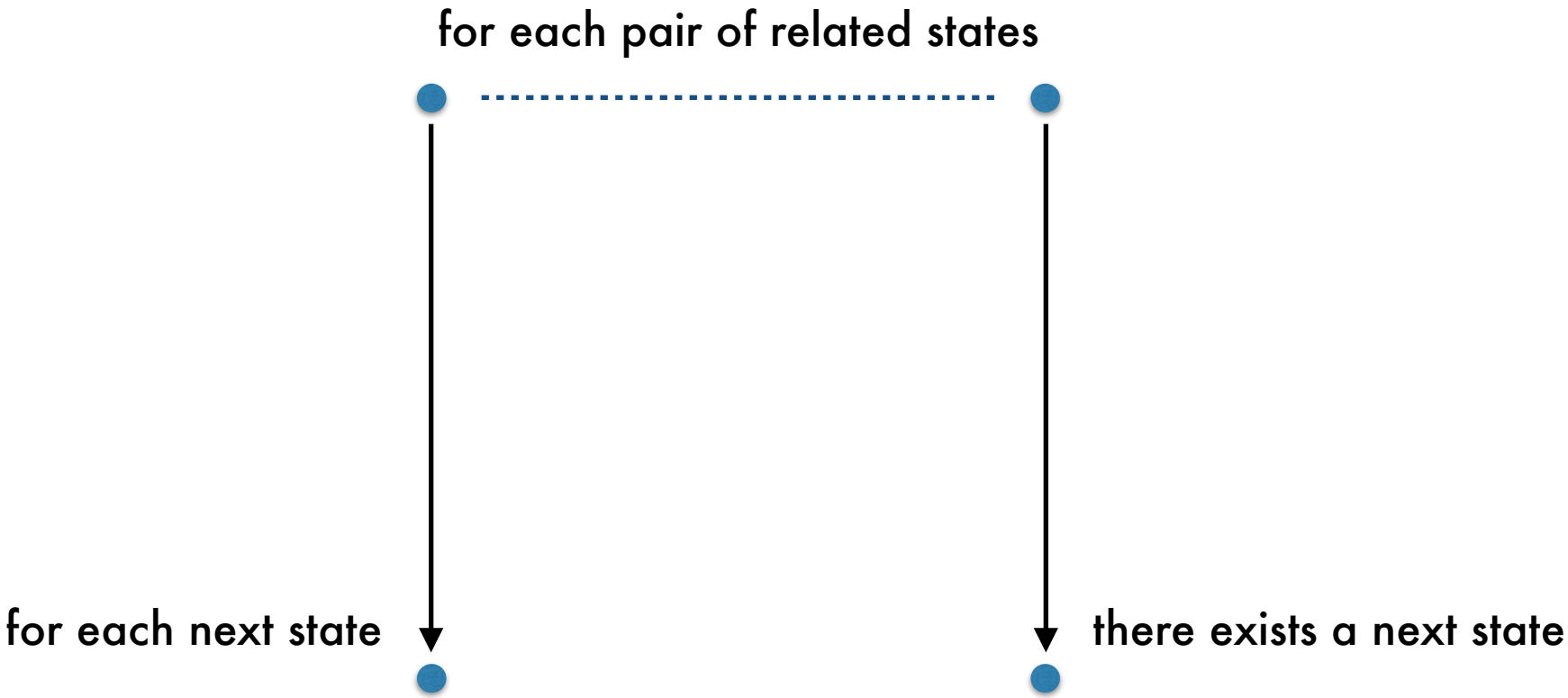
Bisimulation

for each pair of related states

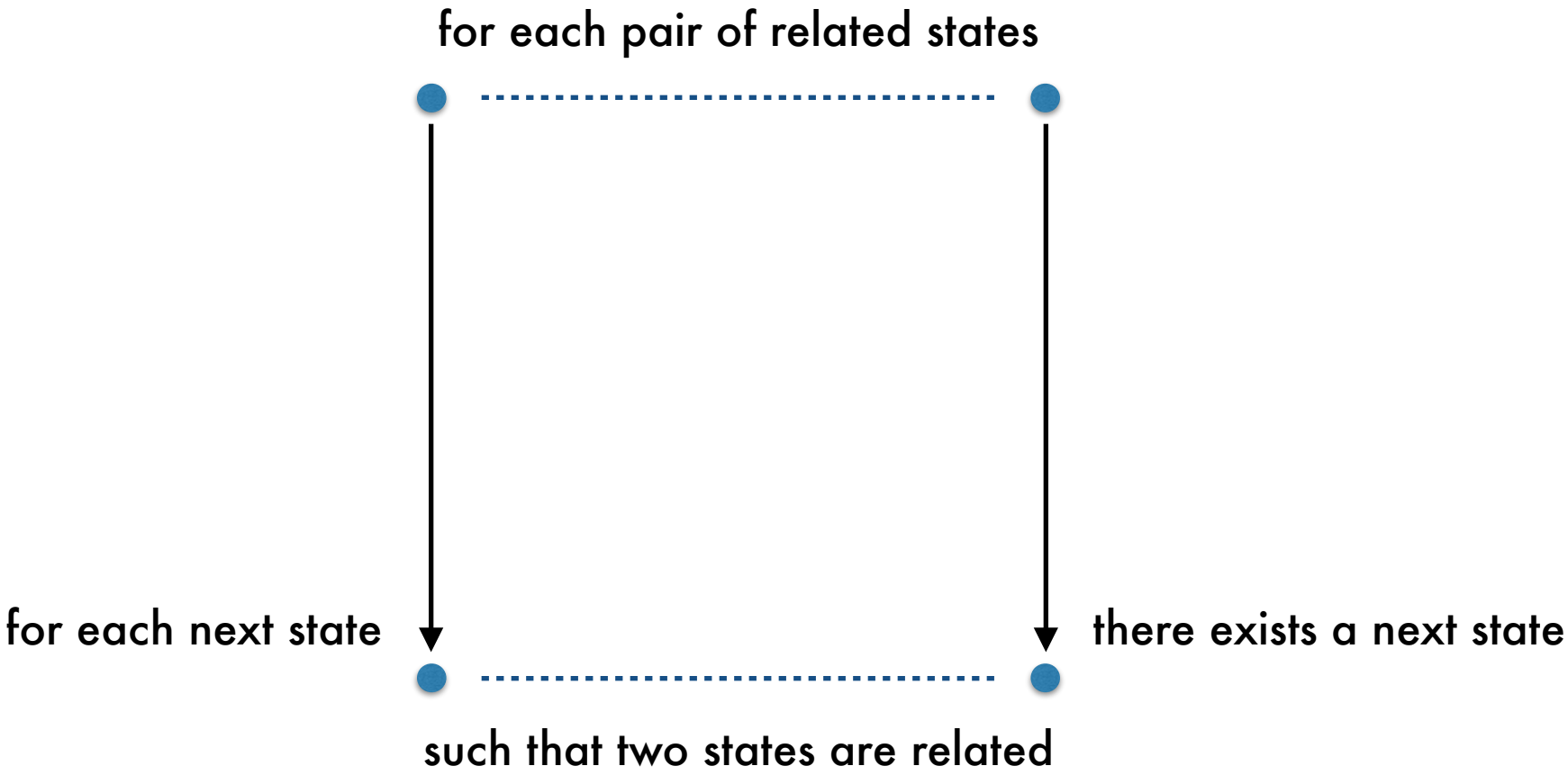
Bisimulation



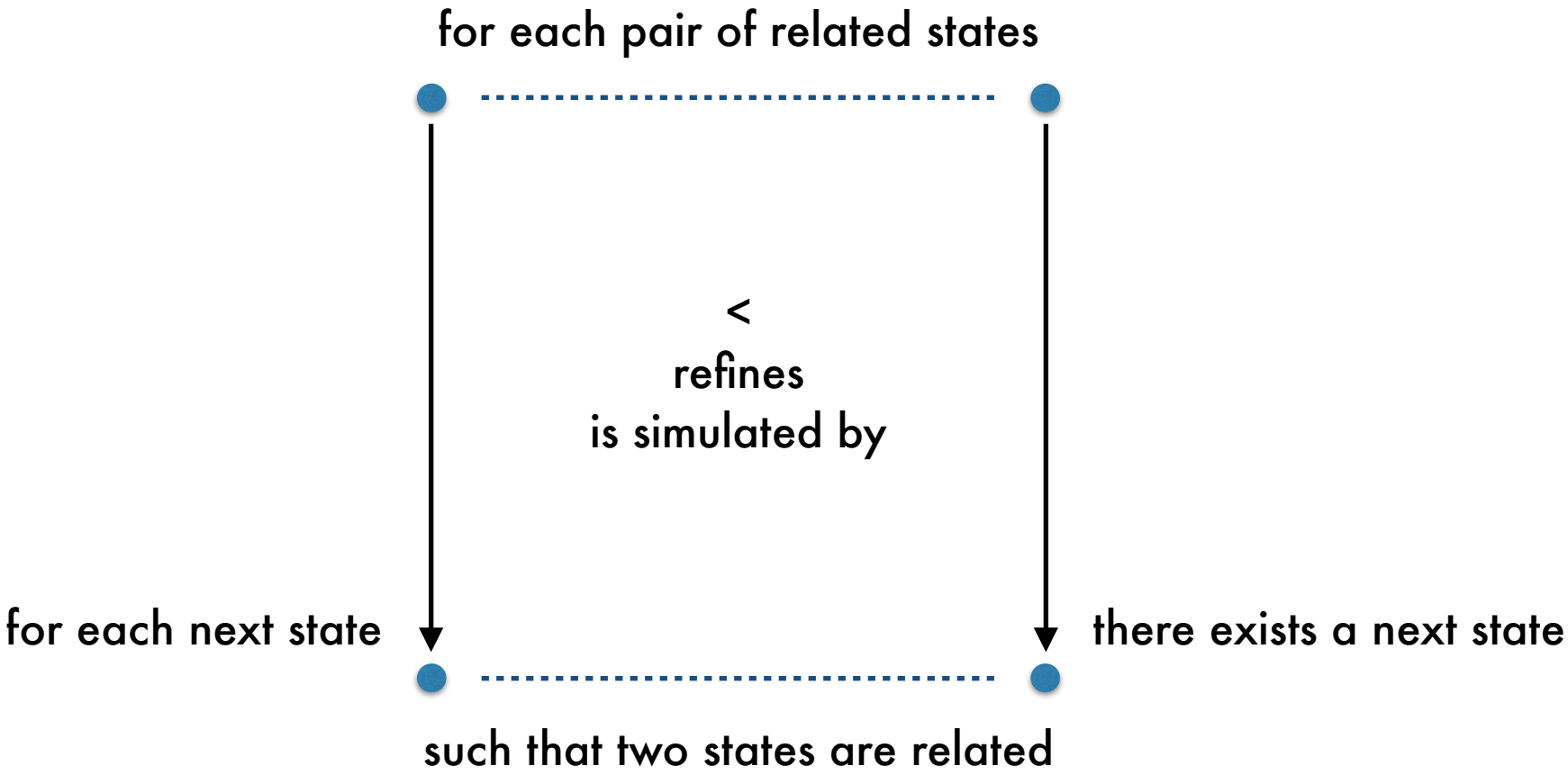
Bisimulation



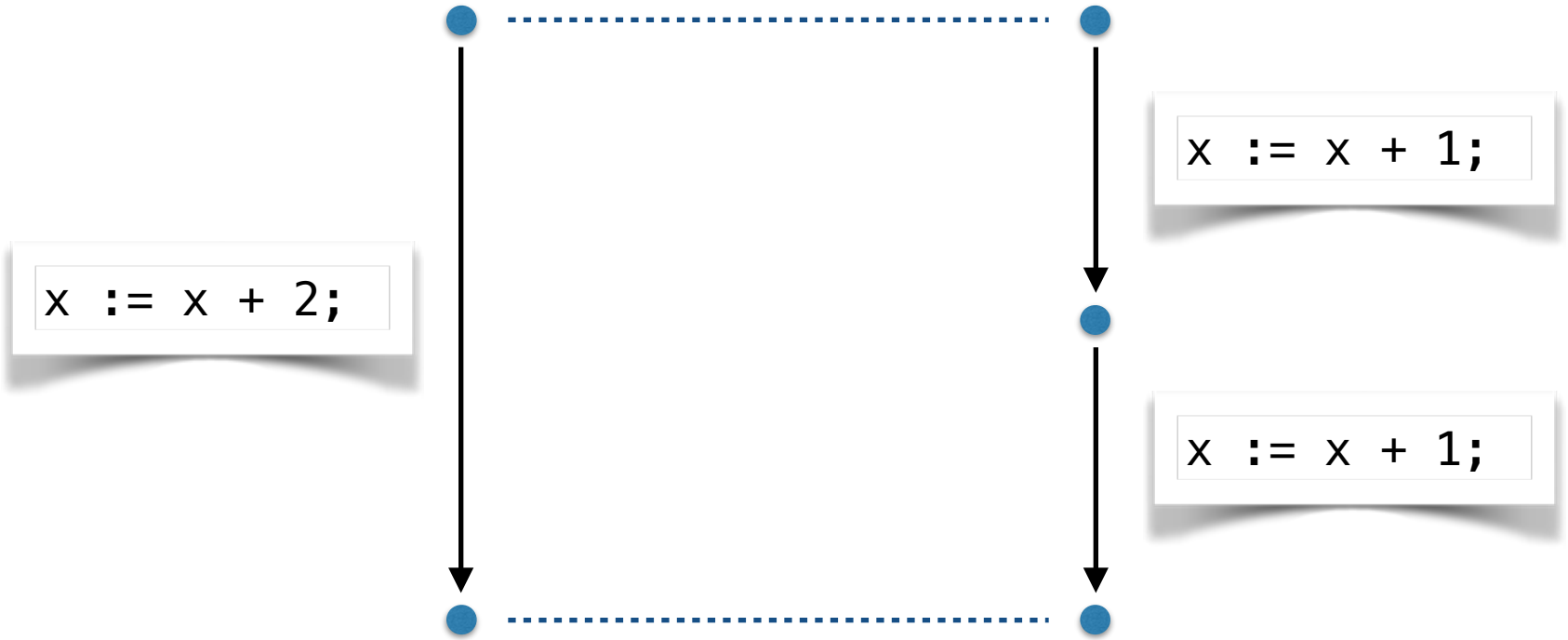
Bisimulation



Bisimulation

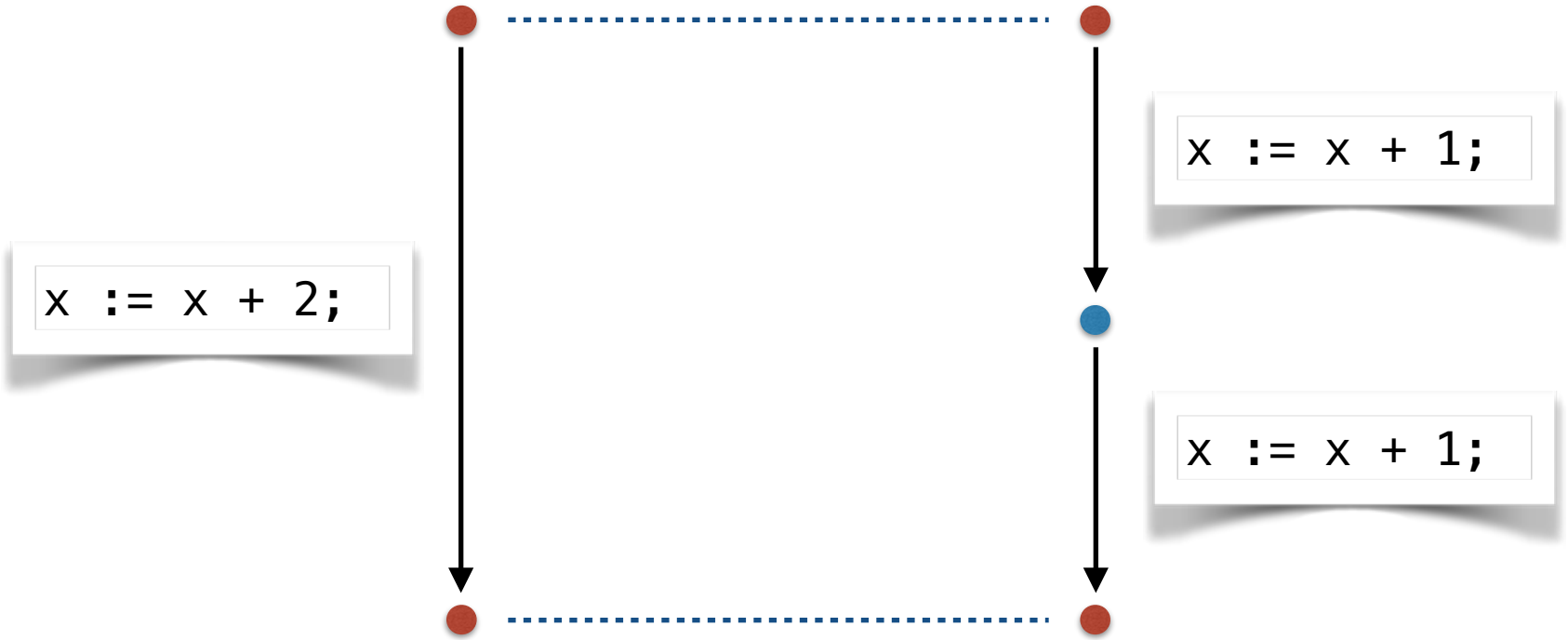


Bisimulation?

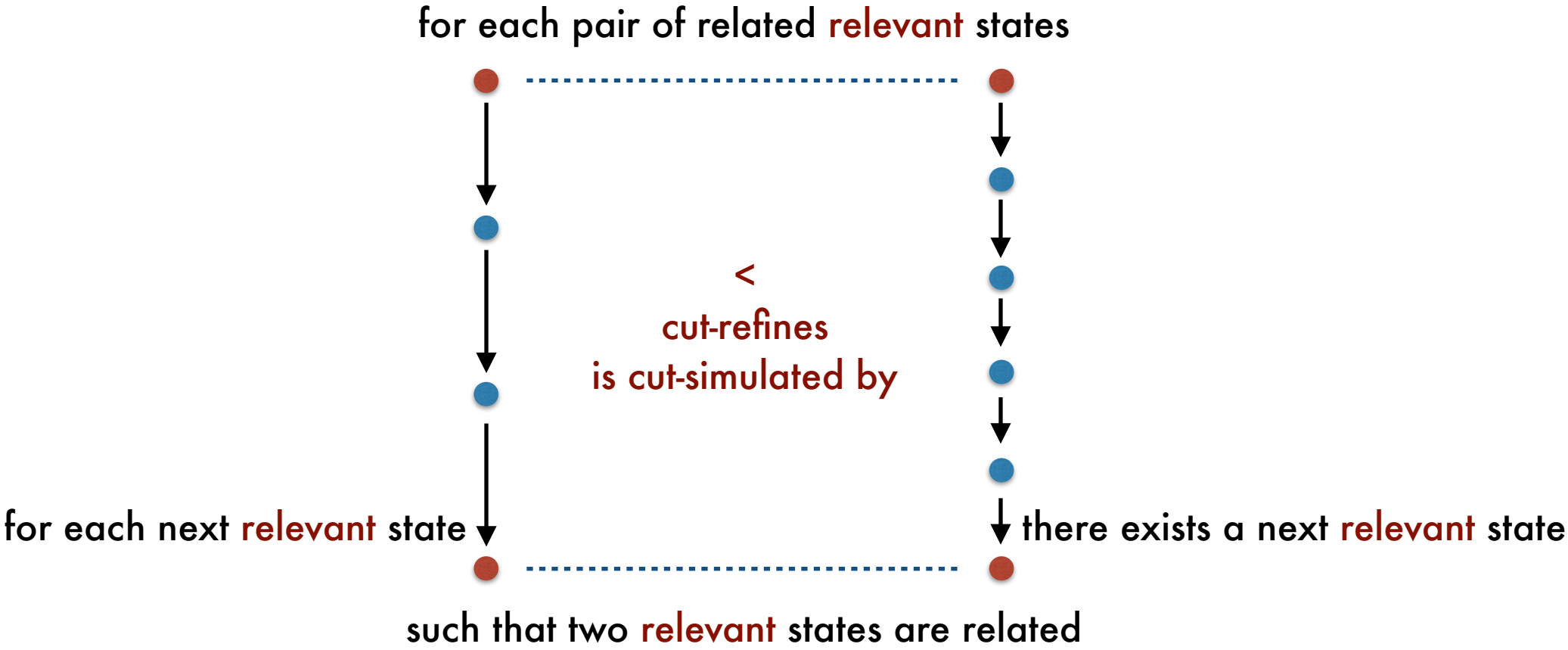


Cut-bisimulation

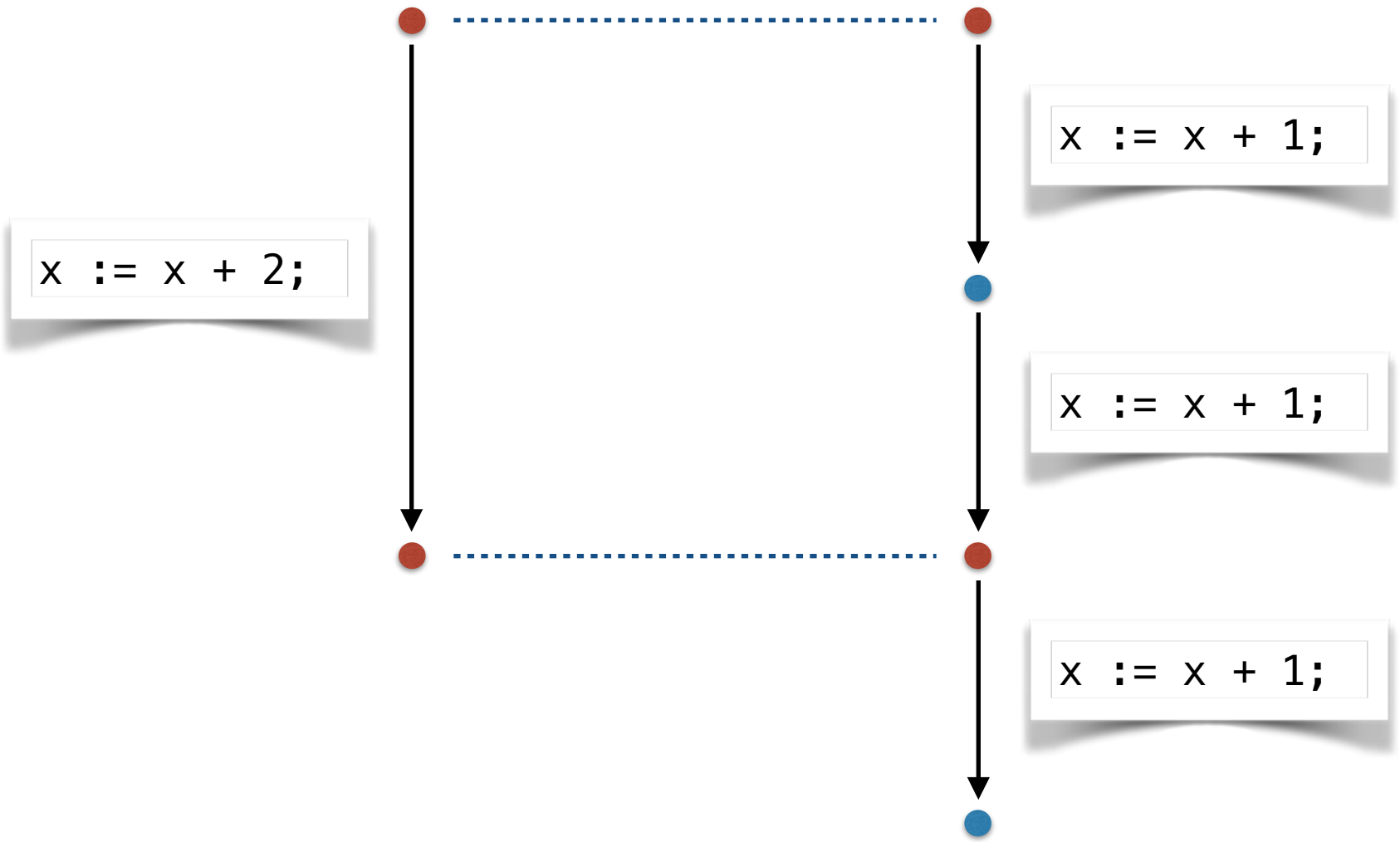
considering only relevant states



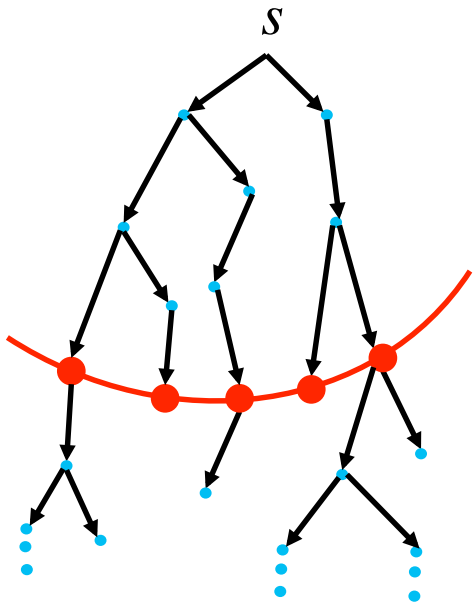
Cut-bisimulation



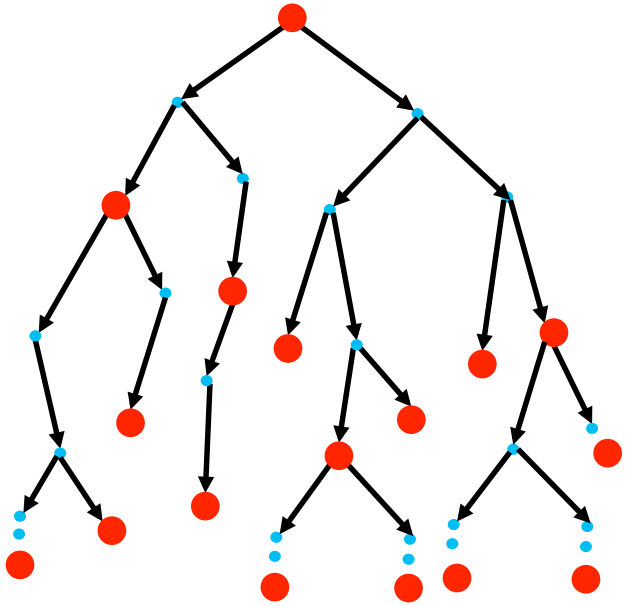
Well-formedness of relevant states



Cut



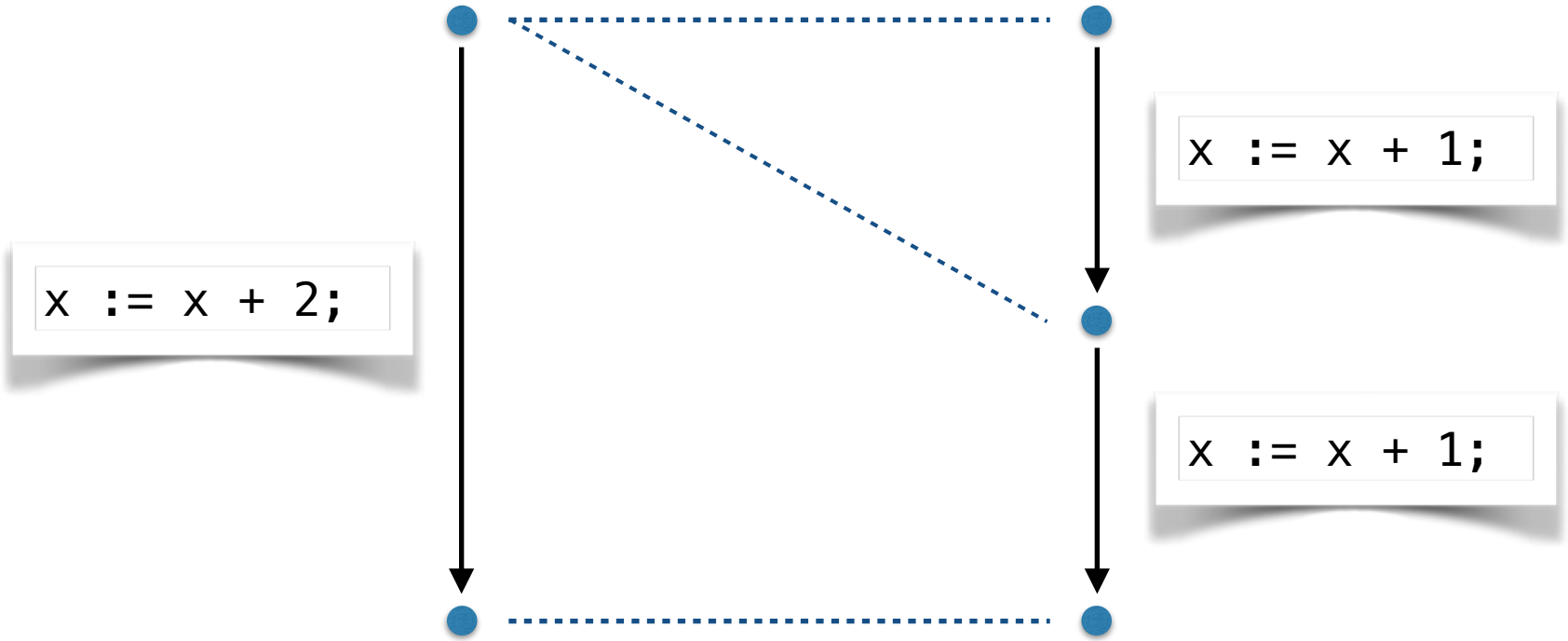
Cut for s



Cut for all

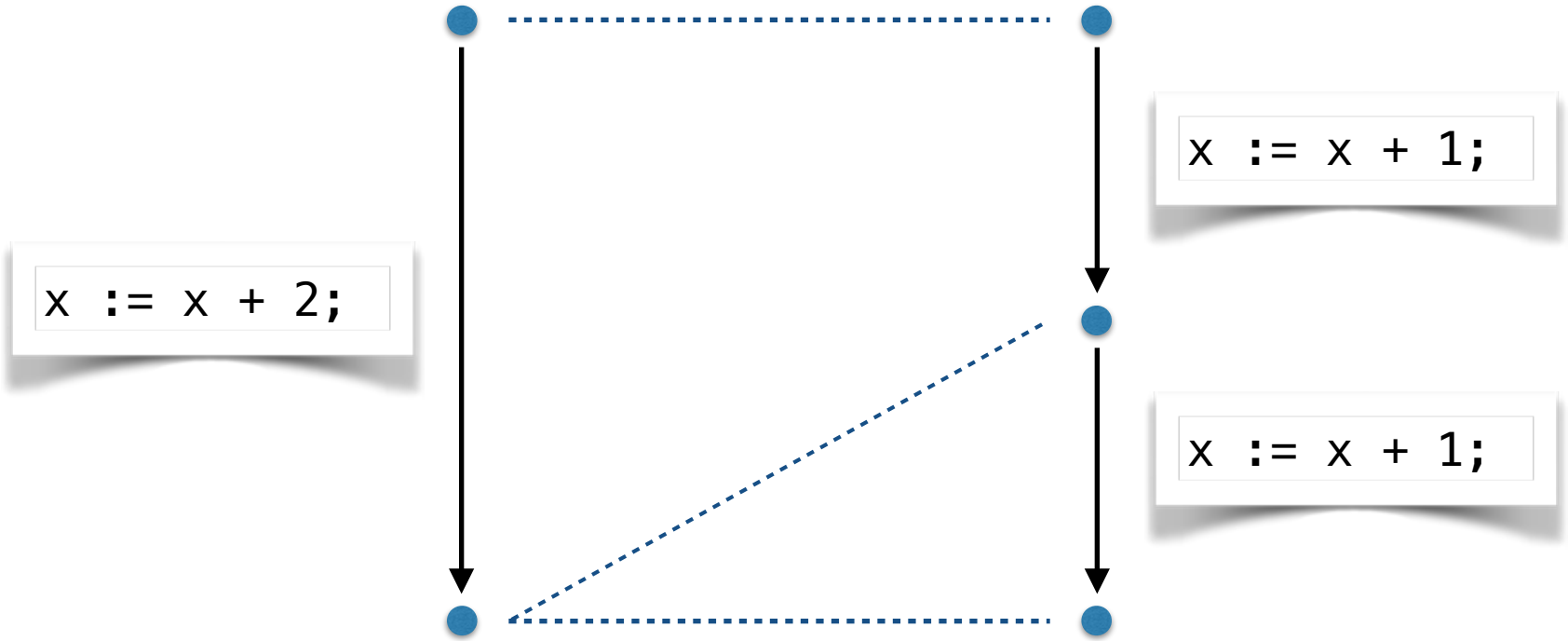
Stuttering bisimulation [BCG'88, dNV'90]

two possible stuttering bisimulations: 1 of 2

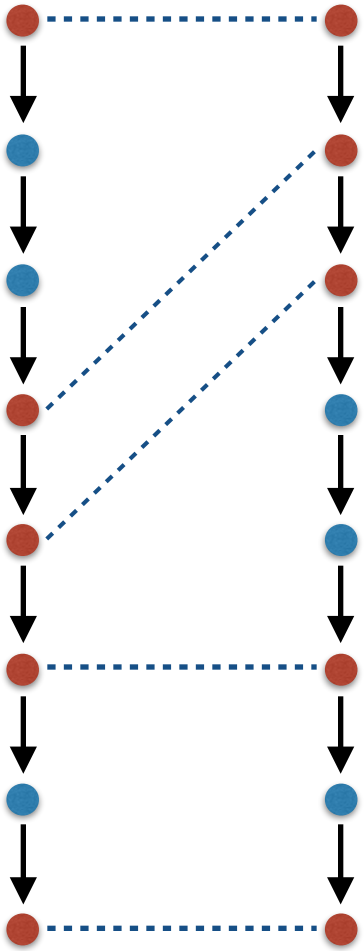


Stuttering bisimulation [BCG'88, dNV'90]

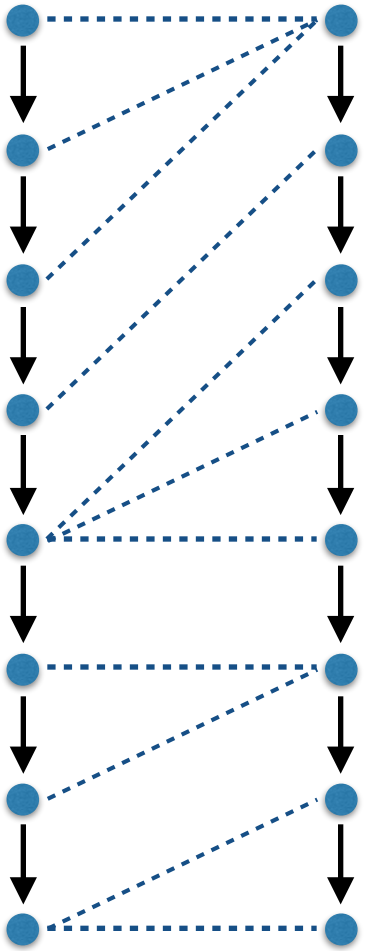
two possible stuttering bisimulations: 2 of 2



Cut-bisimulation vs stuttering bisimulation



vs



[Namjoshi'97]

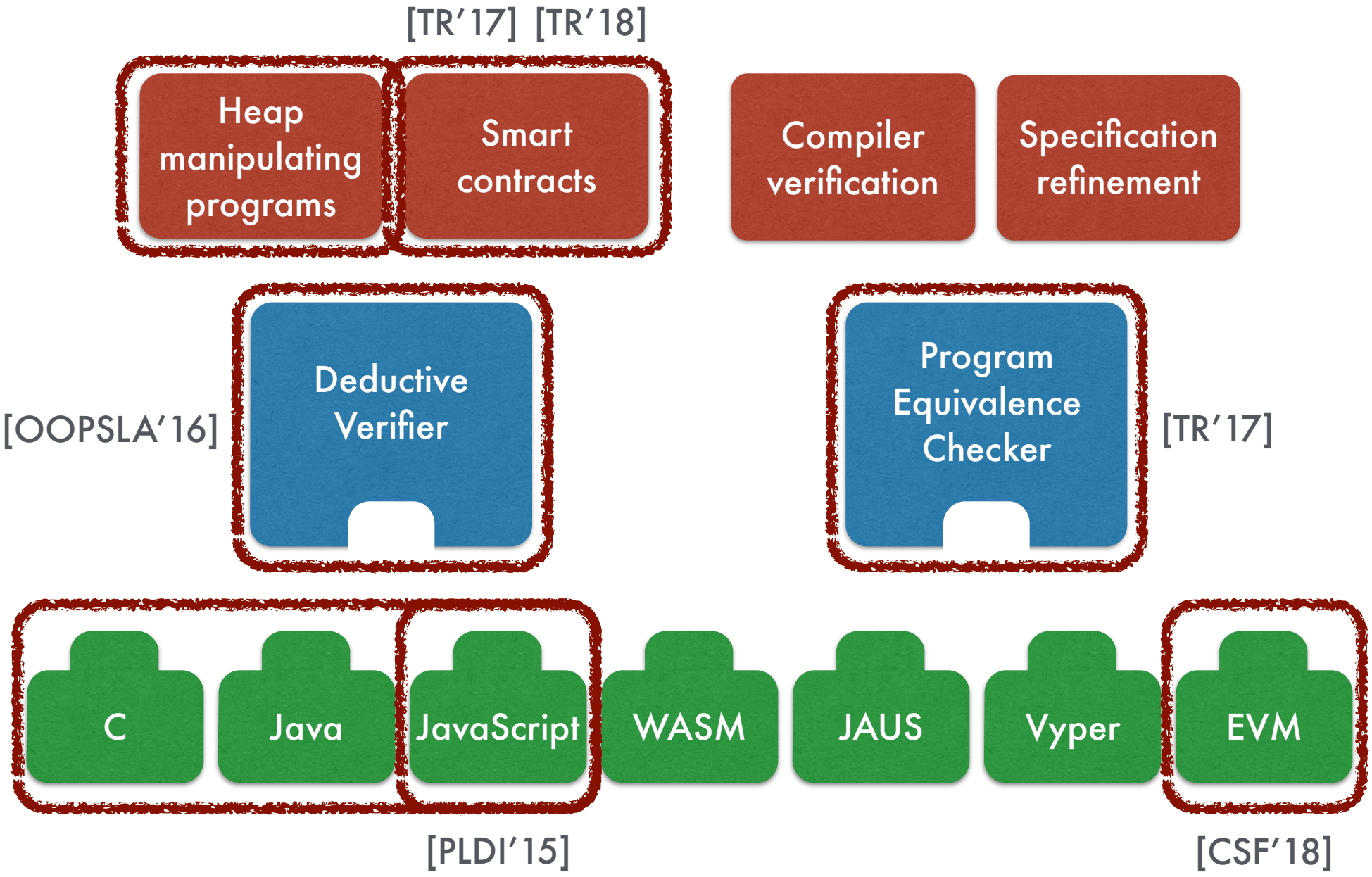
KEQ: universal program equivalence checker

- Inputs
  - Two language semantics
  - Two programs
  - A relation (called synchronization points)
- Output
  - True if the relation is a cut-(bi)simulation
- Can be used for translation validation for compiler

Translation validation [Pnueli *et al.*'98]

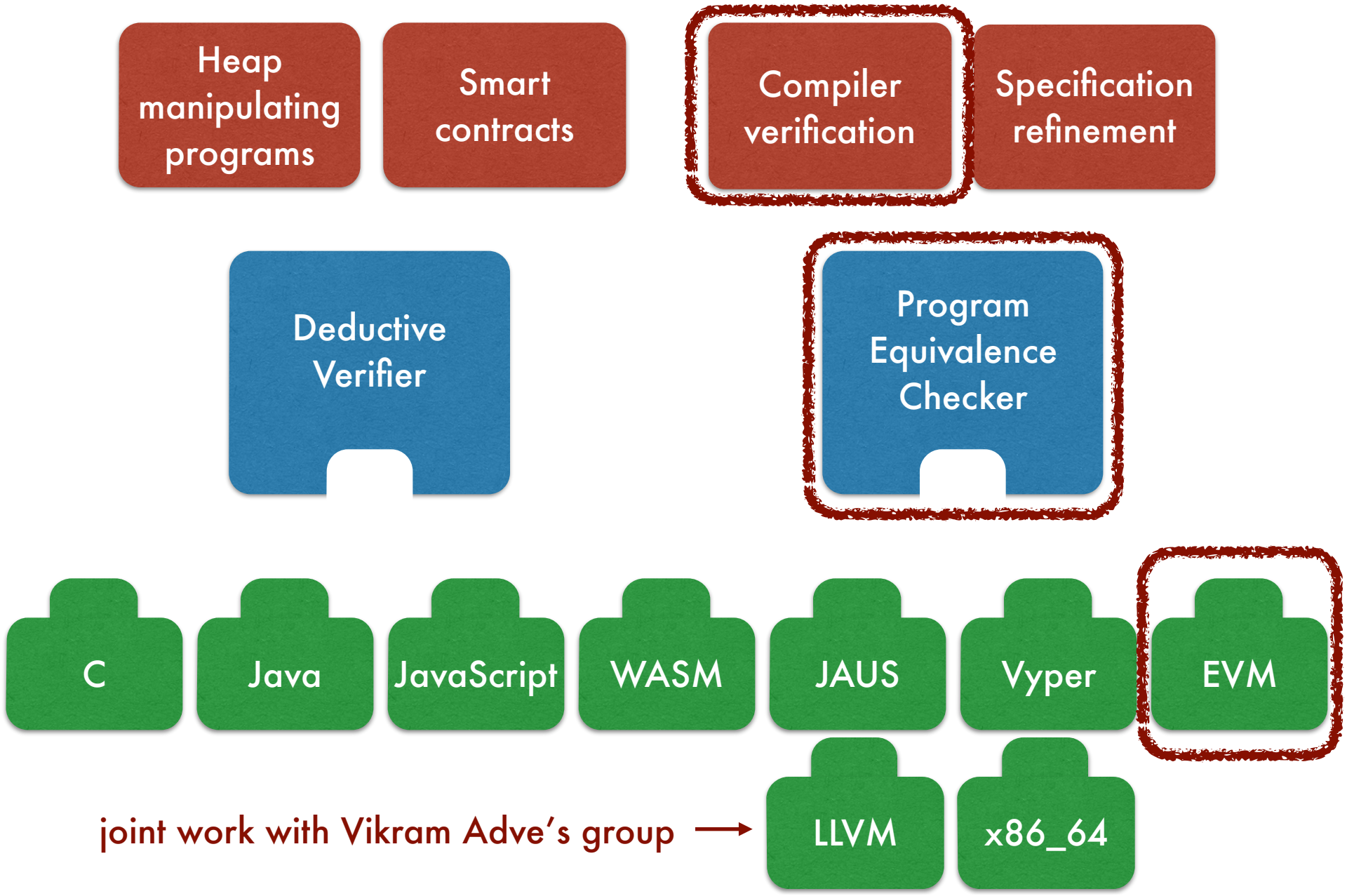
- Validate each instance of compilation
- Verify equivalence of input/output programs
  - KEQ can be used for checking equivalence
- Practical advantages for compiler verification

Current results



Proposed Work

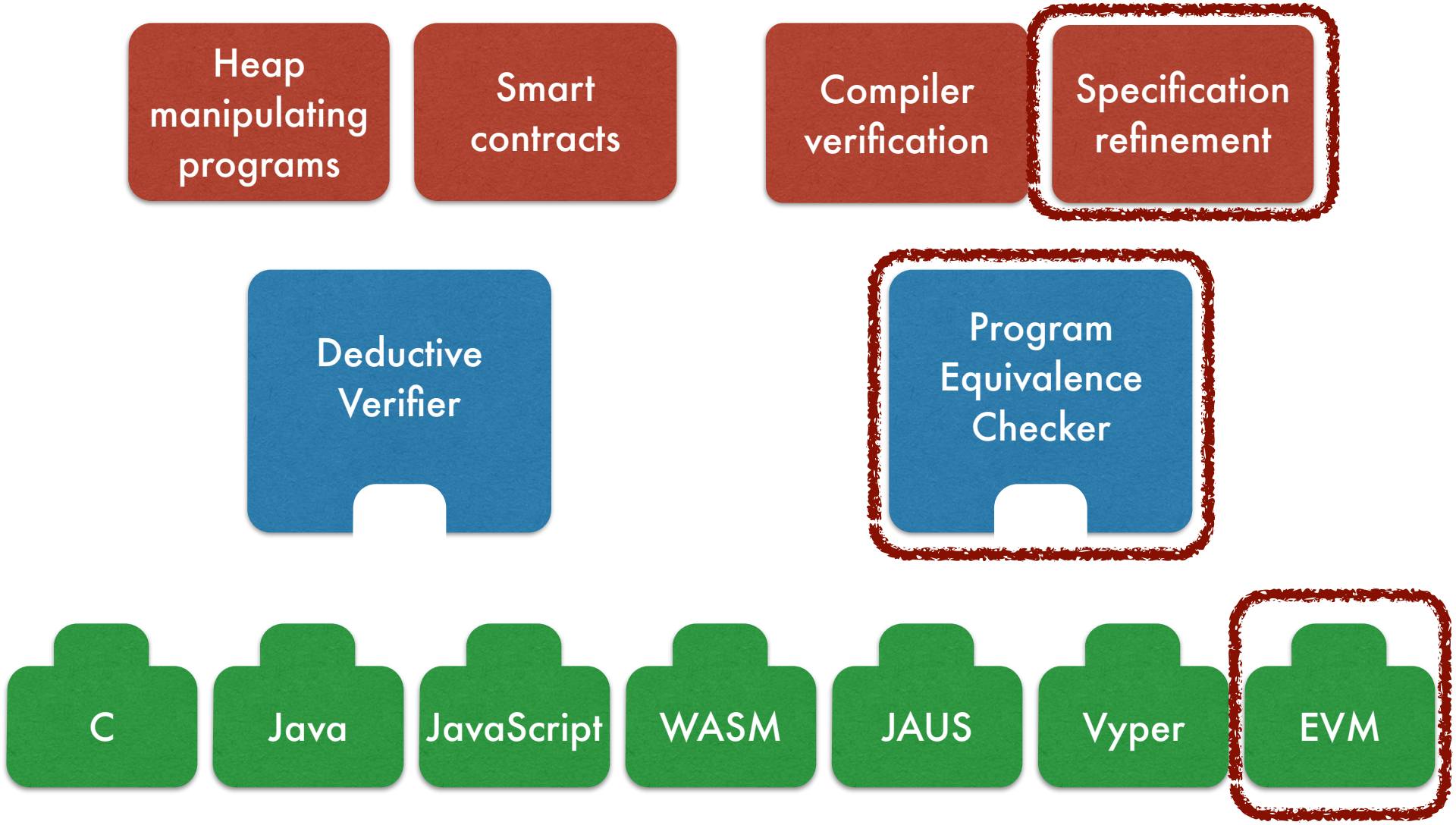
Proposal: translation validation for compiler



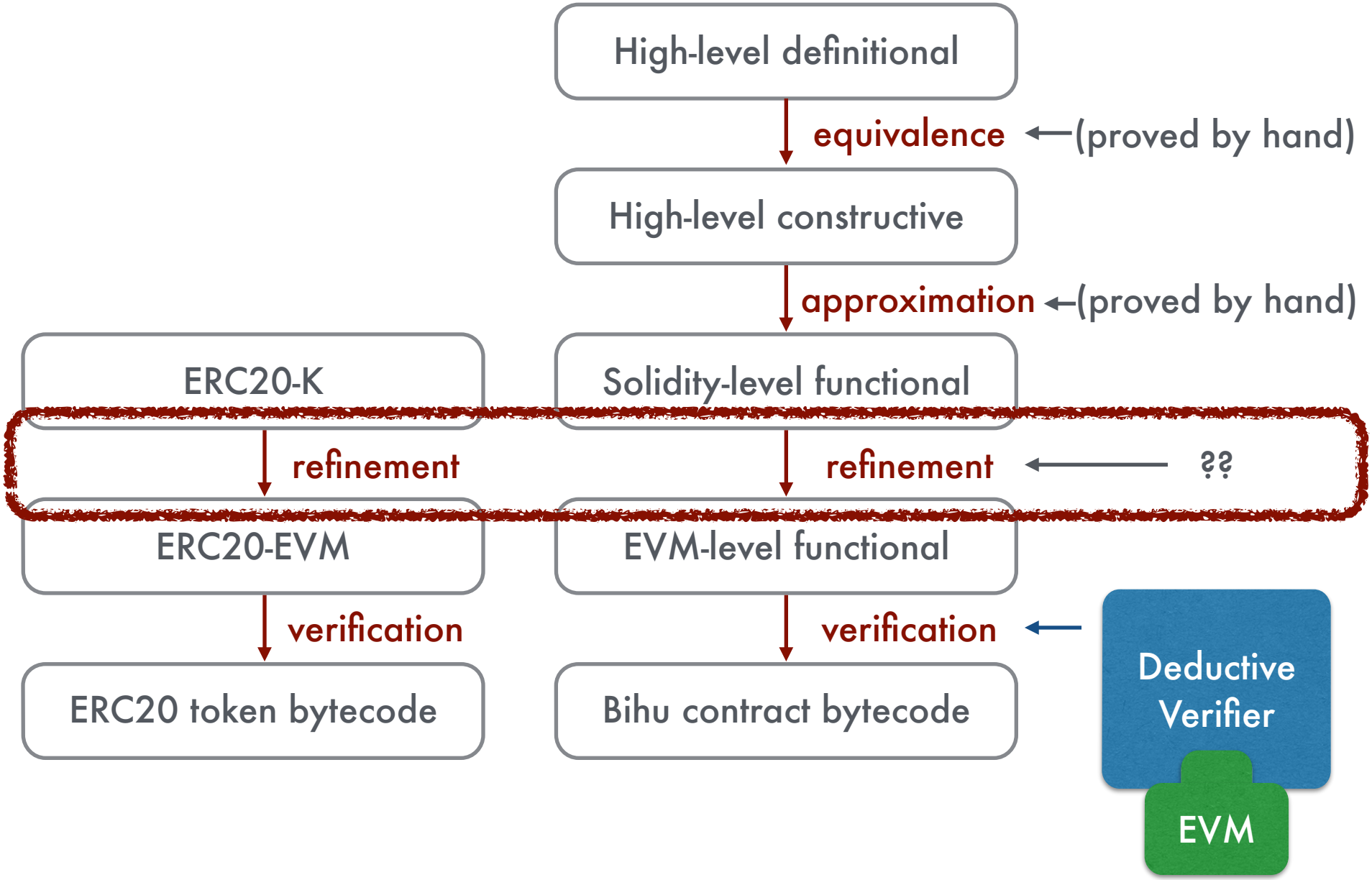
Translation validation for compiler

- Vyper compiler verification
- Verify compatibility between
  - Vyper and Solidity compilers
  - Different versions of the same compiler

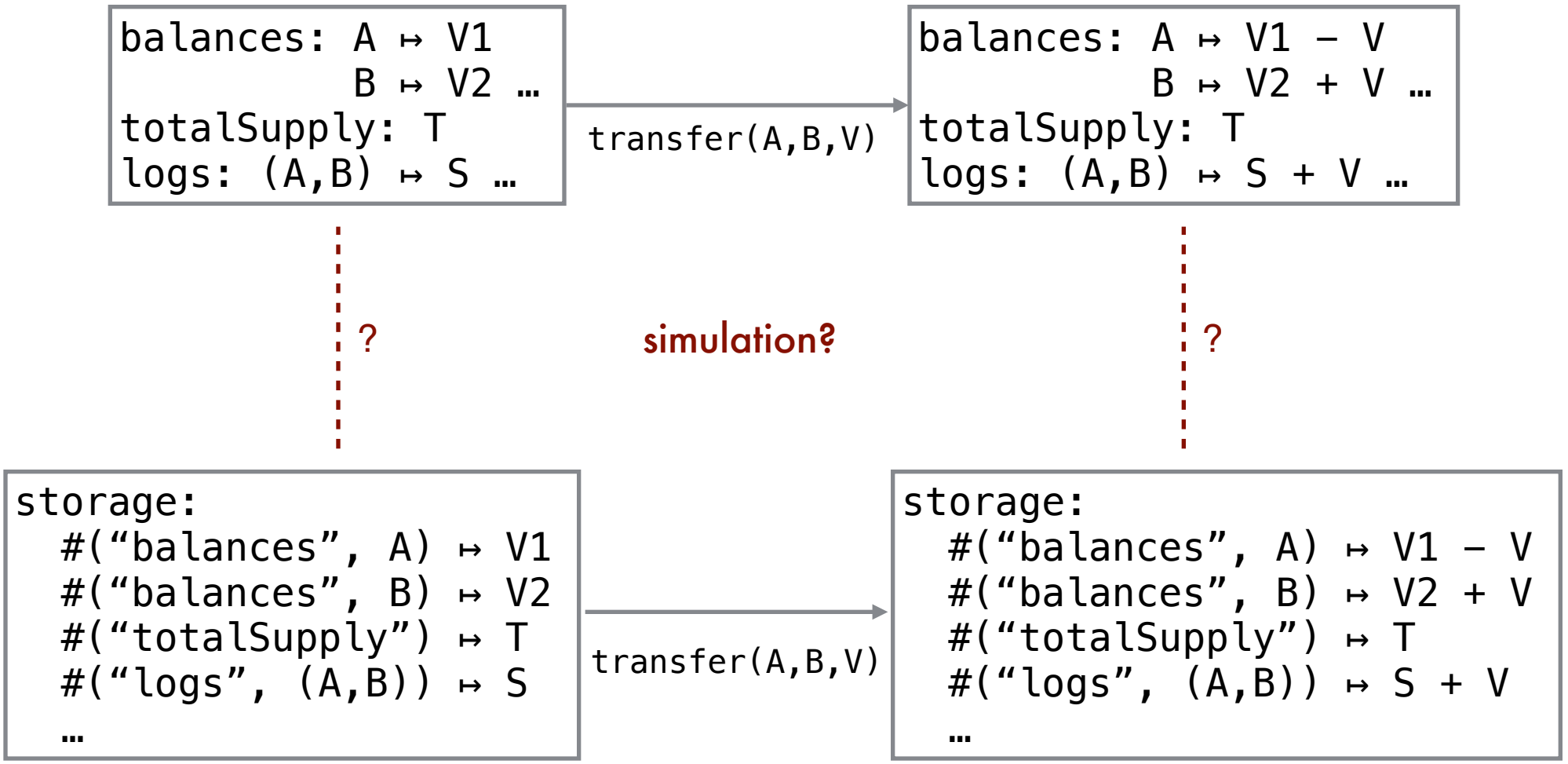
Proposal: verifying spec refinement



Verifying specification refinement

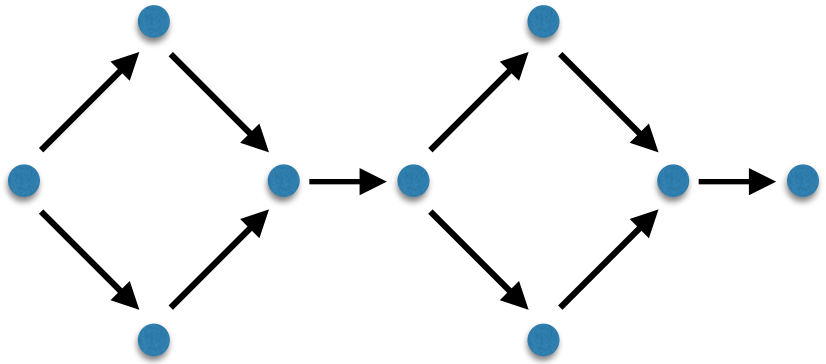


Soundness of specification refinement

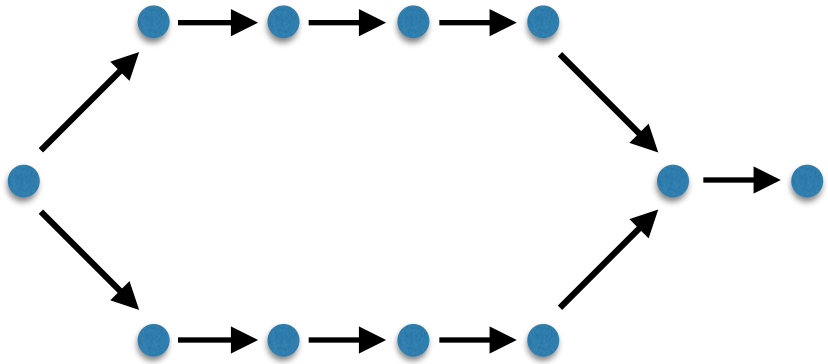


e.g., gas consumption is needed for the high-level spec?

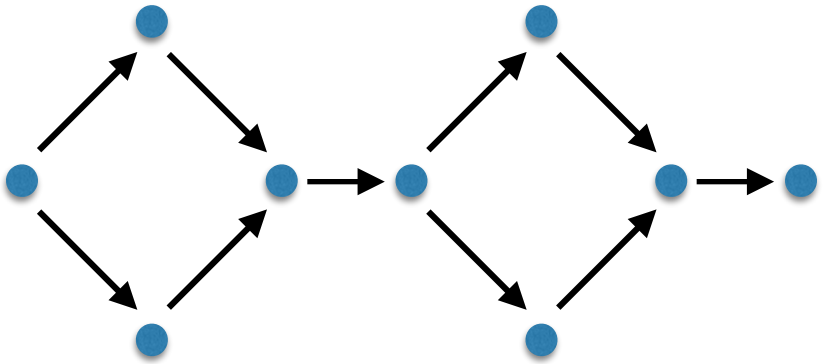
Soundness of specification refinement



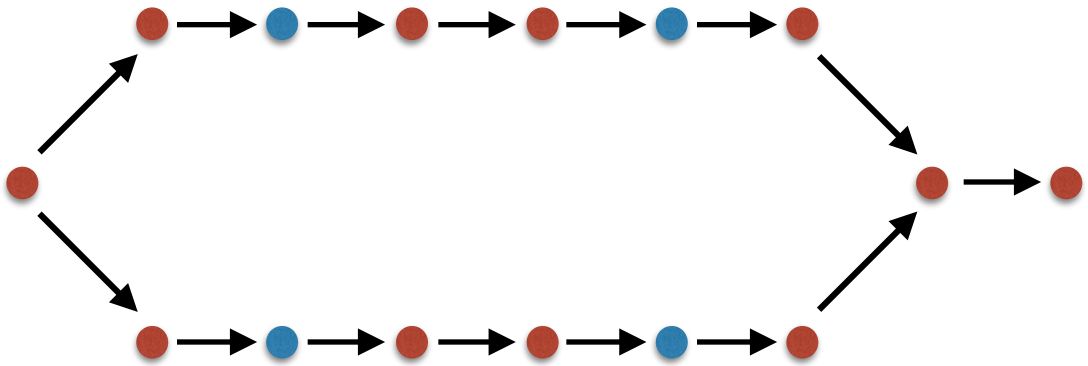
V simulates



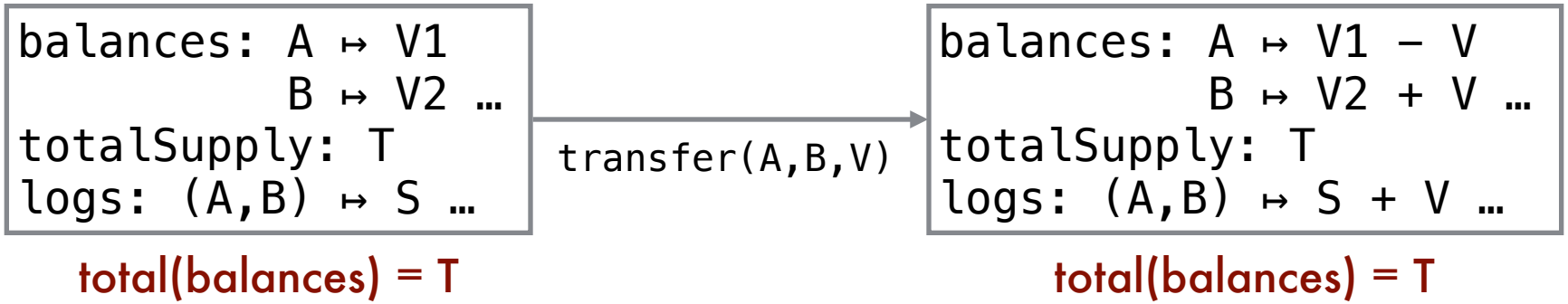
Soundness of specification refinement

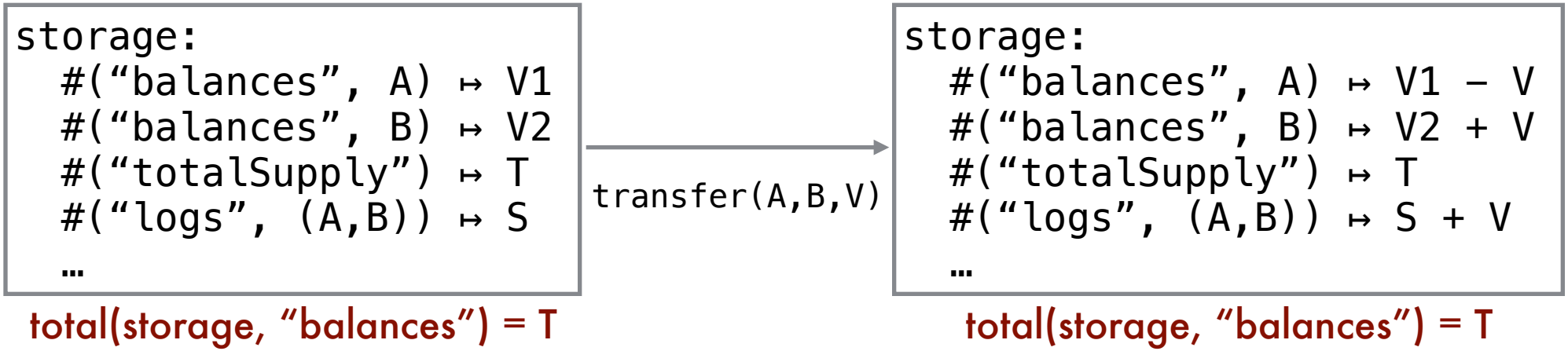


∨ cut-simulates



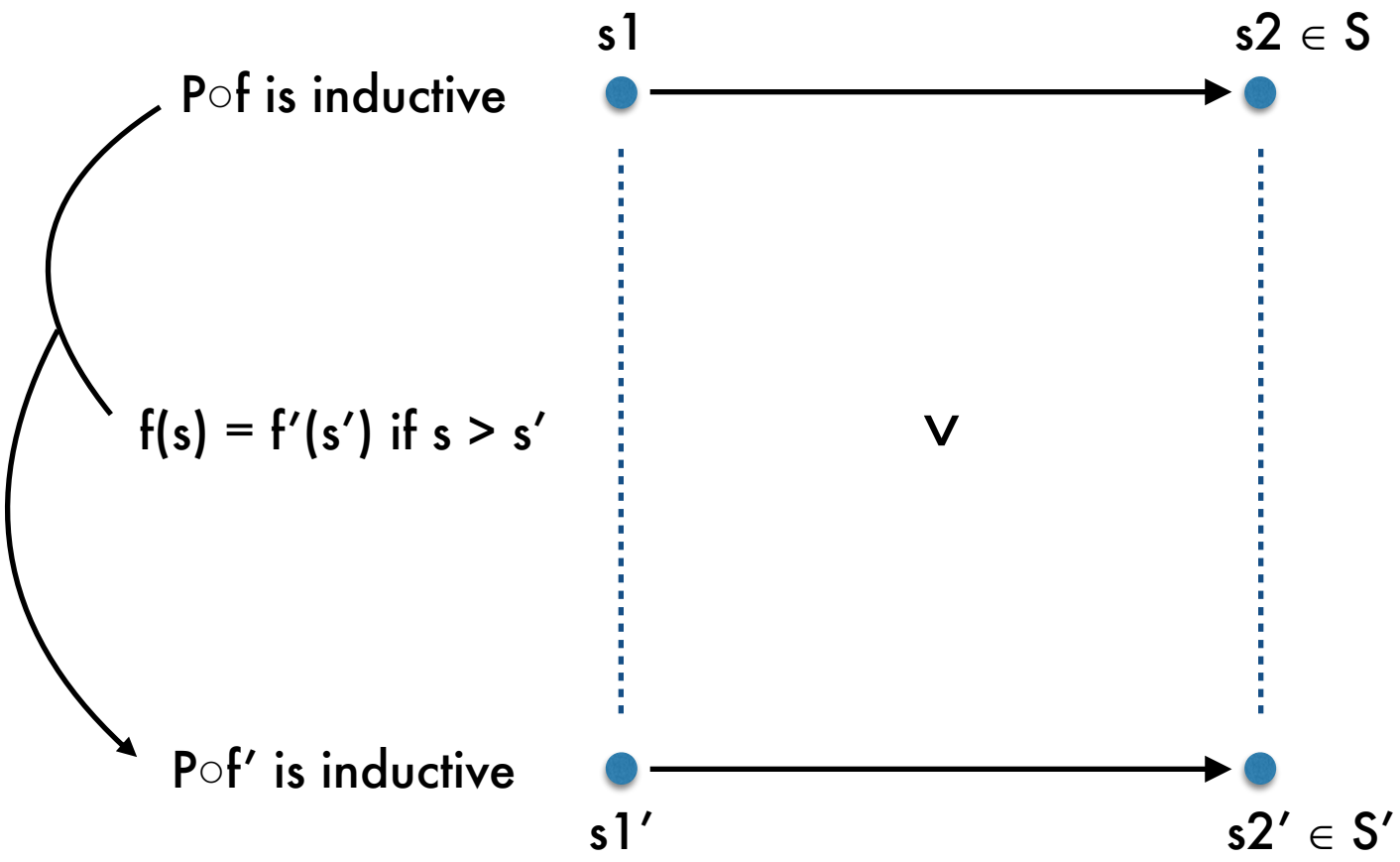
Proposal: property preservation





How to preserve inductive properties?

Property preservation

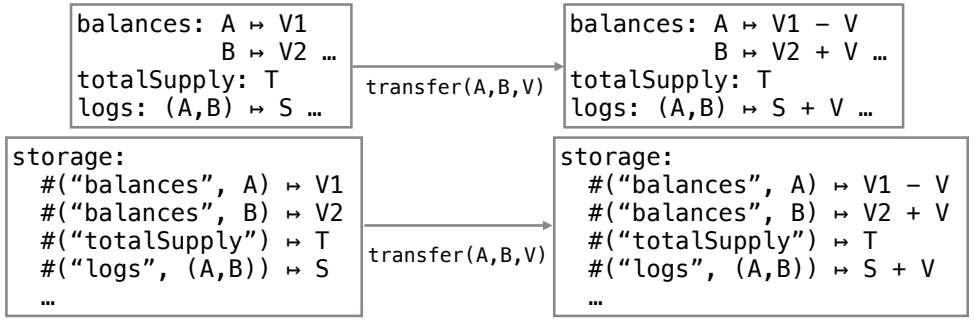


$P : D \rightarrow bool$
$f : S \rightarrow D$
$f' : S' \rightarrow D$

$P \circ f(s) = P(f(s))$
$P \circ f'(s') = P(f'(s'))$

Property preservation

$$P_{\texttt{total}} \stackrel{\text{def}}{=} \texttt{total}(\texttt{balances}) = \texttt{totalSupply}$$
$$P_{\texttt{total}'} \stackrel{\text{def}}{=} \texttt{total}'(\texttt{storage}) = \texttt{totalSupply}$$



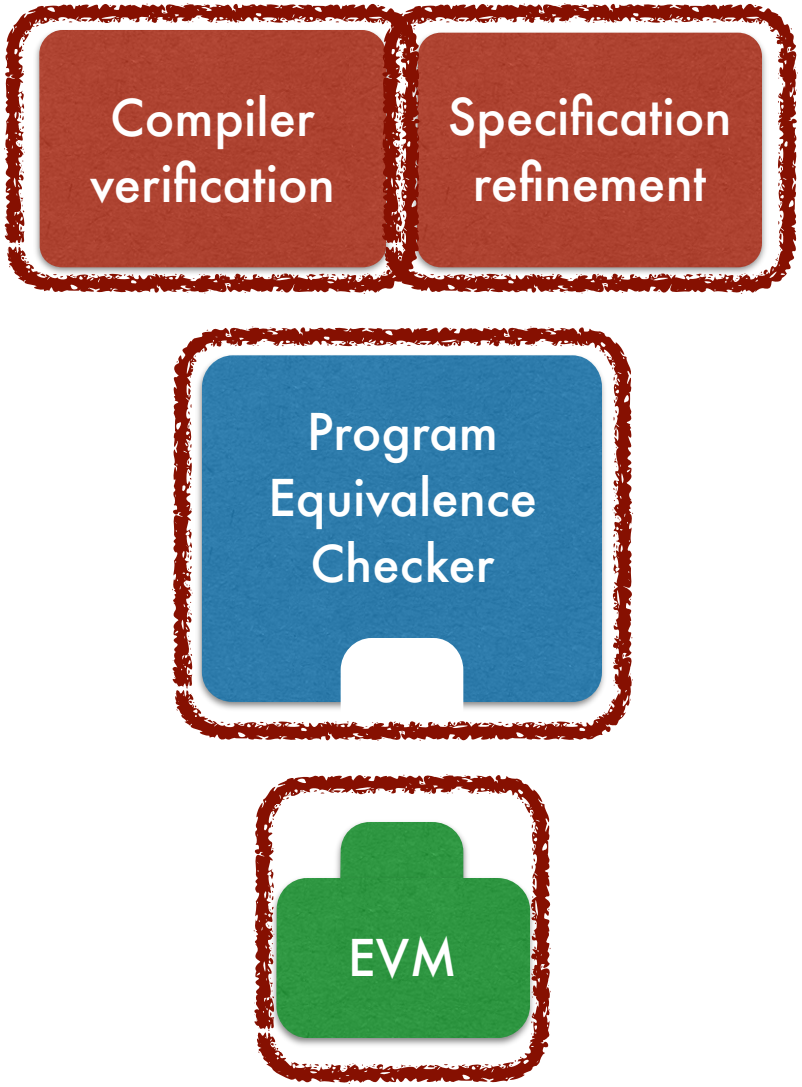
$$\texttt{total}(B) = \Sigma\{v \mid k \mapsto v \in B\}$$
$$\texttt{total}'(B') = \Sigma\{v' \mid k' \mapsto v' \in B' \wedge \exists k.\ \mathsf{hash}(k) = k'\}$$

$$\texttt{total}(\texttt{balances}) = \texttt{total}'(\texttt{storage})$$

$$\texttt{balances} > \texttt{storage} \iff$$
$$\forall k \in \mathit{dom}(\texttt{balances}).\ \texttt{balances}(k) = \texttt{storage}(\mathsf{hash}(k))$$
$$\wedge\ \mathit{dom}(\texttt{balances}) = \{k' \mid k' \in \mathit{dom}(\texttt{storage}) \wedge \exists k.\ \mathsf{hash}(k) = k'\}$$

Proposed work

- Translation validation for compiler
- Verifying specification refinement
- Property preservation of cut-simulation



Conclusion

- Goal: demonstrate/improve the scalability of language-parametric formal methods
- Specified various real-world language semantics
- Improved existing universal deductive program verifier
  - Instantiated with various languages, and applied to high-profile real-world applications
- Newly developed a cross-language program equivalence checker with a novel notion of cut-bisimulation
  - Will evaluate in the domain of Ethereum smart contracts

Thank you