한양대학교 프로그램 합성 연구소개

이 우 석

한양대학교 ERICA 소프트웨어학부

2023 소프트웨어재난연구센터 여름 정기 워크샵

2023. 07. 05

HANYANG UNIVERSITY

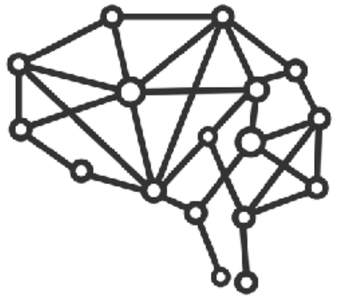
3차년도 하반기 연구 진행상황

- 프로그램 합성 원천 기술
  - 요약해석 기반 합성 가속화 연구 (with 서울대 이광근교수님)
  - LLM (Large Language Model) 솔루션을 참고하는 프로그램 합성 (조한결 박사과정)
- 프로그램 합성 기술 응용
  - 일반화를 보장하는 프로그램 오류 자동수정(이제형 박사과정 with 2세부 이주용교수님)
- 의의
  - 3그룹 : 모델 생성 + 스펙추론에 사용가능한 프로그램 합성 성능 향상
  - 2그룹 : 패치의 품질 제고 목표

요약해석 기반 합성 가속화 연구

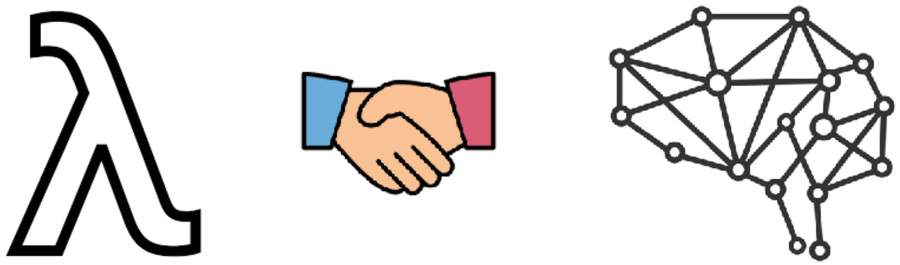
- 프로그램 합성기의 성능을 정적 분석으로 가속화
  - Yongho Yoon, Woosuk Lee, Kwangkeun Yi, Inductive Program Synthesis via Iterative Forward-Backward Abstract Interpretation, PLDI 2023
- 의의: 프로그램 정적 분석기 ⇨ 프로그램 합성기
  - (이론적으로는) 임의의 정확한 정적 분석기가 주어지면 이로부터 고성능 프로그램 합성기가 도출됨
  - 기존에 오랫동안 연구된 정적분석 정확도 향상 기술들을 합성을 위해 사용가능
  - 현재 구현은 비트벡터 프로그램 합성에 특화되어있으나 다른 도메인(문자열, 정수 등)으로 확장 계획 중

LLM 솔루션을 참고하는 프로그램 합성



- 대형언어모델 (GPT, BERT, T5, …) 기반 프로그램 자동 생성
  - 패턴화된 코드 작성에 매우 뛰어나며 비교적 큰 코드 생성 가능
  - 널리 쓰이는 코딩 관습을 따름
  - 그러나 **올바름 보장안됨** (사용자 요구조건 불만족)
  - 학습데이터에 존재하지 않는 새로운 코드 작성 잘 못함
- 탐색알고리즘 기반 프로그램 자동 생성
  - 항상 사용자 요구조건 만족하는 코드 생성
  - 사람이 못찾는 더 나은 코드도 발견 (예: 신경망, LLVM, 동형암호 프로그램 최적화 발견)
  - 큰 코드는 생성 못함

LLM 솔루션을 참고하는 프로그램 합성



- 입출력예제 기반 함수형 프로그램 합성기 Trio 개발 (POPL'23). 큰 코드 생성 잘 못함
- ChatGPT는 잘하나?
  - 자연어 기술 + 입출력 예제를 제공 시 오답 프로그램이 종종 도출됨
  - 비록 오답이지만 정답과 유사하거나, 정답에 사용되는 관용구(code idiom) 포함
- ChatGPT가 생성한 프로그램과 비슷한 솔루션을 찾도록 Trio 의 탐색 안내
- 기대효과
  - GPT가 보장하지 못하는 입출력 예제에 대한 올바름 보장. 과적합 회피

ChatGPT의 오답 예

- Q : give me ocaml function that takes the last n elements of a list l.
  f [1;2;3;4;5] 3 = [3;4;5],
  f [1;2;3] 1 = [3]
  f [3;2;1] 2 = [2;1]

순서바꿈. l n 이어야.

- A :
```
let rec f n l =
        match l with
        | [] -> []
        | hd :: tl ->
          if n <= 0 then l
          else f (n - 1) tl
```

```
if n = 0 then []
else if List.length tl < n
then l
else f tl (n - 1)
```

Correct solution

일반화를 보장하는 프로그램 오류 자동수정

- 2차년도에 프로그램 오류 수정도구 Moses 개발
  - 과적합 흔히 발생 (생성된 패치가 바라던 패치가 아닌)
- 대부분의 오류 수정 도구들: 그럴싸한 패치를 생성하기 위해 휴리스틱 백화점에 의존
  - 대상 실험 벤치마크에 예민하고 깨지기 쉬움
  - 패치가 생성되더라도 얼마만큼 믿을 수 있는지 정보 못줌
- 목표: Crash 버그 패치 생성 시 확률적 보장 주기
  - 바람직한 패치일 확률 X%, 신뢰도 Y%
  - PAC learning theory 접목

SIMBA: Inductive Program Synthesis via Iterative Forward-Backward Abstract Interpretation

Yongho Yoon, Seoul National University & Sparrow Co., Ltd.
Woosuk Lee, Hanyang University
Kwangkeun Yi, Seoul National University

20/06/2023 @ PLDI 2023

PLDI Orlando 2023

Sparrow

Inductive Program Synthesis

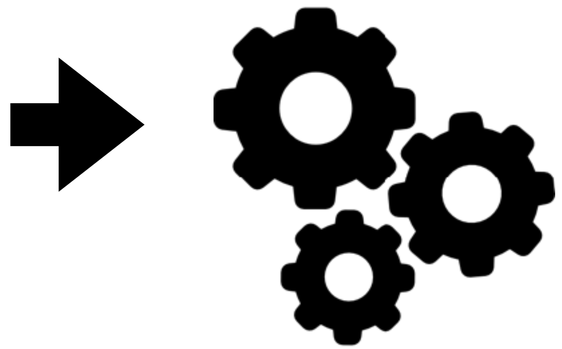
in SyGuS Format

Syntactic Spec

$$
\begin{array}{rcl}
S & \rightarrow & x \mid 0001_2 \\
 & \mid & S \wedge S \mid S \vee S \mid S \oplus S \\
 & \mid & S + S \mid S \times S \mid S / S \mid S >> S
\end{array}
$$

Semantic Spec

$$f(1011_2) = 0011_2$$



$$f(x) = ((x + 0001_2) \oplus x) >> 0001_2$$

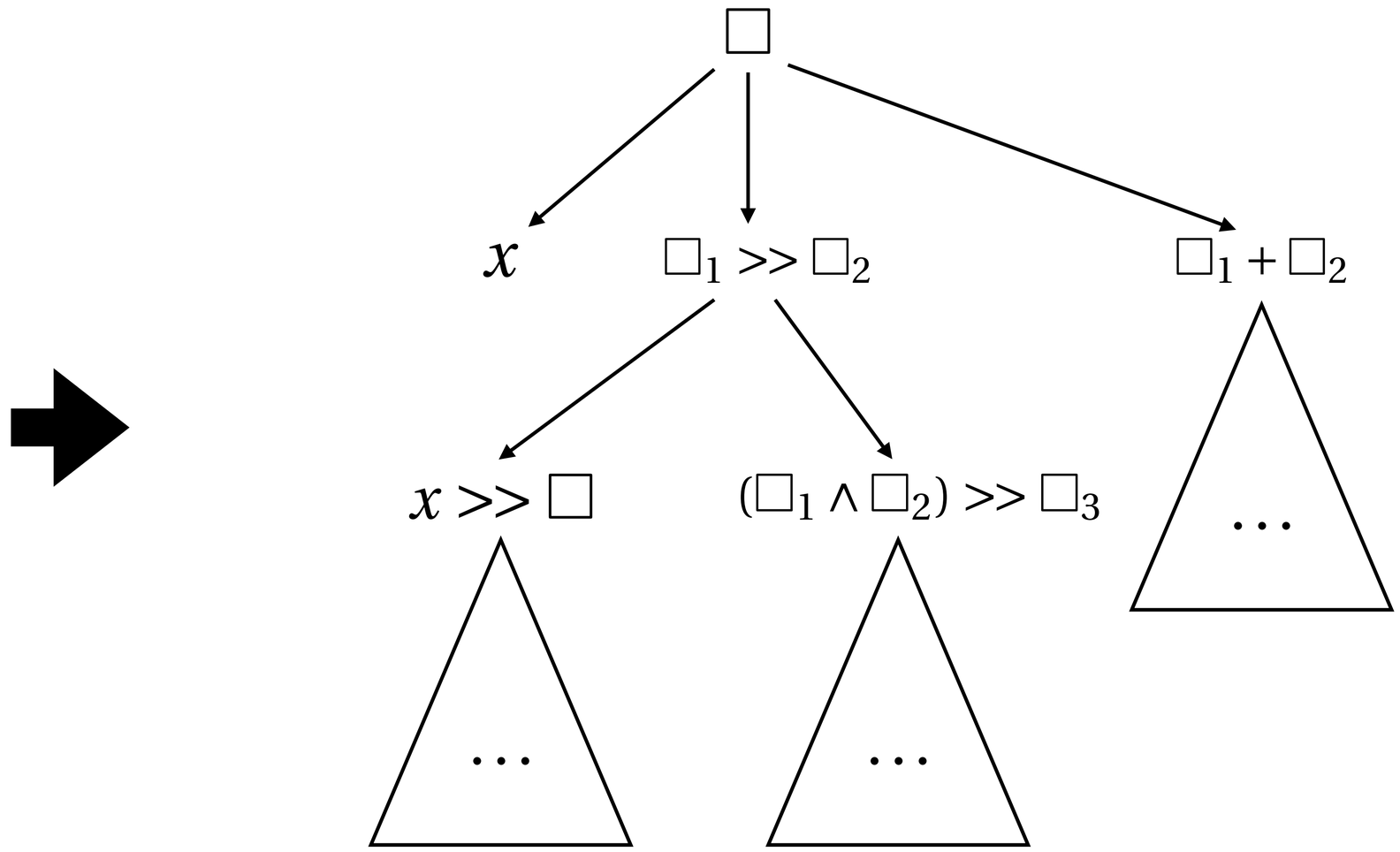
Pruning by Feasibility

Syntactic Spec

$$
\begin{aligned}
S \quad &\rightarrow \quad x \mid 0001_2 \\
&\mid \quad S \wedge S \mid S \vee S \mid S \oplus S \\
&\mid \quad S + S \mid S \times S \mid S / S \mid S >> S
\end{aligned}
$$

Semantic Spec

$f(1011_2) = 0011_2$



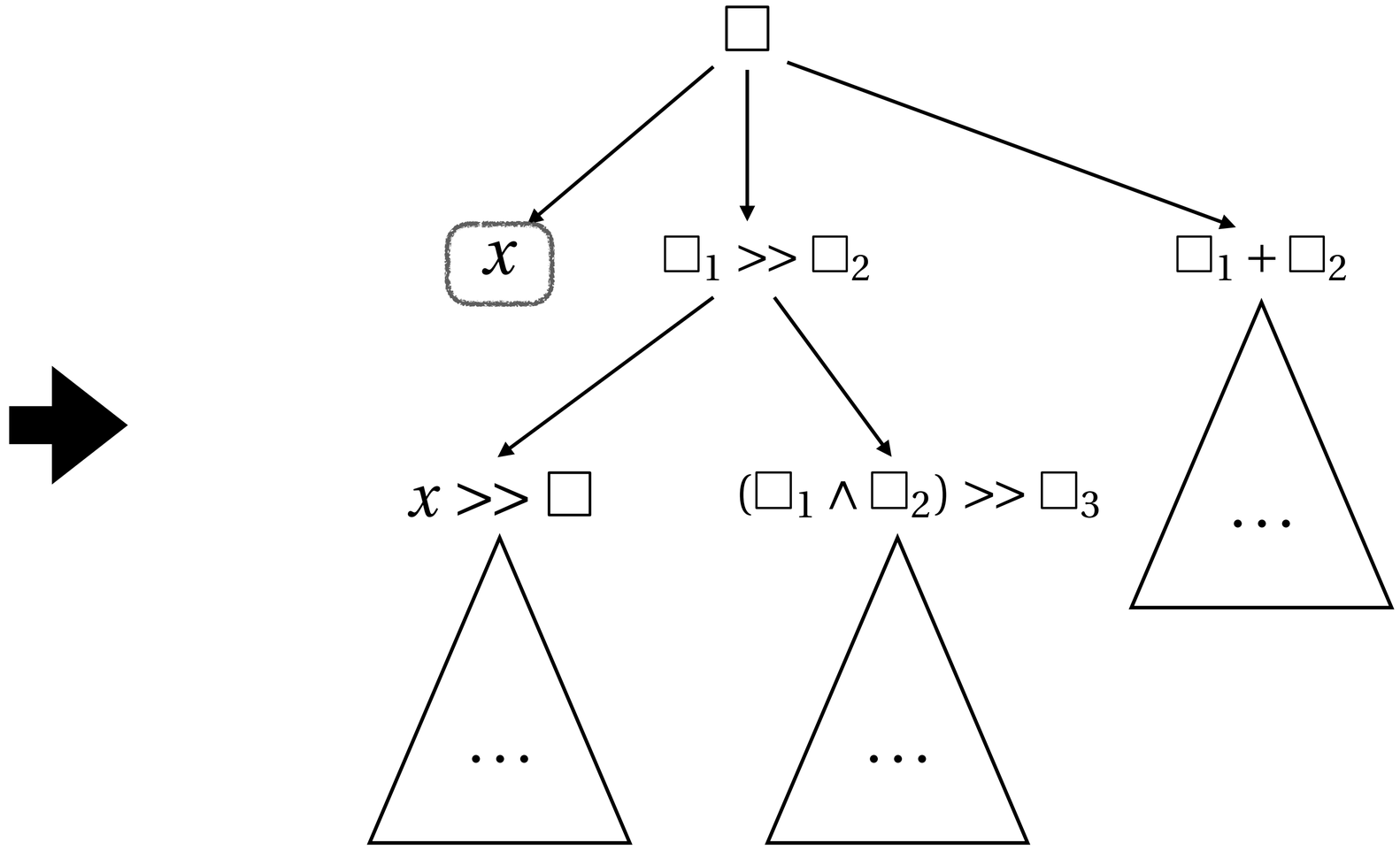
Pruning by Feasibility

Syntactic Spec

$$
\begin{aligned}
S \quad &\rightarrow \quad x \mid 0001_2 \\
&\mid \quad S \wedge S \mid S \vee S \mid S \oplus S \\
&\mid \quad S + S \mid S \times S \mid S / S \mid S >> S
\end{aligned}
$$

Semantic Spec

$f(1011_2) = 0011_2$



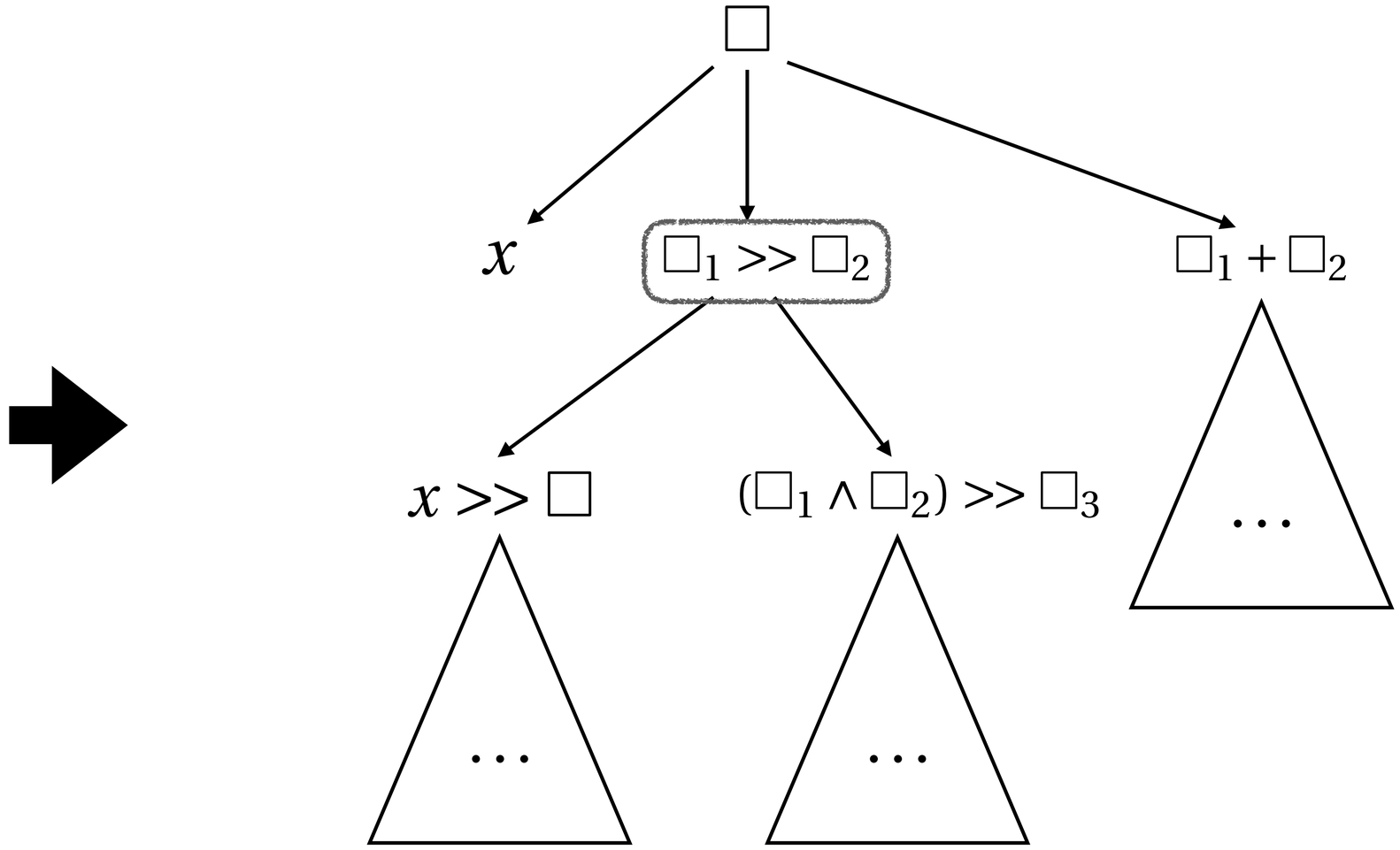
Pruning by Feasibility

Syntactic Spec

$$
\begin{aligned}
S \quad &\rightarrow \quad x \mid 0001_2 \\
&\mid \quad S \wedge S \mid S \vee S \mid S \oplus S \\
&\mid \quad S + S \mid S \times S \mid S / S \mid S >> S
\end{aligned}
$$

Semantic Spec

$f(1011_2) = 0011_2$



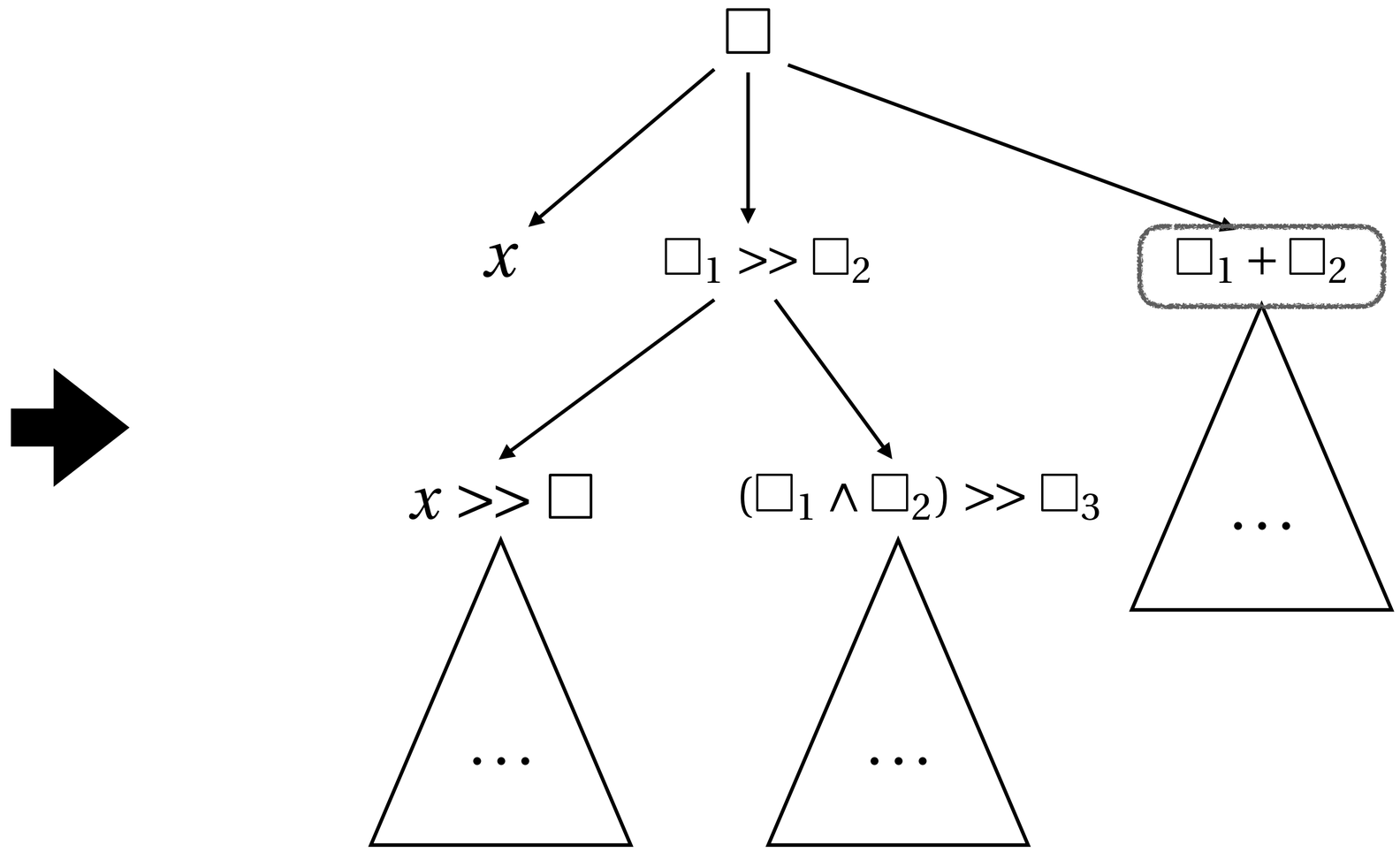
Pruning by Feasibility

Syntactic Spec

$$
\begin{aligned}
S \quad &\rightarrow \quad x \mid 0001_2 \\
&\mid \quad S \wedge S \mid S \vee S \mid S \oplus S \\
&\mid \quad S + S \mid S \times S \mid S / S \mid S >> S
\end{aligned}
$$

Semantic Spec

$f(1011_2) = 0011_2$



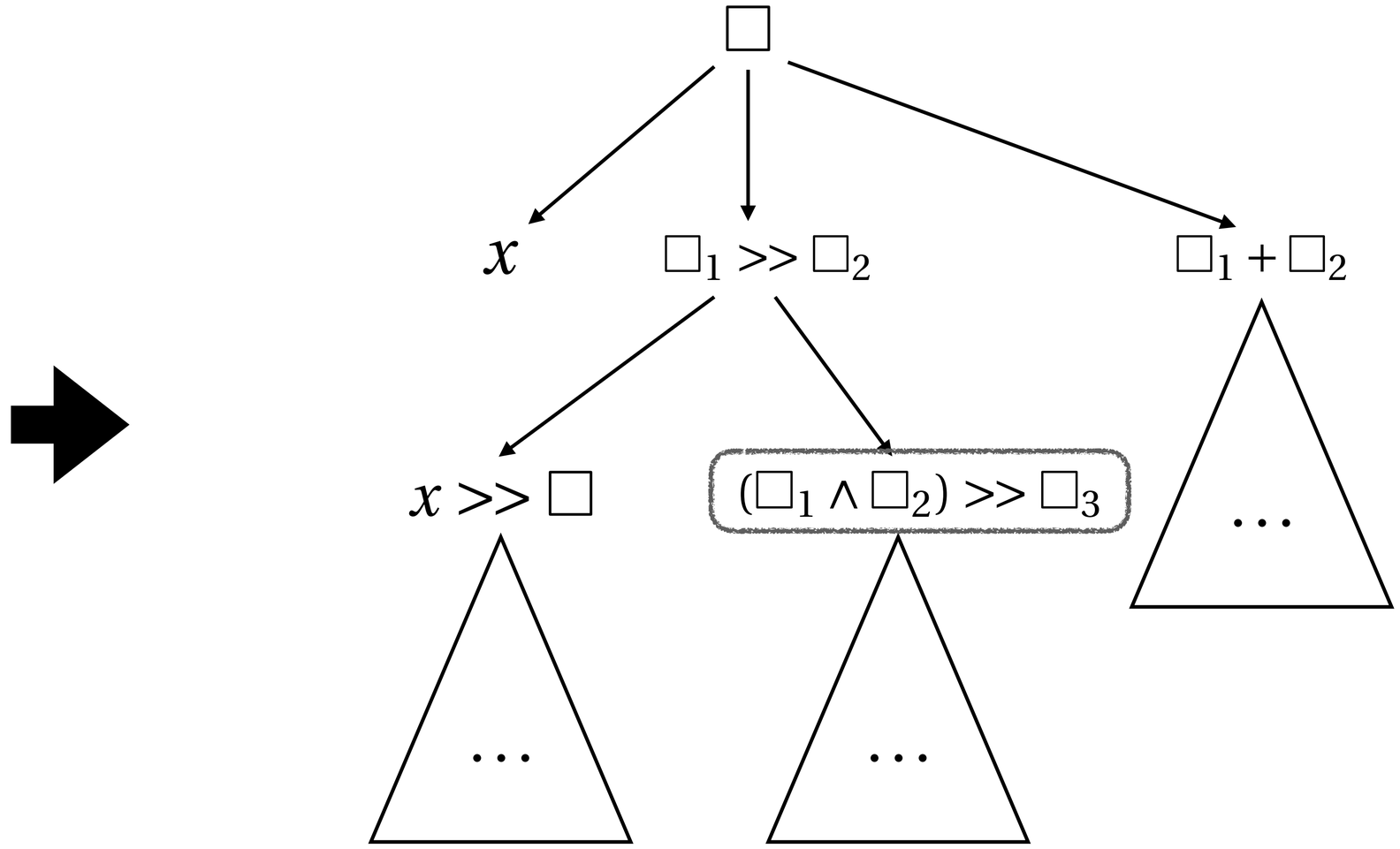
Pruning by Feasibility

Syntactic Spec

$$
\begin{aligned}
S \quad &\rightarrow \quad x \mid 0001_2 \\
&\mid \quad S \wedge S \mid S \vee S \mid S \oplus S \\
&\mid \quad S + S \mid S \times S \mid S / S \mid S >> S
\end{aligned}
$$

Semantic Spec

$f(1011_2) = 0011_2$



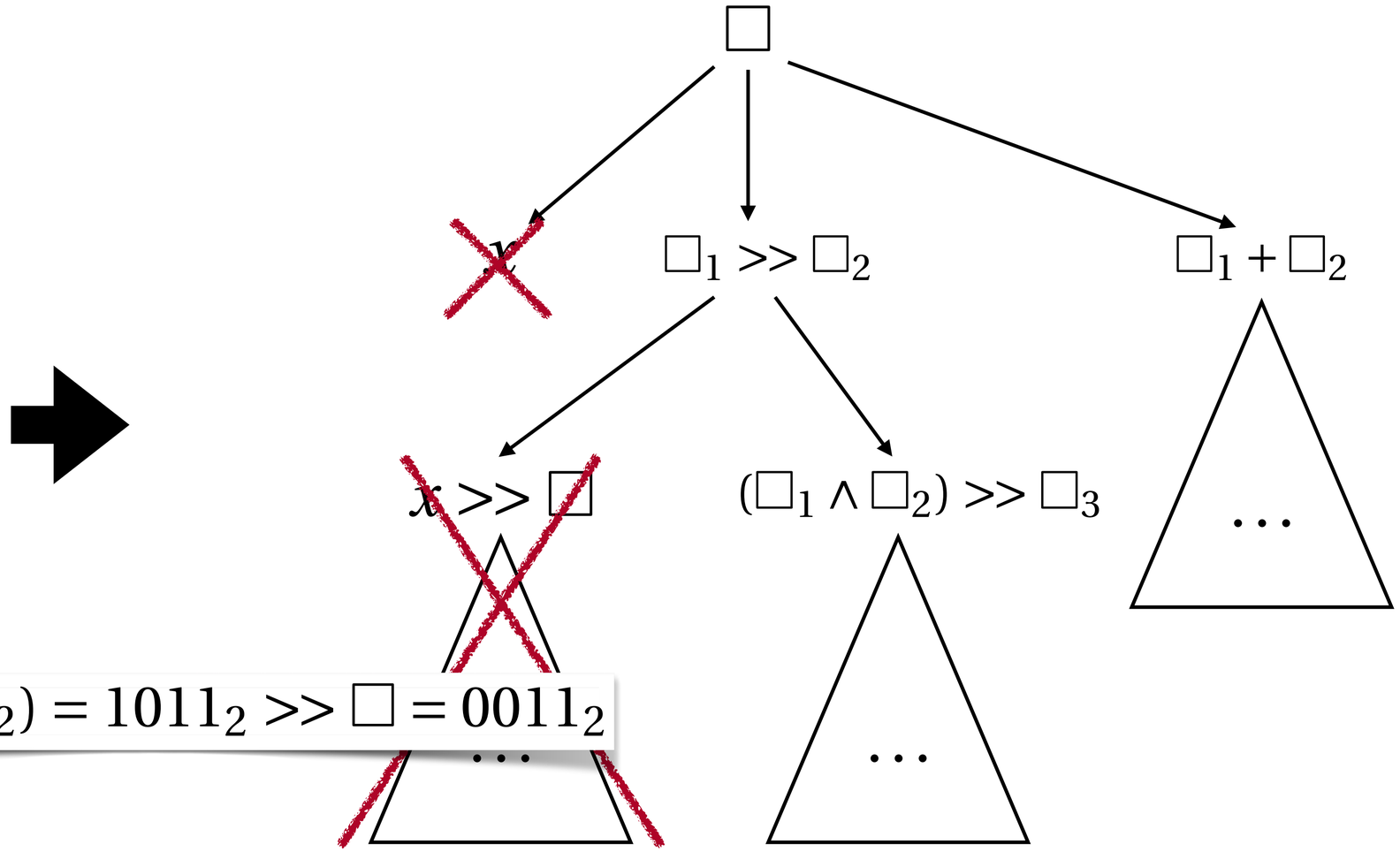
Pruning by Feasibility

Syntactic Spec

$$
\begin{array}{rcl}
S & \rightarrow & x \mid 0001_2 \\
 & \mid & S \wedge S \mid S \vee S \mid S \oplus S \\
 & \mid & S + S \mid S \times S \mid S / S \mid S >> S
\end{array}
$$

Semantic Spec

$f(1011_2) = 0011_2$



$$\nexists \square . f(1011_2) = 1011_2 >> \square = 0011_2$$

Pruning by Feasibility

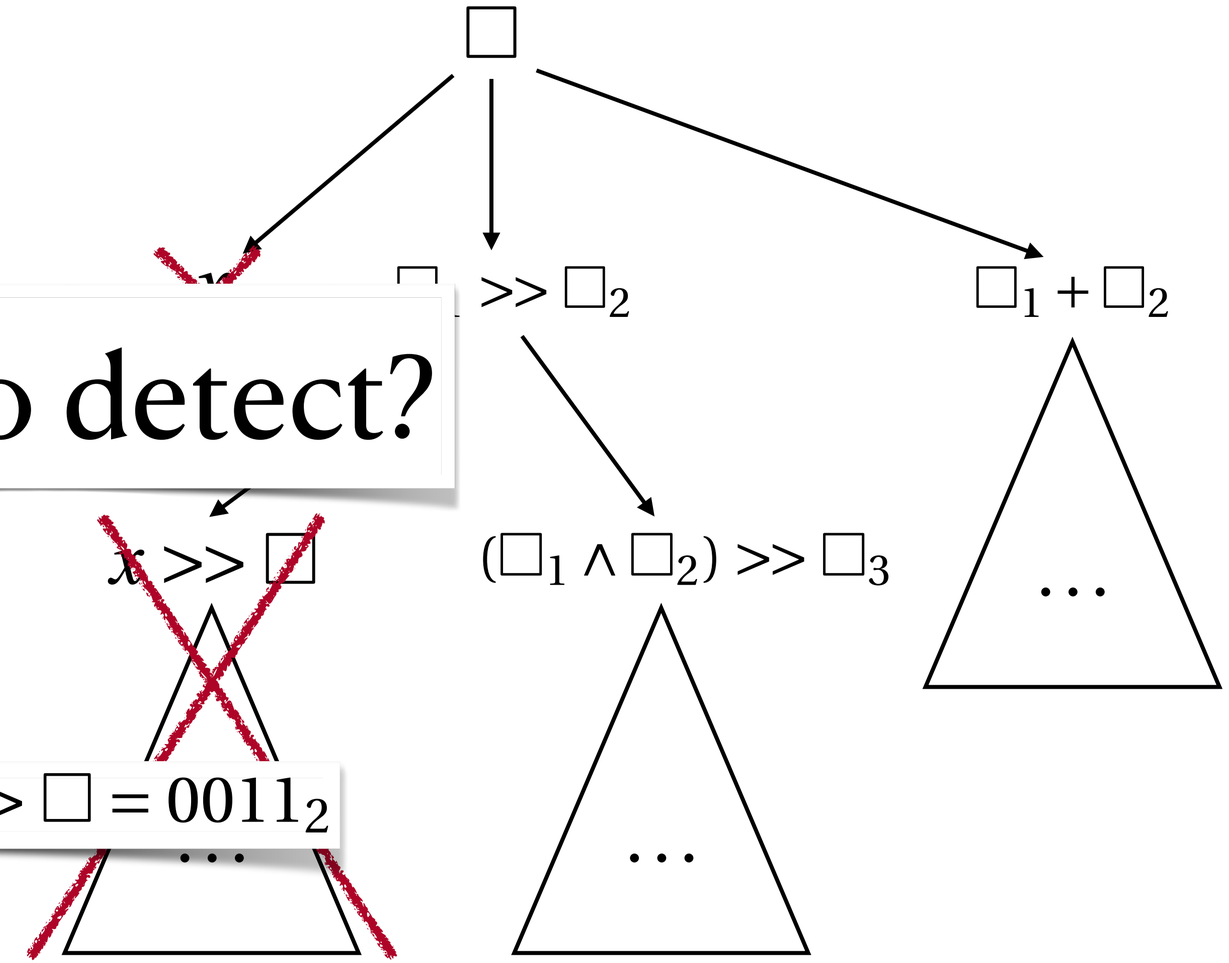
Syntactic Spec

$$
\begin{array}{rcl}
S & \rightarrow & x \mid 0001_2 \\
 & \mid & S \wedge S \mid S \vee S \mid S \oplus S \\
 & \mid & S + S \mid S \times S \mid S / S \mid S >> S
\end{array}
$$

Semantic Spec

$f(1011_2) = 0011_2$

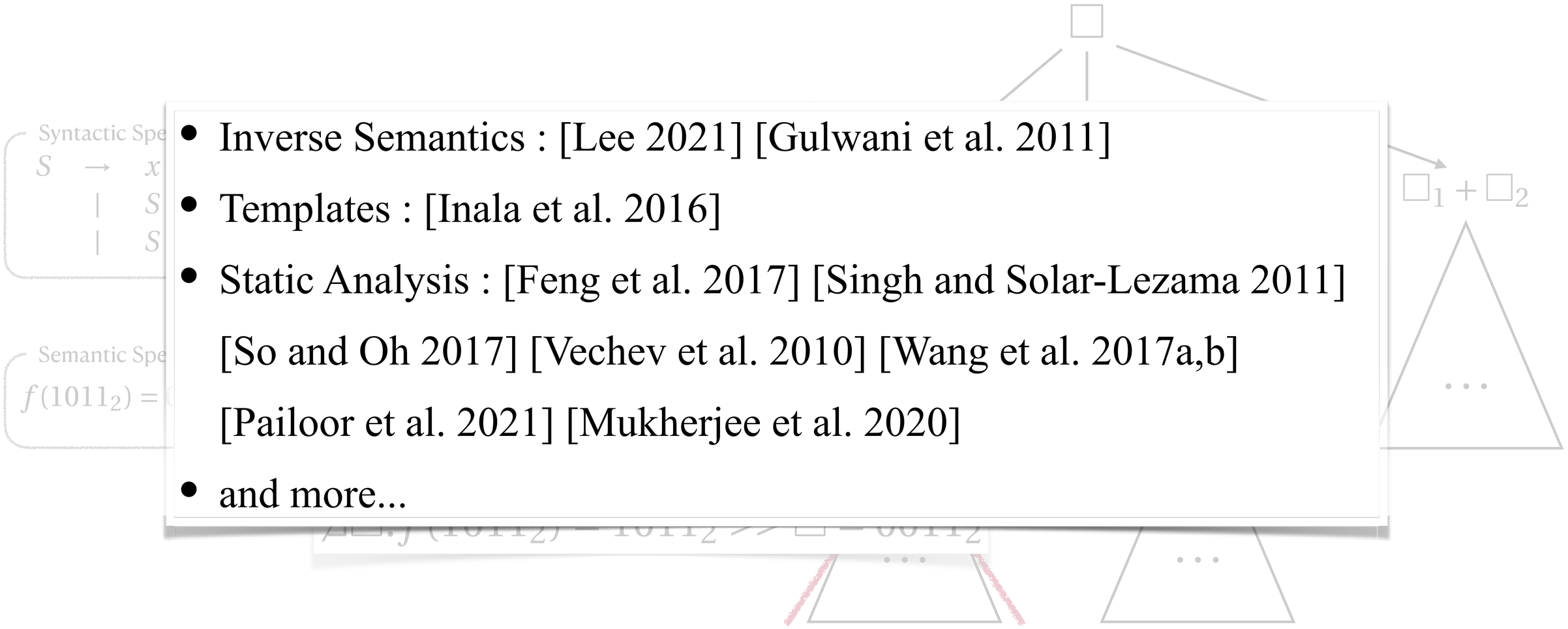
How to detect?



$\nexists \square . f(1011_2) = 1011_2 >> \square = 0011_2$

Pruning by Feasibility

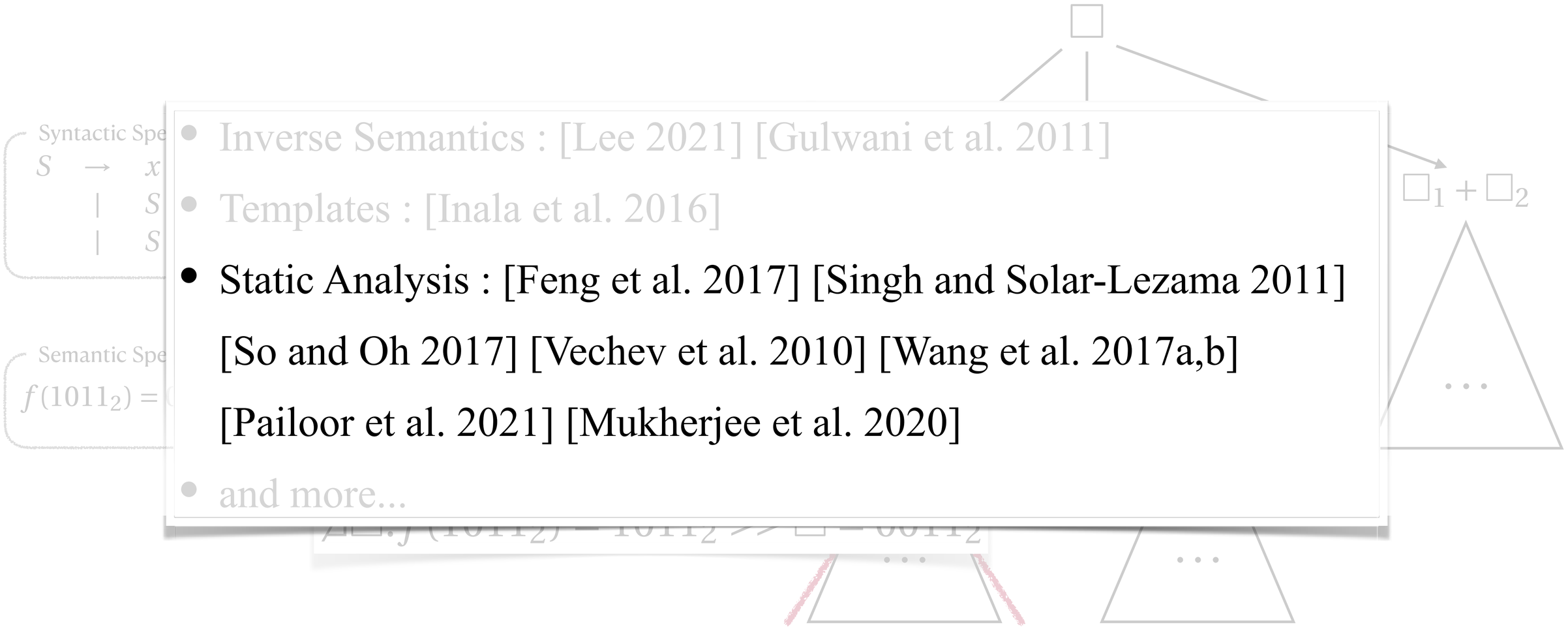
Existing Approaches



- Inverse Semantics : [Lee 2021] [Gulwani et al. 2011]
- Templates : [Inala et al. 2016]
- Static Analysis : [Feng et al. 2017] [Singh and Solar-Lezama 2011] [So and Oh 2017] [Vechev et al. 2010] [Wang et al. 2017a,b] [Pailoor et al. 2021] [Mukherjee et al. 2020]
- and more...

Pruning by Feasibility

Existing Approaches

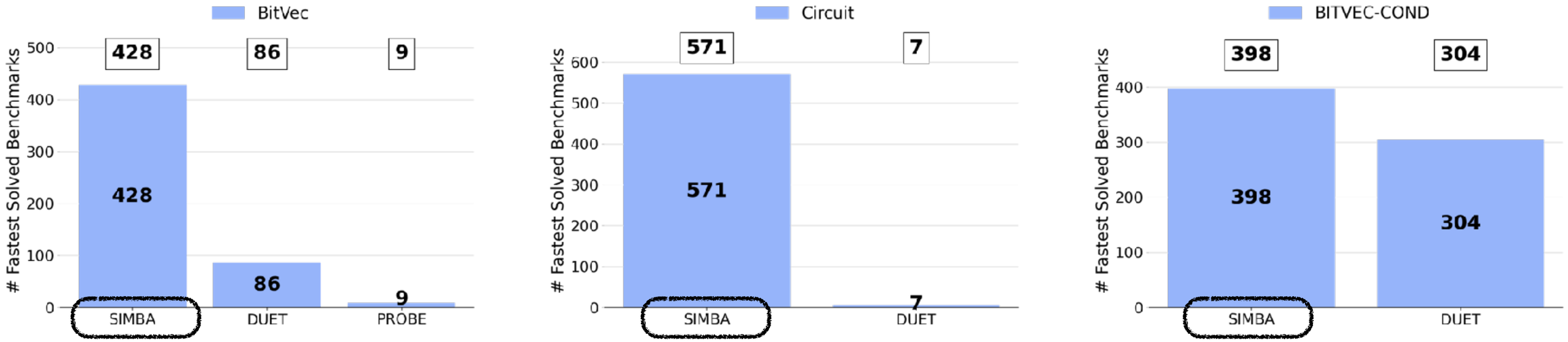


- Inverse Semantics : [Lee 2021] [Gulwani et al. 2011]
- Templates : [Inala et al. 2016]
- Static Analysis : [Feng et al. 2017] [Singh and Solar-Lezama 2011] [So and Oh 2017] [Vechev et al. 2010] [Wang et al. 2017a,b] [Pailoor et al. 2021] [Mukherjee et al. 2020]
- and more...

Our Performance

Comparing to 2 existing tools on 3 benchmark domains

- Our tool SIMBA using [Forward-Backward Static Analysis]
  - Outperformed state-of-the-art tools for 2 branch-free benchmarks
  - Comparable to state-of-the-art tools for a branch benchmark



of the fastest-solved problems for each domain

Using Forward Analysis

Checking only output feasibility

Semantic Spec

$f(1011_2) = 0011_2$

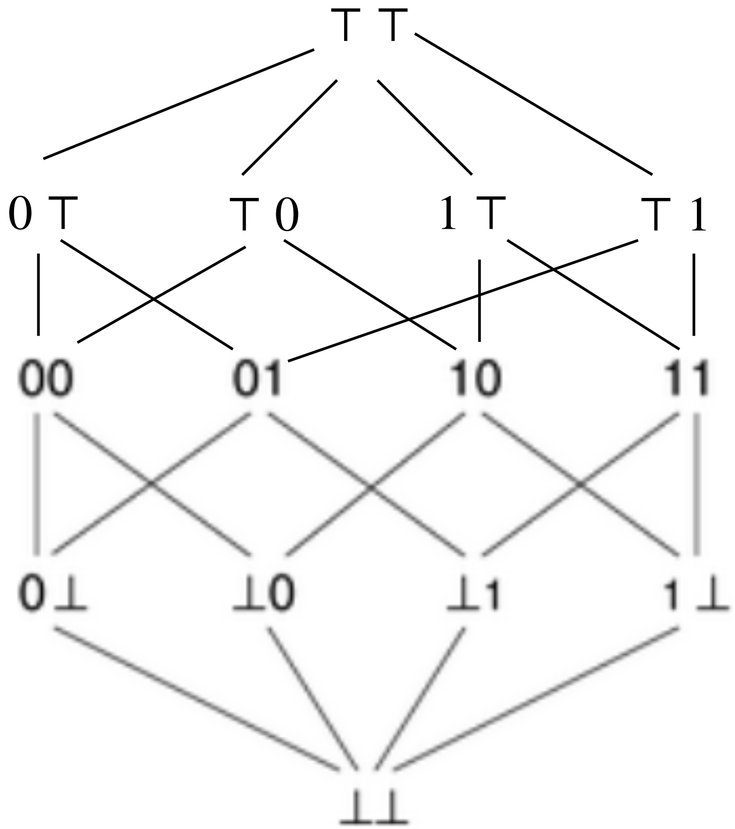
Candidate Partial Program

$f(x) = x \vee \square$

Forward

$x \mapsto 1011$

$\square \mapsto \top\top\top\top$



Using Forward Analysis

Checking only output feasibility

Semantic Spec

$f(1011_2) = 0011_2$

Candidate Partial Program

$f(x) = x \vee \square$

Forward

$$
\begin{array}{rcl}
x & \mapsto & 1011 \\
\square & \mapsto & \top\top\top\top \\
x \vee \square & \mapsto & 1\top 11 \not\sqsupseteq 0011
\end{array}
$$

Infeasible output

Limitation of Forward Analysis

Checking only output feasibility

Semantic Spec

$f(1011_2) = 0011_2$

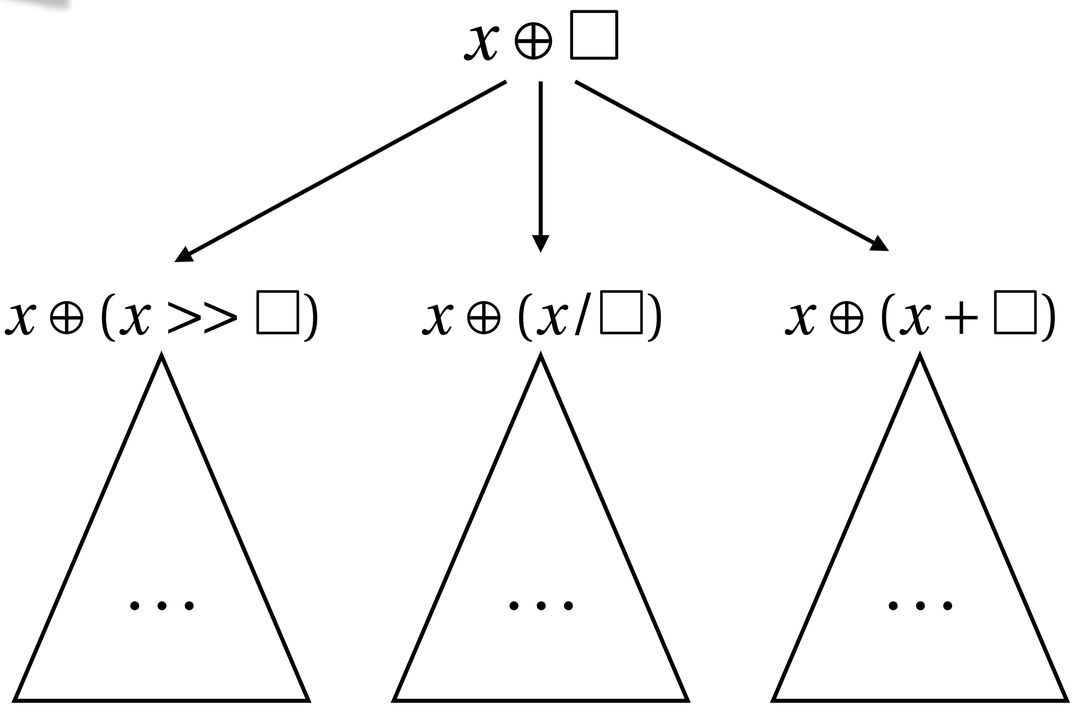
Candidate Partial Program

$f(x) = x \oplus \square$

Forward

$$
\begin{array}{rcl}
x & \mapsto & 1011 \\
\square & \mapsto & \top\top\top\top \\
x \oplus \square & \mapsto & \top\top\top\top \sqsupseteq 0011
\end{array}
$$

Nothing to do



Need: Backward Analysis Too

Output feasibility + Hole Precondition

Semantic Spec

$f(1011_2) = 0011_2$

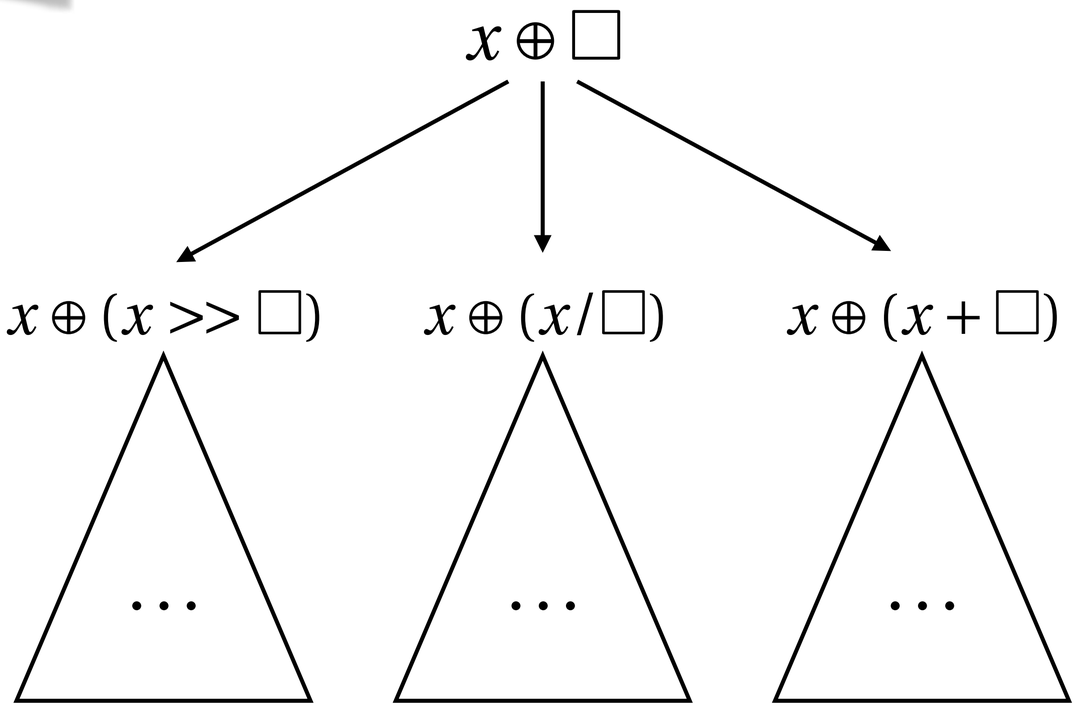
Candidate Partial Program

$f(x) = x \oplus \square$

Forward

$$x \mapsto 1011$$
$$\square \mapsto \top\top\top\top$$
$$x \oplus \square \mapsto 0011$$

Backward



Need: Backward Analysis Too

Output feasibility + Hole Precondition

Semantic Spec

$f(1011_2) = 0011_2$

Candidate Partial Program

$f(x) = x \oplus \square$

Forward

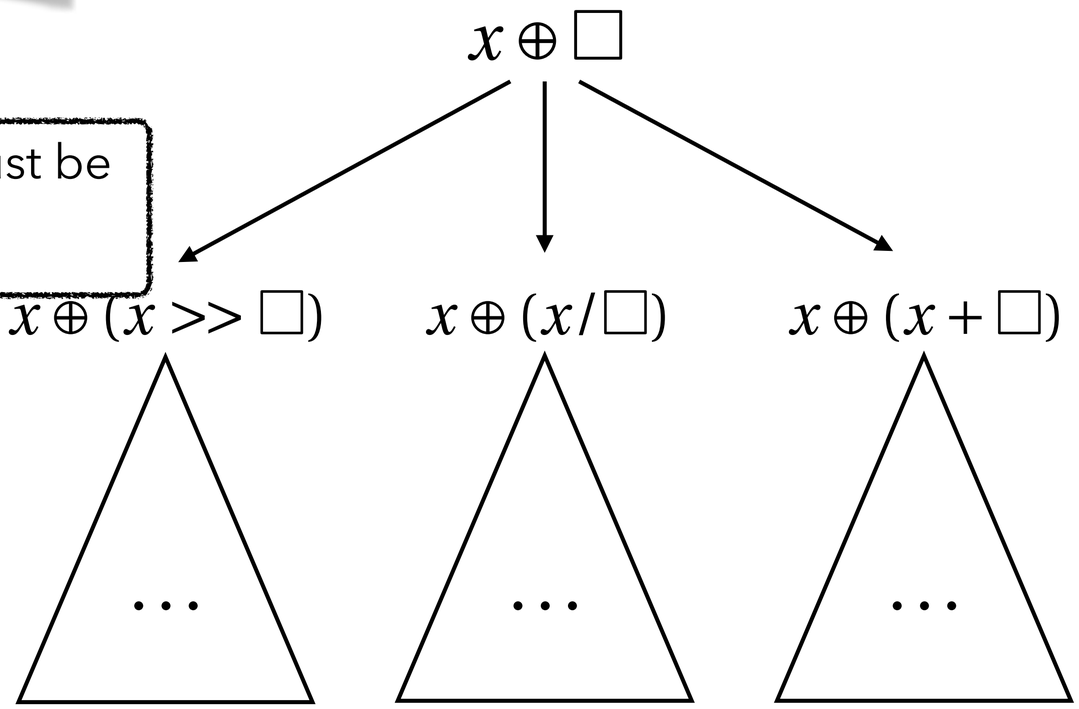
$x \mapsto 1011$

$\square \mapsto 1000$

$x \oplus \square \mapsto 0011$

value of hole must be 1000_2

Backward



Need: Backward Analysis Too

Output feasibility + Hole Precondition

Semantic Spec

$f(1011_2) = 0011_2$

Candidate Partial Program

$f(x) = x \oplus \square$

Forward

$x \mapsto 10$

$\square \mapsto 10$

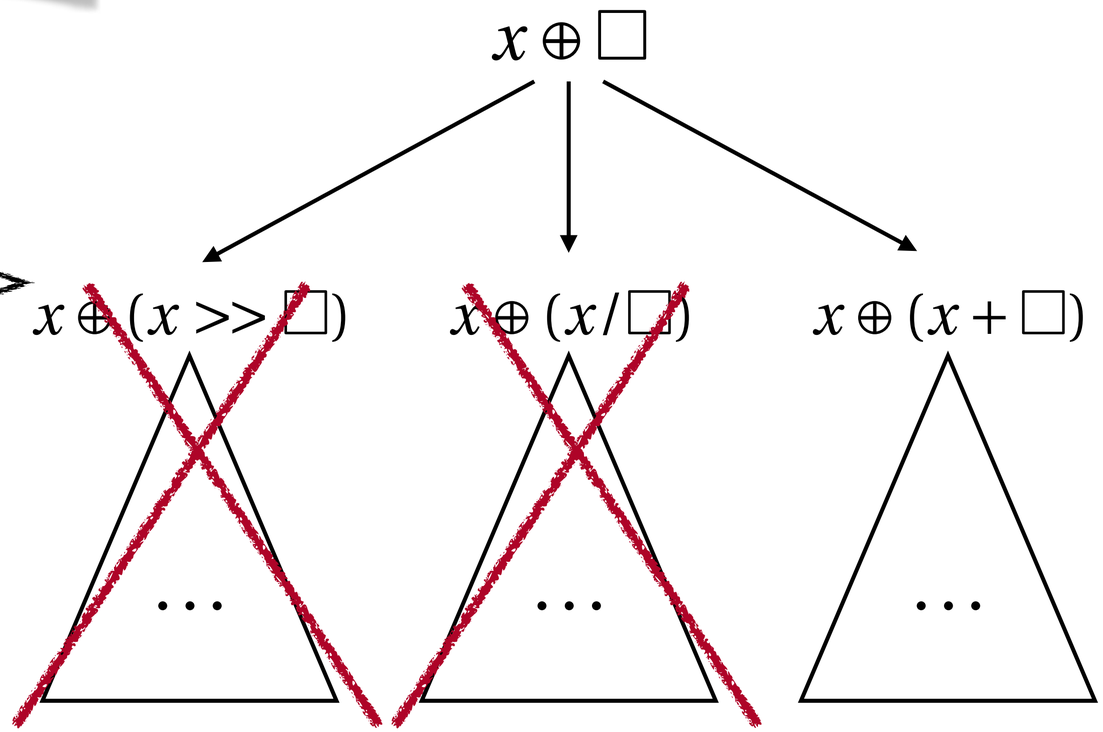
$x \oplus \square \mapsto 00$

prune infeasible parts

$x >> \square \neq 1000_2$

$x / \square$

Backward



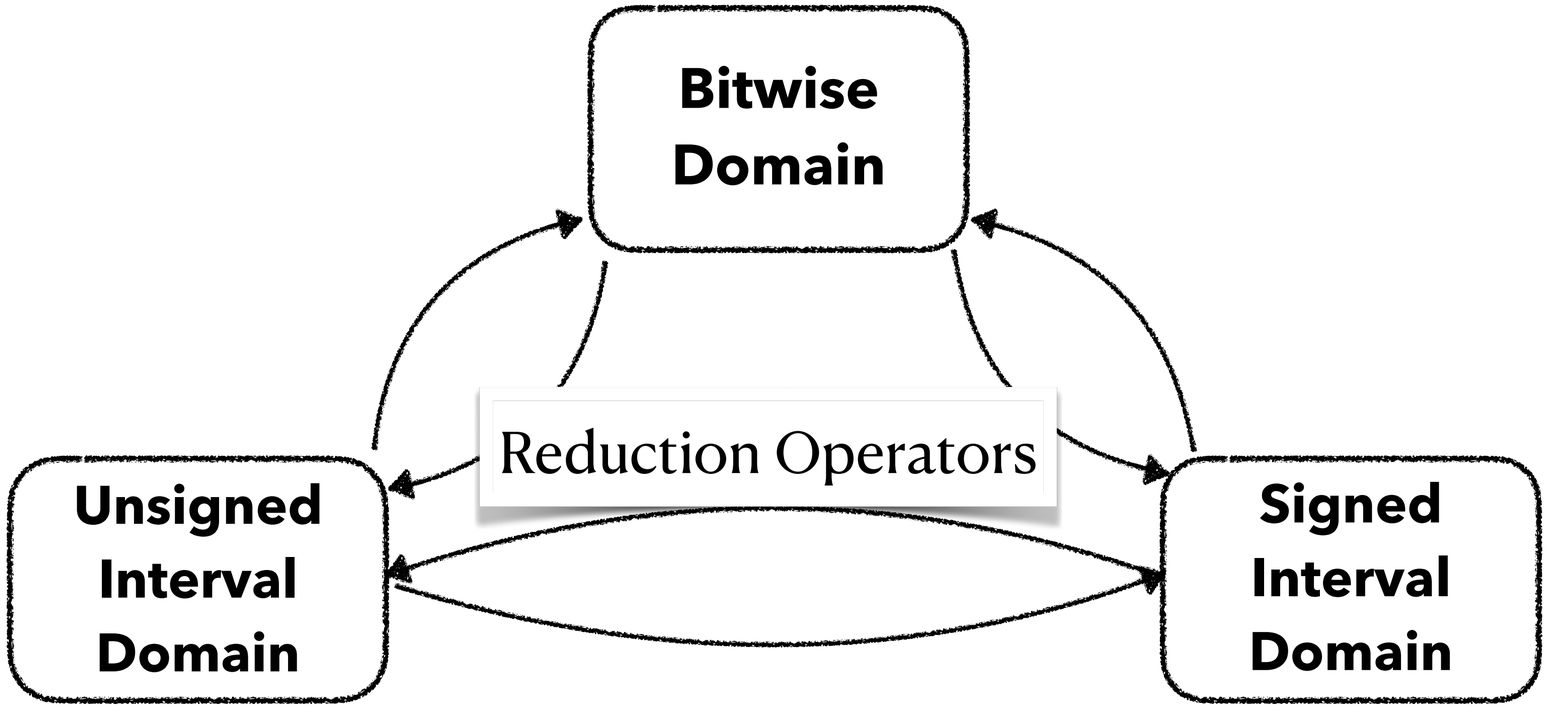
Challenges

Challenge 1: Precision & Cost Balance

Challenge 2: Precise Backward Analysis

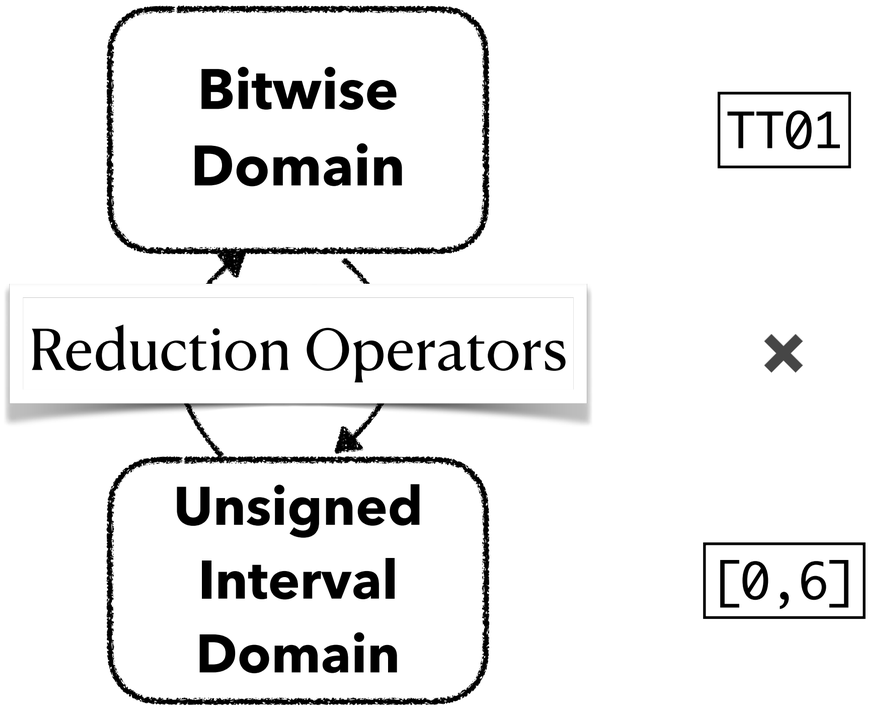
Challenge 1: Precision & Cost Balance

- Reduced Product Domain of well-known and simple domains
- No more complex domain (e.g., relational domain) was necessary



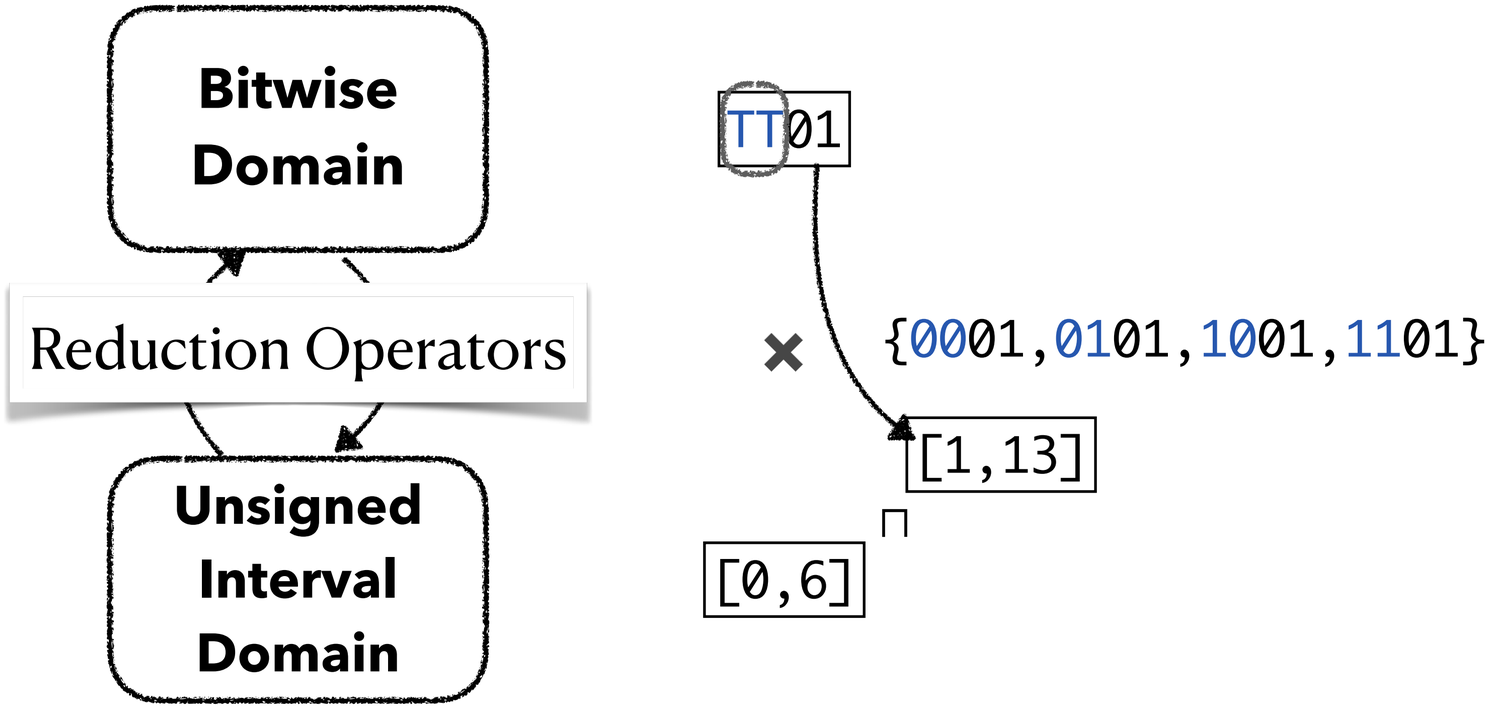
Challenge 1: Precision & Cost Balance

- Reduction Example: Bitwise ✖ Unsigned Interval



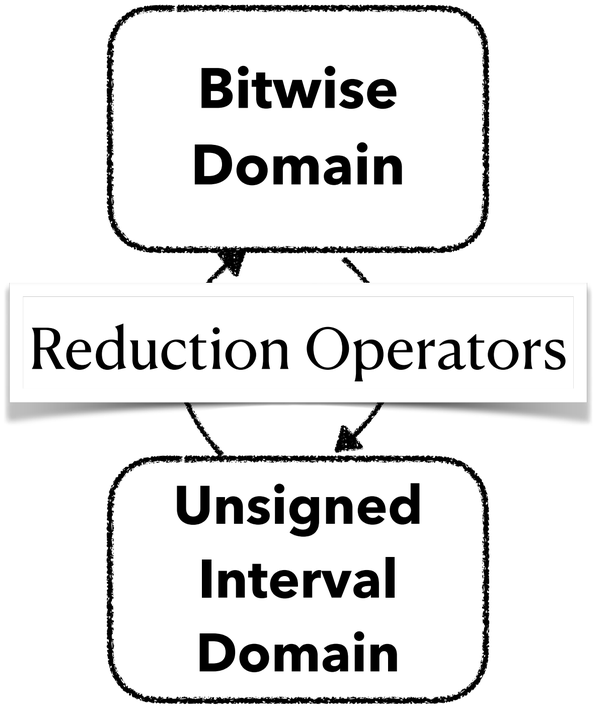
Challenge 1: Precision & Cost Balance

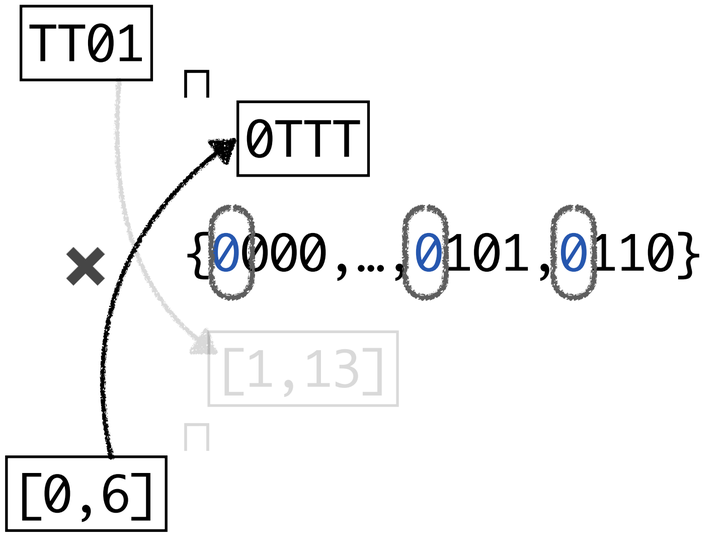
- Reduction Example: Bitwise ✖ Unsigned Interval



Challenge 1: Precision & Cost Balance

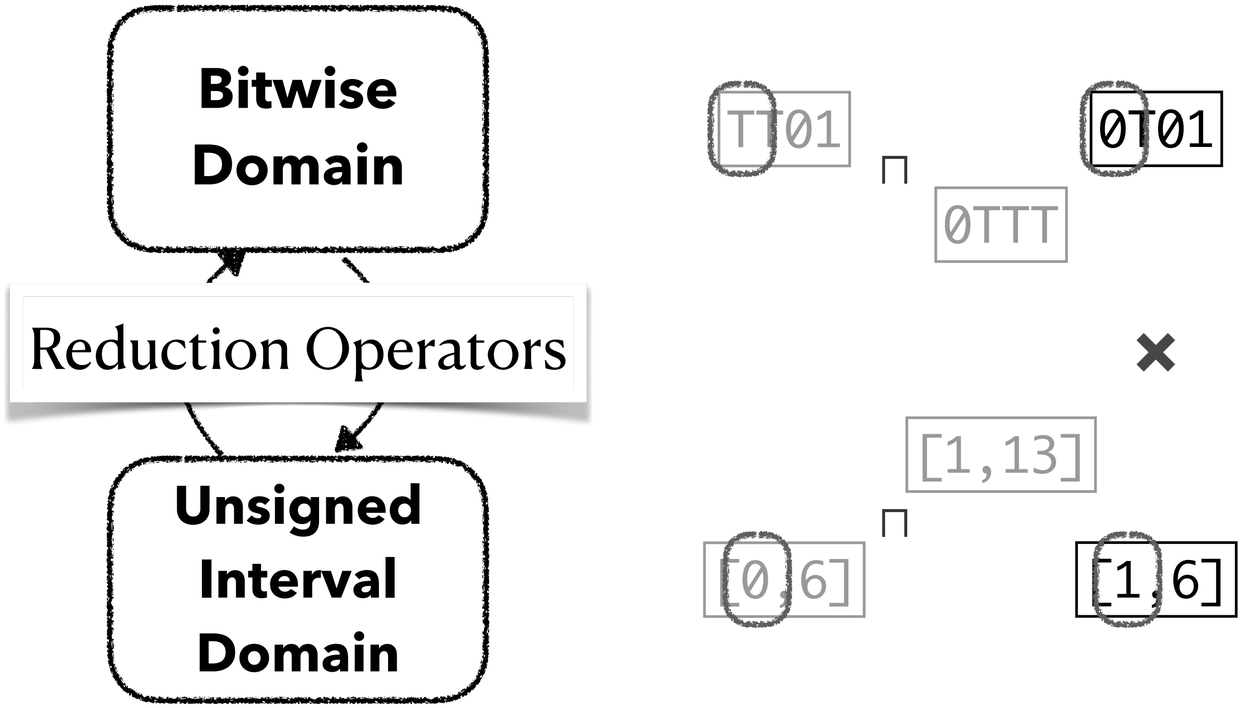
- Reduction Example: Bitwise ✖ Unsigned Interval





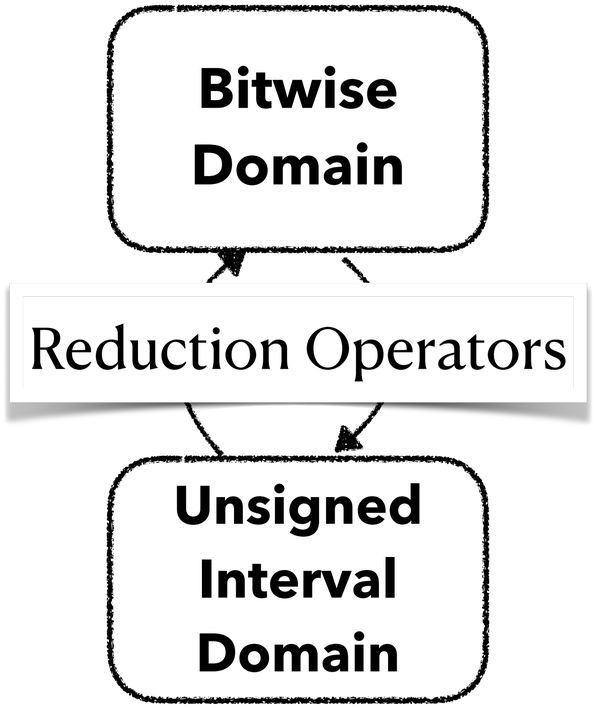
Challenge 1: Precision & Cost Balance

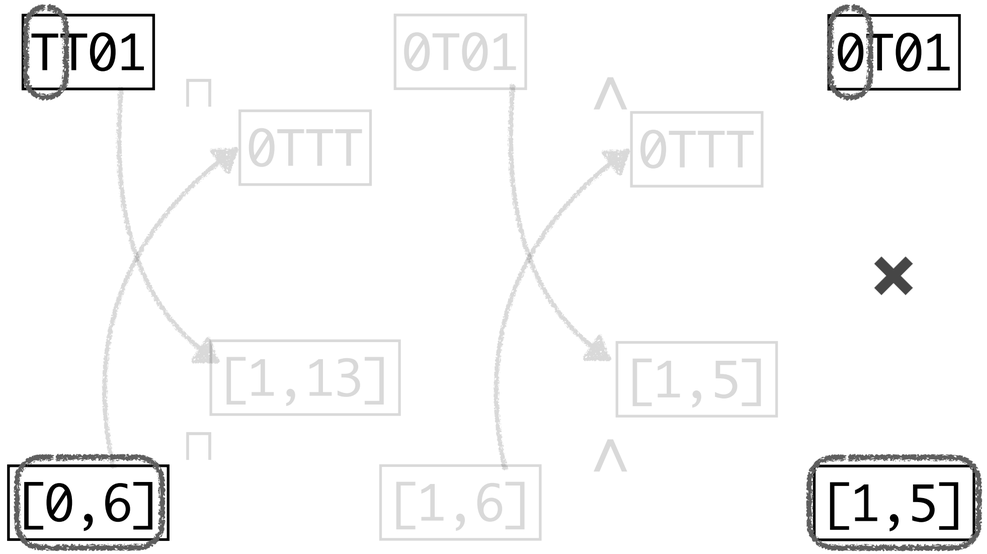
- Reduction Example: Bitwise ✖ Unsigned Interval



Challenge 1: Precision & Cost Balance

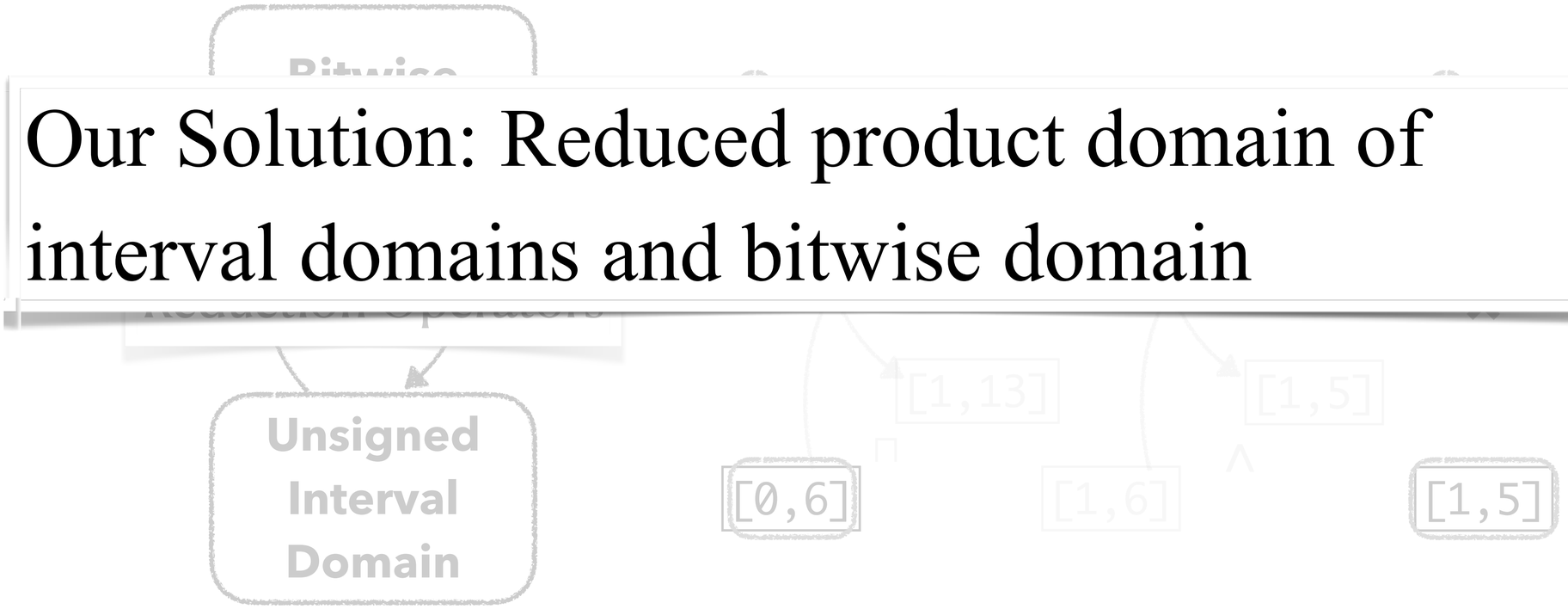
- Reduction Example: Bitwise ✖ Unsigned Interval





Challenge 1: Precision & Cost Balance

- Reduction Example: Bitwise ✖ Unsigned Interval



Our Solution: Reduced product domain of interval domains and bitwise domain

Challenge 2: Precise Backward Analysis by Selective Finite Concretization

Candidate Partial Program

$f(x) = x \wedge (x \times \square)$

Semantic Spec

$f(1011_2) = 0011_2$

Forward

$$
\begin{array}{rcll}
x & \mapsto & 1011 & \\
\square & \mapsto & \top\top\top\top & \\
x \times \square & \mapsto & \top\top\top\top & \\
x \wedge (x \times \square) & \mapsto & \top\top\top\top & \sqsupseteq 0011
\end{array}
$$

Challenge 2: Precise Backward Analysis by Selective Finite Concretization

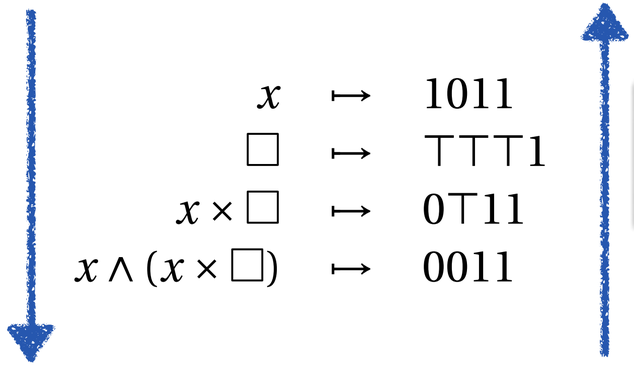
Candidate Partial Program

$f(x) = x \wedge (x \times \square)$

Semantic Spec

$f(1011_2) = 0011_2$

Forward



$$
\begin{array}{rcl}
x & \mapsto & 1011 \\
\square & \mapsto & \top\top\top 1 \\
x \times \square & \mapsto & 0\top 11 \\
x \wedge (x \times \square) & \mapsto & 0011
\end{array}
$$

Can we do better?

Backward

Challenge 2: Precise Backward Analysis by Selective Finite Concretization

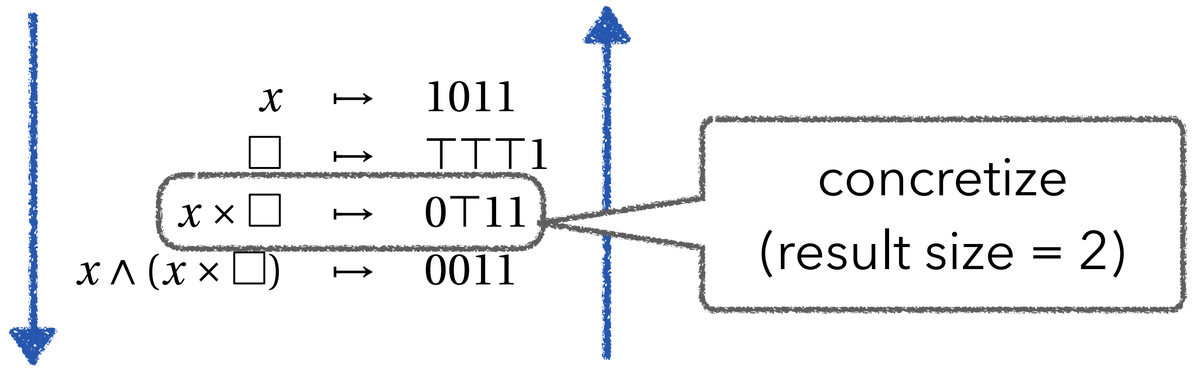
Candidate Partial Program

$f(x) = x \wedge (x \times \square)$

Semantic Spec

$f(1011_2) = 0011_2$

Forward



Backward

Challenge 2: Precise Backward Analysis by Selective Finite Concretization

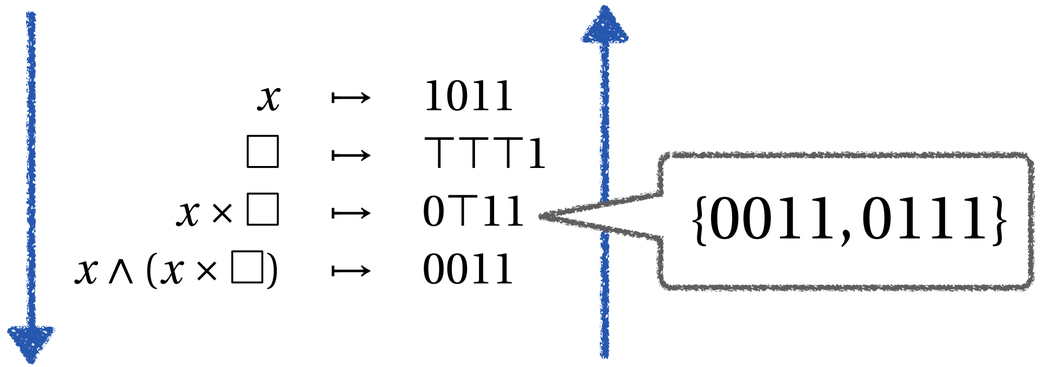
Candidate Partial Program

$f(x) = x \wedge (x \times \square)$

Semantic Spec

$f(1011_2) = 0011_2$

Forward



Backward

Challenge 2: Precise Backward Analysis by Selective Finite Concretization

Candidate Partial Program

$f(x) = x \wedge (x \times \square)$

Semantic Spec

$f(1011_2) = 0011_2$

Forward

$$
\begin{array}{rcl}
x & \mapsto & 1011 \\
\square & \mapsto & \top\top\top 1 \\
x \times \square & \mapsto & 0\top 11 \\
x \wedge (x \times \square) & \mapsto & 0011
\end{array}
$$

we get 2 equations

$$
\begin{array}{lcl}
1011 \times \square_0 & = & 0011 \\
1011 \times \square_1 & = & 0111
\end{array}
$$

Backward

Challenge 2: Precise Backward Analysis by Selective Finite Concretization

Candidate Partial Program

$f(x) = x \wedge (x \times \square)$

Semantic Spec

$f(1011_2) = 0011_2$

Forward

$$
\begin{array}{rcl}
x & \mapsto & 1011 \\
\square & \mapsto & \top\top\top 1 \\
x \times \square & \mapsto & 0\top 11 \\
x \wedge (x \times \square) & \mapsto & 0011
\end{array}
$$

Backward

we get 2 equations

$$
\begin{array}{lcl}
1011 \times \square_0 & = & 0011 \mod 16 \\
1011 \times \square_1 & = & 0111 \mod 16
\end{array}
$$

Incorrect approach: transposition

$$
\begin{array}{lcl}
1011 \times \square_0 & = & 0011/1011 \\
1011 \times \square_1 & = & 0111/1011
\end{array}
$$

Challenge 2: Precise Backward Analysis by Selective Finite Concretization

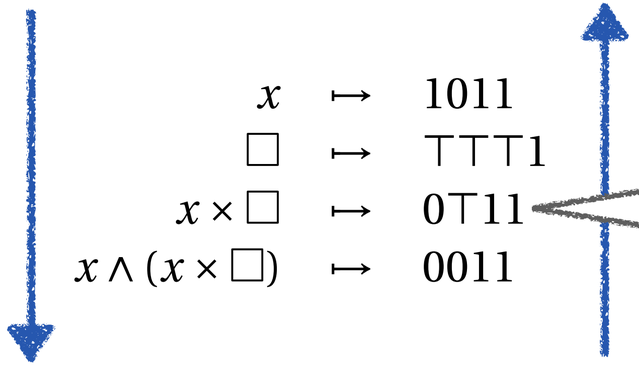
Candidate Partial Program

$f(x) = x \wedge (x \times \square)$

Semantic Spec

$f(1011_2) = 0011_2$

Forward



$$
\begin{array}{rcl}
x & \mapsto & 1011 \\
\square & \mapsto & \top\top\top 1 \\
x \times \square & \mapsto & 0\top 11 \\
x \wedge (x \times \square) & \mapsto & 0011
\end{array}
$$

Backward

we get 2 equations

$$
\begin{array}{lcl}
1011 \times \square_0 & = & 0011 \mod 16 \\
1011 \times \square_1 & = & 0111 \mod 16
\end{array}
$$

by extended Euclidean algorithm,

$$
\begin{array}{lcl}
\square_0 & = & 1001 \\
\square_1 & = & 0101
\end{array}
$$

Challenge 2: Precise Backward Analysis by Selective Finite Concretization

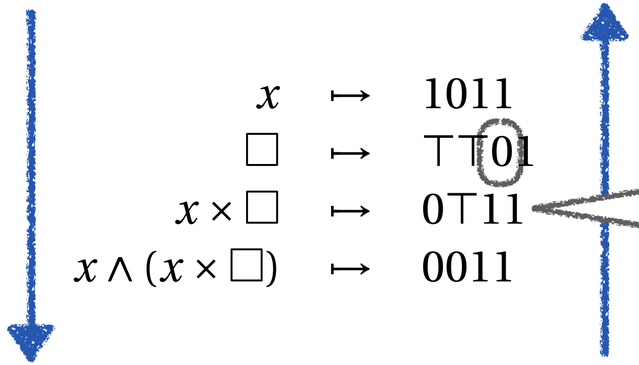
Candidate Partial Program

$f(x) = x \wedge (x \times \square)$

Semantic Spec

$f(1011_2) = 0011_2$

Forward



$$
\begin{array}{rcl}
x & \mapsto & 1011 \\
\square & \mapsto & \top\top 01 \\
x \times \square & \mapsto & 0\top 11 \\
x \wedge (x \times \square) & \mapsto & 0011
\end{array}
$$

Backward

we get 2 equations

$$
\begin{array}{lcll}
1011 \times \square_0 & = & 0011 & \bmod 16 \\
1011 \times \square_1 & = & 0111 & \bmod 16
\end{array}
$$

by extended Euclidean algorithm,

$$
\begin{array}{rcl}
\square_0 & = & 1001 \\
\square_1 & = & 0101 \\
\square & = & \square_0 \sqcup \square_1 = \top\top 01
\end{array}
$$

Challenge 2: Precise Backward Analysis by Selective Finite Concretization

Candidate Partial Program

$f(x) = x \wedge (x \times \square)$

Semantic Spec

$f(1011_2) = 0011_2$

Forward

Backward

$x \wedge (x \times \square) \quad \mapsto \quad 0011$

$1011 \times \square_0 \quad = \quad 0011 \mod 16$

$1011 \times \square_1 \quad = \quad 0111 \mod 16$

$\square_0 \quad = \quad 1001$

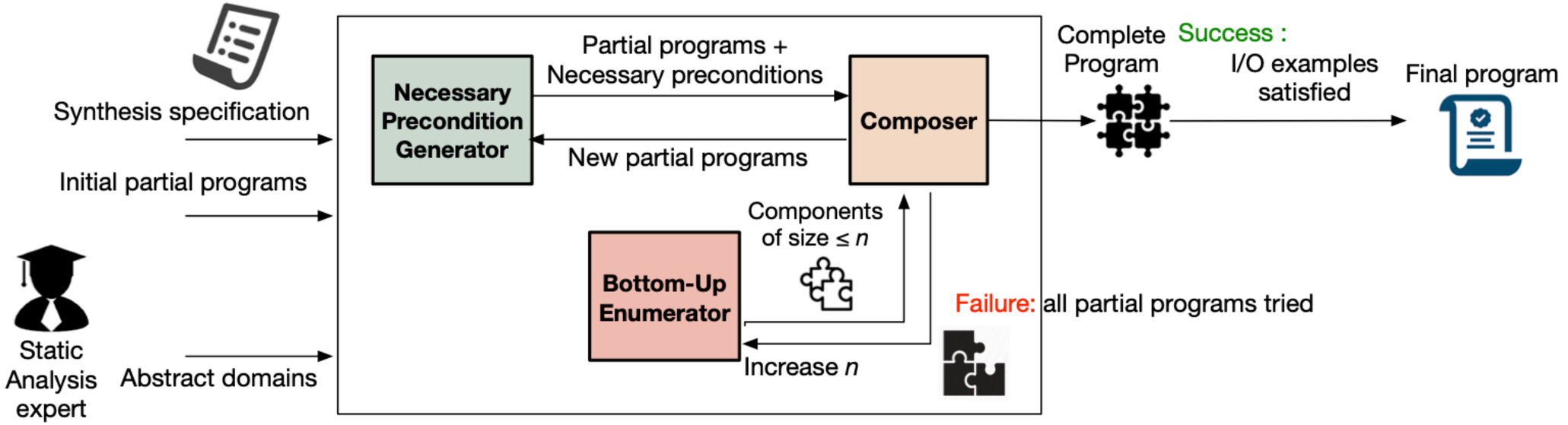
$\square_1 \quad = \quad 0101$

$\square \quad = \quad \square_0 \sqcup \square_1 = \top\top 01$

Our Solution: Use concretization for precise analysis

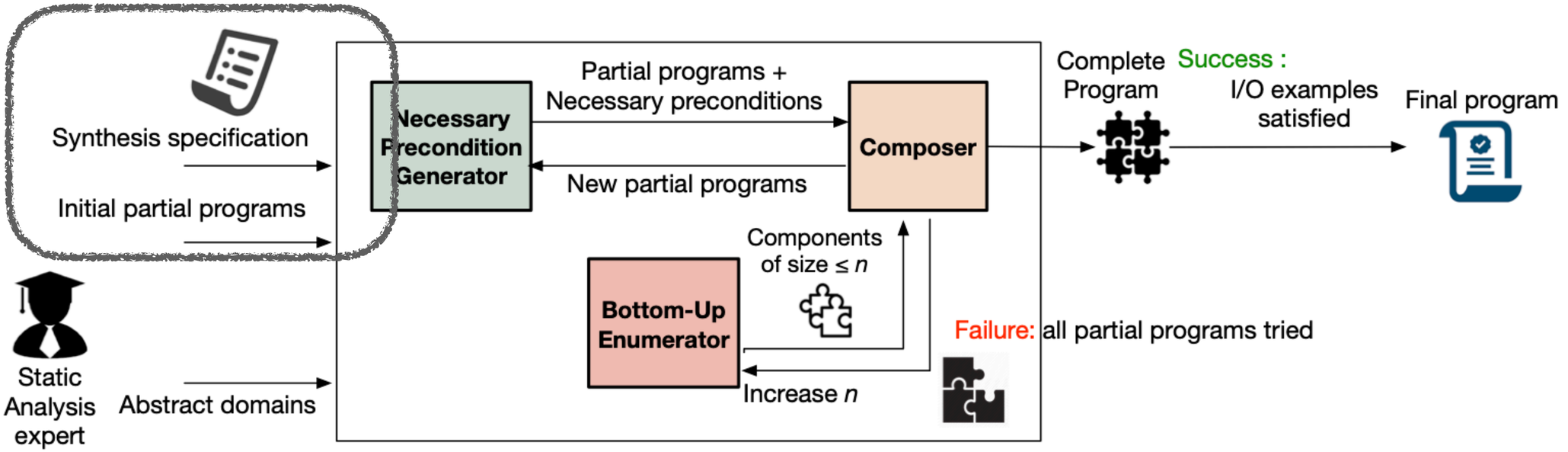
Realized: SIMBA

- **S**ynthesis from **I**nductive specification e**M**powered by **B**idirectional **A**bstract interpretation
- Available at https://github.com/yhyoon/simba



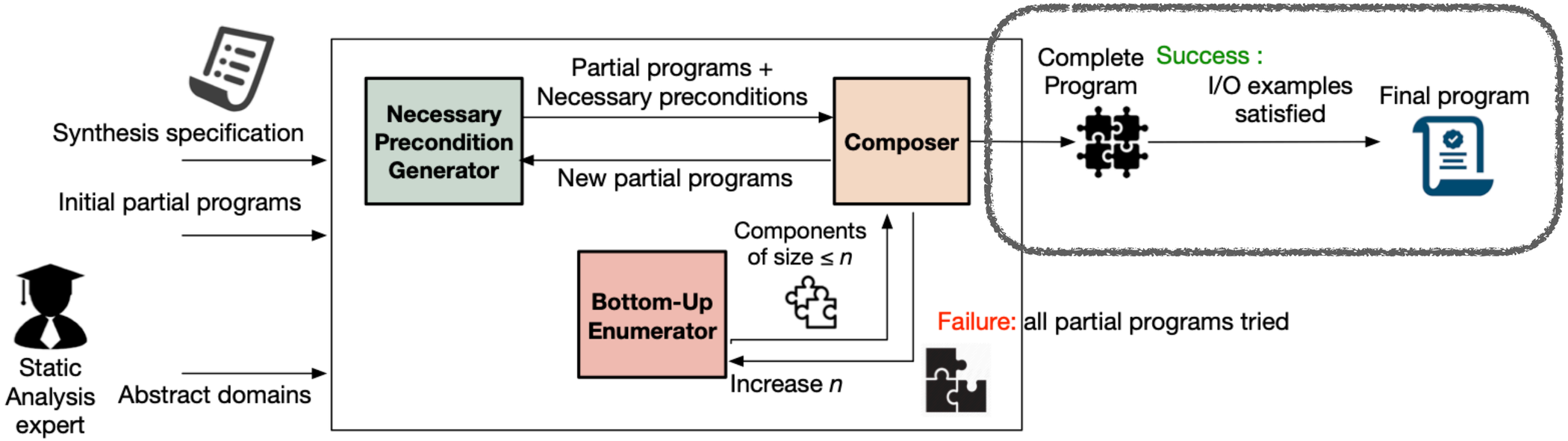
Realized: SIMBA

- **S**ynthesis from **I**nductive specification e**M**powered by **B**idirectional **A**bstract interpretation
- Available at https://github.com/yhyoon/simba

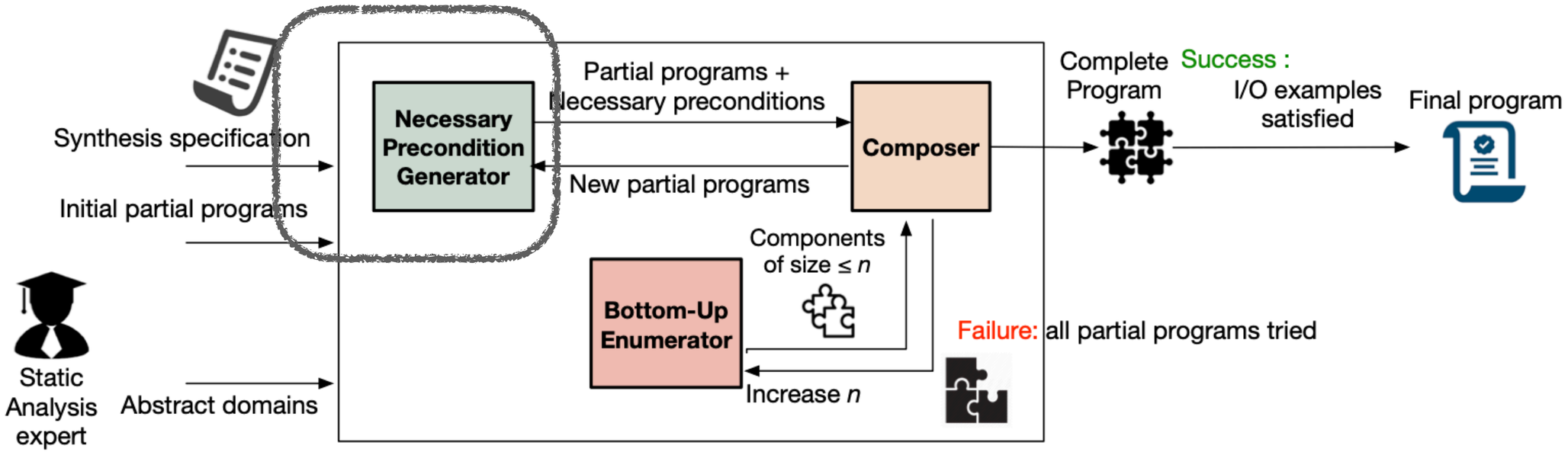


Realized: SIMBA

- **S**ynthesis from **I**nductive specification e**M**powered by **B**idirectional **A**bstract interpretation
- Available at https://github.com/yhyoon/simba



Synthesis Algorithm



Synthesis Algorithm

Example Initial Partial Programs

$(\square \oplus x) >> 0001_2$

$(\square / x) >> 0001_2$

Initial partial programs
chosen for illustration purpose

Spec

$$
\begin{array}{lll}
S & \rightarrow & x \mid 0001_2 \\
 & \mid & S \wedge S \mid S \vee S \mid S \oplus S \\
 & \mid & S + S \mid S \times S \mid S / S \mid S >> S
\end{array}
$$

$f(1011_2) = 0011_2$

Goal

$f(x) = ((x + 0001_2) \oplus x) >> 0001_2$

Synthesis Algorithm

Analyze Partial Programs (1/2)

$(\square \oplus x) >> 0001_2$

Forward Analysis

$\square$	$\mapsto$	$\top\top\top\top$
x	$\mapsto$	1011
$\square \oplus x$	$\mapsto$	$\top\top\top\top$
$(\square \oplus x) >> 0001_2$	$\mapsto$	$\top\top\top\top$

$(\square / x) >> 0001_2$

Spec

$$
\begin{array}{rcl}
S & \rightarrow & x \mid 0001_2 \\
 & \mid & S \wedge S \mid S \vee S \mid S \oplus S \\
 & \mid & S + S \mid S \times S \mid S / S \mid S >> S
\end{array}
$$

$f(1011_2) = 0011_2$

Goal

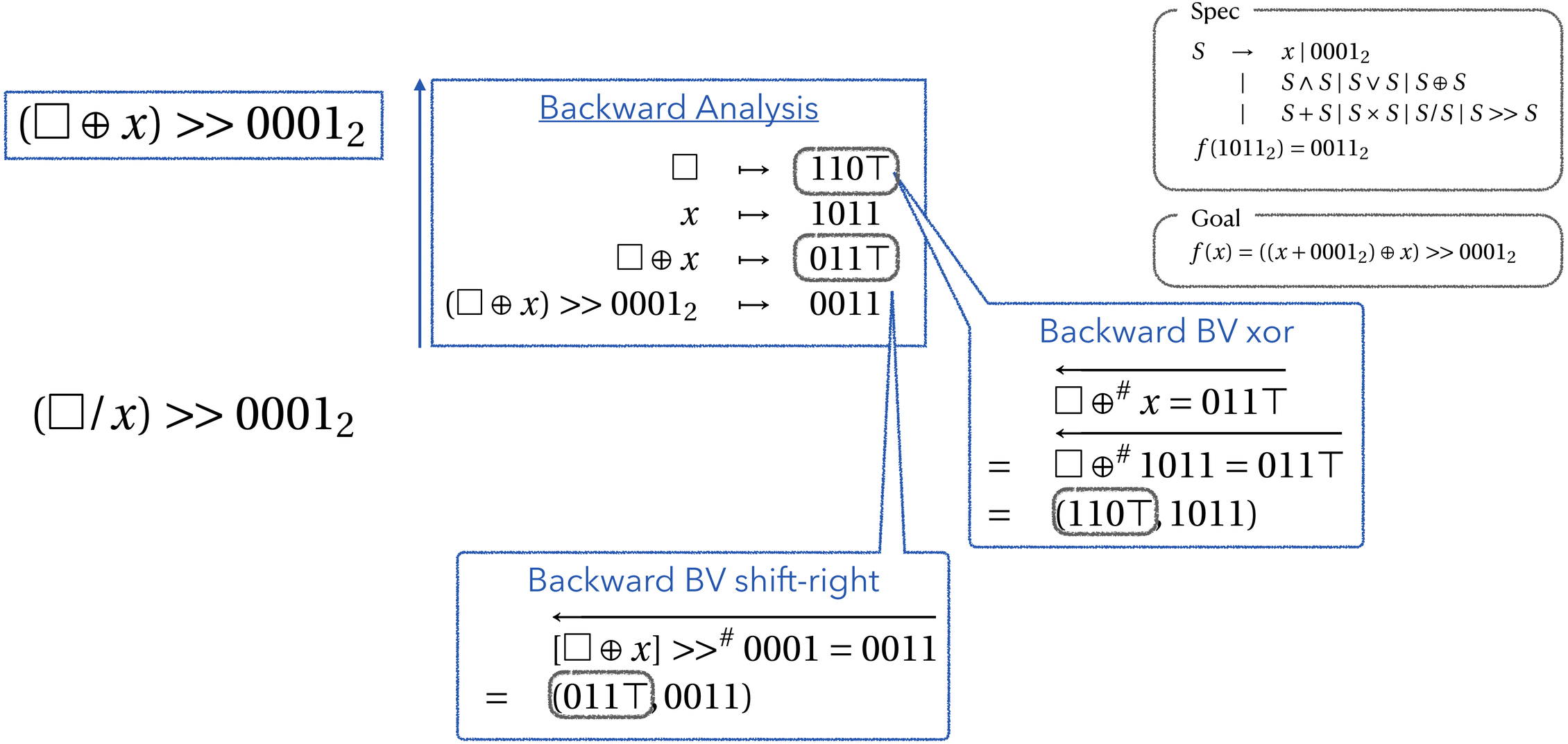
$f(x) = ((x + 0001_2) \oplus x) >> 0001_2$

Synthesis Algorithm

Analyze Partial Programs (1/2)

$(\square \oplus x) >> 0001_2$

$(\square / x) >> 0001_2$



Synthesis Algorithm

Analyze Partial Programs (2/2)

$$(\square \oplus x) >> 0001_2 \rightarrow$$

$$\begin{array}{rcl} \square & \mapsto & 110\top \\ x & \mapsto & 1011 \\ \square \oplus x & \mapsto & 011\top \\ (\square \oplus x) >> 0001_2 & \mapsto & 0011 \end{array}$$

Spec

$$\begin{array}{rcl} S & \rightarrow & x \mid 0001_2 \\ & \mid & S \wedge S \mid S \vee S \mid S \oplus S \\ & \mid & S + S \mid S \times S \mid S / S \mid S >> S \end{array}$$

$$f(1011_2) = 0011_2$$

Goal

$$f(x) = ((x + 0001_2) \oplus x) >> 0001_2$$

$$(\square / x) >> 0001_2$$

Forward Analysis

$$\begin{array}{rcl} \square & \mapsto & \top\top\top\top \\ x & \mapsto & 1011 \\ \square / x & \mapsto & 000\top \end{array}$$

$$\begin{array}{rcl} \square & = & [0, 15] \\ x & = & [11, 11] \\ \xrightarrow{\quad} & & \\ \square /^{\#} x & = & [l_S \,/\, h_x,\ h_S \,/\, l_x] \\ & = & [0 \,/\, 11,\ 15 \,/\, 11] \\ & = & [0, 1] \end{array}$$

Synthesis Algorithm

Analyze Partial Programs (2/2)

$(\square \oplus x) >> 0001_2 \rightarrow$

$\square$	$\mapsto$	$110\top$
x	$\mapsto$	1011
$\square \oplus x$	$\mapsto$	$011\top$
$(\square \oplus x) >> 0001_2$	$\mapsto$	0011

Spec

$$
\begin{array}{rcl}
S & \rightarrow & x \mid 0001_2 \\
 & \mid & S \wedge S \mid S \vee S \mid S \oplus S \\
 & \mid & S + S \mid S \times S \mid S / S \mid S >> S
\end{array}
$$

$f(1011_2) = 0011_2$

Goal

$f(x) = ((x + 0001_2) \oplus x) >> 0001_2$

$(\square / x) >> 0001_2$

Forward Analysis

$\square$	$\mapsto$	$\top\top\top\top$
x	$\mapsto$	1011
$\square / x$	$\mapsto$	$000\top$
$(\square / x) >> 0001_2$	$\mapsto$	$0000 \not\equiv 0011$

Synthesis Algorithm

Analyze Partial Programs (2/2)

$(\square \oplus x) >> 0001_2 \rightarrow$

$\square$	$\mapsto$	$110\top$
x	$\mapsto$	1011
$\square \oplus x$	$\mapsto$	$011\top$
$(\square \oplus x) >> 0001_2$	$\mapsto$	0011

Spec

$$
\begin{array}{rcl}
S & \rightarrow & x \mid 0001_2 \\
 & \mid & S \wedge S \mid S \vee S \mid S \oplus S \\
 & \mid & S + S \mid S \times S \mid S / S \mid S >> S
\end{array}
$$

$f(1011_2) = 0011_2$

Goal

$f(x) = ((x + 0001_2) \oplus x) >> 0001_2$

Discard

$(\square / x) >> 0001_2$

$\square$	$\mapsto$	$\top\top\top\top$
x	$\mapsto$	1011
$\square / x$	$\mapsto$	$000\top$
$(\square / x) >> 0001_2$	$\mapsto$	0000

Synthesis Algorithm

Generate Components (n <= 1)

Partial Programs

$(\square \oplus x) >> 0001_2 \rightarrow$

$$\begin{aligned} \square &\mapsto 101\top \\ x &\mapsto 1011 \\ \square \oplus x &\mapsto 011\top \\ (\square \oplus x) >> 0001_2 &\mapsto 0011 \end{aligned}$$

Components

$x \ (= 1011_2)$

0001_2

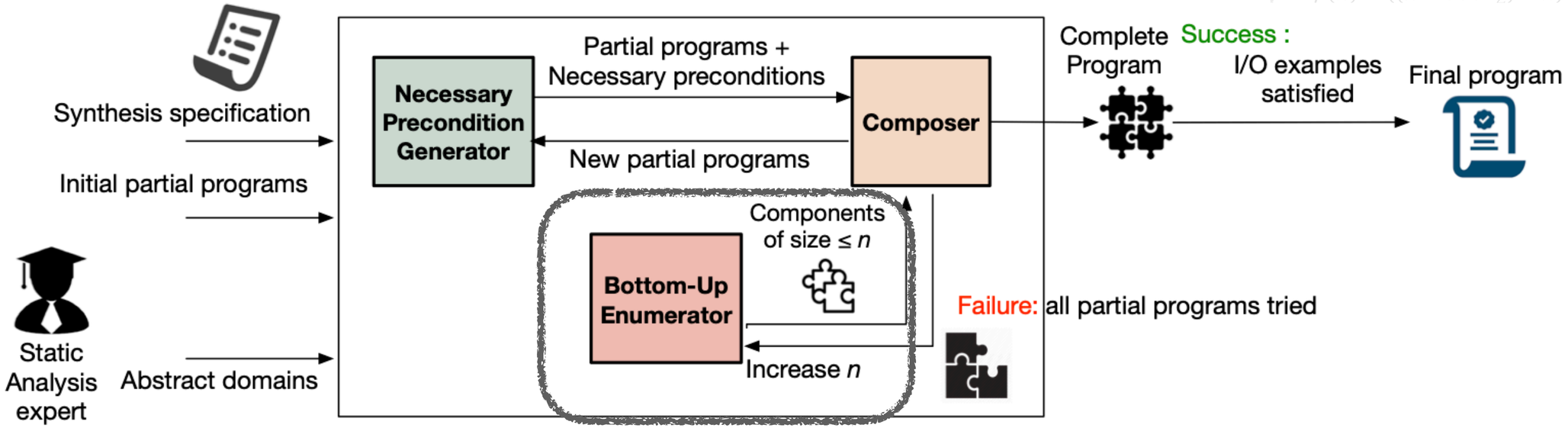
Spec

$$\begin{aligned} S &\rightarrow x \mid 0001_2 \\ &\mid S \wedge S \mid S \vee S \mid S \oplus S \\ &\mid S + S \mid S \times S \mid S / S \mid S >> S \end{aligned}$$

$f(1011_2) = 0011_2$

Goal

$f(x) = ((x + 0001_2) \oplus x) >> 0001_2$



Synthesis Algorithm

Generate Components (n <= 1)

Partial Programs

$(\square \oplus x) >> 0001_2 \rightarrow$

$\square$	$\mapsto$	$110\top$
x	$\mapsto$	1011
$\square \oplus x$	$\mapsto$	$011\top$
$(\square \oplus x) >> 0001_2$	$\mapsto$	0011

Components

$x \ (= 1011_2)$

0001_2

Spec

$$
\begin{aligned}
S &\rightarrow x \mid 0001_2 \\
&\mid S \wedge S \mid S \vee S \mid S \oplus S \\
&\mid S + S \mid S \times S \mid S / S \mid S >> S
\end{aligned}
$$

$f(1011_2) = 0011_2$

Goal

$f(x) = ((x + 0001_2) \oplus x) >> 0001_2$

Synthesis Algorithm

Composition Round 1

Partial Programs

$(\square \oplus x) >> 0001_2 \rightarrow$

$$\begin{array}{rcl} \square & \mapsto & 101\top \\ x & \mapsto & 1011 \\ \square \oplus x & \mapsto & 011\top \\ (\square \oplus x) >> 0001_2 & \mapsto & 0011 \end{array}$$

Components

$x \ (= 1011_2)$

0001_2

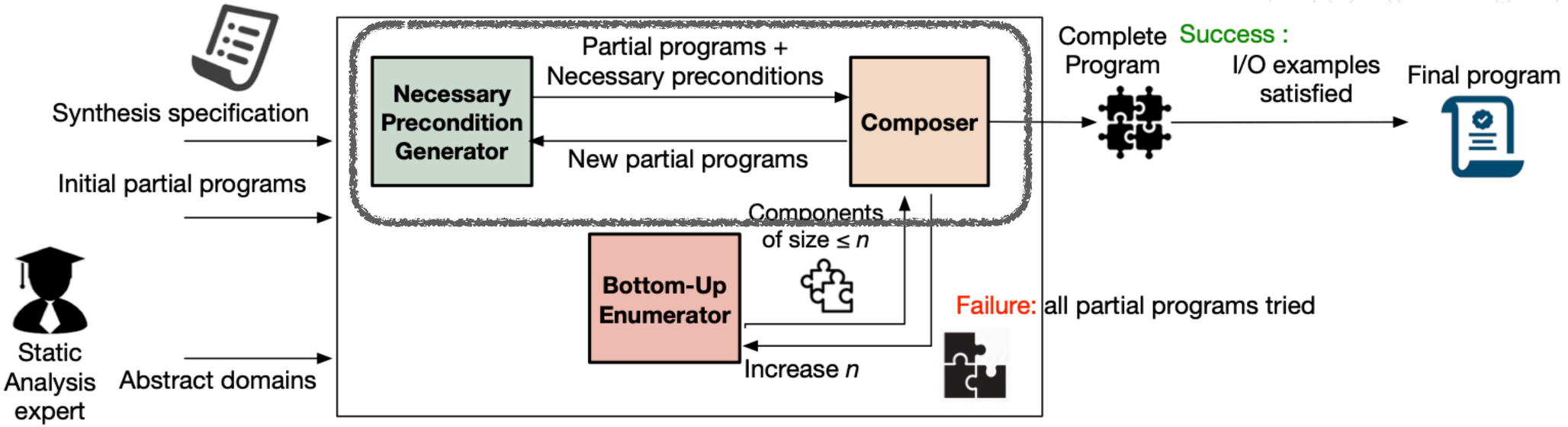
Spec

$$\begin{array}{rcl} S & \rightarrow & x \mid 0001_2 \\ & \mid & S \wedge S \mid S \vee S \mid S \oplus S \\ & \mid & S + S \mid S \times S \mid S / S \mid S >> S \end{array}$$

$f(1011_2) = 0011_2$

Goal

$f(x) = ((x + 0001_2) \oplus x) >> 0001_2$



Synthesis Algorithm

Composition Round 1

Partial Programs

$(\square \oplus x) >> 0001_2 \rightarrow$

$\square$	$\mapsto$	$110\top$
x	$\mapsto$	1011
$\square \oplus x$	$\mapsto$	$011\top$
$(\square \oplus x) >> 0001_2$	$\mapsto$	0011

Components

$x \ (= 1011_2)$

0001_2

No component satisfies precondition -> skip

Spec

$$
\begin{aligned}
S \quad &\rightarrow \quad x \mid 0001_2 \\
&\mid \quad S \wedge S \mid S \vee S \mid S \oplus S \\
&\mid \quad S + S \mid S \times S \mid S / S \mid S >> S
\end{aligned}
$$

$f(1011_2) = 0011_2$

Goal

$f(x) = ((x + 0001_2) \oplus x) >> 0001_2$

Synthesis Algorithm

Generate Components (n <= 3)

Partial Programs

$$(\square \oplus x) >> 0001_2 \rightarrow$$

$$\begin{aligned} \square &\mapsto 101\top \\ x &\mapsto 1011 \\ \square \oplus x &\mapsto 011\top \\ (\square \oplus x) >> 0001_2 &\mapsto 0011 \end{aligned}$$

Components

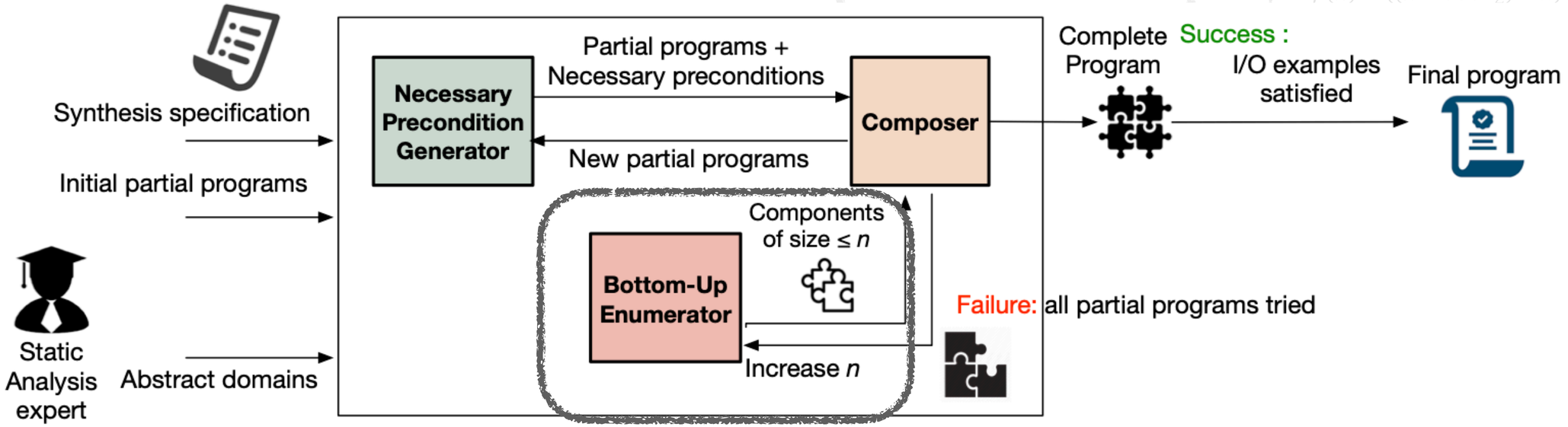
$$x \ (= 1011_2)$$
$$0001_2$$
$$x + 0001_2 \ (= 1100_2)$$
$$x + x \ (= 0110_2)$$
$$x \oplus 0001_2 \ (= 1010_2)$$

Spec

$$\begin{aligned} S &\rightarrow x \mid 0001_2 \\ &\mid S \wedge S \mid S \vee S \mid S \oplus S \\ &\mid S + S \mid S \times S \mid S / S \mid S >> S \end{aligned}$$
$$f(1011_2) = 0011_2$$

Goal

$$f(x) = ((x + 0001_2) \oplus x) >> 0001_2$$



Synthesis Algorithm

Generate Components (n <= 3)

Partial Programs

$(\square \oplus x) >> 0001_2 \rightarrow$

$\square$	$\mapsto$	$110\top$
x	$\mapsto$	1011
$\square \oplus x$	$\mapsto$	$011\top$
$(\square \oplus x) >> 0001_2$	$\mapsto$	0011

Components

$x \ (= 1011_2)$

0001_2

$x + 0001_2 \ (= 1100_2)$

$x + x \ (= 0110_2)$

$x \oplus 0001_2 \ (= 1010_2)$

...

Size increased

Spec

$$
\begin{aligned}
S &\rightarrow x \mid 0001_2 \\
&\mid S \wedge S \mid S \vee S \mid S \oplus S \\
&\mid S + S \mid S \times S \mid S / S \mid S >> S
\end{aligned}
$$

$f(1011_2) = 0011_2$

Goal

$f(x) = ((x + 0001_2) \oplus x) >> 0001_2$

Synthesis Algorithm

Generate Components (n <= 3)

Partial Programs

$(\square \oplus x) >> 0001_2 \rightarrow$

$\square$	$\mapsto$	$110\top$
x	$\mapsto$	1011
$\square \oplus x$	$\mapsto$	$011\top$
$(\square \oplus x) >> 0001_2$	$\mapsto$	0011

Components

$x \ (= 1011_2)$

0001_2

$x + 0001_2 \ (= 1100_2)$

$x + x \ (= 0110_2)$

$x \oplus 0001_2 \ (= 1010_2)$

...

Spec

$S \rightarrow x \mid 0001_2$

$\mid S \wedge S \mid S \vee S \mid S \oplus S$

$\mid S + S \mid S \times S \mid S / S \mid S >> S$

$f(1011_2) = 0011_2$

Goal

$f(x) = ((x + 0001_2) \oplus x) >> 0001_2$

~~$x/1 \ (= 1011_2)$~~

$x \wedge x \ (= 1011_2)$

~~$x \wedge 0001_2 \ (= 1011_2)$~~

~~$x / x \ (= 0001_2)$~~

~~$0001_2 \vee 0001_2 \ (= 0001_2)$~~

Ignore Observational Equivalent Components

Synthesis Algorithm

Composition Round 2

Partial Programs

$(\square \oplus x) >> 0001_2 \rightarrow$

$\square$	$\mapsto$	$110\top$
x	$\mapsto$	1011
$\square \oplus x$	$\mapsto$	$011\top$
$(\square \oplus x) >> 0001_2$	$\mapsto$	0011

Components

$x \ (= 1011_2)$

0001_2

$x + 0001_2 \ (= 1100_2)$

$x + x \ (= 0110_2)$

$x \oplus 0001_2 \ (= 1010_2)$

...

Spec

$$
\begin{aligned}
S \quad &\rightarrow \quad x \mid 0001_2 \\
&\mid \quad S \wedge S \mid S \vee S \mid S \oplus S \\
&\mid \quad S + S \mid S \times S \mid S / S \mid S >> S
\end{aligned}
$$

$f(1011_2) = 0011_2$

Goal

$f(x) = ((x + 0001_2) \oplus x) >> 0001_2$

Synthesis Algorithm

Composition Round 2

Partial Programs

$(\square \oplus x) >> 0001_2 \rightarrow$

$\square$	$\mapsto$	$110\top$
x	$\mapsto$	1011
$\square \oplus x$	$\mapsto$	$011\top$
$(\square \oplus x) >> 0001_2$	$\mapsto$	0011

$\{1100, 1101\} \rightarrow$

Components

$x \ (= 1011_2)$

0001_2

$x + 0001_2 \ (= 1100_2)$

$x + x \ (= 0110_2)$

$x \oplus 0001_2 \ (= 1010_2)$

...

Lookup Table

0001	:	0001_2
0110	:	$x + x$
1010	:	$x \oplus 0001_2$
1011	:	x
1100	:	$x + 0001_2$

Spec

$$
\begin{aligned}
S \ &\rightarrow \ x \mid 0001_2 \\
&\mid \ S \wedge S \mid S \vee S \mid S \oplus S \\
&\mid \ S + S \mid S \times S \mid S / S \mid S >> S
\end{aligned}
$$

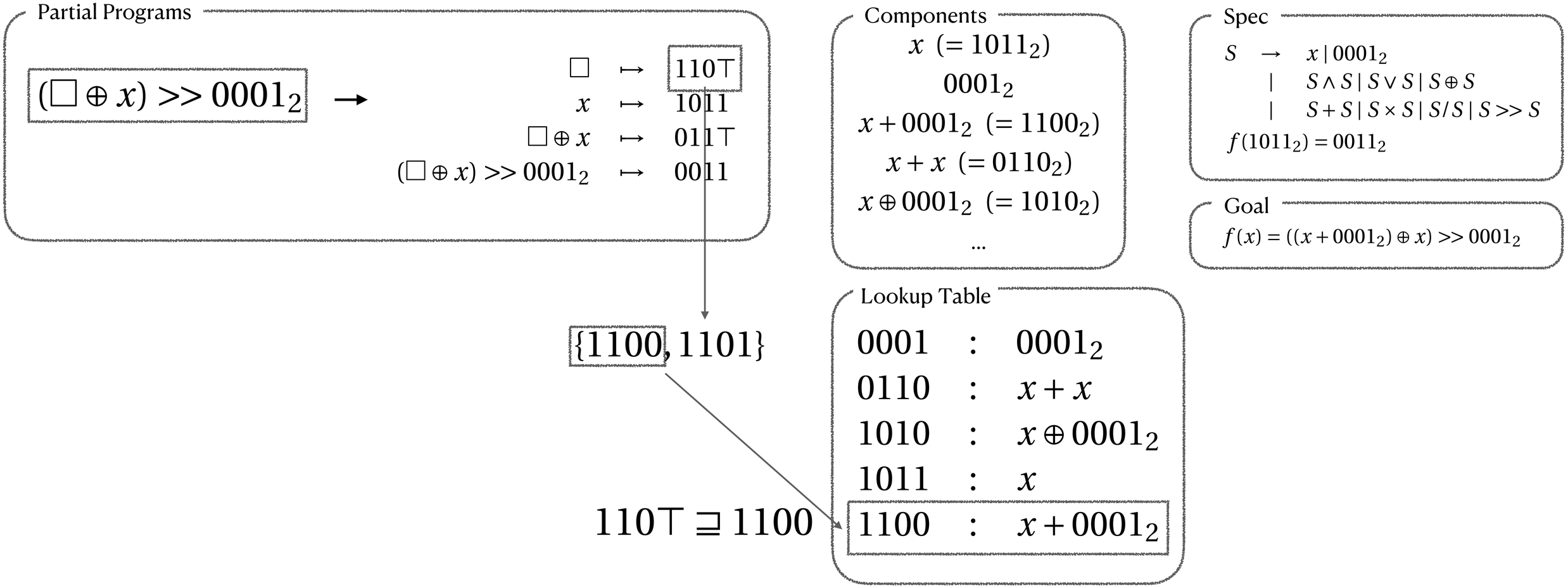
$f(1011_2) = 0011_2$

Goal

$f(x) = ((x + 0001_2) \oplus x) >> 0001_2$

Synthesis Algorithm

Composition Round 2



Synthesis Algorithm

Found Solution

Partial Programs

$(\square \oplus x) >> 0001_2$ →

$\square$	$\mapsto$	$110\top$
x	$\mapsto$	1011
$\square \oplus x$	$\mapsto$	$011\top$
$(\square \oplus x) >> 0001_2$	$\mapsto$	0011

Components

$x \ (= 1011_2)$

0001_2

$x + 0001_2 \ (= 1100_2)$

$x + x \ (= 0110_2)$

$x \oplus 0001_2 \ (= 1010_2)$

...

Spec

$$
\begin{aligned}
S &\rightarrow x \mid 0001_2 \\
&\mid S \wedge S \mid S \vee S \mid S \oplus S \\
&\mid S + S \mid S \times S \mid S / S \mid S >> S
\end{aligned}
$$

$f(1011_2) = 0011_2$

Goal

$f(x) = ((x + 0001_2) \oplus x) >> 0001_2$

New Candidate Solution

$f(x) = ((x + 0001_2) \oplus x) >> 0001_2$

Solution Found!

$$
\begin{aligned}
f(1011_2) &= ((1011_2 + 0001_2) \oplus 1011_2) >> 0001_2 \\
&= (1100_2 \oplus 1011_2) >> 0001_2 \\
&= 0111_2 >> 0001_2 \\
&= 0011_2
\end{aligned}
$$

Synthesizing Conditional Programs

Simba Core + Divde-and-Conquer

- Incorporate the divide-and-conquer approach [Alur et al. 2017] into our algorithm

Syntactic Spec

$$
\begin{array}{rcl}
S & \rightarrow & x \mid 0000_2 \mid 0001_2 \\
 & \mid & S + S \mid \cdots \\
 & \mid & \boxed{\text{if } S \text{ then } S \text{ else } S}
\end{array}
$$

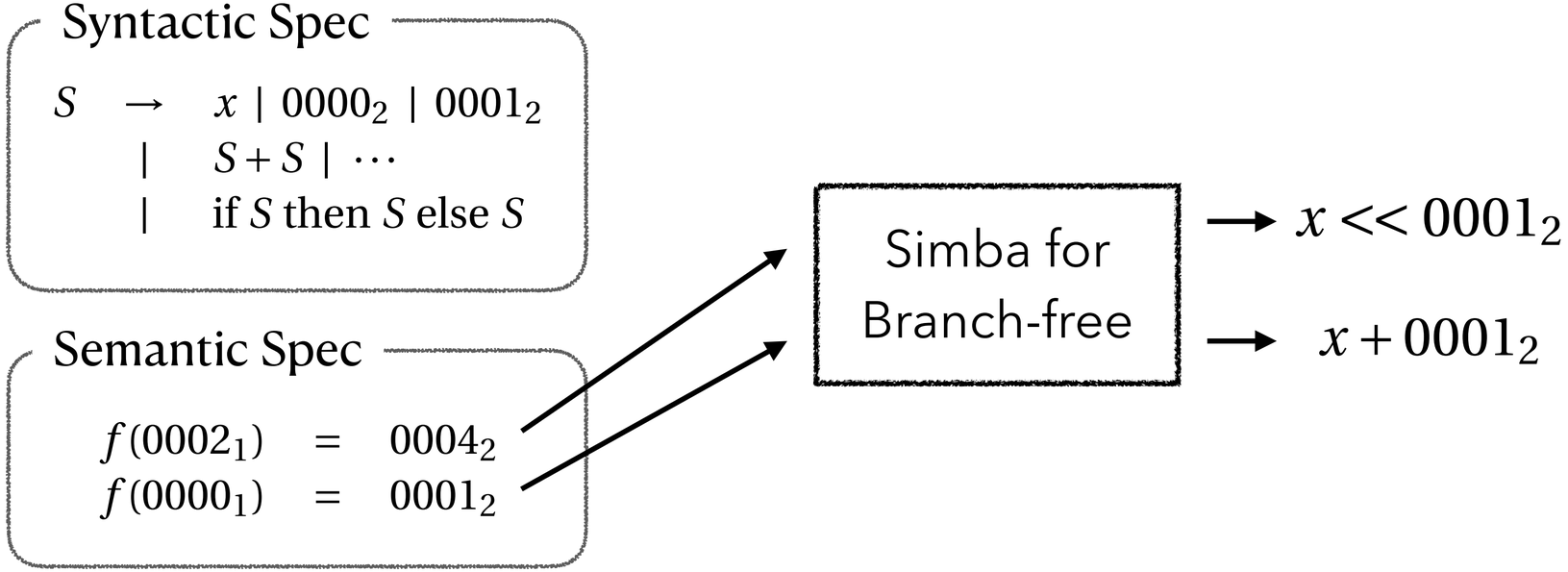
Semantic Spec

$$
\begin{array}{rcl}
f(0002_1) & = & 0004_2 \\
f(0000_1) & = & 0001_2
\end{array}
$$

Synthesizing Conditional Programs

Simba Core + Divde-and-Conquer

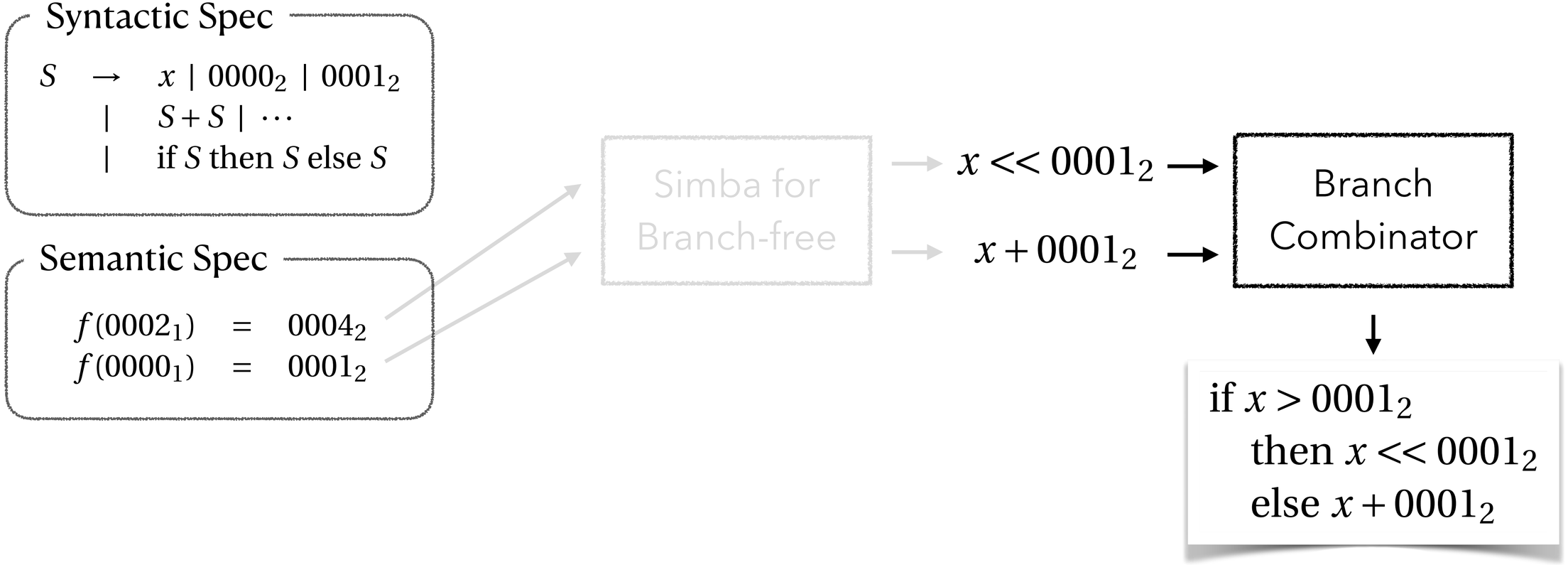
- Incorporate the divide-and-conquer approach [Alur et al. 2017] into our algorithm



Synthesizing Conditional Programs

Simba Core + Divde-and-Conquer

- Incorporate the divide-and-conquer approach [Alur et al. 2017] into our algorithm



Evaluation: Benchmarks

- 1875 synthesis tasks from 3 domains: BitVec, BitVec-Cond, Circuit

Background Theory	**Bit-vector arithmetic**			**SAT**	
Domain	**BitVec (# 544)**		**BitVec-Cond (# 750)**	**Circiut (# 581)**	
Benchmarks	HD	Deobfusc	-	Lobster	Crypto
# Tasks	44	500	750	369	212

Evaluation: Benchmarks

- 1875 synthesis tasks from 3 domains: BitVec, BitVec-Cond, Circuit

Background Theory	**Bit-vector arithmetic**			**SAT**	
Domain	**BitVec (# 544)**		**BitVec-Cond (# 750)**	**Circiut (# 581)**	
Benchmarks	HD	Deobfusc	-	Lobster	Crypto
# Tasks	44	500	750	369	212

Evaluation: Benchmarks

- 1875 synthesis tasks from 3 domains: BitVec, BitVec-Cond, Circuit

Background Theory	**Bit-vector arithmetic**			**SAT**	
Domain	**BitVec (# 544)**		**BitVec-Cond (# 750)**	**Circiut (# 581)**	
Benchmarks	HD	Deobfusc	-	Lobster	Crypto
# Tasks	44	500	750	369	212

Evaluation: Benchmarks

- 1875 synthesis tasks from 3 domains: BitVec, BitVec-Cond, Circuit

Background Theory	**Bit-vector arithmetic**			**SAT**	
Domain	**BitVec (# 544)**		**BitVec-Cond (# 750)**	**Circiut (# 581)**	
Benchmarks	HD	Deobfusc	-	Lobster	Crypto
# Tasks	44	500	750	369	212

Evaluation: Benchmarks

- 1875 synthesis tasks from 3 domains: BitVec, BitVec-Cond, Circuit

Background Theory	**Bit-vector arithmetic**			**SAT**	
Domain	**BitVec (# 544)**		**BitVec-Cond (# 750)**	**Circiut (# 581)**	
Benchmarks	HD	Deobfusc	-	Lobster	Crypto
# Tasks	44	500	750	369	212

Hacker's **D**elight
from SyGuS competition

SyGuS competition PBE-**BitVec**tor
(may contain **cond**itionals)

Circuit transformation
related to **crypto**graphic modules
from SyGus competition

Evaluation: Benchmarks

- 1875 synthesis tasks from 3 domains: BitVec, BitVec-Cond, Circuit

Background Theory	Bit-vector arithmetic			SAT	
Domain	**BitVec (# 544)**		**BitVec-Cond (# 750)**	**Circiut (# 581)**	
Benchmarks	HD	Deobfusc	-	Lobster	Crypto
# Tasks	44	500	750	369	212

Hacker's **D**elight
from SyGuS competition

Deobfuscator
from QSynth [David et al. 2020]

SyGuS competition PBE-**BitVec**tor
(may contain **cond**itionals)

Circuit transformation
from **Lobster** [Lee et al. 2020]

Circuit transformation
related to **crypto**graphic modules
from SyGus competition

Evaluation: Baseline Solvers

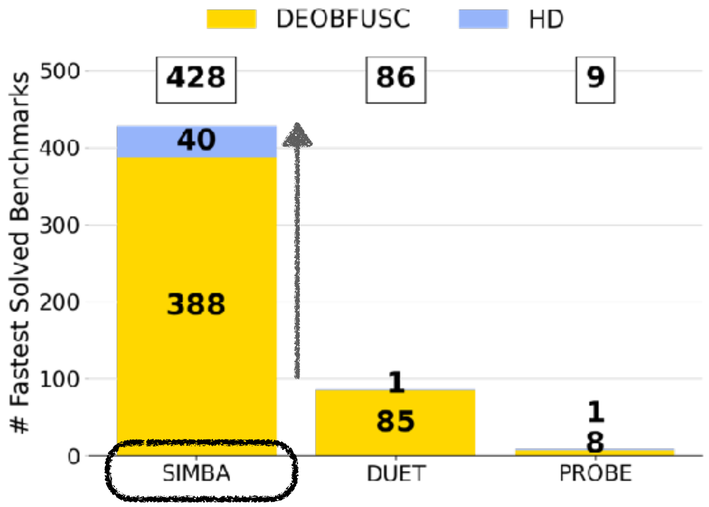
- Duet [Lee 2021]: the state-of-the-art synthesis tool for inductive SyGuS problems
  - Employs a bidirectional(top-down + bottom-up) search strategy
  - Requires *inverse semantics operators* for search space pruning

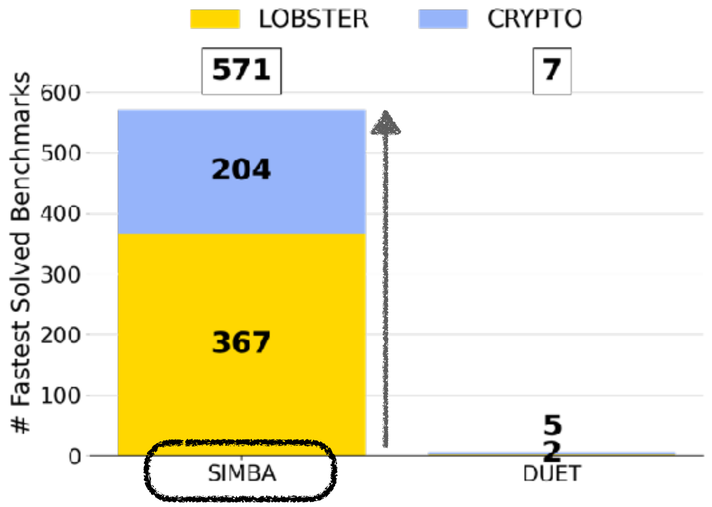
- Probe [Barke et al. 2020]: tool performs a bottom-up search with probabilistic model
  - Probabilistic model is learned *just in time* during the search processes

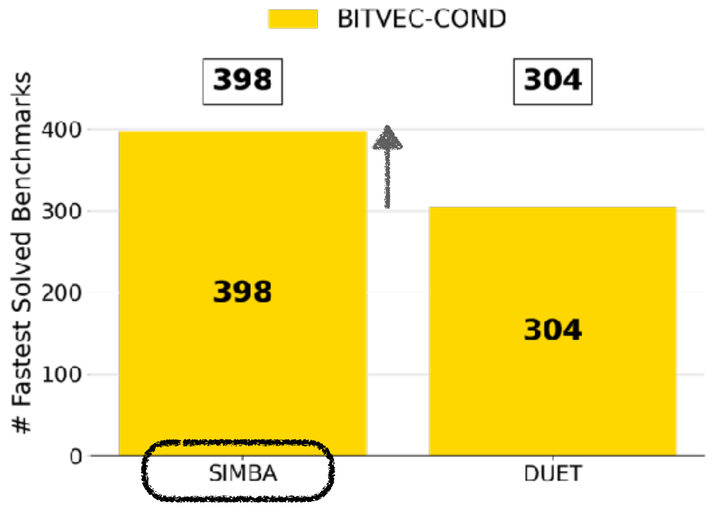
Evaluation: Simba Performance

Solve Faster

- Outperformed baseline solvers for conditional-free program
- Comparable to baseline solvers for conditional program





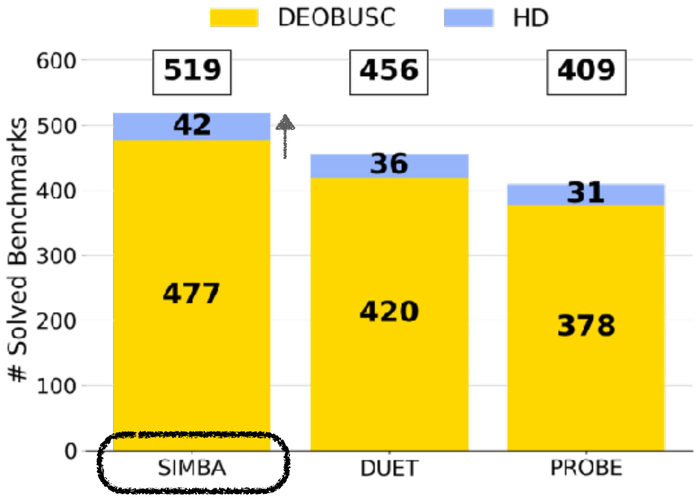


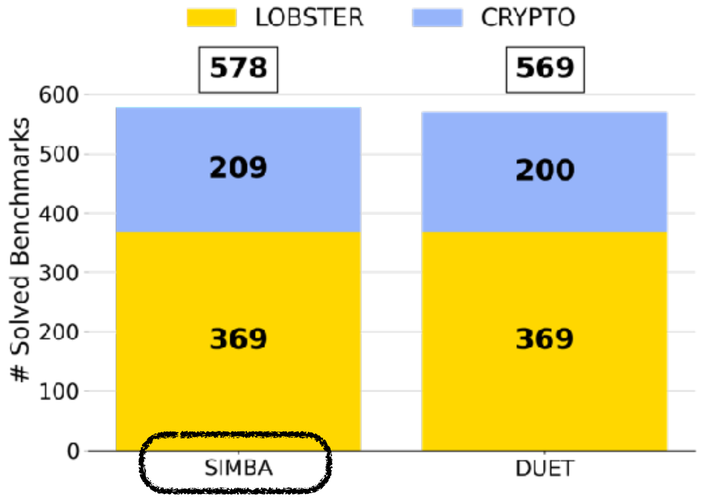
of the fastest-solved problems for each domain

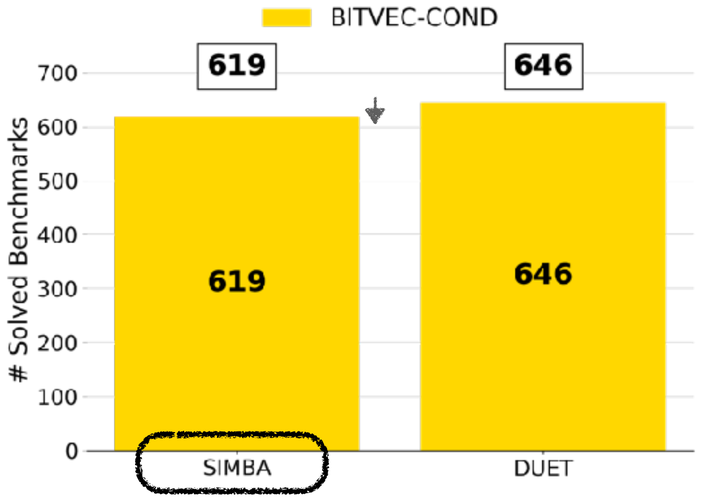
Evaluation: Simba Performance

Solve More

- Outperformed baseline solvers for conditional-free program
- Comparable to baseline solvers for conditional program





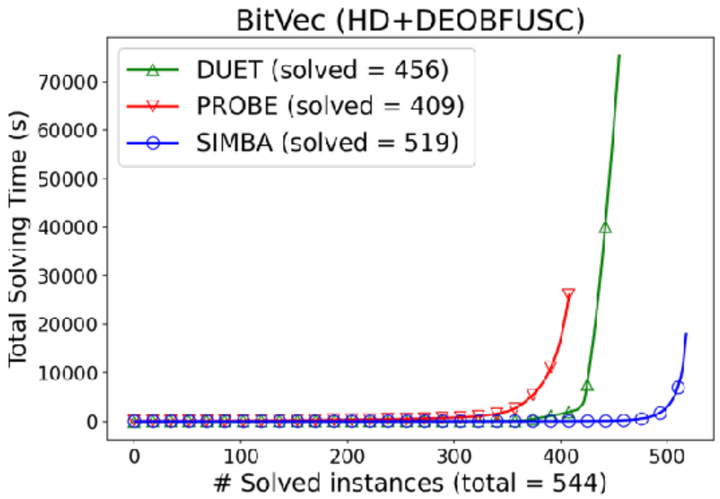


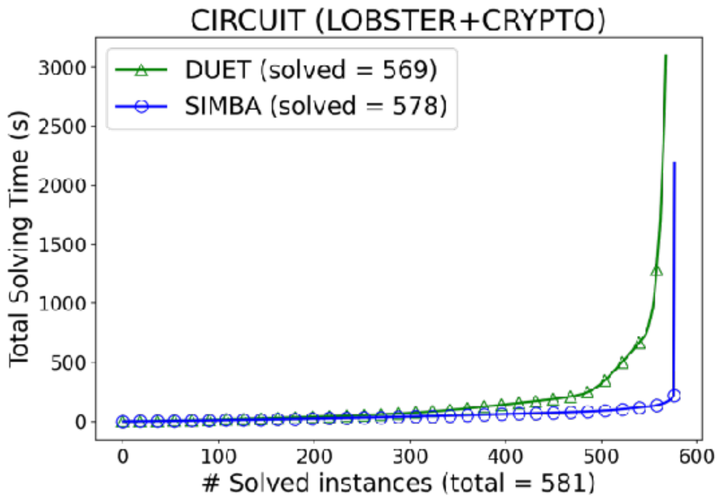
of the solved problems with 1h timeout for each domain

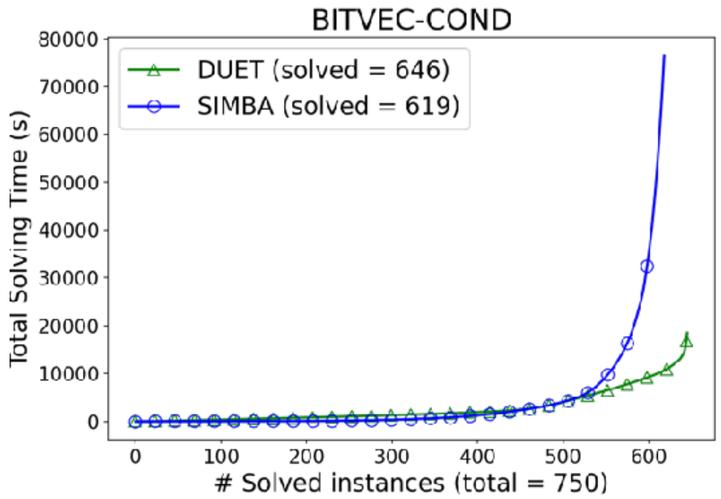
Evaluation: Simba Performance

Overall

- SIMBA(plot ○) shows the best performance for branch-free benchmarks
- DUET(plot △) is better than SIMBA for branch benchmarks in average





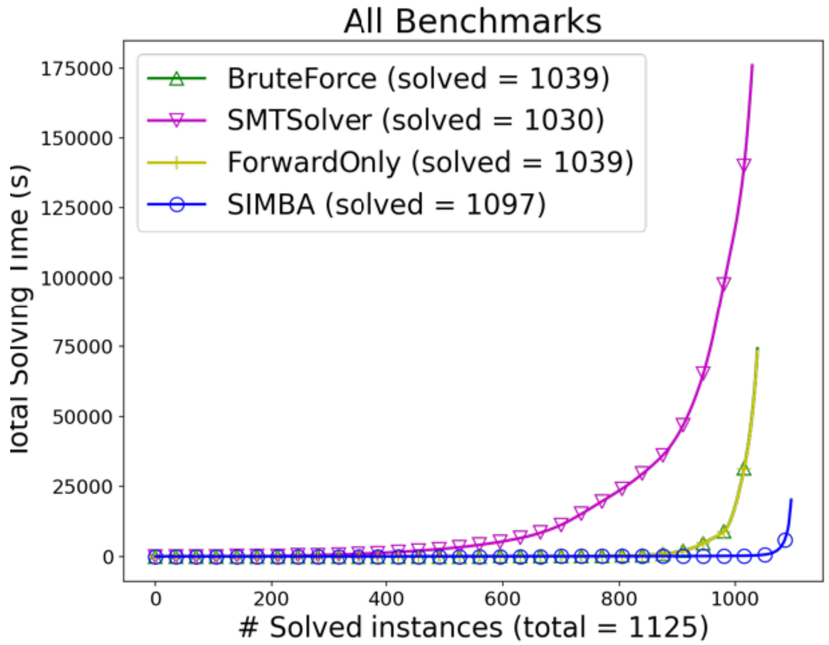


Cactus-plots

Evaluation: Simba Performance

Ablation Study

- Forward-Backward Analysis(plot ○) is necessary to achieve the best performance
- Only forward analysis(plot +) or SMT Solver(plot ▽) is not sufficient
  - Sometimes, even worse than brute force(plot △) because of bad analysis cost



In the paper...

- Formalization (abstract domain, forward-backward transfer functions, etc.)
- Detailed Algorithm of synthesis and forward-backward analysis
- Why our tool outperforms the baseline solvers (case study)
- and more

⬇Paper

😊 Thank You!

GitHub Repo⬇