Program Synthesis via Bi-directional Reduced-product Abstract Interpretation

Kwangkeun Yi
(co-work with Yongho Yoon and Woosuk Lee)

Seoul National University, Korea

01/20/2024 @ N40AI at POPL'24, London



Our Story Upon Abstract Interpretation

- program analysis

- program synthesis

Our Story Upon Abstract Interpretation

- program analysis
  workbench: a more-sound & global analysis for full C

- program synthesis

Our Story Upon Abstract Interpretation

- program analysis
  workbench: a more-sound & global analysis for full C
  - commercialization: SparrowFasoo.com (∼ 2010)
    - safety+security checkers for over 25 programming languages
    - less-sound, lots of compromises

- program synthesis

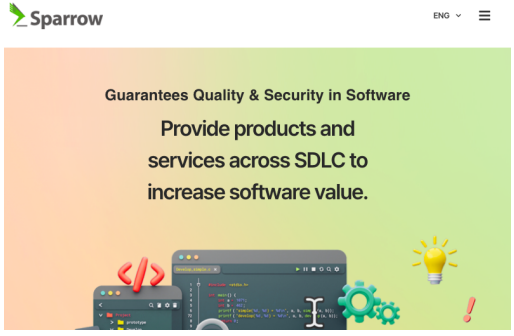
Our Story Upon Abstract Interpretation

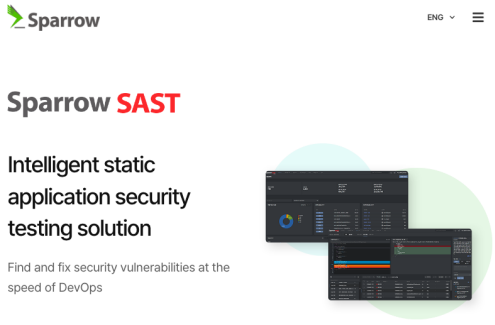
- program analysis
  workbench: a more-sound & global analysis for full C
  - commercialization: SparrowFasoo.com (∼ 2010)
    - safety+security checkers for over 25 programming languages
    - less-sound, lots of compromises
  - scalability: sparse analysis framework (TOPLAS'14, PLDI'12)
  - precision: selective x-sensitivity (TOPLAS'18, TOPLAS'16, PLDI'14)
    - github.com/ropas/sparrow
- program synthesis

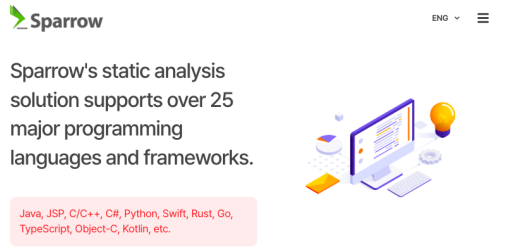
Our Story Upon Abstract Interpretation

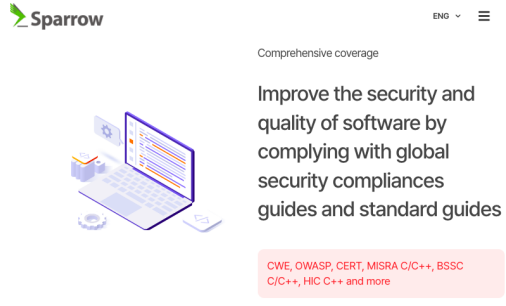
- program analysis
  workbench: a more-sound & global analysis for full C
  - commercialization: SparrowFasoo.com (∼ 2010)
    - safety+security checkers for over 25 programming languages
    - less-sound, lots of compromises
  - scalability: sparse analysis framework (TOPLAS'14, PLDI'12)
  - precision: selective x-sensitivity (TOPLAS'18, TOPLAS'16, PLDI'14)
    - github.com/ropas/sparrow
- program synthesis
  - superb in performance (PLDI'23, PLDI'20)
  - by bi-directional reduced-product AI with finite refinements

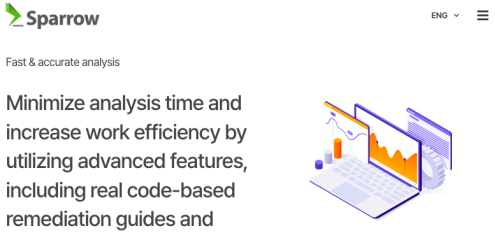
Analysis Commercialization: Our Story Upon Abstract Interpretation

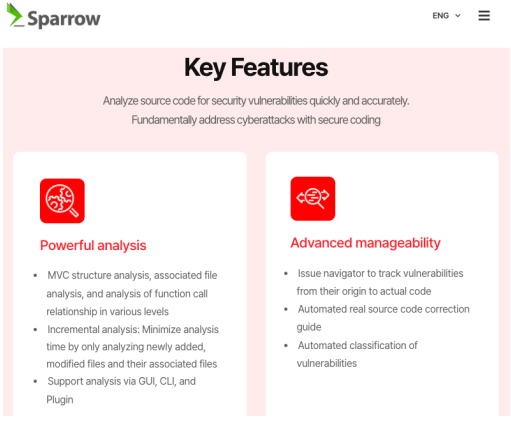










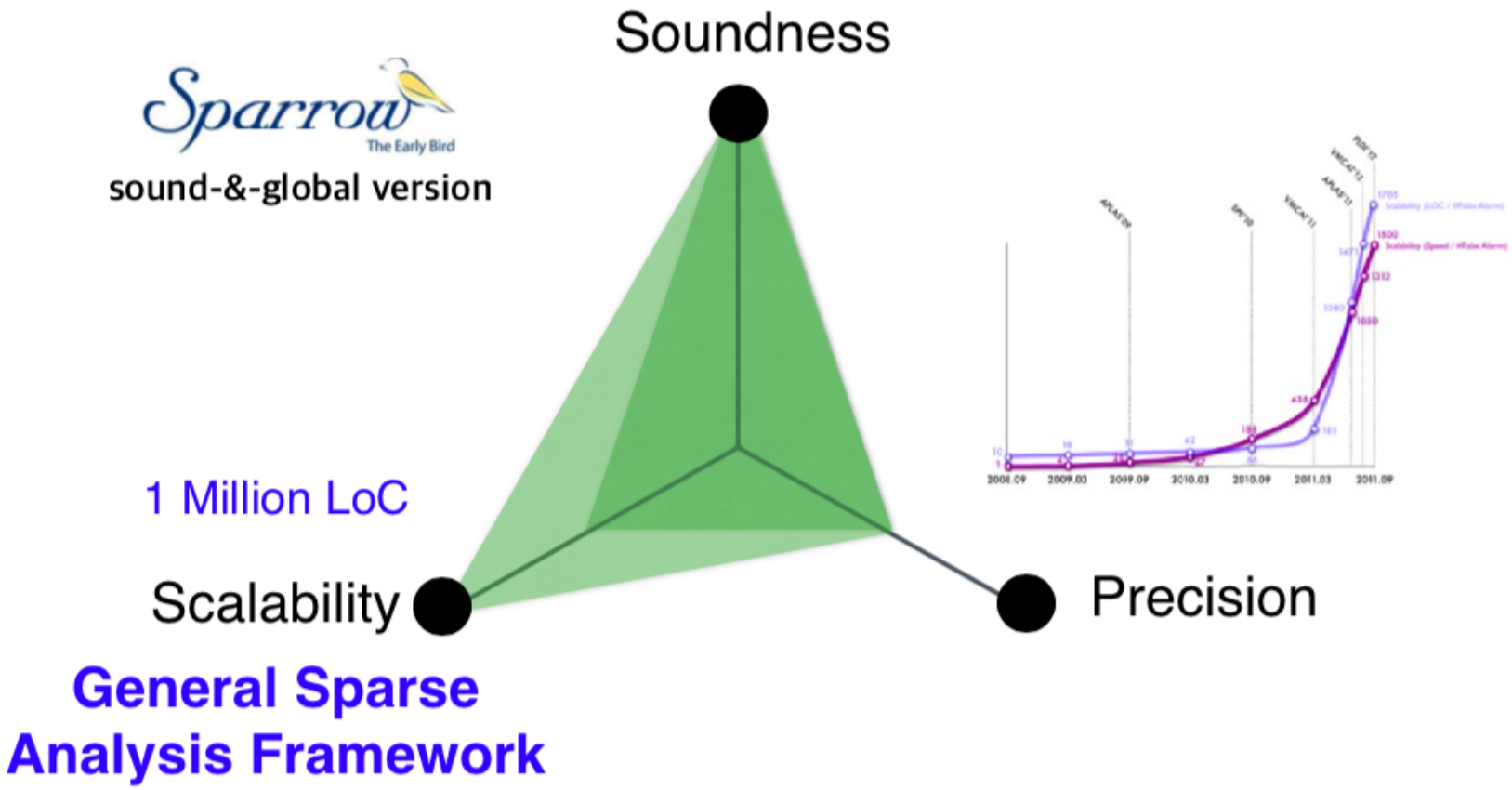


Sparrow
The Early Bird

ROSAEC center
Research On Software Analysis for Error-free Computing
소프트웨어 무결점 연구센터 NRF ERC

, thanks to AI.

Analysis Scalability: Our Story Upon Abstract Interpretation



(TOPLAS'14, PLDI'12)

Analysis Scalability: Our Story Upon Abstract Interpretation

- Idea: exploit the semantic sparsity
  - spatial & temporal sparsities

Analysis Scalability: Our Story Upon Abstract Interpretation

- Idea: exploit the semantic sparsity
  - spatial & temporal sparsities
- Sparse analysis framework
  - general & precision-preserving

$$F^{\#} : D^{\#} \to D^{\#} \quad \overset{\text{sparsify}}{\Longrightarrow} \quad F_s^{\#} : D^{\#} \to D^{\#}$$
$$fix F^{\#} \quad \overset{\text{still}}{=} \quad fix F_s^{\#}$$

(TOPLAS'14, PLDI'12)

Analysis Scalability: Our Story Upon Abstract Interpretation

- Idea: exploit the semantic sparsity
  - spatial & temporal sparsities
- Sparse analysis framework
  - general & precision-preserving

$$F^\# : D^\# \to D^\# \quad \overset{\text{sparsify}}{\Longrightarrow} \quad F_s^\# : D^\# \to D^\#$$
$$fix F^\# \quad \overset{\text{still}}{=} \quad fix F_s^\#$$

(TOPLAS'14, PLDI'12)

A review:

> *"the theoretical result is very general. It could be applied to many other analyses. PLDI papers have been accepted that were simply instances of this framework."*

, thanks to AI.

Spatial & Temporal Localizations

$\hat{X}, \hat{X}' \in \mathbb{C} \to \hat{\mathbb{S}}$

$\hat{f}_c \in \hat{\mathbb{S}} \to \hat{\mathbb{S}}$

$\hat{X} := \hat{X}' := \lambda c.\bot$

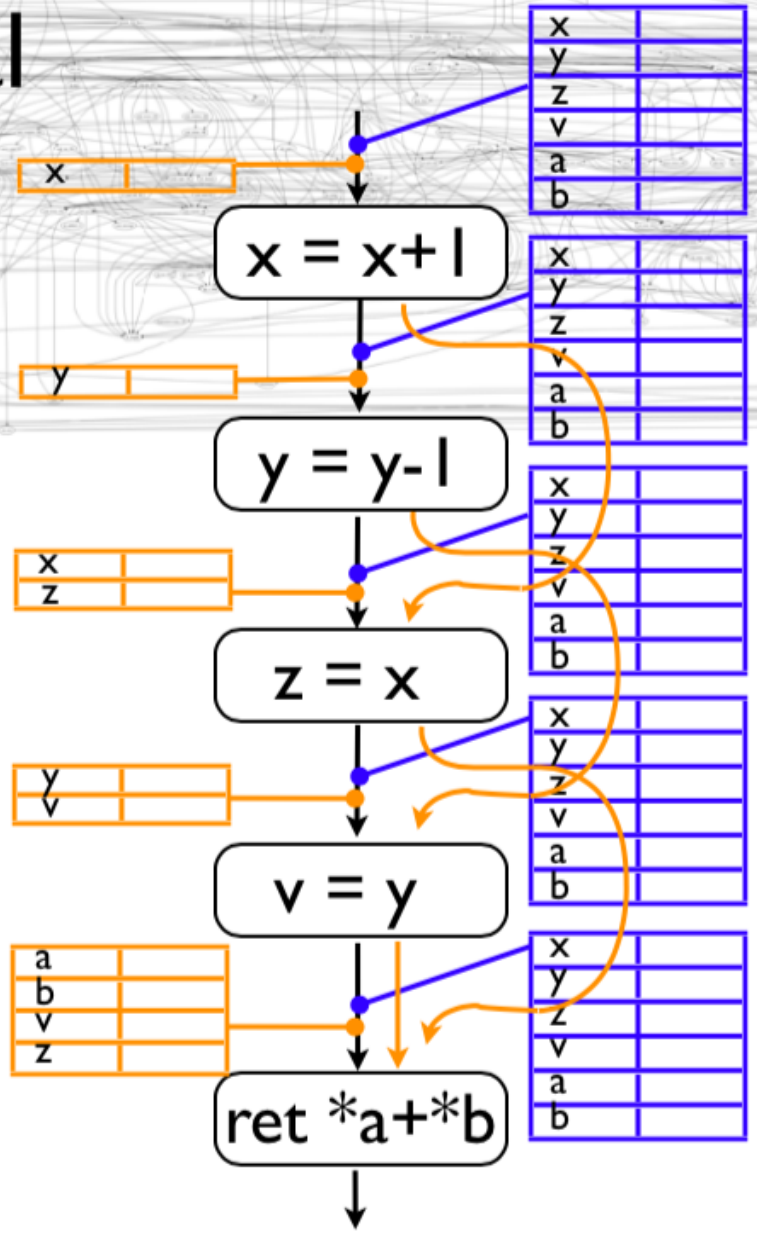
repeat

$\quad \hat{X}' := \hat{X}$

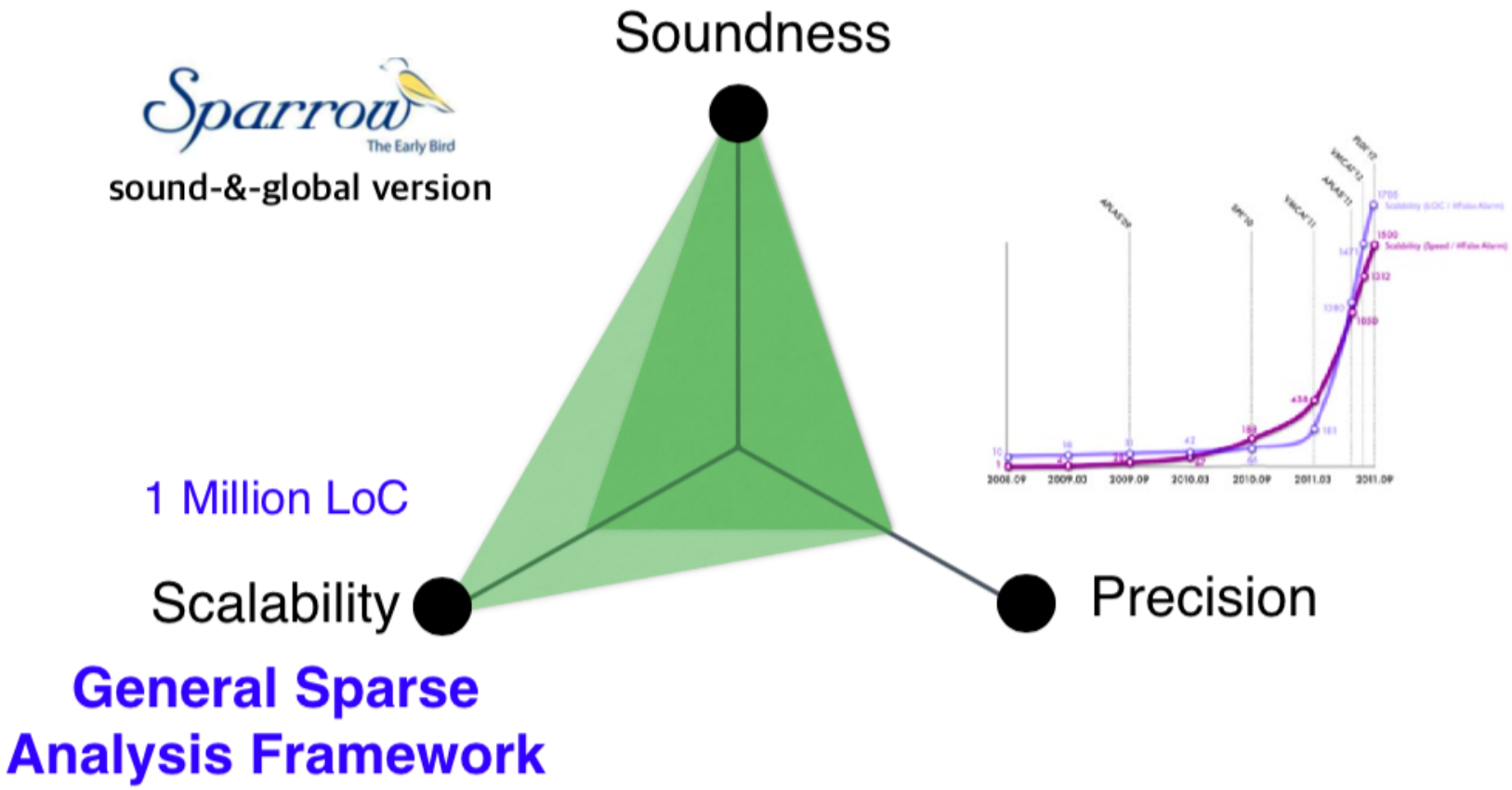
$\quad$ **for all** $c \in \mathbb{C}$ **do**

$\quad\quad \hat{X}(c) := \hat{f}_c(\bigsqcup_{c' \hookrightarrow c} \hat{X}(c'))$

until $\hat{X} \sqsubseteq \hat{X}'$

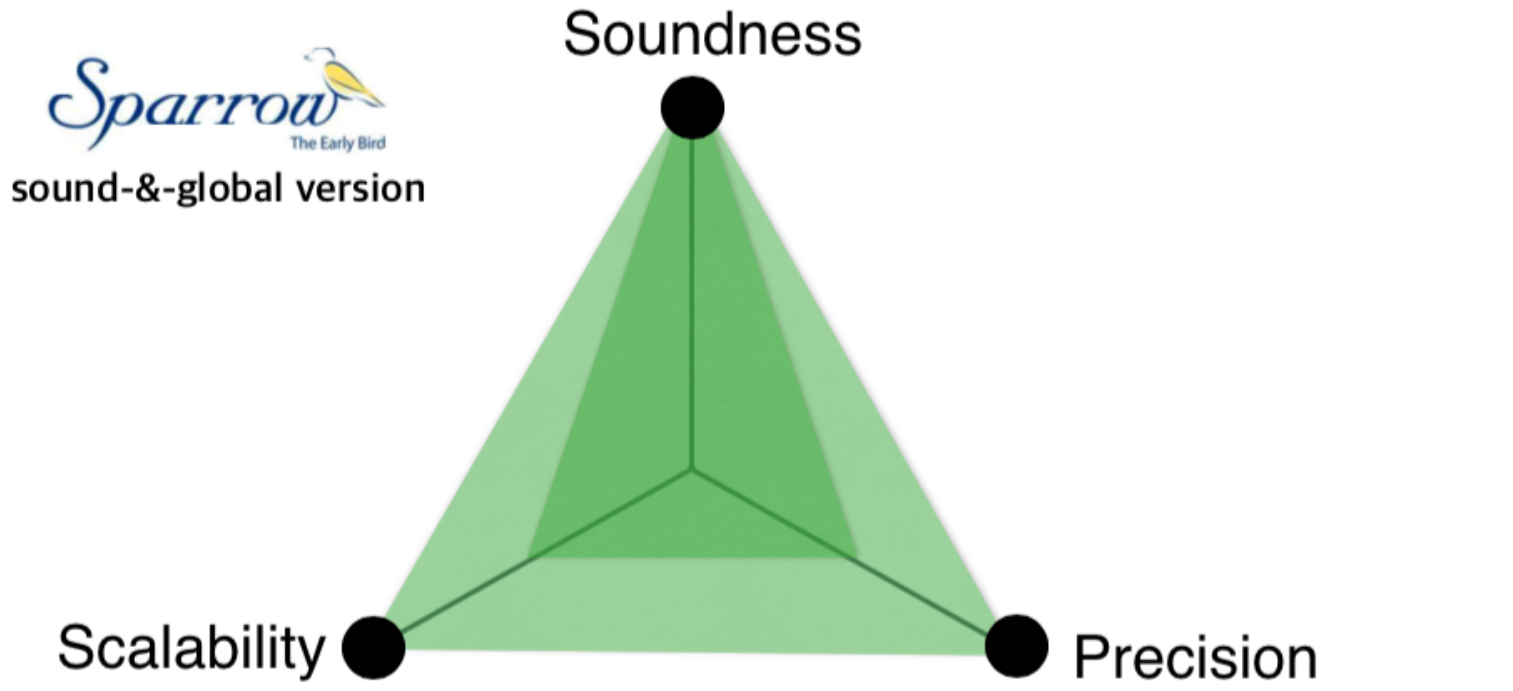


Analysis Scalability: Our Story Upon Abstract Interpretation



(TOPLAS'14, PLDI'12)

Analysis Precision: Our Story Upon Abstract Interpretation

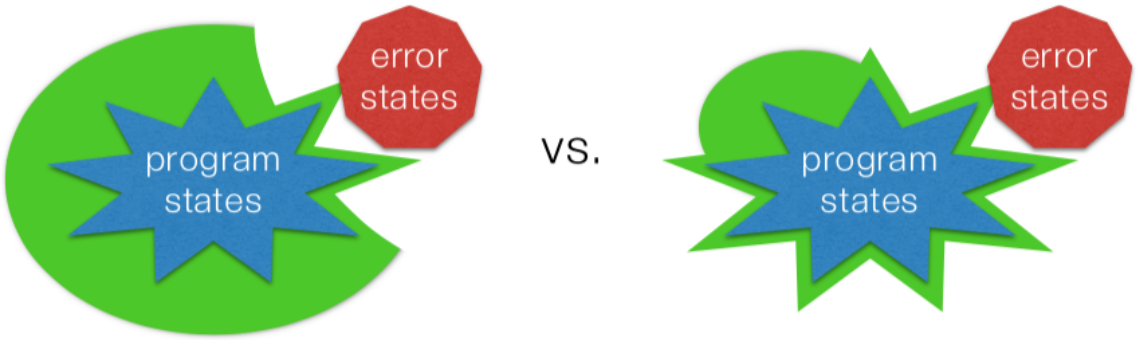


9

(TOPLAS'14, PLDI'12)

(TOPLAS'18, TOPLAS'16, PLDI'14)

Analysis Precision: Our Story Upon Abstract Interpretation



our method: **24%** / **28%**

3-CFA: **24%** / **1300%**

- Selective flow- or context-sensitivity
- i.e., sensitivity only when it would pay off
  - by (design) pre-analysis with 2-element domain (thanks to AI)
  - by (train) machine learning (Bayesian opt.) for linear score ftn
  - side cost: design $<$ train (thanks to AI)

(TOPLAS'18, TOPLAS'16, OOPSLA'16, PLDI'14)

Program Synthesis: Our Story Upon Abstract Interpretation (PLDI'23, PLDI'20)

Inductive synthesis

- input: a grammar + input-ouput pairs
- output: a "right" program in the grammar

Example

- input

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

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

- output

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

Automatic Synthesis Examples

- input

$$\begin{array}{rcl} S & \rightarrow & n \mid x \mid S \wedge S \mid S \vee S \mid S \oplus S \\ & \mid & S \gg S \mid S \ggg S \mid S \ll S \mid S \lll S \\ & \mid & -S \mid S + S \mid S \times S \mid S/S \end{array}$$

  and 5-20 i/o pairs
- ouput (in 0.k sec - 35 min)

Automatic Synthesis Examples

- input

$$\begin{array}{rcl} S & \rightarrow & n \mid x \mid S \wedge S \mid S \vee S \mid S \oplus S \\ & \mid & S \gg S \mid S \ggg S \mid S \ll S \mid S \lll S \\ & \mid & -S \mid S + S \mid S \times S \mid S/S \end{array}$$

  and 5-20 i/o pairs
- ouput (in 0.k sec - 35 min)
  - abs(x) : $8b = (x \ll (x \gg x)) \vee (x \times (x \gg x))$
  - sign(x) : $8b = ((-x) \ggg 7) \vee (x \gg x)$

Automatic Synthesis Examples

- input

$$\begin{aligned} S &\rightarrow n \mid x \mid S \wedge S \mid S \vee S \mid S \oplus S \\ &\mid S \gg S \mid S \ggg S \mid S \ll S \mid S \lll S \\ &\mid -S \mid S + S \mid S \times S \mid S/S \end{aligned}$$

  and 5-20 i/o pairs
- ouput (in 0.k sec - 35 min)
  - abs$(x) : 8b = (x \ll (x \gg x)) \vee (x \times (x \gg x))$
  - sign$(x) : 8b = ((-x) \ggg 7) \vee (x \gg x)$
  - averageFloor$(x, y) : u8b = ((x \oplus y) \ggg 1) + (x \wedge y)$
  - averageCeil$(x, y) : u8b = (x \vee y) - ((x \oplus y) \ggg 1)$

Automatic Synthesis Examples

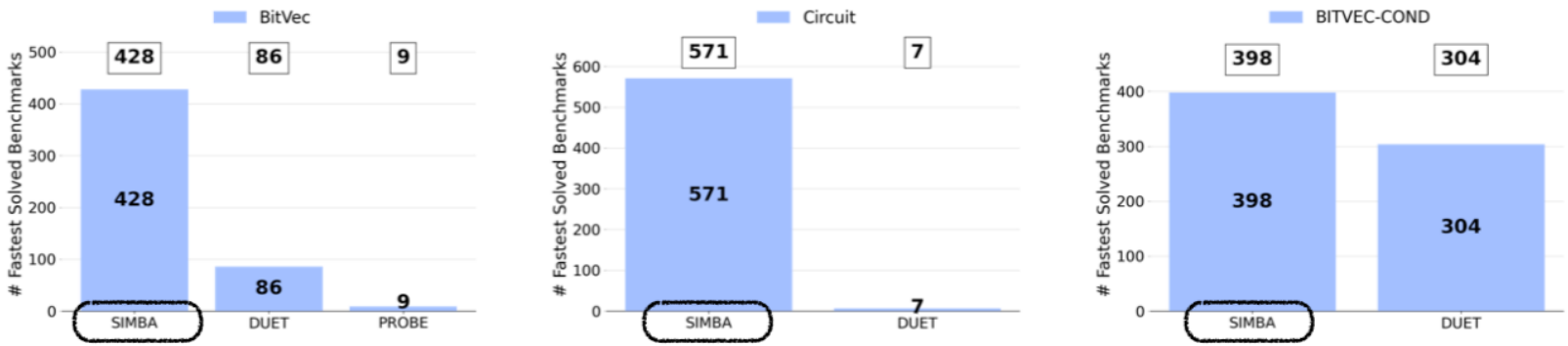
- input

$$\begin{aligned} S &\rightarrow n \mid x \mid S \wedge S \mid S \vee S \mid S \oplus S \\ &\mid S \gg S \mid S \ggg S \mid S \ll S \mid S \lll S \\ &\mid -S \mid S + S \mid S \times S \mid S/S \end{aligned}$$

  and 5-20 i/o pairs
- ouput (in 0.k sec - 35 min)
  - abs$(x) : 8b = (x \ll (x \gg x)) \vee (x \times (x \gg x))$
  - sign$(x) : 8b = ((-x) \ggg 7) \vee (x \gg x)$
  - averageFloor$(x, y) : u8b = ((x \oplus y) \ggg 1) + (x \wedge y)$
  - averageCeil$(x, y) : u8b = (x \vee y) - ((x \oplus y) \ggg 1)$
  - deObfus$(x, y) : 64b =$
    $((x \vee y) \wedge (-x \vee -(y \times y))) - (x \vee y \vee (-x \vee -(y \times y)))$

Synthesis Performance

as of 2023



of the fastest-solved problems for each domain

thanks to bi-directional reduced-product AI with finite refinements

Synthesis Key: Pruning by AI

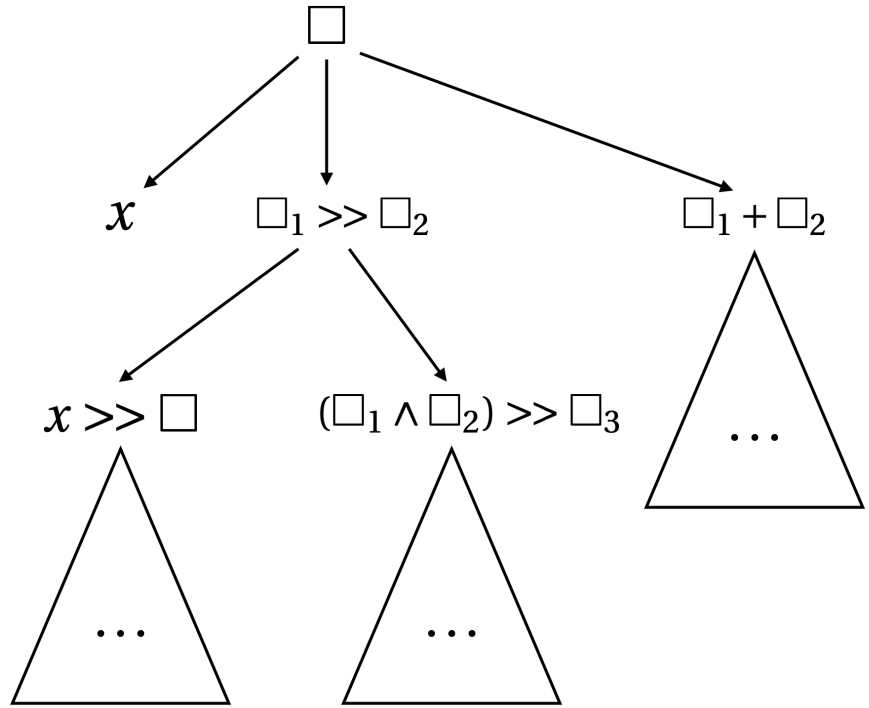
$$
\begin{array}{rcl}
S & \rightarrow & x \mid 0001_2 \mid S \gg S \\
 & \mid & S \wedge S \mid S \vee S \mid S \oplus S \\
 & \mid & S + S \mid S \times S \mid S/S
\end{array}
$$

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

Synthesis Key: Pruning by AI

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

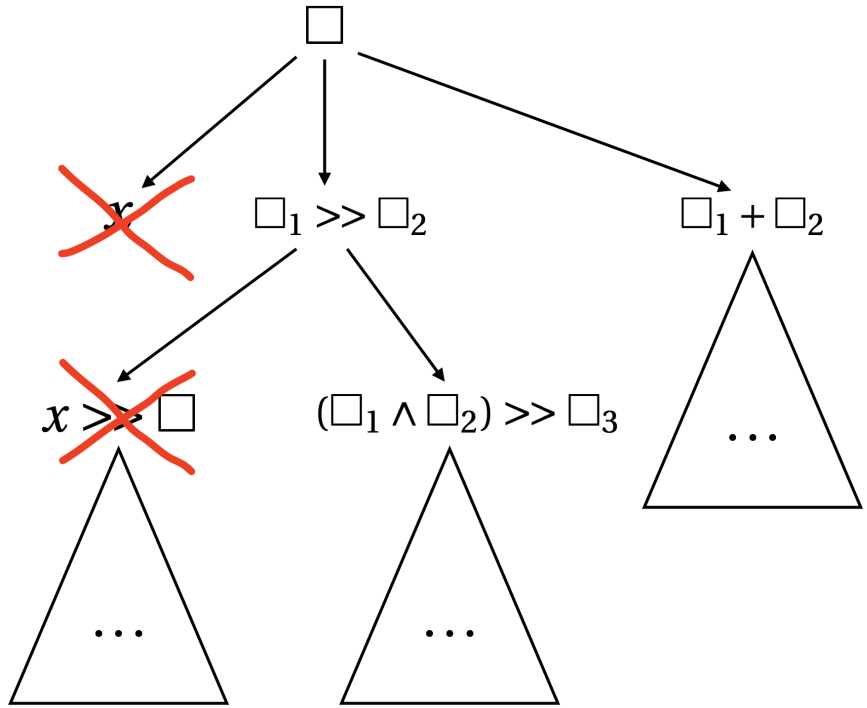
$$f(1011_2) = 0011_2$$



Synthesis Key: Pruning by AI

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

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



because $\nexists \square : f(1011_2) = (1011_2 \gg \square) = 0011_2$

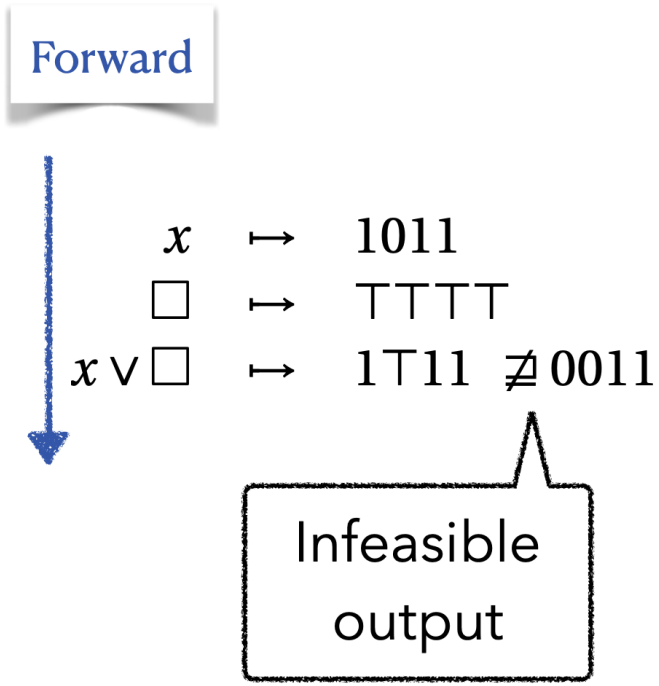
Synthesis Pruning by AI: Forward Analysis

Candidate partial program

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

for

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



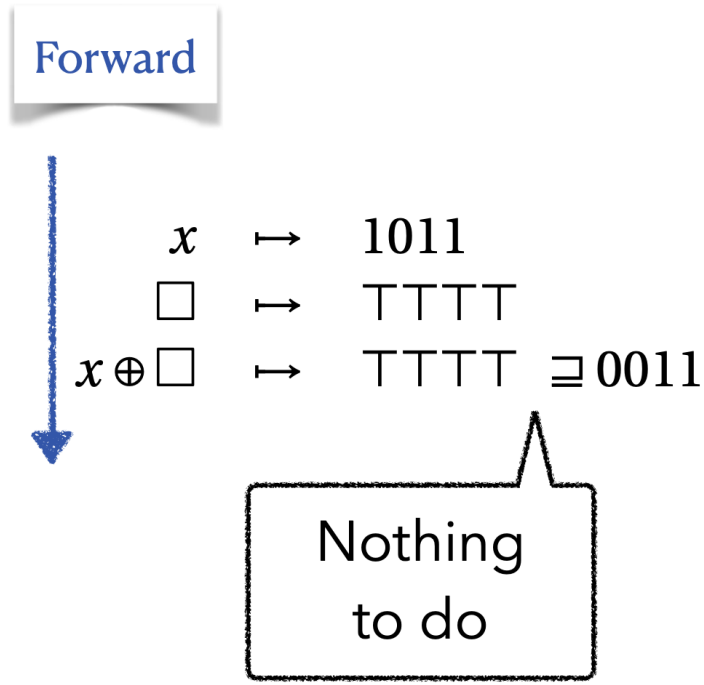
Synthesis Pruning by AI: Forward Analysis Limitation

Candidate partial program

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

for

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



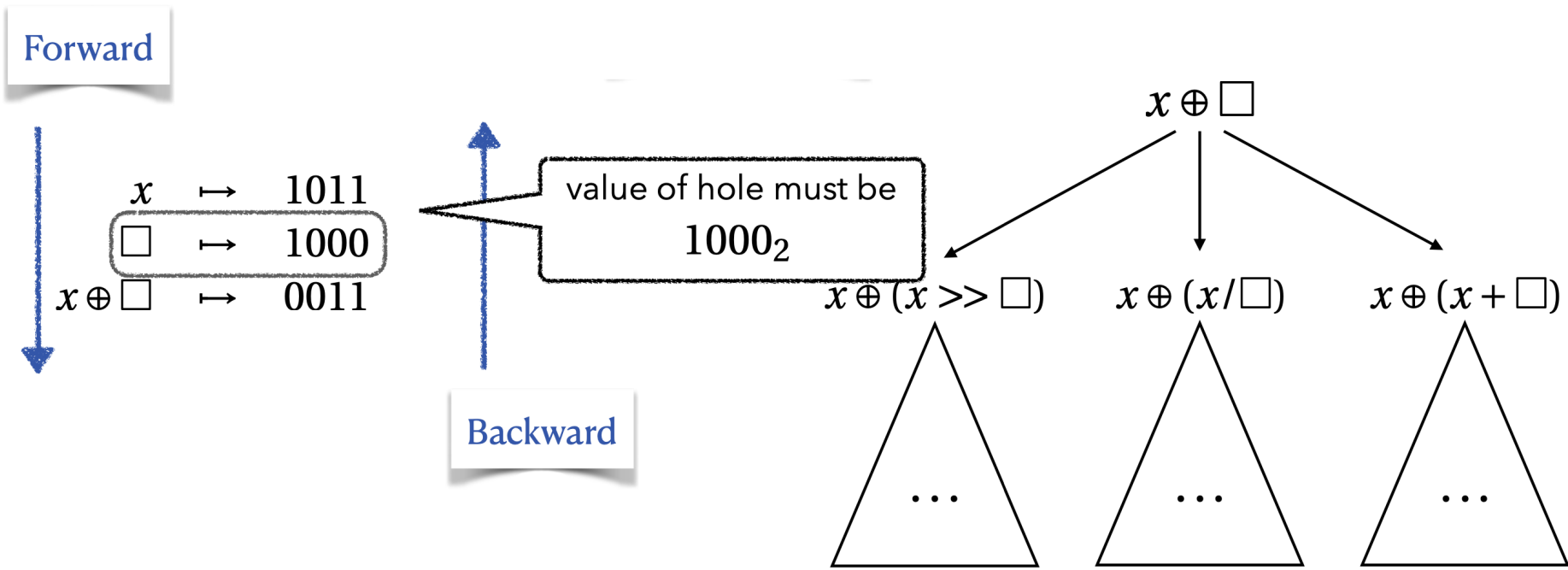
Synthesis Pruning by AI: Backward Analysis

Candidate partial program

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

for

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



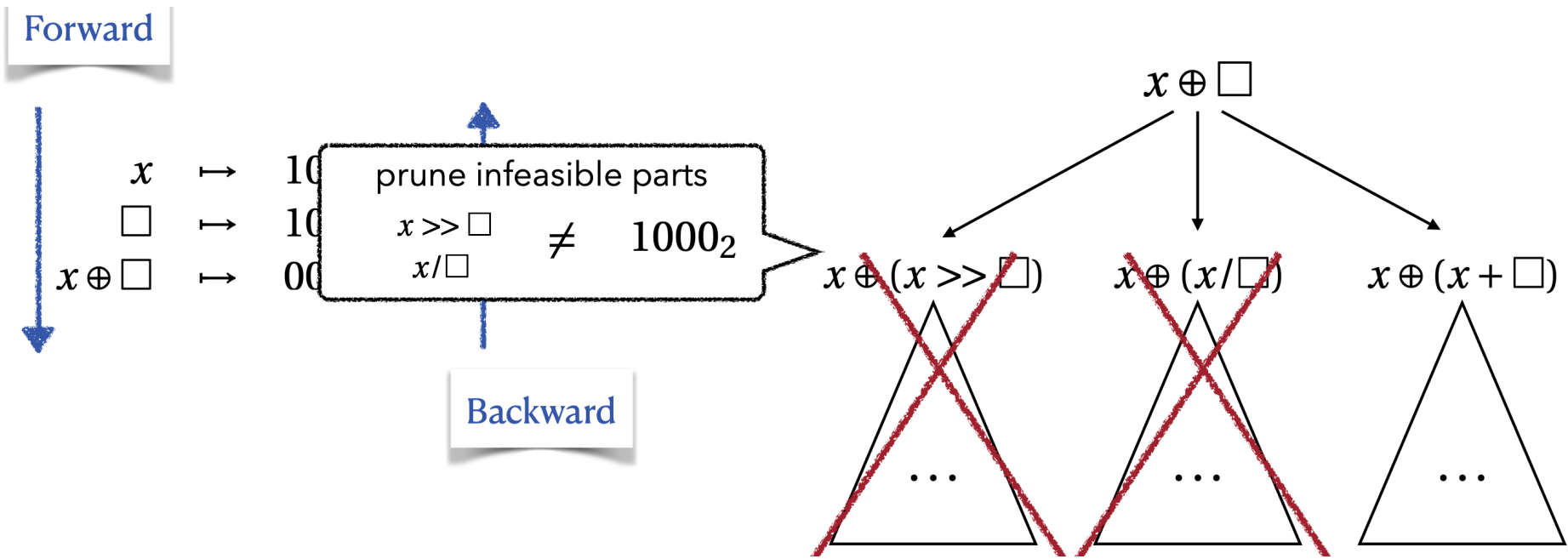
Synthesis Pruning by AI: Backward Analysis

Candidate partial program

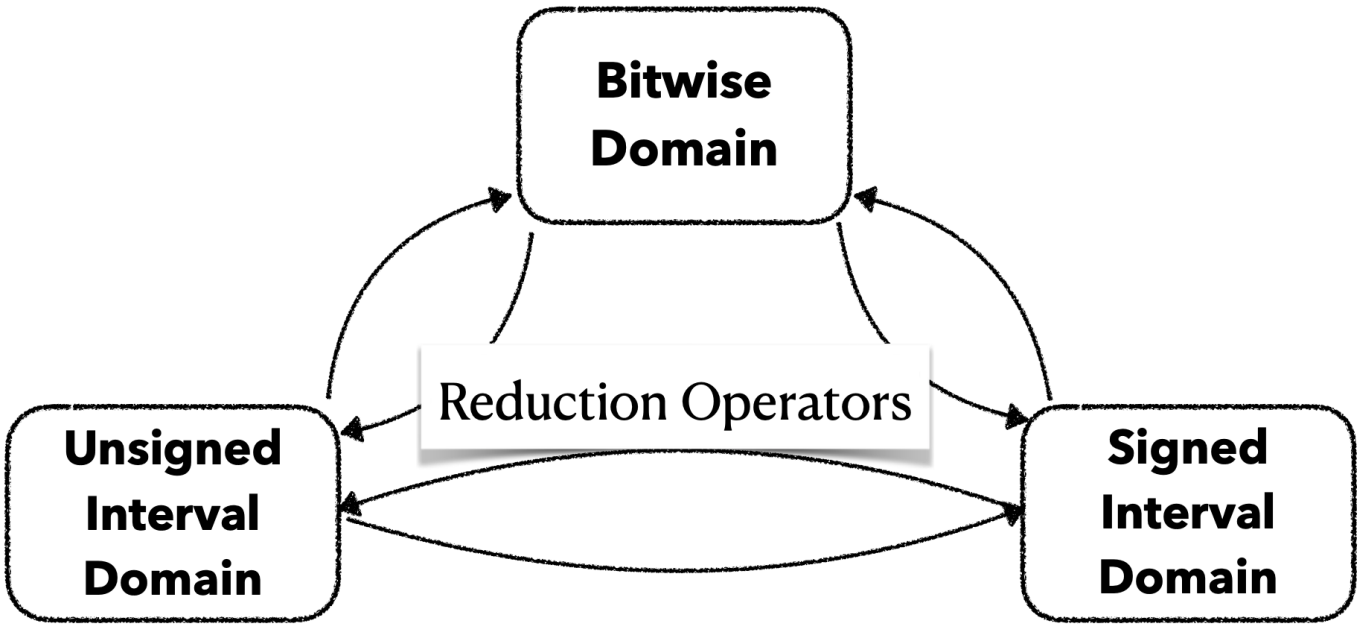
$$f(x) = x \oplus \square$$

for

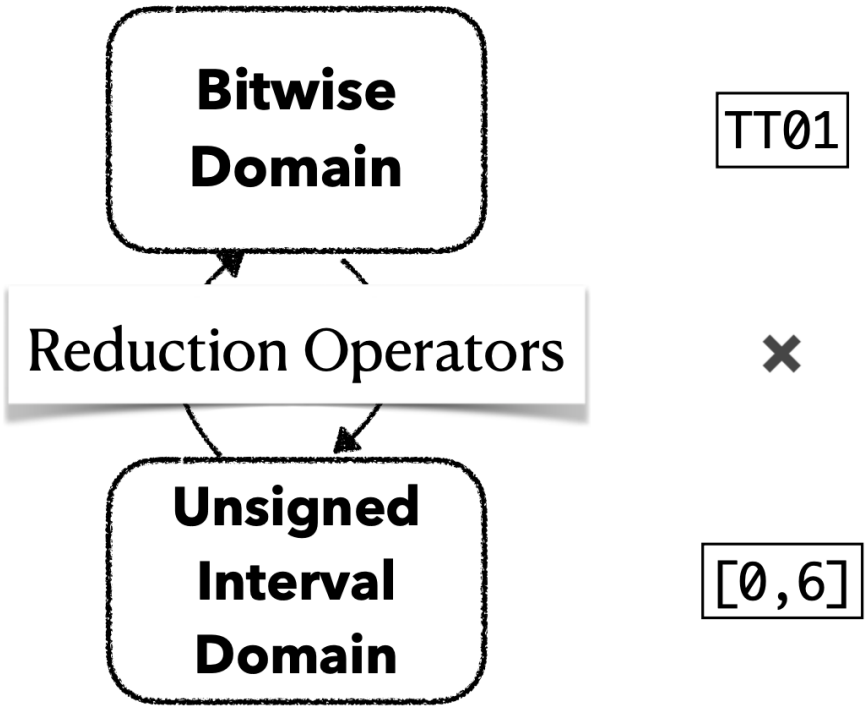
$$f(1011_2) = 0011_2$$



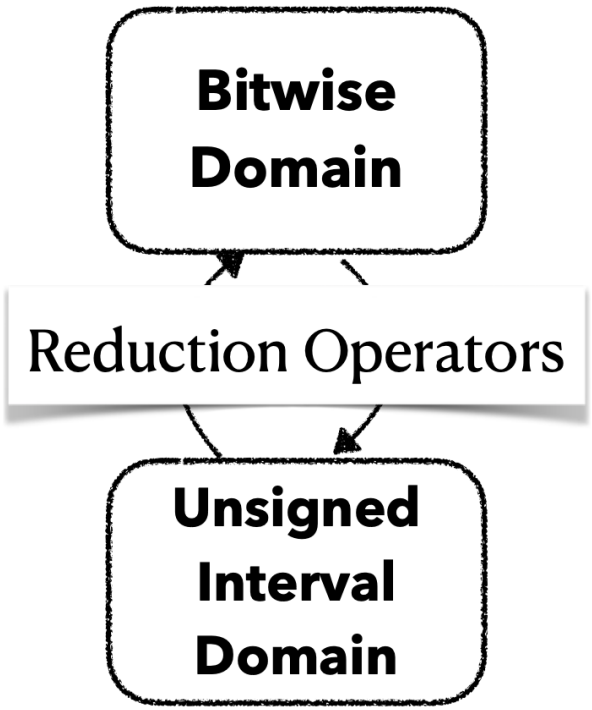
Synthesis Pruning by AI: Reduced Product for Precision & Cost

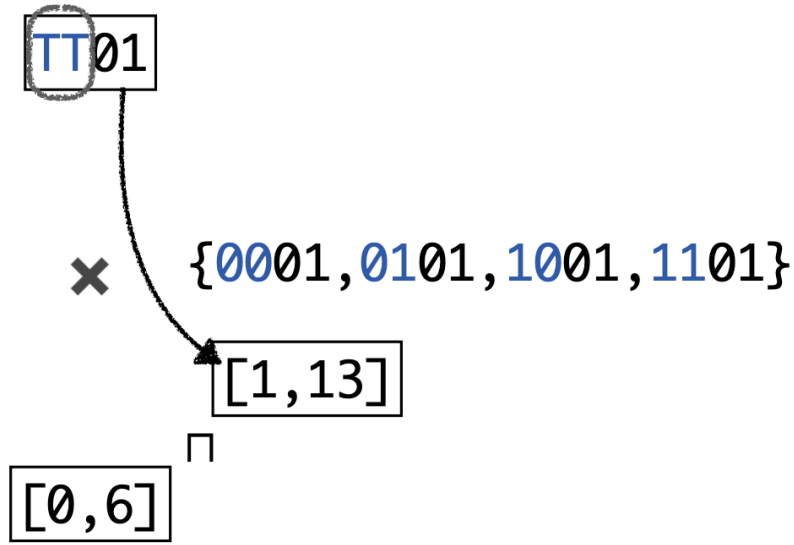


Synthesis Pruning by AI: Reduced Product for Precision & Cost

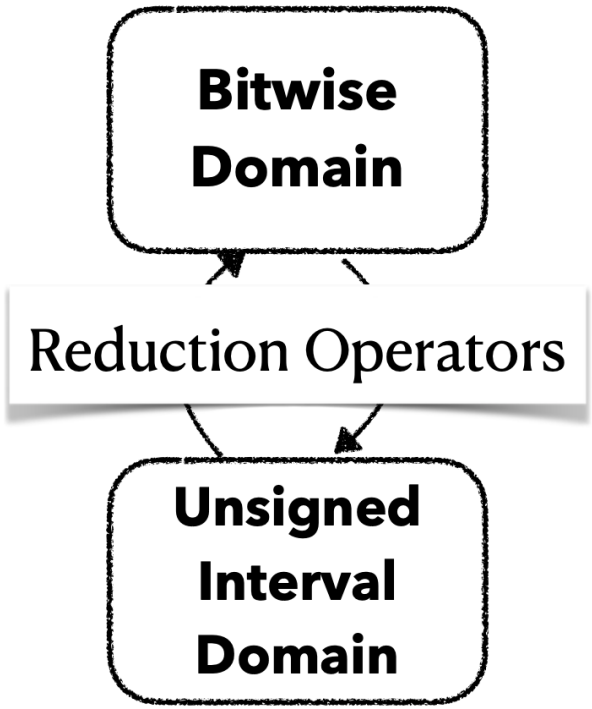


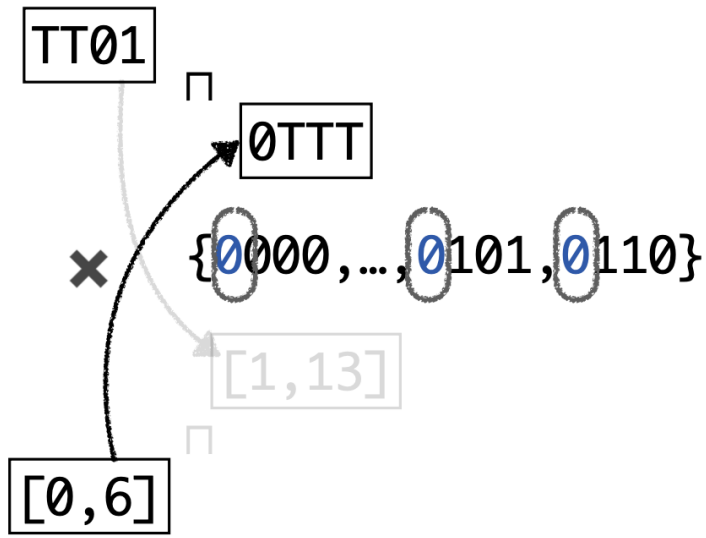
Synthesis Pruning by AI: Reduced Product for Precision & Cost



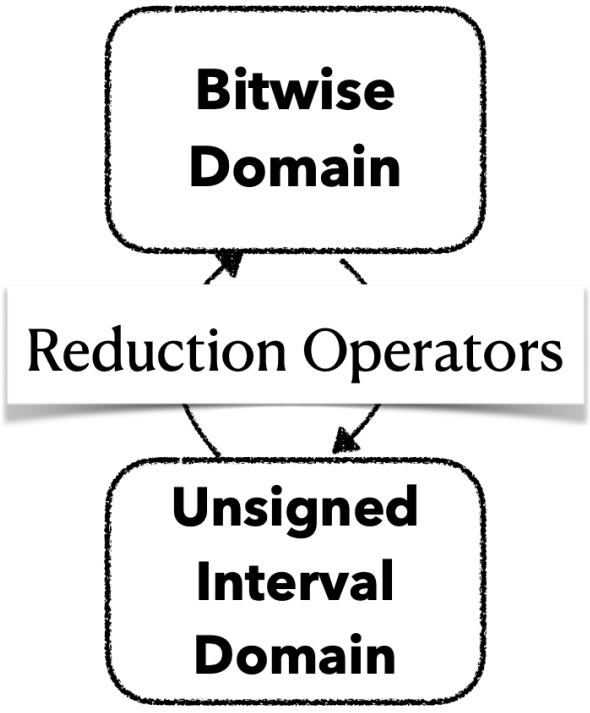


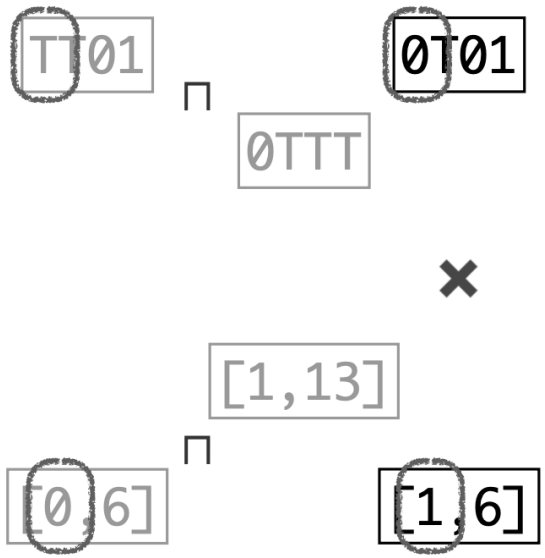
Synthesis Pruning by AI: Reduced Product for Precision & Cost



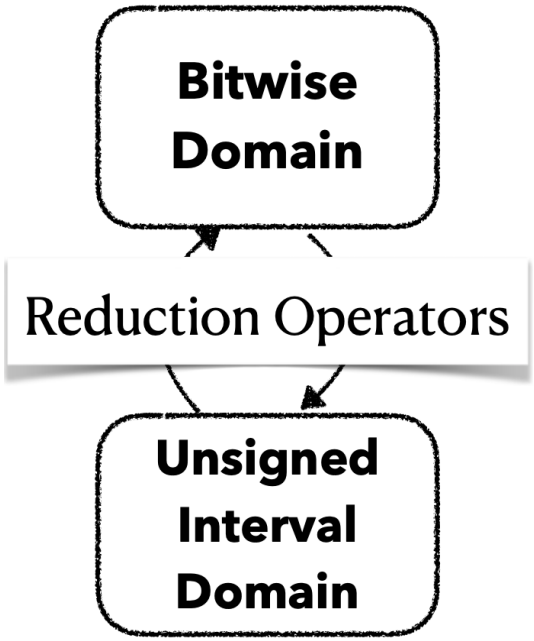


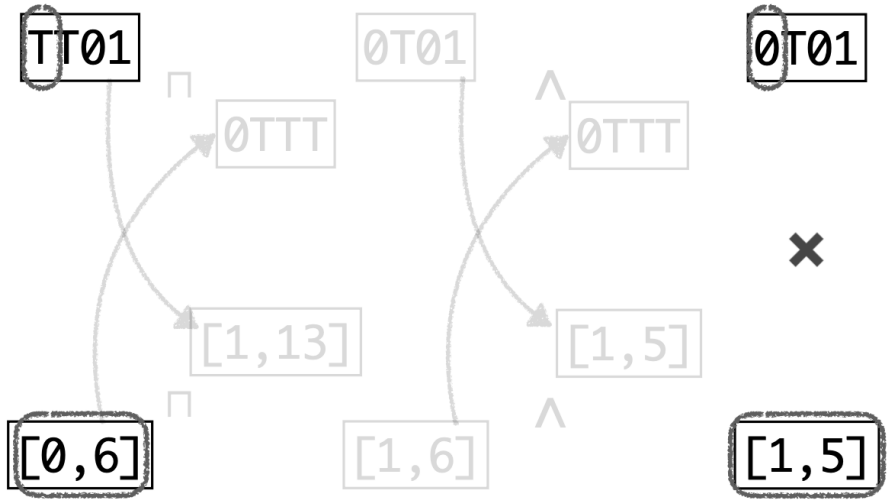
Synthesis Pruning by AI: Reduced Product for Precision & Cost





Synthesis Pruning by AI: Reduced Product for Precision & Cost





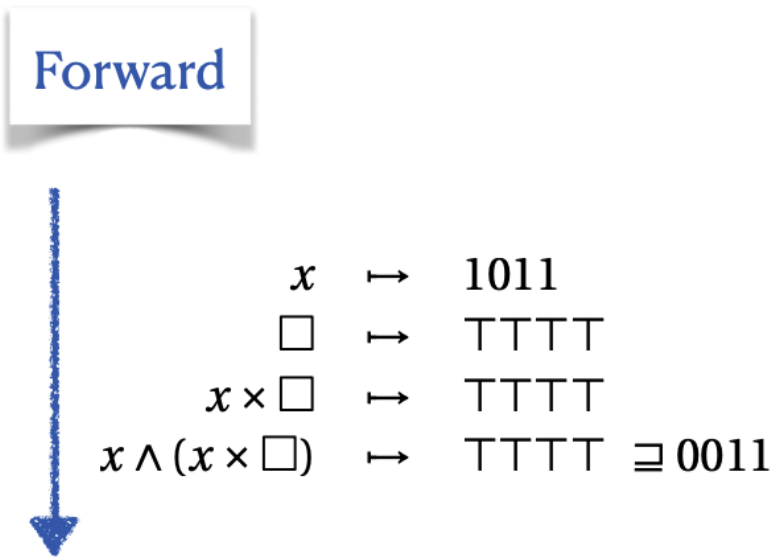
Synthesis Pruning by AI: Finite Refinements for Precision

Candidate partial program

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

for

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



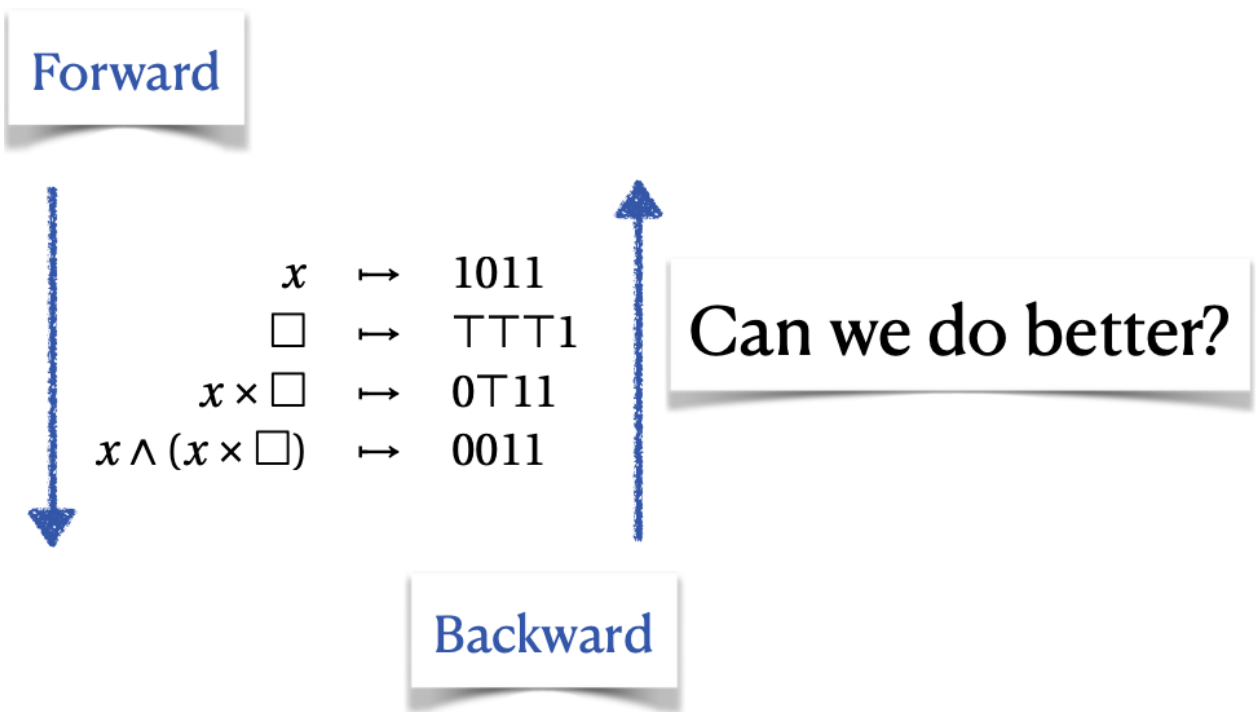
Synthesis Pruning by AI: Finite Refinements for Precision

Candidate partial program

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

for

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



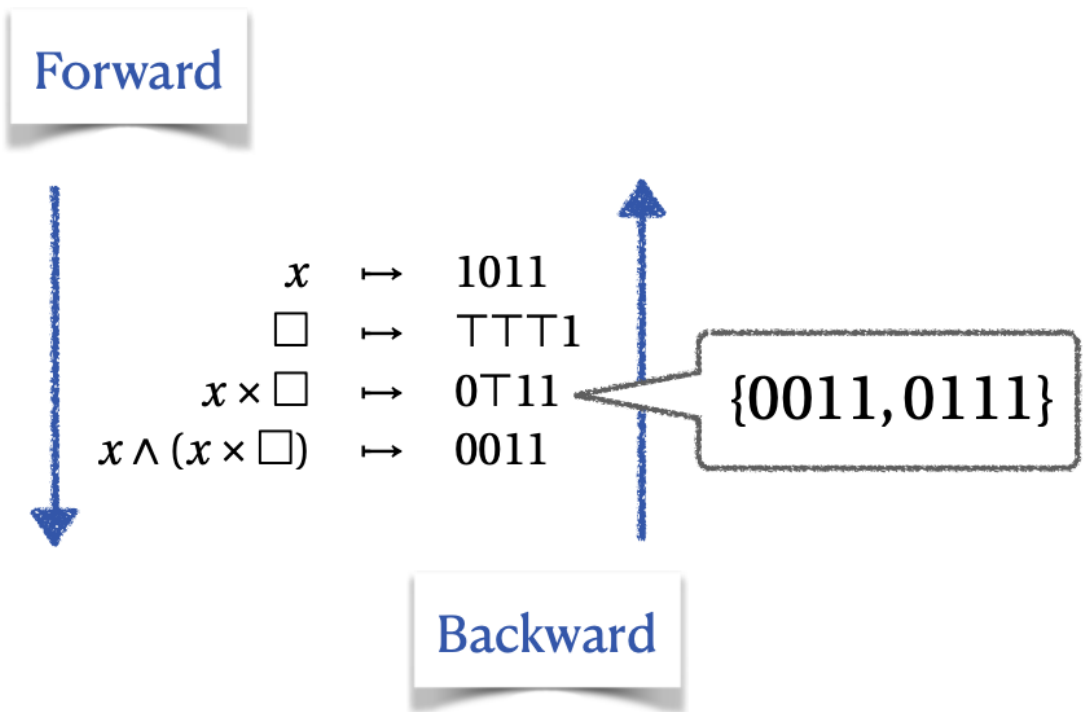
Synthesis Pruning by AI: Finite Refinements for Precision

Candidate partial program

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

for

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



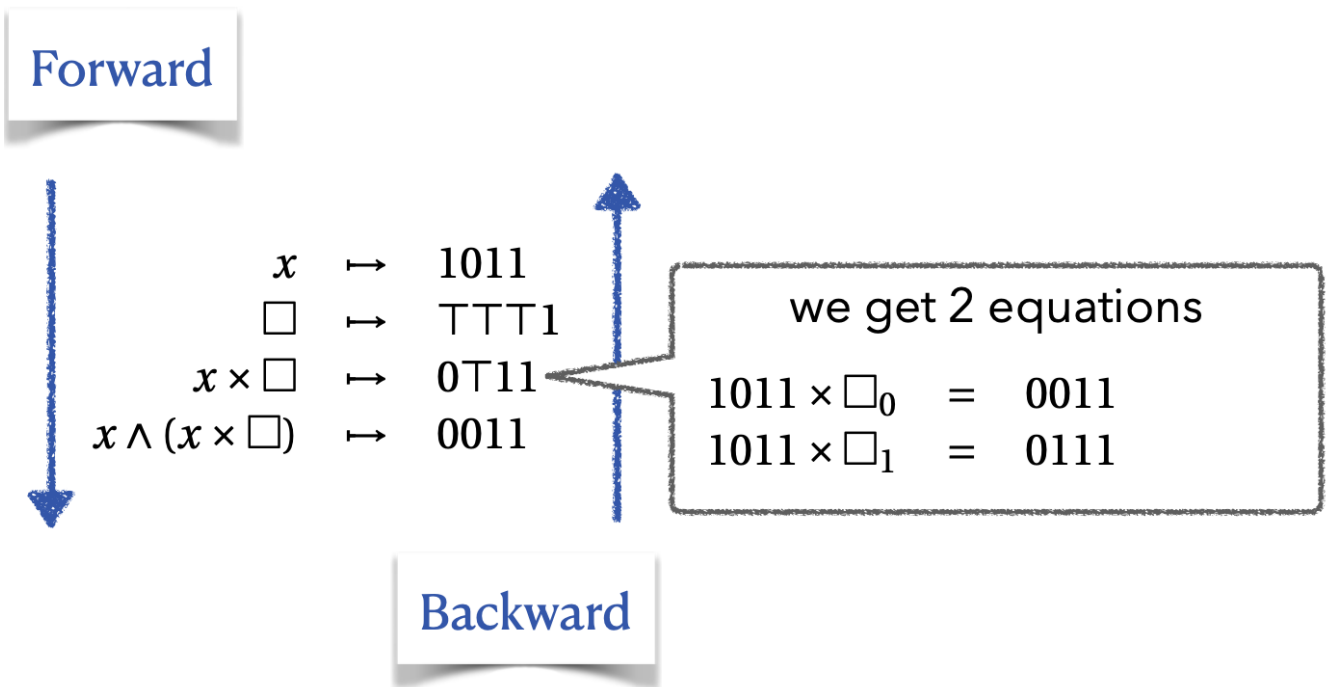
Synthesis Pruning by AI: Finite Refinements for Precision

Candidate partial program

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

for

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



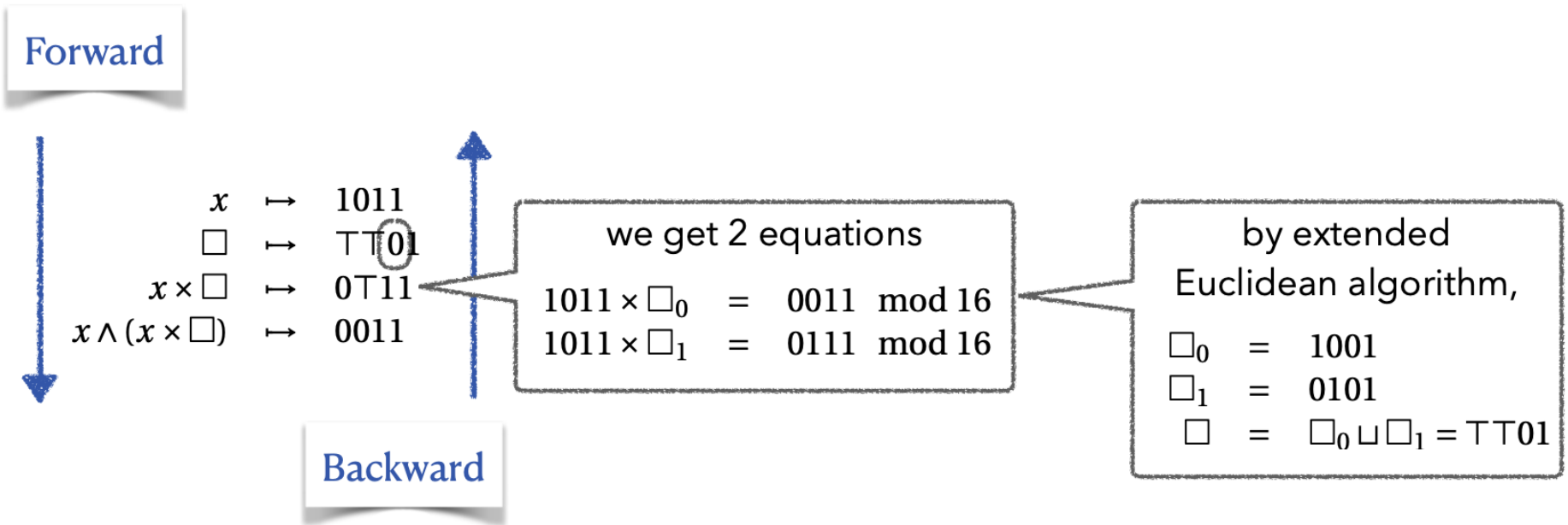
Synthesis Pruning by AI: Finite Refinements for Precision

Candidate partial program

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

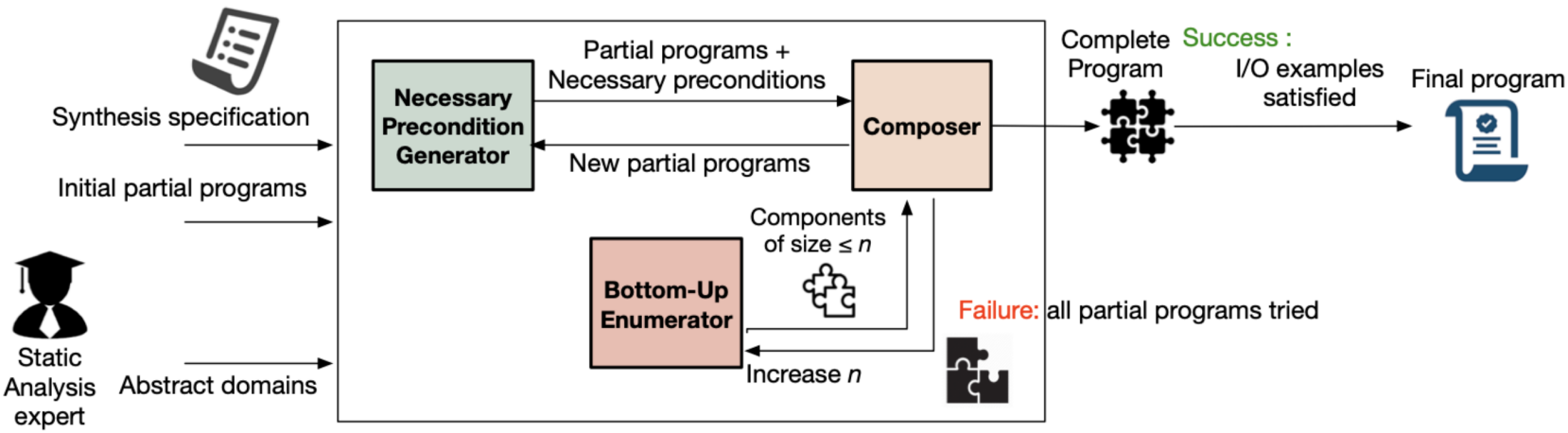
for

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



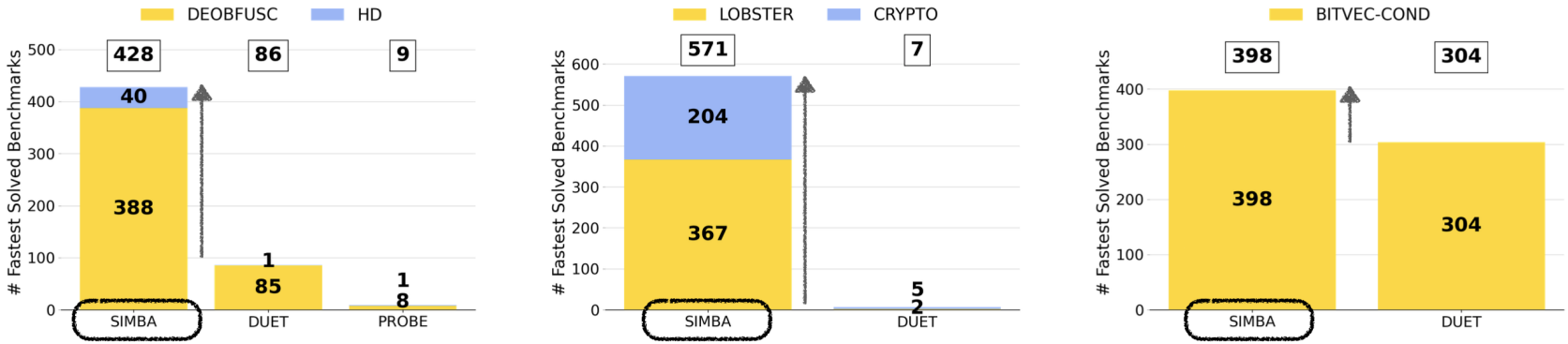
Built a Synthesis System: SIMBA

- **S**ynthesis from **I**nductive spec e**M**powered by **B**idirectional **A**bstract interpretation
- available `https://github.com/yhyoon/simba` **(PLDI'23)**

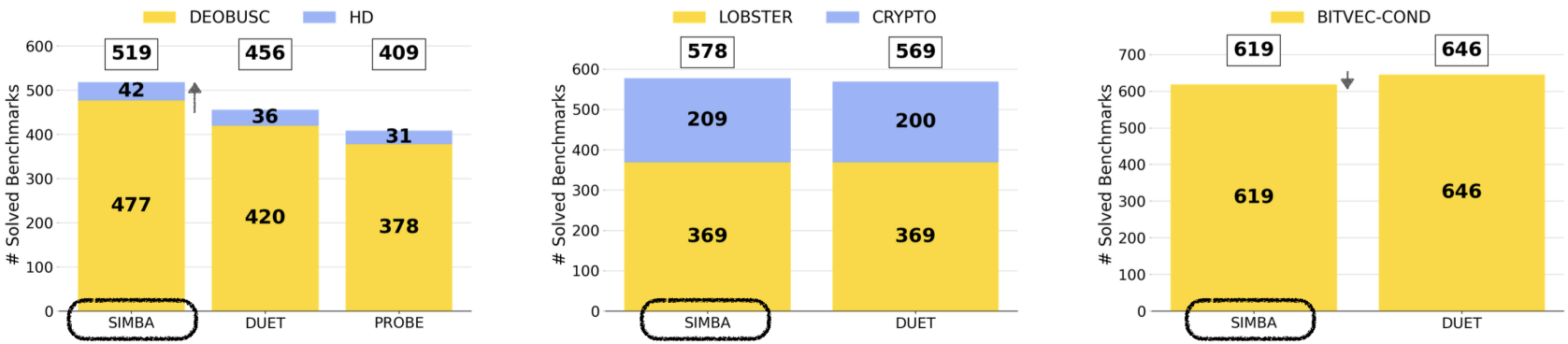


Synthesis Superb Performance

thanks to AI



of the fastest-solved problems for each domain



of the solved problems with 1h timeout for each domain

Our story upon abstract interpretation

- Analysis
  - drive to a "limit": lots of compromises
    - commercialization: SparrowFasoo.com
  - drive to yet another "limit": on sound & global analysis workbench
    - scalability: general sparse analysis framework
    - precision: selective x-sensitivity
- Synthesis
  - superb in performance
  - by simple bi-directional+reduced-product AI with finite refinements

Our story upon abstract interpretation

- Analysis
  - drive to a "limit": lots of compromises
    - commercialization: SparrowFasoo.com
  - drive to yet another "limit": on sound & global analysis workbench
    - scalability: general sparse analysis framework
    - precision: selective x-sensitivity
- Synthesis
  - superb in performance
  - by simple bi-directional+reduced-product AI with finite refinements

---

Without AI framework/concepts/vocabularies, we might have

Our story upon abstract interpretation

- Analysis
  - drive to a "limit": lots of compromises
    - commercialization: SparrowFasoo.com
  - drive to yet another "limit": on sound & global analysis workbench
    - scalability: general sparse analysis framework
    - precision: selective x-sensitivity
- Synthesis
  - superb in performance
  - by simple bi-directional+reduced-product AI with finite refinements

---

Without AI framework/concepts/vocabularies, we might have

- failed to conceive & develop our methods.
- been too scared of the sheer complexties of the reality.

Our story upon abstract interpretation

- Analysis
  - drive to a "limit": lots of compromises
    - commercialization: SparrowFasoo.com
  - drive to yet another "limit": on sound & global analysis workbench
    - scalability: general sparse analysis framework
    - precision: selective x-sensitivity
- Synthesis
  - superb in performance
  - by simple bi-directional+reduced-product AI with finite refinements

---

Without AI framework/concepts/vocabularies, we might have

- failed to conceive & develop our methods.
- been too scared of the sheer complexties of the reality.
- no understanding about what we were doing.
- failed to communicate our methods.

Our story upon abstract interpretation

- Analysis
  - drive to a "limit": lots of compromises
    - commercialization: SparrowFasoo.com
  - drive to yet another "limit": on sound & global analysis workbench
    - scalability: general sparse analysis framework
    - precision: selective x-sensitivity
- Synthesis
  - superb in performance
  - by simple bi-directional+reduced-product AI with finite refinements

---

Without AI framework/concepts/vocabularies, we might have

- failed to conceive & develop our methods.
- been too scared of the sheer complexties of the reality.
- no understanding about what we were doing.
- failed to communicate our methods.

Thank you