Synthesis of Circular Compositional Program Proofs via Abduction

Boyang Li, Işıl Dillig, Tom Dillig (College of William & Mary) Ken McMillan (Microsoft Research) Mooly Sagiv (Tel Aviv University) • Different verification approaches have various strengths and weaknesses





- Different verification approaches have various strengths and weaknesses
- Examples:



- Different verification approaches have various strengths and weaknesses
- Examples:
 - Polyhedra domain is good at inferring linear invariants



- Different verification approaches have various strengths and weaknesses
- Examples:
 - Polyhedra domain is good at inferring linear invariants
 - CEGAR based model checking good at separating paths in programs



- Different verification approaches have various strengths and weaknesses
- Examples:
 - Polyhedra domain is good at inferring linear invariants
 - CEGAR based model checking good at separating paths in programs
 - Interval analysis scales to very large programs



- Different verification approaches have various strengths and weaknesses
- Examples:
 - Polyhedra domain is good at inferring linear invariants
 - CEGAR based model checking good at separating paths in programs
 - Interval analysis scales to very large programs

• . . .



- Different verification approaches have various strengths and weaknesses
- Examples:
 - Polyhedra domain is good at inferring linear invariants
 - CEGAR based model checking good at separating paths in programs
 - Interval analysis scales to very large programs

• . . .

Difficult, if not impossible, to design one approach that is good at everything

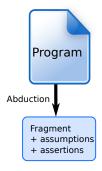
This Talk



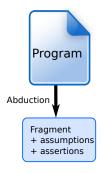
This Talk

New technique for circular compositional program verification

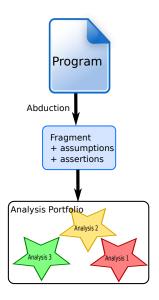
 Decompose the program proofs into small lemmas using logical abduction



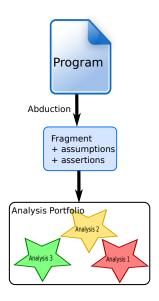
- Decompose the program proofs into small lemmas using logical abduction
- Represent lemmas as code fragments annotated with assertions and assumptions

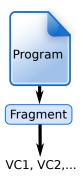


- Decompose the program proofs into small lemmas using logical abduction
- Represent lemmas as code fragments annotated with assertions and assumptions
- Use portfolio of verification techniques to discharge fragments

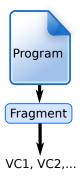


- Decompose the program proofs into small lemmas using logical abduction
- Represent lemmas as code fragments annotated with assertions and assumptions
- Use portfolio of verification techniques to discharge fragments
- Use circular compositional reasoning to turn some assertions into assumptions

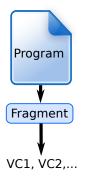




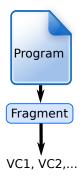
• Compute VCs of assertion on program fragment



- Compute VCs of assertion on program fragment
- For any VC of the form $\phi_1 \Rightarrow \phi_2$ that is not valid, find ψ such that $(\psi \land \phi_1) \Rightarrow \phi_2$ is valid using abduction.



- Compute VCs of assertion on program fragment
- For any VC of the form $\phi_1 \Rightarrow \phi_2$ that is not valid, find ψ such that $(\psi \land \phi_1) \Rightarrow \phi_2$ is valid using abduction.
- Now, introduce ψ as new assertion in program



- Compute VCs of assertion on program fragment
- For any VC of the form $\phi_1 \Rightarrow \phi_2$ that is not valid, find ψ such that $(\psi \land \phi_1) \Rightarrow \phi_2$ is valid using abduction.
- Now, introduce ψ as new assertion in program
- And eliminate old assertion by proving it assuming ψ and converting it to an assumption

• Consider the following code snippet

```
int i=1, j=0;
while(*) {j++; i+=3;}
int z = i - j;
int x=0, y=0, w=0;
while(*) {
  assert(x==y);
  Z += X + Y + W;
  y++;
  x+=z%2;
  w+=2;
```

- Consider the following code snippet
- Code contains assertion in second loop

```
int i=1, j=0;
while(*) {j++; i+=3;}
int z = i - j;
int x=0, y=0, w=0;
while(*) {
  assert(x==y);
  Z + = X + Y + W;
  y++;
  x+=z%2;
  w+=2;
```

- Consider the following code snippet
- Code contains assertion in second loop
- Goal: Discharge assertion using portfolio of analyses on fragments of this code

```
int i=1, j=0;
while(*) { j++; i+=3; }
int z = i - j;
int x=0, y=0, w=0;
while(*) {
  assert(x==y);
  Z += X + Y + W;
  y++;
  x+=z%2;
  w+=2;
```

Decomposition

```
int i=1, j=0;
while(*) {j++; i+=3;}
int z = i - j;
int x=0, y=0, w=0;
while(*) {
  assert(x==y);
  Z += X + Y + W;
  y++;
  x+=z%2;
  w+=2;
}
```

• Want to verify assertion only using highlighted fragment

Decomposition

```
int i=1, j=0;
while(*) { j++; i+=3; }
int z = i - j;
int x=0, y=0, w=0;
while(*) {
  assert(x==y);
  Z += X + Y + W;
  y++;
  x+=z%2;
  w+=2;
}
```

- Want to verify assertion only using highlighted fragment
- But not possible since precondition "z is odd" is missing

Decomposition

```
int i=1, j=0;
while(*) { j++; i+=3; }
int z = i - j;
int x=0, y=0, w=0;
while(*) {
  assert(x==y);
  z+=x+y+w;
  y++;
  x+=z%2;
  w+=2;
}
```

- Want to verify assertion only using highlighted fragment
- But not possible since precondition "z is odd" is missing

Want to solve for missing assumptions required to prove x = y

```
int i=1, j=0;
while(*) { j++; i+=3; }
int z = i - j;
int x=0, y=0, w=0;
assume(\phi_1);
while(*) {
  assert(x==y);
  assume(\phi_2);
  z + = x + y + w;
  y++;
  x+=z%2:
  w+=2;
```

 Use φ₁ and φ₂ to represent unknown assumptions that make the assertion valid

```
int i=1, j=0;
while(*) { j++; i+=3; }
int z = i - j;
int x=0, y=0, w=0;
assume(\phi_1);
while(*) {
  assert(x==y);
  assume(\phi_2);
  Z += X + V + W;
  y++;
  x+=z%2:
  w+=2;
```

- Use φ₁ and φ₂ to represent unknown assumptions that make the assertion valid
- Compute VCs of x = y parametric on φ₁ and φ₂

```
int i=1, j=0;
while(*) { j++; i+=3; }
int z = i - j;
int x=0, y=0, w=0;
assume(\phi_1);
while(*) {
  assert(x==y);
  assume(\phi_2);
  Z += X + V + W;
  y++;
  x+=z%2:
  w+=2;
```

- Use φ₁ and φ₂ to represent unknown assumptions that make the assertion valid
- Compute VCs of x = yparametric on ϕ_1 and ϕ_2

• VC 1:

$$\begin{aligned} (z &= i - j \land x = 0 \land y = 0 \\ \land w &= 0 \land \phi_1) \Rightarrow x = y \end{aligned}$$

```
int i=1, j=0;
while(*) { j++; i+=3; }
int z = i - j;
int x=0, y=0, w=0;
assume(\phi_1);
while(*) {
  assert(x==y);
  assume(\phi_2);
  Z += X + V + W;
  y++;
  x+=z%2:
  w+=2;
```

- Use \$\phi_1\$ and \$\phi_2\$ to represent unknown assumptions that make the assertion valid
- Compute VCs of x = yparametric on ϕ_1 and ϕ_2

$$\begin{aligned} (z &= i - j \wedge x = 0 \wedge y = 0 \\ \wedge w &= 0 \wedge \phi_1) \Rightarrow x = y \end{aligned}$$

• VC 2:

$$(\phi_2 \land x = y) \Rightarrow wp(\sigma, x = y)$$

```
int i=1, j=0;
while(*) { j++; i+=3; }
int z = i - j;
int x=0, y=0, w=0;
assume(\phi_1);
while(*) {
  assert(x==y);
  assume(\phi_2);
  Z += X + V + W;
  y++;
  x+=z%2:
  w+=2;
```

- Use \$\phi_1\$ and \$\phi_2\$ to represent unknown assumptions that make the assertion valid
- Compute VCs of x = yparametric on ϕ_1 and ϕ_2
- VC 1: VALID

$$\begin{aligned} (z &= i - j \wedge x = 0 \wedge y = 0 \\ \wedge w &= 0 \wedge \phi_1) \Rightarrow x = y \end{aligned}$$

• VC 2:

$$(\phi_2 \land x = y) \Rightarrow wp(\sigma, x = y)$$

```
int i=1, j=0:
while(*) { j++; i+=3; }
int z = i - j;
int x=0, y=0, w=0;
assume(\phi_1);
while(*) {
  assert(x==y);
  assume(\phi_2);
  z + = x + y + w;
  y++;
  x+=z%2:
  w+=2;
```

- Use φ₁ and φ₂ to represent unknown assumptions that make the assertion valid
- Compute VCs of x = yparametric on ϕ_1 and ϕ_2
- VC 1: VALID
 - $\begin{aligned} (z = i j \wedge x = 0 \wedge y = 0 \\ \wedge w = 0 \wedge \phi_1) \Rightarrow x = y \end{aligned}$
- VC 2: NOT VALID

 $(\phi_2 \land x = y) \Rightarrow wp(\sigma, x = y)$

 $(\phi_2 \land x = y) \Rightarrow x + (z + x + y + w)\%2 = y + 1$

 $(\phi_2 \land x = y) \Rightarrow x + (z + x + y + w)\%2 = y + 1$

• To prove VC 2, we need to find a ϕ_2 that makes it valid

 $(\phi_2 \land x = y) \Rightarrow x + (z + x + y + w)\%2 = y + 1$

- To prove VC 2, we need to find a ϕ_2 that makes it valid
- But ϕ_2 should not contradict x = y (lemma we want to prove)

 $(\phi_2 \land x = y) \Rightarrow x + (z + x + y + w)\%2 = y + 1$

- To prove VC 2, we need to find a ϕ_2 that makes it valid
- But ϕ_2 should not contradict x = y (lemma we want to prove)
- Therefore, want $\phi_2 \wedge x = y$ to be satisfiable

 $(\phi_2 \land x = y) \Rightarrow x + (z + x + y + w)\%2 = y + 1$

- To prove VC 2, we need to find a ϕ_2 that makes it valid
- But ϕ_2 should not contradict x = y (lemma we want to prove)
- Therefore, want $\phi_2 \wedge x = y$ to be satisfiable

Insight: This is an instance of logical abduction

Abductive Inference



• Given known facts F and desired outcome O, abductive inference finds simple explanatory hypothesis E such that

 $F \wedge E \models O$ and $SAT(F \wedge E)$

9/22

Abductive Inference



abduce

• Given known facts F and desired outcome O, abductive inference finds simple explanatory hypothesis E such that

 $F \wedge E \models O$ and $SAT(F \wedge E)$

• Use abduction to generate simple assumptions that make verification condition valid

Abductive Inference



abduce

• Given known facts F and desired outcome O, abductive inference finds simple explanatory hypothesis E such that

 $F \wedge E \models O$ and $SAT(F \wedge E)$

- Use abduction to generate simple assumptions that make verification condition valid
- Known facts F is verification condition, desired outcome is *true*

Abductive Inference



abduce

• Given known facts F and desired outcome O, abductive inference finds simple explanatory hypothesis E such that

 $F \wedge E \models O$ and $SAT(F \wedge E)$

- Use abduction to generate simple assumptions that make verification condition valid
- Known facts F is verification condition, desired outcome is *true*
- Abductive solution becomes lemma in proof and can now be established separately

• Here, for

$$(\phi_2 \wedge x = y) \Rightarrow x + (z + x + y + w)\%2 = y + 1$$

we compute the solution $\phi_2: (w+z)\% 2 = 1$

int i=1, j=0; while(*) { j++; i+=3; } int z = i - j;int x=0, y=0, w=0; assume(ϕ_1); while(*) { assert(x==y); assume(ϕ_2); z + = x + y + w;σ V++; x+=z%2: w+=2;

• Here, for

$$(\phi_2 \wedge x = y) \Rightarrow x + (z + x + y + w)\%2 = y + 1$$

we compute the solution $\phi_2: (w+z)\% 2 = 1$

```
int i=1, j=0:
while(*) {j++; i+=3;}
int z = i - j;
int x=0, y=0, w=0;
while(*) {
  assert(x==y);
  assume((w+z))(2==1);
  Z += X + Y + W;
  V++;
  x+=z%2:
  w+=2;
```

• Here, for

$$(\phi_2 \wedge x = y) \Rightarrow x + (z + x + y + w)\%2 = y + 1$$

we compute the solution $\phi_2: (w+z)\% 2 = 1$

• Can now show x = y, which turns into an assumption

```
int i=1, j=0;
while(*) { j++; i+=3; }
int z = i - j;
int x=0, y=0, w=0;
while(*) {
  assert(x==y);
  assume((w+z)%2==1);
  Z += X + Y + W;
  V++;
  x+=z%2:
  w+=2;
```

• Here, for

$$(\phi_2 \wedge x = y) \Rightarrow x + (z + x + y + w)\%2 = y + 1$$

we compute the solution $\phi_2: (w+z)\% 2 = 1$

• Can now show x = y, which turns into an assumption

```
int i=1, j=0;
while(*) { j++; i+=3; }
int z = i - j;
int x=0, y=0, w=0;
while(*) {
  assume(x==y);
  assume((w+z)%2==1);
  Z += X + Y + W;
  V++;
  x+=z%2:
  w+=2;
```

• Here, for

$$(\phi_2 \wedge x = y) \Rightarrow x + (z + x + y + w)\%2 = y + 1$$

we compute the solution $\phi_2: (w+z)\% 2 = 1$

- Can now show x = y, which turns into an assumption
- But still need to prove $\phi_2 \Rightarrow$ add as assertion

```
int i=1, j=0;
while(*) {j++; i+=3;}
int z = i - i:
int x=0, y=0, w=0;
while(*) {
  assume(x==y);
  assert((w+z))
  Z + = X + V + W;
  V++;
  x+=z%2:
  w+=2;
```

• Here, for

$$(\phi_2 \wedge x = y) \Rightarrow x + (z + x + y + w)\%2 = y + 1$$

we compute the solution $\phi_2: (w+z)\% 2 = 1$

- Can now show x = y, which turns into an assumption
- But still need to prove $\phi_2 \Rightarrow$ add as assertion
- Circular compositional reasoning at work!

```
int i=1, j=0;
while(*) { j++; i+=3; }
int z = i - j;
int x=0, y=0, w=0;
while(*) {
  assume(x==y);
  assert((w+z))
  Z + = X + V + W;
  V++;
  x+=z%2:
  w+=2;
```

```
int i=1, j=0;
while(*) { j++; i+=3; }
int z = i - j;
int x=0, y=0, w=0;
while(*) {
  assume(x==y);
  assert((w+z)%2==1);
  Z += X + Y + W;
  y++;
  x+=z%2;
  w+=2;
```

• New assertion still not provable since value of *z* unconstrained

```
int i=1, j=0;
while(*) { j++; i+=3; }
int z = i - j;
int x=0, y=0, w=0;
while(*) {
  assume(x==y);
  assert((w+z)%2==1);
  Z += X + Y + W;
  y++;
  x+=z%2;
  w+=2;
```

- New assertion still not provable since value of *z* unconstrained
- Again generate parametric VCs

```
int i=1, j=0;
while(*) { j++; i+=3; }
int z = i - j;
int x=0, y=0, w=0;
assert(\phi_1);
while(*) {
  assume(x==y);
  assert((w+z)%2==1);
  Z += X + Y + W;
  y++;
  x+=z%2;
  w+=2;
```

- New assertion still not provable since value of *z* unconstrained
- Again generate parametric VCs
- First VC introduces assertion before loop

```
int i=1, j=0;
while(*) { j++; i+=3; }
int z = i - j;
int x=0, y=0, w=0;
assert(\phi_1);
while(*) {
  assume(x==y);
  assert((w+z)%2==1);
  Z += X + Y + W;
  y++;
  x+=z%2;
  w+=2;
```

- New assertion still not provable since value of *z* unconstrained
- Again generate parametric VCs
- First VC introduces assertion before loop

```
int i=1, j=0:
while(*) { j++; i+=3; }
int z = i - j:
int x=0, y=0, w=0;
assert(\phi_1);
while(*) {
  assume(x==y);
  assert((w+z)%2==1);
  Z += X + Y + W;
  V++;
  x+=z%2;
  w+=2;
```

- New assertion still not provable since value of *z* unconstrained
- Again generate parametric VCs
- First VC introduces assertion before loop

• Solution computed via abduction $\phi_1: z\%2 = 1$

```
int i=1, j=0;
while(*) { j++; i+=3; }
int z = i - j;
int x=0, y=0, w=0;
assert(z%2==1);
while(*) {
  assume(x==y);
  assert((w+z)%2==1);
  Z += X + Y + W;
  y++;
  x+=z%2;
  w+=2;
```

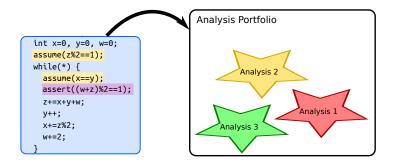
• Now left with two assertions in the code fragment

```
int i=1, j=0;
while(*) { j++; i+=3; }
int z = i - j;
int x=0, y=0, w=0;
assert(z%2==1);
while(*) {
  assume(x==y);
  assert((w+z)%2==1);
  Z += X + Y + W;
  V++;
  x+=z%2;
  w+=2;
```

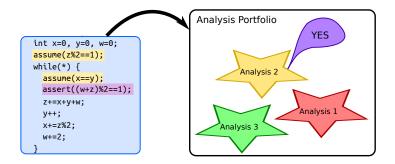
- Now left with two assertions in the code fragment
- Convert first assertion to assumption and invoke our client analyses

```
int i=1, j=0;
while(*) { j++; i+=3; }
int z = i - j;
int x=0, y=0, w=0;
assert(z%2==1);
while(*) {
  assume(x==y);
  assert((w+z)%2==1);
  Z += X + Y + W;
  V++;
  x+=z%2;
  w+=2;
```

- Now left with two assertions in the code fragment
- Convert first assertion to assumption and invoke our client analyses
- Again, circular compositional reasoning at work



• Give fragment with assumptions and assertions to clients



- Give fragment with assumptions and assertions to clients
- Fragment can be locally verified by divisibility analysis

Remaining Code

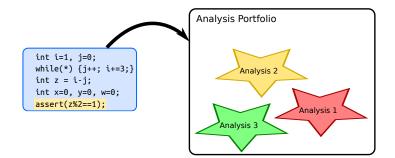
```
int i=1, j=0;
while(*) {j++; i+=3;}
int z = i - j;
int x=0, y=0, w=0;
assert(z%2==1);
while(*) {
  assume(x==y);
  assume((w+z)%2==1);
  Z += X + Y + W;
  y++;
  x+=z%2;
  w+=2;
```

• Now, only one assertion left in our program

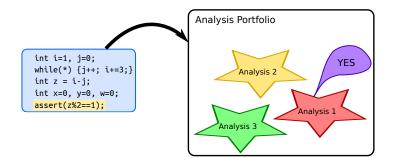
Remaining Code

```
int i=1, j=0;
while(*) {j++; i+=3;}
int z = i - j;
int x=0, y=0, w=0;
assert(z%2==1);
while(*) {
  assume(x==v);
  assume((w+z)%2==1);
  Z += X + V + W;
  V++;
  x+=z%2;
  w+=2:
```

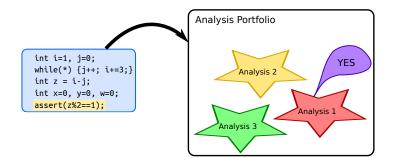
- Now, only one assertion left in our program
- Extract next fragment for this assertion and give to client analyses



Invoke clients on current fragment



- Invoke clients on current fragment
- This assertion can be shown by any client analysis that can establish i=3j+1



- Invoke clients on current fragment
- This assertion can be shown by any client analysis that can establish i = 3j + 1

We have now proven the original assertion



• Technique decomposes proof of program into subgoals on syntactic fragments

Our Technique at a High Level



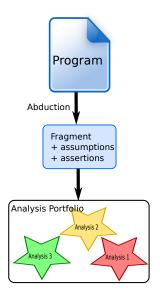
• Technique decomposes proof of program into subgoals on syntactic fragments

 While we show one subgoal, we can safely assume the others ⇒ circular reasoning

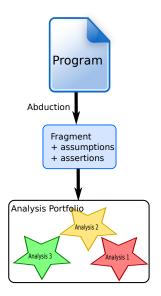
Our Technique at a High Level



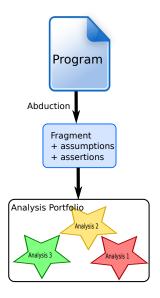
- Technique decomposes proof of program into subgoals on syntactic fragments
- While we show one subgoal, we can safely assume the others ⇒ circular reasoning
- If a subgoal cannot be shown by any client analysis, we backtrack and generate new subgoals using abductive inference



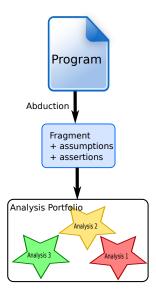
• Clients only analyze (typically) small fragments



- Clients only analyze (typically) small fragments
- Interaction is demand-driven (i.e., lazy)

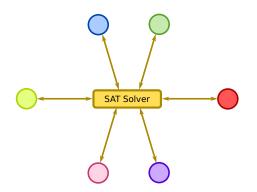


- Clients only analyze (typically) small fragments
- Interaction is demand-driven (i.e., lazy)
- Analyses with complementary strengths can help each other



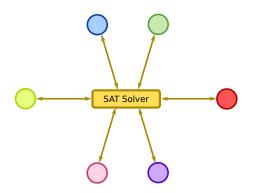
- Clients only analyze (typically) small fragments
- Interaction is demand-driven (i.e., lazy)
- Analyses with complementary strengths can help each other
- Can prove properties no client analysis or eager combination can prove alone

Analogy to SMT Solver



• Can view client analyses as theory solvers

Analogy to SMT Solver



- Can view client analyses as theory solvers
- Proving assertion on program invokes clients on fragments and speculated subgoals, backtracks when needed

• Implemented this techique and used four client tools:

- Implemented this techique and used four client tools:
 - Interproc Polyhedra

- Implemented this techique and used four client tools:
 - Interproc Polyhedra
 - Interproc Linear Congruence

- Implemented this techique and used four client tools:
 - Interproc Polyhedra
 - Interproc Linear Congruence
 - Blast

- Implemented this techique and used four client tools:
 - Interproc Polyhedra
 - Interproc Linear Congruence
 - Blast
 - Compass

• Implemented this techique and used four client tools:

- Interproc Polyhedra
- Interproc Linear Congruence
- Blast
- Compass

These tools have very different strengths and weaknesses

• 10 challenging micro-benchmarks with one assertion each

- 10 challenging micro-benchmarks with one assertion each
- No tool can individually prove any benchmark

- 10 challenging micro-benchmarks with one assertion each
- No tool can individually prove any benchmark

Name	LOC	Time (s)	# queries	Polyhedra	Linear Cong	Blast	Compass
B1	45	0.6	2	×	×	~	×
B2	37	0.2	2	×	~	×	X
B3	51	1.0	2	~	×	~	×
B4	59	0.4	3	~	×	~	X
B5	89	0.6	3	~	×	~	×
B6	60	0.5	5	×	~	X	~
B7	56	0.6	2	×	×	~	~
B8	45	0.2	2	~	×	~	×
B9	59	0.5	1	×	×	~	×
B10	47	0.2	2	~	×	~	~

- 10 challenging micro-benchmarks with one assertion each
- No tool can individually prove any benchmark

Name	LOC	Time (s)	# queries	Polyhedra	Linear Cong	Blast	Compass
B1	45	0.6	2	×	×	~	×
B2	37	0.2	2	×	~	X	×
B3	51	1.0	2	~	×	~	×
B4	59	0.4	3	~	×	~	X
B5	89	0.6	3	~	×	~	×
B6	60	0.5	5	×	~	X	~
B7	56	0.6	2	×	×	~	~
B8	45	0.2	2	~	×	~	×
B9	59	0.5	1	×	×	~	×
B10	47	0.2	2	~	×	~	~

But all benchmarks can be proven when analyses are combined using our technique

We also verified a complicated assertion each on five real programs

We also verified a complicated assertion each on five real programs

Name	LOC	Time (s)	# queries	Avg $\#$ vars in query	Avg LOC in query
Wizardpen Linux Driver	1242	3.8	5	1.5	29
OpenSSH clientloop	1987	2.8	3	2.3	5
Coreutils su	1057	3.0	5	1.7	6
GSL Histogram	526	0.6	4	3.6	15
GSL Matrix	7233	16.9	8	1.8	7

We also verified a complicated assertion each on five real programs

Name	LOC	Time (s)	# queries	Avg $\#$ vars in query	Avg LOC in query
Wizardpen Linux Driver	1242	3.8	5	1.5	29
OpenSSH clientloop	1987	2.8	3	2.3	5
Coreutils su	1057	3.0	5	1.7	6
GSL Histogram	526	0.6	4	3.6	15
GSL Matrix	7233	16.9	8	1.8	7

Fragments extracted for queries small in practice

 New technique to decompose program's correctness proof into small lemmas

- New technique to decompose program's correctness proof into small lemmas
- Allows a portfolio of diverse analyses with different strengths to cooperate

- New technique to decompose program's correctness proof into small lemmas
- Allows a portfolio of diverse analyses with different strengths to cooperate
- Interaction is goal-directed and effective in practice

- New technique to decompose program's correctness proof into small lemmas
- Allows a portfolio of diverse analyses with different strengths to cooperate
- Interaction is goal-directed and effective in practice

