



Advancing Software Analysis via Changed Perspectives

Zhendong Su

ETH Zurich



Mehrdad Afshari Zhoulai Fu

Vu Le Chengnian Sun Qirun Zhang













What is research impact?



Concept

Technique

Insight

Tool

Two instances

□Validate production compilers

Black-box analysis

Analyze floating-point software

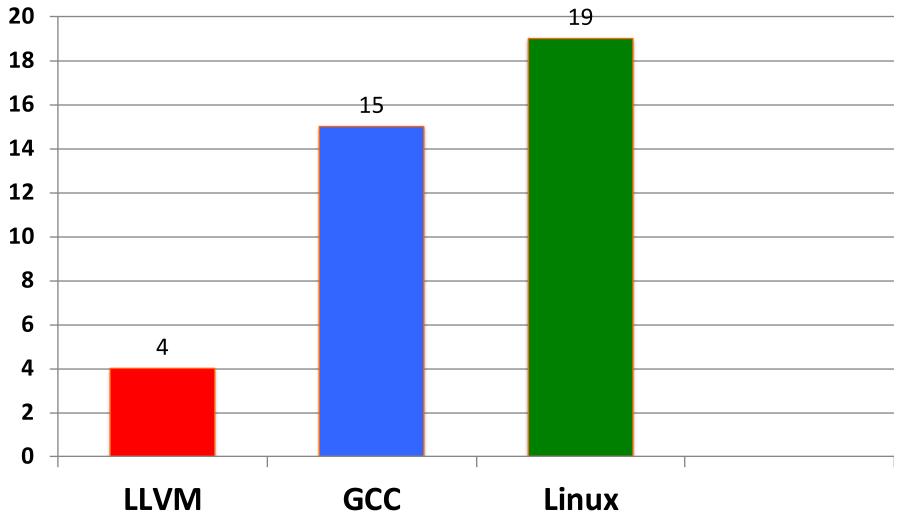
Dynamic analysis

Validate Production Compilers



Compiler complexity

LoC (million)



LLVM bug 14972

```
struct tiny { char c; char d; char e; };
```

```
void foo(struct tiny x) {
    if (x.c != 1) abort();
    if (x.e != 1) abort();
}
```

```
int main() {
    struct tiny s;
    s.c = 1; s.d = 1; s.e = 1;
    foo(s);
    return 0;
}
```

Developer comment

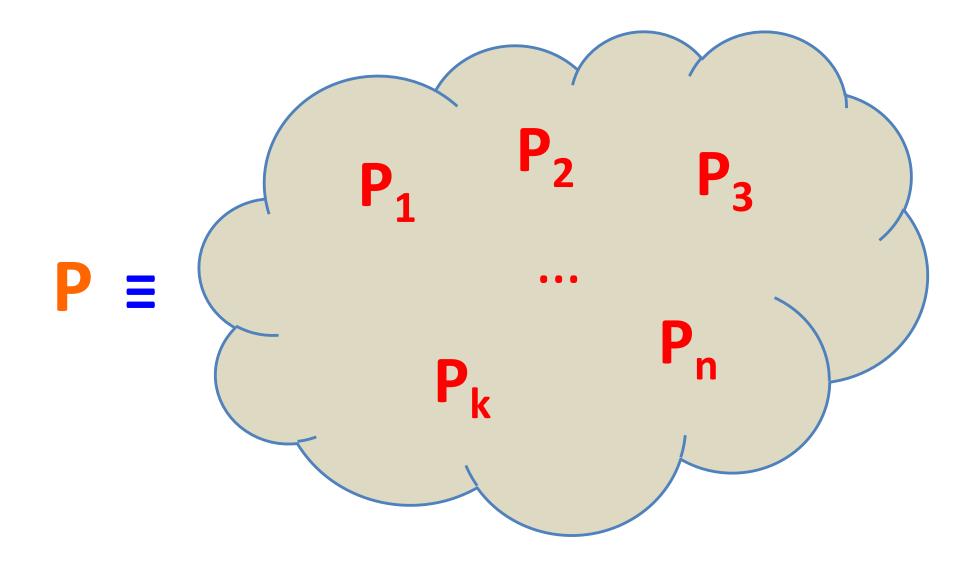
"... very, very concerning when I

got to the root cause, and very

annoying to fix ..."

http://llvm.org/bugs/show_bug.cgi?id=14972





Key challenges

Generation

How to generate different, yet equivalent tests?

□ Validation

How to check that tests are indeed equivalent?

Both are long-standing hard issues

Equiv. modulo inputs

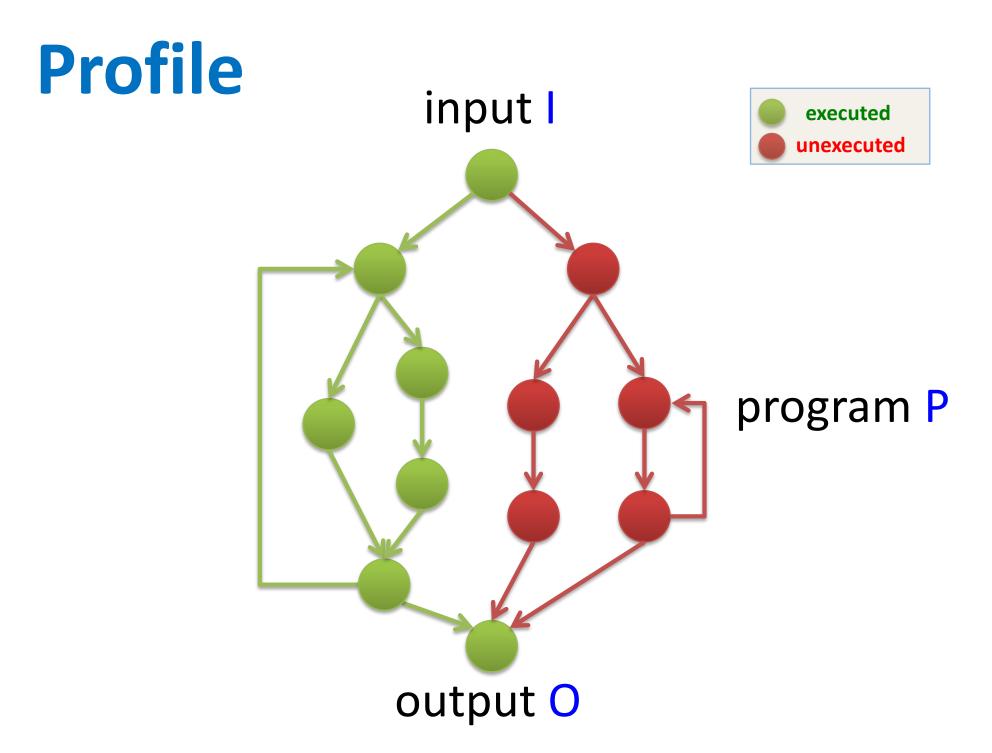
- Relax equiv. wrt a **given input i**
 - Must: P(i) = P_k(i) on input i

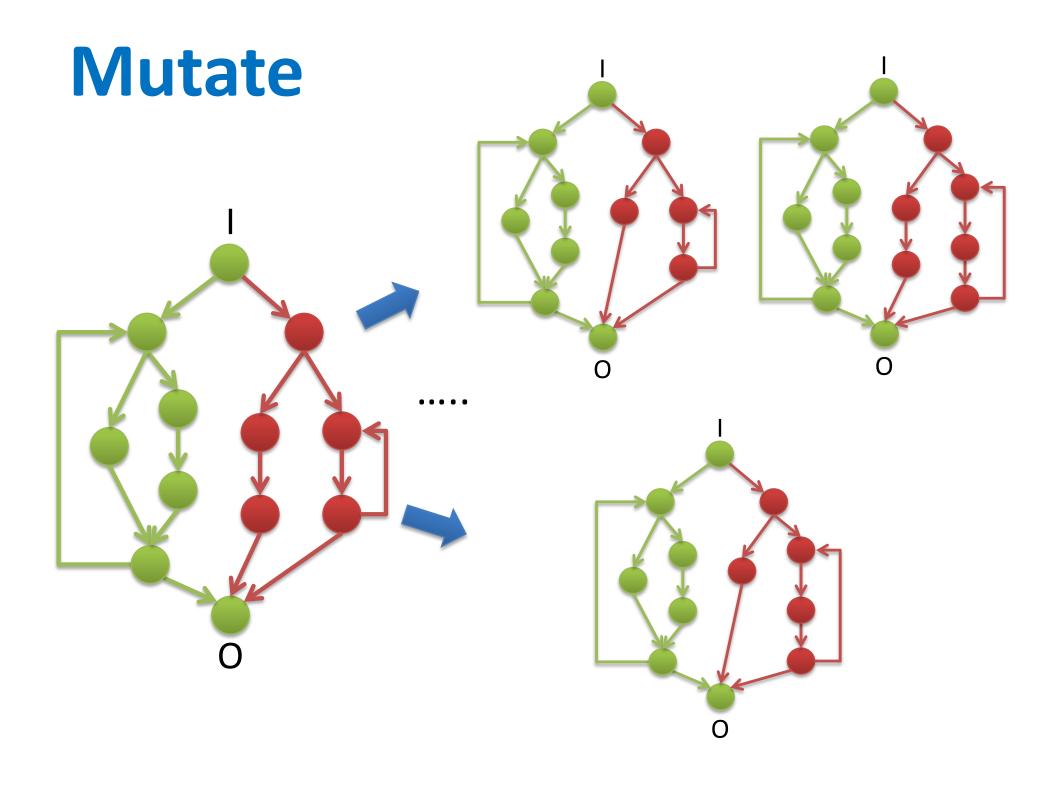
♦ Okay: $P(j) \neq P_k(j)$ on all input $j \neq i$

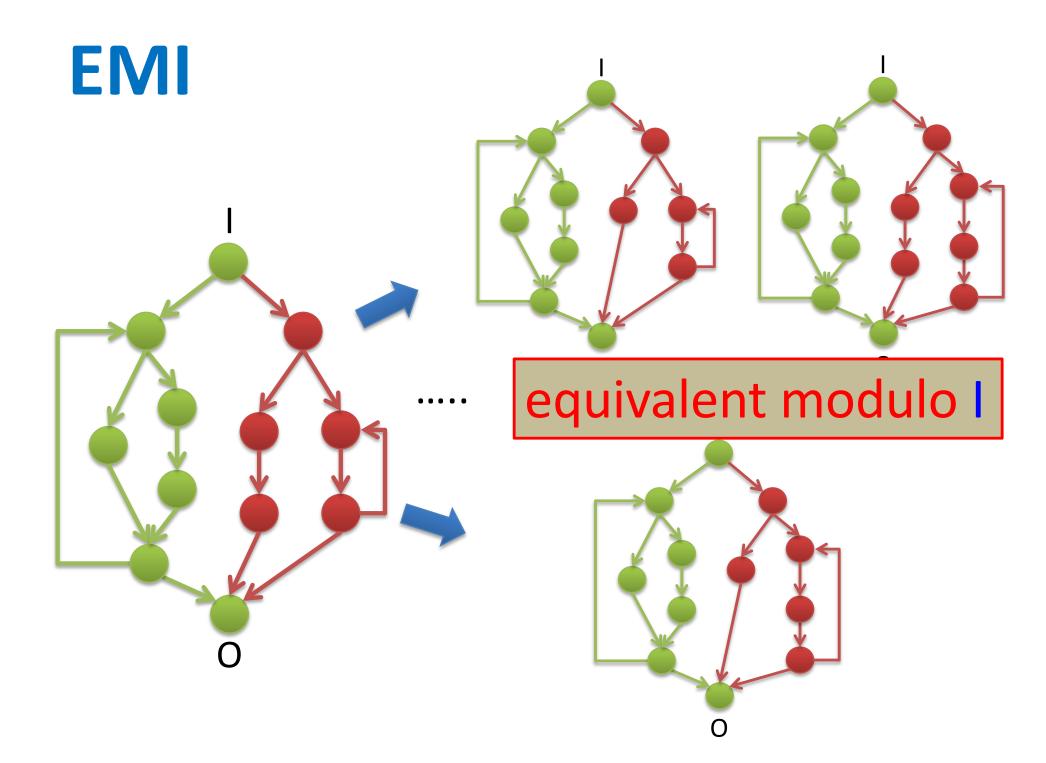
Exploit close interplay between

Dynamic program execution on some input

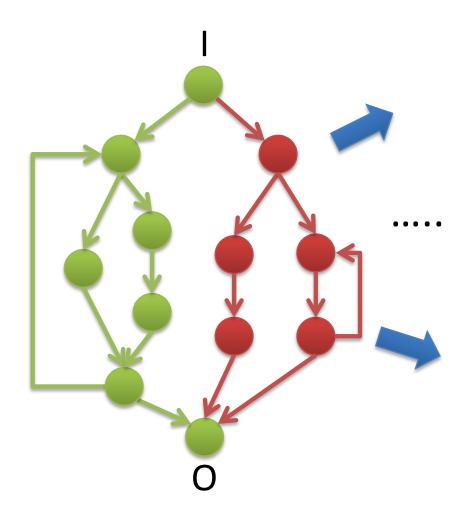
Static compilation for all input

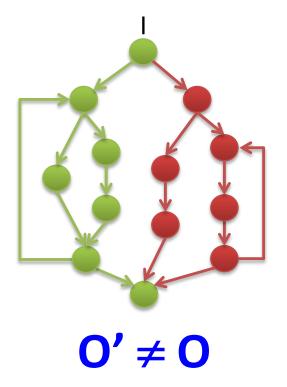














Revisit challenges

Generation (easy)

How to generate different, yet equivalent tests?

□Validation (easy)

How to check that tests are indeed equivalent?

Both are long-standing hard issues

LLVM bug 14972

```
struct tiny { char c; char d; char e; };
```

```
void foo(struct tiny x) {
    if (x.c != 1) abort();
    if (x.e != 1) abort();
}
```

```
int main() {
    struct tiny s;
    s.c = 1; s.d = 1; s.e = 1;
    foo(s);
    return 0;
}
```

Seed file

```
struct tiny { char c; char d; char e; };
  f(int n, struct tiny x, struct tiny y,
    struct tiny z, long 1) {
      if (x.c != 10) abort();
      if (x.d != 20) abort();
      if (x.e != 30) abort();
      if (y.c != 11) abort();
      if (y.d != 21) abort();
      if (y.e != 31) abort();
      if (z.c != 12) abort();
      if (z.d != 22) abort();
      if (z.e != 32) abort();
      if (1 != 123) abort();
  }
  main() {
      struct tiny x[3];
      x[0].c = 10;
      x[1].c = 11;
      x[2].c = 12;
      x[0].d = 20;
      x[1].d = 21;
      x[2].d = 22;
      x[0].e = 30;
      x[1].e = 31;
      x[2].e = 32;
      f(3, x[0], x[1], x[2], (long)123);
      exit(0);
  }
$ clang -m32 -00 test.c ; ./a.out
$ clang -m32 -01 test.c ; ./a.out
```

Seed file

```
struct tiny { char c; char d; char e; };
  f(int n, struct tiny x, struct tiny y,
    struct tiny z, long 1) {
      if (x.c != 10) abort();
      if (x.d != 20) abort();
      if (x.e != 30) abort();
      if (y.c != 11) abort();
      if (y.d != 21) abort();
                                  – unexecuted
      if (y.e != 31) abort();
      if (z.c != 12) abort();
      if (z.d != 22) abort();
      if (z.e != 32) abort();
      if (1 != 123) abort();
  }
  main() {
      struct tiny x[3];
      x[0].c = 10;
      x[1].c = 11;
      x[2].c = 12;
      x[0].d = 20;
      x[1].d = 21;
      x[2].d = 22;
      x[0].e = 30;
      x[1].e = 31;
      x[2].e = 32;
      f(3, x[0], x[1], x[2], (long)123);
      exit(0);
  }
$ clang -m32 -00 test.c ; ./a.out
$ clang -m32 -01 test.c ; ./a.out
```

Transformed file

```
struct tiny { char c; char d; char e; };
  f(int n, struct tiny x, struct tiny y,
    struct tiny z, long 1) {
      if (x.c != 10) /* deleted */;
      if (x.d != 20) abort();
      if (x.e != 30) /* deleted */;
      if (y.c != 11) abort();
      if (y.d != 21) abort();
      if (y.e != 31) /* deleted */;
      if (z.c != 12) abort();
      if (z.d != 22) /* deleted */;
      if (z.e != 32) abort();
      if (1 != 123) /* deleted */;
  }
  main() {
      struct tiny x[3];
      x[0].c = 10;
      x[1].c = 11;
      x[2].c = 12;
      x[0].d = 20;
      x[1].d = 21;
      x[2].d = 22;
      x[0].e = 30;
      x[1].e = 31;
      x[2].e = 32;
      f(3, x[0], x[1], x[2], (long)123);
      exit(0);
  }
$ clang -m32 -00 test.c ; ./a.out
$ clang -m32 -01 test.c ; ./a.out
Aborted (core dumped)
```

Reduced file

```
struct tiny { char c; char d; char e; };
```

```
void foo(struct tiny x) {
    if (x.c != 1) abort();
    if (x.e != 1) abort();
}
```

```
int main() {
    struct tiny s;
    s.c = 1; s.d = 1; s.e = 1;
    foo(s);
    return 0;
}
```

LLVM bug autopsy

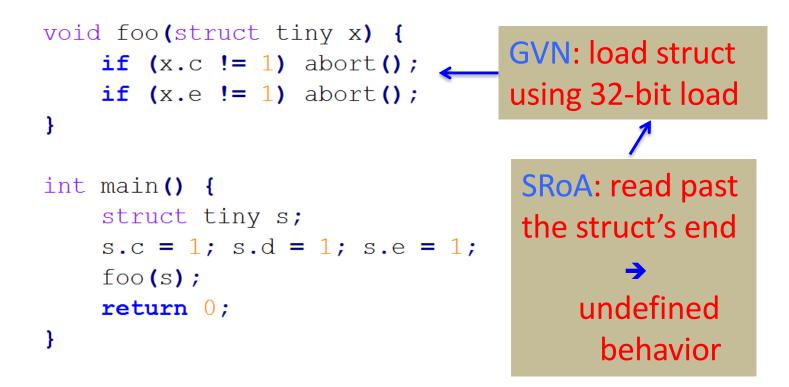
```
struct tiny { char c; char d; char e; };
```

```
void foo(struct tiny x) {
    if (x.c != 1) abort();
    if (x.e != 1) abort();
}
GVN: load struct
using 32-bit load
}
```

```
int main() {
    struct tiny s;
    s.c = 1; s.d = 1; s.e = 1;
    foo(s);
    return 0;
}
```

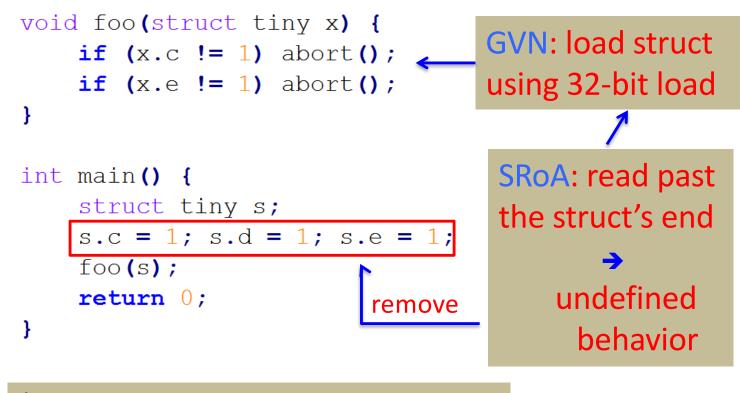
LLVM bug autopsy

```
struct tiny { char c; char d; char e; };
```



LLVM bug autopsy

```
struct tiny { char c; char d; char e; };
```

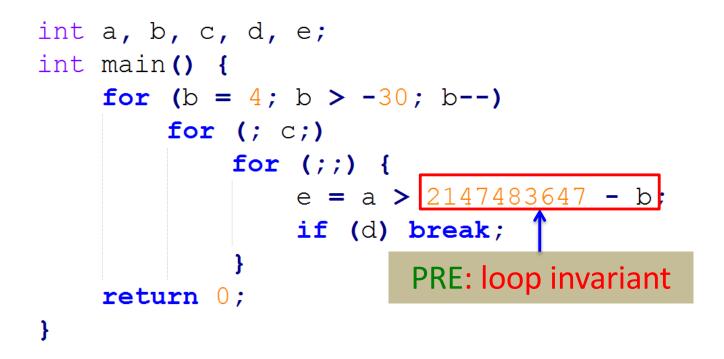


GCC bug 58731

```
int a, b, c, d, e;
int main() {
    for (b = 4; b > -30; b--)
        for (; c;)
        for (;;) {
            e = a > 2147483647 - b;
            if (d) break;
        }
    return 0;
}
```

```
$ gcc -00 test.c ; ./a.out
$ gcc -03 test.c ; ./a.out
^C
```

GCC bug autopsy

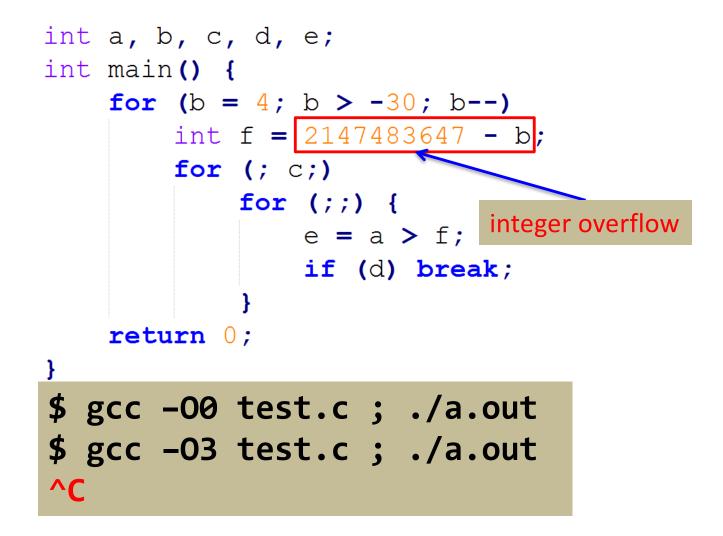


\$ gcc -00 test.c ; ./a.out
\$ gcc -03 test.c ; ./a.out
^C

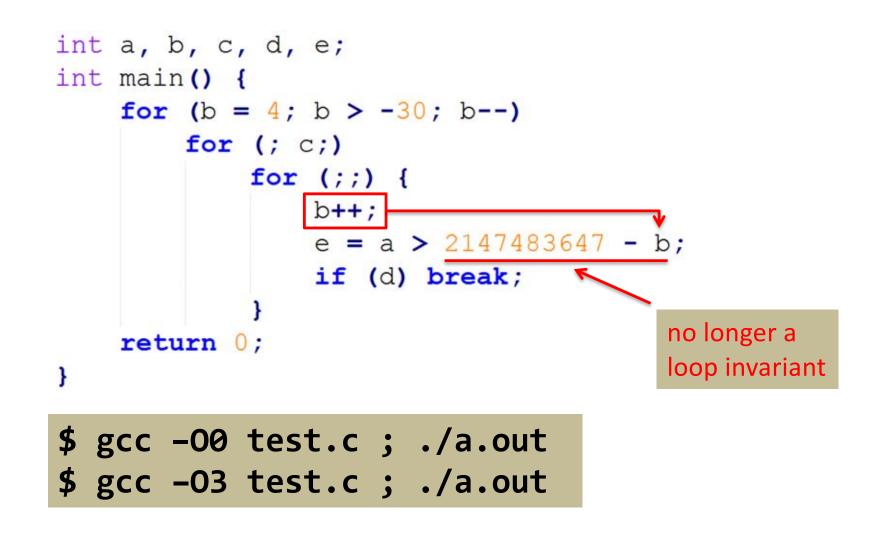
GCC bug autopsy

```
int a, b, c, d, e;
int main() {
    for (b = 4; b > -30; b--)
        int f = 2147483647 - b;
        for (; c;)
                         LIM
            for (;;) {
                e = a > f
                if (d) break;
    return 0;
$ gcc -00 test.c ; ./a.out
$ gcc -03 test.c ; ./a.out
^C
```

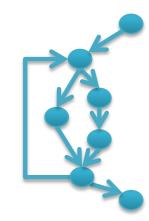
GCC bug autopsy



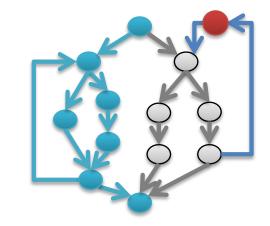
Seed program



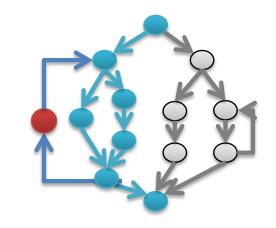
Orion (PLDI'14) Prune dead code



Athena (OOPSLA'15) Prune & inject dead code



Hermes (OOPSLA'16) Mutate live code



bug counts

	GCC	LLVM	TOTAL
Reported	841	781	1622
Fixed	612	419	1031

- ISSTA'15: Stress-testing link-time optimization
- ICSE'16: Analyzing compilers' diagnostic support
- PLDI'17: Skeletal program enumeration (SPE)

LLVM 3.9 & 4.0 Release Notes

"... thanks to Zhendong Su and his team whose fuzz testing prevented many bugs going into the release ..."

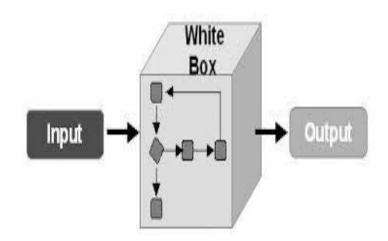
GCC's list of contributors

https://gcc.gnu.org/onlinedocs//gcc/Contributors.html

"Zhendong Su ... for reporting numerous bugs" "Chengnian Sun ... for reporting numerous bugs" "Qirun Zhang ... for reporting numerous bugs"

Take-away: Not only do good research, but be its **loyal**, **continuous** user!

Analyze Floating-Point Software



Floating-point code



Important: bugs can lead to disasters
 Challenging: hard to get right

Why difficult?

FP Math ≠ Real Math Non-linear relations Transcendental functions

sin, log, exp, ...

double foo(double x){ 1 2 **if** (x<=1.0) 3 X++; 4 5 y = x * x;6 **if** (y<=4.0) 7 X--; 8 return x; 9 }

Challenging for all known approaches

New perspective: ME

Analyzing numerical programs

- Coverage-based testing
- Boundary value analysis
- Numerical exception detection

Floating-point constraint solving

Mathematical Execution (ME)

Mathematical optimization (MO)

input **x** drives **p** to satisfy $\phi \leftrightarrow \mathbf{x}$ minimizes **r**

(p, φ)

FP constraints

Solving the floating-point constraint π

$$(SIN(x) = x) \land (x \ge 10^{-10})$$

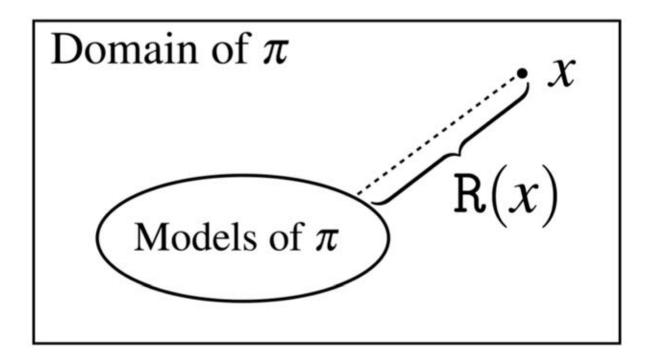
Satisfiable if x is floating-point

For
$$x \in \mathbb{F}$$
, $SIN(x) = x \Leftrightarrow x \simeq 0$

Unsatisfiable if x is real

For $x \in \mathbb{R}$, $SIN(x) = x \Leftrightarrow x = 0$

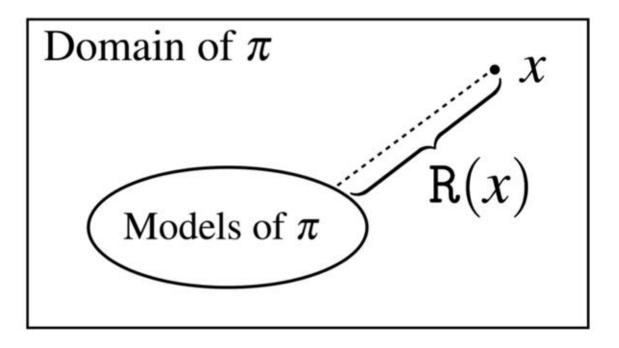




Simulate π with a floating-point program R

- $R(x) \ge 0$ for all x
- $R(x) = 0 \Leftrightarrow x \models \pi$





Minimize R as if it is a mathematical function

Let x* be the minimum point

$$\pi$$
 satisfiable $\Leftrightarrow \mathbf{R}(x^*) = 0$

Construct R

Necessary Conditions to meet :

1.
$$R(x) \ge 0$$
 for all x
2. $R(x) = 0 \Leftrightarrow x \models \pi$

How?

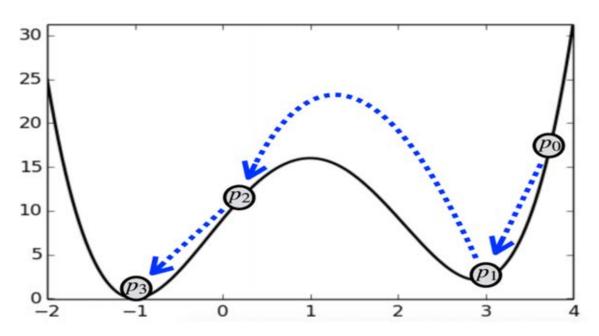
Constraint π	Program R
x == y	$(x-y)^2$
$x \leq y$	$x \le y ? 0 : (x - y)^2$
$\pi_1 \wedge \pi_2$	$R_1 + R_2$
$\pi_1 \lor \pi_2$	$R_1 * R_2$
D	

R can be constructed from a CNF form

Minimize R

Unconstrained programming techniques:

- Local optimization
- Monte Carlo Markov Chain (MCMC)
- We use them as black-box
- Do not analyze π ; execute R



Theoretical guarantees

Let R satisfy (1) $R(x) \ge 0$, and (2) $R(x) = 0 \Leftrightarrow x \models \pi$, and x^* be a minimum point of R. Then

$$\pi$$
 satisfiable $\Leftrightarrow \mathsf{R}(x^*) = 0$.

Threats

- Floating-point inaccuracy when calculating with R
- Sub-optimal x*

Example

 $(SIN(x) = = x) \land (x \ge 10^{-10})$

$(SIN(x) - x)^{2} + \begin{cases} 0 & \text{if } x \ge 10^{-10} \\ (x - 10^{-10})^{2} & \text{otherwise} \end{cases}$

 $x^* = 9.0 * 10^{-9}$ (can be others)

XSat & results

- Developed the ME-based XSat tool
- Evaluated against MathSat and Z3
- Used SMT-Comp 2015 FP benchmarks
- Result summary
 - 100% consistent results
 - 700+X faster than MathSat
 - 800+X faster than Z3

Take-away: Don't be afraid of difficult problems, look at them from **new perspectives**,

even for "damned" problems!

Generalizations

Coverage-based testing of FP code

Boundary value analysis

□ FP exception detection

Path divergence detection

Coverage-based testing

Goal

To generate test inputs to cover all branches of a program like this:

- pointer operations: &, *
- type casting: (int*), (unsigned)
- bit operations ^, &, >>
- floating-point comparison

```
double ieee754 fmod(double x, double y){
  Zero[] = \{0.0, -0.0,\};
  hx = *(1+(int*)\delta x);
  lx = *(int*)\delta x;
  hy = *(1+(int*)\delta y);
  ly = *(int*)&y;
  sx = hx \& 0x 8000000;
  hx ^=sx;
  hy \&= 0x7ffffff;
  if((hy|ly)==0||(hx>=0x7ff00000)||
     ((hy|((ly|-ly)>>31))>0x7ff00000))
    return (x*y)/(x*y);
  if(hx<=hy) {
    if((hx<hy)||(lx<ly)) return x;</pre>
    if(lx==ly)
      return Zero[(unsigned)sx>>31];
 }
  if(hx<0x00100000) {
   if(hx==0) {
```

State-of-the-art & Challenges

Symbolic execution

- Path explosion
- Constraint solving

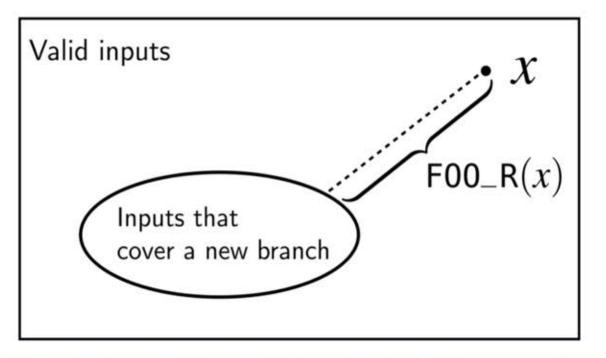
Search-based testing

- Fitness function
- Search strategies

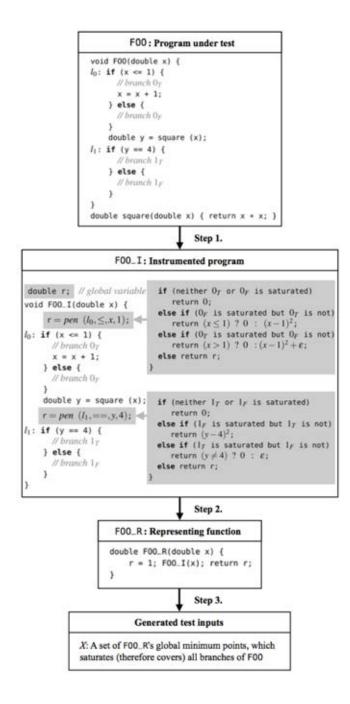
Our approach

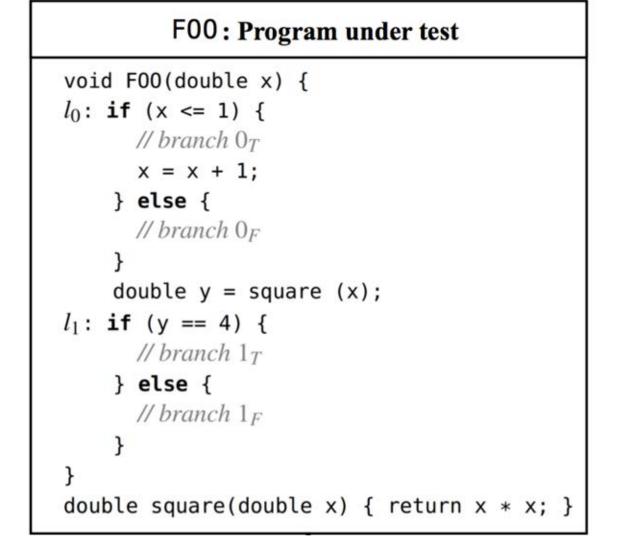
- No path issues
- No need to solve
 - constraints
- Effective for FP
 - programs

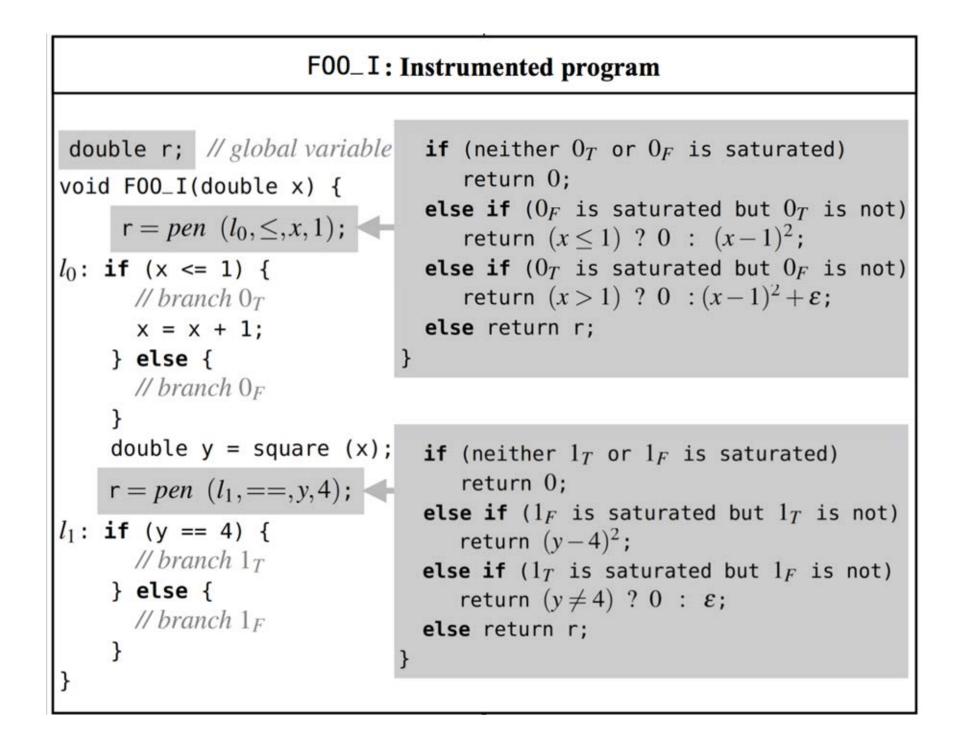
Our approach

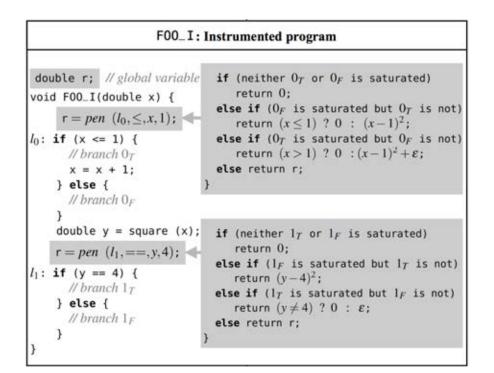


Step 1: Derive a program F00_R from F00 s.t.
F00_R(x) ≥ 0 for all x, and
F00_R(x) = 0 ⇔ x covers a new branch
Step 2: Repeatedly minimize F00_R until > 0









F00_R: Representing function

double F00_R(double x) {
 r = 1; F00_I(x); return r;

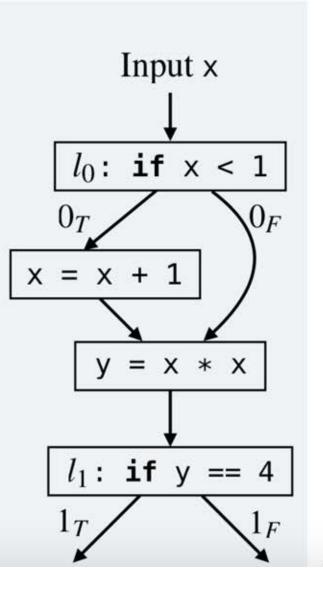
}

Generated test inputs

X: A set of FOO_R 's global minimum points, which saturates (therefore covers) all branches of FOO

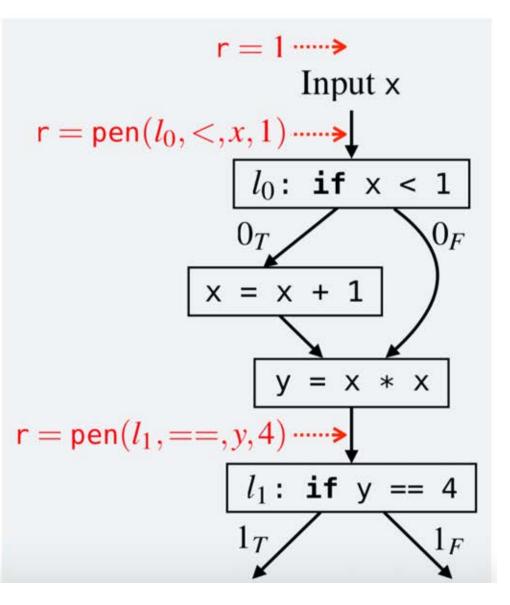
Example

Generate an input set to cover $\{0_T, 0_F, 1_T, 1_F\}$



Step 1: Construct F00_R

- r: global variable
- FOO_R: $x \rightarrow r$
- $R_{a op b}$: Branch distance



Branch distance R_{a op b}

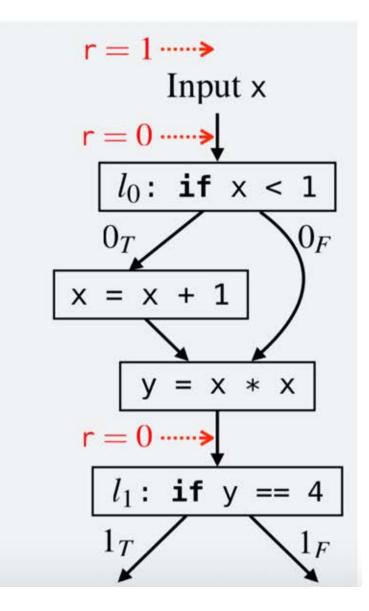
A helper function to quantify how far a and b are from attaining branch *a op b*.

$$R_{a==b}$$
 defined as $(a-b)^2$
 $R_{a\geq b}$ defined as $(a\geq b)$? $0: (a-b)^2$

 $\begin{array}{c} \text{covered at } l_i \ pen(l_i, op, a, b) \\ \emptyset & 0 \\ \{i_F\} & R_{a \ op \ b} \\ \{i_T\} & R_{\neg(a \ op \ b)} \\ \{i_T, i_F\} & r \end{array}$

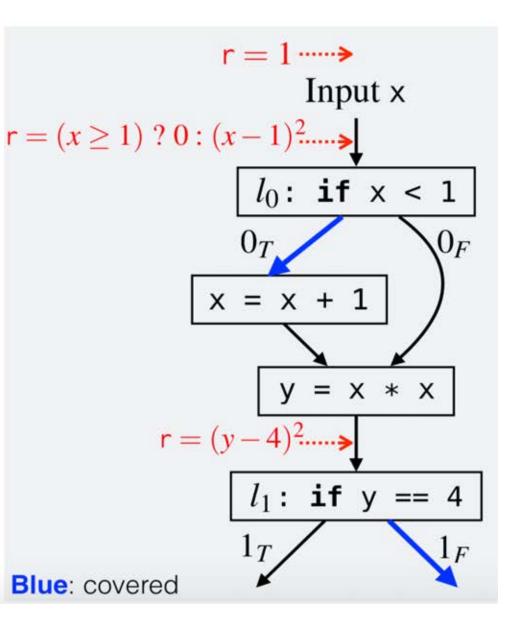
- No branch is covered
- Any input is a minimum point

• Assume
$$x^* = 0.7$$



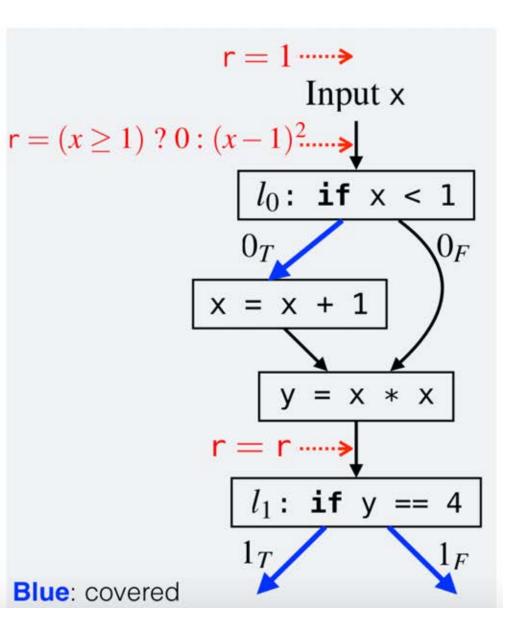
covered at I_i	$pen(l_i, op, a, b)$
Ø	0
$\{i_F\}$	R _{a op b}
$\{i_T\}$	$R_{\neg(a \ op \ b)}$
$\{i_T, i_F\}$	r

- $1_F, 0_T$ are covered
- F00_R attains minimum at -3 or 2
- Assume $x^* = -3$



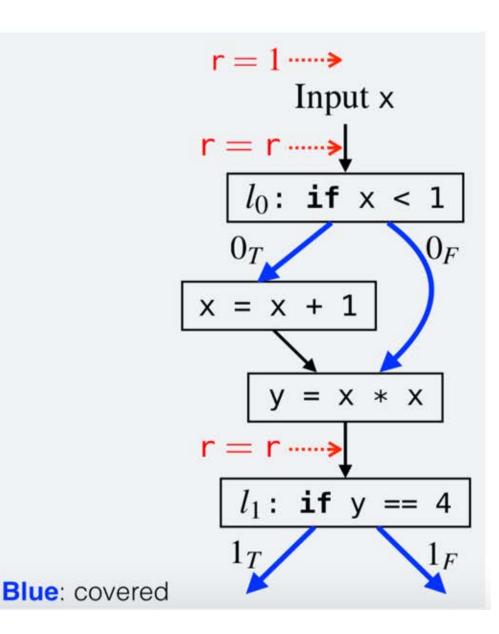
covered at l_i	$pen(l_i, op, a, b)$
Ø	0
$\{i_F\}$	R _{a op b}
$\{i_T\}$	$R_{\neg(a op b)}$
$\{i_{T}, i_{F}\}$	r

- $1_F, 1_T, 0_T$ are covered
- F00_R attains minimum at ≥ 1
- Assume $x^* = 5.1$

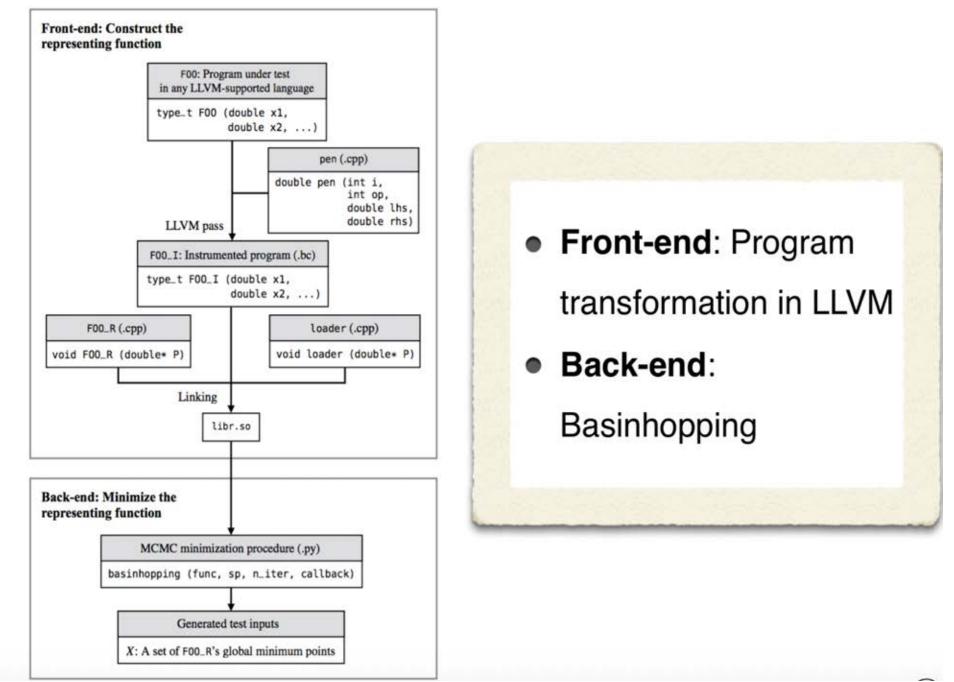


covered at l_i	$pen(l_i, op, a, b)$
Ø	0
$\{i_F\}$	R _{a op b}
$\{i_T\}$	$R_{\neg(a \ op \ b)}$
$\{i_T, i_F\}$	r

- All branches are covered
- $\forall x, FOO_R(x) = 1$
- Termination



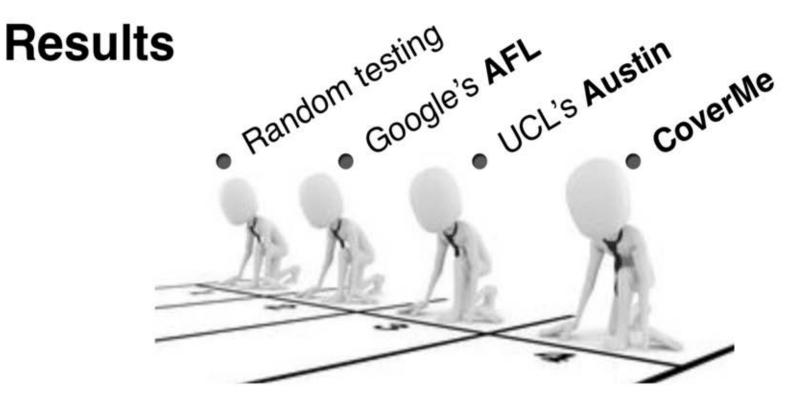
Our implementation CoverMe



Experiments

Benchmarks: Fdlibm

- Sun's math library
- Reference for Java SE 8's math library
- Used in Matlab, JavaScript and Android
- Heavy on branches (max=114, avg=23)



CoverMe covers

- \approx 90% branches in 7 seconds
- \approx 18% more branches than AFL with 1/10 time
- \approx 40% more branches than Austin with speedups of several orders of magnitudes

ME in the long run

Offers a new general analysis paradigm

Complements existing approaches

Random concrete execution (CE)

Symbolic execution (SE)

Abstract execution (AE)

Peeking into the future ...

What is the key mission of

Computer Science?

To help people **turn creative** ideas into working systems

Software research is

central to this mission



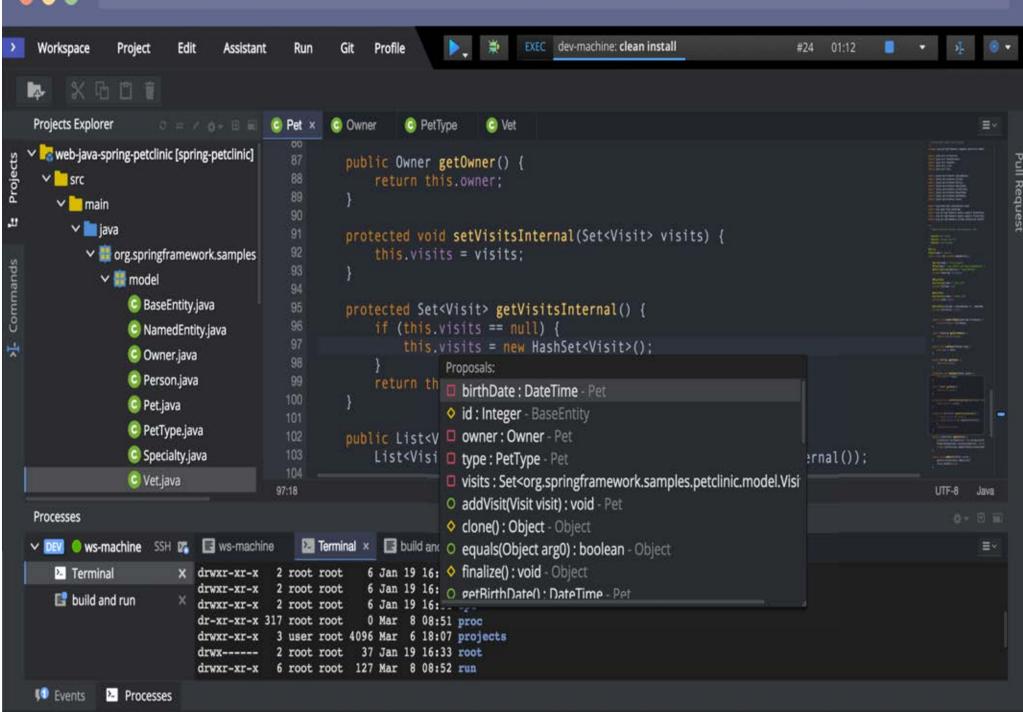
A lot of progress to celebrate for!

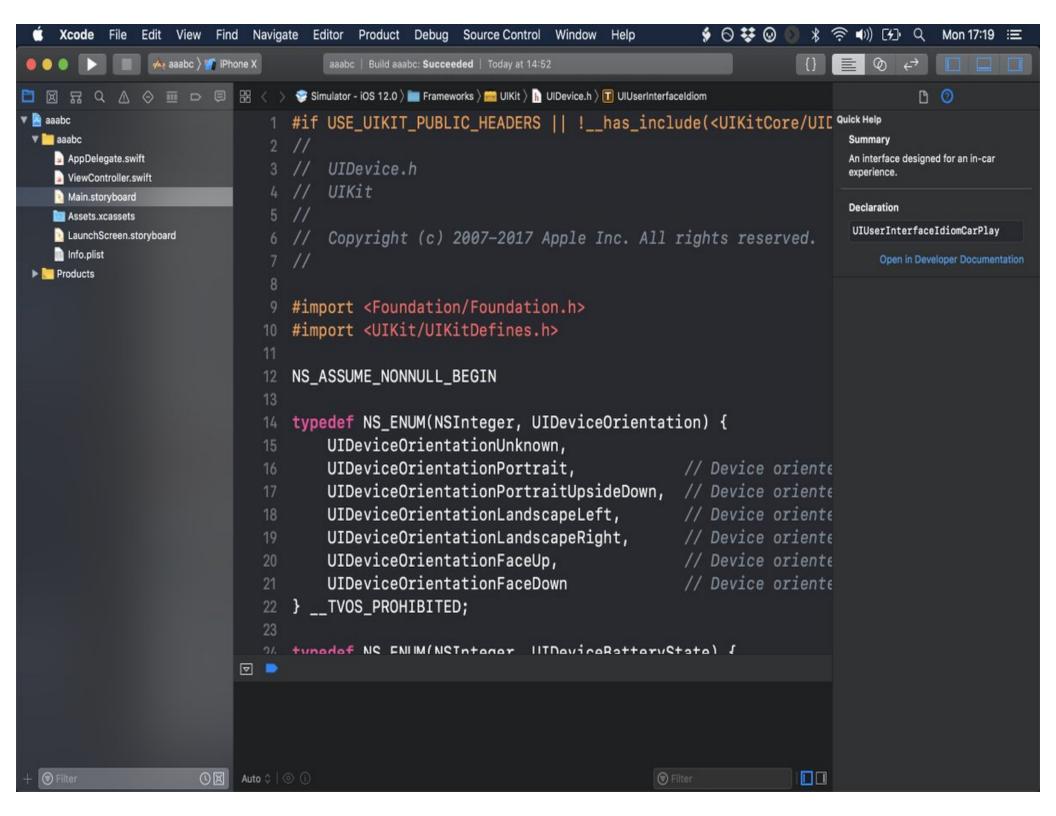


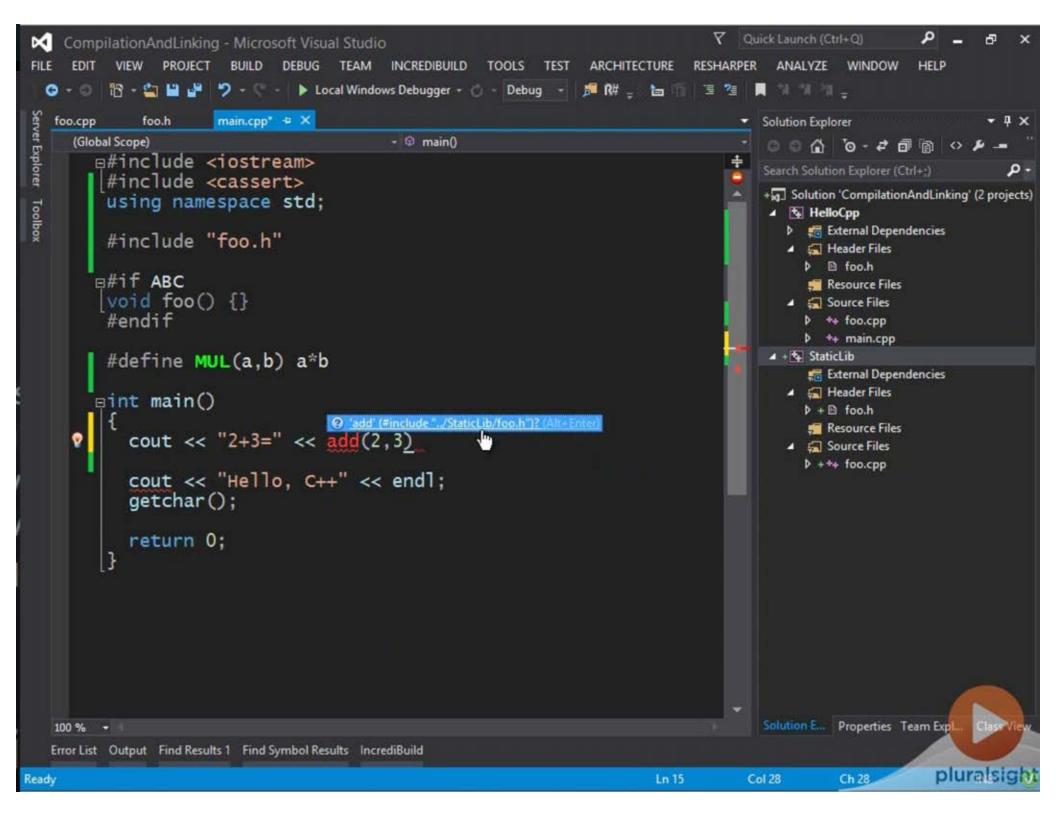
```
1 #include <stdio.h>
 2 #include <string.h>
 3
 4 int main()
       char *word = "everest";
 5
       char reverseword[strlen(word)+1];
 6
       unsigned int letters_remaining = strlen(word);
 7
       char *wordpointer = &word[strlen(word)-1];
 8
       int i = 0;
 9
       while(letters_remaining > 0){
10
           reverseword[i++] = *wordpointer--;
11
           letters_remaining--;
12
       3
13
       reverseword[strlen(word)] = '\0';
14
       printf("So the reversed word is %s\n",reverseword);
15
       return 0;
16
17
```

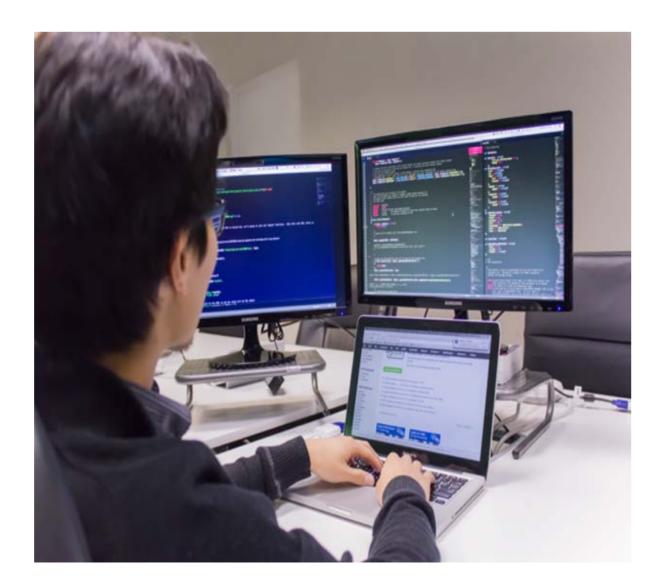
```
#ifndef ___ASM_ALPHA_FPU_H
    #define ASM ALPHA FPU H
   #include <asm/special_insns.h>
    #include <uapi/asm/fpu.h>
    /* The following two functions don't need trapb/excb instructions
       around the mf_fpcr/mt_fpcr instructions because (a) the kernel
       never generates arithmetic faults and (b) call_pal instructions
       are implied trap barriers. */
10
12
    static inline unsigned long
13
    rdfpcr(void)
14
    1
        unsigned long tmp, ret;
    #if defined(CONFIG_ALPHA_EV6) || defined(CONFIG_ALPHA_EV67)
        __asm___volatile__ (
19
            "ftoit $f0,%0\n\t"
20
            "mf_fpcr $f0\n\t"
            "ftoit $f0,%1\n\t"
22
            "itoft %0.$f0"
            : "=r"(tmp), "=r"(ret));
24
    #else
        ___asm____volatile__ (
            "stt $f0,%0\n\t"
26
27
            "mf_fpcr $f0\n\t"
            "stt $f0,%1\n\t"
29
            "ldt $f0,%0"
            : "=m"(tmp), "=m"(ret));
31
    #endif
32
33
        return ret;
34
    3
    static inline void
36
37
   wrfpcr(unsigned long val)
38
    -f
39
        unsigned long tmp;
41
    #if defined(CONFIG_ALPHA_EV6) || defined(CONFIG_ALPHA_EV67)
42
        __asm____volatile__ (
            "ftoit $f0,%0\n\t"
43
            ":+.... #1 #FAL.....
.../linux/arch/alpha/include/asm/fpu.h [Ins]
Eval: START
```

(1, 0) [Top/1.8k]

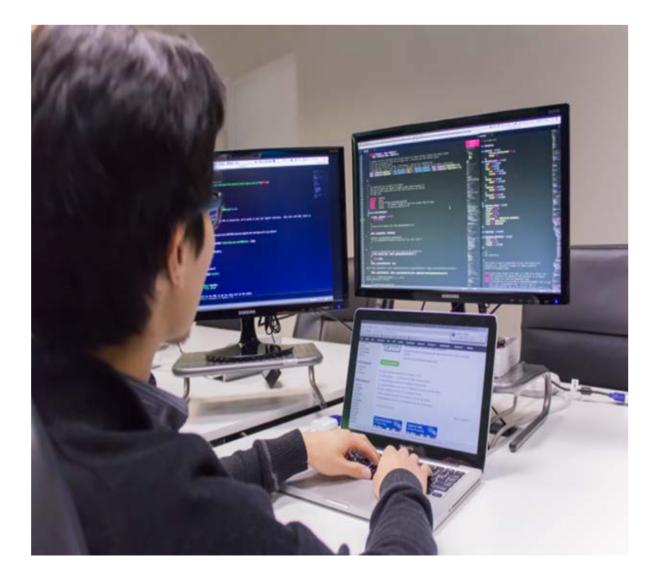


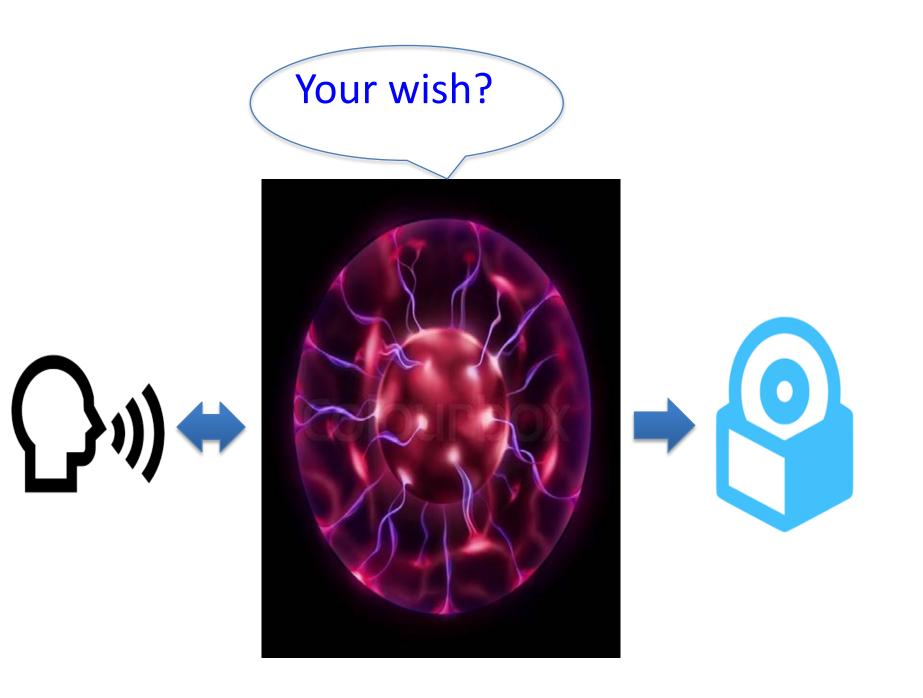


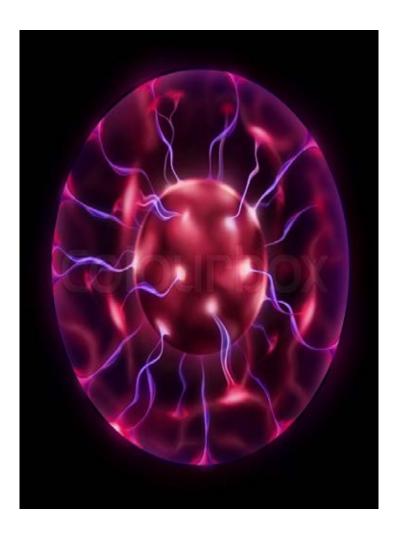




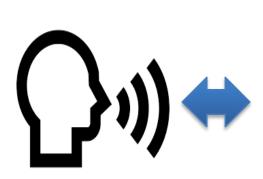
Can we move beyond "coding"?

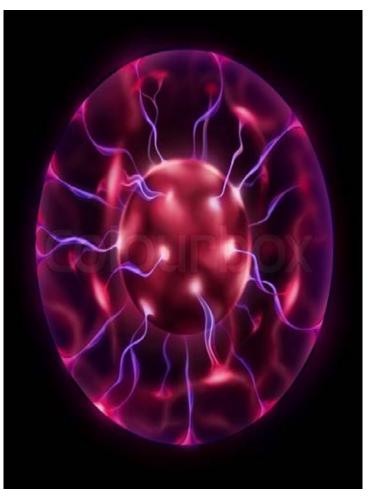






Hardest: communicate the wish







```
function drawTree () {
    var blossomPoints = [];
```

```
resetRandom();
drawBranches(0, -Math.PI/2, canvasWidth/2, canvasHeight, 30,
```

```
resetRandom();
drawBlossoms(blossomPoints);
```

```
function drawBranches (i,angle,x,y,width,blossomPoints) {
    ctx.save();
```

```
var length = tween(i, 1, 62, 12, 3) * random(0.7, 1.3);
if (i == 0) { length = 107; }
```

```
ctx.translate(x,y);
ctx.rotate(angle);
ctx.fillStyle = "#000";
ctx.fillRect(0, -width/2, length, width);
```

```
ctx.restore();
```

```
var tipX = x + (length - width/2) * Math.cos(angle);
var tipY = y + (length - width/2) * Math.sin(angle);
```

```
if (i > 4) {
    blossomPoints.push([x,y,tipX,tipY]);
}
```

```
if (i < 6) {
    drawBranches(i + 1, angle + random(-0.15, -0.05) * Math i
    drawBranches(i + 1, angle + random( 0.15, 0.05) * Math i
}
else if (i < 12) {
    drawBranches(i + 1, angle + random( 0.25, -0.05) * Math.i
</pre>
```



var length = tween(1, 1, 62, 12, 3) + random(0.7, 1.3);
if (1 == 0) { length = 107; }

ctx.translate(x,y);

ctx.fillRect(153, -width/2, length, width);

ctx.restore();

var tipX = x + (length - width/2) * Math.cos(angle); var tipY = y + (length - width/2) * Math.sin(angle); if (i > 4) {

blossomPoints.push([x,y,tipX,tipY]);

if (1 = 6) {

drawBranches(i + 1, angle + random(-0.15, -0.05) + Math.)
drawBranches(i + 1, angle + random(0.15, 0.05) + Math.)
}
else if (i < 12) {
 drawBranches(i + 1, angle + random(0.25, -0.05) + Math.)</pre>

function drawBlossoms (blossomPoints) (

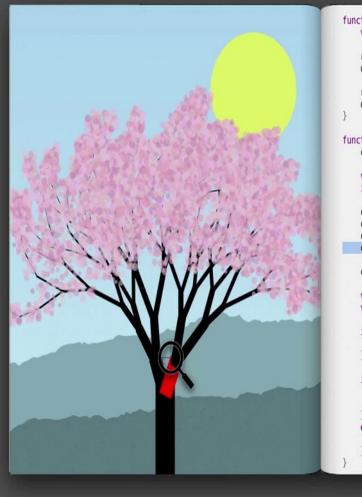
var colors = ("#f5ceea", "#e8d9e4", "#f7c9f3", "#ebb4cc", "# ctx.globalAlpha = 0.60;

```
for (var i = 0; i < blossomPoints.length; i++) {
    var p = blossomPoints[i];
    for (var j = 0; j < 16; j++) {
        var x = lerp(p[0], p[2], random(0,1)) = random(-10,10;
        var y = lerp(p[1], p[3], random(0,1)) + random(-10,10;
        var y = lerp(p[1], p[3], rand
```

ctx.fillStyle = colors(Math.floor(random(0,colors.le ctx.fillCircle(x, y, random(2,5));

Bret Victor: Inventing on Principle





```
function drawTree () {
    var blossomPoints = [];
```

```
resetRandom();
drawBranches(0, -Math.PI/2, canvasWidth/2, canvasHeight, 30,
```

resetRandom(); drawBlossoms(blossomPoints);

function drawBranches (i,angle,x,y,width,blossomPoints) {
 ctx.save();

var length = tween(i, 1, 62, 12, 3) * random(0.7, 1.3); if (i == 0) { length = 107; }

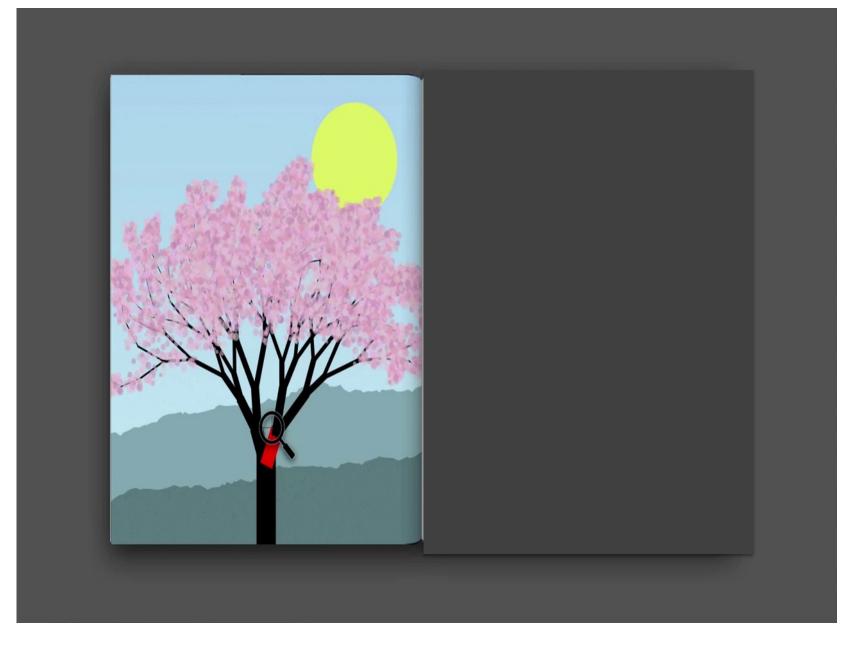
ctx.translate(x,y); ctx.rotate(angle); ctx.fillStyle = "#000"; ctx.fillRect(0, -width/2, length, width);

ctx.restore();

var tipX = x + (length - width/2) * Math.cos(angle); var tipY = y + (length - width/2) * Math.sin(angle);

```
if (i > 4) {
    blossomPoints.push([x,y,tipX,tipY]);
}
```

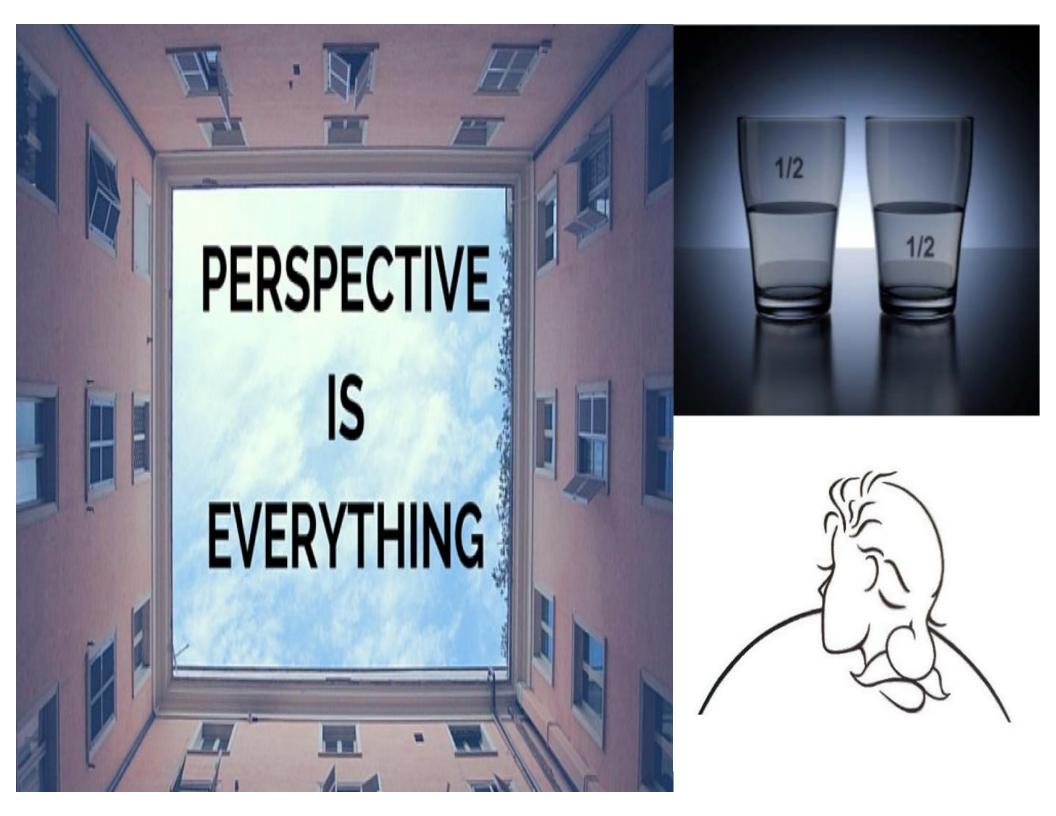
```
if (i < 6) {
    drawBranches(i + 1, angle + random(-0.15, -0.05) * Math I
    drawBranches(i + 1, angle + random( 0.15, 0.05) * Math I
}
else if (i < 12) {
    drawBranches(i + 1, angle + random( 0.25, -0.05) * Math.I</pre>
```



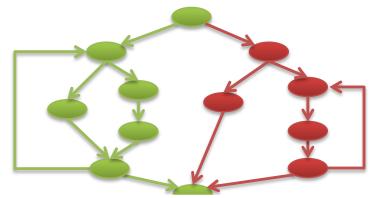
Goal is the object, not the code

Can we directly manipulate & explore the object to express the "wish"? Can we directly manipulate & explore the object to express the "wish"?

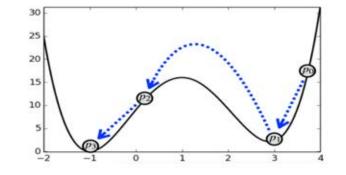
Perhaps via visualization & virtual reality?



Advancing Software Analysis via Changed Perspectives



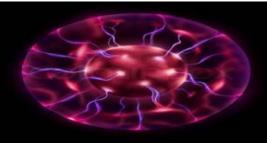
	GCC	LLVM	TOTAL
Reported	841	781	1622
Fixed	612	419	1031



Benchmark		Satisfiability			Time (seconds)					
SMT2-LIB program	size (byte)	#var	MathSat	Z3	Coral	XSat	MathSat	Z3	Coral	XSat
SUMMARY			•	100.0%	54.6%	100.0%	2014.75	2290.05	1.38	2.80

your wish?

 \bigcirc »





Thank you!