Toward a Standard Benchmark Format and Suite for Floating-Point Analysis



Nasrine Damouche, Matthieu Martel, **Pavel Panchekha**, Chen Qiu, Alexander Sanchez-Stern, Zachary Tatlock.

Incredible progress...

Automatic Verification

Fluctuat [SAS'13]

Rosa [POPL'14]

FPTaylor [FM'15]

Optimization

STOKE [PLDI'14]

Improvement

Salsa [FMICS'15]

Herbie [PLDI'15]

Mechanized Proofs

Wave equation [ITP'10]

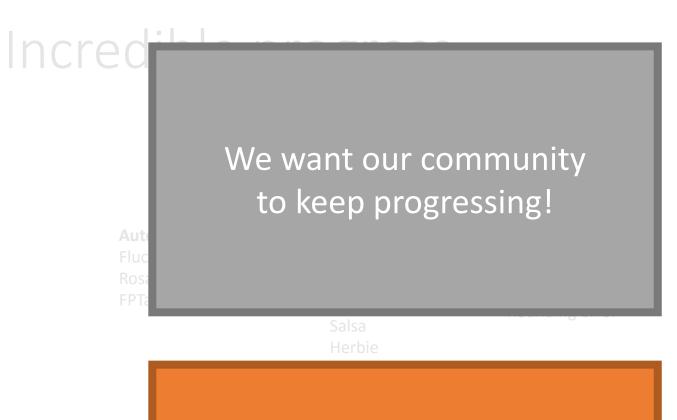
Rounding error [NSV'16]

Rapid improvement in hard problems!

Incredible progress...



Rapid improvement in hard problems!



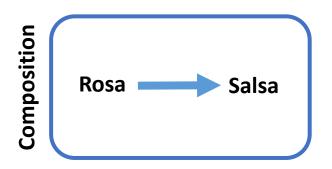
Next

As community grows, growing pains appear

Rapid improvement in hard problems!

Growing pains

Similar growing pains in compilers, HPC, SAT, SMT, ... communities

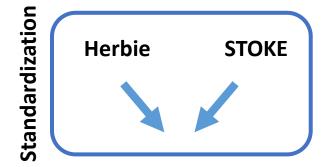


Rosa: def example(x: Double): Double = ...
Salsa: double example(double x) { ... }

```
Fluctuat FPTaylor
```

Fluctuat: Poly, Inv, F1a, F1b, idem, ...

FPTaylor: sine, sqrt, verhulst, ...



Herbie: ulp(NaN, Inf) = UINT_MAX
STOKE: ulp(NaN, Inf) < UINT_MAX</pre>



FPBench is **community infrastructure** for **cooperation** and **comparison** in the FP community.

http://fpbench.org

3

Common format

Benchmark suite

Named measures

β

Common format

Benchmark suite

Named measures

$$\sqrt{x+1} - \sqrt{x}$$

Arguments

(FPCore (x)

S-expression syntax

$$\sqrt{x+1} - \sqrt{x}$$

```
(FPCore (x)
  :name "Sqrt Difference"
  :cite (hamming-87)
  :pre (> x 0)
  (- (sqrt + x 1)) (sqrt x)))
  Preconditions
```

```
(FPCore (x0)
  :name "Sine Newton"
  :cite (darulova-kuncak-2014)
  :pre (< (abs x0) 1)
  (while (< i 10)
    ([i 0 (+ i 1)]
     [x \ x0] Loops (+ (+ (- x (/ (pow x 3) 6.0))
                          (/ (pow x 5) 120.0))
                          (/ (pow x 7) 5040.0))
                    (+ (+ (- 1.0 (/ (* x x) 2.0))
                          (/ (pow x 4) 24.0))
                          (/ (pow x 6) 720.0))))))
    x))
                              Common functions
```

FPCore common format

Simple to use

S-expression syntax
Purely functional
No control flow analysis

Generate from higher-level, imperative FPImp lang.

Expressive

All C, Fortran functions Loops, conditionals Tools support parts

Extensible

Metadata properties
Tool-specific metadata
Input or output format

β

Common format

Benchmark suite

Named measures

Simple to implement

Covers all existing uses

Simple to extend, specialize

β

Common format

Benchmark suite

Named measures

Simple to implement

Covers all existing uses

Simple to extend, specialize

FPBench benchmark suite

72 total benchmarks

Drawn from existing papers

Annotated with source, ranges, description, citation

FPBench benchmark suite

Existing programs

Rich features

Diverse domains

FPTaylor	29
Herbie	28
Rosa	6
Salsa	9

Arith	72
Expt	16
Trig	11
Loop	12
Branch	3

Textbook	59
Math Alg	6
Emb Sys	4
Sci Comp	3

β

Common format

Benchmark suite

Named measures

Simple to implement

Covers all existing uses

Simple to extend, specialize

From existing projects

Cover many domains

Grows over time

β

Common format

Benchmark suite

Named measures

Simple to implement

Covers all existing uses

Simple to extend, specialize

From existing projects

Cover many domains

Grows over time

FPBench measures

Formal definitions of accuracy measures

Described along 5 axes

Standard measures so tools agree

FPBench axes of measurement

Scaling vs. non-scaling

Absolute, relative, ULPs, bits, ...

Forward vs. backward

Fixed input error vs fixed output error

Maximum vs. average

vs -----

Sound vs. statistical

Formal guarantees vs mathematical accuracy

Improvement





β

Common format

Benchmark suite

Named measures

Simple to implement

Covers all existing uses

Simple to extend, specialize

From existing projects

Cover many domains

Grows over time

Terms for measuring error

Standard across tools

Flexible but rigorous

β

Common format

Benchmark suite

Named measures

Simple to implement

Covers all existing uses

Simple to extend, specialize

From existing projects

Cover many domains

Grows over time

Terms for measuring error

Standard across tools

Flexible but rigorous



FPBench is **community infrastructure** for **cooperation** and **comparison** in the FP community.

Common format

Benchmark suite

Named measures

http://fpbench.org