A Three-tier Strategy for Reasoning about Floating-Point Numbers in SMT

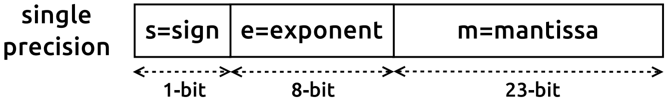
Mohamed Iguernlala

OCaml PRO

Joint Work with : Sylvain Conchon, Kailiang Ji, Guillaume Melquiond, Clément Fumex

(Binary) Floating-Point Arithmetic

- Natural way of approximating Real numbers in computers
- Optimized memory representation and efficient operations



$$normalized\_number = (-1)^s \times (1.m) \times 2^{e-127}$$

FPA pitfalls for common programmers

- What you expect on Reals is not what you get on FP numbers!

$$
\begin{array}{ll}
x = 1.0; & \\
y = 1000000000.; & \\
\textbf{if} \ \ (x + y > y) & \textit{printf}(\text{"}\textit{alright}!\text{"}); \\
\textbf{else} & \textit{printf}(\text{"}\textit{ouch}!\text{"});
\end{array}
$$

The result of an FP computation may arbitrary diverge from expected Real value

FPA reasoning

- Mainly investigated in the context of theorem proving, abstract interpretation, and constraints solving
- Considered only recently in SMT
  - An SMT-LIB theory for FPA, based on IEEE 754-2008
  - Some effort to design (decision) procedures for FPA
  - Some SMT solvers already implement FPA solvers

FPA reasoning in SMT

Bit-blasting (Z3, MathSAT5, SONOLAR)

- circuits encoding (heuristic : reduce FP precision to scale)

Abstract CDCL (MathSAT5)

- CDCL on FP abstract domains

Offline reduction to non-linear RIA (REALIZER+Z3)

- reduction to NRA + rounding operator
- encoding to RIA
- exceptional values not handled !

FPA reasoning in SMT

Bit-blasting (Z3, MathSAT5, SONOLAR)

- circuits encoding (heuristic : reduce FP precision to scale)

Abstract CDCL (MathSAT5)

- CDCL on FP abstract domains

Offline reduction to non-linear RIA (REALIZER+Z3)

- reduction to NRA + rounding operator
- encoding to RIA
- exceptional values not handled !

$\rightarrow$ Some FPA proofs are simpler on reals !

$\rightarrow$ Combination with other theories ? (eg. Reals)

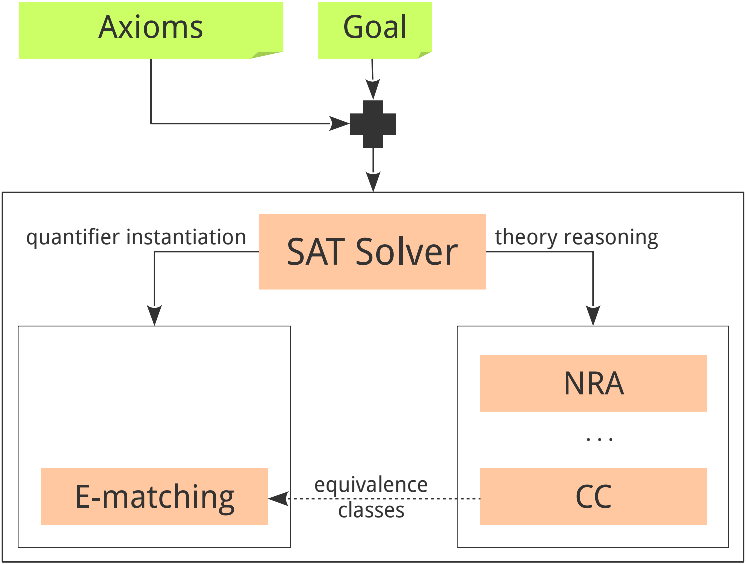
Our idea : Online/Lazy reduction to NRA

Layer 1 Reasoning about high level aspects of FP operations, and possibly reducing FP to rounding operations on reals (as in IEEE-754 standard)

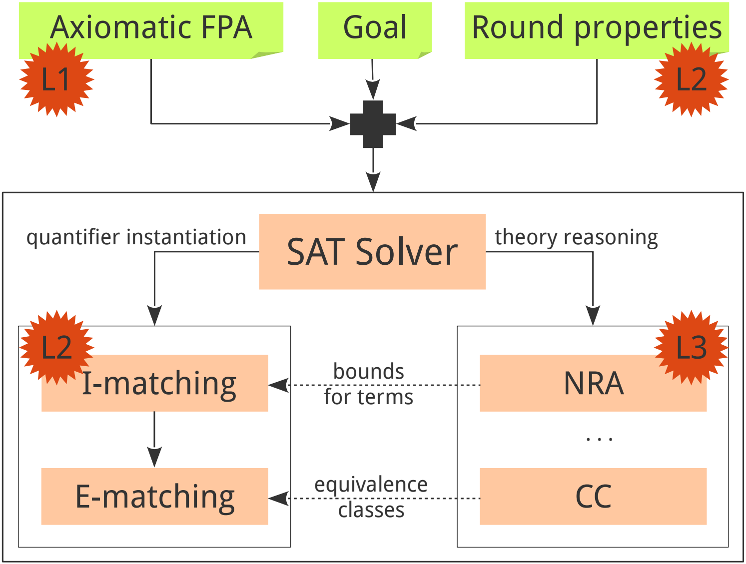
Layer 2 Reasoning about the rounding operator properties

Layer 3 (Non-linear) Real arithmetic reasoning : bounds inference, terms simplification

Current implementation in Alt-Ergo



Current implementation in Alt-Ergo



Our approach on an example

$$(2._{\mathbf{F}} \preceq u \preceq 10._{\mathbf{F}} \ \wedge\ 2._{\mathbf{F}} \preceq v \preceq 10._{\mathbf{F}}) \quad \Longrightarrow$$

$$(\overline{u \oplus v}) - (\overline{u} + \overline{v}) \ \leq\ 0.00000096$$

u and v are simple precision FP variables,

$2._{\mathbf{F}}$ and $10._{\mathbf{F}}$ are two FP constants,

$\preceq$ is the less-or-equal predicate over FP numbers,

$\oplus$ is FP addition,

$\overline{x}$ denotes the Real value of an FP expression x

Example : (some) axioms from "Layer 1"

L1-1 $\quad \forall z.\ \texttt{in\_range}(z) \iff -\text{0x1.FFFFFEp127} \leq z \leq \text{0x1.FFFFFEp127}$

L1-2 $\quad \forall m.\forall x.\forall y.$

$$(\texttt{is\_finite}(x) \land \texttt{is\_finite}(y) \land \texttt{in\_range}(\texttt{round}_m(\overline{x} + \overline{y}))) \implies$$

$$\overline{x \oplus_m y} = \texttt{round}_m(\overline{x} + \overline{y})$$

L1-3 $\quad \forall x.\forall y.\ x \preceq y \implies$

$$\bigvee \begin{pmatrix} \texttt{is\_finite}(x) \land \texttt{is\_finite}(y) \\ \texttt{is\_infinite}(x) \land \texttt{is\_negative}(x) \land \neg\texttt{is\_nan}(y) \\ \texttt{is\_infinite}(y) \land \texttt{is\_positive}(y) \land \neg\texttt{is\_nan}(x) \end{pmatrix}$$

L1-4 $\quad \forall x.\forall y.\ (\texttt{is\_finite}(x) \land \texttt{is\_finite}(y) \land x \preceq y) \implies \overline{x} \leq \overline{y}$

L1-5 $\quad \forall x.\ (\texttt{is\_infinite}(x) \lor \texttt{is\_nan}(x)) \implies \neg\ \texttt{is\_finite}(x)$

L1-6 $\quad \forall x.\ \neg\ (\texttt{is\_negative}(x) \land \texttt{is\_positive}(x))$

Generic axiomatization of FP theory from Why3 [VSTTE'17]

Example : (some) axioms from "Layer 2"

Some mathematical properties about `round` operator

L2-1 $\forall m, i, j, z.$
$$i \le z \le j \implies \mathtt{round}_m(i) \;\le\; \mathtt{round}_m(z) \;\le\; \mathtt{round}_m(j)$$

L2-2 $\forall m, i, j, z.$
$$i \le z \le j \implies -2^{\alpha} \;\le\; \mathtt{round}_m(z) - z \;\le\; 2^{\alpha}$$

$$\text{where} \quad \alpha \equiv \mathtt{ilog}_2(\max(|i|,\, |j|,\, 2^{e_{\min}+prec-1})) - prec$$

$\rightarrow$ For single-precision FP, $prec = 24$ and $e_{min} = -149$

Example : (some) axioms from "Layer 2"

Some mathematical properties about `round` operator

L2-1 $\forall m, i, j, z.$

$$i \leq z \leq j \implies \texttt{round}_m(i) \ \leq \ \texttt{round}_m(z) \ \leq \ \texttt{round}_m(j)$$

L2-2 $\forall m, i, j, z.$

$$i \leq z \leq j \implies -2^{\alpha} \ \leq \ \texttt{round}_m(z) - z \ \leq \ 2^{\alpha}$$

$$\text{where} \quad \alpha \equiv \texttt{ilog}_2(\max(|i|,\ |j|,\ 2^{e_{\min}+prec-1})) - prec$$

$\rightarrow$ For single-precision FP, $prec = 24$ and $e_{min} = -149$

Challenge : how to efficiently instantiate this kind of axioms in SMT

Our approach on an example (very quickly !)

$$(\texttt{2.} \preceq u \preceq \texttt{10.} \land \texttt{2.} \preceq v \preceq \texttt{10.}) \Rightarrow (\overline{u \oplus v}) - (\overline{u} + \overline{v}) \ \leq\ 0.00000096$$

Our approach on an example (very quickly !)

$$(\texttt{2.} \preceq u \preceq \texttt{10.} \wedge \texttt{2.} \preceq v \preceq \texttt{10.}) \Rightarrow (\overline{u \oplus v}) - (\overline{u} + \overline{v}) \leq 0.00000096$$

Our approach on an example (very quickly !)

$$(2. \preceq u \preceq 10. \wedge 2. \preceq v \preceq 10.) \Rightarrow (\overline{u \oplus v}) - (\overline{u} + \overline{v}) \leq 0.00000096$$

Layers	axioms	reasoners	deductions

Our approach on an example (very quickly !)

$$(2. \preceq u \preceq 10. \wedge 2. \preceq v \preceq 10.) \Rightarrow (\overline{u \oplus v}) - (\overline{u} + \overline{v}) \leq 0.00000096$$

$$\text{L1-3}: \forall x. \forall y.\ x \preceq y \implies \bigvee \left(\begin{array}{l} \texttt{is\_finite}(x) \wedge \texttt{is\_finite}(y) \\ \texttt{is\_infinite}(x) \wedge \texttt{is\_negative}(x) \wedge \neg\texttt{is\_nan}(y) \\ \texttt{is\_infinite}(y) \wedge \texttt{is\_positive}(y) \wedge \neg\texttt{is\_nan}(x) \end{array} \right)$$

Layers	axioms	reasoners	deductions
0	H		$\texttt{is\_finite}(2.)$, $\texttt{is\_finite}(10.)$
1	L1-3	EM, SAT	$\texttt{is\_finite}(u) \vee$ $(\texttt{is\_infinite}(u) \wedge \texttt{is\_negative}(u))$
1	L1-3	EM, SAT	$\texttt{is\_finite}(v) \vee$ $(\texttt{is\_infinite}(v) \wedge \texttt{is\_negative}(v))$

Our approach on an example (very quickly !)

$$(2. \preceq u \preceq 10. \wedge 2. \preceq v \preceq 10.) \Rightarrow (\overline{u \oplus v}) - (\overline{u} + \overline{v}) \leq 0.00000096$$

$$\text{L1-6}: \quad \forall x. \ \neg\,(\mathtt{is\_negative}(x) \wedge \mathtt{is\_positive}(x))$$

Layers	axioms	reasoners	deductions
			$\mathtt{is\_finite}(2.)$, $\mathtt{is\_finite}(10.)$
			$\mathtt{is\_finite}(u) \vee (\mathtt{is\_infinite}(u) \wedge \mathtt{is\_negative}(u))$
			$\mathtt{is\_finite}(v) \vee (\mathtt{is\_infinite}(v) \wedge \mathtt{is\_negative}(v))$
1	L1-6	EM, SAT	$\mathtt{is\_finite}(u)$
1	L1-6	EM, SAT	$\mathtt{is\_finite}(v)$

Our approach on an example (very quickly !)

$$(2. \preceq u \preceq 10. \land 2. \preceq v \preceq 10.) \Rightarrow (\overline{u \oplus v}) - (\overline{u} + \overline{v}) \leq 0.00000096$$

$$\text{L1-4}: \quad \forall x. \forall y. \ (\texttt{is\_finite}(x) \land \texttt{is\_finite}(y) \land x \preceq y) \implies \overline{x} \leq \overline{y}$$

Layers	axioms	reasoners	deductions
			$\texttt{is\_finite}(u)$
			$\texttt{is\_finite}(v)$
1	L1-4	EM, SAT	$2 \leq \overline{u}$
1	L1-4	EM, SAT	$2 \leq \overline{v}$
1	L1-4	EM, SAT	$\overline{u} \leq 10$
1	L1-4	EM, SAT	$\overline{v} \leq 10$

Our approach on an example (very quickly !)

$$(2. \preceq u \preceq 10. \land 2. \preceq v \preceq 10.) \Rightarrow (\overline{u \oplus v}) - (\overline{u} + \overline{v}) \leq 0.00000096$$

L1-2 :

$$\forall m.\forall x.\forall y.\ (\mathtt{is\_finite}(x) \land \mathtt{is\_finite}(y) \land \mathtt{in\_range}(\circ_m(\overline{x} + \overline{y}))) \implies \overline{x \oplus_m y} = \circ_m(\overline{x} + \overline{y})$$

Layers	axioms	reasoners	deductions
			$\mathtt{is\_finite}(u)$
			$\mathtt{is\_finite}(v)$
			$2 \leq \overline{u}$
			$2 \leq \overline{v}$
			$\overline{u} \leq 10$
			$\overline{v} \leq 10$
5	L1-2	EM, SAT	$\mathtt{in\_range}(\circ(\overline{u} + \overline{v})) \Rightarrow \overline{u \oplus v} = \circ(\overline{u} + \overline{v})$

Our approach on an example (very quickly !)

$$(\texttt{2.} \preceq u \preceq \texttt{10.} \land \texttt{2.} \preceq v \preceq \texttt{10.}) \Rightarrow (\overline{u \oplus v}) - (\overline{u} + \overline{v}) \;\leq\; 0.00000096$$

Layers	axioms	reasoners	deductions
			$2 \leq \overline{u} \leq 10, \quad 2 \leq \overline{v} \leq 10$
			$\texttt{in\_range}(\circ(\overline{u}+\overline{v})) \Rightarrow \overline{u \oplus v} = \circ(\overline{u}+\overline{v})$
3		NRA	$\overline{u}+\overline{v} \in [4;20]$

Our approach on an example (very quickly !)

$$(2. \preceq u \preceq 10. \wedge 2. \preceq v \preceq 10.) \Rightarrow (\overline{u \oplus v}) - (\overline{u} + \overline{v}) \leq 0.00000096$$

$$\text{L2-2}: \quad \forall z.\forall i.\forall j. \quad i \leq z \leq j \implies -2^{\alpha} \leq \circ(z) - z \leq 2^{\alpha}$$

$$\text{where} \quad \alpha = \mathtt{ilog}_2(\max(|i|, |j|, 2^{e_{\min}+prec-1})) - prec$$

Layers	axioms	reasoners	deductions
			$\mathtt{in\_range}(\circ(\overline{u} + \overline{v})) \Rightarrow \overline{u \oplus v} = \circ(\overline{u} + \overline{v})$
			$\overline{u} + \overline{v} \in [4; 20]$
2	L2-2	EM, IM, SAT	$-2^{-20} \leq \circ(\overline{u} + \overline{v}) - (\overline{u} + \overline{v}) \leq 2^{-20}$

Our approach on an example (very quickly !)

$$(\texttt{2.} \preceq u \preceq \texttt{10.} \wedge \texttt{2.} \preceq v \preceq \texttt{10.}) \Rightarrow (\overline{u \oplus v}) - (\overline{u} + \overline{v}) \ \leq \ 0.00000096$$

Layers	axioms	reasoners	deductions
			$\texttt{in\_range}(\circ(\overline{u}+\overline{v})) \Rightarrow \overline{u \oplus v} = \circ(\overline{u}+\overline{v})$
			$\overline{u}+\overline{v} \in [4;20]$
			$-2^{-20} \leq \circ(\overline{u}+\overline{v}) - (\overline{u}+\overline{v}) \leq 2^{-20}$
3		NRA	$4-2^{-20} \leq \circ(\overline{u}+\overline{v}) \leq 20+2^{-20}$

Our approach on an example (very quickly !)

$$(\texttt{2.} \preceq u \preceq \texttt{10.} \land \texttt{2.} \preceq v \preceq \texttt{10.}) \Rightarrow (\overline{u \oplus v}) - (\overline{u} + \overline{v}) \leq 0.00000096$$

L1-1 : $\forall z.\ \texttt{in\_range}(z) \iff -\text{0x1.FFFFFEp127} \leq z \leq \text{0x1.FFFFFEp127}$

Layers	axioms	reasoners	deductions
			$\texttt{in\_range}(\circ(\overline{u} + \overline{v})) \Rightarrow \overline{u \oplus v} = \circ(\overline{u} + \overline{v})$
			$-2^{-20} \leq \circ(\overline{u} + \overline{v}) - (\overline{u} + \overline{v}) \leq 2^{-20}$
			$4 - 2^{-20} \leq \circ(\overline{u} + \overline{v}) \leq 20 + 2^{-20}$
$1 + 3$	L1-1	EM, SAT, NRA	$\texttt{in\_range}(\circ(\overline{u} + \overline{v}))$

Our approach on an example (very quickly !)

$$(\texttt{2.} \preceq u \preceq \texttt{10.} \wedge \texttt{2.} \preceq v \preceq \texttt{10.}) \Rightarrow (\overline{u \oplus v}) - (\overline{u} + \overline{v}) \ \leq \ 0.00000096$$

Layers	axioms	reasoners	deductions
			$\texttt{in\_range}(\circ(\overline{u}+\overline{v})) \Rightarrow \overline{u \oplus v} = \circ(\overline{u}+\overline{v})$
			$-2^{-20} \leq \circ(\overline{u}+\overline{v}) - (\overline{u}+\overline{v}) \leq 2^{-20}$
			$\texttt{in\_range}(\circ(\overline{u}+\overline{v}))$
		SAT	$\overline{u \oplus v} = \circ(\overline{u}+\overline{v})$

Our approach on an example (very quickly !)

$$(2. \preceq u \preceq 10. \land 2. \preceq v \preceq 10.) \Rightarrow (\overline{u \oplus v}) - (\overline{u} + \overline{v}) \ \leq \ 0.00000096$$

Layers	axioms	reasoners	deductions
			$-2^{-20} \leq \circ(\overline{u} + \overline{v}) - (\overline{u} + \overline{v}) \leq 2^{-20}$
			$\overline{u \oplus v} = \circ(\overline{u} + \overline{v})$
3		NRA	$\overline{u \oplus v} - (\overline{u} + \overline{v}) \leq 2^{-20}$

Our approach on an example (very quickly !)

$$(2. \preceq u \preceq 10. \wedge 2. \preceq v \preceq 10.) \Rightarrow (\overline{u \oplus v}) - (\overline{u} + \overline{v}) \;\leq\; 0.00000096$$

Layers	axioms	reasoners	deductions
			$\overline{u \oplus v} - (\overline{u} + \overline{v}) \leq 2^{-20} < 0.00000096$

Main ingredient : intervals matching

$$\forall z. \forall i. \forall j. \;\; i \le z \le j \implies -2^{\alpha} \;\le\; \mathtt{round}(z) - z \;\le\; 2^{\alpha}$$

where $\alpha \equiv \mathtt{ilog}_2(\max(|i|,\, |j|,\, 2^{e_{\min}+prec-1})) - prec$

Main ingredient : intervals matching

$$\forall z.\forall i.\forall j.\ \ i \leq z \leq j \implies -2^{\alpha} \ \leq\ \mathtt{round}(z) - z \ \leq\ 2^{\alpha}$$

where $\alpha \equiv \mathtt{ilog}_2(\max(|i|,\ |j|,\ 2^{e_{\min}+prec-1})) - prec$

- To handle universally quantified formulas, techniques based on matching need patterns (that cover all quantified variables)
  (eg. $\{\mathtt{round}(z), i, j\}$ or $\{\mathtt{round}(z), \mathtt{abs}(i), \mathtt{abs}(j))\}$

Main ingredient : intervals matching

$$\forall z.\forall i.\forall j.\ \ i \le z \le j \implies -2^{\alpha} \ \le\ \mathtt{round}(z) - z \ \le\ 2^{\alpha}$$

where $\alpha \equiv \mathtt{ilog}_2(\max(|i|,\ |j|,\ 2^{e_{\min}+prec-1})) - prec$

- To handle universally quantified formulas, techniques based on matching need patterns (that cover all quantified variables)
  (eg. $\{\mathtt{round}(z), i, j\}$ or $\{\mathtt{round}(z), \mathtt{abs}(i), \mathtt{abs}(j))\}$
- Syntactic patterns are not well suited to instantiate axioms about rounding properties

Main ingredient : intervals matching

$$\forall z. \forall i. \forall j. \;\; i \leq z \leq j \implies -2^{\alpha} \;\leq\; \mathtt{round}(z) - z \;\leq\; 2^{\alpha}$$

where $\alpha \equiv \mathtt{ilog}_2(\max(|i|,\, |j|,\, 2^{e_{\min}+prec-1})) - prec$

Solution :

- Use a mix of syntactic and semantic patterns to cover universally quantified variables

$$(\text{eg.} \quad \{\, \mathtt{round}(z) \;, \quad z \in [i, j] \,\} \;)$$

- Use intervals information to find relevant instances for variables of semantic patterns (i and j in the example)

Main ingredient : intervals matching

$$\forall z.\forall i.\forall j.\;\; i \le z \le j \implies -2^{\alpha} \;\le\; \texttt{round}(z) - z \;\le\; 2^{\alpha}$$

where $\alpha \equiv \texttt{ilog}_2(\max(|i|, |j|, 2^{e_{\min}+prec-1})) - prec$

Main steps to handle rounding properties

1. annotate them with a mix of syntactic and semantic triggers
2. use generic E-matching with syntactic triggers
3. use intervals matching with semantic triggers and compute upper/lower bounds for terms (on demands)
4. generate ground instances
5. simplify instances (eg. 2^{α} in the axiom will reduce to a constant since i and j will be instantiated by constants)

Evaluation : benchmarks & solvers

307 VCs	C	(S. Boldo, C. Marché)	$\forall, \exists$
1980 VCs	SPARK	(AdaCore)	$\forall, \exists$
20035 VCs	SMTLIB2	(Wintersteiger Unsat)	$\forall, \exists$-Free
114 VCs	SMTLIB2	(Griggio Unsat+Unknown)	$\forall, \exists$-Free

- We are interested in showing unsatisfiability / validity
- Alt-Ergo without FPA reasoning does not prove any VC of these benchmarks

Evaluation : benchmarks & solvers

- Alt-Ergo : Axiomatic FPA (+ eventually other axioms for C and SPARK VCs) + rounding properties

- Z3 : pure QF_FP for SMT-benchmarks, a combination of FP with other theories and quantified axioms for C and SPARK

- Gappa :[1] Axioms of Layer 1, and other quantified axioms are instantiated (best effort), and then abstracted. Non arithmetic constructs are also abstracted

- MS+A and MS+B : like Z3, but quantified axioms are instantiated (best effort) and then abstracted

(See more details in the paper)

1. our approach is inspired by Gappa

Results : C benchmarks

Time limit $= 60$ seconds, Memory limit $= 3$ GB

	CMP-1		CMP-2		
	AE	Z3	Gappa	MS5+B	MS5+A
proved	194	2	**199**	4	2
time	566	< 1	**78**	4	< 1
	230/307 proved with at least one solver				

- The ACSL specification of the C programs uses Reals
- A combination of FPA with Reals is needed to prove the VCs

Results : SPARK benchmarks

Time limit = 60 seconds, Memory limit = 3 GB

	CMP-1		CMP-2		
	AE	Z3	Gappa	MS5+B	MS5+A
proved	**806**	720	488	170	13
time	3090	4142	305	301	1
	1136/1980 proved with at least one solver				

- The SPARK specification uses FPA (Z3 performs better compared to C benchs)
- Techniques are "complementary"

Results : Wintersteiger unsat

Time limit = 60 seconds, Memory limit = 3 GB

	CMP-3	CMP-4	CMP-5		
	AE	Gappa	Z3	MS5+B	MS5+A
proved	19863	18102	**20035**	17201	17200
time	876	44	**65**	66	63
	20035/20035 proved with at least one solver				

- We don't prove 172 VCs because the intervals we compute for square root are not accurate

Results : Griggio unsat+unknown

Time limit = 60 seconds, Memory limit = 3 GB

	CMP-3		CMP-4	CMP-5		
	AE		Gappa	Z3	MS5+B	MS5+A
proved	2		-	**50**	49	5
time	18		-	1337	723	1
	57/114 proved with at least one solver					

- AE's instantiation engine overburdens the SAT solver with plenty of instances from Layer 1, while only some instances and a lot of learning, simplifications and SAT propagations would allow to prove the VCs

Results : Griggio unsat+unknown

Time limit = 60 seconds, Memory limit = 3 GB

	CMP-3		CMP-4	CMP-5		
	AE	AE+ CDCL	Gappa	Z3	MS5+B	MS5+A
proved	2	37	-	**50**	49	5
time	18	733	-	1337	723	1
	57/114 proved with at least one solver					

- AE's instantiation engine overburdens the SAT solver with plenty of instances from Layer 1, while only some instances and a lot of learning, simplifications and SAT propagations would allow to prove the VCs

Conclusion

Pros

- Good results, in particular on VCs coming from deductive programs verification (C and SPARK)
- The technique is complementary compared to others
- Lightweight and non-intrusive extension. Most of added code is not critical (for soundness)

Cons

- SAT benchs are not in the scope of the method (but not an issue for deductive program verification)

Possible/Further improvements

- Inline/mechanize reasoning about (some) axioms of Layers 1 and/or 2