The Strategy Challenge in SMT Solving

Citation for published version:
de Moura, L & Passmore, G 2013, The Strategy Challenge in SMT Solving. in M Bonacina & M Stickel (eds), Automated Reasoning and Mathematics: Essays in Memory of William W. McCune. vol. 7788, Lecture Notes in Computer Science, vol. 7788, Springer Berlin Heidelberg, pp. 15-44. DOI: 10.1007/978-3-642-36675-8_2

Digital Object Identifier (DOI):
10.1007/978-3-642-36675-8_2

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Early version, also known as pre-print

Published In:
Automated Reasoning and Mathematics

The Strategy Challenge in SMT Solving

Leonardo de Moura[1] and Grant Olney Passmore[2,3]
leonardo@microsoft.com, gp351@cam.ac.uk

[1] Microsoft Research, Redmond
[2] Clare Hall, University of Cambridge
[3] LFCS, University of Edinburgh

Abstract. High-performance SMT solvers contain many tightly integrated, hand-crafted heuristic combinations of algorithmic proof methods. While these heuristic combinations tend to be highly tuned for known classes of problems, they may easily perform badly on classes of problems not anticipated by solver developers. This issue is becoming increasingly pressing as SMT solvers begin to gain the attention of practitioners in diverse areas of science and engineering. We present a challenge to the SMT community: to develop methods through which users can exert strategic control over core heuristic aspects of SMT solvers. We present evidence that the adaptation of ideas of strategy prevalent both within the Argonne and LCF theorem proving paradigms can go a long way towards realizing this goal.

1 Introduction

SMT (Satisfiability Modulo Theories) solvers are a powerful class of automated theorem provers which have in recent years seen much academic and industrial uptake [11]. They draw on some of the most fundamental discoveries of computer science and symbolic logic. They combine the Boolean Satisfiability problem with the decision problem for concrete domains such as arithmetical theories studied in convex optimization and term-manipulating symbolic systems. They involve, in an essential way, decision problems, completeness and incompleteness of logical theories and complexity theory.

The standard account of modern SMT solver architecture is given by the so-called *DPLL(T) scheme* [23]. DPLL(T) is a theoretical framework, a rule-based formalism describing, abstractly, how satellite *theory solvers* (T-solvers) for decidable theories such as linear integer arithmetic, arrays and bit-vectors are to be integrated together with DPLL-based SAT solving. Decision procedures (complete T-solvers) for individual theories are combined by the DPLL(T) scheme in such a way that guarantees their combination is a complete decision procedure as well. Because of this, one might get the impression that *heuristics* are not involved in SMT. However, this is not so: heuristics play a vital role in high-performance SMT, a role which is all too rarely discussed or championed.

By design, DPLL(T) abstracts away from many practical implementation issues. High-performance SMT solvers contain many tightly integrated, hand-crafted heuristic combinations of algorithmic proof methods which fall outside the scope of DPLL(T). We shall discuss many examples of such heuristics in this paper, with a focus on our tools **RAHD** [24] and **Z3** [10]. To mention but one class of examples, consider formula pre-processing. This is a vital, heavily heuristic component of modern SMT proof search which occurs outside the purview of DPLL(T). We know of no two SMT solvers which handle formula pre-processing in exactly the same manner. We also know of no tool whose documentation fully describes the heuristics it uses in formula pre-processing, let alone gives end-users principled methods to control these heuristics.

While the heuristic components of SMT solvers tend to be highly tuned for known classes of problems (e.g., SMT-LIB [3] benchmarks), they may easily perform very badly on new classes of problems not anticipated by solver developers. This issue is becoming increasingly pressing as SMT solvers begin to gain the attention of practitioners in diverse areas of science and engineering. In many cases, changes to the prover heuristics can make a tremendous difference in the success of SMT solving within new problem domains. Classically, however, much of the control of these heuristics has been outside the reach[4] of solver end-users[5]. We would like much more control to be placed in the hands of end-users, and for this to be done in a principled way.

We present a challenge to the SMT community: to develop principled methods through which users can exert strategic control over core heuristic aspects of SMT solvers. We present evidence, through research undertaken with the tools **RAHD** and **Z3**, that the adaptation of a few basic ideas of strategy prevalent both within the Argonne and LCF theorem proving paradigms can go a long way towards realizing this goal. In the process, we solidify some foundations for strategy in the context of SMT and pose a number of questions and open problems. We state this challenge as follows:

The Strategy Challenge

> To build theoretical and practical tools allowing users to exert strategic control over core heuristic aspects of high-performance SMT solvers.

In this way, the proof procedure may be tailored to specific problem domains, especially ones very different from those normally considered.

[4] Some SMT solvers expose a vast collection of parameters to control certain behavioral aspects of their core proof procedures. We view these parameters as a primitive way of exerting strategic control over the heuristic aspects of high-performance SMT solvers. As the use of SMT solvers continues to grow and diversify, the number of these options has steadily risen in most solvers. For instance, the number of such options in **Z3** has risen from 40 (v1.0) to 240 (v2.0) to 284 (v3.0). Many of these options have been requested by users. Among users, there seems to be a wide-spread wish for more methods to exert strategic control over the prover's reasoning.

[5] We use the term *end-user* to mean a user of an SMT solver who does not contribute to the essential development of such a solver. End-users regularly embed SMT solvers into their own tools, making SMT solvers a subsidiary theorem proving engine for larger, specialised verification tool-chains.

1.1 Caveat Emptor: What This Paper Is and Is Not

We find it prudent at the outset to make clear what this paper is and is not. Let us first enumerate a few things we shall not do:

– We shall not give a new *theoretical framework* or *rule-based formalism* capturing the semantics of heuristic proof strategies, as is done, for instance, in the influential STRATEGO work on term rewriting strategies [20].
– We shall not prove any theorems about the algebraic structure underlying the collection of heuristic proof strategies, as is done, for instance, in the insightful work on the *proof monad* for interactive proof assistants [18].
– We shall not propose a concrete syntax for heuristic proof strategies in the context of SMT, as, for instance, an extension of the SMT-LIB standard [3].

We shall not attempt any of the (worthy) goals mentioned above because we believe to do so would be premature. It is simply too early to accomplish them in a compelling way. For example, before a standard concrete syntax is proposed for heuristic SMT proof strategies, many different instances of strategy in SMT must be explicated and analyzed, in many different tools, from many contrasting points of view. Before any theorems are proved about the collection of heuristic SMT proof strategies, we must have a firm grasp of their scope and limitations. And so on. Instead, our goals with this Strategy Challenge are much more modest:

– To bring awareness to the crucial role heuristics play in high-performance SMT, and to encourage researchers in the field to be more explicit as to the heuristic strategies involved in their solvers.
– To convince SMT solver developers that providing flexible methods (i.e., a *strategy language*) for end-users to exert fine-grained control over heuristic aspects of their solvers is an important, timely and worthy undertaking.
– To show how the adaptation of some ideas of strategy prevalent both within the Argonne and LCF theorem proving paradigms can go a long way towards realizing these goals.
– To stress that from a scientific point of view, the explication of the actual heuristic strategies used in high-performance SMT solvers is absolutely crucial for enabling the reproducibility of results presented in publications. For instance, if a paper reports on a new decision procedure, including experimental findings, and these experiments rely upon an implementation of the described decision method incorporating some heuristic strategies, then these heuristics should be communicated as well.

Throughout this paper, we shall present many examples in (variants of) the concrete syntaxes we have developed for expressing heuristic proof strategies in our own tools. These are given to lend credence to the practical value of this challenge, not as a proposal for a standard strategy language.

§

As we shall see, our initial approach for meeting this Strategy Challenge is based on an SMT-specific adaptation of the ideas of *tactics* and *tacticals* as found in LCF-style [22,27] proof assistants. This adaptation has a strong relationship with the approach taken by the Argonne group on *theorem proving toolkits*, work that began with the Argonne systems NIUTP1 - NIUTP7 in the 1970s and early 1980s, and has continued through to modern day with Bill McCune's LADR (Library for Automated Deduction Research), the foundation of his powerful Prover9 and MACE4 tools [21]. The way in which we adapt ideas of tactics and tacticals to SMT results in notions of strategy which, though borrowing heavily from both of these sources, are quite distinct from those found in LCF-style proof assistants and Argonne-style theorem provers.

Finally, let us assure the reader that we are *not* proposing a tactic-based approach for implementing "low-level" aspects of SMT solvers such as unit propagation, nor for implementing core reasoning engines such as a SAT solver or Simplex. This would hardly be practical for high-performance SMT. Again, our goals are more modest: We are proposing only the use of this strategy machinery for facilitating the orchestration of "big" reasoning engines, that is, for prescribing heuristic combinations of procedures such as Gaussian Elimination, Gröbner bases, CNF encoders, SAT solvers, Fourier-Motzkin quantifer elimination and the like. In this way, "big" symbolic reasoning steps will be represented as tactics heuristically composable using a language of tacticals.

While we do not propose a particular concrete strategy language for SMT solvers, we will present some key features we believe a reasonable SMT strategy language should have. These features have been implemented in our tools, and examples are presented. One important requirement will be that a strategy language support methods for conditionally invoking different reasoning engines based upon features of the formula being analyzed. We shall exploit this ability heavily. In some cases, when high-level "big" engines contain within themselves heavily heuristic components made up of combinations of other "big" reasoning engines, we also propose that these components be modularized and made replaceable by user-specified heuristic proof strategies given as parameters to the high-level reasoning engine. These ideas will be explained in detail in Sec. 3, where we give the basis of what we call *big-step strategies* in SMT, and in Sec. 4, where we observe that certain strategically parameterized reasoning engines can be realized as tacticals.

1.2 Overview

Finally, let us end this introduction with an overview of the rest of the paper. In Sec. 2 we begin with the question *What is strategy?* and explore some possible answers by recalling important notions of strategy in the history of automated theorem proving. In Sec. 3, we give the foundations of big step strategies in SMT. In Sec. 4, we observe that there is a natural view of some reasoning engines as tacticals, and we present a few examples of this in the context of **RAHD** and **Z3**. In Sec. 5, we show some promising results of implementations of many of these strategy ideas within the two tools. Finally, we conclude in Sec. 6 with a look towards the future.

2 Strategy in Mechanized Proof

There exists a rich history of ideas of *strategy* in the context of mechanized proof. In this section, we work to give a high-level, hopelessly incomplete survey of the many roles strategy has played in automated proof. In the process, we shall keep an eye towards why many of the same issues which motivated members of the non-SMT-based automated reasoning community to develop powerful methods for user-controllable proof search strategies also apply, compellingly, to the case of SMT.

2.1 What is Strategy?

Before discussing strategy any further, we should certainly attempt to define it. *What is strategy, after all?* Even restricted to the realm of mechanized proof, this question is terribly difficult to answer. There are so many aspects of strategy pervasive in modern proof search methods, and there seems to be no obvious precise delineations of their boundaries. Where does one, for instance, draw the line between the “strategic enhancement” of an existing search algorithm and the advent of a new search algorithm altogether?

Despite the difficulties fundamental to defining precisely what *strategy* is, many researchers in automated reasoning have proposed various approaches to incorporating strategy into automated proof. Some approaches have been quite foundational in character, shedding light on the nature of strategy in particular contexts. For instance, in the context of term rewriting, the ideas found within the STRATEGO system have given rise to much fruit, both theoretical and applied, and elucidated deep connections between term rewriting strategy and concurrency theory [20]. In the context of proof strategies in interactive theorem provers, the recent work of Kirchner-Muñoz has given heterogeneous proof strategies found in proof assistants like PVS a firm algebraic foundation using the category-theoretic notion of a monad [18].

Though a general definition of *what strategy is* seems beyond our present faculties, we may still make some progress by describing a few of its most salient aspects. In particular, the following two statements seem a reasonable place to begin:

1. There is a natural view of automated theorem proving as being an exercise in combinatorial search.
2. With this view in mind, then *strategy* may be something like *adaptations of general search mechanisms which reduce the search space by tailoring its exploration to a particular class of problems.*

We are most interested in these adaptations when end-users of theorem proving tools may be given methods to control them.

To gain a more concrete understanding of the importance of strategy in automated theorem proving, it is helpful to consider some key examples of its use. In working to gather and present some of these examples, we have found the vast number of compelling uses of strategy to be quite staggering. There are so many examples, in fact, that it seems fair to say that much of the history of automated reasoning can be interpreted as a series of *strategic advances.* As automated reasoning is such a broad field, let us begin by focusing on the use of strategy in one of its particularly telling strands, the history of mechanized proof search in first-order predicate calculus (FOL).

§

When the field of first-order proof search began and core search algorithms were first being isolated, most interesting strategic advancements were so profound that we today consider them to be the advent of genuinely "new" theorem proving methods. For example, both the Davis-Putnam procedure and Robinson's resolution can be seen to be strategic advancements for the general problem of first-order proof search based on Herbrand's Theorem. But, compared to their predecessors, the changes these strategic enhancements brought forth were of such a revolutionary nature that we consider them to be of quite a different kind than the strategies we want to make user-controllable in the context of SMT.

Once resolution became a dominant focus of first-order theorem proving, however, then more nuanced notions of strategy began to take hold, with each of them using resolution as their foundation. Many of the most lasting ideas in this line of research were developed by the Argonne group. These ideas, including the set of support, term weighting, pick/given ratio, hot lists and hints, did something very interesting: They provided a fixed, general algorithmic search framework upon which end-users could exert some of their own strategic control by prescribing restrictions to guide the general method. Moreover, beginning in the 1970s, the Argonne group promoted the idea of *theorem proving toolkits*, libraries of high-performance reasoning engines one could use to build customized theorem provers. This idea has influenced very much the approach we propose for strategy in SMT.

Let us now turn, in more detail, to uses of strategy in the history of mechanized first-order proof. Following this, we shall then consider some aspects of strategy in the context of LCF-style interactive proof assistants. Ideas of strategy from both of these histories will play into our work.

2.2 Strategy in Automated Theorem Proving over FOL

One cannot attempt to interpret the history of mechanized proof in first-order logic without bearing in mind the following fact: Over two decades before the first computer experiments with first-order proof search were ever performed, the undecidability of FOL was established. This result was well-known to the logicians who began our field. Thankfully, this seemingly negative result was tempered with a more optimistic one: the fact that FOL is *semi-decidable*. This allowed programs such as the British Museum Algorithm (and quickly others of a more sophisticated nature) to be imagined which, in principle, will always find a proof of a conjecture C over an axiomatic theory T if C is in fact true in all models of T.

Early in the field, and long before the advent of algorithmic complexity theory in any modern sense, obviously infeasible approaches like the British Museum were recognized as such. Speaking of the earliest (1950s) research in computer mechanized proof, Davis writes in "The Early History of Automated Deduction" [7]:

> [...] it was all too obvious that an attempt to generate a proof of something non-trivial by beginning with the axioms of some logical system and systematically applying the rules of inference in all possible directions was sure to lead to a gigantic combinatorial explosion.

Thus, though a semi-complete theorem proving method was known (and such a method is in a sense "best possible" from the perspective of computability theory), its search strategy was seen as utterly hopeless. In its place, other search strategies were sought in order to make the theorem proving effort more tractable. This point of view was articulated at least as early as 1958 by Hao Wang, who writes in [32]:

> Even though one could illustrate how much more effective partial strategies can be if we had only a very dreadful general algorithm, it would appear desirable to postpone such considerations till we encounter a more realistic case where there is no general algorithm or nor efficient general algorithm, e.g., in the whole predicate calculus or in number theory. As the interest is presumably in seeing how well a particular procedure can enable us to prove theorems on a machine, it would seem preferable to spend more effort on choosing the more efficient methods rather than on enunciating more or less familiar generalities.

At the famous 1957 five week Summer Institute for Symbolic Logic held at Cornell University, the logician Abraham Robinson[6] gave a remarkably influential short talk [30] in which he singled out Skolem functions and Herbrand's Theorem as potentially useful tools for general-purpose first-order theorem provers [7]. Aspects of this suggestion were taken up by many very quickly, notably Gilmore [14], Davis and Putnam [9], and eventually J.A. Robinson [31]. Let us examine, from the perspective of strategy, a few of the main developments in this exceedingly influential strand.

As noted by Davis [7], the first Herbrand-based theorem provers for FOL employed completely *unguided* search of the Herbrand universe. There was initially

[6] It is useful to note that this Abraham Robinson, the model theorist and inventor of non-standard analysis, is not the same person as John Alan Robinson who would some 8 years later invent the proof search method of first-order resolution.

neither a top-level conversion to a normal form such as CNF nor a systematic use of Skolem functions. When these first Herbrand-based methods were applied, through for instance the important early implementation by Gilmore [14], they proved unsuccessful for all but the simplest of theorems. Contemporaneously, Prawitz observed the same phenomena through his early work on a prover based on a modified semantic tableaux [7]. The lesson was clear: unguided search, even when based on deep results in mathematical logic, is not a viable approach. Again, new *strategies* were sought for controlling search space exploration.

The flurry of theorem proving breakthroughs in the early 1960s led to a wealth of new search strategies (and new notions of strategy) which still form the foundation for much of our field today.

First, based on shortcomings they observed in Gilmore's unguided exploration of the Herband universe, in particular the reliance of his method upon a DNF conversion for (what we now call) SAT solving, Davis and Putnam devised a new Herbrand universe exploration strategy which systematically applied Skolemization to eliminate existential quantifiers and used a CNF input formula representation as the basis to introduce a number of enhanced techniques for recognising the unsatisfiability of ground instances [9]. In the process, they spent much effort on enhancing the tractable recognition of ground unsatisfiability, which they believed at the time to be the biggest practical hurdle in Herbrand-based methods [7]. When further enhanced by Davis, Logemann and Loveland, this propositional component of Davis and Putnam's general first-order search strategy gave us what we now call DPLL, the foundation of modern SAT solving [8]. Nevertheless, once implementations were undertaken and experiments performed, the power of their first-order method was still found completely unsatisfactory. As Davis states [7],

> Although testing for satisfiability was performed very efficiently, it soon became clear that no very interesting results could be obtained without first devising a method for avoiding the generation of spurious elements of the Herbrand universe.

In the same year, Prawitz devised an "on-demand" method by which the generation of unnecessary terms in Herbrand expansions could be avoided, at the cost of sequences of expensive DNF conversions [29]. Though these DNF conversions precluded the practicality of Prawitz's method, his new approach made clear the potential utility of unification in Herbrand-based proof search, and Davis soon proposed [6,7] that effort be put towards

> ... a new kind of procedure which seeks to combine the virtues of the Prawitz procedure and those of the Davis Putnam procedure.

Two years later, Robinson published his discovery of such a method: Resolution, a single, easily mechanizable inference rule (refutationally) complete for FOL. This new method soon became a dominant high-level strategy for first-order proof search [31]. That resolution was a revolutionary improvement over previous methods is without question. But, practitioners soon discovered that the game was by no means won. Even with resolution in hand, the act of proof search was utterly infeasible for the vast majority of nontrivial problems without the introduction of some techniques for guiding the generation of resolvents. It is here that a new class of strategies was born.

In the 1960s, the group centered around Larry Wos at Argonne National Laboratory contributed many fundamental developments to first-order proving. At Argonne, Robinson made his discovery of resolution, first communicated in a privately circulated technical report in 1964, and then published in his influential JACM paper the year that followed. Almost immediately, Wos and his colleagues championed the importance of user-controllable strategies during resolution proof search and began developing methods for their facilitation. At least two important papers in this line were published by the end of 1965: "The Unit Preference Strategy in Theorem Proving" [34] and "Efficiency and Completeness of the Set of Support Strategy in Theorem Proving" [35]. The first gave rise to a strategy that the resolution prover would execute without any influence from the user. The second, however, introduced a strategy of a different kind: a method by which end-users could exert strategic control over the proof search without changing its underlying high-level method or impinging upon its completeness.

In resolution, the idea of set of support is to partition the CNF representation of the negated input formula into two sets of clauses, a satisfiable set A and another set S, and then to restrict the generation of clauses to those who can trace their lineage back to at least one member of S. Importantly, the choice of this division of the input clauses between A and S may be chosen strategically by the end-user. This general approach to heuristic integration — parameterizing a fixed proof search method by user-definable data — has proven enormously influential. Other methods in this class include clause weighting, the use of term orderings to guide rewriting, hot lists, pick/given ratio, and many more.

Beginning in the 1970s, the Argonne group made an important methodological decision. This was the introduction of *theorem proving toolkits.* As Lusk describes [19],

> The notion of a toolkit with which one could build theorem provers was introduced at this time and became another theme for later Argonne development. In this case the approach was used to build a series of systems incorporating ever more complex variations on the closure algorithm without changing the underlying data structures and inference functions. The ultimate system (NIUTP7) provided a set of user-definable theorem-proving "environments," each running a version of the closure algorithm with different controlling parameters, and a meta-language for controlling their interactions. There was enough control, and there were enough built-in predicates, that it became possible to "program" the theorem prover to perform a number of symbolic computation tasks. With these systems, Winker and Wos began the systematic attack on open problems [...].

Finally, we shall end our discussion of strategy in first-order provers with a few high-level observations. Given that the whole endeavor is undecidable, researchers in first-order theorem proving recognized very early that strategy must play an indispensible role in actually finding proofs. In the beginning of the field, new strategies were often so different from their predecessors that we consider them to be genuinely new methods of proof search altogether. But, once a general method such as resolution had taken hold as a dominant foundation, then strategies were sought for allowing users to control specific aspects of this fixed foundation. Abstractly, this was accomplished by providing a resolution proof search loop which accepts strategic data to be given as user-specifiable parameters.

Let us turn our attention now to another important contributor to ideas of strategy in computer-assisted proof, the LCF-style of interactive proof assistants.

2.3 Strategy in LCF-style Interactive Proof Assistants

In the field of interactive proof assistants, strategy appears in many forms. The fact that humans contribute much more to the proof development process in proof assistants than in fully automatic provers gives rise to ideas of strategy quite distinct from those encountered in our discussion of first-order provers. As above, we must limit our discussion to a very small segment of the field. Let us in this section discuss one key exemplar of strategy in proof assistants, the approach of LCF.

Strategy in the LCF Approach The original LCF was an interactive proof checking system designed by Robin Milner at Stanford in 1972. This system, so-named for its mechanization of Scott's Logic of Computable Functions, provided a proof checker with the following high-level functionality [15]:

> Proofs [were] conducted by declaring a main goal (a formula in Scott's logic) and then splitting it into subgoals using a fixed set of subgoaling commands (such as induction to generate the basis and step). Subgoals [were] either solved using a simplifier or split into simpler subgoals until they could be solved directly.

Soon after the birth of Stanford LCF, Milner moved to Edinburgh and built a team to work on its successor. A number of shortcomings had been observed in the original system. In particular, Stanford LCF embodied only one high-level proof strategy: 'backwards' proof, working from goals to subgoals. Moreover, even within a backwards proof, Stanford LCF had only a fixed set of proof construction commands which could not be easily extended [15]. Finding principled techniques to free the system from these strategic shackles became a driving motivation behind the design of Edinburgh LCF.

To address these problems, Milner devised a number of ingenious solutions which still today form the design foundation for many widely-used proof assistants. Fundamentally, Edinburgh LCF took the point of view of treating proofs as computation. New theorems were to be computed from previously established theorems by a fixed set of theorem constructors. To ensure that this computation was always done in a correct way, Milner designed an abstract datatype `thm` whose predefined values were instances of axioms and whose constructors were inference rules [15]. The idea was that strict type-checking would guarantee soundness by making sure that values of type `thm` were always actual theorems. The strictly-typed programming language ML (the **M**eta **L**anguage of Edinburgh LCF) was then designed to facilitate programming *strategies* for constructing values of type `thm`. It was a remarkable achievement. The strategic shackles of Stanford LCF had most certainly been relinquished, but much difficult work remained in order to make this low-level approach to proof construction practical. Many core strategies, both for facilitating high-level proof methods like backwards proof, as well as for implementing proof methods such as simplifiers and decision procedures needed to be built before it would be generally useful to end-users.

As mentioned, with the bare foundaton of a type `thm` and the meta language ML, the system did not directly support backwards proof. To remedy this, Milner introduced *tactics* and *tacticals*. The idea of backwards proof is that one begins with a goal, reduces it to simpler subgoals, and in the process forms a proof tree. When any of these subgoals have been made simple enough to be discharged, then a branch in the proof tree can be closed. A tactic reduces a goal to a set of subgoals such that if every subgoal holds then the goal also holds. If the set of unproved subgoals is empty, then the tactic has proved the goal. A tactic not only reduces a goal to subgoals, but it also returns a proof construction function to justify its action. *Tacticals* are combinators that treat tactics as data, and are used to construct more complex tactics from simpler ones. Gordon summarizes nicely [15]:

> By encoding the logic as an abstract type, Edinburgh LCF directly supported forward proof. The design goal was to implement goal directed proof tools by programs in ML. To make ML convenient for this, the language was made functional so that subgoaling strategies could be represented as functions (called "tactics" by Milner) and operations for combining strategies could be programmed as higher-order functions taking strategies as arguments and returning them as results (called "tacticals"). It was anticipated that strategies might fail (e.g. by being applied to inappropriate goals) so an exception handling mechanism was included in ML.

Since the time of Edinburgh LCF (and its soon-developed successor, Cambridge LCF), the technology within LCF-style proof assistant has grown considerably. However, the underlying design principles centered around user-definable proof strategies have remained more-or-less the same, and are found today in tools like Isabelle [28], HOL [16], Coq, MetaPRL and Matita. For each of these tools, immense work has gone into developing powerful tactics which embody particular proof strategies. Many of them, such as the proof producing real closed field quantifier elimination tactic in HOL-Light, are tactics embodying complete decision procedures for certain logical theories. Others, such as the implementation by Boulton of a tactic based on Boyer-Moore induction heuristics in HOL, are powerful, incomplete heuristic strategies working over undecidable theories.

There are many things to learn from the success of the LCF paradigm. One lesson relevant to our work is the following: By "opening up" strategic aspects of the proof effort and providing principled, sound programming methods through which users may write their own proof strategies, it has been possible for enormously diverse ecosystems of powerful proof strategies to be developed, contributed and shared by communities of users of LCF-style proof assistants. As these ecosystems grow, the theorem proving tools become stronger, and the realizable practical verification efforts scale up significantly more than they would if it these user-specifiable strategic enhancements were not possible.

3 Foundations for Big-Step Strategies

In this section, we propose a methodology for orchestrating reasoning engines where "big" symbolic reasoning steps are represented as functions known as *tactics*, and tactics are composed using combinators known as *tacticals*. We define notions of goals, tactics and tacticals in the context of SMT. Our definitions diverge from the ones found in LCF for the following main reasons:

– in SMT, we are not only interested in proofs of goals, but also in counterexamples (models yielding satisfying instances), and
– we want to support over and under-approximations when defining strategies.

Goals. The SMT decision problem consists of deciding whether a set of formulas S is satifisfiable or not modulo some background theory. We say each one of the formulas in S is an *assumption.* This observation suggests that a goal might be simply a collection of formulas. For practical reasons, a goal will also have a collection of attributes. Later, we describe some of the attributes we use in our systems. Thus, a goal is a pair comprised of a sequence of formulas and a sequence of attributes. Using ML-like syntax, we define this as:

$$\mathit{goal} \qquad = \mathit{formula\ sequence} \times \mathit{attribute\ sequence}$$

We say a goal is *trivially satisfiable* if the formula sequence is empty, and it is *trivially unsatisfiable* if the formula sequence constains the formula *false.* We say a goal is *basic* if it is trivially satisfiable or unsatisfiable.

Tactics. In our approach, when a tactic is applied to some goal G, four different outcomes are possible. The tactic succeeds in showing G to be satisfiable; succeeds in showing G to be unsatisfiable; produces a sequence of subgoals; or fails. A tactic returns a *model* when it shows G to be satisfiable, and a *proof* when it shows G to be unsatisfiable. When reducing a goal G to a sequence of subgoals $G_1, \ldots, G_n$, we face the problems of proof and model conversion. A *proof converter* constructs a proof of unsatisfiability for G using the proofs of unsatisfiability for all subgoals $G_1, \ldots, G_n$. Analogously, a *model converter* construct a model for G using a model for some subgoal G_i.

$$\begin{array}{lll} \textit{proofconv} & = & \textit{proof sequence} \rightarrow \textit{proof} \\ \textit{modelconv} & = & \textit{model} \times \textit{nat} \rightarrow \textit{model} \\ \textit{trt} & = & \mathsf{sat}\ \textit{model} \\ & | & \mathsf{unsat}\ \textit{proof} \\ & | & \mathsf{unknown}\ \textit{goal sequence} \times \textit{modelconv} \times \textit{proofconv} \\ & | & \mathsf{fail} \\ \textit{tactic} & = & \textit{goal} \rightarrow \textit{trt} \end{array}$$

The second parameter of a model converter is a natural number used to communicate to the model converter which subgoal was shown to be satisfiable. We intentionally did not specify how proofs and models are represented. Actually, **RAHD** and **Z3** use different representations. A proof of unsatisfiability may be a full certificate that can be independently checked by a proof checker, or it may be just a subset of the goals (also known as an unsat core) that were used to demonstrate the unsatisfiability.

Let us gain some intuition about tactics and tacticals in the context of SMT with a few simple examples.

The basic tactic returns sat and the empty model if the goal is trivially satisfiable; unsat and a proof of the form *false* $\Rightarrow$ *false* if the goal is trivially unsatisfiable; and fails otherwise. The tactic elim eliminates constants whenever the given goal contains equations of the form $a = t$, where a is a constant and t is a term not containing a. For example, suppose elim is applied to a goal containing the following sequence comprised of three formulas:

$$[\ a = b + 1,\ (a < 0 \vee a > 0),\ b > 3\ ]$$

The result will be unknown(s, mc, pc), where s is a sequence containing the single subgoal:

$$[\ (b + 1 < 0 \vee b + 1 > 0),\ b > 3\ ]$$

The model converter mc is a function s.t. when given a model M for the subgoal above, mc will construct a new model M' equal to M except that the interpretation of a in M' ($M'(a)$) is equal to the interpretation of b in M plus one (i.e., $M(b) + 1$). Similarly, the proof converter pc is a function s.t. given a proof of unsatisfiability for the subgoal will construct a proof of unsatisfiability for the original goal using the fact that $(b + 1 < 0 \vee b + 1 > 0)$ follows from $a = b + 1$ and $(a < 0 \vee a > 0)$.

The tactic split-or splits a disjunction of the form $p_1 \vee \ldots \vee p_n$ into cases and then returns n subgoals. If the disjuction to be split is not specified, the tactic splits the first disjuction occurring in the input goal. For example, given the goal G comprising of the following sequence of formulas:

$$[\ a = b + 1,\ (a < 0 \vee a > 0),\ b > 3\ ]$$

split-or G returns unknown($[G_1, G_2]$, mc, pc), where G_1 and G_2 are the subgoals comprised of the following two formula sequences respectively:

$$[\ a = b + 1,\ a < 0,\ b > 3\ ]$$

$$[\ a = b + 1,\ a > 0,\ b > 3\ ]$$

The model converter *mc* is just the identity function, since any model for G_1 or G_2 is also a model for G. The proof converter *pc* just combines the proofs of unsatisfiability for G_1 and G_2 in a straighforward way. If G does not contain a disjuction, then split-or just returns the input goal unmodified. Another option would be to fail.

RAHD and **Z3** come equipped with several built-in tactics. It is beyond the scope of this paper to document all available tactics. Nonetheless, we list some of them for didactic purposes:

– simplify: Apply simplification rules such as constant folding (e.g., $x+0 \rightsquigarrow x$).
– nnf: Put the formula sequence in negation normal form.
– cnf: Put the formula sequence in conjunctive normal form.
– tseitin: Put the formula sequence in conjunctive normal form, but use fresh Boolean constants and predicates for avoiding exponential blowup. The model converter produced by this tactic "erases" these fresh constants and predicates introduced by it.
– lift-if: Lift term if-then-else's into formula if-then-else's
– bitblast: Reduce bitvector terms into propositional logic.
– gb: Let E be the set of arithmetic equalities in a goal G, gb replaces E with the Gröbner basis induced by E.
– vts: Perform virtual term substitution.
– propagate-bounds: Perform bound propagation using inference rules such as $x < 1 \wedge y < x$ implies $y < 1$.
– propagate-values: Perform value propagation using equalities of the form $t = a$ where a is a numeral.
– split-ineqs: Split inequalities such as $t \leq 0$ into $t = 0 \vee t < 0$.
– som: Put all polynomials in sum of monomials form.
– cad: Apply cylindrical algebraic decomposition.

Tacticals. It is our hope[7] that tactics will be made available in the APIs of next generation SMT solvers. Developers of interactive and automated reasoning systems will be able to combine these tactics using their favorite programming language. Like in LCF, it is useful to provide a set of combinators (tacticals) that are used to combine built-in tactics into more complicated ones. The main advantage of using tacticals is that the resulting tactic is guaranteed to be correct, that is, it is sound if the used building blocks are sound, it connects the model converters and proof converters appropriately, and there is no need to keep track of which subgoals were already proved to be unsatisfiable. We propose the following basic tacticals:

[7] In fact, **Z3** 4.0 is now available with all of the strategy machinery described in this paper. It uses the strategy language internally and publishes a strategy API. Bindings of the strategy API are also available within Python. This Python **Z3** strategy interface can be experimented with on the web at `http://rise4fun.com/Z3Py`.

then : $(\mathit{tactic} \times \mathit{tactic}) \to \mathit{tactic}$
  then(t_1, t_2) applies t_1 to the given goal and t_2 to every subgoal produced by t_1.
then∗ : $(\mathit{tactic} \times \mathit{tactic\ sequence}) \to \mathit{tactic}$
  then∗$(t_1, [t_{2_1}, ..., t_{2_n}])$ applies t_1 to the given goal, producing subgoals $g_1, ..., g_m$. If $n \neq m$, the tactic fails. Otherwise, it applies t_{2_i} to every goal g_i.
orelse : $(\mathit{tactic} \times \mathit{tactic}) \to \mathit{tactic}$
  orelse(t_1, t_2) first applies t_1 to the given goal, if it fails then returns the result of t_2 applied to the given goal.
par : $(\mathit{tactic} \times \mathit{tactic}) \to \mathit{tactic}$
  par(t_1, t_2) executes t_1 and t_2 in parallel.
repeat : $\mathit{tactic} \to \mathit{tactic}$
  Keep applying the given tactic until no subgoal is modified by it.
repeatupto : $\mathit{tactic} \times \mathit{nat} \to \mathit{tactic}$
  Keep applying the given tactic until no subgoal is modified by it, or the maximum number of iterations is reached.
tryfor : $\mathit{tactic} \times \mathit{milliseconds} \to \mathit{tactic}$
  tryfor(t, k) returns the value computed by tactic t applied to the given goal if this value is computed within k milliseconds, otherwise it fails.

The tactic skip is the unit for then: then(skip, t) = then(t, skip) = t; and fail is the unit for orelse: orelse(fail, t) = orelse(t, fail) = t.

Formula Measures. Several SMT solvers use hard-coded strategies that perform different reasoning techniques depending on structural features of the formula being analyzed. For example, **Yices** [13] checks whether a formula is in the difference logic fragment or not. A formula is in the difference logic fragment if all atoms are of the form $x - y \bowtie k$, where x and y are uninterpreted constants, k a numeral, and $\bowtie$ is in $\{\leq, \geq, =\}$. If the formula is in the difference logic fragment, **Yices** checks if the number of inequalities divided by the number of uninterpreted constants is smaller than a threshold k. If this is the case, it uses the Simplex algorithm for processing the arithmetic atoms. Otherwise, it uses an algorithm based on the Floyd-Washall all-pairs shortest distance algorithm. We call such structural features *formula measures*. This type of ad hoc heuristic strategy based upon formula measures is very common.

We use formula measures to create Boolean expressions that are evaluated over goals. The built-in tactic failif : $\mathit{cond} \to \mathit{tactic}$ fails if the given goal does not satisfy the condition *cond*. Many numeric and Boolean measures are available in **RAHD** and **Z3**. Here is an incomplete list for illustrative purposes:

bw: Sum total bit-width of all rational coefficients of polynomials.
diff: True if the formula is in the difference logic fragment.
linear: True if all polynomials are linear.
dim: Number of uninterpreted constants (of sort real or int).
atoms: Number of atoms.
degree: Maximal total multivariate degree of polynomials.

size: Total formula size.

Using formula measures, the **Yices** strategy described above can be encoded as:

$$\mathsf{orelse}(\mathsf{then}(\mathsf{failif}(\mathsf{diff} \wedge \frac{\mathsf{atom}}{\mathsf{dim}} > \mathsf{k}),\ \mathsf{simplex}),\ \mathsf{floydwarshall})$$

Now, we define the combinators if and when based on the combinators and tactics defined so far.

$$\begin{aligned} \mathsf{if}(c,\ t_1,\ t_2) &= \mathsf{orelse}(\mathsf{then}(\mathsf{failif}(\neg c), t_1), t_2) \\ \mathsf{when}(c,\ t) &= \mathsf{if}(c,\ t,\ \mathsf{skip}) \end{aligned}$$

These are often helpful in the construction of strategies based upon formula measures.

Under and over-approximations. Under and over-approximation steps are commonly used in SMT solvers. An *under-approximation step* consists of reducing a set of formulas S to a set S' such that if S' is satisfiable, then so is S, but if S' is unsatisfiable, then nothing can be said about S. For example, any *strengthening* step that adds new formulas not deducible from S into S is an under-approximation.

A more concrete example is found in many SMT solvers for nonlinear integer arithmetic, where lower and upper bounds are added for every uninterpreted constant of sort int, and the resulting set of formulas is then reduced to SAT. Under-approximations are also used in finite model finders for first-order logic formulas, where the universe is assumed to be finite, and the first-order formula is then reduced into SAT. Analogously, an *over-approximation step* consists is reducing a set of formulas S into a set S' such that if S' is unsatisfiable, then so is S, but if S' is satisfiable, then nothing can be said about S. For example, any *weakening* step that removes formulas from S is an over-approximation. Boolean abstraction is another example used in many interactive theorem provers and SMT solvers. This comprises replacing every theory atom with a fresh proposition variable. Clearly, if the resulting set of formulas is unsatisfiable then so is the original set. Of course, given a set of formulas S, arbitrarily applying under and over-approximation steps result in set of formulas S' that cannot be used to establish the satisfiability nor the unsatisfiability of S. To prevent under and over-approximation steps from being incorrectly applied, we associate a precision attribute with every goal. A precision marker is an element of the set $\{\mathsf{prec}, \mathsf{under}, \mathsf{over}\}$. A tactic that applies an under (over) approximation fails if the precision attribute of the input goal is over (under).

4 Parametric Reasoning Engines as Tacticals

Some reasoning engines utilize other engines as subroutines. It is natural to view these higher-level reasoning engines as tacticals. Given a subsidiary engine (a tactic given to the higher-level engine as a parameter), these tacticals produce a new tactic. Let us describe two examples of such parametric engines.

Lazy SMT solvers. We observe three main phases in state-of-the-art SMT solvers: *preprocessing*, *search*, and *final check*.

Preprocessing During preprocessing, also known as pre-solving, several simplifications and rewriting steps are applied. The main goal is to put the problem in a form that is suitable for solving. Another objective is to simplify the problem, eliminate uninterpreted constants, unconstrained terms, and redundancies. Some solvers may also apply reduction techniques such as bit-blasting where bit-vector terms are reduced to propositional logic. Another commonly used reduction technique is Ackermannization [2,5] where uninterpreted function symbols are eliminated at the expense of introducing fresh constants and additional constraints.

Search During the search step, modern SMT solvers combine efficient SAT solving with "cheap" theory propagation techniques. Usually, this combination is an incomplete procedure. For example, consider problems containing arithmetic expressions. Most solvers ignore integrality and nonlinear constraints during the search step. These solvers will only propagate Boolean and linear constraints, and check whether there is a rational assignment that satisfies them. We say the solver is *postponing* the application of "expensive" and complete procedures to the final check step. Solvers, such as **Z3**, only process nonlinear constraints during final check. The word "final" is misleading since it may be executed many times for a give problem. For example, consider the following nonlinear problem comprising of three assumptions (over $\mathbb{R}$):

$$[\ x = 1,\ y \geq x + 1,\ (y \times y < 1 \vee y < 3 \vee y \times y > x + 3)\ ]$$

In the preprocessing step, a solver may decide to eliminate x using Gaussian elimination obtaining:

$$[\ y \geq 2,\ (y \times y < 1 \vee y < 3 \vee y \times y > 4)\ ]$$

During the search step, the solver performs only Boolean propagation and cheap theory propagation such as $y \geq 2$ implies $\neg(y < 3)$. Nonlinear monomials, such as $y \times y$, are treated as fresh uninterpreted constants. Thus, the incomplete solver used during the search may find the candidate assigment $y = 2$ and $y \times y = 0$. This assignment satisfies the atoms $y \geq 2$ and $y \times y < 1$, and all Boolean and linear constraints.

Final check During final check, a complete procedure for nonlinear real arithmetic is used to decide $[\ y \geq 2,\ y \times y < 1\ ]$. The complete procedure finds it to be unsatisfiable, and the solver backtracks and learns the lemma $(\neg y \geq 2 \vee y \times y < 1)$. The search step resumes, and finds a new assignment that satisfies $[\ y \geq 2,\ y \times y > 4\ ]$. The final check step is invoked again, and this time it finds the constraints to be satisfiable and the SMT solver terminates. The procedure above can be encoded as tactic of the form:

$$\mathsf{then}(\mathsf{preprocess},\ \mathsf{smt}(\mathsf{finalcheck}))$$

where preprocess is a tactic corresponding to the preprocessing step, and finalcheck is another tactic corresponding to the final check step, and smt is a tactical. The smt tactical uses a potentially expensive finalcheck tactic to complement an incomplete and fast procedure based on SAT solving and cheap theory propagation.

Abstract Partial Cylindrical Algebraic Decomposition (AP-CAD). AP-CAD [24,26] is an extension of the well-known real closed field (RCF) quantifier elimination procedure *partial cylindrical algebraic decomposition* (P-CAD). In AP-CAD, arbitrary (sound but possibly incomplete) ∃-RCF decision procedures can be given as parameters and used to "short-circuit" certain expensive computations performed during CAD construction. The ∃-RCF decision procedures may be used to reduce the expense of the different phases of the P-CAD procedure. The key idea is to use some *fast*, *sound* and *incomplete* procedure P to improve the performance of a *complete* but potentially very expensive procedure. The procedure P may be the combination of different techniques based on interval constraint propagation, rewriting, Gröbner basis computation, to cite a few. These combinations may be tailored as needed for different application domains. These observations suggest that P is a tactic, and AP-CAD is tactical that given P returns a tactic that implements a complete ∃-RCF decision procedure.

We now illustrate the flexibility of our approach using the following simple strategy for nonlinear arithmetic:

```
then(then(simplify, gaussian), orelse(modelfinder, smt(apcad(icp))))
```

The preprocessing step comprises of two steps: simple rewriting rules such as constant folding and gaussian elimination. Then, a model finder for nonlinear arithmetic based on SAT [36] is used. If it fails, smt is invoked using AP-CAD (apcad) in the final check step. Finally, AP-CAD uses interval constraint propagation (icp) to speedup the P-CAD procedure.

5 Strategies in Action

We demonstrate the practical value of our approach by describing successful strategies used in **RAHD** and **Z3**. We also provide evidence that the overhead due to the use of tactics and tacticals is insignificant, and the gains in performance substantial.

5.1 Z3 QF_LIA strategy

SMT-LIB is a repository of SMT benchmark problems. The benchmarks are divided in different divisions. The QF_LIA division consists of linear integer arithmetic benchmarks. These benchmarks come from different application domains such as: scheduling, hardware verification, software analysis, and bounded-model checking. The structural characteristics of these problems are quite diverse. Some of them contain a rich Boolean structure, and others are just the conjunction of linear equalities and inequalities. Several software analysis benchmark make extensive use of if-then-else terms that need to eliminated during a preprocessing step. A substantial subset of the benchmarks are unsatisfiable even when integrality constraints are ignored, and can be solved using a procedure for linear real arithmetic, such as Simplex. We say a benchmark is *bounded* if every uninterpreted constant a has a lower ($k \leq a$) and upper bound ($a \leq k$), where k is a numeral. A benchmark is said to be *unbounded* if it is not bounded. A bounded benchmark is said to be 0-1 (or *pseudo-boolean*) if the lower (upper) bound of every uninterpreted constant is 0 (1). Moreover, some of the problems in QF_LIA become bounded after interval constraint propagation is applied.

Z3 3.0 won the QF_LIA division in the last SMT competition[8] (SMT-COMP'11). The strategy used by **Z3** can be summarized by the following tactic:

```
then(preamble, orelse(mf, pb, bounded, smt)
```

where the preamble, mf, pb and bounded tactics are defined as

```
preamble = then(simplify, propagate-values, ctx-simplify,
                 lift-if, gaussian, simplify)

mf       = then(failif(not is-ilp), propagate-bounds,
                 orelse(tryfor(mip, 5000),
                        tryfor(smt-no-cut(100), 2000),
                        then(add-bounds(-16, 15), smt),
                        tryfor(smt-no-cut(200), 5000),
                        then(add-bounds(-32, 31), smt),
                        mip))

pb       = then(failif(not is-pb), pb2bv, bv2sat, sat)

bounded  = then(failif(unbounded),
                 orelse(tryfor(smt-no-cut(200), 5000),
                        tryfor(smt-no-cut-no-relevancy(200), 5000),
                        tryfor(smt-no-cut(300), 15000)))
```

The tactic smt is based on the **Yices** approach for linear integer arithmetic. The tactic ctx-simplify performs contextual simplification rules such as:

$$(a \neq t \vee F[a]) \;\rightsquigarrow\; (a \neq t \vee F[t])$$

The tactic mip implements a solver for mixed integer programming. It can only process conjunctions of linear inequalities and equalities. The tactic fails if the input goal contains other Boolean connectives. The tactic smt-no-cut(seed) is a variation of the **Yices** approach where Gomory cuts are not used. The parameter seed is a seed for the pseudo-random number generator. It is used to

[8] http://www.smtcomp.org

randomize the search. The tactic smt-no-cut-no-relevancy(seed) is yet another variation where "don't care" propagation is disabled. The tactic pb2bv converts a pseudo-boolean formula into a bit-vector formula. It fails if the input goal is not pseudo-boolean. Similarly, the tactic bv2sat bitblasts bit-vector terms into propositional logic. The tactic sat implements a SAT solver. Finally, the tactic add-bounds(lower, upper) performs an under-approximation by adding lower and upper bounds to all uninterpreted integer constants. The idea is to guarantee that the branch-and-bound procedure used in smt and mip terminates. The tactic mf is essentially looking for models where all integer variables are assigned to small values. The tactic pb is a specialized 0-1 (Pseudo-Boolean) solver. It fails if the problem is not 0-1.

To demonstrate the benefits of our approach we run all QF_LIA benchmarks using the following variations of the strategy above:

pre	= then(preamble, smt)
pre+pb	= then(preamble, orelse(pb, smt))
pre+bounded	= then(preamble, orelse(bounded, smt))
pre+mf	= then(preamble, orelse(mf, smt))
combined	= then(preamble, orelse(mf, pb, bounded, smt)

All experiments were conducted on an Intel Quad-Xeon (E54xx) processor, with individual runs limited to 2GB of memory and 600 seconds. The results of our experimental evaluation are presented in Table 1. The rows are associated with the individual benchmark families from QF_LIA division, and columns separate different strategies. For each benchmark family we write the number of benchmarks that each strategy failed to solve within the time limit, and the cumulative time for the solved benchmarks.

Overall, the combined strategy is the most effective one, solving the most problems. It fails only on 234 out of 5938 benchmarks. In constrast, the basic smt strategy fails on 978 benchmarks. The results also show which tactics are effective in particular benchmark families. The tactics ctx-simplify and lift-if are particularly effective on the NEC software verification benchmarks (sf nec-smt). The pseudo-boolean strategy reduces by half the number of failures in the industrial pseudo-boolean benchmarks coming from the 2010 pseudo-boolean competition. The convert software verification benchmarks become trivial when Gomory cuts are disabled by the tactic bounded. Finally, the model finder tactic mf is very effective on crafted benchmark families such as CAV 2009, cut lemmas, dillig, prime-cone, and slacks.

Figure 1 contains scatter plots comparing the strategies described above. Each point on the plots represents a benchmark. The plots are in log scale. Points below (above) the diagonal are benchmarks where the strategy on y-axis (x-axis) is faster than the strategy on the x-axis (y-axis). Note that in some cases, the combined strategy has a negative impact, but it overall solves more problems.

We observed several benefits when using tactics and tacticals in **Z3**. First, it is straighforward to create complex strategies using different solvers and techniques. The different solvers can be implemented and maintained independently

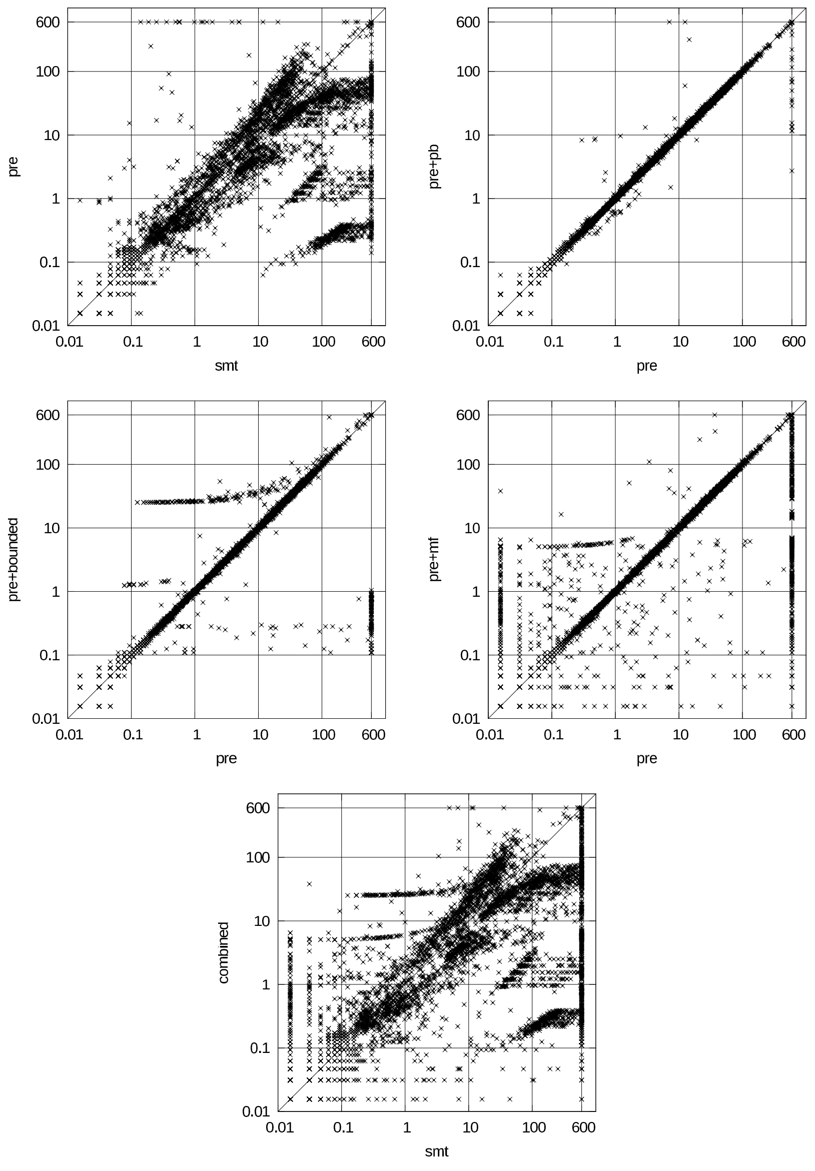

Fig. 1. smt, pre, pre+pb, pre+bounded, pre+mf and combined strategies.

Table 1. Detailed Experimental Results.

	smt		pre		pre+pb		pre+bounded		pre+mf		combined	
benchmark family	failed	time (s)	failed	time (s)	failed	time (s)	failed	time (s)	failed	time (s)	failed	time (s)
Averest (19)	0	4.0	0	5.9	0	6.0	0	5.9	0	5.9	0	5.9
bofill sched (652)	1	1530.7	1	1208.3	1	1191.5	1	1205.4	1	1206.0	1	1205.9
calypto (41)	1	2.0	1	7.7	1	8.0	1	7.8	1	7.9	1	7.8
CAV 2009 (600)	190	1315.3	190	1339.3	190	1329.7	190	1342.7	1	8309.5	1	8208.1
check (5)	0	0.1	0	0.1	0	0.1	0	0.1	0	0.1	0	0.1
CIRC (51)	17	188.4	17	239.8	17	238.2	17	336.1	17	221.5	8	158.66
convert (319)	206	1350.5	190	3060.6	190	3025.9	0	112.7	190	3030.2	0	112.5
cut lemmas (100)	48	2504.0	48	2532.4	48	2509.2	48	2543.4	27	3783.9	27	3709.0
dillig (251)	68	1212.0	68	1237.6	68	1226.9	68	1242.5	3	2677.8	3	2763.9
mathsat (121)	0	171.4	0	150.2	0	149.9	0	151.1	0	150.9	0	150.2
miplib2003 (16)	5	53.8	5	57.7	5	424.4	5	109.5	5	58.8	5	430.5
nec-smt (2780)	147	224149.0	8	59977.4	8	59968.3	8	59929.3	8	60042.1	8	60032.9
pb2010 (81)	43	90.3	43	96.2	25	2581.2	43	146.3	43	96.2	25	2583.1
pidgeons (19)	0	0.3	0	0.4	0	0.4	0	0.3	0	0.3	0	0.3
prime-cone (37)	13	9.6	13	9.5	13	9.5	13	9.7	0	11.0	0	11.0
rings (294)	48	4994.4	46	5973.7	46	6016.2	48	9690.0	46	6024.6	48	9548.2
rings pre (294)	57	441.5	54	1288.7	54	1261.9	54	1260.9	54	1274.7	54	1261.5
RTCL (2)	0	0.1	0	0.1	0	0.1	0	0.1	0	0.1	0	0.1
slacks (251)	135	1132.9	136	550.0	136	545.9	136	550.8	53	8969.3	53	8803.9
total (5938)	978	239153.0	819	77737.4	801	80495.2	631	78646.5	449	95872.4	234	98995.2

of each other. The overhead of using tactics and tacticals is insignificant. We can provide custom strategies to **Z3** users in different application domains. Finally, the number of SMT-LIB problems that could be solved by **Z3** increased dramatically. **Z3** 2.19 uses only the (default) smt tactic, and fails to solve 978 (out of 5938) QF_LIA benchmarks with a 10 minutes timeout. In contrast, **Z3** 3.0 fails in only 234 benchmarks. In **Z3** 4.0, tactic and tacticals are available in the programmatic API and SMT 2.0 frontend.

5.2 Calculemus RAHD strategies

The calculemus **RAHD** strategies[9] combine simplification procedures, interval constraint propagation, Gröbner basis computation, non-strict inequality splitting, DNF conversion, OpenCAD and CAD. OpenCAD is a variation of the CAD procedure that can be applied to problems containing only strict inequalities. OpenCAD is substantially faster than the general CAD procedure because it uses rational numbers instead of algebraic numbers. The key insight in the calculemus strategy is to split non-strict inequalities ($p \leq 0$) appearing in a conjunctive formula F into ($p < 0 \vee p = 0$), resulting in two sub-problems $F_<$ and $F_=$. The branch $F_<$ containing the strict inequality is then closer to being able to be processed using OpenCAD, while the branch $F_=$ containing the equality has an enriched equational structure which is then be used, via Gröbner basis computation, to inject equational information into the polynomials appearing in the strict inequalities in $F_=$. If the ideal generated by the equations in the branch $F_=$ is rich enough and the original formula is unsatisfiable, then this unsatisfiability of $F_=$ may be recognized by applying OpenCAD only to the resulting

[9] Detailed descriptions of these strategies may be found in Passmore's PhD thesis [24].

strict inequational fragment of $F_=$ after this Gröbner basis reduction has been performed.

In this section, we consider the basic calculemus strategy calc-0, and two refinements: calc-1 and calc-2. These refinements use formula measures to control inequality splitting. Moreover, interval constraint propagation is used to close goals before further splitting is performed.

Table 2. The three **RAHD** `calculemus` proof strategies compared with QEPCAD-B and Redlog on twenty-four problems.

			calc-0	calc-1	calc-2	qepcad-b	redlog/rlqe	redlog/rlcad
benchmark	dimension	degree	time (s)	time (s)	time (s)	time (s)	time (s)	time (s)
P0	5	4	0.9	1.6	1.7	416.4	40.4	>600.0
P1	6	4	1.7	3.1	3.4	>600.0	>600.0	>600.0
P2	5	4	1.3	2.4	2.6	>600.0	>600.0	>600.0
P3	5	4	1.5	2.5	2.7	>600.0	>600.0	>600.0
P4	5	4	1.1	2.0	2.7	>600.0	>600.0	>600.0
P5	14	2	0.3	0.3	0.3	>600.0	97.4	>600.0
P6	11	5	147.4	<0.1	<0.1	>600.0	<0.1	<0.1
P7	8	2	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1
P8	7	32	4.5	0.1	<0.1	8.4	<0.1	>600.0
P9	7	16	4.5	0.2	<0.1	0.3	<0.1	6.7
P10	7	12	100.7	20.8	8.9	>600.0	>600.0	>600.0
P11	6	2	1.6	0.5	0.5	<0.1	<0.1	<0.1
P12	5	3	0.8	0.3	0.4	<0.1	<0.1	<0.1
P13	4	10	3.8	3.9	4.0	>600.0	>600.0	>600.0
P14	2	2	4.5	1.7	<0.1	<0.1	>600.0	>600.0
P15	4	3	0.2	0.2	0.1	<0.1	<0.1	<0.1
P16	4	2	10.0	2.2	2.1	<0.1	<0.1	<0.1
P17	4	2	0.6	0.6	0.7	0.3	<0.1	0.6
P18	4	2	1.3	1.3	1.3	<0.1	<0.1	<0.1
P19	3	6	3.3	1.7	2.1	<0.1	<0.1	0.7
P20	3	4	1.2	0.7	0.7	<0.1	<0.1	0.3
P21	3	2	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1
P22	2	4	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1
P23	2	2	<0.1	<0.1	<0.1	<0.1	<0.1	<0.1

Table 2 shows the performance of the calculemus **RAHD** strategies on the twenty-four benchmarks considered in [25] and compares this performance to that of QEPCAD-B [4] and two quantifier elimination procedures available in Reduce/Redlog [1]:

– Rlqe, which is an enhanced implementation by Dolzmann and Sturm of Weispfenning's quadratic virtual term substitution (VTS) [33] , and
– Rlcad, which is an implementation by Seidl, Dolzmann and Sturm of Collins-Hong's partial CAD [12].

For each benchmark we write the dimension and maximal total multivariate degree of polynomials, and the total runtime for each strategy and solver. Experiments were performed on a 2 x 2.4 GHz Quad-Core Intel Xeon PowerMac with 10GB of 1066 MHz DDR3 RAM.

For brevity, let us only compare calc-0 with the QEPCAD-B and Redlog procedures. With this restriction, the results of these experiments can be broadly summarized as follows:

- The `calc-0` strategy is able to solve a number of high-dimension, high-degree problems that QEPCAD-B, Redlog/Rlqe, and Redlog/Rlcad are not. It is interesting that while the `calc-0` strategy involves an exponential blow-up in its reliance on inequality splitting followed by a DNF normalisation, for many benchmarks the increase in complexity caused by this blow-up is overshadowed by the decrease in complexity of the CAD-related computations this process induces.
- Redlog/Rlqe is able to solve a number of high-dimension, high-degree benchmarks that QEPCAD-B and Redlog/Rlcad are not.
- Redlog/Rlqe is able to solve a number of benchmarks significantly faster than the `calc-0` strategy, Redlog/Rlcad, and QEPCAD-B.
- For the benchmarks QEPCAD-B is able to solve directly, using QEPCAD-B directly tends to be much faster than using the `calc-0` strategy.

Overall, the final refinement, `calc-2`, substantially improves upon the strategy `calc-0` on benchmarks P6, P8, P10, P11, P12, P14, P16, P19 and P20, often by many orders of magnitude. On benchmarks P0, P1, P2, P3, P4, `calculemus-2` is slower than `calc-0` by roughly a factor of two. Strategies `calc-1` and `calc-2` are roughly equal for most benchmarks, except for P1 and P19 where `calc-2` is slightly ($\cong$ 10-20%) slower, and P10 and P14 where `calc-2` is substantially ($\cong$ 2-25x) faster.

6 Conclusion

We have demonstrated the practical value of heuristic proof strategies within the context of our **RAHD** and **Z3** tools. We have illustrated that not only is a strategy-language based approach practical in the context of high-performance solvers, it is also desirable. A key take-away message is the following: In difficult (i.e., infeasible or undecidable) theorem proving domains, the situation with heuristic proof strategies is rarely "one size fits all." Instead, given a class of problems to solve, it is often the case that one heuristic combination of reasoning engines is far more suited to the task than another. SMT solver developers cannot anticipate all classes of problems end-users will wish to analyze. By virtue of this, heuristic components of high-performance solvers will never be sufficient in general when they are beyond end-users' control. Without providing end-users mechanisms to control and modify the heuristic components of their solvers, solver developers are inhibiting their chances of success.

Beyond the situation with end-users, let us also make the following anecdotal remarks as solver developers. By introducing a strategy language foundation into our solvers, we have found our productivity radically enhanced, especially when faced with the goal of solving new classes of problems. The strategy language framework allows us to easily modify and experiment with variations of our solving heuristics. Before we had such strategy language machinery in place, with its principled handling of goals, models and proofs, this type of experimentation with new heuristics was cumbersome and error-prone.

We have proposed a Strategy Challenge to the SMT community: To build theoretical and practical tools allowing users to exert strategic control over core heuristic aspects of high-performance SMT solvers. We discussed some of the rich history of ideas of strategy in the context of mechanized proof, and presented an SMT-oriented approach for orchestrating reasoning engines, where "big" symbolic reasoning steps are represented as tactics, and these tactics are composed using combinators known as tacticals. We demonstrated the practical value of this approach by describing a few examples of how tactics and tacticals have been successfully used in our **RAHD** and **Z3** tools.

There are several directions for future work. First, we believe that many other authors of SMT solvers must take up this Strategy Challenge, and much experimentation must be done — from many different points of view and domains of application — before a standard strategy language for SMT should be proposed. When the time is right, we believe that the existence of a strategy standard (extending, for instance, the SMT-LIB standard) and the development and study of theoretical frameworks for SMT strategies could give rise to much progress in the practical efficacy of automated reasoning tools.

Second, we would like to understand how one might *efficiently* exert "small step" strategic control over reasoning engines. Abstract proof procedures, such as Abstract DPLL [23], DPLL(T) [23] and cutsat [17], represent a proof procedure as a set of transition rules. In these cases, a strategy comprises a *recipe* for applying these "small" step rules. Actual implementations of these abstract procedures use carefully chosen efficient data-structures that depend on the preselected strategy. It is not clear to us how to specify a strategy for these abstract procedures so that an efficient implementation can be automatically generated. Another topic for future investigation is to explore different variations of the LCF approach, such as the ones used by the interactive theorem provers Isabelle, HOL, Coq and Matita.

References

1. T. S. A. Dolzmann. Redlog User Manual - Edition 2.0. MIP-9905, 1999.
2. W. Ackermann. Solvable cases of the decision problem. *Studies in Logic and the Foundation of Mathematics*, 1954.
3. C. Barrett, A. Stump, and C. Tinelli. The Satisfiability Modulo Theories Library (SMT-LIB). `www.SMT-LIB.org`, 2010.
4. C. W. Brown. QEPCAD-B: A System for Computing with Semi-algebraic Sets via Cylindrical Algebraic Decomposition. *SIGSAM Bull.*, 38:23–24, March 2004.
5. R. Bruttomesso, A. Cimatti, A. Franzén, A. Griggio, A. Santuari, and R. Sebastiani. To Ackermann-ize or Not to Ackermann-ize? On Efficiently Handling Uninterpreted Function Symbols in *UF(E)*. In *LPAR*, pages 557–571, 2006.
6. M. Davis. Eliminating the Irrelevant from Mechanical Proofs. *Proc. Symp. Applied Math.*, XV:15–30, 1963.
7. M. Davis. The early history of automated deduction. In J. A. Robinson and A. Voronkov, editors, *Handbook of Automated Reasoning*, pages 3–15. Elsevier and MIT Press, 2001.

8. M. Davis, G. Logemann, and D. Loveland. A machine program for theorem-proving. *Commun. ACM*, 5:394–397, July 1962.
9. M. Davis and H. Putnam. A computing procedure for quantification theory. *J. ACM*, 7:201–215, July 1960.
10. L. de Moura and N. Bjørner. Z3: An efficient SMT solver. In *Proc. of TACAS*, number 4963 in LNCS. Springer, 2008.
11. L. M. de Moura and N. Bjørner. Satisfiability modulo theories: introduction and applications. *Commun. ACM*, 54(9):69–77, 2011.
12. A. Dolzmann, A. Seidl, and T. Sturm. Efficient Projection Orders for CAD. In *ISSAC '04: Proceedings of the 2004 international symposium on Symbolic and algebraic computation*, pages 111–118. ACM, 2004.
13. B. Dutertre and L. de Moura. A Fast Linear-Arithmetic Solver for DPLL(T). In *CAV*, pages 81–94, 2006.
14. P. C. Gilmore. A Proof Method for Quantification Theory: its Justification and Realization. *IBM J. Res. Dev.*, 4:28–35, January 1960.
15. M. Gordon. *From LCF to HOL: a short history*, pages 169–185. MIT Press, Cambridge, MA, USA, 2000.
16. M. J. C. Gordon and T. F. Melham. *Introduction to HOL: a theorem-proving environment for higher-order logic.* Cambridge University Press, 1993.
17. D. Jovanovic and L. de Moura. Cutting to the chase solving linear integer arithmetic. In *CADE*, pages 338–353, 2011.
18. F. Kirchner and C. Muñoz. The proof monad. *Journal of Logic and Algebraic Programming*, 79(3–5):264–277, 2010.
19. E. L. Lusk. Controlling Redundancy in Large Search Spaces: Argonne-Style Theorem Proving Through the Years. In *Proceedings of the International Conference on Logic Programming and Automated Reasoning*, LPAR '92, pages 96–106, London, UK, UK, 1992. Springer-Verlag.
20. B. Luttik and E. Visser. Specification of rewriting strategies. In M. P. A. Sellink, editor, *2nd International Workshop on the Theory and Practice of Algebraic Specifications (ASF+SDF 1997)*, Electronic Workshops in Computing, Berlin, November 1997. Springer-Verlag.
21. W. McCune. Prover9 and mace4. `http://www.cs.unm.edu/~mccune/prover9/`, 2005–2010.
22. R. Milner. Logic for computable functions: description of a machine implementation. Technical Report STAN-CS-72-288, Stanford University, 1972.
23. R. Nieuwenhuis, A. Oliveras, and C. Tinelli. Solving SAT and SAT Modulo Theories: From an abstract Davis–Putnam–Logemann–Loveland procedure to DPLL(T). *J. ACM*, 53(6):937–977, 2006.
24. G. O. Passmore. *Combined Decision Procedures for Nonlinear Arithmetics, Real and Complex.* PhD thesis, University of Edinburgh, 2011.
25. G. O. Passmore and P. B. Jackson. Combined Decision Techniques for the Existential Theory of the Reals. In *Proceedings of the 16th Symposium, 8th International Conference. Conference on Intelligent Computer Mathematics*, Calculemus '09/MKM '09, pages 122–137, Berlin, Heidelberg, 2009. Springer-Verlag.
26. G. O. Passmore and P. B. Jackson. Abstract Partial Cylindrical Algebraic Decomposition I: The Lifting Phase. In S. B. Cooper, A. Dawar, and B. Loewe, editors, *Proceedings of Computability in Europe 2012: Turing Centenary (To appear)*. Springer-Verlag, 2012.
27. L. Paulson. *Logic and Computation: Interactive Proof with Cambdrige LCF*, volume 2. Cambridge University Press, 1987.

28. L. Paulson. Isabelle: The next 700 theorem provers. In *Logic and Computer Science*, pages 361–386. Academic Press, 1990.
29. D. Prawitz. An Improved Proof Procedure. *Theoria*, 26(2):102–139, 1960.
30. A. Robinson. Short Lecture. Summer Institute for Symbolic Logic, Cornell University, 1957.
31. J. A. Robinson. A machine-oriented logic based on the resolution principle. *J. ACM*, 12:23–41, January 1965.
32. H. Wang. Toward mechanical mathematics. *IBM J. Res. Dev.*, 4:2–22, January 1960.
33. V. Weispfenning. Quantifier Elimination for Real Algebra - the Quadratic Case and Beyond. *Appl. Algebra Eng. Commun. Comput.*, 8(2):85–101, 1997.
34. L. Wos, D. Carson, and G. Robinson. The unit preference strategy in theorem proving. In *Proceedings of the October 27-29, 1964, fall joint computer conference, part I*, AFIPS '64 (Fall, part I), pages 615–621, New York, NY, USA, 1964. ACM.
35. L. Wos, G. A. Robinson, and D. F. Carson. Efficiency and completeness of the set of support strategy in theorem proving. *J. ACM*, 12:536–541, October 1965.
36. H. Zankl and A. Middeldorp. Satisfiability of non-linear (ir)rational arithmetic. In *Proceedings of the 16th International Conference on Logic for Programming and Automated Reasoning*, volume 6355 of *Lecture Notes in Artificial Intelligence*, pages 481–500, Dakar, 2010. Springer-Verlag.

Acknowledgements. Grant Passmore was supported during this research by the UK Engineering and Physical Sciences Research Council [grant numbers EP/I011005/1 and EP/I010335/1].