Meta-F⋆: Proof Automation with SMT, Tactics, and Metaprograms

Guido Martínez[1,2], Danel Ahman[3], Victor Dumitrescu[4], Nick Giannarakis[5],
Chris Hawblitzel[6], Cătălin Hriţcu[2], Monal Narasimhamurthy[7],
Zoe Paraskevopoulou[5], Clément Pit-Claudel[8], Jonathan Protzenko[6],
Tahina Ramananandro[6], Aseem Rastogi[6], and Nikhil Swamy[6]

[1]CIFASIS-CONICET   [2]Inria Paris   [3]University of Ljubljana
[4]MSR-Inria Joint Centre   [5]Princeton University   [6]Microsoft Research
[7]University of Colorado Boulder   [8]MIT CSAIL

Abstract. We introduce Meta-F⋆, a tactics and metaprogramming framework for the F⋆ program verifier. The main novelty of Meta-F⋆ is allowing the use of tactics and metaprogramming to discharge assertions not solvable by SMT, or to just simplify them into well-behaved SMT fragments. Plus, Meta-F⋆ can be used to generate verified code automatically.

Meta-F⋆ is implemented as an F⋆ *effect*, which, given the powerful effect system of F⋆, heavily increases code reuse and even enables the lightweight verification of metaprograms. Metaprograms can be either interpreted, or compiled to efficient native code that can be dynamically loaded into the F⋆ type-checker and can interoperate with interpreted code. Evaluation on realistic case studies shows that Meta-F⋆ provides substantial gains in proof development, efficiency, and robustness.

Keywords: Tactics · Metaprogramming · Program verification · Verification conditions · SMT solvers · Proof assistants

1 Introduction

Scripting proofs using tactics and metaprogramming has a long tradition in interactive theorem provers (ITPs), starting with Milner's Edinburgh LCF [37]. In this lineage, properties of *pure* programs are specified in expressive higher-order (and often dependently typed) logics, and proofs are conducted using various imperative programming languages, starting originally with ML.

Along a different axis, program verifiers like Dafny [47], VCC [23], Why3 [33], and Liquid Haskell [59] target both pure *and effectful* programs, with side-effects ranging from divergence to concurrency, but provide relatively weak logics for specification (e.g., first-order logic with a few selected theories like linear arithmetic). They work primarily by computing verification conditions (VCs) from programs, usually relying on annotations such as pre- and postconditions, and encoding them to automated theorem provers (ATPs) such as satisfiability modulo theories (SMT) solvers, often providing excellent automation.

These two sub-fields have influenced one another, though the situation is somewhat asymmetric. On the one hand, most interactive provers have gained support for exploiting SMT solvers or other ATPs, providing push-button automation for certain kinds of assertions [26, 31, 43, 44, 54]. On the other hand, recognizing the importance of interactive proofs, Why3 [33] interfaces with ITPs like Coq. However, working over proof obligations translated from Why3 requires users to be familiar not only with both these systems, but also with the specifics of the translation. And beyond Why3 and the tools based on it [25], no other SMT-based program verifiers have full-fledged support for interactive proving, leading to several downsides:

Limits to expressiveness The expressiveness of program verifiers can be limited by the ATP used. When dealing with theories that are undecidable and difficult to automate (e.g., non-linear arithmetic or separation logic), proofs in ATP-based systems may become impossible or, at best, extremely tedious.

Boilerplate To work around this lack of automation, programmers have to construct detailed proofs by hand, often repeating many tedious yet error-prone steps, so as to provide hints to the underlying solver to discover the proof. In contrast, ITPs with metaprogramming facilities excel at expressing domain-specific automation to complete such tedious proofs.

Implicit proof context In most program verifiers, the logical context of a proof is implicit in the program text and depends on the control flow and the pre- and postconditions of preceding computations. Unlike in interactive proof assistants, programmers have no explicit access, neither visual nor programmatic, to this context, making proof structuring and exploration extremely difficult.

In direct response to these drawbacks, we seek a system that successfully combines the convenience of an automated program verifier for the common case, while seamlessly transitioning to an interactive proving experience for those parts of a proof that are hard to automate. Towards this end, we propose Meta-F*, a tactics and metaprogramming framework for the F* [1, 58] program verifier.

Highlights and Contributions of Meta-F*

F* has historically been more deeply rooted as an SMT-based program verifier. Until now, F* discharged VCs exclusively by calling an SMT solver (usually Z3 [28]), providing good automation for many common program verification tasks, but also exhibiting the drawbacks discussed above.

Meta-F* is a framework that allows F* users to manipulate VCs using *tactics*. More generally, it supports *metaprogramming*, allowing programmers to script the construction of programs, by manipulating their syntax and customizing the way they are type-checked. This allows programmers to (1) implement custom procedures for manipulating VCs; (2) eliminate boilerplate in proofs and programs; and (3) to inspect the proof state visually and to manipulate it programmatically, addressing the drawbacks discussed above. SMT still plays a central

role in Meta-F⋆: a typical usage involves implementing tactics to transform VCs, so as to bring them into theories well-supported by SMT, without needing to (re)implement full decision procedures. Further, the generality of Meta-F⋆ allows implementing non-trivial language extensions (e.g., typeclass resolution) entirely as metaprogramming libraries, without changes to the F⋆ type-checker.

The technical **contributions** of our work include the following:

"Meta-" is just an effect (§3.1) Meta-F⋆ is implemented using F⋆'s extensible effect system, which keeps programs and metaprograms properly isolated. Being first-class F⋆ programs, metaprograms are typed, call-by-value, direct-style, higher-order functional programs, much like the original ML. Further, metaprograms can be themselves verified (to a degree, see §3.4) and metaprogrammed.

Reconciling tactics with VC generation (§4.2) In program verifiers the programmer often guides the solver towards the proof by supplying intermediate assertions. Meta-F⋆ retains this style, but additionally allows assertions to be solved by tactics. To this end, a contribution of our work is extracting, from a VC, a proof state encompassing all relevant hypotheses, including those implicit in the program text.

Executing metaprograms efficiently (§5) Metaprograms are executed during type-checking. As a baseline, they can be interpreted using F⋆'s existing (but slow) abstract machine for term normalization, or a faster normalizer based on normalization by evaluation (NbE) [10, 16]. For much faster execution speed, metaprograms can also be run natively. This is achieved by combining the existing extraction mechanism of F⋆ to OCaml with a new framework for safely extending the F⋆ type-checker with such native code.

Examples (§2) and evaluation (§6) We evaluate Meta-F⋆ on several case studies. First, we present a functional correctness proof for the Poly1305 message authentication code (MAC) [11], using a novel combination of proofs by reflection for dealing with non-linear arithmetic and SMT solving for linear arithmetic. We measure a clear gain in proof robustness: SMT-only proofs succeed only rarely (for reasonable timeouts), whereas our tactic+SMT proof is concise, never fails, and is faster. Next, we demonstrate an improvement in expressiveness, by developing a small library for proofs of heap-manipulating programs in separation logic, which was previously out-of-scope for F⋆. Finally, we illustrate the ability to automatically construct verified effectful programs, by introducing a library for metaprogramming verified low-level parsers and serializers with applications to network programming, where verification is accelerated by processing the VC with tactics, and by programmatically tweaking the SMT context.

We conclude that tactics and metaprogramming can be prosperously combined with VC generation and SMT solving to build verified programs with better, more scalable, and more robust automation.

The full version of this paper, including appendices, can be found online in https://www.fstar-lang.org/papers/metafstar.

2   Meta-F* by Example

F* is a general-purpose programming language aimed at program verification. It puts together the automation of an SMT-backed deductive verification tool with the expressive power of a language with full-spectrum dependent types. Briefly, it is a functional, higher-order, effectful, dependently typed language, with syntax loosely based on OCaml. F* supports refinement types and Hoare-style specifications, computing VCs of computations via a type-level weakest precondition (WP) calculus packed within *Dijkstra monads* [57]. F*'s effect system is also user-extensible [1]. Using it, one can model or embed imperative programming in styles ranging from ML to C [55] and assembly [35]. After verification, F* programs can be extracted to efficient OCaml or F# code. A first-order fragment of F*, called Low*, can also be extracted to C via the KreMLin compiler [55].

This paper introduces Meta-F*, a metaprogramming framework for F* that allows users to safely customize and extend F* in many ways. For instance, Meta-F* can be used to preprocess or solve proof obligations; synthesize F* expressions; generate top-level definitions; and resolve implicit arguments in user-defined ways, enabling non-trivial extensions. This paper primarily discusses the first two features. Technically, none of these features deeply increase the expressive power of F*, since one could manually program in F* terms that can now be metaprogrammed. However, as we will see shortly, manually programming terms and their proofs can be so prohibitively costly as to be practically infeasible.

Meta-F* is similar to other tactic frameworks, such as Coq's [29] or Lean's [30], in presenting a set of goals to the programmer, providing commands to break them down, allowing to inspect and build abstract syntax, etc. In this paper, we mostly detail the characteristics where Meta-F* *differs* from other engines.

This section presents Meta-F* informally, displaying its usage through case studies. We present any necessary F* background as needed.

2.1   Tactics for Individual Assertions and Partial Canonicalization

Non-linear arithmetic reasoning is crucially needed for the verification of optimized, low-level cryptographic primitives [18, 64], an important use case for F* [13] and other verification frameworks, including those that rely on SMT solving alone (e.g., Dafny [47]) as well as those that rely exclusively on tactic-based proofs (e.g., FiatCrypto [32]). While both styles have demonstrated significant successes, we make a case for a middle ground, leveraging the SMT solver for the parts of a VC where it is effective, and using tactics only where it is not.

We focus on Poly1305 [11], a widely-used cryptographic MAC that computes a series of integer multiplications and additions modulo a large prime number $p = 2^{130} - 5$. Implementations of the Poly1305 multiplication and mod operations are carefully hand-optimized to represent 130-bit numbers in terms of smaller 32-bit or 64-bit registers, using clever tricks; proving their correctness requires reasoning about long sequences of additions and multiplications.

Previously: Guiding SMT solvers by manually applying lemmas Prior proofs of correctness of Poly1305 and other cryptographic primitives using SMT-based program verifiers, including F* [64] and Dafny [18], use a combination of SMT automation and manual application of lemmas. On the plus side, SMT solvers are excellent at linear arithmetic, so these proofs delegate all associativity-commutativity (AC) reasoning about addition to SMT. Non-linear arithmetic in SMT solvers, even just AC-rewriting and distributivity, are, however, inefficient and unreliable—so much so that the prior efforts above (and other works too [40, 41]) simply turn off support for non-linear arithmetic in the solver, in order not to degrade verification performance across the board due to poor interaction of theories. Instead, users need to explicitly invoke lemmas.[1]

For instance, here is a statement and proof of a lemma about Poly1305 in F*. The property and its proof do not really matter; the lines marked "*(∗argh! ∗)*" do. In this particular proof, working around the solver's inability to effectively reason about non-linear arithmetic, the programmer has spelled out basic facts about distributivity of multiplication and addition, by calling the library lemma distributivity_add_right, in order to guide the solver towards the proof. (Below, p44 and p88 represent 2^{44} and 2^{88} respectively)

```
let lemma_carry_limb_unrolled (a0 a1 a2 : nat) : Lemma (ensures (
    a0 % p44 + p44 ∗ ((a1 + a0 / p44) % p44) + p88 ∗ (a2 + ((a1 + a0 / p44) / p44))
    == a0 + p44 ∗ a1 + p88 ∗ a2)) =
 let z = a0 % p44 + p44 ∗ ((a1 + a0 / p44) % p44)
         + p88 ∗ (a2 + ((a1 + a0 / p44) / p44)) in
 distributivity_add_right p88 a2 ((a1 + a0 / p44) / p44); (∗ argh! ∗)
 pow2_plus 44 44;
 lemma_div_mod (a1 + a0 / p44) p44;
 distributivity_add_right p44 ((a1 + a0 / p44) % p44)
         (p44 ∗ ((a1 + a0 / p44) / p44)); (∗ argh! ∗)
 assert (p44 ∗ ((a1 + a0 / p44) % p44) + p88 ∗ ((a1 + a0 / p44) / p44)
         == p44 ∗ (a1 + a0 / p44) );
 distributivity_add_right p44 a1 (a0 / p44); (∗ argh! ∗)
 lemma_div_mod a0 p44
```

Even at this relatively small scale, needing to explicitly instantiate the distributivity lemma is verbose and error prone. Even worse, the user is blind while doing so: the program text does not display the current set of available facts nor the final goal. Proofs at this level of abstraction are painfully detailed in some aspects, yet also heavily reliant on the SMT solver to fill in the aspects of the proof that are missing.

Given enough time, the solver can sometimes find a proof without the additional hints, but this is usually rare and dependent on context, and almost never robust. In this particular example we find by varying Z3's random seed that, in

---

[1] Lemma (requires pre) (ensures post) is F* notation for the type of a computation proving pre $\implies$ post—we omit pre when it is trivial. In F*'s standard library, math lemmas are proved using SMT with little or no interactions between problematic theory combinations. These lemmas can then be explicitly invoked in larger contexts, and are deleted during extraction.

an isolated setting, the lemma is proven automatically about 32% of the time. The numbers are much worse for more complex proofs, and where the context contains many facts, making this style quickly spiral out of control. For example, a proof of one of the main lemmas in Poly1305, poly_multiply, requires 41 steps of rewriting for associativity-commutativity of multiplication, and distributivity of addition and multiplication—making the proof much too long to show here.

SMT and tactics in Meta-F⋆ The listing below shows the statement and proof of poly_multiply in Meta-F⋆, of which the lemma above was previously only a small part. Again, the specific property proven is not particularly relevant to our discussion. But, this time, the proof contains just two steps.

```
let poly_multiply (n p r h r0 r1 h0 h1 h2 s1 d0 d1 d2 h1 h2 hh : int) : Lemma
    (requires p > 0 ∧ r1 ≥ 0 ∧ n > 0 ∧ 4 * (n * n) == p + 5 ∧ r == r1 * n + r0 ∧
             h == h2 * (n * n) + h1 * n + h0 ∧ s1 == r1 + (r1 / 4) ∧ r1 % 4 == 0 ∧
             d0 == h0 * r0 + h1 * s1 ∧ d1 == h0 * r1 + h1 * r0 + h2 * s1 ∧
             d2 == h2 * r0 ∧ hh == d2 * (n * n) + d1 * n + d0)
    (ensures (h * r) % p == hh % p) =
    let r14 = r1 / 4 in
    let h_r_expand = (h2 * (n * n) + h1 * n + h0) * ((r14 * 4) * n + r0) in
    let hh_expand = (h2 * r0) * (n * n) + (h0 * (r14 * 4) + h1 * r0
                                  + h2 * (5 * r14)) * n + (h0 * r0 + h1 * (5 * r14)) in
    let b = (h2 * n + h1) * r14 in
    modulo_addition_lemma hh_expand p b;
    assert (h_r_expand == hh_expand + b * (n * n * 4 + (−5)))
        by (canon_semiring int_csr) (* Proof of this step by Meta-F* tactic *)
```

First, we call a single lemma about modular addition from F⋆'s standard library. Then, we assert an equality annotated with a tactic (assert..by). Instead of encoding the assertion as-is to the SMT solver, it is preprocessed by the canon_semiring tactic. The tactic is presented with the asserted equality as its goal, in an environment containing not only all variables in scope but also hypotheses for the precondition of poly_multiply and the postcondition of the modulo_addition_lemma call (otherwise, the assertion could not be proven). The tactic will then canonicalize the sides of the equality, but notably only "up to" linear arithmetic conversions. Rather than fully canonicalizing the terms, the tactic just rewrites them into a sum-of-products canonical form, leaving all the remaining work to the SMT solver, which can then easily and robustly discharge the goal using linear arithmetic only.

This tactic works over terms in the commutative semiring of integers (int_csr) using proof-by-reflection [12, 20, 36, 38]. Internally, it is composed of a simpler, also proof-by-reflection based tactic canon_monoid that works over monoids, which is then "stacked" on itself to build canon_semiring. The basic idea of proof-by-reflection is to reduce most of the proof burden to mechanical computation, obtaining much more efficient proofs compared to repeatedly applying lemmas. For canon_monoid, we begin with a type for monoids, a small AST representing monoid values, and a denotation for expressions back into the monoid type.

```
type monoid (a:Type) = { unit : a; mult : (a → a → a); (* + monoid laws ... *) }
```

```
type exp (a:Type) = | Unit : exp a | Var : a → exp a | Mult : exp a → exp a → exp a
(∗ Note on syntax: #a below denotes that a is an implicit argument ∗)
let rec denote (#a:Type) (m:monoid a) (e:exp a) : a =
   match e with
   | Unit → m.unit | Var x → x | Mult x y → m.mult (denote m x) (denote m y)
```

To canonicalize an exp, it is first converted to a list of operands (flatten) and then
reflected back to the monoid (mldenote). The process is proven correct, in the
particular case of equalities, by the monoid_reflect lemma.

```
val flatten : #a:Type → exp a → list a
val mldenote : #a:Type → monoid a → list a → a
let monoid_reflect (#a:Type) (m:monoid a) (e₁ e₂ : exp a)
                 : Lemma (requires (mldenote m (flatten e₁) == mldenote m (flatten e₂)))
                         (ensures (denote m e₁ == denote m e₂)) = ...
```

At this stage, if the goal is $t_1 == t_2$, we require two monoidal expressions e_1 and
e_2 such that $t_1 ==$ denote m e_1 and $t_2 ==$ denote m e_2. They are constructed by
the tactic canon_monoid by inspecting the *syntax* of the goal, using Meta-F⋆'s
reflection capabilities (detailed ahead in §3.3). We have no way to prove once
and for all that the expressions built by canon_monoid correctly denote the terms,
but this fact can be proven automatically at each application of the tactic, by
simple unification. The tactic then applies the lemma monoid_reflect m $e_1 e_2$, and
the goal is changed to mldenote m (flatten e_1) == mldenote m (flatten e_2). Finally,
by normalization, each side will be canonicalized by running flatten and mldenote.
    The canon_semiring tactic follows a similar approach, and is similar to existing
reflective tactics for other proof assistants [9, 38], except that it only canonicalizes
up to linear arithmetic, as explained above. The full VC for poly_multiply contains
many other facts, e.g., that p is non-zero so the division is well-defined and that
the postcondition does indeed hold. These obligations remain in a "skeleton" VC
that is also easily proven by Z3. This proof is much easier for the programmer to
write and much more robust, as detailed ahead in §6.1. The proof of Poly1305's
other main lemma, poly_reduce, is also similarly well automated.

Tactic proofs without SMT Of course, one can verify poly_multiply in Coq,
following the same conceptual proof used in Meta-F⋆, but relying on tactics only.
Our proof (included in the appendix) is 27 lines long, two of which involve the
use of Coq's ring tactic (similar to our canon_semiring tactic) and omega tactic for
solving formulas in Presburger arithmetic. The remaining 25 lines include steps
to destruct the propositional structure of terms, rewrite by equalities, enriching
the context to enable automatic modulo rewriting (Coq does not fully automat-
ically recognize equality modulo p as an equivalence relation compatible with
arithmetic operators). While a mature proof assistant like Coq has libraries and
tools to ease this kind of manipulation, it can still be verbose.
    In contrast, in Meta-F⋆ all of these mundane parts of a proof are simply
dispatched to the SMT solver, which decides linear arithmetic efficiently, beyond
the quantifier-free Presburger fragment supported by tactics like omega, handles
congruence closure natively, etc.

2.2   Tactics for Entire VCs and Separation Logic

A different way to invoke Meta-F* is over an entire VC. While the exact shape
of VCs is hard to predict, users with some experience can write tactics that find
and solve particular sub-assertions within a VC, or simply massage them into
shapes better suited for the SMT solver. We illustrate the idea on proofs for
heap-manipulating programs.

One verification method that has eluded F* until now is separation logic,
the main reason being that the pervasive "frame rule" requires instantiating
existentially quantified heap variables, which is a challenge for SMT solvers, and
simply too tedious for users. With Meta-F*, one can do better. We have written
a (proof-of-concept) embedding of separation logic and a tactic (sl_auto) that
performs heap frame inference automatically.

The approach we follow consists of designing the WP specifications for prim-
itive stateful actions so as to make their footprint syntactically evident. The
tactic then descends through VCs until it finds an existential for heaps aris-
ing from the frame rule. Then, by solving an equality between heap expressions
(which requires canonicalization, for which we use a variant of canon_monoid
targeting *commutative* monoids) the tactic finds the frames and instantiates
the existentials. Notably, as opposed to other tactic frameworks for separation
logic [4, 45, 49, 51], this is *all* our tactic does before dispatching to the SMT
solver, which can now be effective over the instantiated VC.

We now provide some detail on the framework. Below, 'emp' represents the
empty heap, '•' is the separating conjunction and '$r \mapsto v$' is the heaplet with the
single reference r set to value v.[2] Our development distinguishes between a "heap"
and its "memory" for technical reasons, but we will treat the two as equivalent
here. Further, defined is a predicate discriminating valid heaps (as in [52]), i.e.,
those built from separating conjunctions of *actually* disjoint heaps.

We first define the type of WPs and present the WP for the frame rule:

```
let pre = memory → prop (* predicate on initial heaps *)
let post a = a → memory → prop (* predicate on result values and final heaps *)
let wp a = post a → pre (* transformer from postconditions to preconditions *)

let frame_post (#a:Type) (p:post a) (m₀:memory) : post a =
  λx m₁ → defined (m₀ • m₁) ∧ p x (m₀ • m₁)
let frame_wp (#a:Type) (wp:wp a) (post:post a) (m:memory) =
  ∃m₀ m₁. defined (m₀ • m₁) ∧ m == (m₀ • m₁) ∧ wp (frame_post post m₁) m₀
```

Intuitively, frame_post p m_0 behaves as the postcondition p "framed" by m_0, i.e.,
frame_post p m_0 x m_1 holds when the two heaps m_0 and m_1 are disjoint and p
holds over the result value x and the conjoined heaps. Then, frame_wp wp takes
a postcondition p and initial heap m, and requires that m can be split into disjoint
subheaps m_0 (the footprint) and m_1 (the frame), such that the postcondition p,
when properly framed, holds over the footprint.

---

[2] This differs from the usual presentation where these three operators are heap *predi-
cates* instead of heaps.

In order to provide specifications for primitive actions we start in small-footprint style. For instance, below is the WP for reading a reference:

```
let read_wp (#a:Type) (r:ref a) = λpost m₀ → ∃x. m₀ == r ↦ x ∧ post x m₀
```

We then insert framing wrappers around such small-footprint WPs when exposing the corresponding stateful actions to the programmer, e.g.,

```
val (!) : #a:Type → r:ref a → STATE a (λ p m → frame_wp (read_wp r) p m)
```

To verify code written in such style, we annotate the corresponding programs to have their VCs processed by sl_auto. For instance, for the swap function below, the tactic successfully finds the frames for the four occurrences of the frame rule and greatly reduces the solver's work. Even in this simple example, not performing such instantiation would cause the solver to fail.

```
let swap_wp (r₁ r₂ : ref int) =
    λp m → ∃x y. m == (r₁ ↦ x • r₂ ↦ y) ∧ p () (r₁ ↦ y • r₂ ↦ x)
let swap (r₁ r₂ : ref int) : ST unit (swap_wp r₁ r₂) by (sl_auto ()) =
    let x = !r₁ in let y = !r₂ in r₁ := y; r₂ := x
```

The sl_auto tactic: (1) uses syntax inspection to unfold and traverse the goal until it reaches a frame_wp—say, the one for $!r_2$; (2) inspects frame_wp's first explicit argument (here read_wp r_2) to compute the references the current command requires (here r_2); (3) uses unification variables to build a memory expression describing the required framing of input memory (here $r_2 \mapsto ?u_1 \bullet ?u_2$) and instantiates the existentials of frame_wp with these unification variables; (4) builds a goal that equates this memory expression with frame_wp's third argument (here $r_1 \mapsto x \bullet r_2 \mapsto y$); and (5) uses a commutative monoids tactic (similar to §2.1) with the heap algebra (emp, $\bullet$) to canonicalize the equality and sort the heaplets. Next, it can solve for the unification variables component-wise, instantiating $?u_1$ to y and $?u_2$ to $r_1 \mapsto x$, and then proceed to the next frame_wp.

In general, after frames are instantiated, the SMT solver can efficiently prove the remaining assertions, such as the obligations about heap definedness. Thus, with relatively little effort, Meta-F* brings an (albeit simple version of a) widely used yet previously out-of-scope program logic (i.e., separation logic) into F*. To the best of our knowledge, the ability to *script* separation logic into an SMT-based program verifier, without any primitive support, is unique.

2.3 Metaprogramming Verified Low-level Parsers and Serializers

Above, we used Meta-F* to manipulate VCs for user-written code. Here, we focus instead on generating verified code automatically. We loosely refer to the previous setting as using "tactics", and to the current one as "metaprogramming". In most ITPs, tactics and metaprogramming are not distinguished; however in a program verifier like F*, where some proofs are not materialized at all (§4.1), proving VCs of existing terms is distinct from generating new terms.

Metaprogramming in F* involves programmatically generating a (potentially effectful) term (e.g., by constructing its syntax and instructing F* how to type-check it) and processing any VCs that arise via tactics. When applicable (e.g.,

when working in a domain-specific language), metaprogramming verified code can substantially reduce, or even eliminate, the burden of manual proofs.

We illustrate this by automating the generation of parsers and serializers from a type definition. Of course, this is a routine task in many mainstream metaprogramming frameworks (e.g., Template Haskell, camlp4, etc). The novelty here is that we produce imperative parsers and serializers extracted to C, with proofs that they are memory safe, functionally correct, and mutually inverse. This section is slightly simplified, more detail can be found the appendix.

We proceed in several stages. First, we program a library of pure, high-level parser and serializer combinators, proven to be (partial) mutual inverses of each other. A parser for a type t is represented as a function possibly returning a t along with the amount of input bytes consumed. The type of a serializer for a given p:parser t contains a refinement[3] stating that p is an inverse of the serializer. A package is a dependent record of a parser and an associated serializer.

```
let parser t = seq byte → option (t ∗ nat)
let serializer #t (p:parser t) = f:(t → seq byte){∀ x. p (f x) == Some (x, length (f x))}
type package t = { p : parser t ; s : serializer p }
```

Basic combinators in the library include constructs for parsing and serializing base values and pairs, such as the following:

```
val p_u8 : parse u8
val s_u8 : serializer p_u8
val p_pair : parser t1 → parser t2 → parser (t1 ∗ t2)
val s_pair : serializer p1 → serializer p2 → serializer (p_pair p1 p2)
```

Next, we define low-level versions of these combinators, which work over mutable arrays instead of byte sequences. These combinators are coded in the Low* subset of F* (and so can be extracted to C) and are proven to both be memory-safe and respect their high-level variants. The type for low-level parsers, parser_impl (p:parser t), denotes an imperative function that reads from an array of bytes and returns a t, behaving as the specificational parser p. Conversely, a serializer_impl (s:serializer p) writes into an array of bytes, behaving as s.

Given such a library, we would like to build verified, mutually inverse, low-level parsers and serializers for specific data formats. The task is mechanical, yet overwhelmingly tedious by hand, with many auxiliary proof obligations of a predictable structure: a perfect candidate for metaprogramming.

Deriving specifications from a type definition Consider the following F* type, representing lists of exactly 18 pairs of bytes.

```
type sample = nlist 18 (u8 ∗ u8)
```

The first component of our metaprogram is gen_specs, which generates parser and serializer specifications from a type definition.

```
let ps_sample : package sample = _ by (gen_specs (`sample))
```

---

[3] F* syntax for refinements is x:t{ϕ}, denoting the type of all x of type t satisfying ϕ.

The syntax _ by τ is the way to call Meta-F* for code generation. Meta-F* will run the metaprogram τ and, if successful, replace the underscore by the result. In this case, the gen_specs (`sample) inspects the syntax of the sample type (§3.3) and produces the package below (seq_p and seq_s are sequencing combinators):

```
let ps_sample = { p = p_nlist 18 (p_u8 `seq_p` p_u8)
                ; s = s_nlist 18 (s_u8 `seq_s` s_u8) }
```

Deriving low-level implementations that match specifications From this pair of specifications, we can automatically generate Low* implementations for them:

```
let p_low : parser_impl ps_sample.p = _ by gen_parser_impl
let s_low : serializer_impl ps_sample.s = _ by gen_serializer_impl
```

which will produce the following low-level implementations:

```
let p_low = parse_nlist_impl 18ul (parse_u8_impl `seq_pi` parse_u8_impl)
let s_low = serialize_nlist_impl 18ul (serialize_u8_impl `seq_si` serialize_u8_impl)
```

For simple types like the one above, the generated code is fairly simple. However, for more complex types, using the combinator library comes with non-trivial proof obligations. For example, even for a simple enumeration, type color = Red | Green, the parser specification is as follows:

```
parse_synth (parse_bounded_u8 2)
            (λ x2 → mk_if_t (x2 = 0uy) (λ _ → Red) (λ _ → Green))
            (λ x → match x with | Green → 1uy | Red → 0uy)
```

We represent Red with 0uy and Green with 1uy. The parser first parses a "bounded" byte, with only two values. The parse_synth combinator then expects functions between the bounded byte and the datatype being parsed (color), which must be proven to be mutual inverses. This proof is conceptually easy, but for large enumerations nested deep within the structure of other types, it is notoriously hard for SMT solvers. Since the proof is inherently computational, a proof that destructs the inductive type into its cases and then normalizes is much more natural. With our metaprogram, we can produce the term and then discharge these proof obligations with a tactic *on the spot*, eliminating them from the final VC. We also explore simply tweaking the SMT context, again via a tactic, with good results. A quantitative evaluation is provided in §6.2.

3   The Design of Meta-F*

Having caught a glimpse of the use cases for Meta-F*, we now turn to its design. As usual in proof assistants (such as Coq, Lean and Idris), Meta-F* tactics work over a set of goals and apply primitive actions to transform them, possibly solving some goals and generating new goals in the process. Since this is standard, we will focus the most on describing the aspects where Meta-F* differs from other engines. We first describe how metaprograms are modelled as an effect (§3.1) and their runtime model (§3.2). We then detail some of Meta-F*'s syntax inspection and building capabilities (§3.3). Finally, we show how to perform some (lightweight) verification of metaprograms (§3.4) within F*.

3.1   An Effect for Metaprogramming

Meta-F⋆ tactics are, at their core, programs that transform the "proof state", i.e. a set of goals needing to be solved. As in Lean [30] and Idris [22], we define a monad combining exceptions and stateful computations over a proof state, along with actions that can access internal components such as the type-checker. For this we first introduce abstract types for the proof state, goals, terms, environments, etc., together with functions to access them, some of them shown below.

```
type proofstate                          val goals_of : proofstate → list goal
type goal                                val goal_env : goal → env
type term                                val goal_type : goal → term
type env                                 val goal_solution : goal → term
```

We can now define our metaprogramming monad: tac. It combines F⋆'s existing effect for potential divergence (Div), with exceptions and stateful computations over a proofstate. The definition of tac, shown below, is straightforward and given in F⋆'s standard library. Then, we use F⋆'s effect extension capabilities [1] in order to elevate the tac monad and its actions to an effect, dubbed TAC.

```
type error = exn ∗ proofstate (∗ error and proofstate at the time of failure ∗)
type result a = | Success : a → proofstate → result a | Failed : error → result a
let tac a = proofstate → Div (result a)
let t_return #a (x:a) = λps → Success x ps
let t_bind #a #b (m:tac a) (f:a → tac b) : tac b = λps → ... (∗ omitted, yet simple ∗)
let get () : tac proofstate = λps → Success ps ps
let raise #a (e:exn) : tac a = λps → Failed (e, ps)
new_effect { TAC with repr = tac ; return = t_return ; bind = t_bind
                                  ; get = get ; raise = raise }
```

The new_effect declaration introduces *computation types* of the form TAC t wp, where t is the return type and wp a specification. However, until §3.4 we shall only use the derived form Tac t, where the specification is trivial. These computation types are distinct from their underlying monadic representation type tac t—users cannot directly access the proof state except via the actions. The simplest actions stem from the tac monad definition: get : unit → Tac proofstate returns the current proof state and raise: exn → Tac α fails with the given exception[4]. Failures can be handled using catch : (unit → Tac α) → Tac (either exn α), which resets the state on failure, including that of unification metavariables. We emphasize two points here. First, there is no "set" action. This is to forbid metaprograms from arbitrarily replacing their proof state, which would be unsound. Second, the argument to catch must be thunked, since in F⋆ impure un-suspended computations are evaluated before they are passed into functions.

The only aspect differentiating Tac from other user-defined effects is the existence of effect-specific primitive actions, which give access to the metaprogramming engine proper. We list here but a few:

```
val trivial : unit → Tac unit      val tc : term → Tac term       val dump : string → Tac unit
```

---

[4] We use greek letters α, β, ... to abbreviate universally quantified type variables.

All of these are given an interpretation internally by Meta-F⋆. For instance, trivial calls into F⋆'s logical simplifier to check whether the current goal is a trivial proposition and discharges it if so, failing otherwise. The tc primitive queries the type-checker to infer the type of a given term in the current environment (F⋆ types are a kind of terms, hence the codomain of tc is also term). This does not change the proof state; its only purpose is to return useful information to the calling metaprograms. Finally, dump outputs the current proof state to the user in a pretty-printed format, in support of user interaction.

Having introduced the Tac effect and some basic actions, writing metaprograms is as straightforward as writing any other F⋆ code. For instance, here are two metaprogram combinators. The first one repeatedly calls its argument until it fails, returning a list of all the successfully-returned values. The second one behaves similarly, but folds the results with some provided folding function.

```
let rec repeat (τ : unit → Tac α) : Tac (list α) =
   match catch τ with | Inl _ → [] | Inr x → x :: repeat τ

let repeat_fold f e τ = fold_left f e (repeat τ)
```

These two small combinators illustrate a few key points of Meta-F⋆. As for all other F⋆ effects, metaprograms are written in applicative style, without explicit return, bind, or lift of computations (which are inserted under the hood). This also works across different effects: repeat_fold can seamlessly combine the pure fold_left from F⋆'s list library with a metaprogram like repeat. Metaprograms are also type- and effect-inferred: while repeat_fold was not at all annotated, F⋆ infers the polymorphic type $(\beta \to \alpha \to \beta) \to \beta \to (\text{unit} \to \text{Tac } \alpha) \to \text{Tac } \alpha$ for it.

It should be noted that, if lacking an effect extension feature, one could embed metaprograms simply via the (properly abstracted) tac monad instead of the Tac effect. It is just more convenient to use an effect, given we are working within an effectful program verifier already. In what follows, with the exception of §3.4 where we describe specifications for metaprograms, there is little reliance on using an effect; so, the same ideas could be applied in other settings.

3.2   Executing Meta-F⋆ Metaprograms

Running metaprograms involves three steps. First, they are *reified* [1] into their underlying tac representation, i.e. as state-passing functions. User code cannot reify metaprograms: only F⋆ can do so when about to process a goal.

Second, the reified term is applied to an initial proof state, and then simply evaluated according to F⋆'s dynamic semantics, for instance using F⋆'s existing normalizer. For intensive applications, such as proofs by reflection, we provide faster alternatives (§5). In order to perform this second step, the proof state, which up until this moments exists only internally to F⋆, must be *embedded* as a term, i.e., as abstract syntax. Here is where its abstraction pays off: since metaprograms cannot interact with a proof state except through a limited interface, it need not be *deeply* embedded as syntax. By simply wrapping the internal proofstate into a new kind of "alien" term, and making the primitives aware of

this wrapping, we can readily run the metaprogram that safely carries its alien proof state around. This wrapping of proof states is a constant-time operation.

The third step is interpreting the primitives. They are realized by functions of similar types implemented within the F⋆ type-checker, but over an internal tac monad and the concrete definitions for term, proofstate, etc. Hence, there is a translation involved on every call and return, switching between embedded representations and their concrete variants. Take dump, for example, with type string → Tac unit. Its internal implementation, implemented within the F⋆ type-checker, has type string → proofstate → Div (result unit). When interpreting a call to it, the interpreter must *unembed* the arguments (which are representations of F⋆ terms) into a concrete string and a concrete proofstate to pass to the internal implementation of dump. The situation is symmetric for the return value of the call, which must be *embedded* as a term.

3.3   Syntax Inspection, Generation, and Quotation

If metaprograms are to be reusable over different kinds of goals, they must be able to reflect on the goals they are invoked to solve. Like any metaprogramming system, Meta-F⋆ offers a way to inspect and construct the syntax of F⋆ terms. Our representation of terms as an inductive type, and the variants of quotations, are inspired by the ones in Idris [22] and Lean [30].

Inspecting syntax Internally, F⋆ uses a locally-nameless representation [21] with explicit, delayed substitutions. To shield metaprograms from some of this internal bureaucracy, we expose a simplified view [61] of terms. Below we present a few constructors from the term_view type:

```
val inspect : term → Tac term_view   type term_view =
val pack : term_view → term                 | Tv_BVar : v:dbvar → term_view
                                            | Tv_Var : v:name → term_view
                                            | Tv_FVar : v:qname → term_view
                                            | Tv_Abs : bv:binder → body:term → term_view
                                            | Tv_App : hd:term → arg:term → term_view
                                            ...
```

The term_view type provides the "one-level-deep" structure of a term: metaprograms must call inspect to reveal the structure of the term, one constructor at a time. The view exposes three kinds of variables: bound variables, Tv_BVar; named local variables Tv_Var; and top-level fully qualified names, Tv_FVar. Bound variables and local variables are distinguished since the internal abstract syntax is locally nameless. For metaprogramming, it is usually simpler to use a fully-named representation, so we provide inspect and pack functions that open and close binders appropriately to maintain this invariant. Since opening binders requires freshness, inspect has effect Tac.[5] As generating large pieces of syntax via the view easily becomes tedious, we also provide some ways of *quoting* terms:

---

[5] We also provide functions inspect_ln, pack_ln which stay in a locally-nameless representation and are thus pure, total functions.

Static quotations A static quotation `e is just a shorthand for statically calling the F$^\star$ parser to convert e into the abstract syntax of F$^\star$ terms above. For instance, `(f 1 2) is equivalent to the following,

pack (Tv_App (pack (Tv_App (pack (Tv_FVar "f"))
                           (pack (Tv_Const (C_Int 1)))))
             (pack (Tv_Const (C_Int 2)))))

Dynamic quotations A second form of quotation is dquote: #a:Type → a → Tac term, an effectful operation that is interpreted by F*'s normalizer during metaprogram evaluation. It returns the syntax of its argument at the time dquote e is evaluated. Evaluating dquote e substitutes all the free variables in e with their current values in the execution environment, suspends further evaluation, and returns the abstract syntax of the resulting term. For instance, evaluating (λx → dquote (x + 1)) 16 produces the abstract syntax of 16 + 1.

Anti-quotations Static quotations are useful for building big chunks of syntax concisely, but they are of limited use if we cannot combine them with existing bits of syntax. Subterms of a quotation are allowed to "escape" and be substituted by arbitrary expressions. We use the syntax `#t to denote an antiquoted t, where t must be an expression of type term in order for the quotation to be well-typed. For example, `(1 + `#e) creates syntax for an addition where one operand is the integer constant 1 and the other is the term represented by e.

Unquotation Finally, we provide an effectful operation, unquote: #a:Type → t:term → Tac a, which takes a term representation t and an expected type for it a (usually inferred from the context), and calls the F* type-checker to check and elaborate the term representation into a well-typed term.

3.4 Specifying and Verifying Metaprograms

Since we model metaprograms as a particular kind of effectful program within F*, which is a program verifier, a natural question to ask is whether F* can specify and verify metaprograms. The answer is "yes, to a degree".

To do so, we must use the WP calculus for the TAC effect: TAC-computations are given computation types of the form TAC a wp, where a is the computation's result type and wp is a weakest-precondition transformer of type tacwp a = proofstate → (result a → prop) → prop. However, since WPs tend to not be very intuitive, we first define two variants of the TAC effect: TacH in "Hoare-style" with pre- and postconditions and Tac (which we have seen before), which only specifies the return type, but uses trivial pre- and postconditions. The requires and ensures keywords below simply aid readability of pre- and postconditions—they are identity functions.

effect TacH (a:Type) (pre : proofstate → prop) (post : proofstate → result a → prop) =
        TAC a (λ ps post' → pre ps $\land$ ($\forall$ r. post ps r $\implies$ post' r))
effect Tac (a:Type) = TacH a (requires (λ _ → $\top$)) (ensures (λ _ _ → $\top$))

Previously, we only showed the simple type for the raise primitive, namely exn $\rightarrow$ Tac α. In fact, in full detail and Hoare style, its type/specification is:

```
val raise : e:exn→ TacH α (requires (λ _ → ⊤))
                          (ensures (λ ps r → r == Failed (e, ps)))
```

expressing that the primitive has no precondition, always fails with the provided exception, and does not modify the proof state. From the specifications of the primitives, and the automatically obtained Dijkstra monad, $F^\star$ can already prove interesting properties about metaprograms. We show a few simple examples.

The following metaprogram is accepted by $F^\star$ as it can conclude, from the type of raise, that the assertion is unreachable, and hence raise_flow can have a trivial precondition (as Tac unit implies).

```
let raise_flow () : Tac unit = raise SomeExn; assert ⊥
```

For cur_goal_safe below, $F^\star$ verifies that (given the precondition) the pattern match is exhaustive. The postcondition is also asserting that the metaprogram always succeeds without affecting the proof state, returning some unspecified goal. Calls to cur_goal_safe must statically ensure that the goal list is not empty.

```
let cur_goal_safe () : TacH goal (requires (λ ps → ¬(goals_of ps == [])))
                                 (ensures (λ ps r → ∃g. r == Success g ps)) =
    match goals_of (get ()) with | g :: _ → g
```

Finally, the divide combinator below "splits" the goals of a proof state in two at a given index n, and focuses a different metaprogram on each. It includes a runtime check that the given n is non-negative, and raises an exception in the TAC effect otherwise. Afterwards, the call to the (pure) List.splitAt function requires that n be statically known to be non-negative, a fact which can be proven from the specification for raise and the effect definition, which defines the control flow.

```
let divide (n:int) (tl : unit → Tac α) (tr : unit → Tac β) : Tac (α ∗ β) =
    if n < 0 then raise NegativeN;
    let gsl, gsr = List.splitAt n (goals ()) in ...
```

This enables a style of "lightweight" verification of metaprograms, where expressive invariants about their state and control-flow can be encoded. The programmer can exploit dynamic checks (n < 0) and exceptions (raise) or static ones (preconditions), or a mixture of them, as needed.

Due to type abstraction, though, the specifications of most primitives cannot provide complete detail about their behavior, and deeper specifications (such as ensuring a tactic will correctly solve a goal) cannot currently be proven, nor even stated—to do so would require, at least, an internalization of the typing judgment of $F^\star$. While this is an exciting possibility [3], we have for now only focused on verifying basic safety properties of metaprograms, which helps users detect errors early, and whose proofs the SMT can handle well. Although in principle, one can also write tactics to discharge the proof obligations of metaprograms.

4   Meta-F⋆, Formally

We now describe the trust assumptions for Meta-F⋆ (§4.1) and then how we reconcile tactics within a program verifier, where the exact shape of VCs is not given, nor known a priori by the user (§4.2).

4.1   Correctness and Trusted Computing Base (TCB)

As in any proof assistant, tactics and metaprogramming would be rather useless if they allowed to "prove" invalid judgments—care must be taken to ensure soundness. We begin with a taste of the specifics of F⋆'s static semantics, which influence the trust model for Meta-F⋆, and then provide more detail on the TCB.

Proof irrelevance in F⋆ The following two rules for introducing and eliminating refinement types are key in F⋆, as they form the basis of its proof irrelevance.

$$
\frac{\text{T-Refine} \quad}{}
$$

T-Refine
$$\frac{\Gamma \vdash e \ : \ t \quad \Gamma \models \phi[e/x]}{\Gamma \vdash e \ : \ x{:}t\{\phi\}}$$

V-Refine
$$\frac{\Gamma \vdash e \ : \ x{:}t\{\phi\}}{\Gamma \models \phi[e/x]}$$

The $\models$ symbol represents F⋆'s *validity judgment* [1] which, at a high-level, defines a proof-irrelevant, classical, higher-order logic. These validity hypotheses are usually collected by the type-checker, and then encoded to the SMT solver in bulk. Crucially, the irrelevance of validity is what permits efficient interaction with SMT solvers, since reconstructing F⋆ terms from SMT proofs is unneeded.

As evidenced in the rules, validity and typing are mutually recursive, and therefore Meta-F⋆ must also construct validity derivations. In the implementation, we model these validity goals as holes with a "squash" type [5, 53], where squash $\phi = \_{:}\text{unit}\{\phi\}$, i.e., a refinement of unit. Concretely, we model $\Gamma \models \phi$ as $\Gamma \vdash ?u : \text{squash } \phi$ using a unification variable. Meta-F⋆ does not construct deep solutions to squashed goals: if they are proven valid, the variable ?u is simply solved by the unit value '()'. At any point, any such irrelevant goal can be sent to the SMT solver. Relevant goals, on the other hand, cannot be sent to SMT.

Scripting the typing judgment A consequence of validity proofs not being materialized is that type-checking is undecidable in F⋆. For instance: does the unit value () solve the hole $\Gamma \vdash ?u : \text{squash } \phi$? Well, only if ϕ holds—a condition which no type-checker can effectively decide. This implies that the type-checker cannot, in general, rely on proof terms to reconstruct a proof. Hence, the primitives are designed to provide access to the typing judgment of F⋆ directly, instead of building syntax for proof terms. One can think of F⋆'s type-checker as implementing one particular algorithmic heuristic of the typing and validity judgments—a heuristic which happens to work well in practice. For convenience, this default type-checking heuristic is also available to metaprograms: this is in fact precisely what the exact primitive does. Having programmatic access to the typing judgment also provides the flexibility to tweak VC generation as needed, instead of leaving it to the default behavior of F⋆. For instance, the refine_intro

primitive implements T-Refine. When applied, it produces two new goals, including that the refinement actually holds. At that point, a metaprogram can run any arbitrary tactic on it, instead of letting the F⋆ type-checker collect the obligation and send it to the SMT solver in bulk with others.

Trust There are two common approaches for the correctness of tactic engines: (1) the *de Bruijn criterion* [6], which requires constructing full proofs (or proof terms) and checking them at the end, hence reducing trust to an independent proof-checker; and (2) the LCF style, which applies backwards reasoning while constructing validation functions at every step, reducing trust to primitive, forward-style implementations of the system's inference rules.

As we wish to make use of SMT solvers within F⋆, the first approach is not easy. Reconstructing the proofs SMT solvers produce, if any, back into a proper derivation remains a significant challenge (even despite recent progress, e.g. [17, 31]). Further, the logical encoding from F⋆ to SMT, along with the solver itself, are already part of F⋆'s TCB: shielding Meta-F⋆ from them would not significantly increase safety of the combined system.

Instead, we roughly follow the LCF approach and implement F⋆'s typing rules as the basic user-facing metaprogramming actions. However, instead of implementing the rules in forward-style and using them to validate (untrusted) backwards-style tactics, we implement them directly in backwards-style. That is, they run by breaking down goals into subgoals, instead of combining proven facts into new proven facts. Using LCF style makes the primitives part of the TCB. However, given the primitives are sound, any combination of them also is, and any user-provided metaprogram must be safe due to the abstraction imposed by the Tac effect, as discussed next.

Correct Evolutions of the Proof State For soundness, it is imperative that tactics do not arbitrarily drop goals from the proof state, and only discharge them when they are solved, or when they can be solved by other goals tracked in the proof state. For a concrete example, consider the following program:

```
let f : int → int = _ by (intro (); exact (`42))
```

Here, Meta-F⋆ will create an initial proof state with a single goal of the form $[\emptyset \vdash ?u_1 : \text{int} \to \text{int}]$ and begin executing the metaprogram. When applying the intro primitive, the proof state transitions as shown below.

$$[\emptyset \vdash ?u_1 : \text{int} \to \text{int}] \rightsquigarrow [\text{x:int} \vdash ?u_2 : \text{int}]$$

Here, a solution to the original goal has not yet been built, since it *depends* on the solution to the goal on the right hand side. When it is solved with, say, 42, we can solve our original goal with $\lambda x \to 42$. To formalize these dependencies, we say that a proof state ϕ *correctly evolves (via f) to* ψ, denoted $\phi \preceq_f \psi$, when there is a generic transformation f, called a *validation*, from solutions to all of ψ's goals into correct solutions for ϕ's goals. When ϕ has n goals and ψ has m goals, the validation f is a function from term^m into term^n. Validations may be

composed, providing the transitivity of correct evolution, and if a proof state ϕ correctly evolves (in any amount of steps) into a state with no more goals, then we have fully defined solutions to all of ϕ's goals. We emphasize that validations are not constructed explicitly during the execution of metaprograms. Instead we exploit unification metavariables to instantiate the solutions automatically.

Note that validations may construct solutions for more than one goal, i.e., their codomain is not a single term. This is required in Meta-F⋆, where primitive steps may not only decompose goals into subgoals, but actually combine goals as well. Currently, the only primitive providing this behavior is join, which finds a maximal common prefix of the environment of two irrelevant goals, reverts the "extra" binders in both goals and builds their conjunction. Combining goals using join is especially useful for sending multiple goals to the SMT solver in a single call. When there are common obligations within two goals, joining them before calling the SMT solver can result in a significantly faster proof.

We check that every primitive action respects the $\preceq$ preorder. This relies on them modeling F⋆'s typing rules. For example, and unsurprisingly, the following rule for typing abstractions is what justifies the intro primitive:

$$\text{T-Fun} \quad \frac{\Gamma, x : t \vdash e : t'}{\Gamma \vdash \lambda(x : t).e \ : \ (x : t) \to t'}$$

Then, for the proof state evolution above, the validation function f is the (mathematical, meta-level) function taking a term of type int (the solution for $?u_2$) and building syntax for its abstraction over x. Further, the intro primitive respects the correct-evolution preorder, by the very typing rule (T-Fun) from which it is defined. In this manner, every typing rule induces a syntax-building metaprogramming step. Our primitives come from this dual interpretation of typing rules, which ensures that logical consistency is preserved.

Since the $\preceq$ relation is a preorder, and every metaprogramming primitive we provide the user evolves the proof state according $\preceq$, it is trivially the case that the final proof state returned by a (successful) computation is a correct evolution of the initial one. That means that when the metaprogram terminates, one has indeed broken down the proof obligation correctly, and is left with a (hopefully) simpler set of obligations to fulfill. Note that since $\preceq$ is a preorder, Tac provides an interesting example of monotonic state [2].

4.2   Extracting Individual Assertions

As discussed, the logical context of a goal processed by a tactic is not always syntactically evident in the program. And, as shown in the List.splitAt call in divide from §3.4, some obligations crucially depend on the control-flow of the program. Hence, the proof state must crucially include these assumptions if proving the assertion is to succeed. Below, we describe how Meta-F⋆ finds proper contexts in which to prove the assertions, including control-flow information. Notably, this process is defined over logical formulae and does not depend at all on F⋆'s WP calculus or VC generator: we believe it should be applicable to any VC generator.

As seen in §2.1, the basic mechanism by which Meta-F* attaches a tactic to a specific sub-goal is assert ϕ by τ. Our encoding of this expression is built similarly to F*'s existing assert construct, which is simply sugar for a pure function _assert of type ϕ :prop $\to$ Lemma (requires ϕ) (ensures ϕ), which essentially introduces a cut in the generated VC. That is, the term (assert ϕ; e) roughly produces the verification condition $\phi \wedge (\phi \Longrightarrow \text{VC}_e)$, requiring a proof of ϕ at this point, and assuming ϕ in the continuation. For Meta-F*, we aim to keep this style while allowing asserted formulae to be decorated with user-provided tactics that are tasked with proving or pre-processing them. We do this in three steps.

First, we define the following "phantom" predicate:

let with_tactic (ϕ : prop) (τ : unit $\to$ Tac unit) $= \phi$

Here ϕ `with_tactic` τ simply associates the tactic τ with ϕ, and is equivalent to ϕ by its definition. Next, we implement the assert_by_tactic lemma, and desugar assert ϕ by τ into assert_by_tactic ϕ τ. This lemma is trivially provable by F*.

let assert_by_tactic (ϕ : prop) (τ : unit $\to$ Tac unit)
                              : Lemma (requires (ϕ `with_tactic` τ)) (ensures ϕ) = ()

Given this specification, the term (assert ϕ by τ; e) roughly produces the verification condition ϕ `with_tactic` $\tau \wedge (\phi \Longrightarrow \text{VC}_e)$, with a tagged left sub-goal, and ϕ as an hypothesis in the right one. Importantly, F* keeps the with_tactic marker uninterpreted until the VC needs to be discharged. At that point, it may contain several annotated subformulae. For example, suppose the VC is VC0 below, where we distinguish an ambient context of variables and hypotheses Δ:

(VC0) $\Delta \models$ X $\Longrightarrow$ ($\forall$ (x:t). R `with_tactic` $\tau_1 \wedge$ (R $\Longrightarrow$ S))

In order to run the τ_1 tactic on R, it must first be "split out". To do so, all logical information "visible" for τ_1 (i.e. the set of premises of the implications traversed and the binders introduced by quantifiers) must be included. As for any program verifier, these hypotheses include the control flow information, postconditions, and any other logical fact that is known to be valid at the program point where the corresponding assert R by τ_1 was called. All of them are collected into Δ as the term is traversed. In this case, the VC for R is:

(VC1) Δ, _:X, x:t $\models$ R

Afterwards, this obligation is removed from the original VC. This is done by replacing it with $\top$, leaving a "skeleton" VC with all remaining facts.

(VC2) $\Delta \models$ X $\Longrightarrow$ ($\forall$ (x:t). $\top \wedge$ (R $\Longrightarrow$ S))

The validity of VC1 and VC2 implies that of VC0. F* also recursively descends into R and S, in case there are more with_tactic markers in them. Then, tactics are run on the the split VCs (e.g., τ_1 on VC1) to break them down (or solve them). All remaining goals, including the skeleton, are sent to the SMT solver.

Note that while the *obligation* to prove R, in VC1, is preprocessed by the tactic τ_1, the *assumption* R for the continuation of the code, in VC2, is left as-is. This is crucial for tactics such as the canonicalizer from §2.1: if the skeleton VC2

contained an assumption for the canonicalized equality it would not help the SMT solver show the uncanonicalized postcondition.

However, not all nodes marked with with_tactic are proof obligations. Suppose X in the previous VC was given as (Y `with_tactic` τ_2). In this case, one certainly does not want to attempt to prove Y, since it is an hypothesis. While it would be *sound* to prove it and replace it by ⊤, it is useless at best, and usually irreparably affects the system. Consider asserting the tautology (⊥ `with_tactic` τ) ⟹ ⊥.

Hence, F⋆ splits such obligations only in strictly-positive positions. On all others, F⋆ simply drops the with_tactic marker, e.g., by just unfolding the definition of with_tactic. For regular uses of the assert..by construct, however, all occurrences are strictly-positive. It is only when (expert) users use the with_tactic marker directly that the above discussion might become relevant.

Formally, the soundness of this whole approach is given by the following metatheorem, which justifies the splitting out of sub-assertions, and by the correctness of evolution detailed in §4.1. The proof of Theorem 1 is straightforward, and included in the appendix. We expect an analogous property to hold in other verifiers as well (in particular, it holds for first-order logic).

Theorem 1. *Let E be a context with $\Gamma \vdash E : prop \Rightarrow prop$, and ϕ a squashed proposition such that $\Gamma \vdash \phi : prop$. Then the following holds:*

$$\frac{\Gamma \vDash E[\top] \quad \Gamma, \gamma(E) \vDash \phi}{\Gamma \vDash E[\phi]}$$

where $\gamma(E)$ is the set of binders E introduces. If E is strictly-positive, then the reverse implication holds as well.

5   Executing Metaprograms Efficiently

F⋆ provides three complementary mechanisms for running metaprograms. The first two, F⋆'s call-by-name (CBN) interpreter and a (newly implemented) call-by-value (CBV) NbE-based evaluator, support strong reduction—henceforth we refer to these as "normalizers". In addition, we design and implement a new *native plugin* mechanism that allows both normalizers to interface with Meta-F⋆ programs extracted to OCaml, reusing F⋆'s existing extraction pipeline for this purpose. Below we provide a brief overview of the three mechanisms.

5.1   CBN and CBV Strong Reductions

As described in §3.1, metaprograms, once reified, are simply F⋆ terms of type proofstate → Div (result a). As such, they can be reduced using F⋆'s existing computation machinery, a CBN interpreter for strong reductions based on the Krivine abstract machine (KAM) [24, 46]. Although complete and highly configurable, F⋆'s KAM interpreter is slow, designed primarily for converting types during dependent type-checking and higher-order unification.

Shifting focus to long-running metaprograms, such as tactics for proofs by reflection, we implemented an NbE-based strong-reduction evaluator for $F^\star$ computations. The evaluator is implemented in $F^\star$ and extracted to OCaml (as is the rest of $F^\star$), thereby inheriting CBV from OCaml. It is similar to Boespflug et al.'s 2011 NbE-based strong-reduction for Coq, although we do not implement their low-level, OCaml-specific tag-elimination optimizations—nevertheless, it is already vastly more efficient than the KAM-based interpreter.

5.2   Native Plugins & Multi-language Interoperability

Since Meta-$F^\star$ programs are just $F^\star$ programs, they can also be extracted to OCaml and natively compiled. Further, they can be dynamically linked into $F^\star$ as "plugins". Plugins can be directly called from the type-checker, as is done for the primitives, which is much more efficient than interpreting them. However, compilation has a cost, and it is not convenient to compile every single invocation. Instead, Meta-$F^\star$ enables users to choose which metaprograms are to be plugins (presumably those expected to be computation-intensive, e.g. canon_semiring). Users can choose their native plugins, while still quickly scripting their higher-level logic in the interpreter.

This requires (for higher-order metaprograms) a form of multi-language interoperability, converting between representations of terms used in the normalizers and in native code. We designed a small multi-language calculus, with ML-style polymorphism, to model the interaction between normalizers and plugins and conversions between terms. See the appendix for details.

Beyond the notable efficiency gains of running compiled code vs. interpreting it, native metaprograms also require fewer embeddings. Once compiled, metaprograms work over the internal, *concrete* types for proofstate, term, etc., instead of over their $F^\star$ representations (though still treating them abstractly). Hence, compiled metaprograms can call primitives without needing to embed their arguments or unembed their results. Further, they can call each other directly as well. Indeed, operationally there is little operational difference between a primitive and a compiled metaprogram used as a plugin.

Native plugins, however, are not a replacement for the normalizers, for several reasons. First, the overhead in compilation might not be justified by the execution speed-up. Second, extraction to OCaml erases types and proofs. As a result, the $F^\star$ *interface* of the native plugins can only contain types that can also be expressed in OCaml, thereby excluding full-dependent types—internally, however, they can be dependently typed. Third, being OCaml programs, native plugins do not support reducing open terms, which is often required. However, when the programs treat their open arguments parametrically, relying on parametric polymorphism, the normalizers can pass such arguments *as-is*, thereby recovering open reductions in some cases. This allows us to use native datastructure implementations (e.g. List), which is much faster than using the normalizers, even for open terms. See the appendix for details.

6   Experimental evaluation

We now present an experimental evaluation of Meta-F⋆. First, we provide benchmarks comparing our reflective canonicalizer from §2.1 to calling the SMT solver directly without any canonicalization. Then, we return to the parsers and serializers from §2.3 and show how, for VCs that arise, a domain-specific tactic is much more tractable than a SMT-only proof.

6.1   A Reflective Tactic for Partial Canonicalization

In §2.1, we have described the `canon_semiring` tactic that rewrites semiring expressions into sums of products. We find that this tactic significantly improves proof robustness. The table below compares the success rates and times for the `poly_multiply` lemma from §2.1. To test the robustness of each alternative, we run the tests 200 times while varying the SMT solver's random seed. The `smt`i`x` rows represent asking the solver to prove the lemma without any help from tactics, where i represents the resource limit (`rlimit`) multiplier given to the solver. This `rlimit` is memory-allocation based and independent of the particular system or current load. For the `interp` and `native` rows, the `canon_semiring` tactic is used, running it using F⋆'s KAM normalizer and as a native plugin respectively—both with an `rlimit` of 1. For each setup, we display the success rate of verification, the average (CPU) time taken for the SMT queries (not counting the time for parsing/processing the theory) with its standard deviation, and the average total time (its standard deviation coincides with that of the queries). When applicable, the time for tactic execution (which is independent of the seed) is displayed. The `smt` rows show very poor success rates: even when upping the `rlimit` to a whopping 100x, over three quarters of the attempts fail.

Note how the (relative) standard deviation increases with the `rlimit`: this is due to successful runs taking rather random times, and failing ones exhausting their resources in similar times. The setups using the tactic show a clear increase in robustness: canonicalizing the assertion causes this proof to always succeed, even at the default `rlimit`. We recall that the tactic variants still leave

	Rate	Queries	Tactic	Total
smt1x	0.5%	0.216 ± 0.001	–	2.937
smt2x	2%	0.265 ± 0.003	–	2.958
smt3x	4%	0.304 ± 0.004	–	3.022
smt6x	10%	0.401 ± 0.008	–	3.155
smt12x	12.5%	0.596 ± 0.031	–	3.321
smt25x	16.5%	1.063 ± 0.079	–	3.790
smt50x	22%	2.319 ± 0.230	–	5.030
smt100x	24%	5.831 ± 0.776	–	8.550
interp	100%	0.141 ± 0.001	1.156	4.003
native	100%	0.139 ± 0.001	0.212	3.071

goals for SMT solving, namely, the skeleton for the original VC and the canonicalized equality left by the tactic, easily dischargeable by the SMT solver through much more well-behaved linear reasoning. The last column shows that native compilation speeds up this tactic's execution by about 5x.

6.2   Combining SMT and Tactics for the Parser Generator

In §2.3, we presented a library of combinators and a metaprogramming approach to automate the construction of verified, mutually inverse, low-level parsers and

serializers from type descriptions. Beyond generating the code, tactics are used to process and discharge proof obligations that arise when using the combinators.

We present three strategies for discharging these obligations, including those of bijectivity that arise when constructing parsers and serializers for enumerated types. First, we used F⋆'s default strategy to present all of these proofs directly to the SMT solver. Second, we programmed a ∼100 line tactic to discharge these proofs without relying on the SMT solver at all. Finally, we used a hybrid approach where a simple, 5-line tactic is used to prune the context of the proof removing redundant facts before presenting the resulting goals to the SMT solver.

The table alongside shows the total time in seconds for verifying metaprogrammed low-level parsers and serializers for enumerations of different sizes. In short, the hybrid approach scales the best; the tactic-only ap-

Size	SMT only	Tactic only	Hybrid
4	178	17.3	6.6
7	468	38.3	9.8
10	690	63.0	19.4

proach is somewhat slower; while the SMT-only approach scales poorly and is an order of magnitude slower. Our hybrid approach is very simple. With some more work, a more sophisticated hybrid strategy could be more performant still, relying on tactic-based normalization proofs for fragments of the VC best handled computationally (where the SMT solver spends most of its time), while using SMT only for integer arithmetic, congruence closure etc. However, with Meta-F⋆'s ability to manipulate proof contexts programmatically, our simple context-pruning tactic provides a big payoff at a small cost.

7   Related Work

Many SMT-based program verifiers [7, 8, 19, 34, 48], rely on user hints, in the form of assertions and lemmas, to complete proofs. This is the predominant style of proving used in tools like Dafny [47], Liquid Haskell [60], Why3 [33], and F⋆ itself [58]. However, there is a growing trend to augment this style of semi-automated proof with interactive proofs. For example, systems like Why3 [33] allow VCs to be discharged using ITPs such as Coq, Isabelle/HOL, and PVS, but this requires an additional embedding of VCs into the logic of the ITP in question. In recent concurrent work, support for *effectful* reflection proofs was added to Why3 [50], and it would be interesting to investigate if this could also be done in Meta-F⋆. Grov and Tumas [39] present Tacny, a tactic framework for Dafny, which is, however, limited in that it only transforms source code, with the program verifier unchanged. In contrast, Meta-F⋆ combines the benefits of an SMT-based program verifier and those of tactic proofs within a single language.

Moving away from SMT-based verifiers, ITPs have long relied on separate languages for proof scripting, starting with Edinburgh LCF [37] and ML, and continuing with HOL, Isabelle and Coq, which are either extensible via ML, or have dedicated tactic languages [3, 29, 56, 62]. Meta-F⋆ builds instead on a recent idea in the space of dependently typed ITPs [22, 30, 42, 63] of reusing the object-language as the meta-language. This idea first appeared in Mtac, a Coq-based tactics framework for Coq [42, 63], and has many generic benefits in-

cluding reusing the standard library, IDE support, and type checker of the proof assistant. Mtac can additionally check the partial correctness of tactics, which is also sometimes possible in Meta-F⋆ but still rather limited (§3.4). Meta-F⋆'s design is instead more closely inspired by the metaprogramming frameworks of Idris [22] and Lean [30], which provide a deep embedding of terms that metaprograms can inspect and construct at will without dependent types getting in the way. However, F⋆'s effects, its weakest precondition calculus, and its use of SMT solvers distinguish Meta-F⋆ from these other frameworks, presenting both challenges and opportunities, as discussed in this paper.

Some SMT solvers also include tactic engines [27], which allow to process queries in custom ways. However, using SMT tactics from a program verifier is not very practical. To do so effectively, users must become familiar not only with the solver's language and tactic engine, but also with the translation from the program verifier to the solver. Instead, in Meta-F⋆, everything happens within a single language. Also, to our knowledge, these tactics are usually coarsely-grained, and we do not expect them to enable developments such as §2.2. Plus, SMT tactics do not enable metaprogramming.

Finally, ITPs are seeing increasing use of "hammers" such as Sledgehammer [14, 15, 54] in Isabelle/HOL, and similar tools for HOL Light and HOL4 [43], and Mizar [44], to interface with ATPs. This technique is similar to Meta-F⋆, which, given its support for a dependently typed logic is especially related to a recent hammer for Coq [26]. Unlike these hammers, Meta-F⋆ does not aim to reconstruct SMT proofs, gaining efficiency at the cost of trusting the SMT solver. Further, whereas hammers run in the background, lightening the load on a user otherwise tasked with completing the entire proof, Meta-F⋆ relies more heavily on the SMT solver as an end-game tactic in nearly all proofs.

8   Conclusions

A key challenge in program verification is to balance automation and expressiveness. Whereas tactic-based ITPs support highly expressive logics, the tactic author is responsible for all the automation. Conversely, SMT-based program verifiers provide good, scalable automation for comparatively weaker logics, but offer little recourse when verification fails. A design that allows picking the right tool, at the granularity of each verification sub-task, is a worthy area of research. Meta-F⋆ presents a new point in this space: by using hand-written tactics alongside SMT-automation, we have written proofs that were previously impractical in F⋆, and (to the best of our knowledge) in other SMT-based program verifiers.

Bibliography

[1] D. Ahman, C. Hriţcu, K. Maillard, G. Martínez, G. Plotkin, J. Protzenko, A. Rastogi, and N. Swamy. Dijkstra monads for free. *POPL*. 2017.

[2] D. Ahman, C. Fournet, C. Hriţcu, K. Maillard, A. Rastogi, and N. Swamy. Recalling a witness: Foundations and applications of monotonic state. *PACMPL*, 2(POPL):65:1–65:30, 2018.

[3] A. Anand, S. Boulier, C. Cohen, M. Sozeau, and N. Tabareau. Towards certified meta-programming with typed Template-Coq. In *ITP*. 2018.

[4] A. W. Appel. Tactics for separation logic. Early Draft, 2006.

[5] S. Awodey and A. Bauer. Propositions as [Types]. *J. Log. and Comput.*, 14 (4):447–471, 2004.

[6] H. Barendregt and H. Geuvers. Handbook of automated reasoning. chapter Proof-assistants Using Dependent Type Systems, pages 1149–1238. Elsevier Science Publishers B. V., Amsterdam, The Netherlands, 2001.

[7] M. Barnett, B. E. Chang, R. DeLine, B. Jacobs, and K. R. M. Leino. Boogie: A modular reusable verifier for object-oriented programs. *FMCO*. 2005.

[8] M. Barnett, R. DeLine, M. Fähndrich, B. Jacobs, K. R. M. Leino, W. Schulte, and H. Venter. The Spec# programming system: Challenges and directions. *VSTTE*. 2005.

[9] B. Barras, B. Grégoire, A. Mahboubi, and L. Théry. Coq reference manual; chapter 25: The ring and field tactic families. Available at `https://coq.inria.fr/refman/ring.html`.

[10] U. Berger and H. Schwichtenberg. An inverse of the evaluation functional for typed lambda-calculus. *LICS*. 1991.

[11] D. J. Bernstein. The Poly1305-AES message-authentication code. *FSE*. 2005.

[12] F. Besson. Fast reflexive arithmetic tactics the linear case and beyond. *TYPES*. 2006.

[13] K. Bhargavan, B. Bond, A. Delignat-Lavaud, C. Fournet, C. Hawblitzel, C. Hriţcu, S. Ishtiaq, M. Kohlweiss, R. Leino, J. Lorch, K. Maillard, J. Pang, B. Parno, J. Protzenko, T. Ramananandro, A. Rane, A. Rastogi, N. Swamy, L. Thompson, P. Wang, S. Zanella-Béguelin, and J.-K. Zinzindohoué. Everest: Towards a verified, drop-in replacement of HTTPS. *SNAPL*, 2017.

[14] J. C. Blanchette and A. Popescu. Mechanizing the metatheory of Sledgehammer. *FroCoS*. 2013.

[15] J. C. Blanchette, S. Böhme, and L. C. Paulson. Extending Sledgehammer with SMT solvers. *JAR*, 51(1):109–128, 2013.

[16] M. Boespflug, M. Dénès, and B. Grégoire. Full reduction at full throttle. *CPP*. 2011.

[17] S. Böhme and T. Weber. Fast LCF-style proof reconstruction for Z3. *ITP*. 2010.

[18] B. Bond, C. Hawblitzel, M. Kapritsos, K. R. M. Leino, J. R. Lorch, B. Parno, A. Rane, S. T. V. Setty, and L. Thompson. Vale: Verifying high-performance cryptographic assembly code. *USENIX Security*. 2017.

[19] L. Burdy, Y. Cheon, D. R. Cok, M. D. Ernst, J. R. Kiniry, G. T. Leavens, K. R. M. Leino, and E. Poll. An overview of JML tools and applications. *STTT*, 7(3):212–232, 2005.

[20] A. Chaieb and T. Nipkow. Proof synthesis and reflection for linear arithmetic. *J. Autom. Reasoning*, 41(1):33–59, 2008.

[21] A. Charguéraud. The locally nameless representation. *Journal of Automated Reasoning*, 49(3):363–408, 2012.

[22] D. R. Christiansen and E. Brady. Elaborator reflection: extending Idris in Idris. *ICFP*, 2016.

[23] E. Cohen, M. Moskal, W. Schulte, and S. Tobies. Local verification of global invariants in concurrent programs. *CAV*. 2010.

[24] P. Crégut. Strongly reducing variants of the Krivine abstract machine. *HOSC*, 20(3):209–230, 2007.

[25] P. Cuoq, F. Kirchner, N. Kosmatov, V. Prevosto, J. Signoles, and B. Yakobowski. Frama-C - A software analysis perspective. *SEFM*. 2012.

[26] Ł. Czajka and C. Kaliszyk. Hammer for Coq: Automation for dependent type theory. *J. Autom. Reason.*, 61(1-4):423–453, 2018.

[27] L. de Moura and G. O. Passmore. Automated reasoning and mathematics. chapter The Strategy Challenge in SMT Solving, pages 15–44. Springer-Verlag, Berlin, Heidelberg, 2013.

[28] L. M. de Moura and N. Bjørner. Z3: an efficient SMT solver. *TACAS*. 2008.

[29] D. Delahaye. A tactic language for the system Coq. *LPAR*. 2000.

[30] G. Ebner, S. Ullrich, J. Roesch, J. Avigad, and L. de Moura. A metaprogramming framework for formal verification. *PACMPL*, 1(ICFP):34:1–34:29, 2017.

[31] B. Ekici, A. Mebsout, C. Tinelli, C. Keller, G. Katz, A. Reynolds, and C. W. Barrett. SMTCoq: A plug-in for integrating SMT solvers into coq. *CAV*, 2017.

[32] A. Erbsen, J. Philipoom, J. Gross, R. Sloan, and A. Chlipala. Simple high-level code for cryptographic arithmetic - with proofs, without compromises. *IEEE S&P*, 2019.

[33] J.-C. Filliâtre and A. Paskevich. Why3 — where programs meet provers. *ESOP*. 2013.

[34] C. Flanagan, K. R. M. Leino, M. Lillibridge, G. Nelson, J. B. Saxe, and R. Stata. PLDI 2002: Extended static checking for Java. *SIGPLAN Notices*, 48(4S):22–33, 2013.

[35] A. Fromherz, N. Giannarakis, C. Hawblitzel, B. Parno, A. Rastogi, and N. Swamy. A verified, efficient embedding of a verifiable assembly language. *PACMPL*, (POPL), 2019.

[36] G. Gonthier. Formal proof–the four-color theorem. *Notices of the AMS*, 55 (11):1382–1393, 2008.

[37] M. J. C. Gordon, R. Milner, and C. Wadsworth. *Edinburgh LCF: A Mechanized Logic of Computation.* Springer-Verlag, 1979.

[38] B. Grégoire and A. Mahboubi. Proving equalities in a commutative ring done right in Coq. *TPHOLs*. 2005.

[39] G. Grov and V. Tumas. Tactics for the Dafny program verifier. *TACAS*. 2016.

[40] C. Hawblitzel, J. Howell, J. R. Lorch, A. Narayan, B. Parno, D. Zhang, and B. Zill. Ironclad Apps: End-to-end security via automated full-system verification. *OSDI*. 2014.

[41] C. Hawblitzel, J. Howell, M. Kapritsos, J. R. Lorch, B. Parno, M. L. Roberts, S. T. V. Setty, and B. Zill. Ironfleet: proving safety and liveness of practical distributed systems. *CACM*, 60(7):83–92, 2017.

[42] J. Kaiser, B. Ziliani, R. Krebbers, Y. Régis-Gianas, and D. Dreyer. Mtac2: typed tactics for backward reasoning in Coq. *PACMPL*, 2(ICFP):78:1–78:31, 2018.

[43] C. Kaliszyk and J. Urban. Learning-assisted automated reasoning with Flyspeck. *JAR*, 53(2):173–213, 2014.

[44] C. Kaliszyk and J. Urban. MizAR 40 for Mizar 40. *JAR*, 55(3):245–256, 2015.

[45] R. Krebbers, A. Timany, and L. Birkedal. Interactive proofs in higher-order concurrent separation logic. *POPL*. 2017.

[46] J.-L. Krivine. A call-by-name lambda-calculus machine. *Higher Order Symbol. Comput.*, 20(3):199–207, 2007.

[47] K. R. M. Leino. Dafny: An automatic program verifier for functional correctness. *LPAR*. 2010.

[48] K. R. M. Leino and G. Nelson. An extended static checker for Modula-3. *CC*. 1998.

[49] A. McCreight. Practical tactics for separation logic. *TPHOLs*. 2009.

[50] G. Melquiond and R. Rieu-Helft. A Why3 framework for reflection proofs and its application to GMP's algorithms. *IJCAR*. 2018.

[51] A. Nanevski, J. G. Morrisett, and L. Birkedal. Hoare type theory, polymorphism and separation. *JFP*, 18(5-6):865–911, 2008.

[52] A. Nanevski, V. Vafeiadis, and J. Berdine. Structuring the verification of heap-manipulating programs. *POPL*. 2010.

[53] A. Nogin. Quotient types: A modular approach. *TPHOLs*. 2002.

[54] L. C. Paulson and J. C. Blanchette. Three years of experience with Sledgehammer, a practical link between automatic and interactive theorem provers. *IWIL*. 2010.

[55] J. Protzenko, J.-K. Zinzindohoué, A. Rastogi, T. Ramananandro, P. Wang, S. Zanella-Béguelin, A. Delignat-Lavaud, C. Hriţcu, K. Bhargavan, C. Fournet, and N. Swamy. Verified low-level programming embedded in F*. *PACMPL*, 1(ICFP):17:1–17:29, 2017.

[56] A. Stampoulis and Z. Shao. VeriML: typed computation of logical terms inside a language with effects. *ICFP*. 2010.

[57] N. Swamy, J. Weinberger, C. Schlesinger, J. Chen, and B. Livshits. Verifying higher-order programs with the Dijkstra monad. *PLDI*, 2013.

[58] N. Swamy, C. Hriţcu, C. Keller, A. Rastogi, A. Delignat-Lavaud, S. Forest, K. Bhargavan, C. Fournet, P.-Y. Strub, M. Kohlweiss, J.-K. Zinzindohoué, and S. Zanella-Béguelin. Dependent types and multi-monadic effects in F*. *POPL*. 2016.

[59] N. Vazou, E. L. Seidel, R. Jhala, D. Vytiniotis, and S. L. Peyton Jones. Refinement types for Haskell. *ICFP*, 2014.

[60] N. Vazou, A. Tondwalkar, V. Choudhury, R. G. Scott, R. R. Newton, P. Wadler, and R. Jhala. Refinement reflection: complete verification with SMT. *PACMPL*, 2(POPL):53:1–53:31, 2018.

[61] P. Wadler. Views: A way for pattern matching to cohabit with data abstraction. *POPL*. 1987.

[62] M. Wenzel. The Isabelle/Isar reference manual. Available at `http://isabelle.in.tum.de/doc/isar-ref.pdf`, 2017.

[63] B. Ziliani, D. Dreyer, N. R. Krishnaswami, A. Nanevski, and V. Vafeiadis. Mtac: A monad for typed tactic programming in Coq. *JFP*, 25, 2015.

[64] J.-K. Zinzindohoué, K. Bhargavan, J. Protzenko, and B. Beurdouche. HACL*: A verified modern cryptographic library. *CCS*. 2017.