

# Metamath Zero: The Cartesian Theorem Prover

Mario Carneiro

Pure & Applied Logic

Carnegie Mellon University

Pittsburgh, PA, United States

mcarneir@andrew.cmu.edu

## Abstract

As the usage of theorem prover technology expands, so too does the reliance on correctness of the tools. Metamath Zero is a verification system that aims for simplicity of logic and implementation, without compromising on efficiency of verification. It is formally specified in its own language, and supports a number of translations to and from other proof languages. This paper describes the abstract logic of Metamath Zero, essentially a multi-sorted first order logic, as well as the binary proof format and the way in which it can ensure essentially linear time verification while still being concise and efficient at scale. Metamath Zero currently holds the record for fastest verification of the `set.mm` Metamath library of proofs in ZFC (including 71 of Wiedijk’s 100 formalization targets), at less than 200 ms. Ultimately, we intend to use it to verify the correctness of the implementation of the verifier down to binary executable, so it can be used as a root of trust for more complex proof systems.

**CCS Concepts** • **Mathematics of computing** → *Mathematical software*; • **Security and privacy** → *Logic and verification*;

**Keywords** Metamath Zero, mathematics, formal proof, verification, metamathematics

## ACM Reference Format:

Mario Carneiro. 2020. Metamath Zero: The Cartesian Theorem Prover. In *Proceedings of The 9th ACM SIGPLAN International Conference on Certified Programs and Proofs (CPP’20)*. ACM, New York, NY, USA, 13 pages.

## 1 Introduction

The idea of using computers to check mathematical statements has been around almost as long as computers themselves, but the scope of formalizations have grown in recent times, both in pure mathematics and software verification, and it now seems that there is nothing that is really beyond our reach if we aim for it. But at the same time, software faces a crisis of correctness, where more powerful systems lead to more reliance on computers and higher stakes for failure. Software verification stands poised to solve this problem, providing a high level of certainty in correctness for critical components.

But software verification systems are themselves critical components, particularly the popular and effective ones. A proof in such a system is only as good as the software that checks it. How can we bootstrap trust in our systems?

This paper presents a formal system, called Metamath Zero (MM0), which aims to fill this gap, having both a simple extensible logical theory and a straightforward yet efficient proof format. Work to prove the correctness theorem is ongoing, but this paper explains the design of the system and how it relates to other theorem provers, as well as general considerations for any bootstrapping theorem prover.

### 1.1 Who verifies the verifiers?

There are two major sources of untrustworthiness in a verification system: the logic and the implementation. If the logic is unsound, then it may be able to prove absurd statements. This problem is well studied in the mathematical and logical literature, and there are a number of formal systems that are widely believed to be trustworthy, such as Peano Arithmetic (PA) and Zermelo-Fraenkel set theory (ZFC), and moreover the relationship between these theories and others (such as type theory, higher order logic, etc.) are well understood.

However, implementation correctness is much harder to establish. Implementation bugs can exist in the theorem prover itself, the compiler for the language, any additional components used by the compiler (the preprocessor, linker, and assembler, if applicable), as well as the operating system, firmware, and hardware. In this area, mathematics and logic holds little sway, and it is “common knowledge” that no non-trivial program is or can be bug-free. The argument for correctness of these systems is largely a social one: the compiler has compiled many programs without any bugs (that we noticed) (except when we noticed and fixed the bugs), so it must work well enough.

What can we do to solve the impasse? Ignoring for the moment the difficulty of proving facts about such complex systems, what is the dream that we can strive for? Once the goal is clear, experience has shown that it only takes a few people a few years to make it happen, and the human cost of formalization drops every day.

One possibility for the “dream statement” is to write down a description of, say, the transistor arrangement in the computer, and assert that as long as the transistors behave according to our understanding of physics of transistors, within some tolerance, the program running on the computer will

perform its intended function (e.g. proving theorems in some axiomatic system). We will not go as far as this, for a few reasons:

- Most common hardware is not open source, so the precise description of those transistors is not available.
- Hardware micro-architecture changes frequently, so even if some multi-year endeavor resulted in a complete proof along these lines, the result would already be out of date when it is completed.

Instead, we will target an instruction set architecture (ISA), which is much more stable between processors. The dominant ISA in the desktop computer space is Intel x86, which was introduced in 1978 with the 8086 processor and has been slavishly binary backward compatible since then, although many additions have been made to the instruction set in the interim. Another advantage of targeting an ISA is that as an interface to hardware, it is specifiable without the need to incorporate details of the physical model and error correction. Additionally, this level of description matches the actual distributable artifact, which is a binary executable, not a physical machine (which is expensive to distribute) and not source code (unless one intends to have all users compile the program).

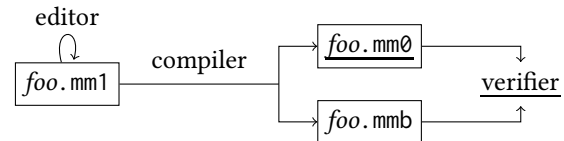
## 1.2 Efficiency matters

Why should it matter if a proof takes hours or days to compile? Besides the obvious problem that no one likes to set up a build job that takes hours, a longer-running proof means a larger window for “attack” from the outside world: more cosmic rays, more OS context switches, more firmware updates, more hardware failures. Generally speaking, getting your work done faster means less possibility of interference of all kinds.

But these are usually negligible concerns. Most of the time a bug either manifests immediately or not at all during a run. An exception is out of memory handling bugs, which are more likely to be exercised in a memory intensive process, so using less memory is one way to mitigate (but not avoid) this problem. Memory allocation bugs are distressingly common, because allocation failure is so rare that the error pathways are almost never tested.

But sometimes, performance is about more than just getting work done a little faster. When something takes *a lot* less time, it changes the way you interact with the computer. A process that takes hours goes on the nightly build server; a process that takes minutes might be a compile that runs on your local machine; a process that takes seconds is a progress bar; and a process that takes milliseconds might happen in an editor between keystrokes.

Furthermore, a verifier is a *component* in a larger system. Program correctness proofs are generally large, so the thing that processes the proof needs to be fast. What is reasonable performance for an end-user tool may not be reasonable for



**Figure 1.** The MM0 workflow. Underlined components are trusted.

a backend library. A program that can verify proofs of OS kernel correctness in under a second can work in the boot stage, providing “secure boot” backed not by social factors but by mathematical proof.

## 1.3 A standalone verifier

Many theorem provers have a “small trusted kernel.” (The term “trusted” here is possibly a misnomer, as it is not necessarily trusted or even trustworthy, but rather correctness-critical. But it is standard usage and we will use it throughout the paper.) The idea is that all trusted code be isolated in a relatively small corner of the program, where it can be inspected for correctness.

But if the goal is end-to-end formal correctness, this approach reveals certain flaws. For example, if the code is written in standard C++, then if undefined behavior is executed anywhere in the program, the entire run could be compromised, including the kernel. This means that correctness of the kernel depends not only on the kernel but also on the correctness (or at least lack of UB) of all code in the project.

We will take a more extreme approach: keep the “small trusted kernel” in its own process, leveraging the process boundary enforced by the operating system. This also means that the verifier is the complete application, so we can reasonably analyze the binary image directly rather than viewing it as a module in a larger code base in which other parts of the code are untrusted. This alternative approach can still be made to work with full formal correctness, but it requires the language to be formalized and the compiler to be proven correct, and the language must not have any “unsafe” features, which limits the capabilities of the untrusted part.

Having an external kernel also frees up everything else in the application from full formal correctness. The user interface to a theorem prover need not be formally correct, even if it contains its own proof checker. Of course it is undesirable for there to be bugs in this code, but errors here are not critical because the external verifier can always pick up the slack. Even if the prover interface is bug-ridden, as long as the exported artifact did not exercise those bugs, and the external verifier checks it, the resulting proof is still correct.

## 1.4 The Metamath Zero architecture

At this point, we have what we need to explain the overall architecture, depicted in Figure 1. As indicated above, the

prover is separated into two pieces. One is the trusted verifier, the MM0 verifier, which checks proofs for correctness. The other is the untrusted prover interface, which produces proof objects from high level code written in an extension of MM0 called MM1, discussed in Section 4.

An essential component of a verifier is the ability to communicate what theorem is being checked. A program that reads an unreadable blob of bytes and reports success or failure only tells the user that *something* is correct without explaining what the *something* is, so either it's not general purpose or it is nearly useless.

There are a few ways to communicate this information: it could be encoded in the input in human readable form, or it could be presented as output by the verifier. We split the input into two parts: the “specification” part (*foo.mm0* in the figure), which is trusted and that contains the human readable statements of theorems, and the “proof” part (*foo.mmb*), which need not be readable or trusted.

Given these general divisions, most of the structural questions answer themselves. The specification file should contain axioms, definitions, and the statement of the main theorem of interest (plus possibly additional theorems to validate that definitions have the appropriate behavior), and the proof file should be some combinatorial structure that guides the verifier to validate that the theorem of interest follows from those axioms.

The remainder of the paper discusses the various components of this process. Section 2 describes the logical framework in which theorems are proved, section 2.4 describes the specification format, Section 3 describes the proof format, and Section 4 discusses how MM0 proof objects can be generated. Section 5 shows work that has been done to connect MM0 to other proof languages.

## 2 The Metamath Zero logic

### 2.1 Metamath

As its name suggests, Metamath Zero is based on Metamath [7], a formal system developed by Norman Megill in 1990. Its largest database, *set.mm*, is the home of over 23000 proofs in ZFC set theory. In the space of theorem prover languages, it is one of the simplest, by design.

The name “Metamath” comes from “metavariable mathematics,” because the core concept is the pervasive use of metavariables over an object logic. An example theorem statement in Metamath is

$$\vdash (\forall x (\varphi \rightarrow \psi) \rightarrow (\forall x \varphi \rightarrow \forall x \psi))$$

which has three “free metavariables:”  $x$ ,  $\varphi$ , and  $\psi$ .  $\varphi$  and  $\psi$  range over formulas of the object logic (let us say first order logic formulas like  $\forall v_0 v_0 = v_1$ ), and  $x$  ranges over variables of the object logic (that is,  $x$  can be  $v_0, v_1, \dots$ ).

However, this object logic never actually appears in actual usage. Rather, a theorem is proved with these metavariables, and then it is later applied with the metavariables (simultaneously) substituted for expressions that will contain more metavariables. For example one could apply the above theorem with the substitution  $\{x \mapsto y, \varphi \mapsto \forall y \varphi, \psi \mapsto x = y\}$  to get:

$$\vdash (\forall y (\forall y \varphi \rightarrow x = y) \rightarrow (\forall y \forall y \varphi \rightarrow \forall y x = y))$$

which again contains metavariables (in this case  $x, y, \varphi$ ) that can be further substituted later.

One consequence of the fact that variables like  $x$  are themselves “variables ranging over variables” is that in a statement like  $\forall x x = y$ , the variable  $y$  may or may not be bound by the  $\forall x$  quantifier, because  $x$  and  $y$  may be the same variable. In order to express that two variables are different, the language includes “disjoint variable provisos”  $A \# B$ , which may be used as preconditions in theorems and assert that variables  $A$  and  $B$  may not be substituted for expressions containing a common variable. This is usually seen in the special cases  $x \# y$ , asserting that  $x$  and  $y$  are not the same variable, and  $x \# \varphi$ , asserting that the substitution to  $\varphi$  does not contain the variable that  $x$  is substituted to.

When a theorem is applied, a substitution  $\sigma$  of all the variables is provided, and for each pair of variables  $A \# B$ , it is checked that for every pair of variables  $v \in \sigma(A), w \in \sigma(B)$ , the disjoint variable condition  $v \# w$  is in the context.

This is essentially the whole algorithm. There is no built in notion of free and bound variable, proper substitution, or alpha renaming — these can all be defined in the logic itself. It turns out that this is not only straightforward to implement (which explains why there are 17 known verifiers written in almost as many languages), but the fundamental operation, substitution, is effectively string interpolation in the sense of `printf`, which can be done very efficiently on modern computers. As a result, Metamath boasts some of the fastest checking times of any theorem prover library; the reference implementation, *metamath.exe*, can check the *set.mm* database of ZFC mathematics in about 8 seconds, and the fastest checker, *smm*, has performed the same feat in 0.7 seconds (on a 2-core Intel i5 1.6GHz).

### 2.2 Shortcomings of Metamath

The primary differences between Metamath and Metamath Zero lie in the handling of first order variables (“variables over variables” from the previous section), expression parsing, and definitions, so some attention is merited to the way these are handled. In each case, Metamath chooses the simplest course of action, possibly at the cost of not making a statement as strong as one would like.

### 2.2.1 Bundling

As has been mentioned already, variables can alias, which leads to a phenomenon known as “bundling” in which a theorem might mean many different things depending on how the variables are substituted. For example,  $\vdash \exists x x = y$  is an axiom in `set.mm` with no disjointness assumptions on  $x$  and  $y$ . There are essentially two different kinds of object language assertions encoded here. If  $i \neq j$ , then  $\vdash \exists v_i v_i = v_j$  asserts that there exists an element equal to  $v_j$ , and when the indices are the same,  $\vdash \exists v_i v_i = v_i$  asserts that there exists an element that is equal to itself. As it happens, in FOL both of these statements are true, so we are comfortable asserting this axiom.

Nevertheless, there is no easy way to render this as a single theorem of FOL, except by taking the conjunction of the two statements, and this generalizes to more variables – a bundled theorem containing  $n$  variables with no disjointness condition is equivalent to  $B_n$  shadow copies of that theorem in FOL, where  $B_n$  is the  $n$ th Bell number. The Bell numbers grow exponentially,  $B_n = e^{O(n \ln n)}$ , so this is at least a theoretical problem.

From the point of view of the Metamath user, this is not actually a problem – this says that Metamath in theory achieves *exponential compression* over more traditional variable handling methods, in which variables with different names are always distinct. However, it is a barrier to translations out of Metamath, because of the resulting exponential explosion.

However, this is not a problem in practice, because the theoretically predicted intricately bundled theorems aren’t written. Usually all or almost all first order variables will be distinct from each other, in which case there is exactly one corresponding FOL theorem (up to alpha renaming). In order to ease translations, MM0 requires that all first order variables be distinct, and shoulders the burden of unbundling in the automatic  $\text{MM} \rightarrow \text{MM0}$  translation (see Section 5).

### 2.2.2 Strings vs trees

Metamath uses strings of tokens in order to represent expressions. That is, the theorem  $\vdash (\varphi \rightarrow \varphi)$  is talking about the provability of the expression consisting of five tokens  $[(, \text{ph}, \rightarrow, \text{ph}, )]$ , with the initial constant  $|$  – distinguishing this judgment from other judgments (for example  $\vdash \varphi$  asserts that  $\varphi$  is provable, while  $\text{wff } \varphi$  asserts that  $\varphi$  is a well formed formula). The upshot of this is that parsing is trivial; spaces between tokens are mandatory so it is often as simple as `tokens = mm_file.split(" ")`. This makes correctness of the verifier simpler because the Metamath specification lines up closely with the internal data representation.

However, this leads to a problem when interpreting expressions as formulas of FOL. The axioms that define the  $\text{wff } \varphi$  judgment can be interpreted as clauses of a context-free grammar, and when that grammar is unambiguous there

is a one-to-one relationship between strings and their parse trees, which are identified with the proofs of  $\text{wff } \varphi$  judgments [2]. So in effect, parsing is not required because the parses are provided with the proof. But unambiguity of a context-free grammar, though true for `set.mm` [1], is undecidable in general, yet is soundness critical – by conflating the two parses of  $\perp \rightarrow \perp \rightarrow \perp$  (if parentheses were omitted in the definition of  $\text{wff } \varphi \rightarrow \psi$ ) it is not difficult to prove a contradiction.

Metamath Zero uses trees (or more accurately dags, directed acyclic graphs) to represent expressions, which has some other side benefits for the proof format (see Section 3). This on its own is enough to prevent ambiguity from leading to unsoundness. However, this means that an MM0 verifier requires a dynamic parser for its operation, which we will discuss in more detail in section 2.4.

### 2.2.3 Definitions

In Metamath, a definition is no more or less than an axiom. Generally a new definition begins with an axiom defining a new syntax constructor, for example  $\text{wff } \exists! x \varphi$ , and an axiom that uses the  $\leftrightarrow$  symbol to relate this syntax constructor with its “definition,” for example

$$y \# x, y \# \varphi \quad \vdash \exists! x \varphi \leftrightarrow \exists y \forall x (\varphi \leftrightarrow x = y).$$

Once again, the correctness of these definitional axioms is soundness critical but not checked by the verifier. Definitions such as the above definition of  $\exists!$  are conservative and eliminable (this is a metatheorem that can be proved outside Metamath), and by convention almost all definitions in `set.mm` have a syntactic form like this, that is, a new constructor  $P(\bar{x})$  is introduced together with an axiom  $\bar{y} \# \bar{y}, \bar{x} \vdash P(\bar{x}) \leftrightarrow \varphi(\bar{x}, \bar{y})$ , where the additional variables  $\bar{y}$  (disjoint from  $\bar{x}$  and each other) are all bound in the FOL sense.

This convention is sufficiently precise that there is a tool that checks these criteria, but this goes beyond the official Metamath specification, and only one of the 17 verifiers supports this check. This effectively means that MM verification in practice extends beyond the narrow definition of MM verification laid out in the standard.

Metamath Zero bakes in a concept of definition, which necessitates a simple convertibility judgment. It also requires an identification of variables that are “bound in the FOL sense,” which means that it can no longer completely ignore the notion of free and bound variables, at least when checking definitions.

## 2.3 The MM0 formal system

MM0 is intended to act as a schematic metatheory over multi-sorted first order logic. This means that it contains *sorts*, two kinds of *variables*, *expressions* constructed from *terms* and *definitions*, and *axioms* and *theorems* using expressions for their hypotheses and conclusion. Theorems have *proofs*, which involve applications of other theorems and axioms.

$\Gamma ::= \cdot \mid \Gamma, x : s \mid \Gamma, \varphi : s \bar{x}$	contexts
$e ::= x \mid \varphi \mid t \bar{e}$	expressions
$A ::= e, \quad \Delta ::= \bar{A}$	statements
$E ::= \bar{\delta}$	environment
$\delta ::= s \text{ sort}$	sorts
$  t : \Gamma \Rightarrow s \bar{x}$	terms
$= \overline{y : s'}. e$	definitions
$  \text{axiom}(\Gamma; \Delta \vdash A)$	axioms
$  \text{thm}(\Gamma; \Delta \vdash A)$	theorems

$\frac{}{\cdot \text{ ctx}}$	$\frac{s \text{ sort} \quad \Gamma \text{ ctx}}{\Gamma, x : s \text{ ctx}}$	$\frac{s \text{ sort} \quad \Gamma \text{ ctx} \quad \overline{x \in \Gamma}}{\Gamma, \varphi : s \bar{x} \text{ ctx}}$
$\frac{(x : s) \in \Gamma}{\Gamma \vdash x : s}$	$\frac{(\varphi : s \bar{x}) \in \Gamma}{\Gamma \vdash \varphi : s}$	$\frac{t : \Gamma' \Rightarrow s \bar{x} \quad \Gamma \vdash \bar{e} :: \Gamma'}{\Gamma \vdash t \bar{e} : s}$
$\frac{}{\Gamma \vdash \cdot :: \cdot}$	$\frac{\Gamma \vdash \bar{e} :: \Gamma' \quad (y : s) \in \Gamma}{\Gamma \vdash \bar{e} y :: \Gamma', x : s}$	$\frac{\Gamma \vdash \bar{e} :: \Gamma' \quad \Gamma \vdash e' : s}{\Gamma \vdash \bar{e} e' :: \Gamma', \varphi : s \bar{x}}$
$\frac{}{s \text{ sort ok}}$	$\frac{s \text{ sort} \quad \Gamma \text{ ctx} \quad \overline{x \in \Gamma}}{t : \Gamma \Rightarrow s \bar{x} \text{ ok}}$	
$\frac{s \text{ sort} \quad \Gamma, \overline{y : s'} \text{ ctx} \quad \overline{x \in \Gamma} \quad \Gamma, \overline{y : s'} \vdash e : s \quad \text{FV}_{\Gamma, \overline{y : s'}}(e) \subseteq \bar{x}}{(t : \Gamma \Rightarrow s \bar{x}) = \overline{y : s'}. e \text{ ok}}$		
$\frac{\Gamma \text{ ctx} \quad \overline{\Gamma \vdash A : s} \quad \Gamma \vdash B : s'}{\text{axiom}(\Gamma; \bar{A} \vdash B) \text{ ok}}$	$\frac{(\Gamma; \bar{A} \vdash B) \text{ ok} \quad \Gamma, \overline{y : s'}; \bar{A} \vdash B}{\text{thm}(\Gamma; \bar{A} \vdash B) \text{ ok}}$	
$\frac{}{\cdot \text{ env}}$	$\frac{E \text{ env} \quad E \vdash \delta \text{ ok}}{E, \delta \text{ env}}$	
$V(x) = \{x\}$		
$V_{\Gamma}(\varphi) = \bar{x}$	where $(\varphi : s \bar{x}) \in \Gamma$	
$V(t \bar{e}) = \bigcup_i V(e_i)$		
$FV(x) = \{x\}$		
$FV_{\Gamma}(\varphi) = \bar{x}$	where $(\varphi : s \bar{x}) \in \Gamma$	
$FV(t \bar{e}) = \underline{FV}(\bar{e} :: \Gamma') \cup \{e_i \mid \Gamma'_i \in \bar{x}\}$	where $t : \Gamma' \Rightarrow s \bar{x}$	
$\underline{FV}(\cdot :: \cdot) = \emptyset$		
$\underline{FV}(\bar{e}, y :: \Gamma', x : s) = \underline{FV}(\bar{e} :: \Gamma')$		
$\underline{FV}(\bar{e}, e' :: \Gamma', \varphi : s \bar{x}) = \underline{FV}(\bar{e} :: \Gamma') \cup (FV(e') \setminus \{e_i \mid \Gamma'_i \in \bar{x}\})$		

**Figure 2.** MM0 syntax and well formedness judgments:  $\Gamma \text{ ctx}$  defines a valid variable context,  $\Gamma \vdash e : s$  is expression typing,  $\Gamma \vdash \bar{e} :: \Gamma'$  is substitution typing,  $\delta \text{ ok}$  checks correctness of an individual statement. All of these have a hidden argument for the global environment  $E$ , which is checked with the  $E \text{ env}$  judgment.  $\text{FV}_{E, \Gamma}(e)$  and  $\text{FV}_{E, \Gamma}(\bar{e} :: \Gamma')$  give the free variables of expressions and substitutions,  $\text{FV}_{E, \Gamma}(e)$  gives all variables. See Figure 3 for the definition of  $\Gamma; \bar{A} \vdash B$ .

$\frac{A \in \Delta}{\vdash A}$	$\frac{(\Gamma'; A \vdash B) \in E \quad \Gamma \vdash \bar{e} :: \Gamma' \text{ safe} \quad \vdash A[\Gamma' \mapsto \bar{e}]}{\vdash B[\Gamma' \mapsto \bar{e}]}$		
$\frac{\vdash A \equiv B \quad \vdash A}{\vdash B}$	$\frac{}{\vdash e \equiv e}$	$\frac{\vdash e \equiv e'}{\vdash e' \equiv e}$	$\frac{\overline{\vdash e \equiv e'}}{\vdash t \bar{e} \equiv t \bar{e}'}$
$(t : \Gamma' \Rightarrow s \bar{x}) = \overline{y : s'}. e' \quad \Gamma \vdash \bar{e}, \bar{z} :: \Gamma', \overline{y : s'} \text{ safe}$			
$\vdash e'[\Gamma', \overline{y : s'} \mapsto \bar{e}, \bar{z}] \equiv e''$			
$\vdash t \bar{e} \equiv e''$			
$\Gamma \vdash \bar{e} :: \Gamma' \text{ safe} \iff \Gamma \vdash \bar{e} :: \Gamma' \quad \text{and}$			
for all $i, j$ , if $\Gamma_i = x \notin V_{\Gamma}(\Gamma_j)$ , then $e_i =: y \notin V_{\Gamma}(e_j)$ .			
$x[\Gamma' \mapsto \bar{e}] = e_i$	where $x = \Gamma'_i$		
$\varphi[\Gamma' \mapsto \bar{e}] = e_i$	where $\varphi = \Gamma'_i$		
$(t \bar{e}')[\Gamma' \mapsto \bar{e}] = t \overline{e'[\Gamma' \mapsto \bar{e}]}$			

**Figure 3.** MM0 proof and convertibility judgments  $\Gamma; \Delta \vdash A$  and  $\Gamma; \Delta \vdash e \equiv e'$ . The arguments  $E, \Gamma, \Delta$  are fixed and hidden.

The remaining sections will go into more detail on each of these points.

### 2.3.1 Sorts

An MM0 file declares a (finite) collection of sorts. Every expression has a unique sort, and an expression can only be substituted for a variable of the same sort. There are no type constructors or function types, so the type system is finite. (Higher order functions are mimicked using open terms, see section 2.3.3.)

### 2.3.2 Variables

MM0 distinguishes between two different kinds of variables. One may variously be called names, first order variables or bound/binding variables. These play the role of “variable variables” from Metamath, and will be denoted in this paper with letters  $x, y, z, \dots$ . They are essentially names that may be bound by quantifiers internal to the logic. “Substitution” of names is alpha renaming; terms cannot be substituted directly for names, although axioms may be used to implement this action indirectly. The other kind of variable may be called a (schematic) metavariable or second order variable, and these may *not* be bound by quantifiers; they are always implicitly universally quantified and held fixed within a single theorem, but unlike names, they may be directly substituted for an expression. We use  $\varphi, \psi, \chi, \dots$  to denote schematic metavariables.

In FOL, notations like  $\varphi(\bar{x})$  are often used to indicate that a metavariable is explicitly permitted to depend on the variables  $\bar{x}$ , and sometimes but not always additional “parameter” variables not under consideration. In MM0, we use a

binder  $(\varphi : s \bar{x})$ , where  $s$  is the sort and  $\bar{x}$  are the *dependencies* of  $\varphi$ , to indicate that  $\varphi$  represents an open term that may reference the variables  $\bar{x}$  declared in the context, as well as any number of “parameter” variables that are not mentioned in the context at all, but not any names that are in the context but missing from the list of dependencies. This is opposite the Metamath convention which requires mentioning all pairs of variables that are *not* dependent, but it is otherwise a merely cosmetic change.

### 2.3.3 Terms

Term declarations are represented in Figure 2 by  $t : \Gamma \Rightarrow s \bar{x}$ . A term is an expression constructor; examples of terms are  $\text{imp} : (\_ : \text{wff}, \_ : \text{wff}) \Rightarrow \text{wff}$ , which defines implication as a binary operator on the sort  $\text{wff}$  (which can be shortened to  $\text{imp} : \text{wff} \Rightarrow \text{wff} \Rightarrow \text{wff}$ ), and  $\text{all} : (x : \text{var}, \varphi : \text{wff } x) \Rightarrow \text{wff}$ , which defines the forall binder. To demonstrate how this actually has the right binding behavior, if we evaluate the definition of  $\text{FV}(\text{all } y \psi)$ , we get

$$\begin{aligned} \text{FV}(\text{all } y \psi) &= \text{FV}(y, \psi :: (x : \text{var}, \varphi : \text{wff } x)) \cup \emptyset \\ &= \text{FV}(y :: (x : \text{var})) \cup (\text{FV}(\psi) \setminus \{(y)_i \mid (x)_i = x\}) \\ &= \emptyset \cup (\text{FV}(\psi) \setminus \{y\}) = \text{FV}(\psi) \setminus \{y\}. \end{aligned}$$

This should be contrasted with  $\text{V}(\text{all } y \psi) = \{y\} \cup \text{V}(\psi)$ . It is easy to see that  $\text{FV}(e) \subseteq \text{V}(e)$  generally; that is, every free variable in an expression  $e$  is present in  $e$ . Metamath, and Metamath Zero, take the somewhat unorthodox approach of using  $\text{V}$  instead of  $\text{FV}$  in the definition of an admissible substitution in theorem application (the  $\Gamma \vdash \bar{e} :: \Gamma'$  safe judgment), but this is clearly sound because  $\text{FV}(e) \subseteq \text{V}(e)$ . This is done because  $\text{V}$  is faster to compute than  $\text{FV}$ , and they are equally expressive, assuming the axioms support alpha conversion, because any expression  $e$  with  $x \in \text{V}(e) \setminus \text{FV}(e)$  is alpha-equivalent to an expression  $e'$  such that  $x \notin \text{V}(e')$ .

### 2.3.4 Definitions

Definitions, denoted by  $(t : \Gamma \Rightarrow s \bar{x}) = \overline{y : s'}$ .  $e$  in Figure 2, are similar to terms in that they are expression constructors, but terms are axiomatic while definitions are conservative, and can be unfolded by the convertibility judgment  $\vdash e \equiv e'$ . One should read the definition as asserting  $\Gamma, \overline{y : s'} \vdash t \Gamma := e$ . The variables  $\bar{y}$  are all required to be bound in  $e$ , but they are added to the context anyway because in MM0 the context must contain all variables in  $\text{V}(e)$ , because it does not expand when traversing binders. The convertibility rule for definitions in Figure 3 substitutes both the variables in  $\Gamma$  as well as the variables  $\bar{y}$ , which provides limited support for alpha renaming.

### 2.3.5 Axioms and theorems

Provable assertions are simply expressions of designated sorts. A general axiom or theorem is really an inference rule  $\Gamma; \Delta \vdash$

$A$ , where  $\Delta$  is a list of hypotheses and  $A$  is a conclusion, and  $\Gamma$  contains the variable declarations used in  $\Delta$  and  $A$ . For example, the Łukaciewicz axioms for propositional logic in this notation are:

$$\begin{aligned} &\varphi \psi : \text{wff}; \cdot \vdash \varphi \rightarrow \psi \rightarrow \varphi \\ &\varphi \psi \chi : \text{wff}; \cdot \vdash (\varphi \rightarrow \psi \rightarrow \chi) \rightarrow (\varphi \rightarrow \psi) \rightarrow (\varphi \rightarrow \chi) \\ &\varphi \psi : \text{wff}; \cdot \vdash (\neg \varphi \rightarrow \neg \psi) \rightarrow (\psi \rightarrow \varphi) \\ &\varphi \psi : \text{wff}; \varphi \rightarrow \psi, \varphi \vdash \psi \end{aligned}$$

Things get more interesting with the FOL axioms:

$$\begin{aligned} &x : \text{var}, \varphi \psi : \text{wff } x; \cdot \vdash \forall x (\varphi \rightarrow \psi) \rightarrow (\forall x \varphi \rightarrow \forall x \psi) \\ &x : \text{var}, \varphi : \text{wff}; \cdot \vdash \varphi \rightarrow \forall x \varphi \end{aligned}$$

Notice that  $\varphi$  has type  $\text{wff } x$  in the first theorem and  $\text{wff}$  in the second, even though  $x$  appears in both statements. This indicates that in the first theorem  $\varphi$  may be substituted with an open term such as  $x < 2$ , while in the second theorem  $\varphi$  must not contain an occurrence of  $x$  (not even a bound occurrence).

One may rightly point out that this restriction seems unnecessary, particularly as we no longer have Metamath’s excuse that the logic has no concept of bound variable. The reason for this choice is twofold. First, this enables compatibility with both Metamath (which would reject such FV-admissible substitutions) and FOL and HOL (which use individual variables rather than names with bundling). Second, it is faster. Theorem application is the hottest loop in the verifier, which has to go through possibly millions of them in a large development, and having a fast path is extremely helpful for this purpose; most theorems don’t need alpha renaming or proper substitution, so those that do can afford a few extra theorem applications, possibly auto-generated by the proof authoring tool, in order to perform the renaming in the logic.

### 2.3.6 Proofs and convertibility

Metamath has only the first two rules of Figure 3: the hypothesis rule, and the application of a theorem after (direct) admissible substitution. Metamath Zero adds the third rule, which consists only of definition unfolding and compatibility rules. Alpha renaming is not directly available, because it is a nonlocal operation, but it can be simulated through the making of definitions (as well as built using theorems, as we endorsed in the previous section).

The rule for  $\text{thm}(\Gamma; \bar{A} \vdash B)$  ok allows additional dummy variables  $\bar{y} : s'$  to be used in the proof, as long as they do not appear in the statement ( $\bar{A}$  and  $B$  must not mention  $\bar{y}$ ). This in particular implies that all sorts are nonempty.

## 2.4 The .mm0 specification format<sup>1</sup>

The .mm0 file is responsible for explaining to the reader what the statement of all relevant theorems is. It closely resembles the axiomatic description of section 2.3, but with a few changes for the sake of clarity.

### 2.4.1 Sort modifiers

Sorts have modifiers that limit what roles they can play. These are enforced by the verifier but not strictly necessary for expressivity.

- Every statement is required to have a provable sort, so that one can assert that if  $x : \text{nat}$  then  $\vdash x$  is nonsense and not permitted.
- The free modifier asserts that a sort cannot be used as a dummy variable, in which case the sort may possibly be empty.
- The strict modifier asserts that the sort cannot be used as a name. This is useful for metavariable-only sorts like wff.
- The pure modifier asserts that the sort has no expression constructors (terms or defs). This is useful for name-only sorts like var.

### 2.4.2 No proofs

As its name implies, the .mm0 specification file is only about specifying axioms and theorems, so it does not contain any proofs. This is an unusual choice for a theorem prover, although some systems like Mizar and Isabelle support exporting an “abstract” of the development, with proofs omitted.

The reason for this comes back to our breakdown of the purpose of the different components of the architecture in section 1.4. Consider the following scenario: You are trying to encourage formalization of some theorem of interest, let's say Fermat's last theorem, so you decide to organize a competition. You write FLT as a .mm0 file, and open it up for the world to submit proof attempts as corresponding .mmb files. *Even in the face of malicious proof attempts*, even if you are receiving gigabytes-long machine learned proofs, you want the assurance that if the verifier accepts it, then the theorem is proved from the axioms you defined. (It would also be nice to know that even if you run the verifier on your local machine containing sensitive information, and auto-run the verifier on network requests directly off the internet, that the verifier will not crash or leak all your data.)

Since an .mm0 file is therefore a *formalization target*, it does not require or even accept proofs of its statements directly in line. Axioms and theorems look exactly the same except for the keyword used to introduce them.

### 2.4.3 Abstract definitions

We can do something similar with definitions. A definition requires a definiens in Figure 2, but we can instead write

a definition with no definiens, so that it looks just like a term declaration. This allows us to assert *the existence* of a term constructor which satisfies any theorems that follow, which gives us a kind of abstraction. Sometimes it is easier to write down characteristic equations for a function rather than an explicit definition, especially in the case of recursive functions.

If we view the entire .mm0 file as a single theorem statement of the metalogic, then this construction corresponds to a second order (constructive) existential quantifier, complementing the second order universal quantifiers that are associated to theorems with free metavariables.

### 2.4.4 Local theorems and definitions

Once one is committed to not proving theorems in the specification file, most dependencies go away. Theorems never reference each other, and only reference terms and definitions involved in their statements. So if focus is given to one theorem, then almost everything else goes away, and even in extreme cases it becomes quite feasible to write down everything up to and including the axiomatic framework in a few hundred lines. In the above example of FLT, the specification file must define the natural numbers and exponentiation, but certainly not modular forms. These are properly the domain of the proof file.

But that means that the proof file must have license to introduce its own definitions and theorems, beyond the ones described in the specification file (but *not* sorts, terms, or axioms). And this is exactly the piece that is missing in Metamath: Forbidding new axioms is necessary in order to prevent a malicious prover from assuming false things, but in MM that also means no new definitions, and that is an untenable expressivity limitation.

### 2.4.5 Notation

In the abstract characterization, we did not concern ourselves with notation, presuming that terms were constructed inductively as trees, but early testing of the concrete syntax revealed that no one likes to read piles of s-expressions, and readability was significantly impacted. The notation system was crafted so as to make parsing as simple as possible to implement, while still ensuring unambiguity, and allowing some simple infix and bracketing notations.

The parser is a precedence parser, with a numeric hierarchy of precedence levels 0, 1, 2, ..., max, forming the order  $\mathbb{N} \cup \{\infty\}$ . (max is the precedence of atoms and parenthesized expressions.) Infix constants are declared with a precedence, and left/right associativity. (An earlier version of MM0 used nonassociative operators and a partial order for precedence levels, but this complicated the parser for no added expressivity. We recognize that overuse of precedence ordering can lead to miscommunication, but this is in the trusted specification file anyway, so the drafter must take care to be clear.)

<sup>1</sup>mm0/mm0.md

General notations are also permitted; these have an arbitrary sequence of constants and variables, and can be used to make composite notations like  $\text{sum\_ } i < n \text{ ai}$ . The only restriction on general notations to make them unambiguous is that they must begin with a unique constant, in this case  $\text{sum\_}$ . This is restrictive, but usually one can get away with a subscript or similar disambiguating mark without significantly hampering readability.

Coercions are functions from one sort to another that have no notation. For example, if we have a sort of set expressions and another sort of class expressions, we might register a coercion  $\text{set} \rightarrow \text{class}$  so that  $x \in y$  makes sense even if  $x$  and  $y$  are sets and  $x \in A$  is a relation between a set and a class. For unambiguity, the verifier requires that the coercion graph have at most one path from any sort to any other.

### 3 The .mmb binary proof file<sup>2</sup>

Having a precise language for specifying formal statements is nice, but it is most powerful when coupled with a method for proving those formal statements. We have indicated several times now design decisions that were made for efficiency reasons. How does MM0 achieve these goals?

In section 1.4, we called the .mmb format a “combinatorial structure” that somehow guides the verifier to a solution. A useful model to keep in mind is that of a powerful but untrustworthy oracle providing hints whenever the verifier needs one, or a nondeterministic Turing machine that receives its nondeterminism from external input.

There are two fundamental principles that guide the design: “don’t search, know the answer,” and “don’t repeat yourself.” By storing more, we end up doing a lot less computation, and by pervasively deduplicating, we can avoid all the exponential blowups that happen in unification. Using these techniques, we managed to translate `set.mm` into MM0 (see Section 5) and verify the resulting binary proof file in  $195 \pm 5$  ms (Intel i7 3.9 GHz, single threaded). While `set.mm` is formidable, at 34 MB / 590 kLOC, we are planning to scale up to larger or less optimized formal libraries to see if it is competitive even on more adversarial inputs.

#### 3.1 High level structure

The proof file is designed to be manipulated in situ; it does not need to be processed into memory structures, as it is already organized like one. It contains a header that declares the sorts, and the number of terms/defs and axioms/theorems, and then links to the beginning of the term table and the theorem table, and the declaration stream.

The term table and theorem table are arrays with fixed size per element, with pointers to additional data for the larger structures. This means that a term lookup is generally a single indexed memory access, and for common terms like

$\sigma ::= e \mid \vdash A \mid e \equiv e' \mid e \stackrel{?}{\equiv} e'$	stack element
$H, S ::= \bar{\sigma}$	heap, stack
Save:	$H; S, \sigma \hookrightarrow H, \sigma; S, \sigma$
Term $t$ :	$S, \bar{e} \hookrightarrow S, e' \quad \left( \begin{array}{l} t : \Gamma' \Rightarrow s \bar{x}, \quad \Gamma \vdash \bar{e} :: \Gamma' \\ e' := \text{alloc}(t \bar{e}) \end{array} \right)$
Ref $i$ :	$H; S \hookrightarrow H; S, H_i$
Dummy $s$ :	$H; S \hookrightarrow H, e; S, e \quad (e := \text{alloc}(x : s), x \text{ fresh})$
Thm $T$ :	$S, \bar{e}^*, A \hookrightarrow S', \vdash A \quad (\text{Unify}(T): S; \bar{e}; A \hookrightarrow_u S')$
Hyp:	$\Delta; S, A \hookrightarrow \Delta, A; S, \vdash A$
Conv:	$S, A, \vdash B \hookrightarrow S, \vdash A, A \stackrel{?}{\equiv} B$
Refl:	$S, e \stackrel{?}{\equiv} e \hookrightarrow S$
Symm:	$S, e \stackrel{?}{\equiv} e' \hookrightarrow S, e' \stackrel{?}{\equiv} e$
Cong:	$S, t \bar{e} \stackrel{?}{\equiv} t \bar{e}' \hookrightarrow S, \bar{e} \stackrel{?}{\equiv} \bar{e}'^*$
Unfold:	$S, t \bar{e}, e' \hookrightarrow S', e' \stackrel{?}{\equiv} e'' \quad (\text{Unify}(t): S; \bar{e}; e' \hookrightarrow_u S', t \bar{e} \stackrel{?}{\equiv} e'')$
ConvCut:	$S, e, e' \hookrightarrow S, e \equiv e', e \stackrel{?}{\equiv} e' \quad \hookrightarrow_u S', t \bar{e} \stackrel{?}{\equiv} e''$
ConvRef $i$ :	$S, e \stackrel{?}{\equiv} e' \hookrightarrow S \quad (H_i = e \equiv e')$
USave:	$U; K, \sigma \hookrightarrow_u U, \sigma; K, \sigma$
UTerm $t$ :	$K, t \bar{e} \hookrightarrow_u K, \bar{e}$
URef $i$ :	$U; K, U_i \hookrightarrow_u U; K$
UDummy $s$ :	$U; K, x \hookrightarrow_u U, x; K \quad (x : s)$
UHyp:	$S, \vdash A; K \hookrightarrow_u S; K, A$

**Figure 4.** Proof stream and unify stream opcodes. Proof steps have the form  $C : \Delta; H; S \hookrightarrow \Delta'; H'; S'$ , and unify steps have the form  $C : S; U; K \hookrightarrow S'; U'; K'$ , but values that do not change are suppressed.  $\bar{e}^*$  denotes the reverse of  $\bar{e}$ .

implication they will invariably already be in cache. This makes type checking for terms  $(\Gamma \vdash e : s)$  extremely fast in practice.

Variable names, term names, and theorem names are all replaced as identifiers with indexes into the relevant arrays. All strings are stored in an index that is placed at the end of the file, linked to from the header, and not touched by the verifier except when it wants to report an error. It is analogous to debugging data stored in executables — it can be stripped without affecting anything except the quality of error reporting.

A term entry contains a table of variable declarations (the context  $\Gamma$  and the target type  $s \bar{x}$ ) followed by a unify stream for definitions, and a theorem entry contains a table of variable declarations (the context  $\Gamma$ ), followed by a unify stream (section 3.2).

<sup>2</sup>`mm0/mm0-c`



### 3.2 The declaration stream

The declaration stream follows the order of declarations in the environment, emitting declarations as it goes. The global state of the verifier is very small; it need only keep track of how many terms, theorems, and sorts have been verified so far, treating some initial segment of the input tables as available for use and the rest as inaccessible. Because terms and theorems are numbered in the same order they appear in the file, when a theorem appears in the declaration stream it is always the one just after the current end of the theorem table.

There are two kinds of opcode streams, proof streams and unify streams. Roughly speaking, a proof stream stores proofs in postfix order (RPN), so for example the proof  $T_1(x, T_2(y))$  would be expressed as Ref  $x$ , Ref  $y$ , Thm  $T_2$ , Ref  $z$ , Term  $f$ , Thm  $T_1$ . A notable difference from metamath is that a theorem with  $n$  variables and  $m$  hypotheses takes  $m + n + 1$  arguments off the stack, where the additional argument is the statement of the theorem to be proved. (Metamath only requires the  $m + n$  arguments, and if first order unification is used only the  $m$  subproofs are truly required to reconstruct the proof.)

By contrast, the unify stream stores expressions (theorem hypotheses and conclusions, and definition bodies) in *prefix* order; so for example  $g(x, f(z))$  would be stored as UTerm  $g$ , URef  $x$ , UTerm  $f$ , URef  $z$ . The reason for this apparent inconsistency is that the proof stream, which is responsible for constructing proofs of intermediate statements, works by *building up* expressions, while the unify stream, which is responsible for proving facts of the form  $e_1[\Gamma \mapsto \bar{e}] = e_2$  (for fixed  $e_1, \Gamma$  and variable  $\bar{e}, e_2$ ), works by *deconstructing* expressions, repeatedly matching the head of  $e_1$  and pushing the pieces using UTerm.

The top level loop reads a declaration from the stream, and does some checking:

- Sorts and terms just check the declaration table and bump the counter.
- A definition  $(t : \Gamma \Rightarrow s \bar{x}) = \overline{y : s'}$ .  $e$  reads a proof stream  $\text{Proof}(t) : \cdot; \Gamma; \cdot \hookrightarrow \cdot; \Gamma, y : s'; e$ , followed by  $\text{Unify}(t) : \cdot; \overline{y : s'}; e \hookrightarrow_u \cdot; \Gamma'; \cdot$ .
- A theorem or axiom  $T : (\Gamma, \Delta \vdash A)$  reads a proof stream  $\text{Proof}(T) : \cdot; \Gamma; \cdot \hookrightarrow \Delta^*; \Gamma, y : s'; \vdash A$  (for axioms, the stack at the end holds  $A$  instead of  $\vdash A$ ), followed by  $\text{Unify}(T) : \cdot; \vdash \Delta; \Gamma; A \hookrightarrow_u \cdot; \Gamma'; \cdot$ .

In short, we build up an expression using the  $\text{Proof}(t)$  proof stream, and then check it against the expression that is in the global space using  $\text{Unify}(t)$ , so that we can safely reread it later.

Convertibility is handled slightly differently than in the abstract formalism. Most of the convertibility rules are inverted, working with a co-convertibility hypothetical. In the absence of  $e \equiv e'$  judgments on the stack, the meaning of the stack is that all  $\vdash A$  statements in it are provable under

the hypotheses  $\Delta$ , but  $S, e \stackrel{?}{\equiv} e'$  means that if  $e \equiv e'$  is provable, then the meaning of  $S$  holds. So for instance, the Conv rule  $S, A, \vdash B \hookrightarrow S, \vdash A, A \equiv B$  says that from  $\vdash B$ , we can deduce that if  $\vdash A \equiv B$  is provable, then  $\vdash A$  holds, which is indeed the conversion rule.

The reason for this inversion is that it makes most unfolding proofs much terser, since all the terms needed in the proof have already been constructed, and the Refl and Cong rules only need to deconstruct those terms.

The ConvCut rule is not strictly necessary, but is available in accordance with the “don’t repeat yourself” principle. It allows for an unfolding proof to be stored and replayed multiple times, which might be useful if it is a frequently appearing subterm. But the proof authoring tool (Section 4) does not generate proofs using it.

Most expressions (elements  $e$  appearing in  $\Delta, H, S, U, K$ ) are pointers into an arena that is cleared after each proof. The nodes themselves store the head and sort of the expression:  $x : s$ ,  $\varphi : s$ , or  $t \bar{e} : s$ , as well as precalculating  $V(e)$  (FV( $e$ ) when constructing definition expressions).

The handling of memory is interesting in that all allocations are “controlled by the user” in the sense that they happen only on Term  $t$  and Dummy  $s$  steps. Because proof streams are processed in one pass, that means that every allocation in the verifier can be identified with a particular opcode in the file. An earlier version of the verifier actually put the data that would otherwise need to be allocated into the instruction itself (i.e. the instruction might be Term  $t \bar{e} V(t \bar{e})$ , and the verifier is responsible for checking that the redundant arguments have the values it expects to see). However, this wastes a lot of space (the  $V(e)$  slots are typically 8 bytes) for ephemeral data. Putting too much data into the proof file means more IO to read it, which can cancel the performance benefits of not having to allocate memory. The memory highwater is under 1 megabyte even after reading the largest proofs in set.mm (which deliberately includes a few stress test theorems), so memory usage doesn’t seem to be a major issue. Nevertheless, it is useful to note that by encoding the heap and stack in the instruction stream, it is possible to perform verification with  $O(1)$  writable memory, streaming almost all of the proof.

But the biggest upshot of letting the user control allocation is that they have complete control over the result of pointer equality. That is, whenever a statement contains a subterm multiple times, for example  $g(f(x), f(x))$ , the user can arrange the proof such that these subterms are always pointers to the same element on the heap (in this example, Ref  $x$ , Term  $f$ , Save, Ref 1, Term  $g$ , assuming that the Save puts  $f(x)$  at index 1). This would not be possible without hash-consing if the verifier “built expressions on its own volition” in the course of performing substitution or applying theorems. As such, the verifier can simply *require* that every

term be constructed at most once (or at least, any expressions that participate in an equality test should be identified), and then expression equality testing in steps like URef and Refl is always constant time.

Verification is not quite linear time, because each Thm  $T$  instruction causes the verifier to read Unify( $T$ ), which is approximately as large as the (deduplicated) statement of  $T$ . It is  $O(mn)$  where  $n$  is the length of the proof and  $m$  is the length of the longest theorem statement, but the statements that exercise the quadratic worst case are rather contrived; they require *one theorem statement* to be on the same order as the whole proof. An example would be if we have a theorem  $T : (a : \text{nat} \vdash a \cdot \bar{n} = 0)$ , where  $\bar{n}$  is a large unary numeral  $S^n(0)$ , and we have a proof that constructs and discards  $T(0), \dots, T(\bar{n})$ , and then proves a triviality. This requires only  $O(n)$  to state (because there are  $O(n)$  expression subterms in the full proof), but each application of  $T(i)$  requires  $O(n)$  to verify because it must match the large term  $\bar{n}$  in the statement of  $T$ .

#### 4 The .mm1 proof authoring file<sup>3</sup>

So far, we have talked about the MM0 verifier, that receives a very explicit proof from some fantasy world in which (untrusted) proofs are easy to produce. But clearly we do not live in that world; formal proofs are becoming more mainstream but are still an order of magnitude more difficult than either informal proofs or unverified but well-tested software. Shouldn't we be focusing more on expanding the use of formal methods, rather than getting needlessly pedantic about what formalization is in the first place?

One response to this objection is simply that it is out of scope for this project. There are many theorem provers that have worked very hard at this problem, but there appears to be no silver bullet for human-computer interaction; a large fraction of formalization is making one's thoughts precise and arguments air-tight, and this is inherent in the problem, so unless we redefine the goal to settle for less, there doesn't seem to be a shortcut.

There are two principal methods for producing .mm0/.mmb pairs: Translate them from another language, or write in a language that is specifically intended for compilation to MM0. (Translations are discussed in Section 5.)

For the bootstrapping project, we used MM0 to specify Peano Arithmetic (PA), and within this axiomatic system we defined the x86 instruction set architecture [3] and the MM0 formal system as defined in section 2.4, to obtain an end-to-end specification from input strings, through lexing, parsing, specification well-formedness, type checking, and proof checking, relating it to the operation of an ELF binary file.

Of these, only the PA framework<sup>4</sup> has been proved thus far, but this already includes about 1000 theorems defining:

- propositional logic,
- natural deduction style,
- first order logic over nat,
- a second-order sort set that ranges over subsets of nat (this is a conservative extension because set is a strict sort; one cannot quantify over sets so they are just syntax sugar over wffs with one free variable),
- a definite description operator the : set  $\rightarrow$  nat (also a conservative extension, allowing the definition of functions like exponentiation from functional relations).
- numerals and arithmetic,
- The Cantor pairing function,
- (signed) integers,
- GCD, Bezout's lemma, the chinese remainder theorem,
- Exponentiation, primitive recursion,
- The Ackermann bijection, finite set theory,
- Functions, lambda and application,
- Lists, recursion and functions on lists.

None of this is particularly difficult, but it does cover the majority of the set-up work for doing metamathematics in PA. But it is enough to get the sense of the scalability of the approach. After compilation, verification takes  $2 \pm 0.05$  ms, which makes sense since it is only a small fraction of the size of set.mm. Compilation is not quite as competitive, at  $5.5 \pm 1$  seconds, but it has not been as aggressively optimized as verification.

The MM1 language, in which peano.mm1 was written, has a syntax which is mostly an extension of MM0 which allows providing proofs of theorems. The main MM0 tool is mm0-hs, a program written in Haskell which provides verification, parsing and translation for all the MM0 family languages, compilation of MM1 files to MMB, and a server compliant with the Language Server Protocol to provide editing support (syntax highlighting, live diagnostics, go-to-definition, hovers, etc.) for Visual Studio Code.

Here we see an important reason for speed: the faster the server can read and execute the file, the faster the response time to live features like diagnostics that the user is relying on for making progress through the proof. We initially intended to add save points in between theorems so that we don't have to process the entire file on each keypress, but the round trip time for diagnostics stayed under half a second throughout the development of peano.mm1, so it never became a sufficiently pressing problem to be worth implementation. (We do not expect that trend to continue, though.)

The MM1 language also contains a Turing-complete meta-programming language based on Scheme. It is intended for writing small "tactics" that construct proofs. Besides a few

<sup>3</sup>mm0/mm0-hs/mm1.mm

<sup>4</sup>mm0/examples/peano.mm1

small quality-of-life improvements, we used it to implement a general algorithm for proving congruence lemmas (theorems of the form  $A = B \rightarrow f(A) = f(B)$ ) for all new definitions.

While MM1 has a long way to go to compete with heavyweights in the theorem proving world like Coq, Isabelle, or Lean, we believe this to be an effective demonstration that even a parsimonious language like Metamath or MM0 can be used as the backend to a theorem prover, and “all” that is necessary is a bit of UI support to add features like a type system, a tactic language, unification, and inference; the mark of a good underlying formal system is that it gets out of your way and lets you say what needs to be said – this is what we mean by “expressivity.”

#### 4.1 Does MM1 generate MM0 files?

The MM1 language directly supports features for being able to generate `.mm0` files. This is one of the reasons why it has a similar syntax; if one deletes the proofs and all local theorems, and additional extensions, then the result is basically a valid `.mm0` file.

Alternatively, one could write a `.mm0` file first, then “fill it out” progressively with proofs until the specification is proved. This is what we did for `peano.mm0`. The approach of generating an `.mm0` file is similar to the “abstract” functionality of Mizar and Isabelle alluded to earlier. But a moment’s consideration of Figure 1 reveals a weakness of this approach: `foo.mm1` is not trusted, but it generates a file `foo.mm0` that is trusted. How is this? It is difficult to trust a build artifact that is hidden away.

We found it helpful to maintain *both* `peano.mm0` and `peano.mm1`, even though they share a lot of common text. When they are both tracked by version control, it makes any changes to the axioms or statements much more obvious, drawing attention to the important parts. The relationship is formally checked, so we need not fear them falling out of alignment. Additionally, it is much easier to make the `.mm0` file look good (clear, unambiguous, well formatted) if it is manually written; much more effort must be put into a formatting tool to get a similar effect.

## 5 MM0 as an interchange format

For a theorem prover to be trustworthy, it must have a semantics. That is, it must be possible for people to look at statements in the `.mm0` file and understand *what they mean*. However, MM0 lets you define your own types, terms, and axioms, and these don’t necessarily make sense. Some theorem provers solve this problem by anointing one axiomatic foundation, and making it either discouraged or impossible to work with others. We believe that this is harmful in the long term, by limiting the ability to perform large scale automated translations.

Instead, we hope to use MM0 to prove correctness of other theorem provers, and vice versa. There is a  $O(n^2)$  problem with having  $n$  mutually supporting bootstraps, as there are  $n^2$  proofs to be done. But the proof of  $B \vdash A$  is correct is closely related to the proof of  $A \vdash A$  is correct; if we had a method for translating proofs in  $A$  to proofs in  $B$ , we would obtain the result immediately. Moreover, proof transformations compose, so it only requires a spider-web of proof connections before we can achieve such a critical mass.

Our work in this area is modest, but it has already been quite helpful. Several times now we have mentioned verification of `set.mm` in MM0, but this is a gigantic library that we would not have a hope of creating without a huge investment of time and effort. Instead, we map MM statements to MM0, and then we obtain tens of thousands of MM0 theorems in one fell swoop, a huge data set for testing that we could not have obtained otherwise.

#### 5.1 Translating MM to MM0

The Haskell verifier `mm0-hs` contains a `from-mm` subcommand that will convert Metamath proofs to MM0. Because of the similarity of the logics, the transformation is mostly cosmetic; unbundling is the most significant logical change. Whenever Metamath proves a theorem of the form  $\vdash T[x, y]$  with no  $x \# y$  assumption, we must generate two theorems,  $\vdash T[x, x]$  and  $\vdash T[x, y]$  (which implicitly assumes  $x \# y$  in MM0). In many cases we can avoid this, for example if  $x$  and  $y$  are not bound by anything, as in  $\vdash x = y \rightarrow y = z \rightarrow x = z$ , we can just make them metavariables instead of names, but some theorems require this treatment, like  $\vdash \forall x x = y \rightarrow \forall y y = x$ .

For definitions, we currently do nothing (we leave them as axioms), but we plan to detect MM style definitional axioms and turn them into MM0 definitions.

#### 5.2 Translating MM0 to HOL systems

The `to-hol` subcommand translates MM0 into a subset of HOL in a very natural way. A metavariable  $\varphi : s \bar{x}$  becomes an  $n$ -ary variable  $\varphi : s_1 \rightarrow \dots \rightarrow s_n \rightarrow s$ , where  $x_i : s_i$ , and all occurrences of  $\varphi$  in statements are replaced by  $\varphi \bar{x}$ . All hypotheses and the conclusion, are universally closed over the names, and the entire implication from hypotheses to conclusion is universally quantified over the metavariables.

For example, the axiom of generalization is

$$x : \text{var}, \varphi : \text{wff } x; \varphi \vdash \text{all } x \varphi,$$

which becomes

$$\forall \varphi : \text{var} \rightarrow \text{wff}, (\forall x : \text{var}, \vdash \varphi x) \Rightarrow \vdash \text{all } (\lambda x : \text{var}, \varphi x)$$

after translation.

The actual output of `mm0-hs to-hol` is a fictional intermediate language (although it has a typechecker), but it is used as a stepping-off point to OpenTheory and Lean. One

of the nice side effects of this work was that Metamath theorems in `set.mm` finally became available to other theorem provers. We demonstrate the utility of this translation by proving Dirichlet’s theorem in Lean<sup>5</sup>, using the number theory library in Metamath for the bulk of the proof and post-processing the statement so that it is expressed in idiomatic Lean style.

## 6 Related work

The idea of a bootstrapping theorem prover is not new. There are a number of notable projects in this space, many of which have influenced the design of MM0. However, none of these projects seem to have recognized (in words or actions) the value of parsimony, specifically as it relates to bootstrapping.

At its heart, a theorem prover that proves it is correct is a type of circular proof. While a proof of correctness can significantly amplify our confidence that we haven’t missed any bugs, we must eventually turn to other methods to ground the argument, and direct inspection is always the fallback. But the effectiveness of direct inspection is inversely proportional to the size of the artifact, so the only way to make a bootstrap argument more airtight is to make it smaller.

### 6.1 CakeML

The most active bootstrapping system today appears to be CakeML [5]. The bootstrap consists of two parts: CakeML is a compiler for ML that is written in the logic of HOL4 [9], and HOL4 is a theorem prover written in ML. Since the completion of the bootstrap in 2014, the CakeML team have expanded downward with *verified stacks* [6], formalizing the hardware of an open source processor design they could implement using an FPGA.

We do not want to diminish the achievements here in any way – it is truly impressive work. But we believe that it does not directly attack the problems that we have set out to deal with. CakeML is a compiler for ML, but one does not need a compiler for ML to have programs that work. In fact, once one has a formalization of low level computer architecture, a sufficiently expressive logic and a theorem prover with metaprogramming capabilities can mostly replace the function of a compiler.

The cost doing more than necessary shows in the compile times: CakeML takes on the order of 14 hours to compile.

### 6.2 Milawa

Milawa [4] is a theorem prover based on ACL2 and developed for Jared Davis’s PhD thesis. It starts with a simple inspectable verifier A which proves the correctness of a more powerful verifier B, which proves verifier C and so on. After another 12 steps or so the verifier becomes practical enough to be able to prove verifier A correct.

From our point of view, the approach taken here suffers from severe blowup problems. Each verifier translates a proof of correctness of verifier A to the next lower level, leading to a constant factor increase in size. As the number of layers grows, the proof grows exponentially, such that the result of all transformations composed together is a proof that is impractical to check by verifier A, which defeats the point.

We diagnose this as a *failure of expressivity*. Davis writes, “to be trustworthy, the Proof Checker takes tiny steps, so proofs are big, and the Theorem Prover is a big system.” But in an expressive logic, one should be able to express *any* computation with a constant factor overhead. In particular, the low level proof checker must be able to mimic the operations of the high level theorem prover without overhead beyond the modeling overhead. A computation that takes linear time should require a proof that is linear in size to verify. MM0 achieves this through the introduction of theorems; a suite of related theorems can be used to perform pattern matching computations, functional updates, and generally achieve the performance of a pure-functional programming language within a constant factor.

This project was later extended by Magnus Myreen to *Jitawa* [8], a Lisp runtime that was verified in HOL4 and can run Milawa, reusing the x86 verification work done for CakeML. Although this isn’t exactly a bootstrap, it is an instance of bootstrap cooperation, of the sort we described in Section 5.

### 6.3 Bootstrappable.org

This is not a theorem prover, but rather a community of projects working on bootstrapping compilers. Here the problem to be solved is that because many compilers are bootstrapped (notably C compilers), the only way to get a C compiler binary is to compile the compiler with a C compiler binary. This leads to some trustworthiness issues because of the *Trusting Trust attack* [10], where malicious content hides in the compiler binary and propagates itself during compilation of the compiler, so that there is no evidence of the bug at all in the source.

To solve this problem, the idea is to have a very simple program, say a hex assembler, which can assemble an assembler, which can assemble a simple compiler, which can compile a more complicated compiler, which can compile gcc. Here we see the tendency to parsimony in force, but correctness is reliant on human verification of all the stages. The advantage of this approach is that while the amount of code that has to be read is significant, at least it is not all machine code.

Our hope is to get verification into the pipeline as early as possible, so that the need to read code and verify correctness is lessened.

<sup>5</sup>`mm0/mm0-lean/mm0/set/post.lean`

## 7 Conclusion

*I think, therefore I prove.*

—René Descartes (paraphrased)

Metamath Zero is a theorem prover which is built to solve the problem of bootstrapping trust into a system. Yet at the same time it is general purpose — it does not use a tailor-made program logic, it uses whatever axioms you give it, so it can support all common formal systems (ZFC, HOL, DTT, PA, really anything recursively enumerable). It is extremely fast, at least on clean inputs, and can handle computer-science-sized problems.

But this is not an attempt to encourage the world to switch to MM0. The nice thing about bootstrapping problems is that language choice is very flexible. We don't all have to commit to one system — different provers, written by different people in different languages to work on different hardware, nevertheless can communicate as long as they share the common language of basic mathematics.

We hope to see a future where all the major theorem provers are either proven correct or can export their proofs to systems that are proven correct, so that when we verify our most important software, we bequeath the highest level of confidence we are capable of providing. It's not an impossible dream — the technology is in our hands; we need only define the problem, and solve it.

## Acknowledgments

I would like to thank Norman Megill for writing Metamath, and André Bacci, Wolf Lammen, David A. Wheeler, Giovanni Mascellani, Seul Baek, and Jeremy Avigad for their input and suggestions during the design phase of MM0. I thank Giovanni Mascellani for his support and for bringing my attention to the Bootstrappable project, and I thank Jeremy Avigad and Jesse Han for their reviews of early versions of this work.

This work was supported in part by AFOSR grant FA9550-18-1-0120 and a grant from the Sloan Foundation.

## References

- [1] Mario Carneiro. 2013. Grammar ambiguity in set.mm. (2013). <http://us.metamath.org/downloads/grammar-ambiguity.txt>
- [2] Mario Carneiro. 2016. Models for Metamath. (2016). arXiv:math.LO/1601.07699 presented at CICM 2016.
- [3] Mario Carneiro. 2019. Specifying verified x86 software from scratch. In *Workshop on Instruction Set Architecture Specification (SpISA 2019)*. [https://www.cl.cam.ac.uk/~jrh13/spisa19/paper\\_07.pdf](https://www.cl.cam.ac.uk/~jrh13/spisa19/paper_07.pdf)
- [4] Jared Curran Davis and J Strother Moore. 2009. *A self-verifying theorem prover*. Ph.D. Dissertation. University of Texas.
- [5] Ramana Kumar, Magnus O. Myreen, Michael Norrish, and Scott Owens. 2014. CakeML: A Verified Implementation of ML. *SIGPLAN Not.* 49, 1 (Jan. 2014), 179–191. <https://doi.org/10.1145/2578855.2535841>
- [6] Andreas Löf, Ramana Kumar, Yong Kiam Tan, Magnus O Myreen, Michael Norrish, Oskar Abrahamsson, and Anthony Fox. 2019. Verified Compilation on a Verified Processor. In *Proceedings of the 40th ACM SIGPLAN Conference on Programming Language Design and Implementation*. ACM, 1041–1053.
- [7] Norman Megill and David A. Wheeler. 2019. *Metamath: A Computer Language for Mathematical Proofs*. Lulu Press.
- [8] Magnus O Myreen and Jared Davis. 2011. A verified runtime for a verified theorem prover. In *International Conference on Interactive Theorem Proving*. Springer, 265–280.
- [9] Konrad Slind and Michael Norrish. 2008. A Brief Overview of HOL4. In *International Conference on Theorem Proving in Higher Order Logics*. Springer, 28–32.
- [10] Ken Thompson et al. 1984. Reflections on trusting trust. *Commun. ACM* 27, 8 (1984), 761–763.