Research Plan

Joseph Tooby-Smith

My research lies at the intersection of computer science, mathematics, and physics. I am particularly interested in building bridges between these disciplines using interactive theorem proving. I hold a PhD in theoretical physics from the University of Cambridge, and have completed a postdoctoral fellowship at Cornell University, where I focused on applying theorem proving software to physics. In the academic year 2024-2025, I am undertaking a postdoctoral position in computer science at the University of Reykjavik.

The past

A significant part of my past research has involved digitizing results from high-energy physics into a Lean 4 library called "HepLean." Lean 4 is an interactive theorem prover that automatically verifies the correctness of proofs. The main motivations behind this project include:

1. A linear-storage of information: Currently information in high energy physics is stored non-linearly, with material related to the same subject spread across papers indexed only by an arXiv number at best. In HepLean results are stored linearly, with results related to the same area of high energy physics stored together. The benefit of this is that look-up of information becomes easier.

2. Automated methods to derive new results: In Lean, one can write "tactics" to attempt automatic completion of proofs. There is also ongoing work on using AI to prove theorems in Lean, with such proofs being automatically certified as correct. Notably, Google DeepMind has explored this in the context of the math olympiad, .

3. Automatic review of results for mathematical correctness: HepLean provides a method for authors and reviewers to be confident of the mathematical correctness of results. In this regard, Lean's automatic proof verification ensures the rigor and accuracy of results.

4. New pedagogy methods: HepLean offers a novel approach to teaching both functional programming and physics, allowing students to engage in active research and learn through direct interaction with theorem proving software.

This is the first anything like this has being attempted in high energy physics. However, there is a similar project for mathematics called Mathlib, which forms the mathematical foundation underlying HepLean.

The future

My future research will continue the development of HepLean, with the aim of making it more accessible to physical scientists. To achieve this overarching goal, I plan to undertake the following specific research steps:

1. Blueprint: Lean 4, a "blueprint" is an English-written document that outlines all the steps needed to convert an English-written proof into a Lean proof. I plan to create such blueprints for various theories in physics, which will serve as pseudo-code, bridging the gap between traditional physics work and interactive theorem provers.

2. AI: While AI has been applied to mathematical problems in Lean (e.g., DeepMind's work on math Olympiad problems), I aim to extend this to physics. I plan to create a dataset of Lean 4-written theorems from physics that can be used for AI testing and training.

3. Tensors: High-energy physics relies heavily on tensors. I plan to develop Lean 4 tactics that can formally verify results related to tensors, which will be crucial for ensuring the correctness of complex calculations in physics.

There will be also be a focus on HepLeans theoretical and computer science foundations. This will involve applying techniques from functional programming, AI, and category theory to create foundational definitions and functions to ease future development.

Plan for undergraduate student involvement

HepLean offers numerous opportunities for involving undergraduate students in research. I plan to develop three lists of undergraduate-level projects around HepLean:

- Functional programming projects: As an example: the handling of lists in Lean to efficiently undertake computations needed for index notation (a notation used by physicists to deal with tensors). Other examples will involve meta-programming in Lean to make the user-experience easier.
- AI in Physics and Mathematics: These projects will explore auto-formalization of theorems in physics (converting human-written results into Lean proofs) and the inverse process, "auto-informalization." While these techniques have been explored in mathematics, they remain largely unexplored in the physical sciences, as dicussed above.
- Interdisciplinary Theorem Proving: These projects will involve proving physics theorems using Lean. Many such problems require minimal prerequisites in physics once the theorem is stated. For instance, formalizing properties of the two-Higgs doublet model potential could be an excellent project for students. Physicists are interested in its properties, such as its minima, whether it is bounded or not etc.

Each of these project lists will include homework-style tasks that can be completed in a few hours and more detailed thesis-level projects. The breadth of HepLean makes it relatively easy to generate a diverse range of projects.

Other research

In addition to my work on HepLean, I have a strong background in applying category theory outside of pure mathematics. Historically, I have used category theory as a language to reframe problems in the physical sciences and derive new results. For example, in high-energy physics, there is a relatively new concept called "generalized symmetry." In we used higher topoi - a special type of higher category - to define and derive new results about these symmetries. Higher topos theory is actually related to homotopy type theory, which led me to interactive theorem proving.

Category theory is also heavily used in functional programming. For instance, the concept of a monad in functional programming is a categorical definition. A monad is a special case of a more general object in the theory of Higher algebra. This is an area I have expertise, since it overlaps with my use of category theory in physics.

In the future, I plan to investigate the role that higher category theory can play in functional programming and, more specifically, in interactive theorem provers for the physical sciences. Currently, I am exploring the theory of modular operads to develop an efficient method for index notation of tensors (mentioned briefly above).