Lambda: The Ultimate Sublanguage (Experience Report)

JEREMY YALLOP, University of Cambridge, UK
LEO WHITE, Jane Street Capital, UK

We describe our experience teaching an advanced typed functional programming course based around the use of System F_ω as a programming language.

CCS Concepts: • Social and professional topics → Computer science education; • Mathematics of computing → Lambda calculus; • Software and its engineering → Functional languages.

Additional Key Words and Phrases: education, pedagogy, types, functional programming, lambda calculus, mental models, sublanguages

ACM Reference Format:

1 INTRODUCTION

Feature-rich programming languages can be challenging to learn. One source of difficulty is the tendency for advanced features to “leak out” unexpectedly in error messages. For example, a student of Haskell, having learned about function composition and \texttt{map}, might make the following erroneous attempt at a function that calculates the successors of a list of integers:

\begin{verbatim}
Prelude> let succlist = map . (1+) and be baffled by the resulting error message:
\end{verbatim}

\begin{verbatim}
Non type-variable argument in the constraint: Num (a -> b)
(Use FlexibleContexts to permit this)
\end{verbatim}

The error itself (composition in place of application) is elementary, but the diagnostic mentions several advanced concepts that are unlikely to help a beginner identify the problem. The root of this communication failure is a mismatch in models: the compiler describes the error in terms of the full language definition, but the beginner’s mental model covers only a subset of the language.

One approach to overcoming the leaky error message problem is to use a sublanguage [Brusilovsky et al. 1994; Pagan 1980]: an implementation of a programming language that supports only a subset of its features. For example, the Helium sublanguage of Haskell [Heeren et al. 2003] excludes type classes, removing a large set of concepts (such as “constraints” and “FlexibleContexts”) from the language, so that errors are more likely to be reported in terms that beginners can understand. Similarly, DrRacket supports several sublanguages (called “language levels”) that remove macros, quasiquotation, higher-order functions, etc., improving the clarity of error messages for programs written by beginners [Marceau et al. 2011].
The problem that sublanguages tackle is central to learning. Learning typically involves building a mental model and incrementally refining it by integrating new facts. By bringing the language implemented by the compiler closer to the language learnt by a beginner, sublanguages make it much easier to integrate facts such as type mismatch reports into the learner’s mental model.

Sublanguages are an effective way to learn the basics of a language. But how can students move from basic competency to mastery — that is, to a point where the intricate details of a language and sophisticated programming techniques can be readily understood? Here, too, the sublanguage approach can help. A strength of most functional languages is their definition in terms of simpler constructs: deep pattern matching may be translated to shallow cases; overloading may be translated away; syntactic sugar may be expanded into a tiny core. Treating this core as a sublanguage can help students to develop intuition — that is, a clear mental model — that provides a solid foundation for understanding advanced language features and techniques that would otherwise appear ad-hoc.

This paper describes our experience teaching a course based on this approach over the last four years. The first week of the course involves a typed lambda calculus, System F_\omega, that supports only abstraction and application. Using this tiny language students explore language features found in full-featured typed functional languages: nested types and GADTs, first-class polymorphism, higher-kindred types, existentials, higher-order modules, etc. Later, the course introduces these features directly, building on the intuition developed in the early parts. The intuitions developed using System F_\omega allow us to cover ground rapidly, and eight weeks later students are familiar with many of the topics beloved of ICFP participants and able to write and understand sophisticated typed functional programs involving indexed data, staging constructs, algebraic effects, and more.

1.1 Outline

Our thesis, then, is that System F_\omega is an effective sublanguage for teaching advanced functional programming, and the remainder of the paper presents our experience as supporting evidence.

Section 2 gives an overview of the course: its principles, structure and content (Section 2.1), its assessment (Section 2.2) and the tools students use to learn System F_\omega (Section 2.3).

Section 3 discusses a broad range of System F_\omega programming exercises, and their relation to the topics covered in the later sections of the course: data types and folds (Section 3.1), type equality (Section 3.2), non-regular types (Section 3.3), GADTs (Section 3.4), proofs (Section 3.5), semirings (Section 3.6), existentials (Section 3.7) and modules (Section 3.8).

Developing and teaching a new course is often an educational experience for instructors as well as students. Section 4 discusses some lessons we learnt from running the course.

Section 5 covers related work, focusing on books and courses that take a similar approach.

2 A COURSE IN ADVANCED FUNCTIONAL PROGRAMMING

Our course is a unit on a one-year masters course; it runs for eight weeks (Figure 1), with two lectures each week. The course focuses on OCaml, including extensions and future features, such as modular implicits [White et al. 2014] (week 6), algebraic effects [Dolan et al. 2015] (week 7) and multi-stage programming [Kiselyov 2014] (week 8).

The aim of the course is to develop familiarity with advanced typed functional programming constructs and techniques: students are expected to reach a level where they can readily grasp much of the research presented at ICFP.

We typically have between 10 and 25 registered students, and other students and staff from the department often join the lectures.
2.1 Course Structure and Content

The first week serves as a kind of précis of the course: many of the ideas that are studied in detail in OCaml later on appear in miniature using System F_\omega_ here. For example, week 5 covers programming with indexed data such as GADTs. However, all the components of GADTs — type equality witnesses (Section 3.2), polymorphic recursion (Section 3.3) and existential types (Section 3.7) appear in the first week, in the context of System F_\omega_. In some years the students have even learned to program with encodings of GADTs directly in System F_\omega_ (Section 3.4) in the first assessed exercise (Section 2.2) before revisiting GADTs using OCaml’s language support in week 5.

Why System F_\omega_? The System F_\omega_ calculus has long played a central role in the theory and practice of typed functional programming. Compilers for ML and Haskell use intermediate languages based on System F_\omega_ [Morrisett et al. 1999; Sulzmann et al. 2007], and many features of functional languages — higher-rank and higher-kindied polymorphism [Jones 1993; Peyton Jones et al. 2007; Vytiniotis et al. 2008; Yallop and White 2014], recursion over non-regular types [Bird and Meertens 1998], type classes [Hall et al. 1996], higher-order modules [Rossberg et al. 2014], existential types (Section 3.7), and more — have a straightforward elaboration into the calculus. Further, System F_\omega_ can be used as the elaboration language for type inference [Pottier 2014], helping to elucidate the limitations of inference, relating to the value restriction [Wright 1995]\(^1\) and its relaxation in OCaml [Garrigue 2004], polymorphic recursion [Henglein 1993] and higher-rank types [Garrigue and Rémy 1999; Peyton Jones et al. 2007; Wells 1994], which we touch on in weeks 2 and 3.

Similarly, System F_\omega_ provides a useful working model for type equality (Section 3.2), with some limitations (e.g. the System F_\omega_ encodings do not support the injectivity principle.)

Most other topics found in the course can be similarly based around System F_\omega_, which provides a useful basis for understanding the Curry-Howard correspondence (Section 3.5) and parametricity [Wadler 2003], can give a semantics to multi-stage programming [Kameyama et al. 2008], and offers features essential for generic programming [Siek and Lumsdaine 2005].

Finally, the remarkable expressive power of System F_\omega_ exposes some limitations of functional programming languages: for example, OCaml’s type language supports only first-order applications, and Haskell’s supports only higher-order applications; neither supports type abstractions (although it is often possible to simulate abstractions via lambda-lifting [Lindley 2012]); contrariwise, the limitations of System F_\omega_, which does not support primitive effects or general recursion, help to make clear exactly what can be accomplished without those facilities (Cf. Turner [2004]).

---

\(^1\)Elaborating Hindley-Milner implicit type schemes into System F_\omega_-style type abstractions justifies Wright’s value restriction: it is only safe to generalize when the body is a value, since in other cases inserting a type abstraction would change the evaluation and sharing behaviour: Kiselyov [2015] further explores connections between sharing and generalization.
The assignments seemed quite hard at first, but everything got clearer once I started reviewing the material. It was very rewarding to understand most of it in the end\textsuperscript{2} (2017)

Without the assignments the lectures are too abstract to understand on anything but a high level. The assignments really tied things together well. (2018)

The course is assessed by three programming exercises (issued around weeks 2, 5 and 8), which students submit two weeks after they are issued. There is no final exam.

The exercises are a crucial part of the course; the lectures and notes present principles, but the exercises build the familiarity that underlies intuition, and ensure that students steadily master functional programming principles as the course progresses. The exercises follow a similar structure each year: the first, whose contents this paper describes in some detail, focuses on System F\(\omega\) programming, establishing a foundation for the remainder of the course.

Balancing pedagogy with assessment is challenging with questions about System F\(\omega\) and typeful programming more generally [Cardelli 1989]. Section 4.3 discusses our response to this challenge.

\subsection*{2.3 Tooling}

We provide a primitive System F\(\omega\) interpreter to support the exercises, based on the implementation by Pierce [2002]. Figure 2 shows the supported syntax: type variables (denoted by Greek letters or capitals), term variables (lowercase), kinds (formed from \(\ast\) and \(\Rightarrow\)), type expressions (variables, function types, quantified types, abstractions and applications), and terms (variables, term abstractions and applications, and type abstractions and applications). Type binders (quantification, type functions, and type abstractions) can bind type variables of arbitrary kind.

There are a few concessions to usability: the interpreter also supports top-level definitions and signatures and some additional types (sums, products and existentials) which could be encoded but are more convenient to have as primitives. Furthermore, kind annotations may be omitted for binders of type \(\ast\); all other binders must be type- or kind-annotated.

Besides a command-line tool, we also provide a simple web interface to the System F\(\omega\) interpreter based on js\_of\_ocaml [Vouillon and Balat 2014]. Section 4.1 discusses the effectiveness of this toolchain.

\section{Programming with System F\(\omega\)}

\subsection{Data Types and Folds in System F\(\omega\)}

Folds, over lists and over other data types, are central to functional programming. They are particularly useful in System F\(\omega\), where they can be used (as typed Church encodings) in place of data types, which are not available in the language. The first exercise on the course often involves a simple fold-based encoding of a tree type together with some functions (such as \texttt{sum} or \texttt{depth}) defined on trees (Figure 3).

As well as providing a useful warm-up, encodings of data types illustrate the usefulness of higher kinds, and the value of the extra expressive power offered by System F\(\omega\). In System F it is possible

\textsuperscript{2}We include excerpts from students’ anonymized feedback throughout the paper.
Fig. 3. Folds for tree types. Note the correspondence between the OCaml and System F\(\omega\) definitions to encode the type \(t\) \texttt{list} for any given \(t\), but not the type constructor \texttt{list} (which has kind \(* \Rightarrow \ast\)) itself. With System F\(\omega\) it becomes possible to encode both \texttt{list} and a wide variety of other data types, including the nested (non-regular) types that appear in advanced functional programming techniques [Bird and Paterson 1999; Hinze 1999] (Section 3.3).

Further, data type encodings based on folds extend neatly to the so-called final tagless style [Carette et al. 2009] commonly used to define DSLs. At first, writing functions compositionally seems unnatural and sometimes inefficient; however, at a larger scale compositionality becomes a boon, supporting elaborate optimizations of DSLs [Kiselyov 2016] and even self-representations of typed languages [Brown and Palsberg 2017].

### 3.2 Type Equality in System F\(\omega\)

Type equality — in particular, first-class equality witnesses — plays a significant role in typed functional programming, in meta-programming [Pasalic 2004], in intermediate language design [Sulzmann et al. 2007], and as the basis for GADTs [Johann and Ghani 2008].

Encoding type equalities in System F\(\omega\) therefore serves as a first step towards implementing GADTs; indeed, decomposing GADTs into their constituent parts (type equality (Section 3.2), existential types (Section 3.7), and non-regularity (Section 3.3)) robs them of much of their mystery. In high-level languages such as OCaml and Haskell, type equalities are automatically introduced into the context by the compiler in GADT pattern matches [Garrigue and Rémy 2013]. However, it is instructive to deal with type equalities that must be explicitly constructed and applied.

Figure 4 presents the elements of an exercise from the 2016 instance of the course, where students were asked to construct a fold-based encoding of the equality GADT and show its equivalence to the more common Leibniz encoding. The so-called Leibniz encoding \(\text{Eq}\) (which is given in roughly this form by Church [1940], and appears more recently in various functional programming papers [Baars and Swierstra 2002; Weirich 2004; Yallop and Kiselyov 2010]) appears at the top left: it says that types \(\alpha\) and \(\beta\) are equal if \(\alpha\) can be turned into \(\beta\) in any context \(\phi\). The fold-based encoding \(\text{Equal}\) appears in the top right: it directly corresponds to the OCaml definition of \(\text{eq}\): the polymorphic argument \(\forall \gamma. \phi \gamma \gamma\) corresponds to the type of the \texttt{Refl} constructor \(\langle \texttt{\_x}, \texttt{\_x} \rangle\) \texttt{eq}, and the result type corresponds to \(\langle \texttt{\_a}, \texttt{\_b} \rangle\) \texttt{eq}. Both encodings fundamentally rely on higher-kinded polymorphism, and cannot be defined in System F.

The students’ task was to provide conversions between \(\text{Eq}\) and \(\text{Equal}\), and further to show that \(\text{Equal}\) represents an equivalence relation by defining values that encode reflexivity, symmetry and transitivity properties; sample solutions are given in Figure 4.
Two representations of $\sigma \equiv \tau$ in System $\mathrm{F}\omega$

<table>
<thead>
<tr>
<th>Eq</th>
<th>Equal</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\texttt{Eq} :: \ast \Rightarrow \ast \Rightarrow \ast$</td>
<td>$\texttt{Equal} :: \ast \Rightarrow \ast \Rightarrow \ast$</td>
</tr>
<tr>
<td>$\texttt{Eq} = \lambda \alpha. \lambda \beta. \forall \phi :: \ast \Rightarrow \ast. \phi \alpha \rightarrow \phi \beta$</td>
<td>$\texttt{Equal} = \lambda \alpha. \lambda \beta. \forall \phi :: \ast \Rightarrow \ast \Rightarrow \ast. (\forall \gamma. \phi \gamma \gamma) \rightarrow \phi \alpha \beta$</td>
</tr>
</tbody>
</table>

\[
\Lambda \alpha. \forall \beta. \texttt{Equal \: \alpha \beta} \rightarrow \texttt{Eq \: \alpha \beta}
\]

\[
\Lambda \alpha. \forall \beta. \texttt{Eq \: \alpha \beta} \rightarrow \texttt{Equal \: \alpha \beta}
\]

\[
\texttt{refl} : \forall \alpha. \texttt{Equal} \: \alpha \alpha
\]

\[
\texttt{refl} = \lambda \alpha. \: \forall \phi :: \ast \Rightarrow \ast \Rightarrow \ast. \lambda \gamma: (\forall \gamma. \phi \gamma \gamma). \texttt{f}[\alpha]
\]

\[
\texttt{symm} : \forall \alpha. \forall \beta. \texttt{Equal} \: \alpha \beta \rightarrow \texttt{Equal} \: \beta \alpha
\]

\[
\texttt{symm} = \lambda \alpha. \lambda \beta. \lambda \phi :: \ast \Rightarrow \ast \Rightarrow \ast. \lambda h: (\forall \gamma. \phi \gamma \gamma). \texttt{e}[\lambda \alpha. \lambda \beta. \phi \beta \alpha] \: \texttt{f}
\]

\[
\texttt{trans} : \forall \alpha. \forall \beta. \forall \gamma. \texttt{Equal} \: \alpha \beta \rightarrow \texttt{Equal} \: \beta \gamma \rightarrow \texttt{Equal} \: \alpha \gamma
\]

\[
\texttt{trans} = \lambda \alpha. \lambda \beta. \lambda \gamma. \lambda \phi: \ast \Rightarrow \ast \Rightarrow \ast. \lambda \delta: (\forall \gamma. \phi \gamma \gamma). \texttt{e}[\lambda \alpha. \lambda \beta. \phi \beta \delta] \: (\lambda \gamma: (\lambda \phi: \phi \alpha \gamma \rightarrow \phi \alpha \delta) \: (\forall \gamma. \lambda x: \phi \gamma x)) \: (\texttt{e}[\phi] \: \texttt{f})
\]

Type equalities in OCaml

\[
\texttt{type} \ ('a, 'b)\ \texttt{eql} = \texttt{Refl} : ('x, 'x)\ \texttt{eql}
\]

\[
\texttt{let}\ \texttt{symm} : \texttt{type} \ a\ b. (a, b)\ \texttt{eql} \rightarrow (b, a)\ \texttt{eql} = \lambda\ \texttt{Refl} \rightarrow \texttt{Refl}
\]

\[
\texttt{let}\ \texttt{trans} : \texttt{type} \ a\ b. (a, b)\ \texttt{eql} \rightarrow (b, c)\ \texttt{eql} \rightarrow (a, c)\ \texttt{eql} = \lambda\ \texttt{Refl} \ \texttt{Refl} \rightarrow \texttt{Refl}
\]

Fig. 4. Type equalities

As well as improving familiarity with type equality, the exercise is also a useful lesson in the power of polymorphic code, since any correctly-typed implementation that converts between $\texttt{Equal}$ and $\texttt{Eq}$ is necessarily correct.

For example, consider the conversion from $\texttt{Eq}$ to $\texttt{Equal}$ which, after expanding the aliases (and renaming variables to avoid confusion) has the following type:

\[
\forall \alpha. \forall \beta. (\forall \psi :: \ast \Rightarrow \ast. \psi \alpha \rightarrow \psi \beta) \rightarrow (\forall \phi :: \ast \Rightarrow \ast \Rightarrow \ast. (\forall \gamma. \phi \gamma \gamma) \rightarrow \phi \alpha \beta)
\]

The term may be built by following the types outside-in, introducing a $\lambda$ for each outer $\forall$ and a $\lambda$ for each outer $\rightarrow$, giving the partially-constructed term:

\[
\lambda \alpha. \lambda \beta. \lambda \phi :: \ast \Rightarrow \ast \Rightarrow \ast. \lambda h: (\forall \gamma. \phi \gamma \gamma). \texttt{??}
\]

Then the remainder of the type requires that the body have type $\phi \alpha \beta$, and so we must construct such a term from $\alpha$, $\beta$, $e$, $\phi$ and $h$. Only two of these, $e$ and $h$, are term variables, and $h$ is unsuitable (since it can only build terms where both arguments of $\phi$ are the same). In order to use $e$ it is sufficient to find a suitable $\psi$ such that $\psi \beta \equiv \phi \alpha \beta$. Fortunately, this equation has an obvious solution ($\psi \equiv \phi \alpha$), which can be used to instantiate $e$ and start building the body:

\[
e[\phi \alpha] : \phi \alpha \alpha \rightarrow \phi \alpha \beta
\]
type 'a perfect =
    ZeroP : 'a -> 'a perfect
  | SuccP : ('a * 'a) perfect -> 'a perfect

Perfect = \alpha. \forall \phi :: \alpha \to \alpha -> \forall \alpha. \phi \alpha \to \phi \alpha

zeroP = \alpha. \lambda x: \alpha. \lambda \phi :: \alpha \to \alpha -> \forall \alpha. \phi \alpha \to \phi \alpha

  s: (\forall \beta. \phi (\beta \times \beta) \to \phi \beta). s[\alpha] (\alpha \times \alpha)

Fig. 5. Perfect trees in System F\omega and OCaml

Finally, completing the puzzle requires a term of type \phi \alpha \alpha to pass to this function. We can construct a suitable term by instantiating h:

e[\phi \alpha] (h[\alpha])

which suffices to complete the exercise. Besides illustrating the utility of polymorphic types in reasoning about programs, exercises of this sort help to distinguish between straightforwardly mechanical programming (e.g. type-directed insertion of introduction forms) and steps that involve some insight (e.g. finding suitable substitutions).

Besides applications to GADTs, encodings of type equalities can also be used as a basis (or, at least, as an inspiration) for encodings of other relationships between types, such as subtyping [Yallop and Dolan 2017]. Equality proofs also plays a central role in many programs based on the Curry-Howard correspondence (Section 3.5).

3.3 Non-Regular Types in System F\omega

Non-regular types [Bird and Meertens 1998] (often called nested types) — i.e. data types that are applied to something other than their parameters within their own definition — appear frequently in advanced functional programming applications [Bird and Paterson 1999; Hinze 1999, 2000; Okasaki 1998]. Figure 5 gives a typical example: the perfect type constructor appears as ('a,'a) perfect (rather than 'a perfect) in the argument to SuccP; every tree of that type therefore has the form illustrated in the figure, with a chain of n SuccP constructors and exactly 2^n values at the leaves.

Encodings of nested types arise naturally in System F\omega using the same scheme used to define the regular tree types (Figure 3). Figure 5 gives a definition: the System F\omega type Perfect closely follows the definition of the OCaml perfect, with the type variable \phi in place of perfect and \alpha in place of 'a.

3.4 GADTs in System F\omega

Generalized algebraic data types (GADTs), whose usefulness once appeared to be restricted to length-indexed vectors and well-typed evaluators, are now commonplace in functional programming, appearing in applications from parsers [Pottier and Régis-Gianas 2006] to foreign function
interfaces [Yallop et al. 2018]. However, novices still find GADTs challenging to understand. Encodings of GADTs in System F_\omega are certainly less convenient than using the facilities provided by high-level functional languages directly but, like other verbose encodings, can serve as a significant aid to understanding. (We cannot recall a single instance from the past four years where a student has had trouble with GADTs after programming in System F_\omega.)

For example, given the standard equality GADT
\[
\text{type } (\text{'a, 'b) eql } = \text{Refl } : (\text{'a, 'a) eql }
\]
and a pattern
\[
(\text{Refl } : (\text{int list, string list) eql) } \to ...
\]
the OCaml compiler reasons as follows [Garrigue and Normand 2015]. First, the definition of Refl requires that the two arguments to eql are equivalent, so there should be a type equation int list \equiv string list, scoped under the pattern. Second, the type constructor list is injective, so there should be a second equation int \equiv string. Finally, since int and string are in fact known to be distinct types, the equation is unsatisfiable and the pattern is redundant.

In OCaml programs this reasoning happens automatically, but with a System F_\omega encoding, each step must be performed explicitly by manipulating values in the program. While verbosity is inconvenient, explicitness can be useful, both for programmers seeking to understand the types of the programs they write, and for language designers. For example, a slightly careless approach to reasoning about these kinds of interactions between injectivity and type equality led to a soundness bug in the OCaml compiler [Garrigue 2013]. Our aim in breaking the convenient GADT matching down into unwieldy explicit components is to help students understand, identify and avoid such problems.

Later parts of the course often make heavy use of GADTs: our exercises have involved building balanced trees (with insertion and deletion) using GADT constraints to ensure balancing, defining a DSL that represents only valid XHTML values [Elsman and Larsen 2004], implementing typed printf and scanf format specifiers [Vaugon 2013], and ensuring invariants for the classic pure functional two-list implementation of queues (e.g. that the length of outq never exceeds the length of inq).

With System F_\omega encodings of GADTs we have been a little less ambitious, given their verbosity and awkwardness. In 2017 we asked the students to define functions based on an encoding of vectors (length-indexed sequences) that approximately follows Atkey [2012]:

\[
\text{Vec } :: \ast \Rightarrow \ast \Rightarrow \ast
\]
\[
\text{Vec } = \lambda \alpha. \lambda M. \forall \phi : \ast \Rightarrow \ast \Rightarrow \ast. \phi Z \to (\forall N. \alpha \to \phi N \to \phi (S N)) \to \phi M
\]

There are two arguments to the function that represents a vector: the first, of type \( \phi Z \), represents an empty vector (indexed by zero); the second, of type \( \forall N. \alpha \to \phi N \to \phi (S N) \), represents a cons cell containing a head (type \( \alpha \)) a tail (type \( \phi N \)) and returning type \( \phi (S N) \) — i.e., a vector indexed by the successor of the length of the tail.

The awkwardness of programming with encodings of GADTs provides a reminder of the various conveniences provided by an in-language implementation. For example, several auxiliary functions and values representing facts about numbers are needed, such as a proof that zero is not the successor of any number:

\[
\text{sz_distinct } : \forall N. \text{Not } (\text{Eql } (S N) Z)
\]
\[
= \lambda N. \lambda eq : \text{Eq } (S N) Z. \text{ eq}[\lambda X. X](\text{inr}[N] \text{ unit})
\]

Although it is clunky, almost everything needed to program with vectors can be encoded in this way, except for a proof that the successor function is injective, which must be added as an axiom.
Fig. 6. The lambda and logic cubes

\[ \begin{align*}
\text{DM}_1 &= \forall \alpha. \forall \beta. \text{Not} (\alpha \times \beta) \rightarrow (\text{Not} \alpha + \text{Not} \beta) \\
\text{DM}_2 &= \forall \alpha. \forall \beta. (\text{Not} \alpha + \text{Not} \beta) \rightarrow \text{Not} (\alpha \times \beta) \\
\text{DM}_3 &= \forall \alpha. \forall \beta. \text{Not} (\alpha + \beta) \rightarrow (\text{Not} \alpha \times \text{Not} \beta) \\
\text{DM}_4 &= \forall \alpha. \forall \beta. (\text{Not} \alpha \times \text{Not} \beta) \rightarrow \text{Not} (\alpha + \beta)
\end{align*} \]

Fig. 7. De Morgan’s laws in System \( F_\omega \)

\[ \begin{align*}
\text{DM}_5 &= \forall \phi :: \ast \Rightarrow \ast. \text{Not} (\forall \alpha. \phi \alpha) \rightarrow (\exists \alpha. \text{Not} (\phi \alpha)) \\
\text{DM}_6 &= \forall \phi :: \ast \Rightarrow \ast. (\exists \alpha. \text{Not} (\phi \alpha)) \rightarrow \text{Not} (\forall \alpha. \phi \alpha) \\
\text{DM}_7 &= \forall \phi :: \ast \Rightarrow \ast. \text{Not} (\exists \alpha. \phi \alpha) \rightarrow (\forall \alpha. \text{Not} (\phi \alpha)) \\
\text{DM}_8 &= \forall \phi :: \ast \Rightarrow \ast. (\forall \alpha. \text{Not} (\phi \alpha)) \rightarrow \text{Not} (\exists \alpha. \phi \alpha)
\end{align*} \]

Fig. 8. Generalization of De Morgan’s laws in System \( F_\omega \)

(Similarly, Brown and Palsberg [2017] extended System \( F_\omega \) with a typecase construct to support injectivity in their self-representation of System \( F_\omega \).)

A particularly tricky function to define is the function that takes the tail of a non-empty vector: \( \text{tail} : \forall \alpha. \forall \mu. \text{Vec} \alpha (S \mu) \rightarrow \text{Vec} \alpha \mu \)

Readers familiar with the predecessor function for Church-encodings of natural numbers may recognise the difficulty. Here we have the additional challenge of programming with a much more richly typed structure. (We discuss this function further in Section 4.3)

3.5 Proofs in System \( F_\omega \)

The famous Curry-Howard correspondence [Wadler 2015] connects propositions with types, programs with proofs, and normalization with proof simplification. It is an interesting curiosity, and occasionally useful for language designers (e.g. [Davies 1996; Mitchell and Plotkin 1988]), but is it a worthwhile addition to a course which is intended to teach programming?

Sheard [2005] answers in the affirmative: the Curry-Howard correspondence can be put to work to build a variety of interesting programs, from composable pipelines to modular arithmetic. Like System \( F_\omega \), the correspondence provides a mental model — a way of thinking about programs and programming languages that can be fruitful for reasoning.

Figure 6 gives one key to understanding the connection between functional programming and logic: for each calculus in the famous lambda cube on the left there is a corresponding logic on the
more well known cube on the right. As the figure shows, System \( F_\omega \) and the other languages on the
left face of the lambda cube all correspond to propositional logics, which can express propositions
about sets but not about individual objects. (The predicate logics, which can express propositions
about objects, correspond to dependently-typed languages, such as \( \lambda C \), the calculus of cons-
tructions.) By considering the differences in the expressiveness between logics, students can develop
intuition for the differences in expressiveness between traditional functional programming lan-
guages and languages with dependent types. Grasping these differences also helps with under-
standing the “singletons” technique (i.e. turning individual values into sets) which is often used
to make dependently-typed techniques (somewhat) accessible in languages without dependent
types [Altenkirch et al. 2005; Eisenberg and Weirich 2012; Lindley and Mcbride 2013]. Exercises
later in the course sometimes build on these encodings (perhaps taking them rather too far): for
example, we ask students to define properties of data types such as a guarantee that the elements
of a tree are kept in order (somewhat in the style of Kiselyov and Shan [2007]).

As with the other aspects of the course, programming in System \( F_\omega \) can help with developing
intuitions about the Curry-Howard isomorphism. For example, it is sometimes useful to know
whether a given type is inhabited; considering the type as a proposition can be a useful guide.
(Viewed this way, the type \( \forall \alpha . \alpha \) is obviously uninhabited, since the corresponding proposition is
not provable; similarly \( \forall \alpha . \alpha \to \alpha \) is obviously inhabited.)

Figures 7 and 8 illustrate a programming exercise from 2017: attempting to witness De Morgan’s
laws in System \( F_\omega \), first using binary connectives, and then using an arbitrary unary predicate. A
catch is that only three of the laws in each case are provable constructively; students are left to
discover the “classical” law, and give proofs (as System \( F_\omega \) terms) for the others.

3.6 Semirings

Yet another useful mental model for programming is the semiring structure of types [Carette and
Sabry 2016]. This is closely related to the logical structure given by the Curry-Howard isomor-
phism, but lacks idempotency. (For example, while \( A \land A \) is equivalent to \( A \) in propositional logic,
\( A \times A \) is not typically equal to \( A \) in other semirings.) As with the Curry-Howard correspondence,
the semiring view can guide the implementation of programs. To give a trivial example, in any
semiring it is the case that \( A \times 1 \equiv A \), and we can indeed easily give a pair of functions that
convert between the types \( A \) and \( A \times 1 \).

\[
f : \forall A . A \times 1 \to A \quad g : \forall A . A \to A \times 1
\]

\[
f = \lambda \cdot \lambda p: A \times 1 . f s t p \quad g = \lambda A \cdot \lambda x: A . \langle x, \langle \rangle \rangle
\]

It is not possible to prove within System \( F_\omega \) that functions so defined do actually form an isomor-
phism — such a proof would need predicate logic. However, external proofs of properties like this
are straightforward using techniques developed later in the course, such as equational reasoning
or parametricity.

In 2018 we asked the students to define terms corresponding to several more complex semiring
equations (such as distributivity), and later to build a variant of generalized functional tries [Elliott

3.7 Existentials in System \( F_\omega \)

Existential types are common in typed functional programs, in the form of abstract types [Mitchell
and Plotkin 1988], and as a way of hiding possibly-heterogeneous data behind a common interface.

Although our System \( F_\omega \) implementation provides existentials as a primitive, it is straightfor-
ward to define them in terms of universals using the well-known encoding

\[
\exists \alpha . \tau ::= \forall \beta . (\forall \alpha . \tau \to \beta) \to \beta
\]
In fact, System $F_\omega$ can encode higher-kind existential types, such as $\exists \alpha :: \ast \Rightarrow \ast \Rightarrow \ast$ in this way. (However, each kind requires a new encoding, since System $F_\omega$ does not support kind polymorphism.)

In 2018 we asked the students to investigate encodings of existential types, along with operations corresponding to the standard pack and open introduction and elimination forms.

\[
\begin{align*}
\text{pack}_b &:: \quad \forall \phi :: \ast \Rightarrow \ast \Rightarrow \ast \Rightarrow \ast . \exists \alpha . \lambda r : \alpha . \lambda s : \alpha . \phi \beta \\
\text{open}_b &:: \quad \forall \phi :: \ast \Rightarrow \ast \Rightarrow \ast \Rightarrow \ast . \exists \alpha . \lambda r : \alpha . \lambda s : \alpha . \phi \\
&\quad \quad \text{and to use the encodings to define various abstract types. Figure 9 gives an example: the code on the left uses existentials to hide the implementation of Bools as a sum of units, and the implementation on the right unpacks an implementation of Bools and performs computation.}
\end{align*}
\]

### 3.8 Modules in System $F_\omega$

Modules play a crucial role in writing programs in ML-family languages, but beginners often make very limited use of the module system, shying away from apparently advanced features such as functors. System $F_\omega$ can be used to give a semantics to large parts of ML module systems, and writing System $F_\omega$ programs is a good way to develop intuition for modules.

In 2018 we asked students to encode a module with four primitive components (three plumbing operations, and a nor gate) for constructing circuits:

\[
\begin{align*}
\text{nor} &:: \quad \forall \beta :: \ast \Rightarrow \ast . T (\text{Bool} \times \text{Bool}) \text{Bool} \\
\text{split} &:: \quad \forall \beta :: \ast \Rightarrow \ast . T (\text{Bool} \times \text{Bool}) \\
\text{join} &:: \quad \forall \beta :: \ast \Rightarrow \ast . T (\text{Bool} \times \text{Bool}) \\
\text{plug} &:: \quad \forall \beta :: \ast \Rightarrow \ast . T (\text{Bool} \times \text{Bool})
\end{align*}
\]

The following System $F_\omega$ definition corresponds to such a module signature; the type parameter $\mathcal{T}$ corresponds to type component that can be instantiated by an implementation of the module. The remainder of the exercise involves using the module to build new gates (And, Not), implementing the signature with implementations that count and interpret gates, and encoding higher-kind existentials (Section 3.7) to make the module type member abstract.

\[
\begin{align*}
\text{Gates} &:: \quad \lambda T :: \ast \Rightarrow \ast \Rightarrow \ast \Rightarrow \ast . (T (\text{Bool} \times \text{Bool}) \text{Bool}) \times \\
&\quad \quad \times (\forall \alpha . T \alpha (\alpha \times \alpha)) \times \\
&\quad \quad \times (\forall \alpha . \forall \beta . T \alpha \beta \rightarrow T (\alpha \times \beta) (\gamma \times \delta)) \times \\
&\quad \quad \times (\forall \alpha . \forall \beta . \forall \gamma . T \alpha \beta \rightarrow T \beta \gamma \rightarrow T \alpha \gamma)
\end{align*}
\]

## 4 LESSONS AND CHALLENGES

I definitely learned a lot, but it was tough getting there (2018)

Planning and teaching the course involves substantial ongoing work; for example, we develop a fresh set of exercises each year, and review the course material, adding or removing lectures, based
on student feedback and other measures of effectiveness. One substantial change in 2018 was the addition of a few sections on dependently-typed programming. In retrospect, this turned out to be a mistake: shifting the focus away from System $F_\omega$-based programming led to a less coherent presentation; in future we plan to stay firmly on the left-hand side of the cube (Figure 10).

Overall, feedback from students has been positive, often extremely so:

- *It was probably my favourite course in the entire tripos*. The material was very cutting-edge, interesting and intellectually stimulating. (2017)

Other students reported finding the course practical:

- *I could see how the material could be usefully applied in practice* (2017)

or inspiring:

- *I have started to really consider research in programming languages and category theory in part due to this course.* (2018)

although some students disagreed:

- *I think there are very few careers (research or applied) that this course would have any bearing on.* (2018)

### 4.1 Tooling

Section 2.3 described our simple batch-mode System $F_\omega$ interpreter. In practice this is adequate, but it is likely that the students’ learning experience could be significantly enhanced with some usability improvements.

The main shortcoming is the quality of error messages, which report type or kind mismatches, but do not describe the exact nature of the incompatibility. Producing good error messages for languages with type inference (such as ML or Haskell) can be challenging but, since every variable in a System $F_\omega$ program is annotated, we have no such excuse! Good error messages are valuable; one student submitted some improvements to the interpreter to improve location reporting.

It would also be helpful to have a selection of modern conveniences found in more serious language implementations: a REPL, typed holes, syntax highlighting, and auto-completion, particularly where the context uniquely determines the possible inhabitants of a type (Section 3.2).

---

$^3$Tripos is a Cambridge term for the degree course.
4.2 Preparation

Actually, the course was a lot more in-depth than I would’ve expected from reading the syllabus—a very pleasant surprise! (2017)

The different backgrounds of the students provide an additional challenge, especially during the early weeks of the course. In practice, students are divided into two (roughly equal) camps: those who have continued from a bachelor’s degree at our university, and those who join the masters course from elsewhere. The local students have usually taken a course on type systems that covers System F$^\omega$, and find little new in the early material:

For those having taken any Type Theory courses in the past, much of the first half of the course was completely familiar. (2018)

In contrast, some students joining from elsewhere have little to no experience with type systems and functional programming. Although we ask that students have some typed functional programming experience before joining the course, some students do not heed the advice, and find the plunge into System F$^\omega$ rather a shock.

Despite this variety in experience, most students complete the course successfully; indeed, in typical years, all students do so. We attribute this high success rate to several factors: a selective admissions process, a supportive environment with opportunities for questions and discussion (during lectures, in office hours, and via a class mailing list), comprehensive lecture notes, small class sizes and a course structure that means that students are often taking only other unit concurrently, allowing them to focus their efforts. Furthermore, assessment via compulsory exercises rather than a final exam makes it easier to discover difficulties in understanding at an early stage, and makes it possible for students who struggle with the material to overcome difficulties with prolonged effort.

However, although these factors make it possible to overcome the difficulties arising from the students’ different backgrounds, we think that it would be even better to address the issue head-on. One way to do this would be to split the course into two parts: an initial optional part that teaches type theory up to System F$^\omega$, and a functional programming section that starts with programming in System F$^\omega$ before moving on to standard functional programming languages such as OCaml or Haskell. At present our course covers both parts, but the initial part is rather fast-paced for students without much type theory experience, and mostly unnecessary for students who have taken a type theory course already. Unfortunately, administrative constraints have so far made it difficult for us to divide the course in this way.

4.3 Assessment

Because the compiler provides so much feedback, it seems most of the marking is check marks. (2018)

Finding a suitable level of difficulty for the assessed exercises has often proved challenging. In languages with powerful type systems such as System F$^\omega$ it is frequently the case that programs are either correct or ill-typed. Students therefore often know before submitting whether their solutions are correct and high scores are common.

Assignments were very well constructed, matched the course material extremely well, and were tricky yet satisfying (2018)

From a pedagogical viewpoint this is ideal! An incentive structure that encourages students to persist until they succeed is likely to result in mastery of the course material. (In practice, this is exactly what happens, and the spread in marks is replaced by a spread in time taken: some students

---

When we started teaching our course, the type systems course only covered System F
spend ten times as long to complete exercises as others, and there is little correlation between the
time taken and the mark achieved.) Furthermore, teaching students to structure programs so that
type checking ensures the absence of errors is a central aim of the course.

Assignments were generally really good. Just long enough to make you understand the
content without dragging it out. (2015)

However, from a practical viewpoint there is sometimes a need to differentiate students — for
example, PhD programs often ask for student rankings. In order to ensure a spread of marks we
consequently include some functions such as `tail` (Section 3.4) that are likely to prove challenging
even for those who have mastered the linguistic aspects of System F\(\omega\).

5 RELATED WORK

We believe that the precise approach described here — using System F\(\omega\) as a sublanguage for
typed functional programming — is unusual. However, since System F\(\omega\) lies at the core of several
typed functional programming languages it is no surprise to find other advanced functional pro-
gramming courses with a substantial System F or System F\(\omega\) component. A complete survey is
impractical, but we mention some representative examples.

François Pottier and colleagues at Inria teach *Functional programming and type systems* (MPRI
2-4), which focuses on the semantics, compilation and metatheory of functional languages based
on System F, but has recently started to incorporate some broader programming ideas such as
effectful programming, tagless interpreters and generic programming.

Kevin Hamlen at the University of Texas teaches *Advanced Programming Languages*, starting
with functional languages and progressing to System F.

Tim Sheard’s *Advanced Functional Programming* course at Portland covers several of the topics
touched on here (e.g. higher-rank polymorphism, quasiquotation, and monads), but using Haskell
as the main teaching language.

Several books use some form of lambda calculus as a means to teach functional programming.
Nordström et al. [1990] teach programming directly in type theory — that is, programming
on the right face of the cube (Figure 10). Similarly, Thompson [1991] focuses on type theory for
programming, drawing connections (analogies and comparisons) with Miranda.

At the opposite end of the type spectrum, Michaelson [2011] introduces functional programming
starting from untyped lambda calculus, before moving to simple types and on to Standard ML.

Felleisen et al. [2004] critically examine the introductory computer science curriculum, focusing
on the shortcomings of the classic *Structure and Interpretation of Computer Programs*, and proposing
an alternative based (in part) upon the use of sub-languages; one inspiration for our course is
their advocacy of that approach.

ACKNOWLEDGMENTS

We are grateful to all those whose help and ideas contributed to the course, including Alan My-
croft, Stephen Dolan, Nada Amin, Neel Krishnaswami, KC Sivaramakrishnan, Dominic Mulligan,
Anil Madhavapeddy, Heidi Howard and, of course, the Advanced Computer Science students who
participated in the course over the past few years and provided valuable feedback. We thank Sally
Fincher and Alan Blackwell for ethical guidance. Finally, we thank the ICFP’19 reviewers for help-
ful comments.

REFERENCES

cs.nott.ac.uk/~psztxa/publ/ydtm.pdf.


Jeremy Yallop and Oleg Kiselyov. 2010. First-class modules: hidden power and tantalizing promises. ACM SIGPLAN Workshop on ML. (September 2010). Baltimore, Maryland, United States.
