

Custom Representations of Inductive Families

Constantine Theocharis^[0009–0001–0198–2750] and
Edwin Brady^[0000–0002–9734–367X]

University of St Andrews, UK
{kt81,ecb10}@st-andrews.ac.uk

Abstract. Inductive families provide a convenient way of programming with dependent types. Yet, when it comes to compilation, their default linked-tree runtime representations, as well as the need to convert between different indexed views of the same data when programming with dependent types, can lead to unsatisfactory runtime performance. In this paper, we aim to introduce a language with dependent types, and inductive families with custom representations. Representations are a version of Wadler’s views [12], refined to inductive families like in Epi-gram [10]. However, representations come with compilation guarantees: a represented inductive family will not leave any runtime traces behind, without having to rely on automated optimisations such as deforestation [13]. This way, we can build a library of convenient inductive families based on a minimal set of primitives, whose re-indexing and conversion functions are erased at compile-time. In addition, we show how we can express inductive data optimisation techniques, such as representing `Nat`-like types as GMP-style big integers, without special casing in the compiler. With dependent types, reasoning about data representations is also possible; for example, we get computationally irrelevant isomorphisms between the original and represented data.

Keywords: Dependent types · Memory representation · Inductive families

1 Introduction

Inductive families are a generalisation of inductive data types found in some programming languages with dependent types. Every inductive definition is equipped with an eliminator that captures the notion of mathematical induction over the data, and in particular, enables structural recursion over the data. This is a powerful tool for programming as well as theorem proving. However, this abstraction has a cost when it comes to compilation: the runtime representation of inductive types is a linked tree structure. This representation is not always the most efficient for all operations, and often forces users to rely on more efficient machine primitives to achieve desirable performance, at the cost of structural recursion and dependent pattern matching. This is the first problem we aim to address in this paper.

Despite advances in the erasure of irrelevant indices in inductive families [5] and the use of theories with irrelevant fragments [3,11], there is still a need

to convert between different indexed views of the same data. For example, the function to convert from `List T` to `Vec T n` by forgetting the length index n is *not* erased by any current language with dependent types, unless vectors are defined as a refinement of lists with an erased length field (which hinders dependent pattern matching due to the presence of non-structural witnesses), or a Church encoding is used in a Curry-style context [9] (which restricts the flexibility of data representation). This is the second problem we aim to address in this paper.

Wadler’s views [12] provide a way to abstract over inductive interfaces, so that different views of the same data can be defined and converted between seamlessly. In the context of inductive families, views have been used in Epigram [10] that utilise the index refinement machinery of dependent pattern matching to avoid certain proof obligations with eliminator-like constructs. While views exhibit a nice way to transport across a bijection between the original data and the viewed data, they do not utilise this bijection to erase the view from the program. Despite deforestation handling this erasure to some extent, it is not guaranteed to erase all traces of the view from the program, and the optimisation might be difficult to predict.

In this paper, we propose an extension λ_{REP} to a core language with dependent types and inductive families λ_{IND} , which allows programmers to define custom, correct-by-construction data representations. This is done through user-defined translations of the constructors and eliminators of an inductive type to a concrete implementation, which form a bijective view of the original data called a ‘representation’. Representations are defined internally to the language, and require coherence properties that ensure a representation is faithful to its the original inductive family. In the final version of the paper, we plan to contribute the following:

- A dependent type system with inductive families λ_{IND} , and its extension by representations λ_{REP} .
- A formulation of common optimisations such as the ‘Nat-hack’, and similarly for other inductive types, as representations.
- A demonstration of zero-cost data reuse when reindexing by using representations.
- A translation from λ_{REP} to λ_{IND} that erases all inductive types with representations from the program.
- An implementation of this system and accompanying examples in SUPERFLUID, a programming language with inductive types and dependent pattern matching.

2 A tour of data representations

A common optimisation done by programming languages with dependent types such as Idris 2 and Lean is to represent natural numbers as GMP-style [1] big integers. The definition of natural numbers looks like

$$\text{data Nat} \left\{ \begin{array}{l} 0 : \text{Nat} \\ 1+ : \text{Nat} \rightarrow \text{Nat} \end{array} \right\} \quad (1)$$

and generates a Peano-style induction principle elim_{Nat} of type¹

$$(P : \text{Nat} \rightarrow \mathcal{U}) \rightarrow P \ 0 \rightarrow ((n : \text{Nat}) \rightarrow \overline{P \ n} \rightarrow P \ (1 + n)) \rightarrow (s : \text{Nat}) \rightarrow P \ s.$$

Without further intervention, the Nat type is represented in unary form, where each digit becomes an empty heap cell at runtime. This is inefficient for a lot of the basic operations on natural numbers, especially since computers are particularly well-equipped to deal with numbers natively, so many real-world implementations will treat Nat specially, swapping the default inductive type representation with one based on GMP integers. This is done by performing the replacements

$$|0| = 0 \tag{2}$$

$$|1 +| = 1 + \tag{3}$$

$$|\text{elim}_{\text{Nat}} P \ m_0 \ m_{1+} \ s| = \text{ubig-elim} \ |s| \ |m_0| \ |m_{1+}| \tag{4}$$

where $|\cdot|$ denotes a source translation into a compilation target language with primitives ubig- .²

In addition to the constructors and eliminators, the compiler might also define translations for commonly used definitions which have a more efficient counterpart in the target, such as recursively-defined addition, multiplication, etc. The recursively-defined functions are well-suited to proofs and reasoning, while the GMP primitives are more efficient for computation.

The issue with this approach is that it only works for the data types which the compiler can recognise as special. Particularly in the presence of dependent types, other data types might end up being equivalent to Nat or another ‘nicely-representable’ type, but in a non-trivial way that the compiler cannot recognise. Hence, one of our goals is to extend this optimisation to work for any data type. To achieve this this, our framework requires that representations are fully typed in a way that ensures the behaviour of the representation of a data type matches the behaviour of the data type itself.

2.1 The well-typed Nat-hack

A representation definition looks like

$$\text{repr Nat as UBig} \left\{ \begin{array}{l} 0 \text{ as } 0 \\ 1 + n \text{ as } 1 + n \\ \text{elim}_{\text{Nat}} \text{ as } \text{ubig-elim} \\ \text{by } \text{ubig-elim-zero-id}, \\ \quad \text{ubig-elim-add-one-id} \end{array} \right\}$$

¹ Recursive parameters like $\overline{P \ n}$ are lazy, which makes the eliminator more efficient when they are not used.

² Idris 2 will in fact look for any ‘Nat-like’ types and apply this optimisation. A Nat-like type is any type with two constructors, one with arity zero and the other with arity one. A similar optimisation is also done with list-like and boolean-like types because they have a canonical representation in the target runtime, Chez Scheme.

Nat is represented as the type **UBig** of GMP-style unlimited-size unsigned integers, with translations for the constructors **0** and **1+**, and the eliminator **elim_{Nat}**. Additionally, the eliminator satisfies the expected computation rules of the **Nat** eliminator, which are postulated as propositional equalities. This representation is valid in a signature containing the primitives

$$\begin{aligned} &0, 1 : \mathbf{UBig} \quad +, \times : \mathbf{UBig} \rightarrow \mathbf{UBig} \rightarrow \mathbf{UBig} \\ &\mathbf{ubig-elim} : (P : \mathbf{UBig} \rightarrow \mathcal{U}) \rightarrow P \ 0 \rightarrow ((n : \mathbf{UBig}) \rightarrow \overline{P \ n} \rightarrow P \ (1 + n)) \\ &\quad \rightarrow (s : \mathbf{UBig}) \rightarrow P \ s \end{aligned}$$

and propositional equalities

$$\begin{aligned} &\mathbf{ubig-elim-zero-id} :_{\forall P b r} \mathbf{ubig-elim} \ P \ b \ r \ 0 = b \\ &\mathbf{ubig-elim-add-one-id} :_{\forall P b r n} \mathbf{ubig-elim} \ P \ b \ r \ (1 + n) = r \ n \ (\lambda _ . \mathbf{ubig-elim} \ P \ b \ r \ n). \end{aligned}$$

Representations can also be defined for functions on **Nat**, such as addition, multiplication, and other numeric operations, in terms of **UBig** primitives.

$$\mathbf{repr} \ \mathbf{add} \ \mathbf{as} \ + \ \mathbf{by} \ +\text{-id} \quad \mathbf{repr} \ \mathbf{mul} \ \mathbf{as} \ \times \ \mathbf{by} \ \times\text{-id}$$

These will be replaced during a translation process back to λ_{IND} , like rewriting rules [7], given that we have the appropriate lemmas to justify them in the signature.

This will effectively erase the **Nat** type from the compiled program, replacing all occurrences with the **UBig** type and its primitives. In a way, the hard work is done by the postulates above; we expect that the underlying implementation of **UBig** indeed satisfies them, which is a separate concern from the correctness of the representation itself. However, postulates are only needed when the representation target is a primitive; the next examples use defined types as targets, so that the coherence of the target eliminator follows from the coherence of other eliminators used in its implementation.

2.2 Vectors are just certain lists

In addition to representing inductive types as primitives, we can use representations to share the underlying data when converting between indexed views of the same data. For example, we can define a representation of **Vec** in terms of **List**, so that the conversion from one to the other is ‘compiled away’. We can do this by representing the indexed type as a refinement of the unindexed type by an appropriate relation. For the case of **Vec**, we know intuitively that

$$\mathbf{Vec} \ T \ n \simeq \{l : \mathbf{List} \ T \mid \mathbf{length} \ l = n\} := \mathbf{List}' \ T \ n$$

so we can start by choosing $\mathbf{List}' \ T \ n$ as the representation of $\mathbf{Vec} \ T \ n$.³ We are then tasked with providing terms that correspond to the constructors of **Vec** but

³ We will take the subset $\{x : A \mid P \ x\}$ to mean a Σ -type $(x : A) \times P \ x$ where the right component is irrelevant and erased at runtime.

for List' . These can be defined as

$$\begin{aligned} \text{nil} &: \text{List}' T 0 & \text{cons} &: T \rightarrow \text{List}' T n \rightarrow \text{List}' T (1+ n) \\ \text{nil} &= (\text{nil}, \text{refl}) & \text{cons } x \ (xs, p) &= (\text{cons } x \ xs, \text{cong } (1+) \ p) \end{aligned}$$

Next we need to define the eliminator for List' , which should have the form

$$\begin{aligned} \text{elim-List}' &: (E : (n : \text{Nat}) \rightarrow \text{List}' T n \rightarrow \text{Type}) \\ &\rightarrow E \ 0 \ \text{nil} \\ &\rightarrow ((x : T) \rightarrow (n : \text{Nat}) \rightarrow (xs : \text{List}' T n) \rightarrow \overline{E \ n \ xs} \rightarrow E \ (1+ \ n) \ (\text{cons } x \ xs)) \\ &\rightarrow (n : \text{Nat}) \rightarrow (v : \text{List}' T n) \rightarrow E \ n \ v \end{aligned}$$

Dependent pattern matching does a lot of the heavy lifting by refining the length index and equality proof by matching on the underlying list. However we still need to substitute the lemma $\text{cong } (1+) \ (1+-\text{inj } p) = p$ in the recursive case.

$$\begin{aligned} \text{elim-List}' \ P \ b \ r \ 0 \ (\text{nil}, \text{refl}) &= b \\ \text{elim-List}' \ P \ b \ r \ (1+ \ m) \ (\text{cons } x \ xs, e) &= \text{subst } (\lambda p. \ P \ (1+ \ m) \ (\text{cons } x \ xs, p)) \\ &\quad (1+-\text{cong-id } e) \ (r \ x \ (xs, 1+-\text{inj } e)) \\ &\quad (\lambda _ . \ \text{elim-List}' \ P \ b \ r \ m \ (xs, 1+-\text{inj } e))) \end{aligned}$$

Finally, we need to prove that the eliminator satisfies the expected computation rules propositionally. These are

$$\begin{aligned} \text{elim-List}'\text{-nil-id} &: \text{elim-List}' \ P \ b \ r \ 0 \ (\text{nil}, \text{refl}) = b \\ \text{elim-List}'\text{-cons-id} &: \text{elim-List}' \ P \ b \ r \ (1+ \ m) \ (\text{cons } x \ xs, \text{cong } (1+) \ p) \\ &= r \ x \ (xs, p) \ (\lambda _ . \ \text{elim-List}' \ P \ b \ r \ m \ (xs, p)) \end{aligned}$$

which we leave as an exercise, though they are evident from the definition of $\text{elim-List}'$. This completes the definition of the representation of Vec as List' , which would be written as

$$\text{repr } \text{Vec } T \ n \ \text{as } \text{List}' T \ n \left\{ \begin{array}{l} \text{nil as nil} \\ \text{cons as cons} \\ \text{elim}_{\text{Vec}} \text{ as elim-List}' \\ \text{by elim-List}'\text{-nil-id,} \\ \text{elim-List}'\text{-cons-id} \end{array} \right\}$$

Now the hard work is done; Every time we are working with a $v : \text{Vec } T \ n$, its form will be (l, p) at runtime, where l is the underlying list and p is the proof that the length of l is n . Under the assumption that the Σ -type's right component is irrelevant and erased at runtime, every vector is simply a list at runtime, where the length proof has been erased. In the full paper we will show how this erasure is achieved in practice in SUPERFLUID using Quantitative Type Theory [3].

We can utilise this representation to convert between `Vec` and `List` at zero runtime cost, by using the `repr` and `unrepr` operators of the language (defined in section 3). Specifically, we can define the functions

```
forget-length : Vec T n → List T
forget-length v = let (l, _) = repr v in l

recall-length : (l : List T) → Vec T (length l)
recall-length l = unrepr (l, refl)
```

and it holds by reflexivity that `forget-length` is a left inverse of `recall-length`.

2.3 General reindexing

The idea from the previous example can be generalised to any data type. In general, suppose that we have two inductive families

$$F : \Xi \rightarrow \mathcal{U} \quad G : \Xi \rightarrow X \rightarrow \mathcal{U}$$

for some index family $X : \Xi \rightarrow \mathcal{U}$. If we hope to represent G as some refinement of F then we must be able to provide a way to compute G 's extra indices X from F , like we computed `Vec`'s extra `Nat` index from `List` with `length` in the previous example. This means that we need to provide a function

$$\text{comp} : \forall \xi. F \xi \rightarrow X \xi$$

which can then be used to form the family

$$F^{\text{comp}} \xi x := \{f : F \xi \mid \text{comp } f = x\}.$$

If G is ‘equivalent’ to the algebraic ornament of F by the algebra defining `comp` (given by an isomorphism between the underlying polynomial functors), then it is also equivalent to the Σ -type above. The ‘recomputation lemma’ of algebraic ornaments [8] then arises from its projections. Our system allows us to *set* the representation of G as F^{comp} , so that the forgetful map from G to F is the identity at runtime.

2.4 Zero-copy deserialisation

The machinery of representations can be used to implement zero-copy deserialisation of data formats into inductive types. For example, consider the following record for a player in a game:

$$\text{data Player} \left\{ \begin{array}{l} \text{player} : (\text{position} : \text{Position}) \\ \quad \rightarrow (\text{direction} : \text{Direction}) \\ \quad \rightarrow (\text{items} : \text{Fin MAX_INVENTORY}) \\ \quad \rightarrow (\text{inventory} : \text{Inventory items}) \rightarrow \text{Player} \end{array} \right\}$$

We can use the `Fin` type to maintain the invariant that the inventory has a maximum size. Additionally, we can index the `Inventory` type by the number of items it contains, which might be defined similarly to `Vec`:

$$\text{data } \text{Inventory} \ (n : \text{Nat}) \left\{ \begin{array}{l} \text{empty} : \text{Inventory } 0 \\ \text{add} : \text{Item} \rightarrow \text{Inventory } n \rightarrow \text{Inventory } (1 + n) \end{array} \right\}$$

We can use the full power of inductive families to model the domain of our problem in the way that is most convenient for us. If we were writing this in a lower-level language, we might choose to use the serialised format directly when manipulating the data, relying on the appropriate pointer arithmetic to access the fields of the serialised data, to avoid copying overhead. Representations allow us to do this while still being able to work with the high-level inductive type.

We can define a representation for `Player` as a pair of a byte buffer and a proof that the byte buffer contents correspond to a player record. Similarly, we can define a representation for `Inventory` as a pair of a byte buffer and a proof that the byte buffer contents correspond to an inventory record of a certain size. The projection `inventory : (p : Player) → Inventory p.items` is compiled into some code to slice into the inventory part of the player's byte buffer. We assume that the standard library already represents `Fin` in the same way as `Nat`, so that reading the `items` field is a constant-time operation (we do not need to build a unary numeral). We can thus define the representation of `Player` as

$$\text{repr } \text{Player} \text{ as } \{ \text{Buf} \mid \text{IsPlayer} \} \left\{ \begin{array}{l} \text{player as buf-is-player} \\ \text{elim}_{\text{Player}} \text{ as elim-buf-is-player} \\ \text{by elim-buf-is-player-id} \end{array} \right\}$$

with an appropriate definition of `IsPlayer` which refines a byte buffer. We will provide the full details of this construction in the final paper.

2.5 Transitivity

Representations are transitive, so in the previous example, the ‘terminal’ representation of `Vec` also depends on the representation of `List`. It is possible to define a custom representation for `List` itself, for example a heap-backed array or a finger tree, and `Vec` would inherit this representation. However it will still be the case that `Repr (Vec T n) ≡ List T`, which means the `repr/Repr` operators only look at the immediate representation of a term, not its terminal representation. Regardless, we can construct predicates that find types which satisfy a certain ‘eventual’ representation. For example, given a `Buf` type of byte buffers, we can consider the set of all types which are eventually represented as a `Buf`:

$$\text{data } \text{ReprBuf} \ (T : \mathcal{U}) \left\{ \begin{array}{l} \text{buf} : \text{ReprBuf } \text{Buf} \\ \text{from} : \text{ReprBuf} \ (\text{Repr } T) \rightarrow \text{ReprBuf } T \\ \text{refined} : \text{ReprBuf } T \rightarrow \text{ReprBuf } \{ t : T \mid P \ t \} \end{array} \right\}$$

Every such type comes with a projection function to the `Buf` type

```

as-buf :  $\forall T. \text{ReprBuf } T \rightarrow \text{Buf}$ 
as-buf buf  $x = x$ 
as-buf (from  $t$ )  $x = \text{as-buf } t (\text{repr } x)$ 
as-buf (refined  $t$ )  $(x, \_ ) = \text{as-buf } t x$ 

```

which eventually computes to the identity function after applying `repr` the appropriate amount of times. Upon compilation, every type is converted to its terminal representation, and all `repr` calls are erased, so the `as-buf` function is effectively the identity function at runtime.⁴

3 A type system for data representations

This section provides an overview of the language λ_{REP} , which has dependent types and representations for inductive families and global function definitions. We start with a core language with inductive families λ_{IND} , that is extended with data representations to form λ_{REP} . All of the examples in the paper are written in a surface language that elaborates to λ_{REP} .

3.1 The core language, λ_{IND}

The language λ_{IND} , is a dependent type theory with Π and a universe \mathcal{U} , extended with strictly positive inductive families and global definitions. The base layer is a standard dependent type theory, defined in an intrinsically well-formed fashion quotiented by conversion. The constructors of the syntax are given in ???. We omit the conversion equations for brevity as they are standard [2].

We will not concern ourselves with universe levels because a sound logical interpretation is not of much interest in this paper. Nevertheless, all of our results should be readily adaptable to a calculus with universe levels.

We follow a similar approach to Cockx and Abel [6] by packaging named inductive constructions and global function definitions into a signature Σ , and indexing contexts by signatures.

To do this, we define signatures, local contexts, and contexts (as well as terms and types) by mutual induction.

3.2 Extending λ_{IND} with representations

We extend the language λ_{IND} to form λ_{REP} , which allows users to define custom representations for inductive types and global functions. First, we add a type former

$$\mathbf{Repr} : \text{Ty } (\Sigma \mid \Gamma) \rightarrow \text{Ty } (\Sigma \mid \Gamma) \quad (5)$$

⁴ We do not guarantee that an invocation of `as-buf` will be entirely erased, but rather that any invocation will eventually produce the identity function without having to perform a case analysis on its T subject.

along with two new terms in the syntax, forming an isomorphism

$$\mathbf{repr} : \mathsf{Tm} (\Sigma \mid \Gamma) T \simeq \mathsf{Tm} (\Sigma \mid \Gamma) (\mathbf{Repr} T) : \mathbf{unrepr} . \quad (6)$$

which holds definitionally and preserves Π and universes. The type $\mathbf{Repr} T$ is the defined representation of the type T . The term \mathbf{repr} takes a term of type T to its representation of type $\mathbf{Repr} T$, and \mathbf{unrepr} undoes the effect of \mathbf{repr} , treating a represented term as an inhabitant of its original type. These new constructs satisfy certain equalities, which are spelled out in . New valid signature items are introduced, corresponding to representation definitions for components of inductive families and global function definitions, shown in fig. 4.

$$\begin{array}{c}
\begin{array}{c} \text{REPR-FORM} \\ \hline \Sigma \mid \Gamma \vdash T \text{ type}_i \\ \hline \Sigma \mid \Gamma \vdash \mathbf{Repr} T \text{ type}_i \end{array} \quad \begin{array}{c} \text{REPR-INTRO} \\ \hline \Sigma \mid \Gamma \vdash t : T \\ \hline \Sigma \mid \Gamma \vdash \mathbf{repr} t : \mathbf{Repr} T \end{array} \quad \begin{array}{c} \text{REPR-ELIM} \\ \hline \Sigma \mid \Gamma \vdash u : \mathbf{Repr} T \\ \hline \Sigma \mid \Gamma \vdash \mathbf{unrepr} u : T \end{array} \\[10pt]
\begin{array}{c} \text{REPR-UNREPR-ID} \\ \hline \Sigma \mid \Gamma \vdash u : \mathbf{Repr} T \\ \hline \Sigma \mid \Gamma \vdash \mathbf{repr} (\mathbf{unrepr} u) \equiv u : \mathbf{Repr} T \end{array} \quad \begin{array}{c} \text{UNREPR-REPR-ID} \\ \hline \Sigma \mid \Gamma \vdash t : T \\ \hline \Sigma \mid \Gamma \vdash \mathbf{unrepr} (\mathbf{repr} t) \equiv t : T \end{array} \\[10pt]
\begin{array}{c} \text{REPR-TYPE-SUBST-ID} \\ \hline \Sigma \mid \Gamma \vdash T \text{ type}_i \quad \Sigma \mid \Delta \vdash \sigma : \Gamma \\ \hline \Sigma \mid \Delta \vdash \mathbf{Repr} (T[\sigma]) \equiv (\mathbf{Repr} T)[\sigma] \end{array} \\[10pt]
\begin{array}{c} \text{REPR-TERM-SUBST-ID} \\ \hline \Sigma \mid \Gamma \vdash t : T \quad \Sigma \mid \Delta \vdash \sigma : \Gamma \\ \hline \Sigma \mid \Delta \vdash \mathbf{repr} (t[\sigma]) \equiv (\mathbf{repr} t)[\sigma] : \mathbf{Repr} (T[\sigma]) \end{array} \\[10pt]
\begin{array}{c} \text{UNREPR-TERM-SUBST-ID} \\ \hline \Sigma \mid \Gamma \vdash t : \mathbf{Repr} T \quad \Sigma \mid \Delta \vdash \sigma : \Gamma \\ \hline \Sigma \mid \Delta \vdash \mathbf{unrepr} (t[\sigma]) \equiv (\mathbf{unrepr} t)[\sigma] : T[\sigma] \end{array}
\end{array}$$

Fig. 1. New term and type formers, and accompanying definitional equalities in λ_{REP} .

We state some basic lemmas below. The proof are left to the full version of the paper, along with the formalisation of computational irrelevance.

Lemma 1. *The term formers \mathbf{repr} and \mathbf{unrepr} are injective, i.e.*

$$\begin{array}{c} \Sigma \mid \Gamma \vdash \mathbf{repr} t \equiv \mathbf{repr} t' : \mathbf{Repr} T \\ \hline \Sigma \mid \Gamma \vdash t \equiv t' : T \end{array} \quad \begin{array}{c} \Sigma \mid \Gamma \vdash \mathbf{unrepr} t \equiv \mathbf{unrepr} t' : T \\ \hline \Sigma \mid \Gamma \vdash t \equiv t' : \mathbf{Repr} T \end{array}$$

Lemma 2. *The type former \mathbf{Repr} is injective up to internal isomorphism, i.e.*

$$\begin{array}{c} \Sigma \mid \Gamma \vdash \mathbf{Repr} T \equiv \mathbf{Repr} T' \text{ type} \\ \hline \Sigma \mid \Gamma \vdash p : T \simeq T' \end{array}$$

Moreover, this isomorphism is computationally irrelevant.

$$\begin{array}{c}
\text{REPR-}\Pi\text{-ID} \\
\frac{\Sigma \mid \Gamma \vdash T \text{ type}_i \quad \Sigma \mid \Gamma, T \vdash U \text{ type}_i}{\Sigma \mid \Gamma \vdash \mathbf{Repr} (\Pi T U) \equiv \Pi T (\mathbf{Repr} U)} \\
\\
\text{REPR-}\lambda\text{-ID} \\
\frac{\Sigma \mid \Gamma, T \vdash u : U}{\Sigma \mid \Gamma \vdash \mathbf{repr} (\lambda u) \equiv \lambda (\mathbf{repr} u) : \mathbf{Repr} (\Pi T U)} \\
\\
\text{UNEPR-}\lambda\text{-ID} \\
\frac{\Sigma \mid \Gamma, T \vdash u : \mathbf{Repr} U}{\Sigma \mid \Gamma \vdash \mathbf{unrepr} (\lambda u) \equiv \lambda (\mathbf{unrepr} u) : \Pi T U} \\
\\
\text{REPR-APP-ID} \\
\frac{\Sigma \mid \Gamma \vdash f : \Pi T U \quad \Sigma \mid \Gamma \vdash t : T}{\Sigma \mid \Gamma \vdash \mathbf{repr} (f t) \equiv (\mathbf{repr} f) t : \mathbf{Repr} U} \\
\\
\text{UNREPR-APP-ID} \\
\frac{\Sigma \mid \Gamma \vdash f : \mathbf{Repr} (\Pi T U) \quad \Sigma \mid \Gamma \vdash t : T}{\Sigma \mid \Gamma \vdash \mathbf{unrepr} (f t) \equiv (\mathbf{unrepr} f) t : U}
\end{array}$$

Fig. 2. Coherence of representation terms with Π -types.

4 Translating from λ_{REP} to λ_{IND}

We can define a translation step \mathcal{R} from λ_{REP} to $\lambda_{\text{IND}}^{\text{EXT}}$, meant to be applied during the compilation process. More specifically, the translation target is the extensional flavour of λ_{IND} by adding the equality reflection rule. We do this by translating well-formed contexts, substitutions, types, and terms in a mutual manner such that definitional equality is preserved. \mathcal{R} preserves the structure of λ_{REP} , but maps constructs to their terminal representations. Eliminator coherence rules are preserved by reflecting the propositional coherence rules provided by the defined representations. We will prove some desired properties of \mathcal{R} [4] such as typing and computational soundness, and preservation of consistency. The final program can then be converted into a simply-typed language which erases irrelevant data. We can recover a program in λ_{IND} by translating extensional typing derivations to intensional proofs [14].

5 Implementation

SUPERFLUID is a programming language with dependent types, $\mathcal{U} : \mathcal{U}$, quantities, inductive families and dependent pattern matching. Its compiler is written in Haskell and the compilation target is JavaScript. Dependent pattern matching in SUPERFLUID is elaborated to a core language with internal eliminators. The \mathcal{R} transformation is then applied to the core program, which erases all inductive constructs with defined representations. This is finally translated to

$$\begin{array}{c}
\text{REPR-}\mathcal{U}_i\text{-ID} \\
\hline
\Sigma \mid \Gamma \vdash \mathbf{Repr} \mathcal{U}_i \equiv \mathcal{U}_i
\end{array}
\qquad
\begin{array}{c}
\text{REPR-CODE-ID} \\
\Sigma \mid \Gamma \vdash T \text{ type}_i \\
\hline
\Sigma \mid \Gamma \vdash \mathbf{repr} (\text{code } T) \equiv \text{code } T : \mathcal{U}_i
\end{array}$$

$$\begin{array}{c}
\text{UNREPR-CODE-ID} \\
\Sigma \mid \Gamma \vdash T \text{ type}_i \\
\hline
\Sigma \mid \Gamma \vdash \mathbf{unrepr} (\text{code } T) \equiv \text{code } T : \mathcal{U}_i
\end{array}$$

Fig. 3. Coherence of representation terms with universes.

$$\begin{array}{c}
\text{REPR-CTOR-ID} \\
\mathbf{repr} \mathbf{c} \Pi \text{ as } \kappa \in \Sigma \\
\hline
\Sigma \mid \Gamma \vdash \mathbf{repr} (\mathbf{c} \delta \pi) \equiv \kappa[\delta, \pi] : A[\delta, \xi[\pi]]
\end{array}
\qquad
\begin{array}{c}
\text{REPR-DATA-ID} \\
\mathbf{repr} \mathbf{D} \Delta \Xi \text{ as } A \in \Sigma \\
\hline
\Sigma \mid \Gamma \vdash \mathbf{Repr} (\mathbf{D} \delta \psi) \equiv A[\delta, \psi] \text{ type}
\end{array}$$

Fig. 4. Definitional equalities for **Repr** and **repr** relating to data types and constructors with defined representations. Similar equalities hold for representations of global function definitions and eliminators, albeit propositionally.

JavaScript, erasing all irrelevant data. As a result, we are able to represent **Nat** as JavaScript's **BigInt**, and **List** T /**SnocList** T /**Vec** T n as JavaScript's arrays with the appropriate index refinement, such that we can convert between them without any runtime overhead.

References

1. The GNU MP bignum library. <https://gmplib.org/>, accessed: 2024-12-8
2. Altenkirch, T., Kaposi, A.: Type theory in type theory using quotient inductive types. In: Proceedings of the 43rd Annual ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages. pp. 18–29. POPL '16, Association for Computing Machinery, New York, NY, USA (Jan 2016), <https://doi.org/10.1145/2837614.2837638>
3. Atkey, R.: Syntax and semantics of quantitative type theory. In: LICS 2018. pp. 56–65. LICS '18, Association for Computing Machinery, New York, NY, USA (Jul 2018)
4. Boulter, S., Pédro, P.M., Tabareau, N.: The next 700 syntactical models of type theory. In: CPP 2017. pp. 182–194. CPP 2017, Association for Computing Machinery, New York, NY, USA (Jan 2017)
5. Brady, E., McBride, C., McKinna, J.: Inductive families need not store their indices. In: Types for Proofs and Programs. pp. 115–129. Springer Berlin Heidelberg (2004)
6. Cockx, J., Abel, A.: Elaborating dependent (co)pattern matching. Proc. ACM Program. Lang. **2**(ICFP), 1–30 (Jul 2018)
7. Cockx, J., Tabareau, N., Winterhalter, T.: The taming of the rew: a type theory with computational assumptions. Proc. ACM Program. Lang. **5**(POPL), 1–29 (Jan 2021)

8. Dagand, P.E., McBride, C.: A categorical treatment of ornaments. arXiv [cs.PL] (Dec 2012)
9. Diehl, L., Firsov, D., Stump, A.: Generic zero-cost reuse for dependent types. Proc. ACM Program. Lang. **2**(ICFP), 1–30 (Jul 2018)
10. McBride, C., Mckinna, J.: The view from the left. JFP **14**(1), 69–111 (Jan 2004)
11. Moon, B., Eades, III, H., Orchard, D.: Graded modal dependent type theory. In: Programming Languages and Systems. pp. 462–490. Springer (2021)
12. Wadler, P.: Views: a way for pattern matching to cohabit with data abstraction. In: Proceedings of the 14th ACM SIGACT-SIGPLAN symposium on Principles of programming languages. pp. 307–313. POPL '87, Association for Computing Machinery, New York, NY, USA (Oct 1987)
13. Wadler, P.: Deforestation: transforming programs to eliminate trees. Theor. Comput. Sci. **73**(2), 231–248 (Jun 1990)
14. Winterhalter, T., Sozeau, M., Tabareau, N.: Eliminating reflection from type theory. In: CPP 2019. ACM, New York, NY, USA (Jan 2019)

6 Appendix

The metatheory is presented internally to an extensional dependent type theory with an internal universe **Set**, dependent function types $(a : A) \rightarrow B$ and dependent pair types $(a : A) \times B$. The metatheory also supports quotient inductive-inductive definitions, which are used to define the syntaxes of the languages presented in this paper.

6.1 The language λ_{IND}

$$\begin{array}{ll}
 \text{Ty} : \text{Con} \rightarrow \mathbf{Set} & \text{Tm} : (\Gamma : \text{Con}) \rightarrow \text{Ty } \Gamma \rightarrow \mathbf{Set} \\
 [] : \text{Ty } \Delta \rightarrow \text{Sub } \Gamma \Delta \rightarrow \text{Ty } \Gamma & _[_] : \text{Tm } \Delta \ A \rightarrow (\sigma : \text{Sub } \Gamma \Delta) \rightarrow \text{Tm } \Gamma \ A[\sigma] \\
 \mathcal{U} : \text{Ty } \Gamma & \pi_2 : (\sigma : \text{Sub } \Gamma (\Delta \triangleright A)) \rightarrow \text{Tm } \Gamma \ A[\pi_1 \sigma] \\
 \Pi : (A : \text{Ty } \Gamma) \rightarrow \text{Ty } (\Gamma \triangleright A) \rightarrow \text{Ty } \Gamma & \lambda : \text{Tm } (\Gamma \triangleright A) \ B \rightarrow \text{Tm } \Gamma \ (\Pi \ A \ B) \\
 \text{El} : \text{Tm } \Gamma \ \mathcal{U} \rightarrow \text{Ty } \Gamma & @ : \text{Tm } \Gamma \ (\Pi \ A \ B) \rightarrow \text{Tm } (\Gamma \triangleright A) \ B \\
 & \text{code} : \text{Ty } \Gamma \rightarrow \text{Tm } \Gamma \ \mathcal{U}
 \end{array}$$

Fig. 5. The type and term formers of the base layer of λ_{IND} .

TODO

Fig. 6. Equations in the base type system.

$\text{Sig} : \mathbf{Set}$ $\epsilon : \text{Sig}$ $\triangleright : (\Sigma : \text{Sig}) \rightarrow (A : \text{Item } \Sigma) \rightarrow \text{Sig}$	$\text{Loc} : \text{Sig} \rightarrow \mathbf{Set}$ $\epsilon : \text{Loc } \Sigma$ $\triangleright : (\Delta : \text{Loc } \Sigma) \rightarrow \text{Ty } (\Sigma, \Delta) \rightarrow \text{Loc } \Sigma$
$\text{Item} : \text{Sig} \rightarrow \mathbf{Set}$ $\text{ind} : \text{Ind } \Sigma \rightarrow \text{Item } \Sigma$ $\text{def} : (A : \text{Ty } (\Sigma, \epsilon)) \rightarrow \text{Tm } (\Sigma, \epsilon) \rightarrow \text{Item } \Sigma$ $\text{post} : \text{Ty } (\Sigma, \epsilon) \rightarrow \text{Item } \Sigma$	$\text{Con} : \mathbf{Set}$ $\text{Con} := (\Sigma : \text{Sig}) \times \text{Loc } \Sigma$

Fig. 7. Signatures, items, local contexts and contexts in λ_{IND} .

$$\begin{aligned}
 &\text{Sub} : \text{Con} \rightarrow \text{Con} \rightarrow \mathbf{Set} \\
 &\text{id} : \text{Sub } \Gamma \Gamma \\
 &\epsilon : \text{Sub } (\Sigma, \Delta) (\Sigma, \epsilon) \\
 &\triangleright : (\sigma : \text{Sub } (\Sigma, \Delta_1) (\Sigma, \Delta_2)) \rightarrow \text{Tm } (\Sigma, \Delta_1) \rightarrow A[\sigma] \rightarrow \text{Sub } (\Sigma, \Delta_1) (\Sigma, \Delta_2 \triangleright A) \\
 &\circ : \text{Sub } \Gamma_1 \Gamma_2 \rightarrow \text{Sub } \Gamma_2 \Gamma_3 \rightarrow \text{Sub } \Gamma_1 \Gamma_3 \\
 &\pi_1 : \text{Sub } (\Sigma, \Delta_1) (\Sigma, \Delta_2 \triangleright A) \rightarrow \text{Sub } (\Sigma, \Delta_1) (\Sigma, \Delta_2)
 \end{aligned}$$

Fig. 8. Substitutions between contexts in λ_{IND} .

Ind $\text{Ind } \Sigma$	$: (\Sigma : \text{Sig}) \rightarrow \mathbf{Set}$ $:= (P : \text{Loc } \Sigma) \times (\Xi : \text{Tel } (\Sigma, P)) \times \text{Theory } (\Sigma, P) \rightarrow \Xi$
Theory $\text{Theory } \Gamma \Xi$	$: (\Gamma : \text{Con}) \rightarrow \text{Tel } \Gamma \rightarrow \mathbf{Set}$ $:= \text{List } (\text{Op } \Gamma \Xi)$
Op Π_{ext} Π_{int} ret	$: (\Gamma : \text{Con}) \rightarrow (\Xi : \text{Tel } \Gamma) \rightarrow \mathbf{Set}$ $: (A : \text{Ty } \Gamma) \rightarrow \text{Op } (\Gamma \triangleright A) \rightarrow \Xi^+ \rightarrow \text{Op } \Gamma \Xi$ $: \text{Tms } \Gamma \Xi \rightarrow \text{Op } \Gamma \Xi \rightarrow \text{Op } \Gamma \Xi$ $: \text{Tms } \Gamma \Xi \rightarrow \text{Op } \Gamma \Xi$

Fig. 9. Inductive types, theories and operations in λ_{IND} .

$$\begin{aligned}
\text{In} & : \text{Op } \Gamma \Xi \rightarrow \text{Ty } (\Gamma \supseteq \Xi) \rightarrow \mathbf{Set} \\
\text{In } (\Pi_{\text{ext}} A B) X & := (a : \mathbf{Tm } \Gamma A) \times \text{In } B[\langle a \rangle] X \\
\text{In } (\Pi_{\text{int}} \xi B) X & := \mathbf{Tm } \Gamma (X[\langle \xi \rangle]) \times \text{In } B X \\
\text{In } (\text{ret } \xi) X & := \mathbb{1} \\
\\
\text{out} & : (O : \text{Op } \Gamma \Xi) \rightarrow \text{In } O X \rightarrow \mathbf{Tms } \Gamma \Xi \\
\text{out } (\Pi_{\text{ext}} A B) (a, r) & := \text{out } B[\langle a \rangle] r \\
\text{out } (\Pi_{\text{int}} \xi B) (a, r) & := \text{out } B r \\
\text{out } (\text{ret } \xi) () & := \xi
\end{aligned}$$

Fig. 10. Interpretation of operation inputs (i.e. constructor arguments) and outputs (constructor return indices).

$$\begin{aligned}
\text{data} & : (\text{ind } (P, \Xi, T) \in \Gamma) \rightarrow (p : \mathbf{Tms } \Gamma P) \\
& \rightarrow \text{Ty } (\Gamma \supseteq \Xi[\langle p \rangle]) \\
\text{ctor} & : (i : \text{ind } (P, \Xi, T) \in \Gamma) \rightarrow (p : \mathbf{Tms } \Gamma P) \\
& \rightarrow (n : \text{Index } T) \rightarrow (a : \text{In } (T!n)) \\
& \rightarrow (r : \text{Rec } (T!n) (\text{data } i p)) \\
& \rightarrow \mathbf{Tm } \Gamma (\text{data } i p)[\langle \text{out } (T!n) a \rangle] \\
\text{elim} & : (i : \text{ind } (P, \Xi, T) \in \Gamma) \rightarrow (p : \mathbf{Tms } \Gamma P) \\
& \rightarrow (M : \text{Ty } (\Gamma \supseteq \Xi[\langle p \rangle])) \rightarrow (m : \mathbf{Methods } \Gamma T M) \\
& \rightarrow \mathbf{Tm } (\Gamma \supseteq \Xi[\langle p \rangle]) \triangleright \text{data } i p M \\
\text{match} & : (\text{elim } i p M m)[\langle \xi, \text{ctor } i p n a r \rangle] = (m!n)[\langle a, \text{rec } i p M m n \rangle]
\end{aligned}$$

Fig. 11. Data type, constructor and eliminator terms in λ_{IND} .

6.2 The language λ_{REP}

$$\begin{aligned} \text{Repr} &: \text{Ty } \Gamma \rightarrow \text{Ty } \Gamma \\ \text{repr} &: \text{Tm } \Gamma \ T \rightarrow \text{Tm } \Gamma \ (\text{Repr } T) \\ \text{unrepr} &: \text{Tm } \Gamma \ (\text{Repr } T) \rightarrow \text{Tm } \Gamma \ T \end{aligned}$$

Fig. 12. New term and type formers in λ_{REP} .

$$\begin{array}{ll} \text{reprr} &: \text{unrepr } (\text{repr } t) \equiv t \\ \text{reprl} &: \text{repr } (\text{unrepr } t) \equiv t \\ \text{Repr-}\Pi &: \text{Repr } (\Pi \ T \ U) \equiv \Pi \ T \ (\text{Repr } U) \\ \text{repr-}\lambda &: \text{repr } (\lambda \ u) \equiv \lambda \ (\text{repr } u) \\ \text{unrepr-}\lambda &: \text{unrepr } (\lambda \ u) \equiv \lambda \ (\text{unrepr } u) \\ \text{repr-}@ &: \text{repr } (f \ @) \equiv (\text{repr } f) \ @ \\ \text{unrepr-}@ &: \text{unrepr } (f \ @) \equiv (\text{unrepr } f) \ @ \\ \text{Repr-}\mathcal{U} &: \text{Repr } \mathcal{U} \equiv \mathcal{U} \\ \text{repr-code} &: \text{repr } (\text{code } T) \equiv \text{code } T \\ \text{unrepr-code} &: \text{unrepr } (\text{code } T) \equiv \text{code } T \\ \text{Repr}[] &: \text{Repr } (T[\sigma]) \equiv (\text{Repr } T)[\sigma] \\ \text{repr}[] &: \text{repr } (t[\sigma]) \equiv (\text{repr } t)[\sigma] \\ \text{unrepr}[] &: \text{unrepr } (t[\sigma]) \equiv (\text{unrepr } t)[\sigma] \end{array}$$

Fig. 13. Coherence of representation terms with substitutions, Π -types and universes. The terms $\text{Repr } (\text{El } t)$, $\text{repr } (\pi_2 \sigma)$ and $\text{unrepr } (\pi_2 \sigma)$ do not reduce, and there is no rule to collapse repeated invocations of Repr , repr and unrepr .

$$\begin{aligned} \text{Rep} &: (\Sigma : \text{Sig}) \rightarrow \text{Item } \Sigma \rightarrow \text{Set} \\ \text{ind-rep} &: \text{IndRep } F \rightarrow \text{Rep } \Sigma \ (\text{ind } F) \\ \text{def-rep} &: (x : \text{Tm } (\Sigma, \epsilon) \ A) \rightarrow \text{Tm } (\Sigma, \epsilon) \ (x = t) \rightarrow \text{Rep } \Sigma \ (\text{def } A \ t) \\ \text{post-rep} &: \text{Tm } (\Sigma, \epsilon) \ A \rightarrow \text{Rep } \Sigma \ (\text{post } A) \end{aligned}$$

Fig. 14. Implementations of items and representations of signatures in λ_{REP} .

6.3 The translation

$$\begin{aligned} \text{Impl} &: \text{Sig}_{\text{REP}} \rightarrow \mathbf{Set} \\ \text{Impl } \Sigma &:= (I \in \Sigma) \rightarrow \text{Maybe } (\text{Rep } I) \end{aligned}$$

$$\begin{aligned} \mathcal{R} &: (\Sigma : \text{Sig}_{\text{REP}}) \rightarrow \text{Impl } \Sigma \rightarrow \text{Sig}_{\text{IND}} \\ \mathcal{R} (\epsilon) f &:= \epsilon \\ \mathcal{R} (\Sigma \triangleright I) f &:= \text{if } f (\Sigma \triangleright I) \text{ here} = \text{just } r \text{ then } \Sigma \text{ else } (\Sigma \triangleright I) \end{aligned}$$

$$\begin{aligned} \mathcal{R}_{f:\text{Impl } \Sigma} &: \text{Ty}_{\text{REP}} (\Sigma, \Delta) \rightarrow \text{Ty}_{\text{IND}} (\mathcal{R}_f (\Sigma, \Delta)) \\ \mathcal{R}_f (T[\sigma]) &:= (\mathcal{R}_f T)[\mathcal{R}_f \sigma] \\ \mathcal{R}_f (\mathcal{U}) &:= \mathcal{U} \\ \mathcal{R}_f (\Pi A B) &:= \Pi (\mathcal{R}_f A) (\mathcal{R}_f B) \\ \mathcal{R}_f (\text{El } t) &:= \text{El } (\mathcal{R}_f t) \\ \mathcal{R}_f (\text{data } i p) &:= \text{if } f \Sigma i = \text{just } (\text{ind-rep } r) \text{ then } \ulcorner r \urcorner \text{ else } (\text{data } i p) \\ \mathcal{R}_f (\text{Repr } T) &:= \mathcal{R}_f T \\ \mathcal{R}_{f:\text{Impl } \Sigma} &: \text{Tm}_{\text{REP}} (\Sigma, \Delta) T \rightarrow \text{Tm}_{\text{IND}} (\mathcal{R}_f (\Sigma, \Delta)) (\mathcal{R}_f T) \\ \mathcal{R}_f (t[\sigma]) &:= (\mathcal{R}_f t)[\mathcal{R}_f \sigma] \\ \mathcal{R}_f (\pi_2 \sigma) &:= \pi_2 (\mathcal{R}_f \sigma) \\ \mathcal{R}_f (\lambda t) &:= \lambda (\mathcal{R}_f t) \\ \mathcal{R}_f (t @) &:= (\mathcal{R}_f t) @ \\ \mathcal{R}_f (\text{code } T) &:= \text{code } (\mathcal{R}_f T) \\ \mathcal{R}_f (\text{ctor } i p n a) &:= \text{if } f \Sigma i = \text{just } (\text{ind-rep } r) \text{ then } \ulcorner r \urcorner \text{ else } (\text{ctor } i p n a) \\ \mathcal{R}_f (\text{elim } i p M m) &:= \text{if } f \Sigma i = \text{just } (\text{ind-rep } r) \text{ then } \ulcorner r \urcorner \text{ else } (\text{elim } i p M m) \\ \mathcal{R}_f (\text{repr } t) &:= \mathcal{R}_f t \\ \mathcal{R}_f (\text{unrepr } t) &:= \mathcal{R}_f t \end{aligned}$$