

Partially static data as free extension of algebras

JEREMY YALLOP, University of Cambridge, England

TAMARA VON GLEHN, University of Cambridge, England

OHAD KAMMAR, University of Oxford, England

Partially-static data structures are a well-known technique for improving binding times. However, they are often defined in an ad-hoc manner, without a unifying framework to ensure full use of the equations associated with each operation.

We present a foundational view of partially-static data structures as free extensions of algebras for suitable equational theories, i.e. the coproduct of an algebra and a free algebra in the category of algebras and their homomorphisms. By precalculating these free extensions, we construct a high-level library of partially static data representations for common algebraic structures. We demonstrate our library with common use-cases from the literature: string and list manipulation, linear algebra, and numerical simplification.

Additional Key Words and Phrases: multi-stage compilation, metaprogramming, partial evaluation, partially static data, universal algebra

1 INTRODUCTION

The defining feature of multi-stage programming is putting fine-grained control over beta reduction in the hands of the programmer. For example, the multi-stage programming language Typed Template Haskell [Peyton Jones 2016] extends Haskell with two constructs. The first construct, quotation (written $\llbracket e \rrbracket$), prevents beta reduction, turning an arbitrary expression into a value:

$$\llbracket f \ x \rrbracket \quad \text{-- } f \ x \text{ should not be reduced}$$

The second construct, antiquote (written $\$(e)$), re-enables beta reduction inside a quotation:

$$\llbracket f \ \$(g \ x) \rrbracket \quad \text{-- } g \ x \text{ should be reduced}$$

A notable property of this style of quotation is support for reduction *under lambda*, with support for quoting *open terms*:

$$\llbracket \lambda x \rightarrow \$(f \ \llbracket x \rrbracket) \rrbracket \quad \text{-- } f \ \llbracket x \rrbracket \text{ should be reduced}$$

This fine-grained control over evaluation provides a basis for optimisation. In the execution of a multi-stage program, each stage executes code constructed from quoted expressions in the previous stage. Careful insertion of quotations and antiquotations allows terms depending only on known values to be reduced during generation, leaving only terms depending on unknown values to be reduced in the following stage.

The standard introductory example of multi-stage programming [Taha 2003] is the power function, implemented in terms of multiplication:

```
power :: Int -> Int -> Int
power x 0 = 1
power x n = x * power x (n - 1)
```

Authors' addresses: Jeremy Yallop, jeremy.yallop@cl.cam.ac.uk, Department of Computer Science and Technology, University of Cambridge, England; Tamara von Glehn, T.L.Von-Glehn@dpmms.cam.ac.uk, Department of Pure Mathematics and Mathematical Statistics, and Newnham College, University of Cambridge, England; Ohad Kammar, ohad.kammar@cs.ox.ac.uk, University of Oxford, Department of Computer Science, and Balliol College, Wolfson Building, Parks Road, Oxford, OX1 3QD, England.

2018. 2475-1421/2018/9-ART104 \$15.00

<https://doi.org/>

Under the assumption that the exponent n will be available to the generating stage, the programmer adds staging annotations and changes the types accordingly. (Here `Code`, an alias for Typed Template Haskell's `Q (TExp a)`, is the type of quoted expressions):

```
power :: Code Int → Int → Code Int
power x 0 = [1]
power x n = [x * $(power [x] (n - 1))]
```

Now `power` is no longer a function on integers, but a code generator: the application of `power` to a variable x and an integer n evaluates to an expression specialized to the exponent, in which the overhead of the recursive call and the branch have been eliminated:

$$\llbracket \lambda x \rightarrow \$(power [x] 6) \rrbracket \rightsquigarrow \llbracket x * (x * (x * (x * (x * (x * 1)))) \rrbracket$$

However, while the generated code is likely to outperform the unstaged `power`, there is evidently room for further improvement. First, there is a needless multiplication by 1. More significantly, the number of multiplications can be further reduced by `let`-binding intermediate results:

```
[ let y = x * x in let z = y * y in z * y ]
```

Evidently, beta reduction alone will not perform these simplifications, no matter how we coax it by inserting quotation and antiquotation annotations. No series of beta reductions will perform these simplifications, which are justified by the algebraic properties of multiplication – in this case, that integers with multiplication form a monoid.

Similar difficulties occur in a wide variety of staged programs, as we illustrate in §4. In programs that are staged by quoting expressions the structure of the generated code is determined by the evaluation structure of the generator, and algebraic laws play no part in code generation.

How might you deal with this difficulty? First, you might change restructure the source program (e.g. defining `power` via an auxiliary function `square`). However, relying on the programmer to take account of algebraic laws does not scale well. Second, you might post-process generated code to perform algebraic simplifications. However, this risks giving up the most desirable properties of multi-stage programming: namely that code transformations can be implemented by the programmer within the language itself. Finally, you might introduce specialized numeric representations that simplify generated code using laws of the underlying algebraic structure. This is, in our view, the most promising approach, and is widely used in the multi-stage programming literature (§7).

This paper builds on this third approach, re-examining the foundations of these *partially-static* structures, which are typically approached in *ad-hoc* fashion, and conceptualising them with a single *universal property*, in terms of the operations and equations we utilise. Universality translates into a functional specification which we need to implement and validate, and replaces the uncertainty of designing a new data structure with the precise activity of implementing a specification.

For concreteness we present our approach as a Haskell library, *flex*; however, the formulation transfers straightforwardly to other settings.

The *flex* library has a number of appealing features. First, *flex* applies algebraic simplification during the normal evaluation of a generating stage of a multi-stage program – i.e. *it turns algebraic simplification into beta reduction*, improving the performance of generated programs. Second, *flex* provides drop-in replacements for many common algebraic structures that make it possible to re-purpose existing polymorphic code for staging. Third, by unifying a wide variety of partially-static structures under the single concept of a *free extension*, *flex* exposes a tiny user-facing interface consisting of just three simple functions together with existing algebraic classes.

The **contributions** of this paper are as follows:


- §2 provides a general introduction to the principles underlying partially-static structures with algebraic laws, starting with binary operations without no equations, and showing the effect on the representation of adding associativity and commutativity laws.
- §3 shows how the considerations in §2 identify each partially static structure with a *free extension* — i.e. the coproduct of an algebra and a free object in the appropriate category. The unifying view provided by *free extensions* transforms programming with partially-static data, as monads have transformed programming with effects: it guides the definition of instances and programs, clarifies semantic properties, and provides a high-level framework in which many common patterns can be uniformly abstracted.
- §4 uses free extensions to define a variety of partially-static instances for algebraic structures: monoids (§4.1), commutative monoids and abelian groups (§4.5), sets (§4.8), commutative rings and semirings (§4.11), distributive lattices (§4.13) and algebraic data types (§4.15). In each case we illustrate the structures with examples drawn from the literature, showing how *frex* can be applied to programming problems such as arithmetic, pretty-printing, linear algebra and list manipulation, to improve generated code by algebraic simplifications. As the representative benchmarks in §6 demonstrate, these simplifications translate directly into faster running times.
- The major part of this paper is intended to be accessible, since we hope that *frex* and its techniques will be widely adopted by functional and multi-stage programmers, as well as in other settings such as compiler optimisation and partial evaluation. However, readers with some familiarity with universal algebra will find a more formal justification for using free extensions to represent partially-static structures in §5.
- §7 contextualizes our contributions among the work on partially-static data.
- Finally, Appendix A sketches a second implementation of *frex* in the multi-stage language BER MetaOCaml [Kiselyov 2014].

2 DEFINING PARTIALLY-STATIC STRUCTURES

How might we build implementations of algebraic operations that support computation with partially-static data? This section sketches a general approach to defining these *partially-static algebras* that can be used as drop-in replacements for standard instances in type-class polymorphic code. §4 describes the implementation of this approach as a high-level, modular and extensible library, which applies directly to several examples drawn from the literature.

We start with the simplest non-trivial algebraic structure. A *magma* consists of a set a along with a binary operation. We define a Haskell class `Magma` and introduce a picture form, representing the binary constructor \bullet as a binary branch in a tree:

```
class Magma a where (•) :: a → a → a
```



There are many instances of `Magma`, since any binary operation for a type forms a magma. The most general instance simply represents the tree structure directly:

```
data Tree a where
  Leaf :: a → Tree a
  Branch :: Tree a → Tree a → Tree a

instance Magma (Tree a) where
  (•) = Branch
```

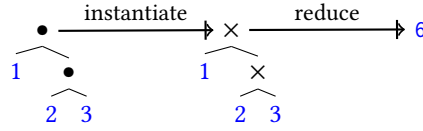
Any binary operator for a type gives rise to a magma. For example, here is a `Magma` instance for multiplication for integers, using a type isomorphism Int_\times to distinguish the instance from magmas for other integer operations such as addition or subtraction:

```

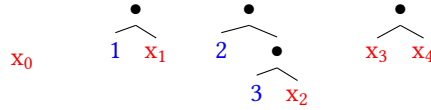
148 newtype Intx = Intx Int
149 instance Magma Intx where (Intx x) • (Intx y) = Intx (x × y)

```

Then we can define trees using the Magma operation together with integer values, and interpret those trees using the Magma instance for Int_x :



So the reduced form of a magma term instantiated to Int_x is (isomorphic to) a single integer. We would also like to compute with open terms containing free variables x_1, x_2 , such as these:



The type `Tree (Either Int (Code Int))` gives a simple representation for these mixed trees. However, this representation keeps all trees, even those without free variables, in unreduced form. **Free variables and binding-time analysis** In our setting free variables correspond to what are called *dynamic* variables in the partial evaluation literature, and integer values correspond to *static* values. The division into static and dynamic also extends to terms: dynamic terms are terms that mention some dynamic variables; other terms are static.

A classification of terms into static and dynamic is called a *binding-time analysis*. Fig. 1 shows a binding-time analysis for a magma tree. The left child contains a dynamic variable, and so it is classified as dynamic. The right child contains only static values, and so it is classified as static. The root tree has a dynamic left child, and so it is classified as dynamic.

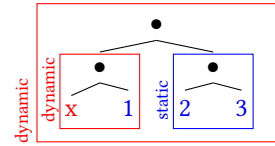
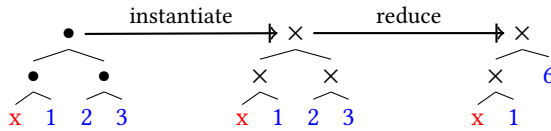


Fig. 1. Binding-time analysis

The aim of the binding-time analysis is to identify static sub-terms, since it is only those terms that can be reduced. Here is an illustration: the right sub-term is static, and consequently reduced, but no further reduction can take place, since all remaining terms are dynamic:



Even before instantiation it was evident that the dynamic terms would block reduction.

We can define a representation for mixed trees that takes binding-times into account, allowing reduction of static sub-trees. We start with a definition of binding-times, `BindingTime`, and an indexed type `BT` that reflects `BindingTime` at the type level, so that we can enforce constraints about binding-times in representations of data:

```

191 data BindingTime =
192     Sta
193     | Dyn
194
195 data BT :: BindingTime -> * where
196     BTSta :: BT Sta
197     BTDyn :: BT Dyn

```

Using `BindingTime` we define an indexed type `SD` to stand in for `Either a` (Code a), and a function `btSD` that computes the binding time for an `SD` value:

```

197 data SD :: BindingTime → * → * where          btSD :: SD bt a → BT bt
198   S :: a → SD Sta a                            btSD (S _) = BTSta
199   D :: Code a → SD Dyn a                       btSD (D _) = BTDyn

```

200 Finally, the Mag type represents magma terms that do not contain unreduced static subtrees.
 201 Leaves may be static or dynamic (LeafM); if the left branch of a tree is static then the right must be
 202 dynamic (Br1); if the left branch is dynamic then the right may be static or dynamic (Br2):

```

203 data Mag :: BindingTime → * → * where          btMag :: Mag bt a → BT bt
204   LeafM :: SD bt a → Mag bt a                  btMag (LeafM m) = btSD m
205   Br1 :: Mag Sta a → Mag Dyn a → Mag Dyn a    btMag (Br1 _ _) = BTDyn
206   Br2 :: Mag Dyn a → Mag r a → Mag Dyn a      btMag (Br2 _ _) = BTDyn

```

207 Here is the magma instance itself. The definition of \bullet uses binding times to reduce the number
 208 of cases to the three circumstances of interest: the case where both operands are static (in which
 209 case the elements must be coalesced), and the two cases that correspond to Br1 and Br2:

```

211 instance Magma a ⇒ Magma (Exists Mag a) where
212   E a • E b = case (btMag a, btMag b, a, b) of
213     (BTSta, BTSta, LeafM (S a), LeafM (S b)) → E (LeafM (S (a • b)))
214     (BTSta, BTDyn, l, r) → E (Br1 l r)
215     (BTDyn, _ , l, r) → E (Br2 l r)

```

216 The Mag instance uses an existential type to hide the BindingTime index in order to conform to the
 217 Magma interface. Existential types will come in useful on several occasions, so we define a general
 218 Exists type that hides the first parameter of a binary type constructor, using kind polymorphism
 219 to support arbitrary index kinds:

```

220 data Exists (f :: k1 → k2 → *) a where E :: f b a → Exists f a

```

222 **Equality and associativity** Trees are considered equal if they have the same shape and if cor-
 223 responding leaves are equal. Furthermore, trees are considered equal if they reduce to equal trees.

224 Adding laws to the algebraic structure groups trees into larger equivalence classes. For example,
 225 an associativity law equates trees with different branching structures:



230 Here a, b, c stand for arbitrary terms.

231 An magma with associativity is called a *semigroup*. We introduce a corresponding Haskell class,
 232 Semigroup, with the same members as Magma, and an additional obligation for instances, stated in
 233 a comment: the implementation of \bullet must be associative:

```

234 class Magma a ⇒ Semigroup a -- a • (b • c) ≡ (a • b) • c

```

235 For example, the instance for integers with multiplication can be made into a valid Semigroup
 236 instance, since multiplication is associative:

```

237
238 instance Semigroup Int×

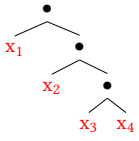
```

239 Magmas for non-associative operations, such as subtraction, do not give rise to semigroups.

240 Standard Haskell does not provide a way of ensuring that instances obey the laws of their class.
 241 However, some extensions, such as Liquid Haskell, provide various means of checking that in-
 242 stances are law-abiding.

243 With the associativity law, trees are equal if they have the same sequence of leaves; branching
 244 structure is no longer relevant.

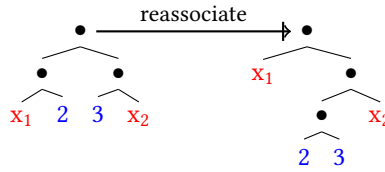
245



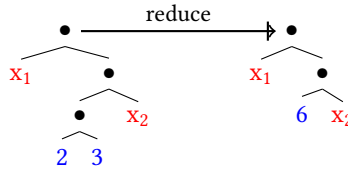
Equivalence and normal forms It is convenient to pick a canonical representative of each equivalence class, i.e. a *normal form*. Without the associativity law the canonical representative of each class is the tree that has been reduced as much as possible.

For trees without any free variables the most-reduced tree is a single static element, such as an element of Int_\times . For trees with *only* free variables, there is no reduction, and a normal form is a fully right-associated tree – equivalent to a non-empty value of type [Code a]. (This is called the *free semigroup*.)

What is the normal form for a tree with both integers and free variables? For magmas the normal form is represented by `Mag`, which ensures that static subtrees are reduced. For semigroups it is additionally possible to reassociate the tree to make adjacent nodes (i.e. nodes that are adjacent in the left-to-right leaf order) into siblings (i.e. nodes with the same parent):



Making adjacent static nodes into siblings makes it possible to reduce the parent node. So a normal form for trees with both static and dynamic elements is a fully right-associated tree with no adjacent static nodes:



In the partial evaluation literature, a change in term structure that increases the number of static terms is called a *binding-time improvement*.

As with `Magma`, we can define a datatype for mixed semigroup trees in normal form. Besides ensuring that there should be no unreduced static subtrees the `Semi` type adds a new constraint that all trees are kept in right-associated form, by making the first arguments of the two binary-branching constructors `ConsS` and `ConsD` atoms rather than trees:

```
data Semi :: BindingTime -> * -> * where
  LeafS :: SD bt a -> Semi bt a
  ConsS ::      a -> Semi Dyn a -> Semi Dyn a
  ConsD :: Code a -> Semi r  a -> Semi Dyn a
```

It is convenient to define auxiliary functions that add static and dynamic elements to the left of the tree. The `consS` function adds a static element to a `Semi` tree; if there is a static element in leftmost position already, `consS` combines the two elements using `•`:

```
consS :: Magma a => a -> Exists Semi a -> Exists Semi a
consS h (E (LeafS (S s))) = E (LeafS (S (h • s)))
consS h (E t@(LeafS (D _))) = E (ConsS h t)
consS h (E (ConsS s t)) = E (ConsS (h • s) t)
consS h (E t@(ConsD _ _)) = E (ConsS h t)
```

The `consD` function is simpler, because dynamic elements may be added to the left of any tree:

```

295         consD :: Code a → Exists Semi a → Exists Semi a
296         consD h (E t) = E (ConsD h t)

```

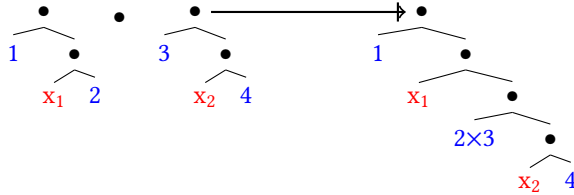
297 Finally, here are Magma and Semigroup instances for Semi:

```

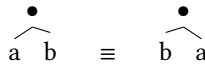
298
299 instance Semigroup a ⇒ Magma (Exists Semi a)      instance Semigroup a ⇒
300   where E (LeafS (S s)) • l = consS s l           Semigroup (Exists Semi a)
301         E (LeafS (D d)) • l = consD d l
302         E (ConsS h t) • l = consS h (E t • l)
303         E (ConsD h t) • l = consD h (E t • l)

```

304 The final two cases of • may traverse the entire left operand to handle the case where the final
305 static element on the left should be coalesced with the initial static element on the right:



313 **Commutativity** Adding an associativity law made trees equivalent that were previously distinct and made trees reducible that were previously irreducible. Adding a second law for commutativity coalesces more equivalence classes and adds further opportunities for reduction:



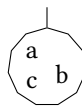
319 The CSemigroup class adds a law to make • commutative:

```

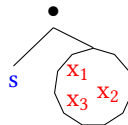
320
321 class Semigroup a ⇒ CSemigroup a    -- a • b ≡ b • a

```

322 Since many types of element have no notion of ordering, we must represent ordering directly
323 in the tree structure. We introduce a new tree constructor, with an unordered bag of *n* children:



328 The ability to reorder subtrees significantly simplifies the normal form. Since all the static elements can be moved to one end, and subsequently reduced, the normal form for a mixed static-dynamic commutative semigroup is simply a pair of an optional single static element and an unordered bag of dynamic variables:



333 Here is the normal form in Haskell, using the standard multiset to represent bags:

```

334
335 data CSemi a = CSemi (Maybe a) (MultiSet (Code a))

```

340 Then $l \bullet r$ is a pair whose components are built from the corresponding parts of l and r using
341 • for the underlying magma lifted into the **Maybe** applicative for the static component and multiset
342 union for the bag of dynamic variables:

```

344 class (algebra a, algebra b, algebra (Coproduct algebra a b)) => Coproduct algebra a b
345   where data family Coprod algebra a b :: *
346         inl :: a -> Coprod algebra a b
347         inr :: b -> Coprod algebra a b
348         eva :: algebra c => (a -> c) -> (b -> c) -> Coprod algebra a b -> c
349

```

Fig. 2. The Coproduct interface

```

351
352 instance CSemigroup a => Magma (CSemi a) where
353   CSemi s1 d1 • CSemi s2 d2 = CSemi (liftA2 (•) s1 s2) (union d1 d2)
354

```

3 PARTIALLY-STATIC DATA: A GENERAL INTERFACE

§2 introduced the ideas behind computing with partially-static algebra. We now look at how to turn these ideas from a design pattern into an general abstraction that can be instantiated with particular algebras to support partially-static computation that turns algebraic equations into beta reduction. The general abstraction is the core of our Haskell library, *flex*, which provides an extensible and modular interface to partially-static data. The ideas behind *flex* are not tied to any particular language, and we also describe a second implementation of *flex* in MetaOCaml (Appendix A).

3.1 Partially-static data: requirements

With the aim of constructing an interface PS to partially static data, we start with a list of requirements, distilled from the discussion up to this point.

First, the PS type is intended to work with a variety of algebraic structures, and so we parameterise it by an algebraic structure and the type of the data it represents:

```

363 PS :: (* -> Constraint) -> * -> *

```

The first parameter is a *constraint* that stands for a type class such as Magma or CSemigroup.

Next, it should be possible to use partially-static values in place of either fully-static or fully-dynamic values, and so PS should support injections from static and dynamic data:

```

374 sta :: algebra a => a -> PS algebra a
375 dyn :: Code a -> PS algebra a

```

(We will extend these type signatures with further constraints in §3.3.) In §2, *sta* and *dyn* took various forms: *LeafM* (S -) and *LeafM* (D -) for magmas, and *LeafS* (S -) and *LeafS* (D -) for semigroups. For commutative semigroups the injections can be written as follows

```

380 staCS = λs -> CSemi (Just s) empty
381 dynCS = λd -> CSemi Nothing (singleton d)

```

Furthermore, PS algebra itself should be an instance of algebra, since it is intended to stand in for contexts where an algebra instance is expected. For example, the partially-static structure for semigroups should be a semigroup instance.

3.2 Coproducts

These considerations suggest designing *flex* around *coproducts* – that is, not the familiar binary sums that Haskell calls *Either*, but the more general notion of coproducts in a category, whose representation varies according to the category.

Fig. 2 defines the coproduct interface as a Haskell type class, Coproduct, with three parameters and four components.

392


```

393     class algebra (FreeA algebra x) ⇒ Free algebra x where
394     data family FreeA algebra x :: *
395     pvar :: x → FreeA algebra x
396     pbind :: algebra c ⇒ FreeA algebra x → (x → c) → c
397
398
399

```

Fig. 3. Free algebraic structures, Free

The three parameters algebra, a, and b, respectively represent a type class for a particular algebraic structure and the two types that comprise the coproduct. For example, Coproduct Monoid **Int String** represents the coproduct of **Int** and **String** in the category of monoids. The first two class constraints algebra a and algebra b constrain the instantiation of the parameters to types that have instances of algebra. For example, the instantiation Coproduct Monoid **Int String** is only allowed if there exist type class instances of Monoid for **Int** and **String**.

The first component, Coprod, is the type of values of the coproduct of a and b in the category algebra. Coprod is an *associated data type* [Chakravarty et al. 2005], whose definition varies with each instance of the Coproduct class. For example a coproduct in the category of *monoids*, Coprod Monoid **Int String**, is an alternating sequence similar to the Semi type of §2 (and defined more precisely in §4.1), whereas the type Coprod Set **Int String**, a coproduct in the category of sets, is the familiar binary sum type (§4.8).

The final class constraint algebra (Coproduct algebra a b) ensures that each instantiation of Coprod is an instance of the algebraic structure algebra — for example, there must be a Monoid instance for the type Coprod Monoid a b. The second and third components, inl and inr, inject values of a and b into Coprod. The final component eva is a kind of fold that generalizes the standard **either** function, producing a value of type c from a value of type Coprod and functions from a and b to c. For example, if algebra has an operation • then eva behaves as follows:

```

417     eva f g (inl s1 • inr d1 • inl s2 • ... ) ∼→ f s1 • g d1 • f s2 • ...
418

```

where the • operations on the left are from Coprod algebra a b and the • operations on the right are from c. (In particular, eva inl inr is the identity.) The constraint algebra c in the type of eva ensures that c is also an instance of the algebraic structure associated with the instance: for example, a coproduct of monoids can only be eliminated into a monoid.

3.3 Free objects and free extensions

§3.2 provides a general interface to coproducts. However, computing with partially-static data requires a particular form of coproduct, where the left type is some type a and the right type is generated by quoted terms Code a. Furthermore, while there are no constraints on the left algebraic structure, the right structure is always *free*, since Code values do not have computational behaviour or additional equations.

In other words, we are interested in what are called *free extensions*: coproducts of algebras and free objects. This section defines a general interface to free algebras and shows how they combine with Coproduct to give free extensions.

Fig. 3 defines a type class Free indexed by a constraint, algebra, and a type x. An instance Free algebra x represents the free algebra for algebra with variables in x; for example, Free Semigroup (Code **Int**) represents the free semigroup with variables in Code **Int**.

There are three class members. First, FreeA is the type of values in the free algebra; as with Coprod, the definition of the type varies with each instance. For instance, the free algebra for Semigroup is a non-empty list of variables, while the free algebra for CSemigroup is a non-empty multiset. Second, pvar injects a variable into FreeA. Finally, the monadic pbind maps a value of FreeA into another algebra c via a function that injects variables into c. The class constraint

```

442 FreeExtC :: (* → Constraint) → * → Constraint
443 type FreeExtC algebra a = Coproduct algebra a (FreeA algebra (Code a))
444
445 FreeExt :: (* → Constraint) → * → *
446 type FreeExt algebra a = Coprod algebra a (FreeA algebra (Code a))
447

```

Fig. 4. Free extensions: constraint alias `FreeExtC` and type alias `FreeExt`

algebra (FreeA algebra x) stipulates that there must be an algebra instance for `FreeA` so that, for example, the type `FreeA Semigroup m` must also support the operation \bullet in addition to `pvar`.

Fig. 4 shows how the `Coproduct` and `Free` classes combine to give the definition of a free extension as a coproduct of an algebra and a corresponding free algebra. There are two definitions: `FreeExtC` defines a constraint that may appear to the left of a fat arrow (\Rightarrow), while `FreeExt` names the type associated with the `FreeExtC` instance. For example, `FreeExt Semigroup Int` defines the type of the free extension of semigroups for integers as an alternating sequence (the coproduct) of integers and lists of integer code values (the free algebra).

The definitions of Fig. 4 form a crucial part of the general partially-static data interface (§3.5).

It is worth emphasizing that the free extension structure may look quite different from both the static and the dynamic instance. For example, in the semigroup of integers with multiplication each expression reduces to a single integer; in the *free* semigroup each expression reduces to a non-empty sequence of names; in the *semigroup free extension* each expression reduces to a sequence of integers and names, with the additional constraint that there are no adjacent integer elements.

These considerations lead to new definitions for `sta` and `dyn`. We drop the provisional name `PS` in favour of using `FreeExt` directly, and write:

```

466 sta :: (algebra a, FreeExtC algebra a) ⇒ a → FreeExt algebra a
467 sta = inl
468 dyn :: (Free algebra (Code a), FreeExtC algebra a) ⇒ Code a → FreeExt algebra a
469 dyn = inr . pvar
470

```

The first definition defines `sta`, the function that builds partially-static representations from static values, as the left injection into the free extension of an algebra. The second definition defines `dyn` as the composition of the injection into the free object and the injection into the free extension. The constraints ensure that there are `algebra` and `Free algebra` instances available for the static and dynamic components and `FreeExtC algebra` instances available for the results.

Viewing partially-static algebraic structures as free extensions makes explicit some additional requirements on the implementation of instances. For example, `sta` should be a homomorphism with respect to each operation \bullet , i.e.:

$$\text{sta } x \bullet \text{sta } y \quad \equiv \quad \text{sta } (x \bullet y)$$

Furthermore, `eva` should preserve the laws of the algebra, so that terms that are equivalent under the laws remain equivalent when `eva` is applied.

3.4 From partially-static to fully dynamic

The `eva` function is a general-purpose destructor for coproducts. The most common use of `eva` with partially-static data is *residualization*: turning partially-static values into fully-dynamic values.

A residualization function can be obtained from `eva` as follows. First, specialize `eva` to `FreeExt`, where the second type parameter to `Coproduct` is instantiated to `FreeA algebra (Code a)`:

```

491 FreeExtC :: (* → Constraint) → * → Constraint
492 type FreeExtC algebra a = Coproduct algebra a (FreeA algebra (Code a))
493
494 FreeExt :: (* → Constraint) → * → *
495 type FreeExt algebra a = Coprod algebra a (FreeA algebra (Code a))
496
497 sta :: (algebra a, FreeExtC algebra a) ⇒ a → FreeExt algebra a
498 sta = inl
499
500 dyn :: (Free algebra (Code a), FreeExtC algebra a) ⇒ Code a → FreeExt algebra a
501 dyn = inr . pvar
502
503 cd :: (Lift a, Free algebra (Code a), algebra (Code a), FreeExtC algebra a) ⇒
504     FreeExt algebra a → Code a
505 cd = eva tlift (`pbind` id)
506

```

Fig. 5. *frex*'s generic interface to partially-static data

```

508 evaFE :: (FreeExtCon algebra a, algebra c) ⇒
509         (a → c) → (FreeA algebra (Code a) → c) → FreeExt algebra a → c
510 evaFE = eva
511

```

Next, instantiate the return type of `evaFE` to `Code a`:

```

512
513 evaCode :: (FreeExtCon algebra a, algebra (Code a)) ⇒
514         (a → Code a) → (FreeA algebra (Code a) → Code a) → FreeExt algebra a → (Code a)
515 evaCode = evaFE
516

```

Finally, supply suitable arguments for the first two parameters of `evaCode`. The first argument to `eva` converts static values to code; this is the purpose of `tlift`, a typed variant of Template Haskell's lift function that provides an interface to cross-stage persistence:

```

517
518
519
520         tlift :: Lift a ⇒ a → Code a
521         tlift = liftM TExp . lift
522

```

The second argument to `eva` builds `Code` values from values of a free object; this can be accomplished with `pbind` (§3.3).

This series of specializations produces the following residualization function, which is the final component of *frex*'s general interface to partially-static data.

```

523
524
525
526         cd :: (Lift a, Free algebra (Code a), algebra (Code a), FreeExtC algebra a) ⇒
527             FreeExt algebra a → Code a
528         cd = eva tlift (`pbind` id)
529

```

Here is `cd` in action:

```

530
531         cd ((dyn [x] • sta 2) • (sta 3 • dyn [y])) ∼ [x × 6 × y]
532

```

And, of course, `cd` preserves the laws of the algebra, since `eva` does, so that equivalent partially-static computations are residualized to equivalent code.

3.5 Using *frex*

Fig. 5 summarises *frex*'s interface to partially-static data.

How does one use *frex* to write programs? In order to use *frex* to program with partially-static representations for an algebraic structure such as `Monoid` or `Ring`, two things are needed: a free

539

```

540     class Monoid t where
541         1 :: t
542         (⊗) : t → t → t
543
544         1 ⊗ x ≡ x ≡ x ⊗ 1
545         x ⊗ (y ⊗ z) ≡ (x ⊗ y) ⊗ z

```

Fig. 6. Monoids and their laws

extension instance for the structure, and instances of the structure for particular types such as `Int` or `String`. In many cases, these requirements will be met by the combination of *frex* and existing Haskell libraries. Here are some common scenarios:

Using *frex* with existing instances *Frex* includes pre-defined free extensions for a number of structures, including sets, monoids, commutative rings, distributive lattices, abelian groups, and F-algebras. If *frex* already defines a free extension for a structure and some other library provides instances of the structure for types in the program, using *frex* is typically a matter of inserting `sta`, `dyn` and `cd`.

For example, since the standard library provides a `String` instance for `Monoid`, nothing more is needed to write programs involving partially-static strings:

```

556     cd ((dyn [x] `mappend` sta "abc") `mappend` (sta "def" `mappend` dyn [y]))
557     ~> [x `mappend` "abcdef" `mappend` y]

```

Creating new instances to use with *frex* Similarly, no special work is needed to add new instances to *frex* when the free extension is already defined. For example, adding a `CMonoid` instance for `()` (building on the existing `Monoid` instance) is sufficient to enable *frex*'s commutative monoid simplifications:

```

562         instance CMonoid ()

```

Adding new classes to *frex* Finally, adding new structures to *frex* is a matter of defining suitable `Coproduct` and `Free` instances; again, no modifications to *frex* internals are needed. §4 provides a number of examples, building on the discussion in §2.

4 INSTANCES AND APPLICATIONS

With the general interface in place (Fig. 5), we now turn to the implementation of free extensions for common algebraic structures: monoids 4.1, commutative monoids and abelian groups 4.5, sets 4.8, commutative rings 4.11, distributive lattices 4.13, and F-algebras 4.15.

In many cases the general coproduct for the structure and the free object can be defined separately, then combined to give the free extension. However, in some cases (e.g. §4.11) where it is not possible to give a general form for the coproduct, we define the free extension directly.

4.1 Coproduct of monoids

Fig. 6 defines a `Monoid` class. The free extension for monoids is an instance of the more general coproduct of monoids, and a slight variant of the partially-static structure for semigroups (§2).

Fig. 8 gives the `Coproduct` instance for the `Monoid` class constraint, built from `Monoid` instances `a` and `b`. As with `Semi`, the `Coprod` type is defined as a sequence of alternating `a` and `b` elements (Fig. 7), indexed by the type of the first element in the list, and with the index hidden by an existential (here `M`) to allow either `a`-prefixed or `b`-prefixed sequences. Unlike `semi`, the sequence may be empty, since `Monoid` adds an identity element. The `inl` and `inr` injections create singleton sequences, and `eva` is a fold over the sequence, using the `⊗` operation of the target monoid to combine the results. As a small optimization, `eva` does not map the `Empty` constructor to `1` except in the case where the input sequence has no elements.

The constraints in the `Coproduct` class specify that each `data` instance `Coprod alg a b` is an instance of `alg`. For example the type `Coprod Monoid a b` should be an instance of `Monoid`.

```

589     data AorB = A | B
590     data Alternate :: AorB → * → * → * where
591         Empty :: Alternate any a b
592         ConsA :: a → Alternate B a b → Alternate A a b
593         ConsB :: b → Alternate A a b → Alternate B a b
594

```

Fig. 7. An alternating sequence of a and b elements

```

596     instance (Monoid a, Monoid b) ⇒ Coproduct Monoid a b where
597         data Coprod Monoid a b where M :: Alt _ a b → Coprod Monoid a b
598         inl a = M (ConsA a Empty)
599         inr b = M (ConsB b Empty)
600         eva (f :: a → d) (g :: b → d) (M c) = eva' c
601         where eva' :: Alternate start a b → d
602               eva' Empty = 1
603               eva' (ConsA a Empty) = f a
604               eva' (ConsB b Empty) = g b
605               eva' (ConsA a m) = f a ⊗ eva' m
606               eva' (ConsB b m) = g b ⊗ eva' m
607

```

Fig. 8. The coproduct of monoids

```

609     instance (Monoid a, Monoid b) ⇒ Monoid (Coproduct Monoid a b) where
610         1 = M Empty
611         M l ⊗ M r = l `mul` r
612         where mul :: (Monoid a, Monoid b) ⇒
613               Alternate s a b → Alternate s' a b → Coprod Monoid a b
614               mul l Empty = M l
615               mul Empty r = M r
616               mul (ConsA a m) r | M m' <- mul m r = M (consA a m')
617               mul (ConsB b m) r | M m' <- mul m r = M (consB b m')
618               consA :: Monoid a ⇒ a → Alternate s a b → Alternate A a b
619               consA a Empty = ConsA a Empty
620               consA a (ConsA a' m) = ConsA (a ⊗ a') m
621               consA a r@(ConsB _ _) = ConsA a r
622               consB :: Monoid b ⇒ b → Alternate s a b → Alternate B a b
623               consB b Empty = ConsB b Empty
624               consB b (ConsB b' m) = ConsB (b ⊗ b') m
625               consB b r@(ConsA _ _) = ConsB b r
626

```

Fig. 9. The coproduct of monoids is a Monoid

Fig. 9 defines the Monoid instance for Coprod Monoid a b, where a and b also have instances of Monoid.

The 1 and ⊗ operations respectively construct an empty sequence and concatenate two sequences. Prepending an a element to an a-prefixed sequence combines the element with the head of the sequence using the ⊗ operation of the a monoid, and similarly for b, *mutatis mutandis*.

```

638 instance Free Monoid x where
639     newtype FreeA Monoid x = P [x] deriving (Monoid)
640     pvar x = P [x]
641     P [] `pbind` f = 1
642     P xs `pbind` f = foldr ((*) . f) 1 xs
643

```

Fig. 10. Free instance for Monoid

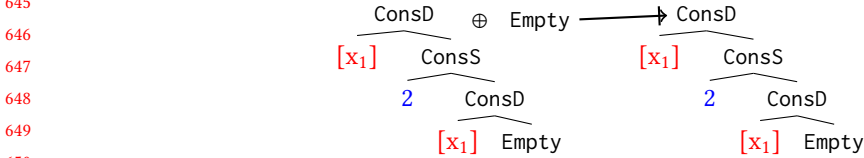


Fig. 11. Partially-static monoid: dropping 1

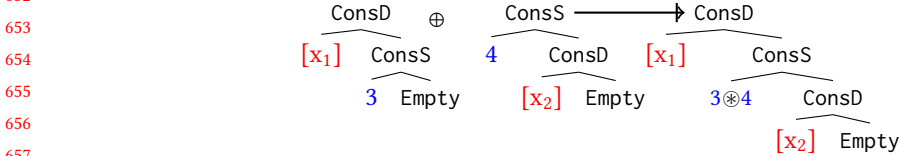


Fig. 12. Partially-static monoid: coalescing adjacent static values

4.2 Free monoids and the free extension

Fig. 10 shows the Free instance for Monoid. The free monoid with variables in x is simply a list of x values. The `pbind` function maps a free monoid value into any other monoid:

$$(\text{pvar } x_1 \otimes \text{pvar } x_2 \otimes \dots \otimes \text{pvar } x_n) \text{ `pbind` } f \rightsquigarrow (f \ x_1 \otimes f \ x_2 \otimes \dots \otimes f \ x_n)$$

Type class resolution combines the Free and Coproduct instances to form the free extension `FreeExt`.

Figs. 11 and 12 illustrate how *frex*'s partially-static monoid performs static reductions using the monoid laws.

4.3 Example: improving printf with partially-static monoids

To show the Monoid free extension in action, we consider an example from the functional programming literature [Asai 2009; Danvy 1998]: typed `sprintf`. It is straightforward to use staging to turn `sprintf` from a function into a code generator [Yallop and White 2015]; however, a naive approach results in code that contains too many catenations.

For example, the following call to `sprintf` generates a function that prints two integer arguments with "ab" interposed:

```
sprintf ((int ++ lit "a") ++ (lit "b" ++ int))
```

When `sprintf` is staged using a simple binding time analysis the result contains four catenations:

```
[ λx y →((( " ++ show x) ++ "a") ++ "b") ++ show y ]
```

Since strings form a monoid under catenation, switching to *frex*'s partially static operations generates the following more efficient code:

```
[ λx y →show x ++ ("ab" ++ show y) ]
```

(In fact, as we shall see, *Frex* can also generate the more efficient code that makes a single call to an n -ary catenation function).

```

687 class Format f where
688   type Acc f a :: *
689   lit :: String → f a a
690   cat :: f b a → f c b → f c a
691   int :: f a (Acc f Int → a)
692   str :: f a (Acc f String → a)
693   sprintf :: f (Acc f String) a → a
694

```

Fig. 13. Format signature

```

696 instance Format Fmt where
697   type Acc Fmt a = a
698   lit x = Fmt $ \k s →k (s ++ x)
699   f `cat` g = Fmt (fmt f . fmt g)
700   int = Fmt $ \k s x →k (s ++ show x)
701   str = Fmt $ \k s x →k (s ++ x)
702   sprintf p = fmt p id ""
703

```

Fig. 14. An unstaged Format implementation

```

instance Format FmtS where
  type Acc FmtS a = Code a
  lit x = FmtS $ \k s →k [ $s ++ x ]
  f `cat` g = FmtS (fmtS f . fmtS g)
  int = FmtS $ \k s x →k [ $s ++ show $x ]
  str = FmtS $ \k s x →k [ $s ++ $x ]
  sprintf p = fmtS p id [ "" ]

```

Fig. 15. An staged Format implementation

```

instance Format FmtPS where
  type Acc FmtPS a = Code a
  lit x = FmtPS $ \k s →k (s ⊗ sta x)
  f `cat` g = FmtPS (fmtPS f . fmtPS g)
  int = FmtPS $ \k s x →k (s ⊗ dyn [ show $x ])
  str = FmtPS $ \k s x →k (s ⊗ dyn x)
  sprintf (FmtPS p) = p cd 1

```

Fig. 16. A partially-static Format implementation

Fig. 13 gives a minimal interface for formatted printing. The type constructor f represents format specifications; its two parameters respectively represent the result and the input type of a `sprintf` instantiation. The following three operations construct format strings: `lit s` is a format string that accepts no arguments and prints s ; `cat x y` concatenates x and y ; `int` is a format string that accepts and prints an integer argument. Finally, `sprintf` combines a format string with corresponding arguments to construct formatted output. Asai [2009] gives further details.

Here is an implementation of Fig. 13 in continuation-passing style, using an accumulator:

```

newtype Fmt r a = Fmt {fmt :: (String → r) → String → a }

```

With this implementation, a format string `Fmt` is a function accepting a continuation argument of type `String → r` and an accumulator of type `String`. Both `lit` and `int` call `k` directly, passing an extended string; `cat` is simply function composition. The function `sprintf` passes the identity function as a top-level continuation along with an empty accumulator (Fig. 14).

Staging `sprintf` is straightforward (Fig. 15). We treat format strings statically; arguments and, consequently, the accumulator, are dynamic. The `cat` function is unchanged, and the rest of the implementation is annotated in accordance with the assignment of static and dynamic classifications:

```

newtype FmtS r a = FmtS {fmtS :: (Code String → r) → (Code String → a)}

```

The generated code (shown at the beginning of this section) is sub-optimal because the staging is simplistic: every concatenation is delayed, even when both operands are statically available.

Staging using *frex*'s partially-static monoid is also straightforward (Fig. 16). The steps are as follows, starting from the unstaged implementation: replace `String` with `FreeExt Monoid String`, replace `cat` and `""` with `⊗` and `1`, insert `sta` and `dyn` to inject static and dynamic expressions, and replace the top-level continuation with the residualization function described above:

```

newtype FmtPS r a = FmtPS {fmtPS :: (PS Monoid String → r) → PS Monoid String → a }

```

This implementation statically constructs a canonical representation before residualizing, eliminating nesting and redundant concatenations with `1`; the result (shown at the beginning of this section) contains only two concatenations rather than the original four.

```

736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784

```

	$\mathbb{1} \otimes x \equiv x$
<code>class Monoid m => CMonoid m</code>	$x \otimes (y \otimes z) \equiv (x \otimes y) \otimes z$
<code>class CMonoid c => CGroup c where cinv :: c -> c</code>	$x \otimes y \equiv y \otimes x$
	$\text{cinv } x \otimes x \equiv \mathbb{1}$

Fig. 17. Commutative monoids and abelian groups

§4.4 gives a second residualization function for partially-static string monoids that generates a single call to n -ary concat rather than a sequence of binary catenations.

4.4 Residualization for monoids

It is often possible to give a more efficient residualization function for a particular instance of a structure. Here we sketch how to give an alternative residualization function for the partially-static monoid of strings. There is no need to step outside the framework provided by *frex*; it is sufficient to instantiate the residualizing call to *eva* with an alternative Monoid instance.

The Monoid interface (Fig. 6) exposes nullary and binary constructors $\mathbb{1}$ and \otimes . However, for some monoids it is most efficient to combine more than two elements in a single operation. The partially-static monoid structure generates the following code for the `printf` example in §4.3:

```
[[ s1 ++ $d ++ s2 ]]
```

However, depending on the representation of strings, it may be more efficient to generate a single call to an n -ary catenation function. (For example, n -ary catenation is more efficient in OCaml, where strings are strict arrays, but not in Haskell, where strings are lazy lists by default.)

```
[[ concat [s1, $d, s2] ]]
```

It is straightforward to write an alternative to `cd` specialized to the monoid of strings that generates this more efficient code. (See the `Printf.cdStrings` function in the supplementary material.)

The opportunity to improve code generation at the point of residualization is one of the advantages of the free extensions view over existing ad-hoc approaches to partially-static representations.

The interface to partially-static data in earlier work (e.g. [Kaloper-Meršinjak and Yallop 2016]) typically provides `cd` as the only way to inspect partially-static data. The coproduct view improves on this approach, providing two additional ways of inspecting partially-static values: the *eva* function, and (in the Haskell implementation) the `Coprod` type. With *eva* and `Coprod` it becomes possible to perform further optimizations at the point of code generation.

§6 shows how *frex*'s simplifications to the code generated by the staged `printf` function example lead to significant performance improvements.

4.4.1 Practical considerations: canonicity. Ideally the partially-static representation should be *canonical*: expressions that are statically equivalent under the laws of the algebra should have the same representation in PS.

Unfortunately, it is not always possible to achieve full canonicity; for example, `ConsS` can store empty monoid elements, even though these could be eliminated according to the unit elimination laws. This kind of deviation from canonicity is sometimes unavoidable, since it is not always possible to determine whether a monoid element should be considered empty. (For example, the monoid of endofunctions does not support equality.)


```

785     instance (CMonoid a, CMonoid b) => Coproduct CMonoid a b where
786         data Coprod CMonoid a b = C a b
787         inl a = C a 1
788         inr b = C 1 b
789         eva f g (C a b) = f a * g b
790

```

Fig. 18. The coproduct of commutative monoids

```

791
792
793     instance Ord x => Free CMonoid x where
794         newtype FreeA CMonoid x = CM (MultiSet x)
795         pvar = CM . singleton
796         CM b `pbind` f = emit (toList b)
797         where emit [] = mempty
798               emit [x] = f x
799               emit (x:xs) = f x `mappend` emit xs
800
801     instance Ord x => Free CGroup x where
802         newtype FreeA CGroup x = CG (Map x Int)
803         pvar x = CG (singleton x 1)
804         CG b `pbind` f = emit (toList b)
805         where emit [] = mempty
806               emit [(x,1)] = f x
807               emit ((x,0):xs) = emit xs
808               emit ((x,n):xs) | n < 0 = cinv (f x) `mappend` emit ((x,n+1):xs)
809                                   | otherwise = f x `mappend` emit ((x,n-1):xs)
810

```

Fig. 19. Free instances for commutative monoids and abelian groups

4.5 Coproducts of commutative monoids and abelian groups

Fig. 17 shows the interface to commutative monoids (CMonoid) and abelian groups (CGroup). The CMonoid inherits the methods from the Monoid interface and adds the commutativity law. The only difference between Monoid and CMonoid is the set of laws tacitly associated with the class.

However, adding the commutativity law to the monoid interface leads to quite a different coproduct structure (Fig. 18). As with the transition from semigroups to commutative semigroups (§2), applying commutativity to the alternating sequence structure of the monoid coproduct allows the elements of each constituent monoid to be brought together and coalesced. The alternating sequence consequently collapses into a two-element sequence – i.e. a Cartesian product of sets. (The coproduct of abelian groups is isomorphic to the commutative monoid coproduct, and not shown.)

4.6 Free commutative monoid and abelian group

Fig. 19 shows implementations of free algebras for commutative monoids and abelian groups.

The CMonoid and CGroup free algebras are multisets (bags) and finite maps (dictionaries) with integer values, respectively:

$$\text{pvar } x \otimes \text{cinv (pvar } y) \otimes \text{pvar } x \otimes \text{cinv (pvar } y) \quad \rightsquigarrow \quad \{ x \mapsto 2, y \mapsto -2 \}$$

In each case the implementation of pbind maps f over each element in the representation and combines the results using mappend; the pbind implementation for abelian groups additionally applies cinv to each output element when the count associated with the element is negative:

```

834                                     instance Coproduct Set a b where
835                                     data Coprod Set a b = Inl a | Inr b
836     class Set a                       inl = Inl
837     instance Set a                     inr = Inr
838                                     eva f g (Inl x) = f x
839                                     eva f g (Inr y) = g y
840

```

Fig. 20. The coproduct of structures with no equations

```

842     (pvar x * cinv (pvar y) * pvar x * cinv (pvar y)) `pbind` f
843     ~> f x * f x * cinv (f y) * cinv (f y)
844

```

As for monoids, the Free and Coproduct instances combine to form the free extension.

4.7 Example: power with partially-static commutative monoids

We are ready to revisit the staged power function from the introduction. The naively-staged power function generates code with a linear sequence of multiplications that concludes with an unnecessary multiplication by 1:

```

851     [ \x -> $(power [x] 6) ] ~> [x * (x * (x * (x * (x * (x * 1)))))] ]
852

```

Here is an implementation of power suitable for use with *frex*:

```

853
854     power :: CMonoid m => m -> Int -> m
855     power x 0 = 1
856     power x n = x * power x (n - 1)
857

```

This new definition illustrates two benefits of *frex*. First, there are no more low-level staging annotations; it is sufficient to make the code more polymorphic by defining power for an arbitrary commutative monoid, then instantiate with *frex*'s predefined free extension. Second, *frex*'s simplifications improve the generated code. With a *cd* function specialized to commutative monoids (in the same vein as the specialized monoid *cd* of §4.4), the six multiplications in the naively-staged version can be reduced to three:

```

863     cdPower (power (dyn [x]) 6) ~> [let y = x*x in let z = y*y in y*z ]
864

```

As in the previous example, these improvements depend only on a specialized *cd*; the implementation of power and the partially-static representations are untouched. The specialized *cd* can be implemented in terms of *eva* by supplying a suitable CMonoid instance, or directly on the CoProd representation.

4.8 Coproduct of sets

One special case of an algebraic structure is Set, the structure with no equations.

Fig. 20 shows the instance for Coproduct in the category of sets. The Set class has no constraints or methods, and a single instance that encompasses every Haskell type. Consequently, the Coproduct instance for Set has no constraints, since all the Coproduct class constraints are satisfied by the single Set instance. The associated Coprod type is simply the familiar type of binary sums (called **Either** in the Haskell standard library), with *inl* and *inr* as its two constructors, and *eva* corresponding to the familiar **either** function¹.

4.9 Free algebra for sets

Fig. 21 shows the free algebra for sets, where a value is a single variable, and *pbind* is application.

¹<https://hackage.haskell.org/package/base-4.10.0.0/docs/Data-Either.html>

```

883         instance Free Set x where
884             newtype FreeA Set x = F x
885             pvar = F
886             F x `pbind` k = k x

```

Fig. 21. Free instance for Set

```

889
890             (a ⊕ b) ⊕ c ≡ a ⊕ (b ⊕ c)
891     class Ring a where           a ⊕ b ≡ b ⊕ a     a ⊕ 0 ≡ a
892     (⊕), (⊗) :: a → a → a       a ⊕ rneg a ≡ 0
893     rneg :: a → a                (a ⊗ b) ⊗ c ≡ a ⊗ (b ⊗ c)
894     0, 1 :: a                    a ⊗ b ≡ b ⊗ a     a ⊗ 1 ≡ a
895                                 a ⊗ (b ⊕ c) ≡ (a ⊗ b) ⊕ (a ⊗ c)

```

Fig. 22. Commutative rings

```

896
897     data Multinomial x a = MN (Map (MultiSet x) a)
898
899     instance Ord x ⇒ Free Ring x where
900         newtype FreeA Ring x = RingA (Multinomial x Int) deriving (Ring)
901         pvar x = RingA (MN (Map.singleton (MultiSet.singleton x) 1))
902         RingA xss `pbind` f = evalMN initMN f xss

```

Fig. 23. Free commutative rings

Although the Set free extension is a rather impoverished structure, it is perhaps the most frequently used representation for partially-static data in the multi-stage programming literature. The Set free extension does not take advantage of any equations; nevertheless, switching from a binary binding-time classification (in which each expression is fixed as always-static or always-dynamic) to the *possibly-static* world of the Set free extension (in which an expression may switch between static and dynamic on different executions) is a significant improvement for many applications.

4.10 Example: possibly-static data

Here is a simple example of possibly-static data. The `isDigitPS` function classifies a possibly-static character using the standard `isDigit` function. If the character is static then the classification is performed immediately; if it is dynamic then the classification is deferred to the next stage.

```

918     isDigitPS :: FreeExt Set Char → FreeExt Set Bool
919     isDigitPS (Inl c) = Inl (Char.isDigit c)
920     isDigitPS (Inr c) = Inr [Char.isDigit $ c ]

```

Without partially-static data it would be necessary either to have two functions for the static and dynamic cases, or to convert every character to a dynamic value, losing the opportunity to perform further static computation on the result.

The idea naturally generalizes to lift arbitrary functions to operate on possibly-static values.

4.11 Free algebra and free extension of commutative rings

Fig. 22 shows the interface Ring and the axioms for commutative rings.

Fig. 23 defines the free commutative ring on a set x , defined as multinomials — i.e finite sums of products of variables in x with integer coefficients (). The call `pvar x` builds a representation of

```

932 instance (Ring a, Ord x) => Coproduct Ring a (FreeA Ring x) where
933   newtype Coprod Ring a (FreeA Ring x) = CR (Multinomial x a)
934   inl a = CR (MN (singleton empty a))
935   inr (RingA (MN x)) = CR (MN (map initMN x))
936   eva f g (CR c) = evalMN f (g . pvar) c
937

```

Fig. 24. The coproduct of a commutative ring and a free commutative ring

```

940 dot [sta 1, sta 0, sta 2, dyn [x]] [dyn [x], dyn [y], dyn [z], dyn [x]]
941 ~> (sta 1 ⊗ dyn [x]) ⊕ (sta 0 ⊗ dyn [y]) ⊕ (sta 2 ⊗ dyn [z]) ⊕ (dyn [x] ⊗ dyn [x])
942 ~> 1x1 + 2z1 + x2 ~> x + 2*z + x*x
943

```

Fig. 25. Reducing partially-static commutative ring expressions

the trivial polynomial x , whose exponent and coefficient are both 1. The call `evalMN initMN f` in the definition of `pbind` interprets both the variables and the coefficients in an arbitrary ring, and combines the results using the ring operators. (See the supplementary material.)

It is not always possible to give a closed form for the coproduct of general algebras. However, we can still sometimes define the free extension, as is the case with commutative rings. The coproduct of a commutative ring A with a free commutative ring consists of multinomials with coefficients in A (Fig. 24).

`inl` maps an element b of A to the constant term b , while `inr` maps each term with coefficient n to the same term with coefficient $(\mathbb{1} \oplus \dots \oplus \mathbb{1})$ (n times, or using \ominus for negative n). As with the `pbind` operation for the free commutative ring, `eva` evaluates the multinomial using the ring operations for addition and multiplication:

$$\text{eva } f \text{ } g \text{ } (a + bx^2y) \quad \rightsquigarrow \quad f \text{ } a \oplus (f \text{ } b \otimes g \text{ } x \otimes g \text{ } x \otimes g \text{ } y)$$

A free commutative *semiring* (an algebra for the operations and axioms of a ring except those involving \ominus) is the same but with natural number coefficients instead of integers, and the free extension is defined analogously.

4.12 Example: linear algebra with partially-static rings

Linear algebra offers many opportunities for optimization via multi-stage specialization and numerical simplification such as the Fast Fourier Transform [Kiselyov et al. 2004], Gaussian elimination [Carette and Kiselyov 2011a], and matrix-vector multiplication [Aktemur et al. 2013]. The inner product illustrates the general principle: given a statically-known vector $s = [1, 0, 2]$ and a dynamic vector $d = [x, y, z]$, a naively-staged inner product function might generate the following code:

$$\llbracket (1 * x) + (0 * y) + (2 * z) \rrbracket$$

There are clear opportunities for improvement: the multiplication by one in the first summand can be omitted, and the multiplication by zero should annihilate the middle summand altogether.

As with power (§4.7), switching from hand-inserted staging annotations to *frex*'s high-level approach means that `dot` can be written as a polymorphic function with no mention of staging:

```

976 dot :: Ring r => [r] -> [r] -> r           sumr :: Ring r => [r] -> r
977 dot xs ys = sumr (zipWith (*) xs ys)       sumr = foldr (+) 0

```

Fig. 25 illustrates the behaviour of `dot` when the `Ring` constraint is instantiated to *frex*'s free extension instance: the call to `dot` constructs multinomials that residualize to multiplications and

additions. Returning to the example at the beginning of this section, *frex*'s ring simplifications lead to the following simpler code:

```
[[ x + (2 * z) ]]
```

§6 considers further examples involving linear algebra, and shows that *frex*'s simplifications lead to significant performance improvements.

4.13 Coproduct of distributive lattices

A distributive lattice is a commutative semiring $(A, 0, \oplus, \mathbb{1}, \otimes)$ with additional absorption rules:

$$a \otimes (a \oplus b) \equiv a \quad a \oplus (a \otimes b) \equiv a$$

For example, booleans form a distributive lattice with $\&\&$ as \otimes and $\|\|$ as \oplus .

As in the case of commutative rings, the coproduct of a distributive lattice A and a free distributive lattice on a set X consists of multinomials over X with coefficients in A . However, the fact that multiplication is idempotent ($a \otimes a \equiv a$) means that duplicates of variables within a term can be ignored, so the `MultiSet` of Fig. 24 is replaced with `Set`. In addition, the second absorption rule means that any term of the sum which is a multiple of another term with the same coefficient is redundant and can be dropped.

4.14 Example: `all` and `any` with partially-static distributive lattices

The examples so far all involve constructing and then residualizing partially-static values. It is also sometimes useful to compute with partially-static values before residualization.

The `all` function takes a predicate p and a list l , and returns true iff every element of l satisfies p . *Frex* supports defining a variant of `all` that operates on partially-static lists, with interleaved static and dynamic portions, and that produces partially-static booleans. Since a single element that does not satisfy p is enough to determine the result of `all`, the result may be static even where the input is partially unknown:

```
allPS even (sta [2, 4] ++ var [[x]] ++ sta [3])
  ~> sta (even 2) * sta (even 4) * dyn [[ all even x ]] * sta (even 3)
  ~> sta false
```

The `allPS` function operates on *frex*'s representations, turning a partially-static monoid (the input list) into a partially-static distributive lattice (the output boolean). The distributive lattice laws, used in *frex*'s free extension instance, reduce the expression to the value `false`, even though one element in the sequence is dynamic. The dual function `anyPS` can be defined similarly.

4.15 Coproduct of initial F-algebras

In addition to the familiar structures discussed above, the algebraic approach naturally subsumes earlier work on staged algebraic data types [Jones et al. 1993; Kaloper-Meršinjak and Yallop 2016; Sheard and Diatchki 2002] that is discussed further in §7.

An algebraic data type is the initial algebra for a presentation consisting of a functor F and no axioms. In other words it is constructed as the free F -algebra over the empty set.

For any algebraic structure, the coproduct of two free algebras is easy to calculate: it is given by the free algebra on the coproduct of their underlying sets. Fig. 26 shows this coproduct for the case of F -algebras.

The free extension of an algebraic data type $T := \text{FreeA } (\text{Alg } f) \text{ Empty}$ is thus of this form, where the type a is the empty type and b is `Code(T)`.

For example, the signature functor

```
IList X := 1 + Int × X
```

```

1030 instance Functor f  $\Rightarrow$  Coproduct (Alg f) (FreeA (Alg f) a) (FreeA (Alg f) b) where
1031 newtype Coprod (Alg f) (FreeA (Alg f) a) (FreeA (Alg f) b) =
1032   L (FreeA (Alg f) (Coproduct Set a b))
1033   inl x = L (fmap Inl x)
1034   inr y = L (fmap Inr y)
1035   eva g h (L e) = e `pbind` eva (g . pvar) (h . pvar)
1036

```

Fig. 26. The coproduct of two free F-algebras

has as initial algebra the type `IntList` of integer lists. The free extension of `IntList` is isomorphic to the free `IList`-algebra over `Code IntList`.

4.16 Example: partially-static algebraic datatypes

More generally, inductive algebraic datatypes can be seen as *initial algebras for a multi-sorted signature*, i.e. free algebras of operations without laws. These datatypes are useful in programs that perform staged computation. Lists with possibly-dynamic tails are a common example of a more general family of partially-static datatypes [Inoue 2014; Kaloper-Meršinjak and Yallop 2016; Sheard and Diatchki 2002].

The free extension for F-algebras in *frex* can be used to define a variant of `sum` that operates on partially-static lists by traversing the initial portion of a list, leaving traversal of the dynamic tail for later:

$$\text{cd} (\text{sum}_{\text{ps}} (1 :_s 2 :_s 3 :_s \text{dyn } \llbracket t \rrbracket)) \rightsquigarrow \llbracket 6 + \text{sum } t \rrbracket$$

4.17 Practical considerations: duplicating and discarding code

Since the aim of partially-static data structures is to avoid unnecessary computation in generated code, it is important to avoid duplicating or discarding expressions. In languages where the evaluation of an expression may have side effects, duplication and discarding are even more crucial to avoid. However, in some of our examples, such as `power` and `dot`, quoted expressions injected with `dyn` may appear either several times or not at all in the output of `cd`.

Fortunately, there are standard techniques available to address this issue. It is common in partial evaluators and multi-stage programming languages to convert programs into a form where every non-trivial expression is `let`-bound [Carette and Kiselyov 2011b; Kiselyov 2014; Yallop 2017] using a function (commonly named `genlet`) that accepts a dynamic expression `e`, inserts a `let`-binding for `e` at some higher point in the code, and returns the bound variable:

$$g (\text{genlet } \llbracket f \ x \rrbracket) \rightsquigarrow \llbracket \text{let } y = f \ x \text{ in } \dots \ $(g \llbracket y \rrbracket) \rrbracket$$

Automatic conversion to ANF form in the LMS multi-stage programming framework [Rompf 2016] serves a similar purpose

`let`-insertion combines straightforwardly with partially-static data; however, we have omitted it from the exposition for simplicity. The MetaOCaml implementation (Appendix A) uses `let`-insertion to avoid duplication.

5 UNIVERSALITY: FREE EXTENSION OF ALGEBRAS

§3 and the examples in §4 show that free extensions, i.e. coproducts with free algebras, provide a natural representation for partially-static data structures. Here we justify more formally why this representation is valid in terms of the universal property of free extensions. We first recall some basic universal algebra, which allows us to discuss classes of algebraic structures uniformly.

5.1 Rudimentary universal algebra

Like datatypes, descriptions of algebraic structures consist of an interface and a functional specification for this interface. The interface is given by an algebraic *signature* Σ : consisting of a set \mathcal{O}_Σ of *operation symbols* where each symbol is assigned a natural number called its *arity*. For example, monoids use the signature where $\mathcal{O}_{\text{mon}} := \{\mathbb{1}, \otimes\}$, $\mathbb{1}$ has arity 0, and \otimes has arity 2. Given a signature Σ , the functional specification is given by a set of *axioms*: equations between terms built from the operation symbols in Σ and according to their corresponding arities. For example, the monoid axioms Ax_{mon} are given in Fig. 6.

Put together, the description of an algebraic structure is called a *presentation* \mathcal{P} , given by a signature $\Sigma_{\mathcal{P}}$ and a set $\text{Ax}_{\mathcal{P}}$ of axioms over this signature. The example signature and axioms above form **mon** – the presentation of monoids.

An *algebra* for a presentation is a mathematical implementation of such specifications. Formally, given a presentation \mathcal{P} , a \mathcal{P} -algebra A is a pair $(|A|, -_A)$ consisting of a set $|A|$, called the *carrier* of the algebra, and, for each operation symbol f of arity n in $\Sigma_{\mathcal{P}}$, an n -ary function $f_A : |A|^n \rightarrow |A|$, such that all the axioms in $\text{Ax}_{\mathcal{P}}$ hold. For example, noting that a nullary function is a constant, a **mon**-algebra is a monoid.

Finally, given two \mathcal{P} -algebras A, B , a \mathcal{P} -homomorphism $h : A \rightarrow B$ is a function between the carriers $h : |A| \rightarrow |B|$ that respects the operations: for each operation symbol $f : n$ in $\Sigma_{\mathcal{P}}$, and for every n -tuple $\vec{a} = (a_1, \dots, a_n)$ of $|A|$ -elements, we have $h(f_A(a_1, \dots, a_n)) = f_B(h(a_1), \dots, h(a_n))$. For example, a **mon**-homomorphism $h : A \rightarrow B$ is a function that satisfies $h(\mathbb{1}_A) = \mathbb{1}_B$ and $h(x \otimes_A y) = h(x) \otimes_B h(y)$, i.e. the familiar notion of a monoid homomorphism.

For each presentation \mathcal{P} , the collection of \mathcal{P} -algebras and \mathcal{P} -homomorphisms between them forms a category $\mathcal{P}\text{-Alg}$, with the identities and composition given by the identity functions and the usual composition of functions. We have an evident functor $|-| : \mathcal{P}\text{-Alg} \rightarrow \text{Set}$ that forgets the algebra structure on objects and the homomorphism requirement on morphisms.

The forgetful functor $|-|$ always has a left adjoint $F_{\mathcal{P}} : \text{Set} \rightarrow \mathcal{P}\text{-Alg}$, the *free \mathcal{P} -algebra functor*. Concretely, its object map on a set X yields the *term algebra over X* : the set of $\Sigma_{\mathcal{P}}$ -terms with variables in X , quotiented by the deductive closure of Ax_{Σ} under the derivations of equational logic. For example, the free monoid over X is the set of finite sequences with X -elements, as described in §4.2. The unit of the adjunction, $\eta^{\mathcal{P}} : X \rightarrow |F_{\mathcal{P}}X|$ maps an element $x \in X$ to its equivalence class as a term. For **mon**, $\eta^{\text{mon}}(x)$ is the one-element sequence $[x]$. The adjunction itself assigns to every function $f : X \rightarrow |A|$ its homomorphic extension $\gg_{\mathcal{P}} f : F_{\mathcal{P}}X \rightarrow A$, which evaluates (the equivalence class of) a term in the algebra A , with X -variables substituted according to f . For example, taking A to be the integers with multiplication:

$$[x; y; z] \gg_{\text{mon}} \{x \mapsto 2, y \mapsto 3, z \mapsto 4\} = 2 \cdot 3 \cdot 4 = 24$$

The categories $\mathcal{P}\text{-Alg}$ have coproducts $A \oplus B$, and their concrete structure is given as follows. The carrier $|A \oplus B|$ is the $\Sigma_{\mathcal{P}}$ -term algebra over the disjoint union $|A| + |B|$ quotiented by the deductive closure of the axioms in \mathcal{P} , together with the equations of the form $f(t_1 a_1, \dots, t_1 a_n) \equiv t_1 f_A(a_1, \dots, a_n)$ for every f of arity n in $\Sigma_{\mathcal{P}}$, a_1, \dots, a_n in $|A|$, and analogous equations for B . The coproduct injection $t_1^{\oplus} : A \rightarrow A \oplus B$ maps a to the equivalence class of $t_1 a$, and similarly for B . For every pair of homomorphisms $h_1 : A \rightarrow C$, $h_2 : B \rightarrow C$, the unique cotupling homomorphism $[h_1, h_2] : A \oplus B \rightarrow C$ interprets a term over $|A| + |B|$ as the corresponding $|C|$ -element, once each variable $t_i x$ is substituted by $h_i(x)$. For example the coproduct of monoids A and B has as its carrier the set of sequences of alternating elements of $|A|$ and $|B|$.

A *free extension* of an algebra A by a set X is the coproduct of the algebra A with the free algebra over X , namely $\text{FreeExt}(A, X) := A \oplus F_{\mathcal{P}}X$. The coproduct injection $\iota : F_{\mathcal{P}}X \rightarrow \text{FreeExt}(A, X)$ corresponds under the adjunction to a function $\iota_X : X \rightarrow |\text{FreeExt}(A, X)|$. So combining the

1128 universal properties of coproducts and adjunctions, a free extension is characterised by an al-
 1129 gebra $\text{FreeExt}(A, X)$ together with a homomorphism $\iota_A : A \rightarrow \text{FreeExt}(A, X)$, and a function
 1130 $\iota_X : X \rightarrow |\text{FreeExt}(A, X)|$, such that for every other pair of a homomorphism $h : A \rightarrow C$ and
 1131 a function $e : X \rightarrow |C|$, there exists a unique homomorphism $\text{eva}(h, e) : \text{FreeExt}(A, X) \rightarrow C$
 1132 satisfying

$$1133 \quad \text{eva}(h, e) \circ \iota_A = h \quad | \text{eva}(h, e) | \circ \iota_X = e.$$

1134

1135 5.2 Conceptual justification

1136 Now suppose the algebra A stands for a static datatype, and the set X stands for a collection of
 1137 dynamically-known values. §3 sets out the minimum requirements we would want for the corre-
 1138 sponding partially-static datatype: an algebra $\text{ps}(A, X)$ together with inclusions

$$1139 \quad \text{sta} : A \rightarrow \text{ps}(A, X) \quad \text{dyn} : X \rightarrow \text{ps}(A, X)$$

1140

1141 such that sta is an algebra homomorphism, and a residualization map

$$1142 \quad \text{cd} : \text{ps}(A, X) \rightarrow X$$

1143

1144 satisfying $\text{cd} \circ \text{dyn} = \text{id}$ and $\text{cd} \circ \text{sta} = \text{lift}$ where lift is the function $A \rightarrow X$ lifting static values to
 1145 dynamic ones.

1146 Certainly the free extension $\text{FreeExt}(A, X)$ meets these requirements when X is the algebra
 1147 Code A , defining

$$1148 \quad \text{sta} := \iota_A \quad \text{dyn} := \iota_X \quad \text{cd} := \text{eva}(\text{lift}, \text{id})$$

1149

1150 as described in 3.4. With just the requirements above it is not the only possible choice. However,
 1151 there are at least two ways in which we could impose reasonable extra conditions on the partially-
 1152 static datatype which would only be satisfied by the free extension.

1153 Firstly, we could ask that in addition to residualization partially-static data should allow post-
 1154 processing. As $\text{ps}(A, X)$ is an algebra, we can consider homomorphisms from $\text{ps}(A, X)$ into other
 1155 algebras C . It is natural to expect that a homomorphism $h : A \rightarrow C$ and a function $e : X \rightarrow |C|$
 1156 should lift to a homomorphism $\text{eva}(h, e) : \text{ps}(A, X) \rightarrow C$ acting as expected on purely static and
 1157 dynamic values. The homomorphism should be unique for a minimal representation of the data.
 1158 This corresponds exactly to the characterization of $\text{FreeExt}(A, X)$ above.

1159 Alternatively, we could impose a uniformity condition on the collection of partially-static datatypes
 1160 for *all* algebras A . The datatype should store representations of static elements in a uniform way so
 1161 that homomorphisms of static data lift to homomorphisms of partially-static datatypes. So assume
 1162 that given a set X ,

- 1162 • partially-static datatypes with sta and dyn exist for every algebra A ,
- 1163 • for any instantiation $e : X \rightarrow A$ of variables in X as elements of A , there is a unique homo-
 1164 morphism $\text{eva}(\text{id}, e) : \text{ps}(A, X) \rightarrow A$ extending e and preserving static values,
- 1165 • for every pair of algebras A, B and homomorphism $h : A \rightarrow B$, there is a unique homomor-
 1166 phism $\text{ps}(h, X) : \text{ps}(A, X) \rightarrow \text{ps}(B, X)$ which acts as h on static values and leaves dynamic
 1167 values unchanged.

1168 Then an algebraic argument shows that $\text{ps}(A, X)$ together with sta , dyn and $\text{eva}(h, e) := \text{eva}(\text{id}, e) \circ$
 1169 $\text{ps}(h, X)$ has the universal property of the free extension $\text{FreeExt}(A, X)$.

1170

1171 6 PERFORMANCE EVALUATION

1172 The central contribution of this paper is a unification of various existing optimizations based
 1173 around partially-static data, although many of the structures and several of the examples in our
 1174 study are novel. Our focus up to this point has been on a rationalised interface to partially static
 1175 data, on the representations for particular algebraic structures, and on the simplifications that they
 1176

1176

1177 produce in code generated by multi-stage programs. It is reassuring to discover that the simplifi-
 1178 cations introduced by *frex* lead directly to improved performance over both the original unstaged
 1179 program and naively staged versions that make no use of partially-static data.

1180 We consider two representative examples: matrix multiplication, in Haskell, and `printf` in OCaml.

1181 The measurements in this section were taken on a Debian Linux system running the 4.9.0 kernel
 1182 on an AMD FX(tm)-8320 eight-core processor with 16GB memory. Haskell code was compiled with
 1183 GHC 8.0.2 using the `-O3` optimization flag, and OCaml code with BER MetaOCaml n104.

1184

1185 6.1 Matrix multiplication

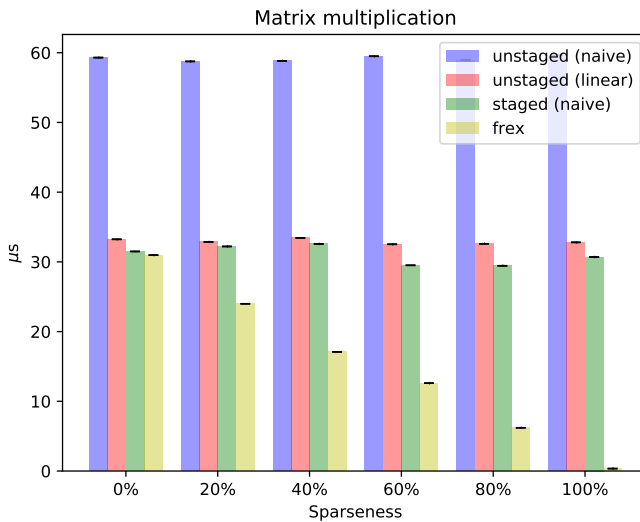
1186

1187

1188

1189

1190



1191

1192

1193

1194

1195

1196

1197

1198

1199

1200

1201

1202

1203

1204

1205

1206

1207 Fig. 27. Matrix multiplication improvements

1208

1209

1210 Fig. 27 shows the performance of four implementations of 10x10 matrix multiplication in Haskell.
 1211 The naive implementation is a one-line function based on a list-of-lists representation of matrices:

1212

```
1212 mmmul m n = [[dot a b | b <- transpose n] | a <- m]
```

1213

1214 The figures for *linear* represent the performance of a popular Haskell library of the same name,
 1215 based on a vector representation. There are two staged implementations, both of which are instan-
 1216 tiations of the one-line function above with appropriate instances. The naive staging unrolls the
 1217 loop, turning the list traversal into an arithmetic expression. The partially-static version takes a
 1218 static and a dynamic input vector, and converts both to lists of lists before passing them to `mmul`:

1219

```
1219 mmmul [ [sta s1, sta s2, ...]... ] [[dyn [[d!0!0]], dyn [[d!0!1]] ...]...]
```

1220

1221 Instantiating `mmul` with *frex*'s free extension instance results in automatic algebraic simplification.

1222 The graph shows measurements for various sparsities (i.e. for matrices with various proportions
 1223 of zero elements). As the sparsity of the matrix increases, the algebraic simplifications performed
 1224 by the partially static version significantly increase its advantage over naive staging.

1225

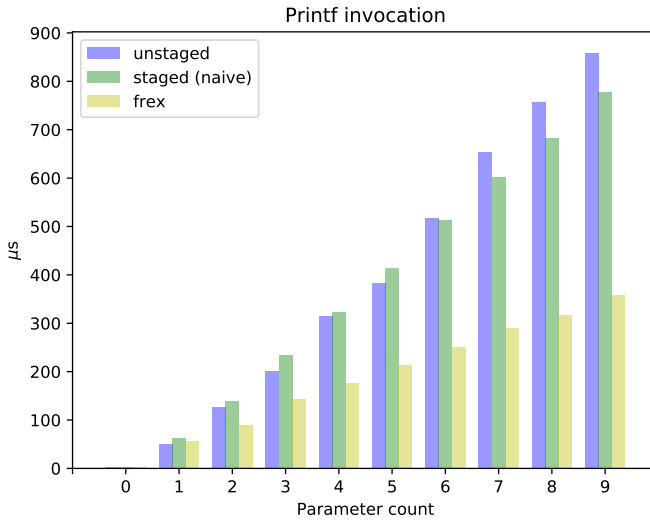


Fig. 28. Printf performance improvements

6.2 Printf

Fig. 28 shows the performance of three implementations of the `printf` function in MetaOCaml — unstaged, staged, and partially static (using *frex*'s free extension for monoids with improved residualization (§4.4)).

In each case the benchmark measures the time to run a call to `printf` with the given number of parameters and a format string that parenthesizes each parameter in the output:

```
printf ((lit "(" ++ str ++ lit ")") ++ (lit "(" ++ str ++ lit ")"))
```

Once again, the partially-static version gains an edge as the opportunities for algebraic simplification increase — in this case, as the number of subexpressions to reassociate and the number of adjacent static strings to merge grows.

7 RELATED WORK

We consider two main classes of related work — previous structured approaches to partially-static data and ad-hoc implementations of particular partially-static structures — before touching briefly on the use of partially-static data in supercompilers, optimizing compilers, and other tools.

Structured approaches to partially-static data. The existence of general schemes for partially-static structures without laws — i.e. datatypes — is well-known. The standard partial evaluation textbook [Jones et al. 1993] informally describes how to generalize partially-static list representations to arbitrary recursive types. Sheard and Diatchki [2002] describe a similar, but more concrete, scheme for deriving the staged versions of particular datatypes in the multi-stage programming language MetaML [Taha and Sheard 1997]. Kaloper-Meršinjak and Yallop [2016] turn Sheard's scheme into a generic programming framework based on the initial algebra view of datatypes.

The present work builds on these foundations, showing how partially-static datatypes arise as an instance of the general view of partially-static algebraic structures as free extensions of algebras.

There have been fewer previous attempts to construct a general view of partially-static structures with laws. Thiemann [2013] considers *partially static operations*, which incorporate algebraic

1275 laws, in a partial evaluation context. Thiemann’s [2013] vision of specialization that considers al-
1276 gebraic structure is an inspiration for this work. However, the implementation is quite different:
1277 Thiemann’s [2013] design operates via repeated rewriting, whereas the structures described in the
1278 present work are reduced using the evaluation mechanism of the host language. It appears unlikely
1279 that this approach to rewriting is suitable for direct use in multi-stage programming.

1280
1281 *Partially-static data in partial evaluation.* Partially-static data has been used in partial evaluation
1282 from early times. Mogensen [1988] introduced the concept in a study of partially-static lists, and
1283 several authors followed suit (e.g. [Hughes 1999; Jones et al. 1993]. Sheard and Diatchki [2002]
1284 notes that the term acquired the more specific meaning of static containers with dynamic elements.

1285 *Partially-static data in multi-stage programming.* Partially-static data is frequently employed
1286 in multi-stage programming, where eliminating unnecessary operations is an essential aspect of
1287 generating optimal code.

1288 The seminal *finally tagless* work by Carette et al. [Carette et al. 2009] uses a static-dynamic type
1289 – i.e. a record that holds a dynamic representation and, optionally, an additional static value of the
1290 same value – to improve generated code in a staged embedded lambda calculus. The implementa-
1291 tion additionally uses partially-static representations that implement ring simplification rules for
1292 zero addition and unit multiplication. Similar unit simplifications for vectors are implemented as
1293 smart constructors in Rompf et al.’s [2013] staging-based compiler optimization framework.

1294 Inspired by abstract interpretation, [Kiselyov et al. 2004] build a staged FFT implementation that
1295 uses partially-static representations that distinguish values from computations and that support
1296 simplification using various laws, including the distributive property, and trigonometric identities.

1297 Carette and Kiselyov [2005, 2011b] describe a modular decomposition of Gaussian Elimination
1298 that abstracts over staged and unstaged implementations of numeric signatures and other aspects
1299 of the algorithm. Partially-static data is used pervasively, primarily in the form of static-dynamic
1300 values. Combining the techniques in this paper with Carette and Kiselyov’s [2005] modular ap-
1301 proach by instantiating numeric signatures with free extensions is a promising avenue for future
1302 exploration.

1303 Yallop [2017] uses several partially-static structures in the staging of an implementation of
1304 the Scrap Your Boilerplate generic programming library, including a partially-static structure for
1305 monoids that reassociates subexpressions similarly to the free extension presented here.

1306 Drawing lessons from supercompilation, Inoue [2014] uses partially-static data that is updated
1307 to reflect equalities between values during the static exploration of the dynamic branches of a
1308 staged program. The primary partially-static structure is a list with a possibly-dynamic tail.

1309 8 CONCLUSION AND FURTHER WORK

1311 We have used free extensions of algebras as a functional specification of partially-static data, and
1312 described a high-level library, *frex*, that uses them to produce efficient staged code. Our approach
1313 combines the following attributes:

1314 **Extensible and modular:** The partially-static interface (*sta*, *dyn*, *cd*) operates uniformly over
1315 algebraic structures. Adding *Coproduct* and *Free* instances for an algebraic structure is sufficient
1316 to make the structure available for use in optimizations.

1317 Similarly, adding an instance of an algebraic class interface is sufficient to make the type available
1318 for use in optimizations. For example, since the standard library provides a *Monoid* instance for
1319 the *Maybe* type of optional values, *frex* will use the monoid laws to optimize programs involving
1320 *Maybe* even though *frex* itself makes no mention of *Maybe*.

1321 **Unifying:** Partially-static data is a well-known technique for binding-time improvement, and
1322 ad-hoc implementations of structures that implement some algebraic simplifications are found
1323

1324 throughout the literature (§7). The observation that partially-static data can be viewed as free ex-
 1325 tensions of algebras exposes and clarifies the structure underlying these ad-hoc implementations.
 1326 **Reusable:** This paper explores a universal view of partially-static data using a concrete library
 1327 (*frex*) in a particular language (Haskell). However, the underlying ideas can be reused in many
 1328 contexts: free extensions can be used to structure optimizers in other multistage languages, opti-
 1329 mizing compilers, partial evaluators, supercompilers, program generators, and so on.
 1330 **Practical:** The effectiveness of algebraic optimization using free extensions for partially-static
 1331 data is evident both from the simplified generated code, and from benchmarks (§6).

1332 In the future we would like to use free extensions of free theories to partially evaluate code
 1333 using effect handlers [Bauer and Pretnar 2015].
 1334

1335 ACKNOWLEDGMENTS

1336 Supported by the European Research Council grant ‘events causality and symmetry – the next-
 1337 generation semantics’, the Engineering and Physical Sciences Research Council grant EP/N007387/1
 1338 ‘Quantum computation as a programming language’, and a Balliol College Oxford Career Devel-
 1339 opment Fellowship. We would like to thank Jacques Carette, Chung-chieh Shan, Sam Staton, and
 1340 Gordon Plotkin for fruitful discussions and suggestions.
 1341

1342 REFERENCES

- 1343 Baris Aktemur, Yuki Yoshi Kameyama, Oleg Kiselyov, and Chung-chieh Shan. 2013. Shonan Challenge for Generative Pro-
 1344 gramming: Short Position Paper. In *Proceedings of the ACM SIGPLAN 2013 Workshop on Partial Evaluation and Program*
 1345 *Manipulation (PEPM '13)*. ACM, New York, NY, USA, 147–154. <https://doi.org/10.1145/2426890.2426917>
- 1346 Kenichi Asai. 2009. On typing delimited continuations: three new solutions to the printf problem. *Higher-Order and*
 1347 *Symbolic Computation* 22, 3 (01 Sep 2009), 275–291. <https://doi.org/10.1007/s10990-009-9049-5>
- 1348 Andrej Bauer and Matija Pretnar. 2015. Programming with algebraic effects and handlers. *J. Log. Algebr. Meth. Program.*
 1349 84, 1 (2015), 108–123.
- 1350 Jacques Carette and Oleg Kiselyov. 2005. Multi-stage Programming with Functors and Monads: Eliminating Abstraction
 1351 Overhead from Generic Code. In *Generative Programming and Component Engineering: 4th International Conference, GPCE 2005, Tallinn, Estonia, September 29 - October 1, 2005. Proceedings*, Robert Glück and Michael Lowry (Eds.). Springer
 1352 Berlin Heidelberg, Berlin, Heidelberg, 256–274. https://doi.org/10.1007/11561347_18
- 1353 Jacques Carette and Oleg Kiselyov. 2011a. Multi-stage programming with functors and monads: Eliminating abstraction
 1354 overhead from generic code. *Sci. Comput. Program.* 76, 5 (2011), 349–375. <https://doi.org/10.1016/j.scico.2008.09.008>
- 1355 Jacques Carette and Oleg Kiselyov. 2011b. Multi-stage Programming with Functors and Monads: Eliminating Abstraction
 1356 Overhead from Generic Code. *Sci. Comput. Program.* 76, 5 (May 2011), 349–375. <https://doi.org/10.1016/j.scico.2008.09.008>
- 1357 Jacques Carette, Oleg Kiselyov, and Chung-chieh Shan. 2009. Finally Tagless, Partially Evaluated: Tagless Staged Inter-
 1358 preters for Simpler Typed Languages. *J. Funct. Program.* 19, 5 (Sept. 2009), 509–543.
- 1359 Manuel M. T. Chakravarty, Gabriele Keller, Simon L. Peyton Jones, and Simon Marlow. 2005. Associated types with class.
 1360 In *Proceedings of the 32nd ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, POPL 2005, Long*
 1361 *Beach, California, USA, January 12-14, 2005*. 1–13. <https://doi.org/10.1145/1040305.1040306>
- 1362 Olivier Danvy. 1998. Functional Unparsing. *J. Funct. Program.* 8, 6 (Nov. 1998), 621–625. <https://doi.org/10.1017/S0956796898003104>
- 1363 John Hughes. 1999. *A Type Specialisation Tutorial*. Springer Berlin Heidelberg, Berlin, Heidelberg, 293–325. https://doi.org/10.1007/3-540-47018-2_12
- 1364 Jun Inoue. 2014. Supercompilation via staging. In *Fourth International Valentin Turchin Workshop on Metacomputation*.
- 1365 Neil D. Jones, Carsten K. Gomard, and Peter Sestoft. 1993. *Partial Evaluation and Automatic Program Generation*. Prentice-
 1366 Hall, Inc., Upper Saddle River, NJ, USA.
- 1367 David Kaloper-Meršinjak and Jeremy Yallop. 2016. Generic Partially-static Data (Extended Abstract). In *Proceedings of the*
 1368 *1st International Workshop on Type-Driven Development (TyDe 2016)*. ACM, New York, NY, USA, 39–40. <https://doi.org/10.1145/2976022.2976028>
- 1369 Oleg Kiselyov. 2014. The Design and Implementation of BER MetaOCaml. In *Functional and Logic Programming (Lecture*
 1370 *Notes in Computer Science)*, Michael Codish and Eijiro Sumii (Eds.), Vol. 8475. Springer International Publishing, 86–102.
- 1371 Oleg Kiselyov, Kedar N. Swadi, and Walid Taha. 2004. A Methodology for Generating Verified Combinatorial Circuits. In
 1372 *Proceedings of the 4th ACM International Conference on Embedded Software (EMSOFT '04)*. ACM, New York, NY, USA,

- 1373 249–258. <https://doi.org/10.1145/1017753.1017794>
- 1374 Torben Ægidius Mogensen. 1988. Partially Static Structures in a Self-Applicable Partial Evaluator. In *Partial Evaluation and*
- 1375 *Mixed Computation*, D. Bjørner, A.P. Ershov, and N.D. Jones (Eds.).
- 1376 Simon Peyton Jones. 2016. Template Haskell, 14 years on. Talk given at the International Summer School on Metaprogram-
- 1377 ming, Cambridge, UK. (August 2016). <https://www.cl.cam.ac.uk/events/metaprog2016/Template-Haskell-Aug16.pptx>.
- 1378 Tiark Rompf. 2016. *The Essence of Multi-stage Evaluation in LMS*. Springer International Publishing, Cham, 318–335.
- 1379 https://doi.org/10.1007/978-3-319-30936-1_17
- 1380 Tiark Rompf, Arvind K. Sujeeth, Nada Amin, Kevin J. Brown, Vojin Jovanovic, HyoukJoong Lee, Manohar Jonnalagedda,
- 1381 Kunle Olukotun, and Martin Odersky. 2013. Optimizing Data Structures in High-level Programs: New Directions for Ex-
- 1382 tensible Compilers Based on Staging. In *Proceedings of the 40th Annual ACM SIGPLAN-SIGACT Symposium on Principles*
- 1383 *of Programming Languages (POPL '13)*. ACM, New York, NY, USA, 497–510. <https://doi.org/10.1145/2429069.2429128>
- 1384 Tim Sheard and Iavor S. Diatchki. 2002. Staging Algebraic Datatypes. Unpublished manuscript. (2002).
- 1385 <http://web.cecs.pdx.edu/~sheard/papers/stagedData.ps>.
- 1386 Walid Taha. 2003. A Gentle Introduction to Multi-stage Programming.. In *Domain-Specific Program Generation (Lecture*
- 1387 *Notes in Computer Science)*, Christian Lengauer, Don S. Batory, Charles Consel, and Martin Odersky (Eds.), Vol. 3016.
- 1388 Springer, 30–50. https://doi.org/10.1007/978-3-540-25935-0_3
- 1389 Walid Taha and Tim Sheard. 1997. Multi-stage Programming with Explicit Annotations. In *Proceedings of the 1997 ACM*
- 1390 *SIGPLAN Symposium on Partial Evaluation and Semantics-based Program Manipulation (PEPM '97)*. ACM, New York, NY,
- 1391 USA, 203–217. <https://doi.org/10.1145/258993.259019>
- 1392 Peter Thiemann. 2013. Partially Static Operations (*PEPM '13*). ACM, New York, NY, USA, 75–76. <https://doi.org/10.1145/2426890.2426906>
- 1393 Jeremy Yallop. 2017. Staged Generic Programming. *Proc. ACM Program. Lang.* 1, ICFP, Article 29 (Aug. 2017), 29:1–
- 1394 29:29 pages. <https://doi.org/10.1145/3110273>
- 1395 Jeremy Yallop and Leo White. 2015. Modular Macros. (September 2015). OCaml Users and Developers Workshop 2015.
- 1396
- 1397
- 1398
- 1399
- 1400
- 1401
- 1402
- 1403
- 1404
- 1405
- 1406
- 1407
- 1408
- 1409
- 1410
- 1411
- 1412
- 1413
- 1414
- 1415
- 1416
- 1417
- 1418
- 1419
- 1420
- 1421

1422 A METAOCAML IMPLEMENTATION

1423 Several features of Haskell — type classes, abundant polymorphism, typed quotation, succinct
 1424 syntax — make it a convenient language for expounding the ideas in this paper. However, the
 1425 view of partially-static data structures as free extensions of algebras is independent of any lan-
 1426 guage, and transfers straightforwardly to a variety of settings. This appendix sketches a second
 1427 implementation in the multi-stage programming language BER MetaOCaml [Kiselyov 2014]. The
 1428 implementation is included as supplementary material.

1429 Haskell is a lazy, pure language with support for compile-time metaprogramming, in which pro-
 1430 grams are structured using implicitly-resolved instances of type classes. In contrast, BER MetaO-
 1431 Caml is an eager, impure language with support for run-time metaprogramming, in which pro-
 1432 grams are structured using explicitly combined modules. Despite these differences, both languages
 1433 provide the basic ingredients for implementing partially-static data structures as free extensions:
 1434 the ability to define datatypes and a distinction between static and dynamic values.

1435

1436 A.1 Parameterization

1437 The various generic components — definitions of coproducts, free extensions and so on, are defined
 1438 within a signature `Sig` containing a parameterized module `Ops` that builds an interface with a set
 1439 of operations for a type:

1440

```
1441     module type Sig = sig
1442       module Ops (X: TYPE) : sig module type OP end
1443       ...
1444     end
```

1445

1446 The partially-static representation for each algebraic structure instantiates `Ops` appropriately.
 1447 For example, here is the definition of the monoid interface:

1448

```
1449     module Monoid_ops (T: TYPE) = struct
1450       module type OP = sig
1451         type t = T.t
1452         val 1 : t
1453         val (⊗) : t → t → t
1454       end
1455     end
1456     module S : Sig with module Ops := Monoid_ops
```

1457

1458 A.2 Algebras

1459 Each additional element in `Sig` depends on `Ops`. An algebra is a pair of a type and a set of operations
 1460 for the type:

1461

```
1462     module type Algebra = sig
1463       module T : TYPE
1464       module Op : Ops(T).OP
1465     end
```

1466

1467 A.3 Coproducts

1468 A coproduct is an aggregation of three instances `T`, `A`, `B`, of an algebra, together with injections
 1469 from `A` and `B` into `T` and an evaluation function that maps values from `T` into any other algebra `C`:

1470

```

1471     module type COPRODUCT = sig
1472       module T : Algebra
1473       module A : Algebra
1474       module B : Algebra
1475
1476       val inl : A.T.t → T.T.t
1477       val inr : B.T.t → T.T.t
1478
1479       module Eva (C: Algebra) : sig
1480         val eva : (A.T.t → C.T.t) → (B.T.t → C.T.t) → T.T.t → C.T.t
1481       end
1482     end

```

The various coproduct implementations are expressed as parameterized modules that build instances of COPRODUCT, constraining A and B to match the parameters. Here is the signature of the parameterized module that builds a coproduct of monoids:

```

1487     module Coproduct_monoid (A : S.Algebra) (B : S.Algebra) :
1488       S.COPRODUCT with module A.T = A.T with module B.T = B.T

```

1490 A.4 Presentations and free algebras

1491 A presentation of an algebra is a definition of the free algebra with a function var for injection of
 1492 variables, and a bind function that performs substitution:
 1493

```

1494     module type PRES = sig
1495       module Free(X : Setoid) : sig
1496         module Alg : Algebra
1497         val var : X.t → Alg.T.t
1498       end
1499
1500       module Bind(X : Setoid) (C : Algebra) : sig
1501         val (>=>) : Free(X).Alg.T.t → (X.t → C.T.t) → C.T.t
1502       end
1503     end

```

1504 The partially-static structure for an algebra A aggregates A, a second implementation of the
 1505 algebra, the var and sta injections, and the cd and eva destructors. Since MetaOCaml is an impure
 1506 language, it is crucial to avoid duplicating or discarding expressions, and so dynamic injections
 1507 are from a distinct type var rather than arbitrary code values (Cf. §4.17).
 1508

```

1509     module type PS = sig
1510       module A : Algebra
1511       include Algebra
1512       val sta : A.T.t → T.t
1513       val var : A.T.t var → T.t
1514       val cd : T.t → A.T.t code
1515       module Eva (C : Algebra) : sig
1516         val eva : (A.T.t → C.T.t) → (A.T.t var → C.T.t) → T.t → C.T.t
1517       end
1518     end

```

1519

1520 A.5 Programming with partially-static data

1521 Although the set of abstraction mechanisms available in the language leads to a structure that is
1522 quite different from the Haskell implementation, the programming experience is rather similar.
1523 For example, the `sprintf` example is expressed identically:

```
1524         sprintf ((int ++ lit "a") ++ (lit "b" ++ int))
```

1526

1527

1528

1529

1530

1531

1532

1533

1534

1535

1536

1537

1538

1539

1540

1541

1542

1543

1544

1545

1546

1547

1548

1549

1550

1551

1552

1553

1554

1555

1556

1557

1558

1559

1560

1561

1562

1563

1564

1565

1566

1567

1568