Modular Transactional Memory

Matthew Brecknell  
matt@brecknell.net  
Queensland University of Technology, Australia

Paul Roe  
p.roe@qut.edu.au

Abstract

Software transactional memory has the potential to greatly simplify development of concurrent software, by supporting safe composition of concurrent shared-state abstractions. However, STM semantics are defined in terms of low-level reads and writes on individual memory locations, so implementations are unable to take advantage of the properties of user-defined abstractions. Consequently, the performance of transactions over some structures can be disappointing.

We present Modular Transactional Memory, our framework which allows programmers to extend STM with concurrency control algorithms tailored to the data structures they use in concurrent programs. We describe our implementation in Concurrent Haskell, and two example structures: a finite map which allows concurrent transactions to operate on disjoint sets of keys, and a non-deterministic channel which supports concurrent sources and sinks.

Our approach is based on previous work by others on boosted and open-nested transactions, with one significant development: transactions are given types which denote the concurrency control algorithms they employ. Typed transactions offer a higher level of assurance for programmers reusing transactional code, and allow more flexible abstract concurrency control.

Keywords: Software transactional memory, concurrency control, transactional boosting, open-nested transactions.

1 Introduction

Software transactional memory (Shavit & Touitou 1997) has emerged as a promising alternative to lock-based techniques for controlling concurrent access to shared data structures. STM provides an abstraction of physical memory, in which arbitrary sequences of memory operations can be executed as transactions, with the expected semantics of atomicity, consistency and isolation. STM saves the programmer from many of the pitfalls of lock-based programming, such as race conditions (failing to take the correct locks) and deadlocks (taking locks in the wrong order). Importantly, STM transactions are composable (Harris et al. 2005), whereas structures built using locks and condition variables typically are not. The latter has major benefits for the modularity and reusability of concurrent software.

Yet, if we care about performance, compositability can be illusory. The semantics of transactional memory are typically defined by whether read and write operations on individual memory variables commute. While this is sufficient to ensure correctness under composition, it is not always necessary. Abstractions already limit the admissible sequences of memory operations, often in ways that would allow more efficient concurrency control. STM implementations are unable to take advantage of this fact, so the performance of concurrent abstractions can be worse than one might intuitively expect.

Transactional boosting (Herlihy & Koskinen 2008) and open-nesting (Ni et al. 2007) are methodologies which address this issue, allowing programmers to improve concurrency without losing compositability. Boosting allows existing data structures to be used in transactional contexts, and open-nesting allows such data structures to be constructed using STM. Both allow programmers to replace low-level concurrency control on individual memory locations with higher-level abstract concurrency control.

Previous implementations providing boosted and open-nested transactions allow execution of arbitrary code in transactional contexts. This is necessary for abstract concurrency control, but it weakens the STM abstraction. The use of untyped transactions also limits the expressivity of abstract concurrency control, resulting in suboptimal implementations. In this paper, we develop a more disciplined approach, making the following contributions:

- We propose types for modular transactional concurrency control. Transactions are typed according to the concurrency control algorithms they employ, for improved assurance and flexibility of abstract concurrency control.
- We describe our implementation in Concurrent Haskell, combining transactional boosting and open-nested transactions with modular blocking and choice.
- Since our system uses a two-phase commit protocol internally, we provide a simple two-phase commit mechanism to users at no extra cost.

This paper is organised as follows. After some critical background (§2), we give an overview of our system (§3), some example applications (§4), and a discussion of future directions (§5). Finally, we draw conclusions (§6).

2 Background and related work

To make the work accessible to a general audience, we include the following background material. We introduce software transactional memory (§2.1), and show
how transactional boosting and open-nested transactions can be used to improve concurrency without losing composability. We also provide a brief overview of Concurrent Haskell (§2.2), since our system makes crucial use of certain features of the language.

2.1 Software transactional memory

STM is an abstraction of shared memory, in which arbitrary sequences of memory operations are executed as transactions, such that concurrent transactions appear to execute serially. Programmers can use familiar sequential reasoning within transactions, and only need to ensure that transactions take consistent states to consistent states.

For example, in a transaction to debit an account, we can be sure that concurrent updates will not modify the balance between the times we read and write the balance. In pseudo-code:

```plaintext
atomic debit(amount) {
    balance := balance - amount;
}
```

Although this appears like a critical section synchronised on the account, note that critical sections often do not compose correctly. For example, to atomically debit multiple accounts, it would be necessary to lock all accounts before any can be modified.

Transactions do compose naturally, and concurrent processes cannot observe intermediate states. For example, during the following transfer, no other process can observe a state where the amount is absent from both accounts:

```plaintext
atomic transfer(amount, debtor, creditor) {
    debtor.debit(amount);
    creditor.debit(-amount);
}
```

An atomic method may contain primitive operations (reads and writes) on transactional memory, as well as calls to pure functions and other atomic methods. An atomic method called from an ordinary method is executed as a top-level transaction, while an atomic method called from another atomic method is executed in the context of the calling transaction.

There is no need to specify locks; the runtime system determines which locks are required by observing the memory locations accessed during execution. No deadlock is possible, even when there are concurrent transfers in opposite directions; the system aborts and restarts transactions as necessary to ensure isolation and global progress.

2.1.1 Informal semantics

STM ensures opacity\(^1\) of concurrent execution traces (Guerraoui & Kapalla 2008). Opacity is similar to the serialisability criteria used to describe database transactions, but also requires that even transactions which abort only observe consistent states. Informally, an execution trace is opaque, or serialisable, if it is equivalent, in a certain sense, to some sequential execution of the same transactions.

Equivalence of execution traces can be defined in terms of conflicting operations. Concurrent operations conflict when they cannot be reordered without potentially changing the outcome of some subsequent operation. For simple read-write memory variables, a write conflicts with any other operation on the same variable, while all other pairs of operations are without conflict.

Executions are conflict-equivalent\(^2\) if they contain the same transactions, and the partial order among conflicting operations is the same. An implementation must ensure that if operations of one transaction conflict with those of another, then all the conflicting operations of one must come before the conflicting operations of the other.

2.1.2 The limits of composability

In a semantics defined by conflicts between read and write operations on simple memory variables, the notion of conflict is defined without reference to any semantics of variable contents. A write conflicts with any other operation on the same variable, irrespective of the value written. This provides a simple abstraction, but it can result in disappointing performance. Consider a simple integer map:

```plaintext
class map_stm {
    atomic int lookup(int key);
    atomic insert(int key, int value);
    atomic delete(int key);
}
```

In an abstract sense, operations on distinct keys can be reordered without affecting subsequent operations. However, an STM implementation of this structure is likely to generate conflicts between operations on distinct keys. This can easily be seen in a sorted linked list (figure 1), where deletion of the first node requires a write to the head, which necessarily conflicts with any other operation on the structure. Similar conflicts occur even in more sophisticated structures, such as hash tables and skip lists.

At the level of the operations themselves, such conflicts are unavoidable. The problem is that the impact of these conflicts on concurrency may leak far beyond the scope of those operations. For example:

```plaintext
atomic update(int key, map_stm src) {
    value := src.lookup(key);
    new_value := expensive_computation(key, value);
    src.insert(key, new_value);
}
```

Conflicts generated internally by map_stm operations mean that update operations on distinct keys might not be able to execute concurrently. That is, update operations are unnecessarily serialised, including any unrelated but expensive computations they contain.

To rectify this loss of composability, the notion of conflict must be generalised to include the operations of user-defined abstractions, for abstract concurrency control. The following sections describe two existing approaches.

2.1.3 Transactional boosting

Transactional boosting (Herlihy & Koskinen 2008) allows existing non-transactional data structures to be used in transactional contexts; for example, the containers in java.util.concurrent. Boosting is possible for structures where each operation has an inverse.

\(^1\)Most existing transactional memory implementations ensure opacity (Guerraoui et al. 2008), so opacity might be considered the de-facto standard semantics for STM.

\(^2\)Opacity actually requires a slightly stronger notion of equivalence, but we find the notion of conflict useful for the discussion of abstract concurrency control which follows.

Figure 1: Deletion from the front of a linked list
and is linearisable\(^3\), and where the commutativity of operations can be identified. Boosting makes use of compensating actions to undo the effects of operations when a transaction must abort, and abstract locks to ensure that non-commutative operations occur in a serialisable (opaque) order.

Our pseudo-code language is extended with new keywords, boosted and onAbort. A boosted method may be called from an atomic method, and contains non-transactional code to access the boosted structure and acquire abstract locks. Within a boosted method, onAbort may be used to record compensating actions to execute if the transaction is aborted. Abstract locks are acquired in a lock manager which holds locks until the acquiring transaction commits or aborts, and which aborts transactions as necessary to avoid deadlock.

Given an integer-indexed lock structure \texttt{lock_idx}, an existing linearisable integer map may be boosted as follows:

```haskell
class map_boosted {
  map_linear map;
  lock_idx lock;
  //...
  boosted delete(int key) {
    lock.shared(key);
    value = map.lookup(key);
    if (value is not null) {
      lock.exclusive(key);
      onAbort { map.insert(key, value); }
      map.delete(key);
    }
  }
}
```

Operations on the underlying structure are already linearisable, and so do not need any further synchronisation. However, locks are required to ensure abstract serialisability among boosted transactions. Since operations on distinct keys commute, one lock per key is sufficient for this purpose. Further, shared locks can be used for read-only operations.

The benefit of boosting this structure is that operations like \texttt{update} (§2.1.2) can reach a higher degree of concurrency. Using existing highly concurrent data structures, the concurrency between individual map operations can be significantly improved. More importantly, abstract locking may also allow other parts of transactions to run with improved concurrency. If those other parts are expensive, the improvement can be dramatic.

In our framework (§3), typed transactions provide the same benefits as transactional boosting, but with greater flexibility. For example, typed transactions allow individual compensating actions to be aggregated into more efficient bulk operations.

### 2.1.4 Open-nested transactions

Open nested transactions (Ni et al. 2007) are nested transactions which commit as soon as they complete, making their effects visible to other processes before their parent transactions commit. Together with abstract locking, and compensating actions in case the parent transaction aborts, open-nested transactions provide similar benefits to transactional boosting.

In full generality, open-nested transactions do not have clear semantics. For example, there is no reasonable answer to the question of how an implementation

\(^3\)Linearisability is a property of individual instances of an abstract data type. Informally, an object is linearisable if each operation appears to take effect "instantaneously at some point between its invocation and response" (Herlihy & Wing 1990). In this context, linearisability is a requirement on the data structure underlying a boosted transaction, while opacity and serialisability are global properties of the transactional system.
newIORef :: a -> IO (IORef a)
readIORef :: IORef a -> IO a
writeIORef :: IORef a -> a -> IO ()

Concurrent threads are started by forkIO:

forkIO :: IO () -> IO ThreadId

Synchronisation is provided by MVar, which acts as a simple one-place blocking channel:

newMVar :: IO (MVar a)
putMVar :: MVar a -> a -> IO ()
takeMVar :: MVar a -> IO a

IO also provides an exception handling mechanism.

### 2.2.1 Haskell STM

Concurrent Haskell includes an STM implementation (Harris et al. 2005). It is unique in its provision of a static guarantee that transactions only contain operations on memory, without requiring special language support.

Like IO actions, STM transactions are described by an abstract datatype, STM. As for IO, constructing an STM transaction does not cause its execution. A transaction can only be executed as an IO action, via a call to **atomically**:

atomically :: STM a -> IO a

Primitive transactions operate on transactional memory variables of type TVar:

newTVar :: a -> STM (TVar a)
readTVar :: TVar a -> STM a
writeTVar :: TVar a -> a -> STM ()

The !do!-notation is overloaded so transactions may be composed. For example, the **debit** function (§2.1) is written:

```haskell
debit :: Int -> TVar Int -> STM ()

debit amount account = do
  balance <- readTVar account
  writeTVar account (balance - amount)
```

Another primitive transaction provides a general and composable form of conditional blocking. A transaction which calls **retry** blocks until at least one variable previously read by the transaction has been modified by another thread, and then aborts and restarts. It can be used to wait for arbitrary conditions. For example, the following transaction waits for clear funds before making a debit:

```haskell
debitWhereClear :: Int -> TVar Int -> STM ()

debitWhereClear amount account = do
  balance <- readTVar account
  if amount <= balance
    then debit amount account
    else retry
```

Deterministic choice is provided by **orElse**, which takes two transactions, and runs the first, unless it would block with a call to **retry**, in which case it runs the second. For example, the following will debit the first of two accounts to have clear funds:

```haskell
debitFirst :: Int -> TVar Int -> TVar Int -> STM ()

debitFirst act1 act2 = debitWhereClear act1
  'orElse' debitWhereClear act2
```

Note one of the benefits of a pure type system: transactions and IO actions have distinct types, so the type rules ensure that transactions only contain operations which can be performed atomically. Our work aims to realise similar benefits for abstract concurrency control.

### 2.2.2 Type classes

Haskell type classes\(^\text{2}\) provide a disciplined approach to overloading. For example, following is a type class for equality comparison:

```haskell
class Eq a where

  (==) :: a -> a -> Bool

  True == True = True
  False == False = False
  _ == _ = False
```

This defines a new type class **Eq** with an overloaded operator method (**==**). The type variable (**a**) stands for each type which will instantiate the class, and must appear somewhere in the type of each method. A type instantiates a class with an **instance** declaration which provides a suitable implementation for each method, specific to that type. For example, **Bool** equality can be defined by **pattern-matching** on the constructors of the type:

```haskell
instance Eq Bool where

  True == True = True
  False == False = False
  _ == _ = False
```

The **Eq** type class allows functions which work for all types supporting equality. For example, to test if an item is contained in a **cons**-list, by pattern-matching on the constructors for the list type:

```haskell
elem :: Eq a => a -> [a] -> Bool

elem x (y:ys) = (x == y) || elem x ys

elem x [] = False
```

Here, the first part of the type signature (**Eq a =>**) indicates that the type variable (**a**) is **constrained**. That is, **elem** is only defined for types supporting equality. Square brackets around a type (**[a]**) indicate a list of that type; a single colon (**x :: xs**) constructs a cons-cell which prepends an item to an existing list; and empty square brackets (**[]**) construct an empty list.

Instances may be defined in terms of simpler instances. For example, equality for lists is defined in terms of equality of elements:

```haskell
instance Eq a => Eq [a] where

  (x:xs) == (y:ys) = (x == y) && (xs == ys)

  [] == [] = True
  _ == _ = False
```

This enables a form of **type-directed** programming.

### 3 Modular transactional memory

This section introduces our system. The design is inspired by Haskell STM, with support for abstract concurrency control. Transactions are attributed with descriptive types, to reconcile the need for arbitrary side-effects during abstract concurrency control, with the benefits of static guarantees provided by a pure type system. Typed transactions can also improve the expressivity and efficiency of abstract concurrency control.

#### 3.1 Typed transactions

We begin by describing transactions and how they can be composed, without reference to any particular transactional data structures. A new type parameter is added to the type of transactions, to represent the concurrency control algorithms used, giving the type (**Tx t a**) of transactions using algorithms described by type (**t**) and returning type (**a**).

In this context, an **algorithm** is the concurrency control (locking, compensating actions, etc.) for some abstract structure. For example, if **Hem** (defined elsewhere) is the type representing the concurrency control for ordinary transactional variables, then we can

\(^2\)Haskell type classes are almost entirely unlike the classes of object-oriented programming. In particular, they are not types.
form the type \((\mathit{Tx \mathit{Mem} a)\) of transactions over ordinary variables. Indeed, the latter corresponds to the type \((\mathit{STM})\) in the existing Haskell STM.

\(\mathit{Tx}\) is defined as an IO action, which executes in the context of a central transaction log of type \(\mathit{Log}\) (§3.3), and a local transaction log of type \(\mathit{t}\):

\[
\text{data Result } a = \text{Abort} \mid \text{Retry} \mid \text{Result } a
\]

\[
\text{newtype } \mathit{Tx} \mathit{t} \mathit{a} = \mathit{Tx} \{ \text{runTx} :: \mathit{Log} \rightarrow \text{t} \rightarrow \mathit{IO} \; \text{(Result } a) \}
\]

The \(\text{Result}\) type is used internally to distinguish between transactions which abort and restart due to a conflict, those which call \(\text{retry}\), and those which return a result normally.

Thus, if concurrency control algorithms are types, then their values are \emph{local transaction logs}, specialised to process transactions over the corresponding data structures. For composition of transactions, a transaction log must provide certain operations, including initialisation, nesting and finalisation. These requirements are represented as a type-class:

\[
\text{class } \mathit{TLog} \mathit{t} \text{ where}
\]

\[
\begin{align*}
\text{run} &:: (\mathit{t} \rightarrow \mathit{Undo} \rightarrow \mathit{Commit} \rightarrow \mathit{IO} \; \mathit{a}) \rightarrow \mathit{IO} \; \mathit{a} \\
\text{nest} &:: \mathit{t} \rightarrow (\mathit{Undo} \rightarrow \mathit{IO} \; \mathit{a}) \rightarrow \mathit{IO} \; \mathit{a}
\end{align*}
\]

\(\mathit{TLog}\) methods are called by the framework, which passes an IO action as a callback. A method implementation should execute the callback, passing the required parameters back to the framework. Typically, method implementations will also wrap the callback with resource management and exception handlers.

The type \(\mathit{Commit}\) describes the actions to perform when a transaction commits, and will be explained later (§3.4). The type \(\mathit{Undo}\) describes an action which reverses the effects of a transaction when it aborts, and is a simple type alias:

\[
\text{type } \mathit{Undo} = \mathit{IO} \; ()
\]

The framework calls \(\text{run}\) to begin execution of a transaction. The implementation of \(\text{run}\) must initialise a new transaction log (type \(\mathit{t}\)), as well as global abort and commit actions (types \(\mathit{Undo}\) and \(\mathit{Commit}\)) to pass back to the body. The framework then calls \(\text{nest}\) to obtain a local undo action for every call to \(\text{catch}\) and \(\text{orElse}\).

An instance of \(\mathit{TLog}\) is sufficient to implement all of the core combinators of Haskell STM, including overloading the do-notation, and these functions:

\[
\begin{align*}
\text{catch} &:: (\mathit{TLog} \mathit{t}, \mathit{Exception } \mathit{e}) \rightarrow \mathit{Tx} \mathit{t} \mathit{a} \rightarrow (\mathit{e} \rightarrow \mathit{Tx} \mathit{t} \mathit{a}) \rightarrow \mathit{Tx} \mathit{t} \mathit{a} \\
\text{retry} &:: \mathit{TLog} \mathit{t} \rightarrow \mathit{Tx} \mathit{t} \mathit{a} \\
\text{orElse} &:: \mathit{TLog} \mathit{t} \rightarrow \mathit{Tx} \mathit{t} \mathit{a} \rightarrow \mathit{Tx} \mathit{t} \mathit{a} \rightarrow \mathit{Tx} \mathit{t} \mathit{a} \\
\text{atomically} &:: \mathit{TLog} \mathit{t} \rightarrow \mathit{Tx} \mathit{t} \mathit{a} \rightarrow \mathit{IO} \; \mathit{a}
\end{align*}
\]

For example, the following shows the implementation of \(\text{catch}\), which presents nested transactions as an exception-handling mechanism. Here, backslash (\(\backslash\)) introduces a lambda (anonymous function):

\[
\text{catch body handler } = \mathit{Tx} \{ \\lambda \mathit{t} \rightarrow \text{nest } \mathit{t} \; (\lambda \mathit{undo} \rightarrow \mathit{Control}.\mathit{Exception}.\text{catch} \; (\text{runTx body } \mathit{l} \; \mathit{t}) \; ((\mathit{e} \rightarrow \text{do} \{ \mathit{undo}; \text{runTx (handler } \mathit{e}) \}; \mathit{l} \; \mathit{t}) )\)
\]

Haskell programmers might be alarmed at the pervasive use of IO in the definitions of \(\mathit{TLog}\) and \(\mathit{Tx}\). Do we fail to take advantage of the benefits of a pure type system? We do not believe so, because every use of IO in the definition of a transactional data structure is labelled with the type \(\mathit{t}\) of the local transaction log. Thus, transaction types act as \emph{certificates of origin}, which can help a user to understand the nature of some transactional code, without the need to study all of its parts.

### 3.2 Combining transaction types

The real value of this framework is in its support for composing transactions over arbitrary \emph{combinations} of data structures, with abstract concurrency control. Since we can only compose transactions of the same local log type, we provide a type-level operator (\(\&:\&\)) to construct \emph{combined} local log types, and a mechanism to inject transactions into these combined types. For example, if \(\mathit{Map}\) is the type representing concurrency control for associative maps, then we can form the type \(\mathit{Tx} \; (\mathit{Mem} :\&: \mathit{Map} \; a)\) of transactions over ordinary variables and associative maps. The operator is implemented as a simple right-associative pair type, which may be right-nested to arbitrary depth:

\[
\begin{align*}
\text{data } l :\&: r & = 1 :\&: r \\
\text{infixr } 5 &:\&
\end{align*}
\]

It is straightforward to provide a \(\mathit{TLog}\) instance for a combined type, based on \(\mathit{TLog}\) instances for the component types:

\[
\begin{align*}
\text{instance } (\mathit{TLog} \; l, \mathit{TLog} \; r) \Rightarrow \mathit{TLog} \; (1 :\&: r)
\end{align*}
\]

It is also easy to implement the following explicit injections, which could be used to compose two transactions over different structures, in sequence:

\[
\begin{align*}
\text{injl} :: \mathit{Tx} \; l \; a \Rightarrow \mathit{Tx} \; (1 :\&: r) \; a \\
\text{injr} :: \mathit{Tx} \; r \; a \Rightarrow \mathit{Tx} \; (1 :\&: r) \; a
\end{align*}
\]

However, explicit injection functions are cumbersome to use, so we define an inclusion relation which automates injections. This is implemented as a multi-parameter type class (\(<:\)) with suitable instances for combined types, following Swierstra (2008):

\[
\begin{align*}
\text{class } t :< c \text{ where}
\end{align*}
\]

\[
\begin{align*}
\text{inject} &:: \mathit{Tx} \; t \; a \rightarrow \mathit{Tx} \; c \; a
\end{align*}
\]

For example, given transactions over \(\mathit{Mem}\) and \(\mathit{Map}\), with the following types:

\[
\begin{align*}
\text{example_mem_basic} &:: \mathit{Tx} \; \mathit{Mem} () \\
\text{example_map_basic} &:: \mathit{Tx} \; \mathit{Map} ()
\end{align*}
\]

We can \(\text{inject}\) into an unspecified, but constrained, combined transaction type, resulting in transactions with the following types:

\[
\begin{align*}
\text{example_mem} &:: (\mathit{Mem} :< c) \Rightarrow \mathit{Tx} \; c () \\
\text{example_mem} &:: \text{inject} \; \text{example_mem_basic} \\
\text{example_map} &:: (\mathit{Map} :< c) \Rightarrow \mathit{Tx} \; c () \\
\text{example_map} &:: \text{inject} \; \text{example_map_basic}
\end{align*}
\]

These can then be composed in sequence, resulting in a transaction with the following type:

\[
\begin{align*}
\text{example} &:: (\mathit{Mem} :< c, \mathit{Map} :< c) \Rightarrow \mathit{Tx} \; c () \\
\text{example} &:: \text{do} \; (\text{example_mem; example_map})
\end{align*}
\]

The type can be read: \(\text{example}\) is a transaction over any set of structure types containing \emph{at least} \(\mathit{Mem}\) and \(\mathit{Map}\). This \(\text{example}\) transaction could be further composed with other transactions, with new constraints added to the transaction type as necessary.

Finally, at the point of execution, the transaction must be given a concrete type, to complete the automatic injection:

\[
\begin{align*}
\text{example_exec} = \text{atomically} \; (\text{example :: } \mathit{Tx} \; (\mathit{Mem} :\&: \mathit{Map}) \; O)
\end{align*}
\]

By combining abstractions this way, we assume that the areas of memory accessed by those abstractions are non-overlapping. This is a reasonable assumption, provided abstractions are properly encapsulated. It is also consistent with the restriction on the use of open-nested transactions (§2.1.4).
3.3 Primitive transactions

So far, we have described transactions and how they compose in a very general sense, without saying anything about how they are implemented. In particular, we have not shown how to provide primitive transactions for any structure of interest.

At this point, we are forced to make some significant implementation choices. There are numerous approaches to STM implementation, with equally many trade-offs (Larus & Rajwar 2007). Many approaches suit certain workloads, but not others. Since we are still in the early stages of developing this framework, our choices have been determined primarily by the need for simplicity, and secondly by a bias toward long transactions and high-contention workloads, since we expect this to be the area where abstract concurrency control will provide the greatest benefit. We intend to revisit these choices as we continue development.

We have opted to implement transactions with visible readers. Whenever a transaction reads state from a data structure, it must register with that structure to receive notifications when updates by other transactions invalidate that observation. For serialisability, we abort any transaction which receives such a notification before it is ready to commit. Processor cache performance can be inhibited by visible readers, but we currently accept this for the sake of simplicity.

The system allows both deferred and direct updates, although support for direct updates is limited. In a deferred update, a transaction records its intention to update to its local transaction log, but the actual update is deferred until the transaction commits. Deferred updates only require locks to be acquired during the commit process, but read operations must consult the transaction log before the data structure in memory. During a direct update, a transaction immediately acquires a lock on the structure, records the previous state in its transaction log (in case the transaction aborts), and writes the update directly to the structure. Transitions using direct updates must take care to avoid deadlock, usually by aborting the transaction if a lock cannot be acquired within a short time limit. Currently, there is no support for dead-lock detection, lock pre-emption or cascading aborts for direct updates.

A primitive transaction for a structure is created using the primitive function, which has the following type:

\[
\text{primitive} :: \text{Cert } t \rightarrow (\text{Log} \rightarrow t \rightarrow \text{IO } \text{Result } a) \rightarrow \text{Tx } t \cdot a
\]

Cert is a type family used to restrict the construction of a primitive transaction to the module which defines the corresponding structure. Constructors for values of type \(\text{Cert } t\) should be private to the module which defines type \(t\). This parameter thus certifies that the calling module is authorised to define a transaction of the given type.

The second argument provides the implementation for the primitive transaction. It is an \(\text{IO}\) action with access to the central transaction log (type Log) and the local transaction log (type \(t\)) for the respective structure type, and which returns a Result type. This means that a primitive transaction may return an ordinary result, or alternatively, \text{Abort} or \text{Retry}. Returning \text{Abort} causes the transaction to abort, for example, if a lock for a direct-mode update could not be acquired. \text{Retry} indicates that the primitive transaction should behave like a call to retry.

The local transaction log (of type \(t\)) is defined entirely by the module providing that transaction type. Typically, it records updates, in either deferred or direct mode. Examples given later (§4) show how local transaction logs can improve the expressiveness of abstract concurrency control.

The central transaction log (type Log) is used to implement the visible reader protocol. A transaction which observes the state of a structure should register its Log with that structure. Concurrent transactions which update the structure must notify any registered transactions which are invalidated by those updates, by calling invalidate, which has type:

\[
\text{invalidate} :: \text{Log} \rightarrow \text{IO } ()
\]

To maximise the benefit of abstract concurrency control, the implementation of a transactional data structure should aim to send notifications only to transactions whose observations have in fact been invalidated, according to the abstract semantics of the structure.

For transactions over multiple structure types, it is important that invalidation is handled centrally, since an update to any previously observed structure could invalidate the transaction. The implementation of primitive checks that the transaction has not been invalidated after executing the body of each primitive transaction.

Conveniently, this mechanism also serves to implement conditional blocking. A transaction which calls retry simply waits for invalidation, and then aborts and restarts.

3.4 Commit actions

As a transaction executes, each local log accumulates updates in deferred or direct mode. When finished, it must either commit or abort. It can only commit if it can acquire suitable locks for deferred updates, and if it has not been invalidated by an update performed by another transaction.

The abort process is straightforward. Deferred-mode updates can simply be discarded. Direct-mode updates must be reverted, but this can be done safely, since locks have already been acquired. Locks must then be released, and reader registrations should be cleared.

The commit process is more complicated, and proceeds in the following phases:

1. Locks must be acquired for all deferred updates across all local transaction logs. If it is not possible to acquire all locks, then the transaction must abort without any updates being committed.
2. The transaction must be validated, with a check that the the transaction has not been invalidated by updates committed by another transaction. Otherwise, it must abort.
3. Only if the transaction has not aborted, then deferred updates must be written across all local transaction logs, and notifications sent to invalidated transactions.
4. In any case, locks must be released, and reader registrations cleared.

We refer to the first and third phases as the prepare and update phases, respectively. They are performed by local transaction logs, while the second (validation) phase is performed centrally.

The system collects actions to execute for the prepare and update phases from local transaction logs, via the run method of the TLog class (§3.1). The actions are described by the Commit data type:

\[
\text{data Commit} = \text{Commit} (\text{prepare} :: \text{IO } () \rightarrow \text{IO } (), \text{update} :: \text{IO } () )
\]
To begin the commit process, the framework constructs a chain of calls to \texttt{prepare} and \texttt{update} for all local logs involved in the transaction, such that each call to \texttt{prepare} is passed a callback consisting of any remaining calls to \texttt{prepare}, followed by validation and a sequence of \texttt{update} actions. Each \texttt{prepare} action should execute its callback \textit{only} if it successfully acquires all locks necessary to ensure that the corresponding \texttt{update} can execute correctly. When the callback returns, the transaction has either aborted or committed, so \texttt{prepare} may release locks it previously acquired.

### 3.5 User-level two-phase commit

With the internal use of two-phase commit, it is possible to provide a form of two-phase commit directly to the user, without any additional cost. We provide \texttt{twophase}, with the following type:

\[
\texttt{twophase} :: \texttt{TLog} \; t \Rightarrow \texttt{Tx} \; t \; (\texttt{IO} \; a) \Rightarrow \texttt{IO} \; a
\]

This is similar to \texttt{atomically}, but instead of taking a transaction returning an ordinary value, \texttt{twophase} takes a transaction returning an \texttt{IO} action. We refer to this as the \texttt{validate} action.

When executing a transaction, \texttt{twophase} performs the validate action between the prepare and update phases of the commit protocol. This guarantees that the transaction can commit, but there is still the option to abort before updates become visible to other transactions.

We define the operation of \texttt{twophase} as follows: if the validate action returns normally, then the transaction commits, and \texttt{twophase} returns the same result as the validate action; if the validate action throws an exception, then the transaction aborts, and \texttt{twophase} propagates the exception.

For example, the following requires confirmation from the user before debiting an account:

\[
\begin{align*}
debitConfirm \; \text{amount} \; \text{account} &= \text{twophase} \; (\text{do} \; \text{balance} <- \text{readTVar} \; \text{account} \; (\text{balance} = \text{amount}) \; \text{return} \; (\text{do} \; \text{confirm} \leftarrow \text{prompt} \; (\"\text{Balance} = \" \; \text{show} \; \text{balance}) \; \text{if} \; \text{not} \; \text{confirm} \; \text{then} \; \text{throw} \; \text{RejectDebit} \; \text{else} \; \text{return} \; () )))
\end{align*}
\]

Without \texttt{twophase}, it would not be possible to ensure that the balance shown to the user is the same balance to which the debit is applied.

### 3.6 Blocking open-nested transactions

Open-nested transactions require special treatment, if they are to be allowed to include calls to \texttt{retry}.

Specifically, if an open-nested transaction blocks with a call to \texttt{retry}, and the parent transaction is invalidated by an update committed by another transaction, then the open-nested transaction should return control to the parent, to allow it to restart. The converse is not necessary: if an open-nested transaction is directly invalidated, but its parent has not been invalidated, then only the open-nested transaction need restart.

Further, if an open-nested transaction calls \texttt{retry} when the parent transaction is in the first branch of the choice operator (or \texttt{orElse}), then the open-nested transaction should not block, but should immediately abort and return control to the parent, to allow it to execute the alternate \texttt{orElse} branch.

These requirements are satisfied by \texttt{open}, which has the following type:

\[
\begin{align*}
\texttt{open} :: \texttt{TLog} \; t \Rightarrow \texttt{Log} \Rightarrow \texttt{Tx} \; t \; a \Rightarrow \texttt{IO} \; (\texttt{Result} \; a)
\end{align*}
\]

Like \texttt{atomically}, \texttt{open} executes a transaction, but requires an additional parameter for the parent transaction \texttt{Log}. The implementation of \texttt{open} uses this to register with the parent, to receive its invalidation signals, and to determine whether the parent is in the first branch of \texttt{orElse}. An open-nested transaction will only block if the parent is not in the first branch of \texttt{orElse}.

A call to \texttt{open} returns a \texttt{Result} type. \texttt{Retry} is returned when the parent is in the first branch of \texttt{orElse}, and the open-nested transaction calls \texttt{retry}. \texttt{Abort} is returned when the parent transaction has been invalidated. This result should be returned to the parent transaction (via \texttt{primitive}), after performing appropriate abstract concurrency control.

### 4 Example structures

This section describes several example transactional structures. First, we develop the prototypical structure, the simple transactional variable, and then contrast this with a boosted associative map structure. Finally, we develop a non-deterministic channel using open-nested transactions. Readers should bear in mind that the combining operators provided by the framework (§3.2) allow transactions over these structures to be composed together.

#### 4.1 Simple transactional variables

In this framework, simple transactional variables are defined as any other structure. We define the structure itself, its local transaction log, and its primitive transactions. Using deferred-mode updates, the \texttt{TVar} structure may be defined as follows:

\[
\begin{align*}
\texttt{data TVar} \; a &= \texttt{TVar} \; (tvar_data :: \texttt{IORef} \; a, \; tvar_lock :: \texttt{MVar} \; (), \; tvar_readers :: \texttt{CList} \; \texttt{Log})
\end{align*}
\]

Recalling that \texttt{IORef} is a mutable variable, and \texttt{MVar} provides synchronisation, the purpose of the first two fields should be clear. The third registers transactions which have observed the value of the \texttt{TVar}, for notification of updates. \texttt{CList} is a list type which supports concurrent insertion and deletion, and is necessary because many transactions may attempt concurrent reads from a \texttt{TVar}. The local transaction log can then be defined as follows:

\[
\begin{align*}
\texttt{data Write} &= \texttt{forall} \; a. \; \texttt{Write} \; (\texttt{Tvar} \; a) \; a
\end{align*}
\]

\[
\begin{align*}
\texttt{data Mem} &= \texttt{Mem} \; (\texttt{mem_write} :: \texttt{IORef} \; (\texttt{Write}), \; \texttt{mem_read} :: \texttt{IORef} \; [\texttt{CNode} \; \texttt{Log}])
\end{align*}
\]

The \texttt{mem_write} field records write operations in deferred mode, and is used during subsequent read operations, as well as the commit process. Each record includes the \texttt{TVar} written to, and the value written. The \texttt{mem_read} field records reader registrations for all variables read during the transaction, and is used to clear those registrations at the end of the transaction. A \texttt{CNode} is simply a handle to a \texttt{CList} node.

The local transaction log, \texttt{Mem}, must be made an instance of \texttt{TLog}, so we must provide implementations for \texttt{run} and \texttt{nest}. Of these, \texttt{nest} is the most interesting:

\[
\begin{align*}
\texttt{nest mem body} &= \texttt{do}
\texttt{saved_writes} &\leftarrow \texttt{readIORef} \; (\texttt{mem_write mem})
\texttt{body} &\leftarrow \texttt{IORef} \; (\texttt{mem_write mem}) \; \texttt{(saved_writes)}
\end{align*}
\]
For a nested transaction, the framework calls \texttt{nest} with the local transaction log (\texttt{mem}) and the body of the nested transaction. \texttt{first} saves a copy of the write log, and then calls the body, passing an undo action which may be used to restore the saved write log in the event that the nested transaction aborts. Recall that constructing an IO action does not cause it to be executed immediately. If needed, the framework will execute the undo action at the appropriate time; otherwise it will be discarded.

Implementations of primitive transactions must certify their authority to create a transaction of the required type. A private type definition, declared as a \texttt{Cert} instance, serves this purpose:

```haskell
data MemCert = MC

type instance Cert Mem = MemCert
```

The \texttt{readTVar} primitive may now be defined:

```haskell
readTVar :: TVar k a -> Tx Mem a
readTVar var = primitive MC (Vlog t -> do
  look_aside <- write_log_lookup var t
  case look_aside of
    Just value -> return (Result value)
    Nothing -> readDirect log t var)
```

In case the transaction has previously performed an uncommitted write to the same variable, a read operation first examines the write log. If the \texttt{TVar} is not found, \texttt{readDirect} is called to insert the transaction log into the \texttt{tvar_readers} field of the \texttt{TVar}, add the registration to the \texttt{mem_read} field of the local transaction log, and return the current value of the variable. To ensure reads are properly sequenced with respect to writes in other transactions, \texttt{readDirect} takes a lock on the \texttt{TVar}.

The \texttt{writeTVar} primitive simply adds a \texttt{Write} entry to the \texttt{mem_write} field of the local transaction log, or creates an existing entry for the same \texttt{TVar}. Then, at commit time, the \texttt{prepare} action attempts to acquire locks for all \texttt{TVar}s in the write log:

```haskell
prepare callback = do
  write_log <- readIORef (mem_write t)
  foldr tryWithLock callback write_log
```

Here, \texttt{foldr} reduces the write log with \texttt{tryWithLock}, which takes a \texttt{Write} entry and a callback, and executes the callback only if it acquires a lock on the corresponding \texttt{TVar}. If \texttt{prepare} succeeds, \texttt{update} writes each new value in the write log to the corresponding \texttt{TVar}, and invalidates any transactions which have registered reads on each \texttt{TVar}. The abort action is empty, but the \texttt{run} method clears reader registrations which have accumulated in the \texttt{mem_read} field of the local transaction log, at the end of the transaction.

### 4.2 Finite map

The map structure presented here is the moral equivalent of the boosted map shown earlier (§2.1.3), since it consists of an underlying linearisable data structure wrapped with abstract concurrency control. However, there are substantial differences in our method: the typed local transaction log is accessible to read operations, and therefore supports deferred-mode updates; it also allows us to aggregate operations on each map structure, for a more efficient commit process.

In fact, this implementation is very similar to that of simple transactional variables, so we only show key differences. Unfortunately, there are no readily available concurrent map implementations for Haskell, so we use a purely functional map (shown as \texttt{M.Map} below) protected by a single lock. Although not concurrent, it is linearisable, and so still meets the requirements for boosting. Despite the absence of operation-level concurrency, abstract concurrency control is still valuable, since it may increase concurrency between other parts of transactions.

The \texttt{TMap} structure is defined as follows. The important difference is that \texttt{tmap_readers} is defined on a per-key basis, so that operations on distinct keys do not conflict.

```haskell
data TMap k v = TMap {
  tmap_data :: IORef (M.Map k v),
  tmap_lock :: MVar (),
  tmap_readers :: IORef (M.Map k (CList Log))
}
```

This complicates the process of clearing reader registrations, which must ensure that \texttt{tmap_readers} does not accumulate empty registration lists. The local transaction log therefore stores each reader registration with its associated key and \texttt{TMap}:

```haskell
data Write = forall k v. Write (TMap k v) (M.Map k (Maybe v))
```

```haskell
data Read = forall k v.
  Read k (TMap k v) (CNode Log)
```

Note that each \texttt{Write} entry aggregates all the updates for a particular map, with deletions represented by a null value.

**Primitive transactions, nesting and finalisation are similar to those for simple transactional variables, so we do not reproduce them here.**

### 4.3 Non-deterministic unordered channel

This section describes an application of open-nested transactions, including those which block (§3.6), as well as direct-mode updates.

More importantly, it also demonstrates the value of typed transactions to the \texttt{consumer} of transactions. To improve concurrency, we allow this structure to exhibit a higher level of non-determinism than is generally allowed by the semantics of closed word-based transactional memory. With this structure, different executions of the same transactions may yield different results, even if the serialisations are the same.

We believe that controlled non-determinism has a useful role in concurrent programming, because it can provide higher efficiency and concurrency. Nevertheless, it is a fundamental change in semantics, so transactions which may exhibit this behaviour should be clearly identified. Our framework achieves this by ensuring that every transaction which operates on this structure has a type which includes its local log type.

The \texttt{NDChan} structure, with local transaction log type \texttt{NDC}, is defined as an unordered, unbounded collection with operations \texttt{put} and \texttt{take}, where \texttt{put} inserts an item, and \texttt{take} either blocks or returns an item which it removes from the collection:

```haskell
put :: NDChan a -> a -> Tx NDC ()
take :: NDChan a -> Tx NDC a
```

The operation of \texttt{take} is non-deterministic, so it may choose any element from the collection, and may even block (\texttt{retry}) when the collection is not empty. We do want the structure to be useful, so we insist that \texttt{take} only blocks when there are \emph{concurrent} transactions performing \texttt{take}, or the collection is empty.

Note that non-deterministic behaviour cannot be localised, but may infect the whole transaction. For example, since \texttt{take} may \texttt{retry} instead of returning an item, any \texttt{orElse} containing a \texttt{take} becomes a non-deterministic choice.
The underlying structure, TChan, is built with simple transactional variables. This is just the multicast channel described by Harris et al. (2005), with the important parts reproduced here:

```haskell
type Link a = TVar (Node a)
data Node a = Empty | Full a (Link a)

data TChan a = TChan {
  take_end :: TVar (Link a),
  put_end :: TVar (Link a) 
}

putTChan :: TChan a -> a -> Tx Mem ()
putTChan chan item = do
  link <- readTVar (put_end chan)
  empty <- newTVar Empty
  writeTVar link (Full item empty)
  writeTVar (put_end chan) empty

takeTChan :: TChan a -> Tx Mem a
takeTChan chan = do
  link <- readTVar (take_end chan)
  node <- readTVar link
  case node of
    Empty -> retry
    Full item next -> do
      modifyIORef (take_end chan) (node:)
      writeTVar link (Empty next)
      return item

While putTChan and takeTChan only conflict with each other if the channel is empty, each will always conflict with itself. This limits the concurrency between transactional performances on the same operation on this structure, even if most of the work in those transactions is performed elsewhere. This problem is similar to the update operation described earlier (§2.1.2), and the solution is the same. While we cannot increase the concurrency between individual TChan operations, abstract concurrency control may allow other parts of transactions to run concurrently.

The non-deterministic specification of take affords a lot of freedom. We implement put in deferred-mode, and take as a direct-mode open-transactional. As abstract operations, put and take do not conflict with themselves, or each other, with one exception: a put conflicts with a take which has blocked. This is the only situation in which we need to register a read operation. We do need to record items extracted by take, so that they may be returned if the transaction aborts. NDChan therefore contains a TChan, and a list of reader registrations:

```haskell
data NDChan a = NDChan {
  ndchan :: TChan a,
  ndchan_read :: CList Log
}
```

The local transaction log (type NDC) records items extracted by take, deferred put operations, and reader registrations:

```haskell
data Buffer = forall a. Buffer (NDChan a) [a]
data NDC = NDC {
  ndc_take :: IORef [Buffer],
  ndc_put :: IORef [Buffer],
  ndc_read :: IORef [CNode Log]
}
```

Here, a Buffer is an association between a channel and a list of items of the same type.

The put primitive inserts a buffer to the ndc_put field of the local transaction log, or augments an existing buffer if the NDChan is already present.

The take primitive first searches the ndc_put field of the local transaction log. If take finds a buffer for the required NDChan, it removes and returns an item from that buffer. Otherwise, take executes an open-transactional to extract an item from the underlying TChan, and buffers the item in the ndc_take field of the local transaction log:

```haskell
take :: NDChan a -> Tx NDC a
take = primitive NDCert (Vlog t -> do
  look_aside <- extract_put chan (ndc_put t)
  case look_aside of
    Just item -> return (Result item)
    Nothing -> do
      result <- open log (takeTChan (ndcchan chan))
      case result of
        Abort -> return Abort
        Retry -> do
          node <- cinserter log (ndchan_read chan)
          modifyIORef (ndc_read t) (node:)
          return Retry
        Result item -> do
          buffer_take chan (ndc_take t) item
          return (Result item))
  return (Result item))
```

Careful readers will observe a race condition between open and the reader registration in the Retry branch, which may allow a put in another transaction to go unnoticed. Our current solution is to rerun open after registering the read, but we omit this for brevity. In a future revision, it may be possible to improve our implementation of open-nested transactions, such that this is not necessary.

The remaining work is done in the abort and commit actions. To abort a transaction, items buffered in ndc_take must be returned to their associated channels, and readers registered on those channels must be invalidated:

```haskell
unbuffer :: Buffer -> IO ()
unbuffer (Buffer chan items) = do
  atomically (mapM_ unbuffer take_buffers)
  invalidate_clist (ndchan_read chan)
  abort = do
    take_buffers <- readIORef (ndc_take t)
    mapM_ unbuffer take_buffers
```

Note, we use atomically rather than open, since the abort action must not fail.

For commit, the prepare action is empty; it simply calls the body it is passed. The update action is similar to the abort action, inserting items buffered in ndc_put to their associated channels, and invalidating registered readers.

Finally, nest performs similar local manipulations of ndc_put and ndc_take buffers.

This example demonstrates that we can not only raise the level of abstraction of concurrency control, but we can relax and redefine its semantics to achieve improved concurrency. A typed local transaction log provides the necessary flexibility, acting as a medium for primitive operations to communicate with each other.

5 Discussion

Our vision, like that of other work on abstract concurrency control for transactional memory, is for concurrent software which reflects the structure of the application domain. Programming with locks and condition variables risks catastrophic structural failure as a system grows and requirements change. In contrast, composable memory transactions provide a safe foundation for evolution of concurrent software, but concurrency may suffer as transactions become more complex. Adding abstract concurrency control allows hot-spots to be removed with local changes only, and without changing the overall semantics of the program. Typed transactions provide a particularly flexible approach to abstract concurrency control.

Our system is still in the early stages of development, and much remains to be done. So far, our focus has been raising the level of abstraction for transactional programming, without sacrificing concurrency. This focus will continue as we experiment
with new data structures and concurrent programming patterns, but must eventually shift to other concerns.

We are yet to examine performance. Our current implementation is high-level and naïve, so we do not yet expect good absolute performance, nor to compete with other current STM implementations. Nevertheless, we believe that typed local transaction logs should provide the means to generalise many of the techniques used in high-performance STM implementations, to support abstract concurrency control. Some of the techniques we are interested in exploring include dependence-aware transactional memory (Ramadan et al. 2009) and aggressive transactional boosting (Koskinen & Herlihy 2008). We would also like to improve support for lock-based approaches (Dragojević et al. 2009), and contention management (Guerraoui et al. 2005). It remains to be seen whether it is possible (let alone desirable) to support multiple approaches to concurrency control in the one framework.

For many structures, like the finite map example (§4.2), development is likely to be similar to that of simple transactional variables. A natural evolution of the framework would be to factor these common parts into a library implemented in low-level C. Cooperation with the run-time thread scheduler may be beneficial for some parts, such as lock management.

Our development and presentation has so far been informal. Given the pitfalls which plague concurrent software, it will be particularly important to formalise our approach. This is not just to verify the soundness of the framework, but to provide correctness criteria to guide the development of new transactional structures. The formalisations given by Herlihy & Koskinen (2008) and Agrawal et al. (2008) give us confidence that this should be possible for a broad class of data structures. Our semantics would also need to account for modular blocking (retry) and choice (orElse) in the presence of open-nested transactions, and should accommodate structures with relaxed semantics, such as the non-deterministic channel (§4.3).

Certain features of the Haskell language are crucial to our design: its purity, and its ability to embed imperative sublanguages as abstract datatypes. Could our approach work in other languages? We think so, but probably only with language extensions. Thus, while our approach may only be accessible to Haskell programmers for the time being, we hope that in the longer term, it can contribute to the development of future concurrent programming languages. In particular, we believe that our work adds weight to arguments in favour of type systems which capture side-effects, particularly mutation, in a compositional way.

6 Conclusion

We have demonstrated our approach to abstract concurrency control, which allows concurrent data structures to be composed safely, without the loss of concurrency which affects transactional memories based on simple transactional variables. Our primary contribution is that transactions should be typed according to the structures they manipulate. Typed transactions admit abstract concurrency control without completely abandoning static safety guarantees. Typed local transaction logs also provide more flexible, and potentially more efficient abstract concurrency control.

References


