

Parameterised Linearisability

Andrea Cerone¹, Alexey Gotsman¹, and Hongseok Yang²

¹ IMDEA Software Institute

² University of Oxford

Abstract Many concurrent libraries are parameterised, meaning that they implement generic algorithms that take another library as a parameter. In such cases, the standard way of stating the correctness of concurrent libraries via linearisability is inapplicable. We generalise linearisability to parameterised libraries and investigate subtle trade-offs between the assumptions that such libraries can make about their environment and the conditions that linearisability has to impose on different types of interactions with it. We prove that the resulting parameterised linearisability is closed under instantiating parameter libraries and composing several non-interacting libraries, and furthermore implies observational refinement. These results allow modularising the reasoning about concurrent programs using parameterised libraries and confirm the appropriateness of the proposed definitions. We illustrate the applicability of our results by proving the correctness of a parameterised library implementing flat combining.

1 Introduction

Concurrent libraries encapsulate high-performance concurrent algorithms and data structures behind a well-defined interface, providing a set of methods for clients to call. Many such libraries [6,7,13] are *parameterised*, meaning that they implement generic algorithms that take another library as a parameter and use it to implement more complex functionality (we give a concrete example in §2). Reasoning about the correctness of parameterised libraries is challenging, as it requires considering all possible libraries that they can take as parameters.

Correctness of concurrent libraries is usually stated using *linearisability* [8], which fixes a certain correspondence between the *concrete* library implementation and a (possibly simpler) *abstract* library, whose behaviour the concrete one is supposed to simulate. For example, a high-performance concurrent stack that allows multiple push and pop operations to access the data structure at the same time may be specified by an abstract library where each operation takes effect atomically. However, linearisability considers only *ground* libraries, where all of the library implementation is given, and is thus inapplicable to parameterised ones. In this paper we propose a notion of *parameterised linearisability* (§3 and §4) that lifts this limitation. The key idea is to take into account not only interactions of a library with its client, but also with its parameter library, with the two types of interactions being subject to different conditions.

A challenge we have to deal with while generalising linearisability in this way is that parameterised libraries are often correct only under some assumptions about the context in which they are used. Thus, a parameterised library may assume that the library it takes as a parameter is *encapsulated*, meaning that clients cannot call its methods directly. A parameterised library may also accept as a parameter only libraries satisfying certain properties. For this reason, we actually present three notions of parameterised

linearisability, appropriate for different situations: a general one, which does not make any assumptions about the client or the parameter library, a notion appropriate for the case when the parameter library is encapsulated, and *up-to linearisability*, which allows making assumptions about the parameter library. These notions differ in subtle ways: we find that there is a trade-off between the assumptions that parameterised libraries make about their environment and the conditions that a notion of linearisability has to impose on different types of interactions with it.

We prove that the proposed notions of parameterised linearisability are *contextual* (§5), i.e., closed under parameter instantiation. This includes the case when the parameter library is itself parameterised. On the other hand, when the parameter is an ordinary ground library, this result allows us to derive the classical linearisability of the instantiated library from our notion for the parameterised one. We also prove that parameterised linearisability is *compositional* (§5): if several non-interacting libraries are linearisable, so is their composition. Finally, we show that parameterised linearisability implies *observational refinement* (§6): the behaviours of any complete program using a concrete parameterised library can be reproduced if the program uses a corresponding abstract one instead. All these results allow modularising the reasoning about concurrent programs using parameterised libraries: contextuality and compositionality break the reasoning about complex parameterised libraries into that about individual libraries from which they are constructed; observational refinement then lifts this to complete programs, including clients. The properties of parameterised linearisability we establish also serve to confirm the appropriateness of the proposed definitions.

We illustrate the applicability of our results by proving the up-to linearisability of flat combining [6] (§4), a generic algorithm for converting hard-to-parallelise sequential data structures into concurrent ones.

Due to space constraints, we defer the proofs of most theorems to [1, §B].

2 Parameterised Libraries

We consider *parameterised libraries* (or simply libraries) L , which provide some *public methods* to their *clients*. The latter are multi-threaded programs that can call the methods in parallel. In §4 and §6 we introduce a particular syntax for libraries and clients; for now it suffices to treat them abstractly. Our libraries are called parameterised because we allow their method implementations to call *abstract methods*, whose implementation is left unspecified. Abstract methods are meant to be implemented by another library provided by L 's client, which we call the *parameter library* of L .

We identify methods by names from a set \mathcal{M} , ranged over by m , and threads by identifiers from a set \mathcal{T} , ranged over by t . For the sake of simplicity, we assume that methods take a single integer as a parameter and always return an integer. We annotate libraries with types as in $L : M \rightarrow M'$, where $M, M' \subseteq \mathcal{M}$ give the sets of abstract and public methods of L , respectively. If $M = \emptyset$ we call L a *ground library*. The sets M and M' do not have to be disjoint: methods in $M \cap M'$ may be called by L 's clients, but their implementation is inherited from the one given by the parameter library.

Example: Flat Combining. Flat combining [6] is a recent synchronisation paradigm, which can be viewed [14] as a parameterised library $\text{FC} : \{m_i\}_{i=1}^n \rightarrow \{\text{do_}m_i\}_{i=1}^n$ for a given set of methods $\{m_i\}_{i=1}^n$. In Figure 1 we show a pseudocode of its implementation, which simplifies the original one in ways orthogonal to our goals. FC takes a library,

whose methods m_i are meant to be executed sequentially, and efficiently turns it into a library with methods $\text{do_}m_i$ that can be called concurrently.

As usual, this is achieved by means of mutual exclusion, implemented using a lock, but in a way that is more sophisticated than just acquiring it before calling a method m_i . A thread executing $\text{do_}m_i$ first publishes the operation it would like to execute and its parameter in its entry of the requests array. It then spins, trying to acquire the global lock. Having acquired a lock, the thread becomes a *combiner*: it performs the operations requested by all threads, stored in requests, by calling methods m_i of the parameter library and writing the values returned into the retval field of the corresponding entries in requests. Each spinning thread periodically checks this field and stops if some other thread has performed the operation it requested (for simplicity, we assume that nil is a special value that is never returned by any method). This algorithm benefits from cache locality when the combiner executes several operations in sequence, and thus yields good performance even for hard-to-parallelise data structures, such as stacks and queues.

In this paper, we develop a framework for specifying and verifying parameterised concurrent libraries. For flat combining, our framework suggests using an *abstract* library FC^\sharp : $\{m_i\}_{i=1}^n \rightarrow \{\text{do_}m_i\}_{i=1}^n$ in Figure 2 as a specification for the *concrete* library in Figure 1. FC^\sharp specifies the expected behaviour of flat combining by using the naive mutual exclusion. Showing that the implementation satisfies this specification in our framework amounts to proving that it is related to FC^\sharp by *parameterised linearisability*, which we present next.

```
LOCK lock;
struct{op,param,retval} requests[NThread];

do_m_i(int z):
  requests[mytid()].op = i;
  requests[mytid()].param = z;
  requests[mytid()].retval = nil;
  do:
    if (lock.tryacquire()):
      for (t = 0; t < NThread; t++):
        if (requests[t].retval == nil):
          int j = requests[t].op;
          int w = requests[t].param;
          requests[t].retval = m_j(w);
        lock.release();
  while (requests[mytid()].retval == nil);
  return requests[mytid()].retval;
```

Figure 1. Flat combining: implementation FC.

```
LOCK lock;
do_m_i(int z):
  lock.acquire();
  int retval = m_i(z);
  lock.release();
  return retval;
```

Figure 2. Flat combining: specification FC^\sharp .

3 Histories and Parameterised Linearisability

Histories. Informally, for a concrete library (such as the one in Figure 1) to be correct with respect to an abstract one (such as the one in Figure 2), the two should interact with their environment—the client and the parameter library—in similar ways. In this paper, we assume that different libraries and their clients access disjoint portions of memory, and thus interactions between them are limited to passing parameters and return values at method calls and returns. This is a standard assumption [8], which we believe can be relaxed using existing techniques [5]; see §7 for discussion. We record interactions of a parameterised library $L : M \rightarrow M'$ with its environment using *histories* (Definition 1 below), which are certain sequences of *actions* of the form

$$\text{Act} ::= (t, \text{call? } m'(z)) \mid (t, \text{ret! } m'(z)) \mid (t, \text{call! } m(z)) \mid (t, \text{ret? } m(z)),$$

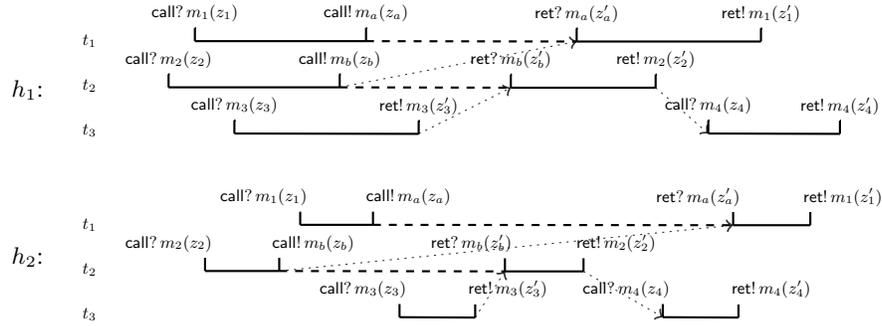


Figure 4. Illustration of histories and parameterised linearisability. A solid line represents a thread executing the code of the parameterised library, and a dashed one, the parameter library.

where $t \in \mathcal{T}$ is the thread performing the action, $m' \in M'$ or $m \in M$ is the method involved, and $z \in \mathbb{Z}$ is the method parameter or a return value.

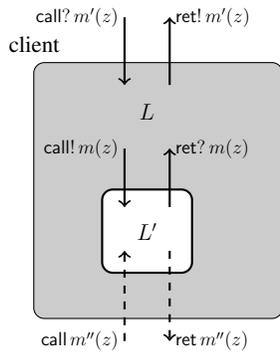


Figure 3. Interactions of a library L with its client and parameter library L' .

where $t \in \mathcal{T}$ is the thread performing the action, $m' \in M'$ or $m \in M$ is the method involved, and $z \in \mathbb{Z}$ is the method parameter or a return value. We illustrate the meaning of the actions in Figure 3: call? and ret! describe the client invoking public methods m' of the parameterised library L , and call! and ret? the library L invoking implementations of abstract methods m provided by a parameter library L' . We denote the sets of actions corresponding to interactions with these two entities by ClAct and AbsAct , respectively. In the spirit of the opponent-proponent distinction in game semantics [9,11], we annotate actions by $!$ or $?$ depending on whether the action was initiated by L or by an external entity, and we denote the corresponding sets of actions by Act! and Act? . We also use sets ActCall? , ActRet! , ActCall! and ActRet? with the expected meaning. Clients can also call methods $m'' \in M \cap M'$ directly, as represented by the dashed lines in the figure. Since such interactions do not involve the library L , we do not include them into Act . Histories are finite sequences of actions with invocations of abstract methods properly nested inside those of public ones.

DEFINITION 1 (Histories) A **history** $h : M \rightarrow M'$ is a finite sequence of actions such that for every t , the projection of h to t 's actions is a prefix of a sequence generated by the grammar SHist below, where $m \in M$ and $m' \in M'$:

$$\begin{aligned} \text{SHist} & ::= \varepsilon \mid (t, \text{call? } m'(z)) \text{IntSHist}(t, \text{ret! } m'(z')) \mid \text{SHist SHist} \\ \text{IntSHist} & ::= \varepsilon \mid (t, \text{call! } m(z)) (t, \text{ret? } m(z')) \mid \text{IntSHist IntSHist} \end{aligned}$$

We denote the set of histories by Hist . See Figure 4 for examples. In this paper, we focus on safety properties of libraries and thus let histories be finite. This assumption is also taken by the classical notion of linearisability [8] and can be relaxed as described in [4] (§7). For a history h and $A \subseteq \text{Act}$, we let $h|_A$ be the projection of h onto actions in A and we denote the i -th action in h by $h(i)$.

Parameterised Linearisability. We would like the notion of correctness of a concrete library $L : M \rightarrow M'$ with respect to an abstract one $L^\sharp : M \rightarrow M'$ to imply *observational refinement*. Informally, this property means that L^\sharp can be used to replace L in any program (consisting of a client, the library and an instantiation of the parameter library) while keeping its observable behaviours reproducible; a formal definition is given in §6. While this notion is intuitive, establishing it between two libraries directly is challenging because of the quantification over all possible programs they can be used by. We therefore set out to find a correctness criterion that compares the concrete and abstract libraries in isolation and thus avoids this quantification. For ground libraries, *linearisability* [8] formulates such a criterion by matching a history h_1 of L with a history h_2 of L^\sharp that yields the same client-observable behaviour. The following definition generalises it to parameterised libraries.

DEFINITION 2 (Parameterised linearisability: general case) *A history $h_1 : M \rightarrow M'$ is linearised by another one $h_2 : M \rightarrow M'$, written $h_1 \sqsubseteq h_2$, if there exists a permutation $\pi : \mathbb{N} \rightarrow \mathbb{N}$ such that*

$$\begin{aligned} \forall i. h_1(i) = h_2(\pi(i)) \wedge (\forall j. i < j \wedge ((\exists t. h_1(i) = (t, -) \wedge h_2(j) = (t, -)) \vee \\ (h_1(i) \in \text{Act!} \wedge h_1(j) \in \text{Act?})) \implies \pi(i) < \pi(j)). \end{aligned}$$

For sets of histories H_1, H_2 we let $H_1 \sqsubseteq H_2 \iff \forall h_1 \in H_1. \exists h_2 \in H_2. h_1 \sqsubseteq h_2$.

In §4 we show how to generate all histories of a library in a particular language and define linearisability on libraries by the \sqsubseteq relation on their sets of histories. For now we explain the above abstract definition. According to it, a history h_1 is linearised by a history h_2 when the latter is a permutation of the former preserving the order of actions within threads and the precedence relation between the actions initiated by the library and those initiated by its environment. As we explain below, we have $h_1 \sqsubseteq h_2$ for the histories h_1, h_2 in Figure 4. Hence, parameterised linearisability is able to match a history of a concurrent library with a simpler one where every contiguous block of library execution (e.g., the one between $(t_1, \text{call? } m_1(z_1))$ and $(t_1, \text{call! } m_a(z_a))$) is executed without interleaving with other such blocks. On the other hand, $h_2 \not\sqsubseteq h_1$, since $(t_1, \text{call! } m_a(z_a))$ precedes $(t_3, \text{call? } m_3(z_3))$ in h_2 , but not in h_1 .

When $h_1, h_2 : \emptyset \rightarrow M'$, i.e., these are histories of a ground library and thus contain only call? and ret! actions, Definition 2 coincides with a variant of the classical linearisability [8], which requires preserving the order between ret! and call? actions. For example, Definition 2 requires preserving the order between $(t_2, \text{ret! } m_2(z'_2))$ and $(t_3, \text{call? } m_4(z_4))$ in h_1 from Figure 4 (shown by a diagonal arrow). This requirement is needed for linearisability to imply observational refinement: informally, during the interval of time between $(t_2, \text{ret! } m_2(z'_2))$ and $(t_3, \text{call? } m_4(z_4))$ in an execution of a program producing h_1 , both threads t_2 and t_3 execute pieces of client code, which can communicate via the client memory. To preserve the behaviour of the client when replacing the concrete library in the program by an abstract one in observational refinement, this communication must not be affected, and, for this, the abstract library has to admit a history in which the order between the above actions is preserved.

When $h_1, h_2 : M \rightarrow M'$ correspond to a non-ground parameterised library, i.e., $M \neq \emptyset$, a similar situation arises with communication between the methods of the parameter library executing in different threads. For this reason, our generalisation of linearisability requires preserving the order between call! and ret? actions, such as

$(t_2, \text{call! } m_b(z_b))$ and $(t_1, \text{ret? } m_a(z'_a))$ in Figure 4; this requirement is dual to the one considered in classical linearisability. It is not enough, however. Definition 2 also requires preserving the order between call! and call? , as well as ret! and ret? actions, e.g., $(t_3, \text{ret! } m_3(z'_3))$ and $(t_2, \text{ret? } m_b(z'_b))$ in Figure 4. In the case when $M \cap M' \neq \emptyset$, this is also required to validate observational refinement. For example, during the interval of time between $(t_3, \text{ret! } m_3(z'_3))$ and $(t_2, \text{ret? } m_b(z'_b))$ in an execution producing h_1 , the client code in thread t_3 can call a method $m'_b \in M \cap M'$ of the parameter library (cf. the dashed arrows in Figure 3). The code of the method m'_b executed by t_3 can then communicate with that of the method m_b executed by t_2 , and to preserve this communication, we need to preserve the order between $(t_3, \text{ret! } m_3(z'_3))$ and $(t_2, \text{ret? } m_b(z'_b))$.

In §5 and §6 we prove that the above notion of linearisability indeed validates observational refinement. If the library $L : M \rightarrow M'$ producing the histories h_1, h_2 in Definition 2 is such that $M \cap M' = \emptyset$, then the client cannot directly call methods of its parameter library, and, as we show, parameterised linearisability can be weakened without invalidating observational refinement.

DEFINITION 3 (Parameterised linearisability: encapsulated case) *For $h_1, h_2 : M \rightarrow M'$ with $M \cap M' = \emptyset$ we let $h_1 \sqsubseteq_e h_2$ if there exists a permutation $\pi : \mathbb{N} \rightarrow \mathbb{N}$ such that*

$$\forall i. h_1(i) = h_2(\pi(i)) \wedge (\forall j. i < j \wedge ((\exists t. h_1(i) = (t, -) \wedge h_2(j) = (t, -)) \vee (h_1(i), h_1(j)) \in (\text{ActRet!} \times \text{ActCall?}) \cup (\text{ActCall!} \times \text{ActRet?}))) \implies \pi(i) < \pi(j)).$$

Since this definition does not take into account the order between $(t_1, \text{call! } m_a(z_a))$ and $(t_3, \text{call? } m_3(z_3))$ in h_2 from Figure 4, we have $h_2 \sqsubseteq_e h_1$ even though $h_2 \not\sqsubseteq h_1$.

Definitions 2 and 3 do not make any assumptions about the implementation of the parameter library. However, sometimes the correctness of a parameterised library can only be established under certain assumptions about the behaviour of its parameter. In particular, this is the case for the flat combining library from §2. In its implementation FC from Figure 1, a request by a thread t to execute a method m_i of the parameter library can be fulfilled by another thread t' who happens to act as a combiner; in contrast, the specification FC^\sharp in Figure 2 pretends that m_i is executed in the requesting thread. Thus, FC and FC^\sharp will behave differently if we supply as their parameter a library whose methods depend on the identifiers of executing threads (e.g., with m_i implemented as “return mytid(”)”). As a consequence, FC does not simulate FC^\sharp . On the other hand, this will be the case if we restrict ourselves to parameter libraries whose behaviour is independent of thread identifiers. The following version of parameterised linearisability allows us to use such assumptions, formulated as closure properties on histories of interactions between a parameterised library and its parameter. Given a history h , let \bar{h} be the history obtained by swapping ! and ? actions in h .

DEFINITION 4 (Up-to linearisability) *For $h_1, h_2 : M \rightarrow M'$ such that $M \cap M' = \emptyset$ and a binary relation \mathcal{R} on histories of type $\emptyset \rightarrow M$, we say that h_1 is **linearised by** h_2 **up to** \mathcal{R} , written $h_1 \sqsubseteq_{\mathcal{R}} h_2$, if $(h_1|_{\text{CIAct}}) \sqsubseteq (h_2|_{\text{CIAct}})$ and $(\overline{h_1|_{\text{AbsAct}}}) \mathcal{R} (\overline{h_2|_{\text{AbsAct}}})$.*

For flat combining, a suitable relation \mathcal{R}_t relates two histories if one can be obtained from the other by replacing thread identifiers of some pairs of a call and a corresponding (if any) return action. There are other useful choices of \mathcal{R} , such as equivalence up to commuting abstract method invocations [7].

So far we have defined our notions of linearisability abstractly, on sets of histories. We next introduce a language for parameterised libraries and show how to generate sets

of histories of a library in this language. This lets us lift the notion of linearisability to libraries and prove that FC in Figure 1 is indeed linearised up to \mathcal{R}_τ by FC^\sharp in Figure 2.

4 Lifting Linearisability to Libraries

Library Syntax. We use the following language to define libraries:

$$\begin{aligned} L &::= \langle \text{public} : B; \text{private} : B \rangle & B &::= \varepsilon \mid (m \leftarrow C); B \mid (\text{abstract } m); B \\ C &::= c \mid m() \mid C; C \mid \text{if}(E) \text{ then } C \text{ else } C \mid \text{while}(E) C \end{aligned}$$

A parameterised library L is a collection of methods, some implemented by commands C and others declared as abstract, meant to be implemented by a parameter library. Methods can be public or private, with only the former made available to clients. In §5 and §6 we extend the language to complete programs, consisting of a multithreaded client using a parameterised library with its parameter instantiated. In particular, we introduce private methods here to define parameter library instantiation in §5.

In commands, c ranges over *primitive commands* from a set PComm, and E over expressions, whose set we leave unspecified. The command $m()$ invokes the method m ; it does not mention its parameter or return value, since, as we explain below, these are passed via dedicated thread-local memory locations. We consider only well-formed libraries where a method is declared at most once and every method called is declared. We identify libraries up to the order of method declarations and α -renaming of private non-abstract methods. For a library $L = \langle \text{public} : B_{\text{pub}}; \text{private} : B_{\text{pvt}} \rangle$ we have $L : \text{Abs}(L) \rightarrow \text{Pub}(L)$, where $\text{Pub}(L)$ is the set of methods declared in B_{pub} , and $\text{Abs}(L)$ of those declared as abstract in B_{pub} or B_{pvt} .

Linearisability on Libraries and the Semantics Idea. We now show how to generate the set of histories $\llbracket L \rrbracket \in 2^{\text{Hist}}$ of a library L . Then we let a library L_1 be *linearised by* a library L_2 , written $L_1 \sqsubseteq L_2$, if $\llbracket L_1 \rrbracket \sqsubseteq \llbracket L_2 \rrbracket$; similarly for \sqsubseteq_e and $\sqsubseteq_{\mathcal{R}}$.

We actually generate all library *traces*, which, unlike histories, also record its internal actions. Let us extend the set of actions Act with elements of the forms (t, c) for $c \in \text{PComm}$, $(t, \text{call } m(z))$ and $(t, \text{ret } m(z))$, leading to a set TrAct. The latter two kinds of actions correspond to calls and returns between methods implemented inside the library. A *trace* τ is a finite sequence of elements in TrAct; we let $\text{Traces} = \text{TrAct}^*$.

The denotation $\llbracket L \rrbracket$ of a library $L : M \rightarrow M'$ includes the histories extracted from traces that L produces in any possible environment, i.e., assuming that client threads perform any sequences of calls to methods in M' with arbitrary parameter values and that abstract methods in M return arbitrary values. The definition of $\llbracket L \rrbracket$ follows the intuitive semantics of our programming language. An impatient reader can skip it on first reading and jump directly to Theorem 1 at the end of this section.

Heaps and Primitive Command Semantics. Let Locs be the set of memory locations. As we noted in §3, we impose a standard restriction that different libraries and their clients access different sets of memory locations, except the ones used for method parameter passing. Formally, we assume that each library L is associated with a set of its locations $\text{Locs}_L \subseteq \text{Locs}$. The state of L is thus given by a *heap* $\sigma \in \text{Locs}_L \rightarrow \mathbb{Z}$. We assume a special subset of locations $\{\text{arg}_t\}_{t \in \mathcal{T}}$ belonging to every Locs_L , which we use to pass parameters and return values for method invocations in thread t .

We assume that the execution of primitive commands and the evaluation of expressions are atomic. The semantics of a primitive command $c \in \text{PComm}$ used by a

Traces of commands	$\langle\langle C \rangle\rangle_t : (\mathcal{M} \times \mathcal{T} \rightarrow 2^{\text{Traces}}) \rightarrow 2^{\text{Traces}}$
$\langle\langle c \rangle\rangle_t \eta = \{(t, c)\}$	$\langle\langle C_1; C_2 \rangle\rangle_t \eta = \{\tau_1 \tau_2 \mid \tau_1 \in \langle\langle C_1 \rangle\rangle_t \eta \wedge \tau_2 \in \langle\langle C_2 \rangle\rangle_t \eta\}$
$\langle\langle \text{if}(E) \text{ then } C_1 \text{ else } C_2 \rangle\rangle_t \eta = (t, \text{assume}(E))(\langle\langle C_1 \rangle\rangle_t \eta) \cup (t, \text{assume}(!E))(\langle\langle C_2 \rangle\rangle_t \eta)$	
$\langle\langle \text{while}(E) C \rangle\rangle_t \eta = ((t, \text{assume}(E))(\langle\langle C \rangle\rangle_t \eta))^*(t, \text{assume}(!E))$	
$\langle\langle m(\cdot) \rangle\rangle_t \eta = \begin{cases} \{(t, \text{call! } m(z)) \tau (t, \text{ret? } m(z')) \mid \tau \in \eta(m, t) \wedge z, z' \in \mathbb{Z}\}, & \text{if } m \in M \\ \{(t, \text{call } m(z)) \tau (t, \text{ret } m(z')) \mid \tau \in \eta(m, t) \wedge z, z' \in \mathbb{Z}\}, & \text{otherwise} \end{cases}$	
Traces of library bodies	
$\mathcal{F} : (\mathcal{M} \times \mathcal{T} \rightarrow 2^{\text{Traces}}) \rightarrow (\mathcal{M} \times \mathcal{T} \rightarrow 2^{\text{Traces}})$	$\langle\langle B \rangle\rangle : \mathcal{M} \times \mathcal{T} \rightarrow 2^{\text{Traces}}$
$\langle\mathcal{F}(\eta)\rangle(m, t) = \begin{cases} \eta(m, t) \cup (\langle\langle C \rangle\rangle_t \eta), & \text{if } (m \Leftarrow C) \text{ appears in } L \\ \{\varepsilon\}, & \text{if } m \in M \\ \emptyset, & \text{otherwise} \end{cases}$	$\langle\langle B_{\text{pub}}; B_{\text{pvt}} \rangle\rangle = \text{lfp}(\mathcal{F})$
Traces of libraries	$\langle\langle L : M \rightarrow M' \rangle\rangle : 2^{\text{Traces}}$
$\langle\langle L \rangle\rangle = \text{prefix} \left(\bigcup_{k>0} \parallel_{t=1}^k \left(\bigcup_{\substack{z, z' \in \mathbb{Z} \\ m \in \tilde{M} \setminus M}} (t, \text{call? } m(z)) (\langle\langle B_{\text{pub}}; B_{\text{pvt}} \rangle\rangle(m, t)) (t, \text{ret! } m(z')) \right) \right)^*$	

Figure 5. Possible traces of a library $L = \langle \text{public} : B_{\text{pub}}; \text{private} : B_{\text{pvt}} \rangle : M \rightarrow M'$. Here $\parallel_{t=1}^k T_t$ denotes the set of all interleavings of traces from the sets T_1, \dots, T_k .

$$\begin{array}{ll}
\sigma \rightsquigarrow_{\text{call } m(z), t}^L \sigma' \text{ iff } \sigma' = \sigma, \sigma(\text{arg}_t) = z & \sigma \rightsquigarrow_{\text{ret } m(z), t}^L \sigma' \text{ iff } \sigma' = \sigma, \sigma(\text{arg}_t) = z \\
\sigma \rightsquigarrow_{\text{call? } m(z), t}^L \sigma' \text{ iff } \sigma' = \sigma[\text{arg}_t \mapsto z] & \sigma \rightsquigarrow_{\text{ret! } m(z), t}^L \sigma' \text{ iff } \sigma' = \sigma, \sigma(\text{arg}_t) = z \\
\sigma \rightsquigarrow_{\text{call! } m(z), t}^L \sigma' \text{ iff } \sigma' = \sigma, \sigma(\text{arg}_t) = z & \sigma \rightsquigarrow_{\text{ret? } m(z), t}^L \sigma' \text{ iff } \sigma' = \sigma[\text{arg}_t \mapsto z]
\end{array}$$

Figure 6. Transformers for calls and returns to, from and inside a library L .

library L is defined by a family of transformers $\{\rightsquigarrow_{c,t}^L\}_{t \in \mathcal{T}}$, where $\rightsquigarrow_{c,t}^L \subseteq (\text{Locs}_L \rightarrow \mathbb{Z}) \times (\text{Locs}_L \rightarrow \mathbb{Z})$ describes how c affects the state of the library. The fact that the transformers are defined on locations from Locs_L formalises our assumption that L accesses only these locations. We assume that the transformers satisfy some standard properties [15], deferred to [1, §A] due to space constraints. To define the semantics of expressions, we assume that for each E the set PComm contains a special command $\text{assume}(E)$, used only in defining the semantics, that allows the computation to proceed only if E is non-zero: $\sigma \rightsquigarrow_{\text{assume}(E), t}^L \sigma' \text{ iff } \sigma' = \sigma \text{ and } E \text{ is non-zero in } \sigma$.

Library Denotations. The set of traces of a library is generated in two stages. First, we generate a superset $\langle\langle L \rangle\rangle \subseteq 2^{\text{Traces}}$ of traces produced by L , defined in Figure 5. If we think of commands as control-flow graphs, these traces contain interleavings of all possible paths through the control-flow graphs of L 's methods, invoked in an arbitrary sequence. We then select those traces in $\langle\langle L \rangle\rangle$ that correspond to valid executions starting in a given heap using a predicate $\llbracket \tau \rrbracket_L : (\text{Locs}_L \rightarrow \mathbb{Z}) \rightarrow \{\text{true}, \text{false}\}$. We define $\llbracket \cdot \rrbracket_L$ by generalising \rightsquigarrow to calls and returns as shown in Figure 6 and letting

$$\llbracket \varepsilon \rrbracket_L \sigma = \text{true}; \quad \llbracket (t, a) \tau \rrbracket_L \sigma = \text{if } (\exists \sigma'. \sigma \rightsquigarrow_{a,t}^L \sigma' \wedge \llbracket \tau \rrbracket_L \sigma' = \text{true}) \text{ then true else false.}$$

Finally, we let the set of histories $\llbracket L \rrbracket$ of a library L consist of those obtained from traces representing its valid executions from a heap with all locations set to 0:

$$\llbracket L \rrbracket = \text{history}(\{\tau \in \langle\langle L \rangle\rangle \mid \llbracket \tau \rrbracket_L(\lambda x \in \text{Locs}_L. 0) = \text{true}\}),$$

where history projects to actions in Act .

THEOREM 1 (Correctness of flat combining) *For the libraries FC in Figure 1 and FC[#] in Figure 2 and the relation \mathcal{R}_t from §3 we have $FC \sqsubseteq_{\mathcal{R}_t} FC^\#$.*

PROOF SKETCH. Consider $h \in \llbracket FC \rrbracket$. In such a history, any invocation of an abstract method $(t, \text{call! } m_i(z_i)) (t, \text{ret? } m_i(z'_i))$ happens within the execution of the corresponding wrapper method $(t', \text{call? } \text{do_}m_i(z_i)) (t', \text{ret! } \text{do_}m_i(z'_i))$ (or just $(t', \text{call? } \text{do_}m_i(z_i))$ if the execution of the method is uncompleted in h), though not necessarily in the same thread. This correspondence is one-to-one, as different invocations of abstract methods correspond to different requests to perform them. Furthermore, abstract methods in h are executed sequentially. We then construct a history h' by replacing every abstract method call $(t, \text{call! } m_i(z_i)) (t, \text{ret? } m_i(z'_i))$ in $h|_{\text{AbsAct}}$ by

$$(t', \text{call? } \text{do_}m_i(z_i)) (t', \text{call! } m_i(z_i)) (t', \text{ret? } m_i(z'_i)) (t', \text{ret! } \text{do_}m_i(z'_i)),$$

where t' is the thread identifier of the corresponding wrapper method invocation (similarly for uncompleted invocations). It is easy to see that $(h|_{\text{AbsAct}}) \mathcal{R}_t (h'|_{\text{AbsAct}})$ and $h' \in \llbracket FC^\# \rrbracket$. Since the execution of an abstract method in h happens within the execution of the corresponding wrapper method, we also have $(h|_{\text{ClAct}}) \sqsubseteq (h'|_{\text{ClAct}})$. \square

5 Instantiating Library Parameters and Contextuality

We now define how library parameters are instantiated and show that our notions of linearisability are preserved under such instantiations. To this end, we introduce a partial operation \circ on libraries of §4: informally, for $L_1 : M \rightarrow M'$ and $L_2 : M' \rightarrow M''$ the library $L_2 \circ L_1 : M \rightarrow M''$ is obtained by instantiating abstract methods in L_2 with their implementations from L_1 . Note that L_1 can itself have abstract methods M , which are left unimplemented in $L_2 \circ L_1$. Since we assume that different libraries operate in disjoint address spaces, for \circ to be defined we require that the sets of locations of L_1 and L_2 be disjoint, with the exception of those used for method parameter passing. To avoid name clashes, we also require that public non-abstract methods of L_2 not be declared as abstract in L_1 (private non-abstract methods are not an issue, since we identify libraries up to their α -renaming); this also disallows recursion between L_2 and L_1 .

DEFINITION 5 (Parameter library instantiation) *Consider $L_1 : M \rightarrow M'$ and $L_2 : M' \rightarrow M''$ such that $(M'' \setminus M') \cap M = \emptyset$ and $\text{Locs}_{L_1} \cap \text{Locs}_{L_2} = \{\text{arg}_t\}_{t \in \mathcal{T}}$. Then $L_2 \circ L_1 : M \rightarrow M''$ is the library with $\text{Locs}_{L_2 \circ L_1} = \text{Locs}_{L_1} \cup \text{Locs}_{L_2}$ obtained by erasing the declarations for methods in M' from L_2 , reclassifying the methods from $M' \setminus M''$ in L_1 as private, and concatenating the method declarations of the resulting two libraries. We write $(L_2 \circ L_1)_\downarrow$ when $L_2 \circ L_1$ is defined.*

We now show that the notions of parameterised linearisability we proposed are **contextual**, i.e., closed under library instantiations. This property is useful in that it allows us to break the reasoning about a complex library into that about individual libraries from which it is constructed. As we show in §6, contextuality also helps us establish observational refinement.

THEOREM 2 (Contextuality of parameterised linearisability: general case) *For $L_1, L_2 : M \rightarrow M'$ such that $L_1 \sqsubseteq L_2$:*

- (i) $\forall L : M'' \rightarrow M. (L_1 \circ L)_\downarrow \wedge (L_2 \circ L)_\downarrow \implies L_1 \circ L \sqsubseteq L_2 \circ L.$
- (ii) $\forall L : M' \rightarrow M''. (L \circ L_1)_\downarrow \wedge (L \circ L_2)_\downarrow \implies L \circ L_1 \sqsubseteq L \circ L_2.$

THEOREM 3 (Contextuality of parameterised linearisability: encapsulated case) For $L_1, L_2 : M \rightarrow M'$ such that $M \cap M' = \emptyset$ and $L_1 \sqsubseteq_e L_2$:

- (i) $\forall L : M'' \rightarrow M. (L_1 \circ L) \downarrow \wedge (L_2 \circ L) \downarrow \implies L_1 \circ L \sqsubseteq_e L_2 \circ L.$
- (ii) $\forall L : M' \rightarrow M''. (L \circ L_1) \downarrow \wedge (L \circ L_2) \downarrow \implies L \circ L_1 \sqsubseteq_e L \circ L_2.$

The restriction on method names in Definition 5 ensures that the library compositions in Theorem 3 have no public abstract methods and can thus be compared by \sqsubseteq_e . Note that if L is ground, then so are $L_1 \circ L$ and $L_2 \circ L$. In this case, Theorems 2(i) and 3(i) allow us to establish classical linearisability from parameterised one.

Stating the contextuality of $\sqsubseteq_{\mathcal{R}}$ is more subtle. The relationship $L_1 \sqsubseteq_{\mathcal{R}} L_2$ allows the use of abstract methods by L_1 and L_2 to differ according to \mathcal{R} . As a consequence, for a non-ground parameter library L , their use by $L_1 \circ L$ and $L_2 \circ L$ may also differ according to another relation \mathcal{G} . We now introduce a property of L ensuring that a change in L 's interactions with its client according to \mathcal{R} (the *rely*) leads to a change in L 's interactions with its abstract methods according to \mathcal{G} (the *guarantee*).

DEFINITION 6 (Rely-guarantee closure) Let \mathcal{R}, \mathcal{G} be relations between histories of type $\emptyset \rightarrow M'$ and $\emptyset \rightarrow M$, respectively. A library $L : M \rightarrow M'$ is $(\frac{\mathcal{R}}{\mathcal{G}})$ -closed if for all $h \in \llbracket L \rrbracket$ and $h' : \emptyset \rightarrow M'$ we have

$$(h|_{\text{ClAct}}) \mathcal{R} h' \implies \exists h'' \in \llbracket L \rrbracket. (h''|_{\text{ClAct}} = h') \wedge \overline{(h|_{\text{AbsAct}})} \mathcal{G} \overline{(h''|_{\text{AbsAct}})}.$$

Due to space constraints, we state contextuality of $\sqsubseteq_{\mathcal{R}}$ only for the case in which library parameters do not have public abstract methods. A more general statement which relaxes this assumption is given in [1, §B].

THEOREM 4 (Contextuality of linearisability up to \mathcal{R}) For $L_1, L_2 : M \rightarrow M'$ such that $M \cap M' = \emptyset$ and a relation \mathcal{R} such that $L_1 \sqsubseteq_{\mathcal{R}} L_2$:

- (i) $\forall L : M'' \rightarrow M. \forall \mathcal{G}. M'' \cap M = \emptyset \wedge (L \text{ is } (\frac{\mathcal{R}}{\mathcal{G}})\text{-closed}) \wedge (L_1 \circ L) \downarrow \wedge (L_2 \circ L) \downarrow \implies L_1 \circ L \sqsubseteq_{\mathcal{G}} L_2 \circ L.$
- (ii) $\forall L : M' \rightarrow M''. (L \circ L_1) \downarrow \wedge (L \circ L_2) \downarrow \implies L \circ L_1 \sqsubseteq_{\mathcal{R}} L \circ L_2.$

When L in Theorem 4(i) is ground, \mathcal{G} becomes irrelevant. In this case we say that L is \mathcal{R} -closed if it is $(\frac{\mathcal{R}}{\{(\varepsilon, \varepsilon)\}})$ -closed. Hence, from Theorems 1 and 4(i) we get that for any \mathcal{R}_t -closed (§3) library L we have $\text{FC} \circ L \sqsubseteq \text{FC}^\sharp \circ L$: instantiating flat combining with a library insensitive to thread identifiers, e.g., a sequential stack or a queue, yields a concurrent library linearisable in the classical sense.

Given two libraries $L_1 : M_1 \rightarrow M'_1$ and $L_2 : M_2 \rightarrow M'_2$ that do not interact, i.e., $(M_1 \cup M'_1) \cap (M_2 \cup M'_2) = \emptyset$, we may wish to compose them by merging their method declarations into a library $L_1 \uplus L_2 : M_1 \uplus M_2 \rightarrow M'_1 \uplus M'_2$, as originally proposed in [8]. Our notions of linearisability are also closed under this composition.

THEOREM 5 (Compositionality of parameterised linearisability) For $L_1, L'_1 : M_1 \rightarrow M'_1$ and $L_2, L'_2 : M_2 \rightarrow M'_2$ such that $(M_1 \cup M'_1) \cap (M_2 \cup M'_2) = \emptyset$:

- (i) $L_1 \sqsubseteq L'_1 \wedge L_2 \sqsubseteq L'_2 \implies L_1 \uplus L_2 \sqsubseteq L'_1 \uplus L'_2.$
- (ii) $L_1 \sqsubseteq_e L'_1 \wedge L_2 \sqsubseteq_e L'_2 \implies L_1 \uplus L_2 \sqsubseteq_e L'_1 \uplus L'_2.$
- (iii) $\forall \mathcal{R}, \mathcal{G}. L_1 \sqsubseteq_{\mathcal{R}} L'_1 \wedge L_2 \sqsubseteq_{\mathcal{G}} L'_2 \implies L_1 \uplus L_2 \sqsubseteq_{\mathcal{R} \otimes \mathcal{G}} L'_1 \uplus L'_2$, where $\mathcal{R} \otimes \mathcal{G}$ relates histories if their projections to M_1 actions are related by \mathcal{R} and the projections to M_2 actions are related by \mathcal{G} .

6 Clients and Observational Refinement

A **program** P has the form $\text{let } L \text{ in } C_1 \parallel \dots \parallel C_n$, where $L : \emptyset \rightarrow M$ is a ground library and $C_1 \parallel \dots \parallel C_n$ is a client such that C_1, \dots, C_n call only methods in M , written $(C_1 \parallel \dots \parallel C_n) : M$. Using the contextuality results from §5, we now show that our notions of linearisability imply observational refinement for such programs.

The semantics of a program P is given by the set of its traces $\llbracket P \rrbracket \in 2^{\text{Traces}}$, which include actions (t, c) recording the execution of primitive commands c by client threads C_t and the library L , as well as $(t, \text{call } m(z))$ and $(t, \text{ret } m(z))$ actions corresponding to the former invoking methods of the latter. The semantics $\llbracket P \rrbracket$ is defined similarly to that of libraries in §4. In particular, we assume that client threads C_t access only locations in a set $\text{Locs}_{\text{client}}$ such that $\text{Locs}_{\text{client}} \cap \text{Locs}_L = \{\text{arg}_t\}_{t \in \mathcal{T}}$ for any L . Due to space constraints, we defer the definition of $\llbracket P \rrbracket$ to [1, §A]. We define the **observable behaviour** $\text{obs}(\tau)$ of a trace $\tau \in \llbracket P \rrbracket$ as its projection to client actions, i.e., those outside method invocations, and lift obs to sets of traces as expected.

DEFINITION 7 (Observational refinement) *For $L_1, L_2 : M \rightarrow M'$ we say that L_1 **observationally refines** L_2 , written $L_1 \sqsubseteq_{\text{obs}} L_2$, if for any ground library $L : \emptyset \rightarrow M$ and client $(C_1 \parallel \dots \parallel C_n) : M'$ we have*

$$\text{obs}(\llbracket \text{let } (L_1 \circ L) \text{ in } C_1 \parallel \dots \parallel C_n \rrbracket) \subseteq \text{obs}(\llbracket \text{let } (L_2 \circ L) \text{ in } C_1 \parallel \dots \parallel C_n \rrbracket).$$

*For a binary relation \mathcal{R} on histories we say that L_1 **observationally refines L_2 up to \mathcal{R}** , written $L_1 \sqsubseteq_{\text{obs}}^{\mathcal{R}} L_2$, if the above is true under the assumption that L is \mathcal{R} -closed.*

Thus, $L_1 \sqsubseteq_{\text{obs}} L_2$ means that L_1 can be replaced by L_2 in any program that uses it while keeping observable behaviours reproducible. This allows us to check a property of a program using L_1 (e.g., the flat combining implementation in Figure 1) by checking this property on a program with L_1 replaced by a possibly simpler L_2 (e.g., the flat combining specification in Figure 2). Using Theorems 2–4, we can show that our notions of linearisability validate observational refinement.

THEOREM 6 (Observational refinement) *For any libraries $L_1, L_2 : M \rightarrow M'$:*

- (i) $L_1 \sqsubseteq L_2 \implies L_1 \sqsubseteq_{\text{obs}} L_2$.
- (ii) $M \cap M' = \emptyset \wedge L_1 \sqsubseteq_e L_2 \implies L_1 \sqsubseteq_{\text{obs}} L_2$.
- (iii) $\forall \mathcal{R}. M \cap M' = \emptyset \wedge L_1 \sqsubseteq_{\mathcal{R}} L_2 \implies L_1 \sqsubseteq_{\text{obs}}^{\mathcal{R}} L_2$.

7 Related Work

Linearisability has recently been extended to handle liveness properties, ownership transfer and weak memory models [4,5,10]. Most of these extensions have exploited the connection between linearisability and observational refinement [2]. The same methodology is adopted in the present work, but for studying two previously unexplored topics: parameterised libraries and the impact that common restrictions on their contexts have on the definition of linearisability. We believe that our results are compatible with the existing ones and can thus be extended to cover liveness and ownership transfer [4,5].

Our work shares techniques with game semantics of concurrent programming languages [12,3] and Jeffrey and Rathke’s semantics of concurrent objects [11] (in particular, we use the $?$ and $!$ notation from the latter). The proofs of our contextuality theorems rely on the fact that library denotations satisfy certain closure properties related to $\sqsubseteq, \sqsubseteq_e$

and $\sqsubseteq_{\mathcal{R}}$, which are similar to those exploited in these prior works. However, there are two important differences. First, prior work has not studied common restrictions on library contexts (such as the encapsulation and closure conditions in Definitions 3 and 4) and the induced stronger notions of refinement between libraries, the two key topics of this paper. Second, prior works have considered all higher-order functions, while our parameterised libraries are limited to second order. Our motivation for constraining the setting in this way is to use a simple semantics and study the key issues involved in linearisability of parameterised libraries without using sophisticated machinery from game semantics, such as justification pointers and views [9], designed for accurately modelling higher-order features. However, it is definitely a promising direction to look for appropriate notions of linearisability for full higher-order concurrent libraries by combining the ideas from this paper with those from game semantics.

Turon et al. proposed CaReSL [14], a logic that allows proving observational refinements between higher-order concurrent programs directly, without going via linearisability. Their work is complimentary to ours: it provides efficient proof techniques, whereas we identify obligations to prove, independent of a particular proof system.

Acknowledgements. We thank Thomas Dinsdale-Young and Ilya Sergey for comments that helped improve the paper. This work was supported by the EU FET project AD-VENT.

References

1. Andrea Cerone, Alexey Gotsman, and Hongseok Yang. Parameterised linearisability (extended version). Available from <http://software.imdea.org/~gotsman/>.
2. Ivana Filipovic, Peter W. O’Hearn, Noam Rinetzkly, and Hongseok Yang. Abstraction for concurrent objects. *Theor. Comput. Sci.*, 411(51-52), 2010.
3. Dan R. Ghica and Andrzej S. Murawski. Angelic semantics of fine-grained concurrency. *Ann. Pure Appl. Logic*, 151(2-3), 2008.
4. Alexey Gotsman and Hongseok Yang. Liveness-preserving atomicity abstraction. In *ICALP*, 2011.
5. Alexey Gotsman and Hongseok Yang. Linearizability with ownership transfer. In *CONCUR*, 2012.
6. Danny Hendler, Itai Incze, Nir Shavit, and Moran Tzafrir. Flat combining and the synchronization-parallelism tradeoff. In *SPAA*, 2010.
7. Maurice Herlihy and Eric Koskinen. Transactional boosting: a methodology for highly-concurrent transactional objects. In *PPOPP*, 2008.
8. Maurice Herlihy and Jeannette M. Wing. Linearizability: A correctness condition for concurrent objects. *ACM Trans. Program. Lang. Syst.*, 12(3), 1990.
9. J. M. E. Hyland and C.-H. Luke Ong. On full abstraction for PCF: I, II, and III. *Inf. Comput.*, 163(2), 2000.
10. Radha Jagadeesan, Gustavo Petri, Corin Pitcher, and James Riely. Quarantining weakness - compositional reasoning under relaxed memory models. In *ESOP*, 2013.
11. Alan Jeffrey and Julian Rathke. A fully abstract may testing semantics for concurrent objects. *Theor. Comput. Sci.*, 338(1-3), 2005.
12. James Laird. A game semantics of idealized CSP. *ENTCS*, 45, 2001.
13. Claudio Russo. The Joins concurrency library. In *PADL*, 2007.
14. Aaron Turon, Derek Dreyer, and Lars Birkedal. Unifying refinement and Hoare-style reasoning in a logic for higher-order concurrency. In *ICFP*, 2013.
15. Hongseok Yang and Peter W. O’Hearn. A semantic basis for local reasoning. In *FoSSaCS*, 2002.