Optimism for Boosting Concurrency

Vincent Gramoli\textsuperscript{1}  Petr Kuznetsov\textsuperscript{2}  Srivatsan Ravi\textsuperscript{3}\textsuperscript{*}

\textsuperscript{1} NICTA and University of Sydney\textsuperscript{2}  Télécom ParisTech\textsuperscript{3}  TU Berlin/Telekom Innovation Laboratories

Abstract

Modern concurrent programming benefits from a large variety of synchronization techniques. These include conventional pessimistic locking, as well as optimistic techniques based on conditional synchronization primitives or transactional memory. Yet, it is unclear which of these approaches better leverage the concurrency inherent to multi-cores.

In this paper, we compare the level of concurrency one can obtain by converting a sequential program into a concurrent one using optimistic or pessimistic techniques. To establish fair comparison of such implementations, we introduce a new correctness criterion for concurrent programs, defined independently of the synchronization techniques they use.

We treat a program’s concurrency as its ability to accept a concurrent schedule, a metric inspired by the theories of both databases and transactional memory. We show that pessimistic locking can provide strictly higher concurrency than transactions for some applications whereas transactions can provide strictly higher concurrency than pessimistic locks for others. Finally, we show that combining the benefits of the two synchronization techniques can provide strictly more concurrency than any of them individually. We propose a list-based set algorithm that is optimal in the sense that it accepts all correct concurrent schedules. As we show via experimentation, the optimality in terms of concurrency is reflected by scalability gains.

[Regular and student paper]
1 Introduction

To exploit concurrency provided by modern multi-cores, conventional lock-based synchronization pessimistically protects accesses to the shared memory before executing them. Speculative synchronization, achieved using transactional memory (TM) or conditional primitives, such as CAS or LL/SC, optimistically executes memory operations with a risk of aborting them in the future. A programmer typically uses these synchronization techniques as “wrappers” to allow every process (or thread) to locally run its sequential code while ensuring that the resulting concurrent execution is globally correct.

Unfortunately, it is difficult for programmers to tell in advance which of the techniques will establish more concurrency in their resulting programs. By speculatively executing concurrent accesses that would have to block in a lock-based implementation, TMs [16, 21, 29] seemingly provide high concurrency. However, TMs conventionally ensure serializability [28] or even stronger properties [15], which may prohibit concurrent scenarios allowed by the sequential specification of the specific data structure we intend to implement [11].

In this paper, we analyze the “amount of concurrency” one can obtain by turning a sequential program into a concurrent one. In particular, we compare the use of optimistic and pessimistic synchronization techniques, whose popular examples are transactions and locking, respectively. To fairly compare concurrency provided by implementations based on various techniques, one has (1) to define what it means for a concurrent program to be correct regardless of the type of synchronization it uses and (2) to define a metric of concurrency.

Correctness. We begin by defining a novel consistency criterion, namely locally-serializable linearizability. We say that a concurrent implementation of a given sequential data type is locally serializable if it ensures that the local execution of each operation is equivalent to some execution of its sequential implementation. This condition is weaker than serializability since it does not require that there exists a single sequential execution that is consistent with all local executions. It is however sufficient to guarantee that optimistic executions do not observe an inconsistent transient state that could lead, for example, to a fatal error like division-by-zero.

Furthermore, the implementation should “make sense” globally, given the sequential type of the data structure we implement. The high-level history of every execution of a concurrent implementation must be linearizable [5, 24] with respect to this sequential type. The combination of local serializability and linearizability gives a correctness criterion that we call LS-linearizability, where LS stands for “locally serializable”. We show that LS-linearizability is, as the original linearizability, compositional [22, 24]: a composition of LS-linearizable implementations is also LS-linearizable. Unlike linearizability, however, it is not non-blocking: local serializability may prevent an operation in a finite LS-linearizable history from completing in a non-blocking manner.

We apply the criterion of LS-linearizability to two broad classes of pessimistic and optimistic synchronization techniques. Pessimistic implementations capture what can be achieved using classic conservative locks like mutexes, spinlocks, reader-writer locks. In contrast, optimistic implementations proceed speculatively and may roll back in the case of conflicts, e.g., relying on classical TMs, like TinySTM [9] or NOrec [7], or more relaxed forms of optimistic techniques, such as “lazy” synchronization [18], elastic transactions [10] or view transactions [1].

Concurrency metric. We measure the amount of concurrency provided by an LS-linearizable implementation as the set of schedules it accepts. To this end, we define a concurrency metric inspired by the analysis of parallelism in database concurrency control [19, 34] and transactional memory [12]. More specifically, we assume an external scheduler that defines which processes execute which steps of the corresponding sequential program in a dynamic and unpredictable fashion. This allows us to define concurrency provided by an implementation as the set of schedules (interleavings of steps of concurrent sequential operations) it accepts (is able to effectively process).
Our concurrency metric is platform-independent and it allows for measuring relative concurrency of LS-linearizable implementations using arbitrary synchronization techniques. We do not claim that this metric necessarily captures efficiency, as it does not account for other factors, like cache sizes, cache coherence protocols, or computational costs of validating a schedule, which may also affect performance on multi-core architectures. However, our experimental evaluations show that the gain in concurrency may translate into better scalability.

Measuring concurrency. This paper provides a framework to compare the concurrency one can get by choosing a particular synchronization technique for a specific data type. For the first time, we analytically capture the inherent incomparability of TM-based and pessimism-based implementations in exploiting concurrency. We illustrate this using a popular sequential list-based set implementation \[22\], concurrent implementations of which are our running examples. More precisely, we show that there exist TM-based implementations that, for some workloads, allow for more concurrency than any pessimistic implementation, but we also show that there exist pessimistic implementations that, for other workloads, allow for more concurrency than any TM-based implementation.

Intuitively, an implementation based on transactions may abort an operation based on the way concurrent steps are scheduled, while a pessimistic implementation has to proceed eagerly without knowing about how future steps will be scheduled, sometimes over-conservatively rejecting a potentially acceptable schedule. By contrast, pessimistic implementations designed to exploit the semantics of the data type can supersede the “semantics-oblivious” TM-based implementations.

More surprisingly, we demonstrate that combining the benefit of pessimistic implementations, namely their semantics awareness, and the benefit of transactions, namely their optimism, enables implementations that are strictly better-suited for exploiting concurrency than any of them individually. We describe a generic optimistic implementation of the list-based set that is optimal with respect to our concurrency metric: we show that, essentially, it accepts all correct concurrent schedules. Our implementation, designed with our theoretical concurrency metric in mind, is surprisingly reminiscent of the state-of-the-art “pragmatic” list-based set implementations \[17, 18\]. Indeed, our experimental results confirm that optimal concurrency leads to higher performance than popular pessimistic algorithms, like hand-over-hand list-based sets, or generic TM-based optimistic ones.

Our concurrency analysis is focused on a specific example of a list-based set, but our findings demonstrate the potential of the concurrency-based approach in analyzing and comparing wider classes of LS-linearizable data structures.

2 Preliminaries

Sequential types and implementations. An object type \(\tau\) is a tuple \((\Phi, \Gamma, Q, q_0, \delta)\) where \(\Phi\) is a set of operations, \(\Gamma\) is a set of responses, \(Q\) is a set of states, \(q_0 \in Q\) is an initial state and \(\delta \subseteq Q \times \Phi \times Q \times \Gamma\) is a transition relation that determines, for each state and each operation, the set of possible resulting states and produced responses \[2\]. For any type \(\tau\), each high-level object \(O_{\tau}\) of this type has a sequential implementation. For each operation \(\pi \in \Phi\), IS specifies a deterministic procedure that performs reads and writes on a collection of objects \(X_1, \ldots, X_m\) that encode a state of \(O_{\tau}\), and returns a response \(r \in \Gamma\).

As a running example, we consider the sorted linked-list based implementation of the set type, commonly referred to as the list-based set \[22\]. The set type exports operations insert\((v)\), remove\((v)\) and contains\((v)\), with \(v \in \mathbb{Z}\). We consider a sequential implementation \(LL\) of the set type using a sorted linked list where each element (or object) stores an integer value, \(val\), and a pointer to its successor, \(next\), so that elements are sorted in the ascending order of their value. Both element fields are accessed atomically. Every operation invoked with a parameter \(v\) traverses the list starting from the head up to the element storing value \(v' \geq v\). If \(v' = v\), then contains\((v)\) returns true, remove\((v)\)
unlinks the corresponding element and returns true, and insert(v) returns false. Otherwise, contains(v) and remove(v) return false while insert(v) adds a new element with value v to the list and returns true. The list-based set is denoted by (LL, set) (cf. formal definition in Appendix [A]).

**Concurrent implementations.** We tackle the problem of turning the sequential implementation IS of type τ into a concurrent one, shared by n processes p1, . . . , pn (n ∈ N). The implementation provides the processes with algorithms for the reads and writes on objects. We refer to the resulting implementation as a concurrent implementation of (IS, τ). We assume an asynchronous shared-memory system in which the processes communicate by applying primitives on shared base objects [20]. We place no upper bounds on the number of versions an object may maintain or on the size of this object. Throughout this paper, the term operation refers to some high-level operation of the type, while read-write operations on objects are referred simply as reads and writes.

An implemented read or write may abort by returning a special response ⊥. In this case we say that the corresponding high-level operation is aborted. The ⊥ event is treated both as the response event of the read or write operation and as the response of the corresponding high-level operation.

**Executions and histories.** An execution of a concurrent implementation is a sequence of invocations and responses of high-level operations of type τ, invocations and responses of read and write operations, and invocations and responses of base-object primitives. We assume that executions are well-formed: no process invokes a new read or write, or high-level operation before the previous read or write, or a high-level operation, resp., returns, or takes steps outside its read or write operation’s interval.

Let α|πi denote the subsequence of an execution α restricted to the events of process πi. Executions α and α′ are equivalent if for every process πi, α|πi = α′|πi. An operation π precedes another operation π′ in an execution α, denoted π →α π′, if the response of π occurs before the invocation of π′. Two operations are concurrent if neither precedes the other. An execution is sequential if it has no concurrent operations. A sequential execution α is legal if for every object X, every read of X in α returns the latest written value of X. An operation is complete in α if the invocation event is followed by a matching (non-⊥) response or aborted; otherwise, it is incomplete in α. Execution α is complete if every operation is complete in α.

The history exported by an execution α is the subsequence of α reduced to the invocations and responses of operations, reads and writes, except for the reads and writes that return ⊥.

**High-level histories and linearizability.** A high-level history H of an execution α is the subsequence of α consisting of all invocations and responses of non-aborted operations. A complete high-level history H is linearizable with respect to an object type τ if there exists a sequential high-level history S equivalent to H such that (1) →H ⊆→S and (2) S is consistent with the sequential specification of type τ. Now a high-level history H is linearizable if it can be completed (by adding matching responses to a subset of incomplete operations in H and removing the rest) to a linearizable high-level history [5][24].

**Obedient implementations.** We only consider implementations that satisfy the following condition: Let α be any complete sequential execution of a concurrent implementation I. Then in every execution of I of the form α · ρ1 · · · ρk where each ρi (i = 1, . . . , k) is the complete execution of a read, every read returns the value written by the last write that does not belong to an aborted operation.

Intuitively, this assumption restricts our scope to “obedient” implementations of reads and writes, where no read value may depend on some future write. In particular, we filter out implementations in which the complete execution of a high-level operation is performed within the first read or write of its sequential algorithm.

**Pessimistic implementations.** Informally, a concurrent implementation is pessimistic if the exported history contains every read-write event that appears in the execution. More precisely, no execution of a pessimistic implementation includes operations that returned ⊥.
For example, a class of pessimistic implementations are those based on locks. A lock provides shared or exclusive access to an object \( X \) through synchronization primitives \( \text{lock}^S(X) \) (shared mode), \( \text{lock}(X) \) (exclusive mode), and \( \text{unlock}(X) \). When \( \text{lock}^S(X) \) (resp. \( \text{lock}(X) \)) invoked by a process \( p_i \) returns, we say that \( p_i \text{ holds a lock on } X \text{ in shared (resp. exclusive) mode} \). A process releases the object it holds by invoking \( \text{unlock}(X) \). If no process holds a shared or exclusive lock on \( X \), then \( \text{lock}(X) \) eventually returns; if no process holds an exclusive lock on \( X \), then \( \text{lock}^S(X) \) eventually returns; and if no process holds a lock on \( X \) forever, then every \( \text{lock}(X) \) or \( \text{lock}^S(X) \) eventually returns. Given a sequential implementation of a data type, a corresponding lock-based concurrent one is derived by inserting the synchronization primitives to provide read-write access to an object.

**Optimistic implementations.** In contrast with pessimistic ones, optimistic implementations may, under certain conditions, abort an operation: some read or write may return \( \perp \), in which case the corresponding operation also returns \( \perp \).

Popular classes of optimistic implementations are those based on “lazy synchronization” \([18,22,29]\) (with the ability of returning \( \perp \) and re-invoking an operation) or transactional memory (TM) \([16,21]\). A TM provides access to a collection of objects via transactions. A transaction is a sequence of read and write operations on objects. A transaction may commit, or one of the read or write performed by the transaction may abort. Given a sequential implementation of a data type, a corresponding TM-based concurrent one puts each sequential operation within a transaction and replaces each read and write of an object \( X \) with the transactional read and write implementations, respectively. If the transaction commits, then the result of the operation is returned to the user; otherwise if one of the transactional operations aborts, \( \perp \) is returned.

## 3 Locally serializable linearizability

We are now ready to define the correctness criterion that we impose on our concurrent implementations.

Let \( H \) be a history and let \( \pi \) be a high-level operation in \( H \). Then \( H|\pi \) denotes the subsequence of \( H \) consisting of the events of \( \pi \), except for the last aborted read or write, if any. Let \( IS \) be a sequential implementation of an object of type \( \tau \) and \( \Sigma_{IS} \), the set of histories of \( IS \).

**Definition 1 (LS-linearizability)** A history \( H \) is locally serializable with respect to \( IS \) if for every high-level operation \( \pi \) in \( H \), there exists \( S \in \Sigma_{IS} \) such that \( H|\pi = S|\pi \). A history \( H \) is LS-linearizable with respect to \( IS,\tau \) (we also write \( H \) is \( IS,\tau \)-LSL) if: (1) \( H \) is locally serializable with respect to \( IS \) and (2) the corresponding high-level history \( \tilde{H} \) is linearizable with respect to \( \tau \).

Observe that local serializability stipulates that the execution is witnessed sequential by every operation. Two different operations (even when invoked by the same process) are not required to witness mutually consistent sequential executions.

A concurrent implementation \( I \) is LS-linearizable with respect to \( IS,\tau \) (we also write \( I \) is \( IS,\tau \)-LSL) if every history exported by \( I \) is \( IS,\tau \)-LSL. Throughout this paper, when we refer to a concurrent implementation of \( IS,\tau \), we assume that it is LS-linearizable with respect to \( IS,\tau \).

Just as linearizability, LS-linearizability is compositional \([22,24]\): a composition of LSL implementations is also LSL. (cf. Appendix \([13]\). However, it is not non-blocking: local serializability may prevent an operation in a finite LSL history from completing in a non-blocking manner.

**LS-linearizability versus other criteria.** LS-linearizability is a two-level consistency criterion which makes it suitable to compare concurrent implementations of a sequential data structure, regardless of synchronization techniques they use. It is quite distinct from related criteria designed for database and software transactions, such as serializability \([28,33]\) and multilevel serializability \([32,33]\).

For example, serializability \([28]\) prevents sequences of reads and writes from conflicting in a cyclic way, establishing a global order of transactions. Reasoning only at the level of reads and writes may be
Figure 1: A concurrency scenario for a list-based set, initially \{1, 3, 4\}, where value \(i\) is stored at node \(X_i\); insert(2) and insert(5) can proceed concurrently with contains(5), the history is LS-linearizable but not serializable. (We only depict important read-write events here.)

Consider an execution of a concurrent list-based set depicted in Figure 1. We assume here that the set initial state is \{1, 3, 4\}. Operation contains(5) is concurrent, first with operation insert(2) and then with operation insert(5). The history is not serializable: insert(5) sees the effect of insert(2) because \(R(X_1)\) by insert(5) returns the value of \(X_1\) that is updated by insert(2) and thus should be serialized after it. But contains(5) misses element 2 in the linked list, but must see the effect of insert(5) to perform the read of \(X_5\), i.e., the element created by insert(5). However, this history is LSL since each of the three local histories is consistent with some sequential history of LL.

Multilevel serializability \[32, 33\] was proposed to reason in terms of multiple semantic levels in the same execution. LS-linearizability, being defined for two levels only, does not require a global serialization of low-level operations as 2-level serializability does. LS-linearizability simply requires each process to observe a local serialization, which can be different from one process to another. Also, to make it more suitable for concurrency analysis of a concrete data structure, instead of semantic-based commutativity \[31\], we use the sequential specification of the high-level behavior of the object \[24\]. Linearizability \[5, 24\] only accounts for high-level behavior of a data structure, so it does not imply LS-linearizability. For example, Herlihy’s universal construction \[20\] provides a linearizable implementation for any given object type, but does not guarantee that each execution locally appears sequential with respect to any sequential implementation of the type. Local serializability, by itself, does not require any synchronization between processes and can be trivially implemented without communication among the processes. Therefore, the two parts of LS-linearizability indeed complement each other.

4 The concurrency metric

To characterize the ability of a concurrent implementation to process arbitrary interleavings of sequential code, we introduce the notion of a schedule. Intuitively, a schedule describes the order in which complete high-level operations, and sequential reads and writes are invoked by the user. More precisely, a schedule is an equivalence class of complete histories that agree on the order of invocation and response events of reads, writes and high-level operations, but not necessarily on read values or high-level responses. Thus, a schedule can be treated as a history, where responses of reads and operations are not specified.

We say that an implementation \(I\) accepts a schedule \(\sigma\) if it exports a history \(H\) such that complete\((H)\) exhibits the order of \(\sigma\), where complete\((H)\) is the subsequence of \(H\) that consists of the events of the complete operations that returned a matching response. We then say that the execution (or history) exports \(\sigma\). A schedule \(\sigma\) is (\(IS, \tau\))-LSL if there exists an (\(IS, \tau\))-LSL history that exports \(\sigma\).

A synchronization technique is a set of concurrent implementations. We define below a specific optimistic synchronization technique and then a specific pessimistic one.
The class $\mathcal{SM}$. Let $\alpha$ denote the execution of a TM implementation and $\text{ops}(\alpha)$, the set of transactions each of which performs at least one event in $\alpha$. Let $\alpha^k$ denote the prefix of $\alpha$ up to the last event of transaction $\pi_k$. Let $\text{Cseq}(\alpha)$ denote the set of subsequences of $\alpha$ that consist of all the events of transactions that are committed and some transactions that started committing in $\alpha$. We say that $\alpha$ is strictly serializable if there exists a legal sequential execution $\alpha'$ equivalent to a sequence in $\sigma \in \text{Cseq}(\alpha)$ such that $\sigma \leq \sigma'$.

This paper focuses on TM-based implementations that are strictly serializable and, in addition, guarantee that every transaction (even aborted or incomplete) observes correct (serial) behavior. More precisely, an execution $\alpha$ is safe-strict serializable if (1) $\alpha$ is strictly serializable, and (2) for each operation $\pi_k$ that is incomplete or returned $\bot$ in $\alpha$, there exist a legal sequential execution of transactions $\alpha' = \pi_0 \cdots \pi_i \cdot \pi_k$ and $\sigma \in \text{Cseq}(\alpha^k)$ such that $\{\pi_0, \cdots, \pi_i\} \subseteq \text{ops}(\sigma)$ and $\forall \pi_m \in \text{ops}(\alpha') : \alpha'[m] = \alpha^k[m]$.

Safe-strict serializability captures nicely both local serializability and linearizability. If we transform a sequential implementation $\text{IS}$ of a type $\tau$ into a concurrent one using any safe-strict serializable TM, we obtain an LSL TM-based implementation of $(\text{IS}, \tau)$. Indeed, by running each operation of $\text{IS}$ within a transaction of a safe-strict serializable TM, we make sure that operations in committed transactions witness the same execution of $\text{IS}$, and every operation that returned $\bot$ is consistent with some execution of $\text{IS}$ based on previously completed operations. Formally, $\mathcal{SM}$ denotes the set of TM-based LSL implementations. (We discuss the relations to similar but stronger TM criteria, such as opacity [15], TMS1 [8] and VWC [25] in Section 7.)

The class $\mathcal{P}$. This denotes the set of deadlock-free pessimistic LSL implementations: assuming that every process takes enough steps, at least one of the concurrent operations return a matching response [23]. Note that $\mathcal{P}$ includes implementations that are not necessarily safe-strict serializable. In the next section, we describe a pessimistic implementation of the list-based set that accepts non-serializable schedules by fine-tuning to the semantics of the set type.

5 On the incomparability of synchronization techniques

We now provide a concurrency analysis of synchronization techniques $\mathcal{SM}$ and $\mathcal{P}$ in the context of the list-based set. We describe a pessimistic implementation of $(\text{LL, set})$, $I^H \in \mathcal{P}$, that accepts non-serializable schedules: each read operation performed by contains acquires the shared lock on the object, reads the next field of the element before releasing the shared lock on the predecessor element in a hand-over-hand manner [6]. Update operations (insert and remove) acquire the exclusive lock on the head during read(head) and release it at the end. Every other read operation performed by update operations simply reads the element next field to traverse the list. The write operation performed by an insert or a remove acquires the exclusive lock, writes the value to the element and releases the lock. There is no real concurrency between any two update operations since the process holds the exclusive lock on the head throughout the operation execution. Note that $I^H$ is deadlock-free and (LL, set)-LSL.
On the one hand, the schedule of \((LL, set)\) depicted in Figure 1, which we denote by \(\sigma_0\), is not serializable as explained in Section 3 and must be rejected by any implementation in \(SM\). However, there exists an execution of \(I^R\) that exports \(\sigma_0\) since there is no read-write conflict on any two consecutive elements accessed.

On the other hand, consider the schedule \(\sigma\) of \((LL, set)\) in Figure 2(a). Clearly, \(\sigma\) is serializable and is accepted by most (progressive [14]) TM-based implementations since there is no read-write conflict. However, we prove that \(\sigma\) is not accepted by any implementation in \(P\). Our proof technique is interesting in its own right: we show that if there exists any implementation in \(P\) that accepts \(\sigma\), it must also accept the schedule \(\sigma'\) depicted in Figure 2(b). In \(\sigma'\), insert(2) overwrites the write on head performed by insert performed by insert(1) resulting in a lost update. By deadlock-freedom, there exists an extension of \(\sigma'\) in which a contains(1) returns false; but this is not a linearizable schedule.

**Theorem 2 (Incomparability)** There exist schedules \(\sigma_0\) and \(\sigma\) of \((LL, set)\) such that (1) \(\sigma_0\) is accepted by an \((LL, set)\)-LSL implementation \(I^H \in P\) but not accepted by any \((LL, set)\)-LSL implementation in \(SM\), and (2) \(\sigma\) is accepted by an \((LL, set)\)-LSL implementation \(I^C \in SM\) but not accepted by any \((LL, set)\)-LSL implementation in \(P\). (The proof is in Appendix C.1.)

The second part of Theorem 2 may look surprising, as the class \(P\) includes implementations that are relaxed (not safe-strict serializable) and fine-tuned to the semantics of the type whereas implementations in the class \(SM\) are oblivious to the semantics of the data type. However, since TM-based implementations are optimistic, i.e., every read-write operation remains tentative, the implementation does not need to be overly conservative and could return \(\bot\) in case a matching response to the operation cannot be returned.

6 On the benefits of being optimistic and relaxed

We now combine the benefits of relaxation and optimism to derive an optimistic implementation of the list-based set that supersedes every implementation in classes \(P\) and \(SM\) in terms of concurrency. Our implementation, denoted \(I^{RM}\), provides processes with algorithms for implementing read and write operations on the elements of the list for each operation of the list-based set (Algorithm 1).

Every object (or element) \(X_\ell\) is specified by the following shared variables: \(t\text{-}\text{var}[\ell]\) stores the value \(v \in V\) of \(X_\ell\), \(r[\ell]\) stores a boolean indicating if \(X_\ell\) is marked for deletion, \(L[\ell]\) stores a tuple of the version number of \(X_\ell\) and a locked flag; the latter indicates whether a concurrent process is performing a write to \(X_\ell\).

Any operation with input parameter \(v\) traverses the list starting from the head element up to the element storing value \(v' \geq v\) without writing to shared memory. If a read operation on an element conflicts with a write operation to the same element or if the element is marked for deletion, the operation terminates by returning \(\bot\). While traversing the list, the process maintains the last two read elements and their version numbers in the local rotating buffer \(rbuf\). If none of the read operations performed by \(\text{contains}(v)\) return \(\bot\) and if \(v' = v\), then \(\text{contains}(v)\) returns \(\text{true}\); otherwise it returns \(\text{false}\). Thus, the \(\text{contains}\) does not write to shared memory.

To perform write operation to an element as part of an update operation (insert and remove), the process first retrieves the version of the object that belongs to its rotating buffer. It returns \(\bot\) if the version has been changed since the previous read of the element or if a concurrent process is executing a write to the same element. The write operation performed by the remove operation, additionally checks if the element to be removed from the list is locked by another process; if not, it sets a flag on the element to mark it for deletion. If none of the read or write operations performed during the insert\((v)\) or remove\((v)\) returned \(\bot\), appropriate matching responses are returned as prescribed by the sequential implementation \(LL\). Any update operation of \(I^{RM}\) uses at most two expensive synchronization patterns [3].
Theorem 3 (Optimality) \( I_{RM} \) accepts all schedules that are observable with respect to \((LL, set)\).

Proof sketch. We prove that any schedule rejected by \( I_{RM} \) is not observable. We go through the cases when a read or write returns \( \perp \) (implying the operation fails to return a matching response) and thus the current schedule is rejected: (1) \( \text{read}(X_\ell) \) returns \( \perp \) in line 16 when \( r[\ell] = \text{true} \) or when \( \text{ver}_1 \neq \text{ver}_2 \), (2) \( \text{write}(X_\ell) \) performed by \( \text{remove}(v) \) either returns \( \perp \) in line 22 when the \( \text{cas} \) operation on \( L[\ell] \) returns false or returns \( \perp \) in line 25 when the \( \text{cas} \) operation on the element that stores \( v \) returns false, and (3) \( \text{write}(X_\ell) \) performed by \( \text{insert} \) returns \( \perp \) in line 35 when the \( \text{cas} \) operation on \( L[\ell] \) returns false.

Consider the subcase (1a), \( r[\ell] \) is set \( \text{true} \) by a preceding or concurrent \( \text{write}(X_\ell) \) (line 27). The high-level operation performing this write is a remove that marks the corresponding list element as removed. Since no removed element can be read in a sequential execution of \( LL \), the corresponding history is not locally serializable. Alternatively, in subcase (1b), the version of \( X_\ell \) read previously in line 11 has changed. Thus, an update operation has concurrently performed a write to \( X_\ell \). However, there exist executions that export such schedules.
In case (2), the write performed by a remove operation returns ⊥. In subcase (2a), \(X_ℓ\) is currently locked. Thus, a concurrent high-level operation has previously locked \(X_ℓ\) (by successfully performing \(L[ℓ].\text{cas}()\) in line 22) and has not yet released the lock (by writing \(⟨\text{ver}', \text{false}⟩\) to \(L[ℓ]\) in line 29). In subcase (2b), the current version of \(X_ℓ\) (stored in \(L[ℓ]\)) differs from the version of \(X_ℓ\) witnessed by a preceding read. Thus, a concurrent high-level operation completed a write to \(X_ℓ\) after the current high-level operation \(π\) performed a read of \(X_ℓ\). In both (2a) and (2b), a concurrent high-level updating operation \(π'\) (remove or insert) has written or is about to perform a write to \(X_ℓ\). In subcase (2c), the \text{cas} on the element \(X_ℓ\) (element that stores the value \(v\)) executed by remove\((v)\) returns false (line 25). Recall that by the sequential implementation \(LL\), remove\((v)\) performs a read of \(X_ℓ'\) prior to the write\((X_ℓ)\), where \(X_ℓ.next\) refers to \(X_ℓ\). If the \text{cas} on \(X_ℓ\) fails, there exists a process that concurrently performed a write to \(X_ℓ\), but after the read of \(X_ℓ\) by remove\((v)\). In all cases, we observe that if we did not abort the write to \(X_ℓ\), then the schedule extended by a complete execution of contains is not LSL.

In case (3), the write performed by an insert operation returns ⊥. Similar arguments to case (2) prove that any schedule rejected is not observable LSL.

Theorem 3 implies that the schedules exported by the histories in Figures 1 and 2(a) and that are not accepted by any \(I' \in SM\) and any \(I \in P\), respectively, are indeed accepted by \(RM\). But it is easy to see that implementations in \(SM\) and \(P\) can only accept observable schedules. As a result, \(RM\) can be shown to strictly supersede any pessimistic or TM-based implementation of the list-based set.

Corollary 4 \(RM\) accepts every schedule accepted by any implementation in \(P\) and \(SM\). Moreover, \(RM\) accepts schedules \(σ\) and \(σ'\) that are rejected by any implementation in \(P\) and \(SM\), respectively.

One take-away from these results is that generic optimistic implementations, appropriately relaxed, are able to provide strictly more concurrency than pessimistic or strongly consistent optimistic ones. Our implementation \(RM\) is in fact optimal with respect to concurrency, while still incurring minimal cost in terms of step-complexity and use of expensive synchronization patterns.

7 Related work

Sets of accepted schedules are commonly used as a metric of concurrency provided by a shared-memory implementation. For static database transactions, Kung and Papadimitriou [26] use the metric to capture the parallelism of a locking scheme. While acknowledging that the metric is theoretical, they insist that it may have “practical significance as well, if the schedulers in question have relatively small scheduling times as compared with waiting and execution times.” Herlihy [19] employed the metric to compare various optimistic and pessimistic synchronization techniques using commutativity of operations constituting high-level transactions. A synchronization technique is implicitly considered in [19] as highly concurrent, namely “optimal”, if no other technique accepts more schedules.

By contrast, we focus here on a dynamic model where the scheduler cannot use the prior knowledge of all the shared addresses to be accessed. Also, unlike [19, 26], we require all operations, including aborted ones, to observe (locally) consistent states. As we confirm experimentally, our (provably optimal) optimistic implementation incurs negligible scheduling overhead, which makes the motivation of the metric proposed in [26] applicable.

Gramoli et al. [12] defined a concurrency metric, the input acceptance, as the ratio of committed transactions over aborted transactions when the Unlike our metric, TM executes the given schedule. Input acceptance does not apply to lock-based programs.

Similar to other relaxations of opacity [15] like TMS1 [8] and VWC [25], safe-strict serializable implementations (SM) require that every transaction (even aborted and incomplete) observes “correct” serial behavior. However, unlike TMS1, we do not require the local serial executions to always respect
the real-time order among transactions. Unlike VWC, we model transactional operations as intervals with an invocation and a response and does not assume unique writes (needed to define causal past in VWC). Therefore, \( \mathcal{SM} \) appears weaker than both TMS1 and VWC. Though weak and possibly not very pragmatic, it still allows us to show that resulting TM-based LSL implementations reject some schedules accepted by pessimistic locks. However, we can easily extend our results to show that even opaque TMs accept some schedules rejected by any pessimistic algorithm.

The problem of transforming a sequential implementation of a list-based set into a concurrent one was considered before in special settings. Vechev and Yahav [30] considered using locks while Felber et al. [10] considered using elastic transactions. Our framework applies to generic concurrent transformations of sequential implementations using arbitrary synchronization techniques.

8 Discussion and concluding remarks

To confirm the practicality of our optimistic list-based set, \( I^{RM} \), we compared its Java implementation against a pessimistic implementation \( HOHL \) that uses hand-over-hand locking (we adopted the Java pseudocode by Herlihy and Shavit [22, Chapter 9]). Figure 3 presents the throughput (the number of completed operations per millisecond) of the two algorithms on a 32-way machine where up to 32 threads run 5% updates (either remove or insert with the same probability) and 95% contains operations on a list, initially populated with 512 integer values. \( I^{RM} \) outperforms both HOHL and E-STM [10], and the reason for this could be that \( I^{RM} \) is optimal in terms of concurrency (Theorem 3), while the HOHL is serializable [4] and, thus, rejects large classes of correct schedules (e.g., of the kind of \( \sigma_0 \) in Figure 1). E-STM does not provide optimal concurrency but rejects less schedules than HOHL. Additionally, our implementation uses asymptotically less expensive memory barriers and read-modify-write primitives [3] than HOHL. We deduce that accepting all observable schedules may be quite efficient on some applications. For the sake of simplicity, we did not optimize our code, e.g., by removing wrappers or using partial aborts. (More experimental results are given in Appendix E.)

\( I^{RM} \) is surprisingly reminiscent of the list-based set implementations in [18] and [17] (state-of-the-art, to the best of our knowledge). However, because of specific optimizations of the \( \text{contains} \) operation, strictly speaking, none of these two algorithms is locally serializable, and thus LSL. The implementations in [17, 18] use the logical deletion technique to associate a \( \text{marked} \) field to an element to indicate if it is contained in the list. The \( \text{contains} \) of the lazy list-based set [18] may read an element marked for deletion, whereas the \( \text{contains} \) of Harris list-based set [17] even uses \( \text{cas} \) to remove logically deleted nodes. But the apparent similarities between our \( I^{RM} \) and the algorithms in [17, 18] suggest that looking for a concurrency optimal LS-linearizable implementation also helps in optimizing performance. An implementation, once proven to be concurrency optimal, may be optimized further to boost performance, as is possible with \( I^{RM} \).

We derived our concurrency lower bounds in the context of the list-based set, a data structure that is suitable for exploiting concurrency because of its localized updates, but we believe that it should be possible to generalize our results to a wider class of search structures. This paper provides some preliminary hints in the quest for the “right” synchronization technique to develop highly concurrent and efficient implementations of data types. Our results are relevant to programmers leveraging multicore architectures as well as to computer manufacturers who aim at defining new instruction sets. This work suggests directions to identify the “killer” application for existing and emerging synchronization techniques.
References

[1] Y. Afek, A. Morrison, and M. Tzafrir. View transactions: Transactional model with relaxed consistency checks. In PODC, 2010.

[2] M. K. Aguilera, S. Frølund, V. Hadzilacos, S. L. Horn, and S. Toueg. Abortable and query-abortable objects and their efficient implementation. In PODC, pages 23–32, 2007.

[3] H. Attiya, R. Guerraoui, D. Hendler, P. Kuznetsov, M. Michael, and M. Vechev. Laws of order: Expensive synchronization in concurrent algorithms cannot be eliminated. In POPL, pages 487–498, 2011.

[4] H. Attiya, G. Ramalingam, and N. Rinetzky. Sequential verification of serializability. In Proceedings of the 37th annual ACM SIGPLAN-SIGACT symposium on Principles of programming languages, pages 31–42, 2010.

[5] H. Attiya and J. Welch. Distributed Computing. Fundamentals, Simulations, and Advanced Topics. John Wiley & Sons, 2004.

[6] R. Bayer and M. Schkolnick. Concurrency of operations on B-trees. In Readings in database systems, pages 129–139. Morgan Kaufmann Publishers Inc., 1988.

[7] L. Dalessandro, M. F. Spear, and M. L. Scott. NOrec: streamlining STM by abolishing ownership records. In PPOPP, pages 67–78, 2010.

[8] S. Doherty, L. Groves, V. Luchangco, and M. Moir. Towards formally specifying and verifying transactional memory. Electron. Notes Theor. Comput. Sci., 259:245–261, Dec. 2009.

[9] P. Felber, C. Fetzer, and T. Riegel. Dynamic performance tuning of word-based software transactional memory. In PPoPP, pages 237–246, 2008.

[10] P. Felber, V. Gramoli, and R. Guerraoui. Elastic transactions. In DISC, pages 93–107, 2009.

[11] V. Gramoli and R. Guerraoui. Democratizing transactional programming. Commun. ACM, 57(1):86–93, Jan 2014.

[12] V. Gramoli, D. Harmanci, and P. Felber. On the input acceptance of transactional memory. Parallel Processing Letters, 20(1):31–50, 2010.

[13] R. Guerraoui and M. Kapalka. The semantics of progress in lock-based transactional memory. In POPL, pages 404–415, 2009.

[14] R. Guerraoui and M. Kapalka. Transactional memory: Glimmer of a theory. In Proceedings of the 21st International Conference on Computer Aided Verification, CAV ’09, pages 1–15, Berlin, Heidelberg, 2009. Springer-Verlag.

[15] R. Guerraoui and M. Kapalka. Principles of Transactional Memory, Synthesis Lectures on Distributed Computing Theory. Morgan and Claypool, 2010.

[16] T. Harris, J. R. Larus, and R. Rajwar. Transactional Memory, 2nd edition. Synthesis Lectures on Computer Architecture. Morgan & Claypool Publishers, 2010.

[17] T. L. Harris. A pragmatic implementation of non-blocking linked-lists. In DISC, pages 300–314, 2001.
A Sequential implementation of the set type

Recall that an object type $\tau$ is a tuple $(\Phi, \Gamma, Q, q_0, \delta)$ where $\Phi$ is a set of operations, $\Gamma$ is a set of responses, $Q$ is a set of states, $q_0 \in Q$ is an initial state and $\delta \subseteq Q \times \Phi \times Q \times \Gamma$ is a transition relation that determines, for each state, and each operation, the set of possible resulting states and produced responses. Hence, $(q, \pi, q', r) \in \delta$ implies that when an operation $\pi \in \Phi$ is applied on an object of type $\tau$ in state $q$, the object moves to state $q'$ and returns a response $r$. We consider only types that are total i.e., for every $q \in Q$, $\pi \in \Phi$, there exist $q' \in Q$ and $r \in \Gamma$ such that $(q, \pi, q', r) \in \delta$. We assume
that every type \( \tau = (\Phi, \Gamma, Q, q_0, \delta) \) is \textit{computable}, i.e., there exists a Turing machine that, for each input \((q, \pi), q \in Q, \pi \in \Phi\), computes a pair \((q', r)\) such that \((q, \pi, q', r) \in \delta\).

Formally, the \textit{set} type is defined by the tuple \((\Phi, \Gamma, Q, q_0, \delta)\) where:

\[
\Phi = \{\text{insert}(v), \text{remove}(v), \text{contains}(v)\}; \ v \in \mathbb{Z}
\]

\[
\Gamma = \{\text{true}, \text{false}\}
\]

\(Q\) is the set of all finite subsets of \(\mathbb{Z}\); \(q_0 = \emptyset\)

\(\delta\) is defined as follows:

\[
(1): (q, \text{contains}(v), q, (v \in q))
\]

\[
(2): (q, \text{insert}(v), q \cup \{v\}, (v \not\in q))
\]

\[
(3): (q, \text{remove}(v), q \setminus \{v\}, (v \in q))
\]

The sequential implementation \(LL\) of the \textit{set} type is presented in Algorithm 2. The implementation uses a \textit{sorted linked list} data structure in which each element (except the \textit{tail}) maintains a \textit{next} field to provide a pointer to the successor node. Initially, the \textit{next} field of the \textit{head} element points to \textit{tail}; \textit{head} (resp. \textit{tail}) is initialized with values \(-\infty\) (resp. \(+\infty\)) that is smaller (resp. greater) than the value of any other element in the list.

## B LS-linearizability is compositional

We define the composition of two distinct object types \(\tau_1\) and \(\tau_2\) as a type \(\tau_1 \times \tau_2 = (\Phi, \Gamma, Q, q_0, \delta)\) as follows: \(\Phi = \Phi_1 \cup \Phi_2\), \(\Gamma = \Gamma_1 \cup \Gamma_2\), \(Q = Q_1 \times Q_2\), \(q_0 = (q_0_1, q_0_2)\), and \(\delta \subseteq Q \times \Phi \times Q \times \Gamma\) is such that \(((q_1, q_2), \pi, (q'_1, q'_2), r) \in \delta\) if and only if for \(i \in \{1, 2\}, \) if \(\pi \in \Phi_i\) then \((q_i, \pi, q'_i, r) \in \delta_i \land q_3-i = q'_3-i\).

---

\[\text{Algorithm 2: Sequential implementation } LL (\text{sorted linked list}) \text{ of set type}\]
Every sequential implementation $IS$ of an object $O_1 \times O_2$ of a composed type $\tau_1 \times \tau_2$ naturally induces two sequential implementations $IS_1$ and $IS_2$ of objects $O_1$ and $O_2$, respectively. Now a correctness criterion $\Psi$ is *compositional* if for every history $H$ on an object composition $O_1 \times O_2$, if $\Psi$ holds for $H|O_i$ with respect to $IS_i$, for $i \in \{1, 2\}$, then $\Psi$ holds for $H$ with respect to $IS = IS_1 \times IS_2$. Here, $H|O_i$ denotes the subsequence of $H$ consisting of events on $O_i$.

**Theorem 5** LS-linearizability is compositional.

**Proof.** Let $H$, a history on $O_1 \times O_2$, be LS-linearizable with respect to $IS$. Let each $H|O_i$, $i \in \{1, 2\}$, be LS-linearizable with respect to $IS_i$. Without loss of generality, we assume that $H$ is complete (if $H$ is incomplete, we consider any completion of it containing LS-linearizable completions of $H|O_1$ and $H|O_2$).

Let $\tilde{H}$ be a completion of the high-level history corresponding to $H$ such that $\tilde{H}|O_1$ and $\tilde{H}|O_2$ are linearizable with respect to $\tau_1$ and $\tau_2$, respectively. Since linearity is compositional [22,24], $\tilde{H}$ is linearizable with respect to $\tau_1 \times \tau_2$.

Now let, for each operation $\pi$, $S^1_{\pi}$ and $S^2_{\pi}$ be any two sequential histories of $IS_1$ and $IS_2$ such that $H|\pi|O_j = S^j_{\pi}|\pi$, for $j \in \{1, 2\}$ (since $H|O_1$ and $H|O_2$ are LS-linearizable such histories exist). We construct a sequential history $S_\pi$ by interleaving events of $S^1_{\pi}$ and $S^2_{\pi}$ so that $S_{\pi}|O_j = S^j_{\pi}$, $j \in \{1, 2\}$.

Since each $S^j_{\pi}$ acts on a distinct component $O_j$ of $O_1 \times O_2$, every such $S_{\pi}$ is a sequential history of $IS$. We pick one $S_{\pi}$ that respects the local history $H|\pi$, which is possible, since $H|\pi$ is consistent with both $S^1_{\pi}$ and $S^2_{\pi}$.

Thus, for each $\pi$, we obtain a history of $IS$ that agrees with $H|\pi$. Moreover, the high-level history of $H$ is linearizable. Thus, $H$ is LS-linearizable with respect to $IS$. 

\[\square\]

C Formal proofs

C.1 SM vs. $P$

A pessimistic implementation $I^H \in P$ of $(LL, set)$. The implementation $I^H$ is a lock-based implementation that associates every object with a distinct lock and another base object that stores the value of the object. In $I^H$, the contains operation uses shared hand-over-hand locking [6,22]. Each read operation performed by contains acquires the shared lock on the object, reads the next field of the element before releasing the shared lock on the predecessor element. Update operations (insert and remove) acquire the exclusive lock on the head during read(head) and release it at the end. Every other read operation performed by an update simply reads the next field of the element to traverse the list. The write operation performed by a insert or remove acquires the exclusive lock, writes the value to the element and releases the lock. There is no real concurrency between any two update operations since the process holds the exclusive lock on the head throughout the operation execution. Intuitively, it is easy to observe that $I^H$ is LS-linearizable with respect to $(LL, set)$.

**Theorem 6 (Part 1 of Theorem 2)** There exists a schedule $\sigma_0$ of $(LL, set)$ that is accepted by $I^H \in PL$, but not accepted by any $(LL, set)$-LSL implementation $I \in SM$.

**Proof.** Let $\sigma_0$ be the schedule of $(LL, set)$ depicted in Figure 1. Suppose by contradiction that $\sigma_0 \in S(I)$, where $I$ is an implementation of $(LL, set)$ based on any safe-strict serializable TM. Thus, there exists an execution $\alpha$ of $I$ that exports $\sigma_0$. Now consider two cases:

- Suppose that the read of $X_4$ by contains(5) returns the value of $X_4$ that is updated by insert(5). Since insert(2) $\rightarrow_\alpha$ insert(5), insert(2) must precede insert(5) in any sequential execution $\alpha'$ equivalent to $\alpha$. Also, since contains(5) reads $X_1$ prior to its update by insert(2), contains(5) must
Consider the pessimistic implementation \( I^P \) shared hand-over-hand locking, the process prior to the acquisition of the exclusive lock on sequential execution of operations alizable, and (2) for each operation \( \pi \) lock on \( \pi cseq \) while still holding the shared lock on element \( X \) in \( \alpha \). An optimistic implementation \( I^C \) are concurrent, both access the same object \( T \) tion [13]: if a transaction \( \exists \) holds the exclusive lock on \( X_1 \) by insert(2). Similarly, \( p_1 \) can acquire the shared lock on \( X_4 \) immediately after the release of the exclusive lock on \( X_4 \) by the process executing insert(5) while still holding the shared lock on element \( X_3 \). Thus, there exists an execution of \( I^H \) that exports \( \sigma_0 \). □

An optimistic implementation \( I^C \in SM \) of \((LL, set)\). Recall that \( SM \) denotes the set of concurrent implementations based on TMs that ensure the following safety condition in every execution.

Let \( \alpha \) denote the execution of a TM-based implementation and \( ops(\alpha) \), the set of operations each of which performs at least one event in \( \alpha \). Let \( \alpha^k \) denote the prefix of \( \alpha \) up to the last event of operation \( \pi_k \). Let \( cseq(\alpha) \) denote the subsequence of \( \alpha \) that consists of the events of the complete operations in \( \alpha \). We say that \( \alpha \) is strictly serializable if there exists a legal sequential execution \( \alpha' \) equivalent to \( cseq(\alpha) \) such that \( \rightarrow_{cseq(\alpha)} \subseteq \rightarrow_{\alpha'} \).

An execution \( \alpha \) of a TM-based implementation is safe-strict serializable if (1) \( \alpha \) is strictly serializable, and (2) for each operation \( \pi_k \) that is incomplete or returned \( \perp \) in \( \alpha \), there exists a legal sequential execution of operations \( \alpha' = \pi_0 \cdots \pi_i \cdot \pi_k \) such that \( \{ \pi_0, \cdots, \pi_i \} \subseteq ops(cseq(\alpha^k)) \) and \( \forall \pi_m \in ops(\alpha') : \alpha'|m = \alpha^k|m \).

The implementation \( I^C \) is based on a TM that ensures the following condition in every execution [13]: if a transaction \( T_i \) aborts, then it encounters a conflict with a transaction \( T_j \), i.e., \( T_i \) and \( T_j \) are concurrent, both access the same object \( X \), and at least one of these is a write. An implementation of such a safe-strict serializable TM can be found in [27].

**Theorem 7 (Part 2 of Theorem [2])** There exists a schedule \( \sigma \) of \((LL, set)\) that is accepted by an \((LL, set)\)-LSL implementation \( I^C \in SM \), but not accepted by any \((LL, set)\)-LSL implementation in \( P \).

**Proof.** We show first that the schedule \( \sigma \) of \((LL, set)\) depicted in Figure [2]a is not accepted by any implementation in \( P \). Suppose the contrary and let \( \sigma \) be exported by an execution \( \alpha \). Here \( \alpha \) starts with three sequential insert operations with parameters 1, 2, and 3. The resulting “state” of the set is \( \{1,2,3\} \), where value \( i \in \{1,2,3\} \) is stored in object \( X_i \).

Suppose, by contradiction, that some \( I \in P \) accepts \( \sigma \). We show that \( I \) then accepts the schedule \( \sigma' \) depicted in Figure [2]b, which starts with a sequential execution of insert(3) storing value 3 in object \( X_1 \).

Let \( \alpha' \) be any history of \( I \) that exports \( \sigma' \). Recall that we only consider obedient implementations: in \( \alpha' \): the read of \( head \) by insert(2) in \( \alpha' \) refers to \( X_1 \) (the next element to be read by insert(2)). In \( \alpha \), element \( X_1 \) stores value 1, i.e., insert(1) can safely return false, while in \( \alpha' \), \( X_1 \) stores value 3, i.e., the next step of insert(1) must be a write to \( head \). Thus, no process can distinguish \( \alpha \) and \( \alpha' \) before the read operations on \( X_1 \) return. Let \( \alpha'' \) be the prefix of \( \alpha' \) ending with \( R(X_1) \) executed by insert(2). Since \( I \) is deadlock-free, we have an extension of \( \alpha'' \) in which both insert(1) and insert(2) terminate; we show that this extension violates linearizability. Since \( I \) is locally-serializable, to respect our sequential implementation of \((LL, set)\), both operations should complete the write to \( head \) before returning. Let \( \pi_1 = \text{insert}(1) \) be the first operation to write to \( head \) in this extended execution. Let \( \pi_2 = \text{insert}(2) \)
be the other insert operation. It is clear that \( \pi_1 \) returns \( \text{true} \) even though \( \pi_2 \) overwrites the update of \( \pi_1 \) on \( \text{head} \) and also returns \( \text{true} \). Recall that implementations in \( P \) are deadlock-free. Thus, we can further extend the execution with a complete contains(1) that will return \( \text{false} \) (the element inserted to the list by \( \pi_1 \) is lost)—a contradiction since \( I \) is linearizable with respect to \( \text{set} \). Thus, \( \sigma \not\in S(I) \) for any \( I \in P \).

On the other hand, the schedule \( \sigma \) is accepted by \( IC \in SM \), since there is no conflict between the two concurrent update operations. \( \square \)

## D Relaxed optimistic implementation: proof of correctness

Let \( \alpha \) be an execution of \( IR^M \) and \( \prec_\alpha \) denote the total-order on events in \( \alpha \). For simplicity, we assume that \( \alpha \) starts with an artificial sequential execution of an insert operation \( \pi_0 \) that inserts \( \text{tail} \) and sets \( \text{head.next} = \text{tail} \). Let \( H \) be the history exported by \( \alpha \), where all reads and writes are sequential. We construct \( H \) by associating a linearization point \( \ell_{op} \) with each non-aborted read or write operation \( op \) performed in \( \alpha \) as follows:

- if \( op \) is a read, then performed by process \( pk \), \( \ell_{op} \) is the base-object \( \text{read} \) in line 12.
- if \( op \) is a write within an \( \text{insert} \) operation, \( \ell_{op} \) is the base-object \( \text{cas} \) in line 22.
- if \( op \) is a write within a \( \text{remove} \) operation, \( \ell_{op} \) is the base-object \( \text{cas} \) in line 35.

We say that a \( \text{read} \) of an element \( X \) within an operation \( \pi \) is \( \text{valid} \) in \( H \) (we also say that \( X \) is \( \text{valid} \)) if there does not exist any \( \text{remove} \) operation \( \pi_1 \) that \( \text{deallocates} \) \( X \) (removes \( X \) from the list) such that \( \ell_{\pi_1.\text{write}(X)} \prec_{\alpha} \ell_{\pi.\text{read}(X)} \).

### Lemma 8

Let \( \pi \) be any operation performing \( \text{read}(X) \) followed by \( \text{read}(Y) \) in \( H \). Then (1) there exists an \( \text{insert} \) operation that sets \( X.\text{next} = Y \) prior to \( \pi.\text{read}(X) \), and (2) \( \pi.\text{read}(X) \) and \( \pi.\text{read}(Y) \) are valid in \( H \).

**Proof.** Let \( \pi \) be any operation in \( IR^M \) that performs \( \text{read}(X) \) followed by \( \text{read}(Y) \). If \( X \) and \( Y \) are \( \text{head} \) and \( \text{tail} \) respectively, \( X.\text{next} = \text{tail} \) (by assumption). Since no \( \text{remove} \) operation deallocates the \( \text{head} \) or \( \text{tail} \), the \( \text{read} \) of \( X \) and \( Y \) are \( \text{valid} \) in \( H \).

Now, let \( X \) be the \( \text{head} \) element and suppose that \( \pi \) performs \( \text{read}(X) \) followed by \( \text{read}(Y) \); \( Y \neq \text{tail} \) in \( H \). Clearly, if \( \pi \) performs a \( \text{read}(Y) \), there exists an operation \( \pi' = \text{insert} \) that has previously set \( \text{head.next} = Y \). More specifically, \( \pi.\text{read}(X) \) performs the action in line 12 after the write to shared memory by \( \pi' \) in line 37. By the assignment of linearization points to \( \text{tx} \)-operations, \( \ell_{\pi'.\text{write}(X)} \prec_{\alpha} \ell_{\pi.\text{read}(X)} \). Thus, there exists an \( \text{insert} \) operation that sets \( X.\text{next} = Y \) prior to \( \pi.\text{read}(X) \) in \( H \).

For the second claim, we need to prove that the \( \text{read}(Y) \) by \( \pi \) is \( \text{valid} \) in \( H \). Suppose by contradiction that \( Y \) has been deallocated by some \( \pi'' = \text{remove} \) operation prior to \( \text{read}(Y) \) by \( \pi \). By the rules for linearization of read and write operations, the action in line 28 precedes the action in line 12. However, \( \pi \) proceeds to perform the check in line 15, and returns \( \bot \) since the flag corresponding to the element \( Y \) is previously set by \( \pi'' \). Thus, \( H \) does not contain \( \pi.\text{read}(Y) \)—contradiction.

Inductively, by the above arguments, every non-\( \text{head} \) \( \text{read} \) by \( \pi \) is performed on an element previously created by an \( \text{insert} \) operation and is valid in \( H \). \( \square \)

### Lemma 9

\( H \) is locally serializable with respect to \( \text{LL} \).
Proof. By Lemma 8, every element \( X \) read within an operation \( \pi \) is previously created by an insert operation and is valid in \( H \). Moreover, if the read operation on \( X \) returns \( v' \), then \( X.next \) stores a pointer to another valid element that stores an integer value \( v'' > v' \). Note that the series of reads performed by \( \pi \) terminates as soon as an element storing value \( v \) or higher is found. Thus, \( \pi \) performs at most \( O(|v - v_0|) \) reads, where \( v_0 \) is the value of the second element read by \( \pi \). Now we construct \( S^\pi \) as a sequence of insert operations, that insert values read by \( \pi \), one by one, followed by \( \pi \). By construction, \( S^\pi \in \Sigma_{IL} \).

It is sufficient for us to prove that every finite high-level history \( H \) of \( I^{RM} \) is linearizable. First, we obtain a completion \( \bar{H} \) of \( H \) as follows. The invocation of an incomplete contains operation is discarded. The invocation of an incomplete \( \pi = \text{insert} \lor \text{remove} \) operation that has not returned successfully from the write operation is discarded; otherwise, it is completed with response true.

We obtain a sequential high-level history \( \bar{S} \) equivalent to \( \bar{H} \) by associating a linearization point \( \ell_\pi \) with each operation \( \pi \) as follows. For each \( \pi = \text{insert} \lor \text{remove} \) that returns true in \( \bar{H} \), \( \ell_\pi \) is associated with the first write performed by \( \pi \) in \( H \); otherwise \( \ell_\pi \) is associated with the last read performed by \( \pi \) in \( H \). For \( \pi = \text{contains} \) that returns true, \( \ell_\pi \) is associated with the last read performed in \( I^{RM} \); otherwise \( \ell_\pi \) is associated with the read of head. Since linearization points are chosen within the intervals of operations of \( I^{RM} \), for any two operations \( \pi_i \) and \( \pi_j \) in \( \bar{H} \), if \( \pi_i \rightarrow H \pi_j \), then \( \pi_i \rightarrow \bar{S} \pi_j \).

Lemma 10 \( \bar{S} \) is consistent with the sequential specification of type set.

Proof. Let \( \bar{S}^k \) be the prefix of \( \bar{S} \) consisting of the first \( k \) complete operations. We associate each \( \bar{S}^k \) with a set \( q^k \) of objects that were successfully inserted and not subsequently successfully removed in \( \bar{S}^k \). We show by induction on \( k \) that the sequence of state transitions in \( \bar{S}^k \) is consistent with operations’ responses in \( \bar{S}^k \) with respect to the set type.

The base case \( k = 1 \) is trivial: the tail element containing \( +\infty \) is successfully inserted. Suppose that \( \bar{S}^k \) is consistent with the set type and let \( \pi_1 \) with argument \( v \in Z \) and response \( r_\pi_1 \) be the last operation of \( \bar{S}^{k+1} \). We want to show that \((q^k, \pi_1, q^{k+1}, r_\pi_1)\) is consistent with the set type.

1. If \( \pi_1 = \text{insert}(v) \) returns true in \( \bar{S}^{k+1} \), there does not exist any other \( \pi_2 = \text{insert}(v) \) that returns true in \( \bar{S}^{k+1} \) such that there does not exist any remove(v) that returns true; \( \pi_2 \rightarrow \bar{S}^{k+1} \). Suppose by contradiction that such a \( \pi_1 \) and \( \pi_2 \) exist. Every successful insert(v) operation performs its penultimate read on an element \( X \) that stores a value \( v' < v \) and the last read is performed on an element that stores a value \( v'' > v \). Clearly, \( \pi_1 \) also performs a write on \( X \). By construction of \( \bar{S} \), \( \pi_1 \) is linearized at the release of the cas lock on element \( X \). Observe that \( \pi_2 \) must also perform a write to the element \( X \) (otherwise one of \( \pi_1 \) or \( \pi_2 \) would return false). By assumption, the write to \( X \) in shared-memory by \( \pi_2 \) (line 37) precedes the corresponding write to \( X \) in shared-memory by \( \pi_2 \). If \( \ell_\pi_2 \prec_\alpha \ell_{\pi_1, \text{read}(X)} \), then \( \pi_1 \) cannot return true—a contradiction. Otherwise, if \( \ell_{\pi_1, \text{read}(X)} \prec_\alpha \ell_\pi_2 \), then \( \pi_1 \) reaches line 22 and return \( \bot \). This is because either \( \pi_1 \) attempts to acquire the cas lock on \( X \) while it is still held by \( \pi_2 \) or the value of \( X \) contained in the rbuf of the process executing \( \pi_1 \) has changed—a contradiction.

If \( \pi_1 = \text{insert}(v) \) returns false in \( \bar{S}^{k+1} \), there exists a \( \pi_2 = \text{insert}(v) \) that returns true in \( \bar{S}^{k+1} \) such that there does not exist any \( \pi_3 = \text{remove}(v) \) that returns true; \( \pi_2 \rightarrow \bar{S}^{k+1} \pi_3 \rightarrow \bar{S}^{k+1} \). Suppose that such a \( \pi_2 \) does not exist. Thus, \( \pi_1 \) must perform its last read on an element that stores value \( v'' > v \), perform the action in Line 37 and return true—a contradiction.

It is easy to verify that the conjunction of the above two claims prove that \( \forall q \in Q; \forall v \in Z \), \( \bar{S}^{k+1} \) satisfies \((q, \text{insert}(v), q \cup \{v\}, (v \notin q))\).

2. If \( \pi_1 = \text{remove}(v) \), similar arguments as applied to insert(v) prove that \( \forall q \in Q; \forall v \in Z \), \( \bar{S}^{k+1} \) satisfies \((q, \text{remove}(v), q \setminus \{v\}, (v \in q))\).
(3) If \( \pi_1 = \text{contains}(v) \) returns \text{true} in \( \tilde{S}^{k+1} \), there exists \( \pi_2 = \text{insert}(v) \) that returns \text{true} in \( \tilde{S}^{k+1} \) such that there does not exist any \( \text{remove}(v) \) that returns \text{true} in \( \tilde{S}^{k+1} \) such that \( \pi_2 \rightarrow \tilde{S}^{k+1} \). The proof of this claim immediately follows from Lemma 8.

Now, if \( \pi_1 = \text{contains}(v) \) returns \text{false} in \( \tilde{S}^{k+1} \), there does not exist an \( \pi_2 = \text{insert}(v) \) that returns \text{true} such that there does not exist any \( \text{remove}(v) \) that returns \text{true}; \( \pi_2 \rightarrow \tilde{S}^{k+1} \). Suppose by contradiction that such a \( \pi_1 \) and \( \pi_2 \) exist. Thus, the action in line 37 by the \text{insert}(v) operation that updates some element, say \( X \) precedes the action in line 12 by \text{contains}(v) that is associated with its first \text{read} (the \text{head}). We claim that \text{contains}(v) must read the element \( X' \) newly created by \text{insert}(v) and return \text{true}—a contradiction to the initial assumption that it returns \text{false}. The only case when this can happen is if there exists a \text{remove} operation that forces \( X' \) to be unreachable from \text{head} i.e. concurrent to the \text{write} to \( X \) by \text{insert}, there exists a \text{remove} that sets \( X'' \rightarrow \text{next} \) after the action in line 37 by \text{insert}. But this is not possible since the \text{cas} on \( X \) performed by the \text{remove} would return \text{false}.

Thus, inductively, the sequence of state transitions in \( \tilde{S} \) satisfies the sequential specification of the \text{set} type.

\[ \Box \]

Lemmas 9 and 10 imply:

**Theorem 11** \( I^{\text{RM}} \) is LS-linearizable with respect to (LL, set).

### E Complementary experiments

To confirm the practicality of our highly concurrent optimistic list-based set algorithm we compared its performance against the state-of-the-art list-based set synchronized with hand-over-hand locking (or lock coupling). To this end, we implemented the pseudocode of Algorithm 1 in Java without further optimizations and compared it against the Java code from Herlihy and Shavit of the hand-over-hand (or lock coupling). To this end, we implemented the pseudocode of Algorithm 1 in Java without further optimizations and compared it against the Java code from Herlihy and Shavit of the hand-over-hand lock-based linked list [22].

Figure 4 gives the throughput as the number of operations per millisecond by having from 1 to 64 threads running between 5% and 20% of updates (either remove or insert with same probability) and the rest of contains operations. The list is initially populated with 512 values that are integers taken from 0 to 1024 with uniform distribution. The machine has 2 8-core Intel Xeon E5-2450 running at 2.1GHz (32-way as each code is hyperthreaded). Java is 1.7.0 55 and the JVM is the OpenJDK 64-Bit Server VM. Each point of the graph results from the average of 10 runs of 5 seconds plus 5 seconds to warmup the JVM.

We can observe that our optimistic algorithm \( I^{\text{RM}} \) outperforms, in most cases, the list based set synchronized with hand-over-hand locking (HOHL). This performance is due to the optimal concurrency of our algorithm (all correct schedules are accepted as shown by Theorem 3) and the low overhead of our algorithm (a valid schedule is efficiently identified and a constant number of \text{cas} are needed, only during update operations). We also observe that the peak performance of E-STM is better than the peak performance of \( I^{\text{RM}} \) after 10% updates. This is due to the fact that our code, kept intentionally simple, is not optimized with partial aborts, hence it always restarts from the beginning of the list upon abort. By contrast, E-STM is optimized to re-read only one node upon some conflict detection [10]. The optimal concurrency of our algorithm makes it scale despite contention whereas E-STM does not scale even starting at 5% updates. Worth noting is that there is a large body of work on concurrent list-based set algorithms, and we are not claiming our algorithm to be the most efficient. Algorithms that are not optimal with respect to concurrency can achieve better results on some workload with a lower overhead. An interesting question is how far can our implementation be optimized. In particular, we know that partial aborts and wrappers inlining could boost the performance of our algorithm while retaining its concurrency optimality.
Figure 4: Performance of the list-based set synchronized with our Algorithm $I^PM$, with hand-over-hand locking (HOHL) and with elastic transactions (E-STM)