1991

J.T. Tromp, P.M.B. Vitányi

Randomized wait-free test-and-set

Computer Science/Department of Algorithmics and Architecture Report CS-R9113 March

CWI, nationaal instituut voor onderzoek op het gebied van wiskunde en informatica

The Centre for Mathematics and Computer Science is a research institute of the Stichting Mathematisch Centrum, which was founded on February 11, 1946, as a nonprofit institution aiming at the promotion of mathematics, computer science, and their applications. It is sponsored by the Dutch Government through the Netherlands Organization for the Advancement of Research (N.W.O.).

Randomized Wait-Free Test-and-Set

John Tromp, Paul M.B. Vitányi*

Centre for Mathematics and Computer Science
P.O.Box 4079, 1009 AB Amsterdam, The Netherlands

It is known to be impossible to implement wait-free test-and-set deterministically in a concurrent setting using only atomic shared variables. We present the first explicit *randomized* algorithm for *any* wait-free concurrent object: a test-and-set bit shared between 2 processes, implemented with two 4-valued single writer single reader atomic variables. The worst-case (over all adversary schedulers) expected number of steps to execute a test-and-set between two processes is at most 11, while the reset takes exactly 1 step. Based on a finite-state analysis, the proofs of correctness and expected length are compressed into one table.

1980 Mathematics Subject Classification: 68C05, 68C25, 68A05, 68B20.
CR Categories: B.3.2, B.4.3, D.4.1, D.4.4.
Keywords and Phrases: Test and Set, Randomized, Shared variable (register), atomicity, verification, automata.
Note: This paper is submitted for publication elsewhere.

1 Introduction

A concurrent system consists of n processes communicating through concurrent data objects $R_0, \ldots, R_{m-1}$. An implementation of a new concurrent data object x (rather, a family of objects, one for each n) is *wait-free* if there is a total function f, such that each process can complete any operation associated with x within $f(n)$ steps, irrespective of the timing and execution speeds of the other processes[1]. Here we take a step to be a single access to one of the R_i's. Local events, including coin-flips, are not counted. The paper develops a similar definition of wait-freeness for randomized protocols based on the worst case expected length of an operation.

A concurrent object is *constructible* if it can be implemented deterministically with boundedly many safe bits, the mathematical analogues of electronic hardware 'flip-flops', [16]. What concurrent wait-free object is the most powerful constructible one? It has been shown

*also: Faculteit Wiskunde en Informatica, Universiteit van Amsterdam. email: tromp/paulv@cwi.nl

[1] This is called *bounded* wait-free in the terminology of [12]

Report CS-R9113
Centre for Mathematics and Computer Science
P.O.Box 4079, 1009 AB Amsterdam, The Netherlands

that wait-free atomic multi-user variables, and atomic snapshot objects, are constructible [23, 16, 29, 18, 3, 2]. In contrast, wait-free consensus—viewed as an object on which each of n processes can execute just one operation—is not constructible, although randomized implementations are possible [13, 1]. Wait-free concurrent 2-process test-and-set can deterministically implement 2-process wait-free consensus, and therefore is not deterministically constructible [17, 11, 12, 15]. This raises the question of whether randomized algorithms for test-and-set exist.

1.1 Results

The main results in this paper are:

1. We present the first explicit randomized algorithm for wait-free concurrent test-and-set between 2 processes, directly implemented in boundedly many, bounded, atomic read/write variables Randomization means that the algorithm contains a branch conditioned on the outcome of a fair coin flip (as in [27]). This is the first randomized algorithm for *any* concurrent wait-free re-usable object. The use of randomization necessitates a re-defenition of wait-free-ness, based on the expected number of steps an operation takes to complete, against an adversary scheduler. The latter is conveniently defined as a restricted probability distrubution on possible histories.

2. We develop a finite-state based proof technique for verifying correctness and worst-case expected execution length. This part is delegated to the Appendix.

3. The solution uses two 4-valued 1-writer 1-reader atomic variables. The worst-case expected number of steps for a test-and-set execution is 11. Reset always takes 1 step.

1.2 Comparison with Related Work

Implementations of multi-user variables are given in [23, 16, 29, 26, 8, 7, 25, 18]. Even more powerful 'snapshot' objects can be implemented in multi-user variables [3, 2]. Impossibility of deterministic implementations of test-and-set in terms of multi-user variables is shown in [4, 17, 11, 15]. In [9, 1, 5], randomized algorithms for wait-free consensus using atomic shared variables are given. For clarity on the issues involved, an implementation of a *reusable* object like 'test-and-set' yields an implementation of a *single-use* object like 'consensus', but not vice versa. In [11] a proper hierarchy of wait-free concurrent objects is established. Objects can be implemented by non-randomized algorithms in terms of objects higher in the hierarchy, but not in terms of lower objects. For instance, test-and-set is higher than multi-writer variables. Universal objects (like fetch-and-cons) on top of the hierarchy are identified, and it is established that the ability to achieve consensus between n processes is necessary and suffices to implement wait-free universal objects, *provided* an *infinite array* of the hardware required for a single concensus bit is used. That is, each operation execution uses a separate hardware implementation of a consensus bit. Here we are interested in constructible solutions (finite hardware). In [24] such a solution is claimed to result by combining several intermediate constructions, but details are not given.

2 Preliminaries

A *test-and-set* bit is a concurrent data object X shared between n processes $0, \ldots, n-1$. The value of X is 0 or 1. Each process i has a local binary variable x_i. At any time exactly one of $X, x_0, \ldots, x_{n-1}$ has value 0, all others have value 1. A process i with $x_i = 1$ can atomically execute a test-and-set operation

read $x_i := X$; write $X := 1$; return x_i.

A process i with $x_i = 0$ can execute a reset operation

$x_i := 1$; write $X := 0$.

This naturally leads to the definition of the *state* of the test-and-set bit, or *0-owner* as a member of $\{\perp, 0, \ldots, n-1\}$ according to which of X and the x_i's is 0.

2.1 Test-and-Set Definition

Instead of assuming an atomic test-and-set, we want to implement it with actions that are sequences of accesses to atomic shared variables, $R_0, \ldots, R_{m-1}$, executed by processes $0, \ldots, n-1$, according to some *protocol*. Since the executions by the different processes happen concurrently and asynchronously, the implementation should guarantee that each system execution of the implementation is equivalent to a system execution of the above defined construct. This leads to the following set of definitions.

Definition. For a given execution of the system, denote the set of *actions* that have been started as $A = R \cup T, T = T0 \cup T1$, where R is the set of *resets*, and Tx is the set of *test-and-sets* returning x, $x = 0, 1$. By $r, t, t0, t1$ we denote elements from $R, T, T0, T1$, respectively. We partition the set of actions according to the processes executing them: define A_i to be the actions by process i, and similarly define $R_i = A_i \cap R$, and $Tx_i = A_i \cap Tx$, $x = \epsilon, 0, 1$. Define an *event* as an execution of a statement in a protocol, that is, a write or a read on a R_i or a coin-flip. A read event is qualified by the value obtained and a coin-flip by its outcome. Every new execution of a statement represents a unique event. Number the events of a test-and-set or reset action $a \in A$ as $a.1, a.2, \ldots$. Let $l = l(a)$, the *length* of a, be its number of events in the execution (possibly infinite). Denote $a.1$, the *start* of a, as $s(a)$. If a finishes during the execution, then denote $a.l$, the *finish* of a, as $f(a)$. Each event is assumed to execute *atomically*. The sequence of the events of all actions in A is called the *history*. A history induces a partial ordering of the actions: $a \to b$ iff $f(a) < s(b)$ (the last event of a precedes the first of b in the history). The number of b such that $b \to a$ is assumed to be finite for each a.

The pair $(A, \to)$ is called a *run*. An implementation of a concurrent object shared between processes $0, \ldots, n-1$, such that each run $(A, \to)$ satisfies the following atomicity axiom, is an atomic test-and-set.

Atomicity: We can extend $\to$ on A to a total order $\Rightarrow$ on A in which the sequence of actions satisfies the test-and-set semantics:

1. the system is initially in state $\perp$.

2. from state $\perp$, an action $t0 \in T0_i$ moves the system to state i.

3. from state i, an action $r \in R_i$ moves the system to state $\perp$.

4. from state i, an action $t1 \in T1 - T1_i$ leaves the system in state i.

5. no other state transitions than the above are allowed.

2.2 Randomization, Adversaries and Wait-Freeness

In the above definition of atomicity we did not use the notion of adversary. The reason is that atomicity must hold for all possible histories, and hence for all possible outcomes of coin-flips. The adversary is introduced to enable a quantification of the *wait-freeness*. While it is inevitable that for some histories a test-and-set action may last arbitrarily many steps, the probability of such histories occurring should be minimized. This leads us to define the probability of a certain history occurring.

Fix a protocol P. Let H (H^∞) be the set of finite (infinite) histories that can arise from this protocol. I.e. the set of h such that

1. for all $i < n$, $h|A_i$, the restriction of h to events by process i, satisfies the protocol for process i, and

2. for all $j < m$, $h|R_j$, the restriction of h to events that access R_j, satisfies the usual semantics of such an atomic variable (a read event returns the value written by the last write event preceding it).

For $h \in H$, let the *cylinder* Γ_h be the set of all histories in H^∞ that start with h. Write he to denote history h followed by event e.

An *adversary* is then defined as a probability measure μ on H^∞ satisfying:

1. $\mu(\Gamma_\epsilon) = 1$, where ϵ is the empty history;

2. $\mu(\Gamma_h) = \sum_{h\,e \in H} \mu(\Gamma_{h\,e})$, for $h \in H$ and e is a single event; and

3. $\mu(\Gamma_{h\,c(\text{heads})}) = \mu(\Gamma_{h\,c(\text{tails})})$, for each coin-flip event $c()$ with $h\,c() \in H$.

The first two conditions—already implied by the notion of probability measure—are included for completeness. The third condition ensures that the adversary has no control over the outcome of a fair coin flip: both outcomes are equally likely. This definition is readily generalized to biased coins and multi-branch decisions.

Note that this notion of adversary is the strongest possible short of allowing it to predict the future. For example, it includes nonrecursive adversaries using omniscient oracles and randomization.

Now that adversaries have been defined, we can define the expected length $E(h,i)$ of process i's current (next if idle) action following a finite initial history segment h. Let $\omega \in \Gamma_h$ be an infinite history starting with h. Let $l_{h,i}(\omega)$ be the length (number of events) of process i's current action following h in ω. If process i is idle at h, then by 'current' we mean 'next', leaving $l_{h,i}$ undefined for ω in which a doesn't start a new action. Define

$$E(h,i) = \sum_{k=1}^{k=\infty} k \cdot \frac{\mu(\{\omega \in \Gamma_h : l_{h,i}(\omega) = k\})}{\mu(\Gamma_h)}.$$

The summation includes the case $k = \infty$ so that the expected length is infinite if (but not necessarily only if) the set of infinite histories in which an operation execution has infinitely many events, has positive measure. The normalization w.r.t. h gives the adversary a free choice of 'starting' configuration.

Definition. An implementation of a concurrent object shared between n processes is *wait-free*, if there is a constant $f(n)$ bounding the expected length $E(h, i)$, for all h, i, under all adversaries.

3 Solution for Two Processes

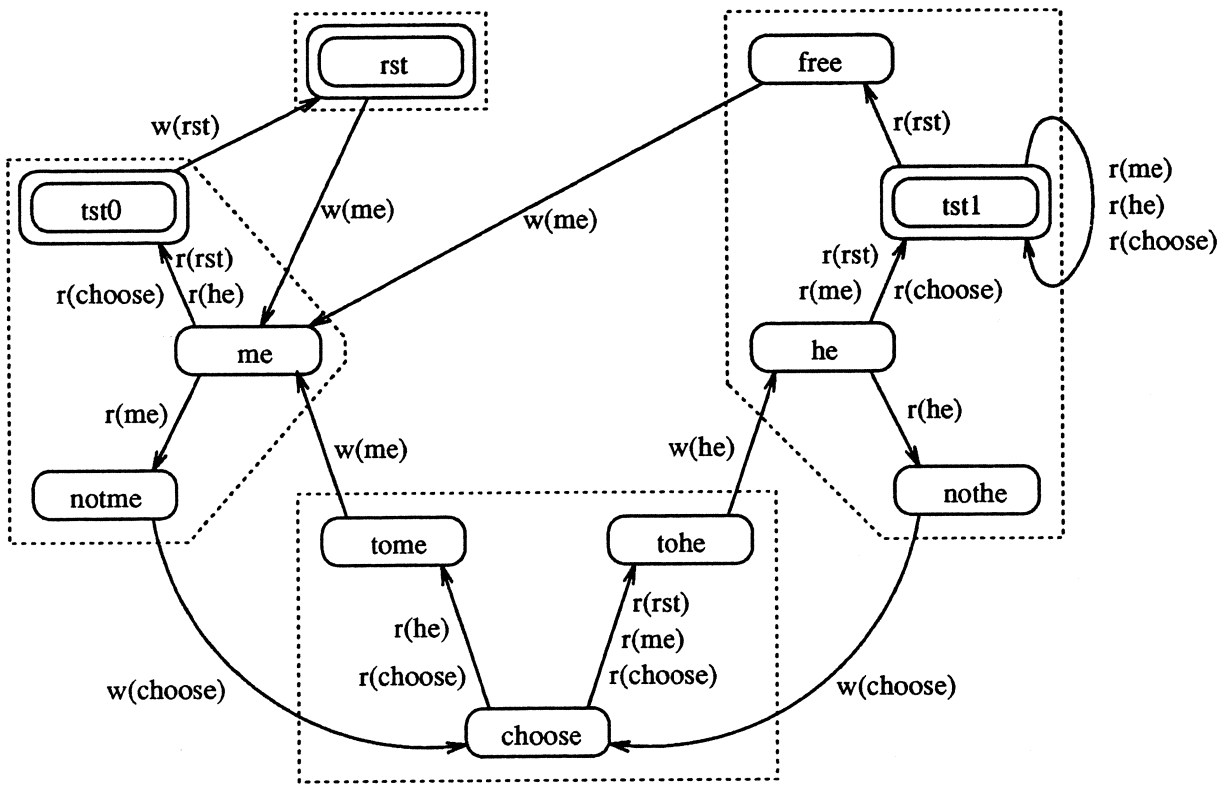

FIGURE 1. test-and-set protocol

We give a test-and-set implementation between two processes, process 0 and process 1. The construction uses two 4-valued shared variable objects, R_0 and R_1. The four values are 'me', 'he', 'choose', 'rst'. Process i solely writes variable R_i, its *own* variable, and solely reads R_{1-i}. For this reason the reads and writes in the protocol don't need to be qualified by the shared variables they access. The protocol, for process i, is first presented as a finite state transition diagram, figure 1. The transitions are labeled with reads $r(value)$ and writes $w(value)$ of the shared variables, where *value* denotes the value read or written. The 11 states of the protocol are split into 4 groups enclosed by dotted lines. Each group is an equivalence class consisting of the set of states in which that process's variable has the same value. That is, the states in a group are equivalent in the sense that process $1 - i$ cannot distinguish between them by reading R_i. Accordingly, the inter-group events are writes to R_i, whereas the intra-group events are reads of R_{1-i}. Each group is named after the corresponding value of the shared variable. The diagram is deterministic, but for a coin flip which is modelled by the two $r(choose)$ transitions from the choose state.

A more conventional representation of the protocol, for process i, is given below. An occurrence of R_i not preceded by 'write' (resp. R_{1-i} not preceded by 'read') refers to the last value written to it (resp. read from it), stored in correspondingly named local variables. The conditional 'rnd(true,false)' represents the boolean outcome 'true' or 'false' of a fair coin flip. The system is initialized with all local and global variables in state rst.

```
test_and_set:

if R_i = he AND read R_{1-i} ≠ rst
then return 1
write R_i := me
while read R_{1-i} = R_i do
  write R_i := choose
  if read R_{1-i} = he OR (R_{1-i} = choose AND rnd(true,false))
  then write R_i := me
  else write R_i := he
if R_i = me
then return 0
else return 1

reset:

write R_i := rst
```

It can be verified in the usual way that the transition diagram represents the operation of the program. The intuition behind the protocol is as follows. The default situation is where both processes are idle in the rst state. If process i starts a test-and-set then it writes $R_i :=$ me (indicating its desire to take the 0), and checks whether process $1-i$ agrees (by *not* having $R_{1-i} :=$ me). If so, then it has succesfully completed a test-and-set of 0. It is easy to see that in this case process $1-i$ can not get 0 until process i does a reset by writing $R_i :=$ rst. While $R_i =$ me, process $1-i$ can only move from state 'me' to state 'notme' and on via states 'choose', 'tohe' and 'he' to 'tst1', where it completes a test-and-set of 1.

Problems arise only if both processes see each other's variable equal to 'me'. In this case they are said to *disagree* or *in conflict.* They then proceed to the choose state from where they decide between going for 0 or 1, according to what the other process is seen to be doing. (It is essential that this decision be made in a neutral state, i.e. without a claim of preference for either 0 or 1. If, for example, on seeing a conflict, a process would change preference at random, then a process cannot know for sure whether the other one agrees or is about to write a changed preference.)

The deterministic choices, those made if the other's variable reads different from 'choose', can be seen to lead to a correct resolution of the conflict. A process ending up in the tst1 state makes sure that its test-and-set of 1 is justified, by remaining in that state until it can be sure that the other process has taken the 0. Only if the other process is seen to be in the rst state need it try to take the 0 itself.

Suppose now that process i has read R_{1-i} = choose and is about to flip a coin. Assume that process $1-i$ has already moved to one of the states tome/tohe (or else reason with the processes interchanged). With 50 percent chance, process i will move to the same state as process $1-i$ did and thus the conflict will be resolved.

So, intuitively, the probability of each loop through the choose state is at most one half and the expected number of 'choices' (transitions from state choose) at most two. This shows that the worst case expected test-and-set length is 11. Namely, starting from the tst1 state, it takes 4 steps to get to state choose, another 4 steps to loop back to choose and 3 more steps to reach tst0/tst1. The reset operation always takes 1 step.

In the Appendix, the construction will be proven correct rigorously, together with the upperbounds.

4 On the Difficulty of Multi Process Test And Set

The obvious way to extend the given solution to more than 2 processes would be to arrange them at the leafs of a binary tree. Then, a process wishing to execute an n-process test-and-set, would enter a tournament, as in [21], by executing a separate 2-process test-and-set for each node on the path up to the root. When one of these fails, it would again descend, resetting all the tas-bits on which it succeeded, and return 1. When it succeeds ascending up to the root, it would return 0 and leave the resetting descend to its n-process reset.

The intuition behind this tree approach is that if a process i fails the test-and-set at some node N, then another process j will get to the root succesfully and thus justify the value 1 returned by the former.

The worst case expected length of the n-process operations is only $\log n$ time more than that of the 2-process case.

Unfortunately, this straightforward extension does not work. The problem is that the other process j need not be the one responsible for the failure at node N, and might have started its n-process test-and-set only after process i completes its own. Clearly, the resulting history cannot be linearized.

We conjecture that, as in the case of n-writer read/write variables, their exists a large gap between the complexity of a solution for the case $n = 2$ versus $n > 2$.

References

[1] K. Abrahamson, *On achieving consensus using shared memory*, Proc. 7th ACM Symposium on Principles of Distributed Computing, 1988, pp. 291–302.

[2] Y. Afek, H. Attiya, D. Dolev, E. Gafni, M. Merritt, N. Shavit, *Atomic snapshots of shared memory*, Proc. 9th ACM Symposium on Principles of Distributed Computing, 1990, pp. 1-13.

[3] J.H. Anderson, *Composite registers*, Proc. 9th ACM Symposium on Principles of Distributed Computing, 1990, pp. 15-29.

[4] J.H. Anderson, M.G. Gouda, *The virtue of patience: concurrent programming with and without waiting*, Unpublished Manuscript, Comp. Sci. Dept. University of Texas, Austin, 1987.

[5] J. Aspnes, *Time- and space-efficient randomized consensus*, Proc. 9th ACM Symposium on Principles of Distributed Computing, 1990, pp. 325-331.

[6] B. Awerbuch, L. Kirousis, E. Kranakis, P.M.B. Vitányi, *A proof technique for register atomicity*, Proc. 8th Conference on Foundations of Software Technology & Theoretical Computer Science, Lecture Notes in Computer Science, vol. 338, pp. 286–303, Springer Verlag, 1988.

[7] B. Bloom, *Constructing Two-writer Atomic Registers*, IEEE Transactions on Computers, vol. 37, pp. 1506–1514, 1988.

[8] J.E. Burns and G.L. Peterson, *Constructing Multi-reader Atomic Values From Nonatomic Values*, Proc. 6th ACM Symposium on Principles of Distributed Computing, pp. 222–231, 1987.

[9] B. Chor, A. Israeli, M. Li, *On processor coordination using asynchronous hardware*, Proc. 6th ACM Symposium on Principles of Distributed Computing, pp. 86–97, 1987.

[10] D. Dolev and N. Shavit, *Bounded Concurrent Time-Stamp Systems Are Constructible*, Proc. 21th ACM Symposium on Theory of Computing, pp. 454–466, 1989.

[11] M.P. Herlihy, *Impossibility and Universality Results for Wait-Free Synchronization*, Proc. 7th ACM Symposium on Principles of Distributed Computing, 1988, pp. 276–290.

[12] M.P. Herlihy, *Wait-Free Synchronization*, ACM Trans. Progr. Lang. Syst., 1991, to appear.

[13] A. Israeli and M. Li, *Bounded Time-Stamps*, Proc. 28th IEEE Symposium on Foundations of Computer Science, pp. 371–382, 1987.

[14] L.M. Kirousis, E. Kranakis, P.M.B. Vitányi, *Atomic Multireader Register*, Proc. 2nd International Workshop on Distributed Computing, Amsterdam, J. van Leeuwen (ed.), Springer Verlag Lecture Notes in Computer Science, vol. 312, pp. 278–296, July 1987.

[15] E. Kranakis, *Functional dependencies of variables in wait-free programs*, Proc. 3rd Int. Workshop on Distributed Algorithms, Lecture Notes in Computer Science, vol. 392, Springer Verlag, pp. , 1989.

[16] L. Lamport, *On Interprocess Communication Parts I and II*, Distributed Computing, vol. 1, pp. 77–101, 1986.

[17] M. Loui, H.H. Abu-Amara, *Memory requirements for agreement among unreliable asynchronous processes*, pp. 163–183 in: Advances in Computing Research, JAI Press, 1987.

[18] M. Li, J. Tromp, P.M.B. Vitányi, *How to construct concurrent wait-free variables*, Tech. Rept. CS-8916, CWI, Amsterdam, April 1989. Previous versions: M. Li, P.M.B. Vitányi, *A very simple construction for atomic multiwriter register*, Techn. Report TR 01-87, Aiken

Comp. Lab., Harvard University, November 1987. Published as pp. 488-505 in: *Proc. International Colloquium on Automata, Languages, and Programming*, Lecture Notes in Computer Science, Vol. 372, Springer Verlag, 1989.

[19] N. Lynch and M. Tuttle, *Hierarchical correctness proofs for distributed algorithms*, Proc. 6th ACM Symposium on Principles of Distributed Computing, 1987.

[20] R. Newman-Wolfe, *A Protocol for Wait-Free, Atomic, Multi-Reader Shared Variables*, Proc. 6th ACM Symposium on Principles of Distributed Computing, pp. 232–248, 1987.

[21] G.L. Peterson, M. Fischer, *Economical solutions for the critical section problem in a distributed system*, Proc. 9th ACM Symp. on Theory of Computing, pp. 91-97, 1977.

[22] G.L. Peterson and J.E. Burns, *Concurrent reading while writing II: the multiwriter case*, Proc. 28th IEEE Symposium on Foundations of Computer Science, pp. 383–392, 1987.

[23] G.L. Peterson, *Concurrent reading while writing*, ACM Transactions on Programming Languages and Systems, vol. 5, No. 1, pp. 46–55, 1983.

[24] S. Plotkin, *Sticky bits and universality of consensus*, Proc. 8th ACM Symposium on Principles of Distributed Computing, pp. 159-175.

[25] R. Schaffer, *On the correctness of atomic multi-writer registers*, Technical Report MIT/LCS/TM-364, MIT lab. for Computer Science, June 1988.

[26] A.K. Singh, J.H. Anderson, M.G. Gouda, *The Elusive Atomic Register Revisited*, Proc. 6th ACM Symposium on Principles of Distributed Computing, pp. 206–221, 1987.

[27] M.O. Rabin, *The choice coordination problem.* Acta Informatica, vol. 17, pp. 121–134, 1982.

[28] K. Vidyasankar, *Converting Lamport's Regular Register to an atomic register*, Information Processing Letters, vol. 28, pp. 287–290, 1988.

[29] P.M.B. Vitányi, B. Awerbuch, *Atomic Shared Register Access by Asynchronous Hardware*, Proc. 27th IEEE Symposium on Foundations of Computer Science, pp. 233–243, 1986. (Errata, Ibid.,1987)

5 Appendix: Proof of the 2-Process Solution

Let h be a history corresponding to a run $(A, \rightarrow)$ of our implementation. Let $B = \{s(t), f(t) : t \in T\} \cup \{s(r) = f(r) : r \in R\}$ be the set of events which start or finish an action. Note that $h|B$, the restriction of h to events in B, completely determines the partial order of actions $\rightarrow$. Let $C = \{t* : t \in T\} \cup R$ be the set of atomic occurences of actions.

The definition of atomic test-and-set for 2 processes, process 0 and process 1, is captured by DFA2, the DFA in figure 2, which accepts all possible sequences of atomic operations (all states final). The states are labeled with the owner of the bit. The arcs representing actions of process 1 are labeled, whereas the non-labeled arcs represent the corresponding actions of process 0.

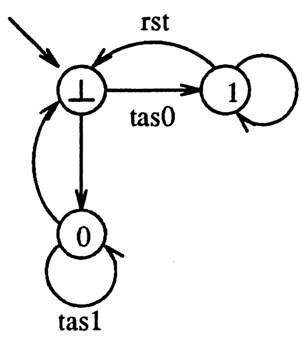

FIGURE 2. DFA2: specification of 2-process atomic test-and-set

Figure 3 shows the DFA, DFA3, that accepts the possible sequences of the following events of one process (all states final):

- the start of a test-and-set action, denoted s(tas),
- the atomic occurrence of a test-and-set 0, denoted tas0,
- the atomic occurrence of a test-and-set 1, denoted tas1,
- the finish of a test-and-set 0 action, denoted f(tas0),
- the finish of a test-and-set 1 action, denoted f(tas1),
- the reset action, denoted rst.

These are the events in $B \cup C$. The reason for not splitting a reset action into start, atomic occurrence, and finish is that it's implemented in our protocol as a single atomic write where the above three transitions coincide.

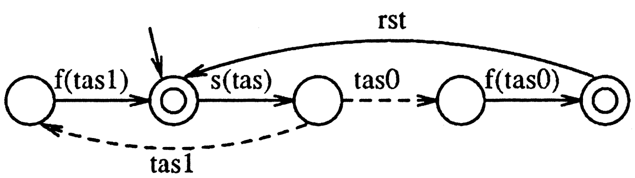

FIGURE 3. DFA3: non-atomic specification of 1-process test-and-set

The proof is based on the finite state diagram DFA4 in figure 4 below (again all states are final).

It is drawn as a cartesian product of the two component processes—transitions of process 0 are drawn vertically and those of process 1 horizontally. For clarity, the transition names are only given once and only for process 1. Identifying the starts and finishes of test-and-set executions with their atomic occurrences by collapsing the $s()$ and $f()$ arcs, the diagram reduces to the atomic test-and-set diagram. Identifying all nodes in the same column (row) reduces the diagram to the diagram of process 0 (process 1).

In the states labeled 'a' through 'h', neither process owns the 0; the bit is in state $\perp$. In the states labeled 'i' through 'n', process 1 owns the 0; the bit is in state 1. In the states labeled 'o' through 't', process 0 owns the 0; and the bit is in state 0.

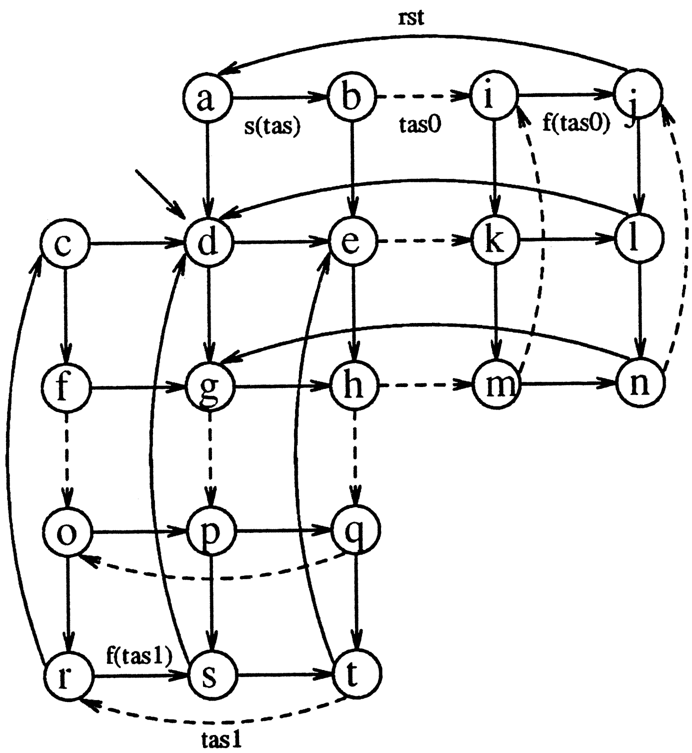

FIGURE 4. DFA4: non-atomic specification of 2-process test-and-set

Formally [19], DFA4 is the composition of DFA2 with 2 copies of DFA3, in the I/O Automata framework.

Let NFA4 be the NFA obtained from DFA4 by turning the broken transitions of figure 4 into ϵ-steps.

We claim that acceptance of $h|B$ by NFA4 implies atomicity of $(A, \rightarrow)$. This is proven as follows. If NFA4 accepts $h|B$, then, corresponding to the ϵ transitions, we can augment $h|B$ with an atomic transition $t*$ between the start $s(t)$ and finish $f(t)$ of each test-and-set action $t \in T$, to get a history h' accepted by DFA4. Therefore, DFA2, which composes DFA4, accepts $h'|C$,the sequence of atomic events in h'. Furthermore, if $a \rightarrow b$, then $a* \leq f(a) \rightarrow s(b) \leq b*$, so $\Rightarrow$, the total order of actions in $h'|C$, extends $\rightarrow$. This proves atomicity of $(A, \rightarrow)$.

To show that for all histories $h \in H$ of our implementation, $h|B$ is accepted by NFA4, and thus the correctness of our construction, we assign to each reacheable combination of process states (s_0, s_1) a nonempty set S_{s_0,s_1} of NFA4 states, such that: for each history h ending in process states (s_0, s_1), the set of states in which NFA4 can be after processing $h|B$ contains S_{s_0,s_1} (*). The assignment is given in figure 5. The table entries were chosen so as to minimize the number of ϵ-steps that can be made from each assigned set of NFA4 states. This gives the most insight into the workings of the protocol.

In the table below each row (column) is labeled with a state of process 1 (process 0) as in diagram figure 1. An entry in the table is labeled with roman letters representing a set of atomicity states in figure 4, assigned to that row/column pair of process states. The number ending an entry gives $E(h, 0)$, the expected number of steps to finish the current operation execution of process 0.

We use induction on the length of the history to check (*):

Base: After processing the empty history, NFA4 can be in $\{initialstate\} \supseteq \{d\} = S_{rst,rst}$.

Induction Step: This reduces to checking whether for all transitions (s_0, s_1) to (t_0, t_1) and all NFA4 states $y \in S_{t_0,t_1}$, there is an NFA4 state $x \in S_{s_0,s_1}$, such that NFA4 can move from x to y by processing: either the event corresponding to the transition if it belongs to B, or no event otherwise (there is a sequence of ϵ-steps from x to y).

	rst	tst0	notme	me	tome	choose	tohe	he	nothe	tst1	free
rst	d10	l10	cek10	ek10	ek10	c10	c10	c10	c10	d10	ek10
tst0	s1	*	rt1	rt1	rt1	r1	r1	r1	r1	s1	rt1
notme	agp8	jn8	imoq8	imoq8	*	imoq8	imoq8	o4	*	p4	*
me	gp9	jn9	imoq9	imoq9	imoq9	o1	o1	o1	o1	p1	imoq9
tome	gp10	jn10	*	imoq10	imoq10	imoq6	o2	o2	imoq6	p2	*
choose	a3	j3	imoq7	i3	imoq7	imoq7	imoq7	o3	imoq7	p3	*
tohe	a2	j2	imoq6	i2	i2	imoq6	imoq10	imoq10	*	p6	*
he	a1	j1	i1	i1	i1	i1	imoq9	imoq9	imoq9	p5	*
nothe	a4	j4	*	i4	imoq8	imoq8	*	imoq8	imoq8	p4	*
tst1	d11	l11	k11	k11	k11	k11	k11	k11	k11	*	*
free	gp10	jn10	*	imoq10	*	*	*	*	*	*	*

FIGURE 5. table to verify correctness and wait-freeness

It is straightforward to check all transitions (state process 0, state process 1) to (newstate process 0, state process 1) or to (state process 0, newstate process 1), corresponding to the atomic transitions in the two copies of protocol figure 1 concerned, to do the induction on the length of the runs to verify correctness, explained above. Simultaneously, the wait-freeness can be checked.

We give an example of checking a few transitions below, and give the interpretation. Verification consists in checking all transitions in the table.

In the default state both processes are in state rst. The table entry $d10$ gives corresponding state d, the start state, in figure 4. The worst-case expected number of steps for a test-and-set by process 0 is 10. Process 0 can start a test-and-set by executing $w(me)$ and entering state me. The corresponding table entry $gp9$ indicates in figure 4 that the system is now either in state g meaning that process 0 has executed $s(tas)$, or in state p meaning that process 0 has executed $s(tas)$ and also $tas0$ atomically. The expected number of steps is now $9 \leq 10 - 1$. Suppose process 1 now starts a test-and-set: it executes $w(me)$ and moves to state me. The corresponding table entry $imoq9$ gives the system state as one possibility in $\{i, m, o, q\}$ in figure 4 and the expected number of steps for execution of test-and-set by process 0 is still 9. State m says process 1 has executed $s(tas)$ and $tas0$ atomically, while process 0 has only executed $s(tas)$—hence the system was previously in state g and not in state p. State i says process 1 has executed $s(tas)$ and $tas0$ atomically, while process 0 has executed $s(tas)$ and $tas1$ atomically—and hence the system was previously in state g and not state p. States o and q imply the same state of affairs with the roles of process 0 and process 1 interchanged, and the previous system state is either p or g.

Note that it is also consistent for the system to be in state h—neither process having executed tas. However, if both processes have started a test-and-set execution, then necessarily, one of them must return 0. We have optimized the table entries by eliminating such spurious states.

Process 0 might now read $R_1 = me$, and move via state notme (table entry *imoq*8) by writing $R_0 := choose$, to state choose. Process 1 is idle in the meantime. The table entry is now *i*3. This says that process 1 has atomically executed *tst*0, and process 0 has atomically executed *tst*1. Namely, all subsequent schedules lead in 3 steps of process 0 to state tst1—hence the expectation 3.

The expected number of remaining steps of process 0's test-and-set has dropped from 8 to 3 by the last step since 8 was the worst-case which could be forced by the adversary. Namely, from the system in state $(notme, me)$, the adversary can schedule process 1 to move to $(notme, notme)$ with table entry *imoq*8, followed by a move of process 1 to state $(notme, choose)$ with table entry *imoq*8, followed by a move of process 0 to state $(choose, choose)$ with table entry *imoq*7. Suppose the adversary now schedules process 0. It now flips a fair coin to obtain the conditional boolean $rnd(true, false)$. If the outcome is *true*, then the system moves to state $(tome, choose)$ with entry *imoq*6. If the outcome is *false*, then the system moves to state $(tohe, choose)$ with table entry *imoq*6. Given a fair coin, this step of process 0 correctly decrements the expected number of steps. Suppose the adversary schedules process 1 in state $(choose, choose)$. Process 1 flips a fair coin. If the outcome is *true* the system moves to state $(choose, tome)$ with table entry *imoq*7; if the outcome is *false* then the system moves to state $(choose, tohe)$ with table entry *imoq*7.

This way the correctness of the implementation can be checked exhaustively by hand. We have done the verification by hand, to optimize the entries, and again by machine.

For the finite-state system as we described, the expected number of remaining steps in a test-and-set execution is *always* bounded by a fixed number. The table shows that, trivially, $1 \leq E(h, 0) \leq 11$ Hence the algorithm is wait-free.