

Übungsblatt zu Haskell

Learn You a Haskell for Great Good!

1 Mehr zu Monaden

Aufgabe 1. Umgekehrte polnische Notation

In der umgekehrten polnischen Notation schreibt man einen mathematischen Ausdruck wie

$$(1+2) \cdot 3 - 4 \cdot 5$$

so:

Der Vorteil an dieser Notation ist, dass man keine Klammern und keine Regeln für die Präzedenz von Operatoren ("Punkt vor Strich") benötigt. Die Auswertung eines Ausdrucks in umgekehrter polnischer Notation erfolgt mit einem Stack. Konstanten pushen jeweils ihren Wert auf den Stack. Rechenoperationen poppen die obersten zwei Werte des Stacks und pushen das Ergebnis.

Wir formulieren das in Haskell wie folgt:

- a) Schreibe eine Funktion exec :: Instr a -> Stack a -> Stack a , die eine Instruktion als Argument nimmt und diese ausführt: also einen gegegeben Stack in einen neuen transformiert. Im Fehlerfall (Division durch Null, zu wenige Argumente auf dem Stack) soll die Funktion einfach mit error terminieren.
- b) Schreibe die Funktion nun unter Verwendung der State-Monade um. Der neue Typ soll also exec':: (Num a) => Instr a -> State (Stack a) () sein. Denk daran, das Modul Control.Monad.State zu importieren. Schreibe und verwende Hilfsfunktionen push:: a -> State (Stack a) () und pop:: State (Stack a) a . Ein Fall in der Definition von exec' soll so aussehen:

```
exec' Add = do
    x <- pop
    y <- pop
    push (x + y)</pre>
```

c) Kombiniere die Funktion mapM_ :: (Monad m) => (a -> m b) -> [a] -> m [b] mit deiner Funktion exec' und der Funktion execState aus Control.Monad.State, um einen Interpreter für Ausdrücke in umgekehrter polnischer Notation zu erhalten: eine Funktion vom Typ (Num a) => [Instr a] -> a . (Zurückgegeben werden soll das oberste Element des Stacks.)

Aufgabe 2. Vermeidung von unübersichtlichen Fehlerbehandlungen I

a) Implementiere zum Aufwärmen folgende drei Funktionen.

```
safeHead :: [a] -> Maybe a
safeLast :: [a] -> Maybe a
safeTail :: [a] -> Maybe [a]
```

b) Folgender Code verwendet diese Funktionen, um das erste Element einer gegebenen Liste mit dem vorletzten zu multiplizieren. Bewundere, wie unübersichtlich der Code durch die verschachtelte Fehlerbehandlung ist.

```
ex :: (Num a) => [a] -> Maybe a
ex xs = case safeHead xs of
   Nothing -> Nothing
   Just x -> case safeTail xs of
        Nothing -> Nothing
   Just xs' -> case safeLast xs' of
        Nothing -> Nothing
   Just y -> Just (x * y)
```

c) Schreibe den Code nun mit Hilfe der do-Notation in der Maybe-Monade um! Er soll dann keine Fallunterscheidungen mehr enthalten und etwa so aussehen:

```
ex :: (Num a) => [a] -> Maybe a
ex xs = do
... -- hier irgendwas mit safeHead
... -- hier irgendwas mit safeTail
... -- hier irgendwas mit safeLast
... -- hier irgendwas mit x * y
```

- d) Schreibe eine generische Funktion iterateM :: (Monad m) => Int -> (a -> m a) -> a -> m a.
- e) Kombiniere safeHead, safeTail und iterateM zu einer Funktion safeNth :: [a] -> Int -> Maybe a.

Aufgabe 3. Vermeidung von unübersichtlichen Fehlerbehandlungen II

Unsere übersichtliche Lösung aus der vorherigen Aufgabe hat ein Manko: Im Fehlerfall gibt sie keine Fehlermeldungen aus. Diesen Missstand können wir beheben, indem wir nicht die Maybe-, sondern die Either String-Monade einsetzen.

a) Ändere die Funktionen aus Teilaufgabe a) der vorherigen Aufgabe so ab, dass sie im Fehlerfall mit Left eine Fehlermeldung zurückgeben. Ihre Typen sollen sein:

```
safeHead :: [a] -> Either String a
safeLast :: [a] -> Either String a
safeTail :: [a] -> Either String [a]
```

- b) Ändere analog die Funktion ex aus der zweiten Teilaufgabe ab. Ihr neuer Typ soll (Num a) => [a] -> Either String a sein. Spüre den dabei entstehenden Schmerz.
- c) Bewundere, wie sich die Lösung aus Teilaufgabe c) der vorherigen Aufgabe mühelos auf die neue Situation überträgt. Nur die (ohnehin optionale) Typsignatur muss angepasst werden!

Aufgabe 4. Logging mit der Writer-Monade

Aufgabe 5. Implementierung der Maybe-Monade

Implementiere die Monaden-Instanz von Maybe neu. Vervollständige also folgendes Programm:

```
-- kein 'import Data.Maybe' hier!

data Maybe a = Nothing | Just a deriving (Show,Eq)

instance Functor Maybe where
   fmap f Nothing = ...
   fmap f (Just x) = ...

instance Applicative Maybe where
   pure x = ...
   Nothing <*> _ = ...
   Just f <*> Nothing = ...
   Just f <*> Just x = ...

instance Monad Maybe where
   Nothing >>= g = ...
   Just x >>= g = ...
```

Aufgabe 6. Ein erster Monadenturm

Wir haben gesehen, dass man mit der State-Monade veränderlichen Zustand durchfädeln kann. Mit der Funktion

```
runState :: State s a \rightarrow s \rightarrow (a,s)
```

kann man eine State-Aktion ausführen. Außerdem haben wir gesehen, dass man mit der Maybe-Monade Fehlerfälle übersichtlicher behandeln kann. Gibt es eine Möglichkeit, diese beiden Fähigkeiten zu kombinieren?

Ja, die gibt es! Und zwar baut man sich den *Monadenturm* StateT s Maybe . Man verlässt diese Monade mit der Funktion

```
runStateT :: StateT s Maybe a -> Maybe (a, s) .
```

Dabei ist StateT s ein *Monadentransformer*: Gegeben eine Monade M, so ist StateT s M eine neue Monade, die die Fähigkeiten von M sowie veränderlichen Zustand vom Typ s unterstützt.

Schreibe die Funktion exec' :: (Num a) => Instr a -> State (Stack a) () aus der Aufgabe zur umgekehrten polnischen Notation so um, dass im Fehlerfall nicht mit error eine Ausnahme geworfen wird, sondern Nothing zurückgegeben wird. Der neue Typ soll also

```
exec'' :: (Num a) => Instr a -> StateT (Stack a) Maybe () Sein.
```

```
Tipp. Wegen Typklassenmagie kannst du weiterhin die Funktionen get und put verwenden. Diese haben hier die Typen get :: StateT (Stack a) Maybe (Stack a) und put :: Stack a -> StateT (Stack a) Maybe () .

Mit der Funktion lift :: Maybe r -> StateT (Stack a) Maybe r kannst du Werte aus der Basismonade, in unserem Fall also Maybe , in die transformierte Monade heben. Nützlich wird vor allem lift Nothing :: StateT (Stack a) Maybe sein.
```

Aufgabe 7. Für Unerschrockene: Die Continuation-Monade

XXX: Listen-Monade!

2 Statistisch aussagekräftige Benchmarks

Die Haskell-Community liebt die Bibliothek *Criterion*, um die Laufzeit von Haskell-Programmen zu messen. Diese macht viel mehr, als nur gegebenen Code mehrmals auszuführen und dann die durchschnittliche Laufzeit zu berechnen. Sie bestimmt auch die Standardabweichung der Laufzeit und gibt ein statistisches Maß für die Verlässlichkeit der geschätzten Werte aus.

Ihre Benutzung ist kinderleicht:

- 1. stack install criterion
- 2. Folgende Vorlage anpassen:

```
import Criterion.Main

fib :: Integer -> Integer
fib = ...

main = defaultMain
   [ bgroup "fib"
        [ bench "10" $ whnf fib 10
        , bench "20" $ whnf fib 20
        , bench "30" $ whnf fib 30
        ]
    ]
```

3. Das Programm ausführen. Wenn man dabei die Option --output foo.html übergibt, erstellt die Criterion-Bibliothek eine interaktive HTML-Seite, der man unter anderem die Verteilung der Messwerte entnehmen kann.

3 Erste Schritte mit Nebenläufigkeit

Da es in Haskell keinen veränderlichen Zustand gibt, können Haskell-Ausdrücke in erster Näherung in beliebiger Reihenfolge und auf beliebigen Prozessorkernen ausgewertet werden. GHC verteilt aber nicht von selbst Aufgaben auf mehrere Kerne.

Es gibt in Haskell vier verschiedene Möglichkeiten, Nebenläufigkeit zu erreichen, die man je nach Anwendungszweck einsetzen kann.

• Parallelisierungsannotationen. Puren Code kann man einfach und ohne Umstrukturierung des Programms mit Auswertungsannotationen versehen, wie zum Beispiel "führe das folgende map parallel aus" oder "falls Ressourcen vorhanden sind, beginne die Auswertung des folgenden Ausdrucks im Hintergrund".

- Threads. Wie in anderen Sprachen auch kann man explizit Threads erstellen. Threads können auf diverse Weisen miteinander kommunizieren, zum Beispiel mittels gemeinsamer veränderlicher Variablen (MVar) und Channels (Chan). Dieser recht explizite Zugang zu Nebenläufigkeit ist also sehr ähnlich zum Zugang von anderen Sprachen wie Python oder JavaScript mit Node.js. Anders als in diesen Sprachen gibt es aber keine "Callback-Hölle".
- Shared Transactional Memory (STM). Parallelisierungsannotationen helfen nicht bei Code, der Nebenwirkungen verursachen muss. Wenn man aber auf Threads zurückgreifen würde, müsste man wie in anderen Sprachen auch auf korrektes Locking und Race Conditions achten; das ist mühsam und fehleranfällig. STM ist eine Technik, mit der man vorgeben kann, dass speziell gekennzeichneter Code so abläuft, als wäre das Programm rein sequenziell geschrieben. Der große Vorteil an STM ist Kompositionalität: Man kann Code rein lokal verstehen und kombinieren, ohne auf Auswirkungen von parallel ablaufenden Programmteilen achten zu müssen.
- Data Parallel Haskell (DPH). Dabei kümmern sich der Compiler und die Laufzeitumgebung selbstständig um eine effiziente Verteilung des auszuwertenden Codes. DPH
 ist ein Forschungsprojekt, das noch nicht seinen Weg in die aktuelle GHC-Version
 gefunden hat.

3.1 Parallelisierungsannotationen

Aus dem ersten Workshop ist ja die Funktion seq :: a \rightarrow b bekannt. Wird der Ausdruck seq x y ausgewertet, so wird zunächst x ausgewertet, das Ergebnis verworfen, und dann y zurückgegeben.

Ein Aufruf wie seq 42 y ist nicht besonders sinnvoll. Wenn aber die Auswertung von z die Auswertung von Teilen einer Datenstruktur anstößt, so bleiben die Ergebnisse gespeichert. Die folgende GHCi-Sitzung illustriert das:

```
> let x = fib 30 -- kehrt sofort zurück
> seq x "Hallo" -- dauert lange, da 'x' ausgewertet wird
"Hallo"
> x -- kehrt sofort zurück
832040
```

Nun gibt es neben seq auch die Funktion par :: a -> b -> b aus dem Modul Control.Parallel (aus dem Paket parallel). Semantisch ist par x y identisch zu y . Als Nebenwirkung wird aber ein Spark erzeugt, der x im Hintergrund parallel auswertet.

In GHCi sieht das zum Beispiel so aus:

Wichtig: Standardmäßig verwendet die Laufzeitumgebung nur einen einzigen Betriebssystem-Thread. Damit können keine Sparks im Hintergrund ausgeführt werden. Man muss seinen Code mit der Option -threaded kompilieren und beim Ausführen dem Laufzeitsystem mitteilen, dass es mehrere Betriebssystem-Threads verwenden soll:

```
# Kompilieren mit:
$ ghc --make -02 -threaded Main

# Ausführen mit:
$ ./Main +RTS -N4 -RTS  # genau vier Betriebssystem-Threads verwenden
$ ./Main +RTS -N -RTS  # sinnvolle Anzahl Betriebssystem-Threads verwenden

# Interaktive Shell:
$ ghci +RTS -N -RTS
```

Aufgabe 8. Was bedeutet eigentlich Auswertung?

Erkläre, wieso in folgender GHCi-Sitzung scheinbar b nicht im Hintergrund ausgewertet wird. Denke daran, GHCi mit der Option +RTS -N -RTS zu starten.

```
> import Control.Parallel
> let a = [fib 30]
> let b = [fib 30]
> b 'par' (a,b)
([832040],[832040])
```

Aufgabe 9. Paralleles Map

Schreibe eine Funktion parMap :: (a -> b) -> [a] -> [b], die semantisch identisch zu map ist, aber alle Werte parallel berechnet.

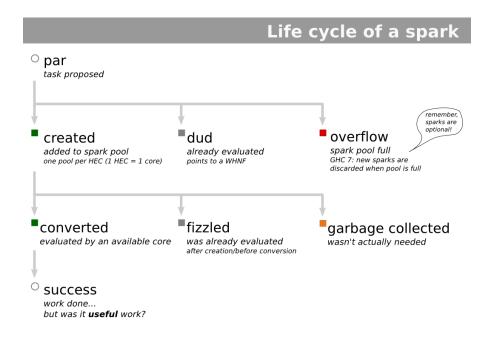
Auf einem Mehrkern-Computer sollte also

```
> parMap fib [30,30,30,30]
```

deutlich schneller ablaufen als map fib [30,30,30,30].

Zu Sparks ist noch viel mehr zu sagen. An dieser Stelle nur zwei Bemerkungen: Startet man sein Programm mit den Optionen +RTS -N -s -RTS, so werden nach Beendigung Statistiken ausgegeben. Diese beinhalten unter anderem, wie viele Sparks erzeugt wurden und wie viele fizzelten – das heißt, dass der zu berechnende Wert schon vom Hauptthread angefordert wurde, noch bevor der Spark loslegen konnte.

Außerdem gibt es *ThreadScope*, mit dem die Auslastung durch Threads und Sparks visualisiert werden kann.



Der Lebenszyklus eines Sparks. Quelle: https://wiki.haskell.org/ThreadScope_Tour/Spark

3.2 Threads

Mit forkIO :: IO a -> IO ThreadId aus dem Modul Control.Concurrent erzeugt man einen leichtgewichtigen Thread. Die übergebene IO-Aktion wird in diesem Thread ausgeführt; Rückgabewert ist ein Wert vom Typ ThreadId, mit dem man den Thread noch nachträglich kontrollieren (etwa vorzeitig beenden) kann.

Die Laufzeitumgebung kommt mit sehr vielen – Millionen – von leichtgewichtigen Threads klar. Sie werden auf eine kleine Anzahl echter Threads auf Betriebssystem-Level verteilt.

In speziellen Anwendungsfällen ist es nötig, Betriebssystem-Threads statt leichtgewichtiger Threads zu erzeugen. Das ist mittels der Funktion forkOS :: IO a -> IO ThreadId ebenfalls möglich.

Wichtig: Wenn man Threads nur verwenden möchte, um mit simultan stattfindenden IO-Aktionen umzugehen (etwa Anforderungen mehrerer gleichzeitig verbundener Clients über das Netzwerk entgegennehmen), genügt prinzipiell ein einzelner Betriebssystem-Thread. Wenn man aber mit Threads tatsächlich auch mehrere Berechnungen parallel ausführen möchte, muss man wie im vorherigen Abschnitt beschrieben seinen Code mit der Option -threaded kompilieren und beim Ausführen dem Laufzeitsystem mit +RTS-N -RTS mitteilen, mehrere Betriebssystem-Threads zu verwenden.

Aufgabe 10. Hallo Welt aus zwei Threads

Schreibe ein Haskell-Programm, das einen leichtgewichtigen Thread erzeugt und den ausführenden Lambroiden sowohl vom Hauptthread als auch dem erzeugten Thread mit putStrLm grüßt.

Aufgabe 11. Vermischte Ausgabe

Schreibe ein Haskell-Programm, das zwei leichtgewichtige Threads erzeugt. Der eine Thread soll tausendmal das Zeichen 'a' ausgeben, der andere das Zeichen 'b'. Was passiert?

Aufgabe 12. Sleep Sort

Implementiere Sleep Sort: Erzeuge für jedes Element \mathbf{x} einer gegebenen Liste von (kleinen) natürlichen Zahlen einen Thread, der sich gleich nach seiner Erstellung für eine zu \mathbf{x} proportionale Zeit schlafen legt und anschließend \mathbf{x} auf dem Terminal ausgibt.

Tipp. Verwende die Funktion threadDelay :: Int -> IO (), die den momentan laufenden Thread für eine gegebene Anzahl Mikrosekunden schlafen legt.

Eine primitive Möglichkeit der Kommunikation zwischen Threads sind (thread-sichere) veränderliche Variablen. Eine solche kann zu jedem Zeitpunkt leer sein oder einen Wert enthalten. Man erstellt sie mit newEmptyMVar :: IO (MVar a) oder, wenn man die Variable gleich initialiseren möchte, mit newMVar :: a -> IO (MVar a).

Mit readMvar :: Mvar a -> 10 a holt man den aktuellen Wert einer übergebenen Variable. Sollte die Variable leer sein, blockiert dieser Aufruf so lange, bis die Variable durch einen anderen Thread gefüllt wird.

Eine Variante ist die Funktion takeMVar :: MVar a -> IO a , die nach Auslesen der Variable diese leert.

Mit putMvar :: Mvar a -> a -> 10 () setzt man den Inhalt einer Variable. Wenn diese zum Zeitpunkt des Aufrufs nicht leer sein sollte, wird der vorhandene Inhalt nicht überschrieben. Stattdessen wird der ausführende Thread so schlafen gelegt, bis ein anderer Thread die Variable mit takeMvar leert. (Es gibt auch tryPutMvar :: Mvar -> a -> 10 Bool , das den Thread nicht schlafen geht und den Erfolg durch den Rückgabewert anzeigt.)

Aufgabe 13. Lesen aus einer dauerhaft leeren Variable

Was macht folgender Code? Wie reagiert das Laufzeitsystem von GHC?

```
import Control.Concurrent
main = do
    ref <- newEmptyMVar
    takeMVar ref</pre>
```

Aufgabe 14. Ein einfaches Beispiel zu Variablen

Schreibe ein Programm, das zwei leichtgewichtigen Threads erzeugt, die je eine große Fibonacci-Zahl berechnen und das Ergebnis in je einer Variable speichern. Der Hauptthread soll dann die beiden Ergebnisse ausgeben.

Aufgabe 15. Vorsicht vor Deadlocks

import Control.Concurrent

Was macht folgender Code? Wie reagiert das Laufzeitsystem von GHC?

```
main = do
    ref1 <- newEmptyMVar
    ref2 <- newEmptyMVar
    forkIO $ takeMVar ref2 >> putMVar ref1 "Hallo Welt"
    putStrLn =<< takeMVar ref1</pre>
```

Aufgabe 16. Warten auf Kinder

Oft möchte man im Hauptthread die Beendigung gestarteter Threads abwarten. Das ist zum Beispiel mit folgendem Code möglich (den es natürlich auch schon in verpackter Form im Modul Control.Concurrent.Async gibt). Vollziehe ihn nach!

```
import Control.Monad
import Control.Concurrent

forkThread :: IO () -> IO (MVar ())
forkThread proc = do
    ref <- newEmptyMVar
    forkFinally proc $ \_ -> putMVar ref ()
    return ref

main = do
    jobs <- mapM forkThread [...]
    mapM_ takeMVar jobs</pre>
```

Neben veränderlichen Variablen gibt es noch *Kanäle* zur Kommunikation zwischen Threads. Kanäle können anders als Variablen mehr als einen Wert zwischenspeichern. Man erzeugt einen Kanal mit newChan :: IO (Chan a) , pusht einen Wert durch writeChan :: Chan a -> a -> IO () und poppt den vordersten Wert mit readChan :: Chan a -> IO a . Der Aufruf von readChan blockiert, falls der Kanal leer ist.

Aufgabe 17. Sleep Sort kanalbasiert

Modifiziere deinen Sleep-Sort-Algorithmus derart, dass die sortierten Werte nicht auf dem Terminal ausgegeben, sondern in einen Kanal geschrieben werden. Dieser soll dann in einem Rutsch ausgegeben werden.

Zum Ende dieses Abschnitts sei bemerkt, dass man selten auf der Ebene dieser Aufgaben programmieren muss. Für viele Einsatzgebiete gibt es schon fertige Kombinatoren-Bibliotheken zum nebenläufigen Programmieren.

Aufgabe 18. Projekt: Ein einfacher Chat-Server

Vervollständige folgende Vorlage für einen einfachen Chat-Server. Clients sollen sich mit ihm auf TCP-Port 4242 verbinden können. Eingehende Nachrichten sollen an alle verbundenen Clients weitergeleitet werden.

Diese Vorlage ist auf einem niedrigen Level, mit expliziten Socket-Operationen, geschrieben. Normalerweise würde man eine High-Level-Streaming-Bibliothek wie Conduits oder Pipes verwenden. Diese kümmern sich auch automatisch um ordnungsgemäßes Abmelden von Clients

Tipp. Verwende die Funktion dupChan :: Chan a \rightarrow IO (Chan a) . Bonusaufgabe. Identifiziere das Speicherleckproblem und löse es.

```
-- Setze einen Kanal auf. Was in diesen Kanal geschrieben wird,
-- soll an alle verbundenen Clients weitergeleitet werden.
-- Die Hauptschleife: Akzeptiere laufend neue Verbindungen und
-- bearbeite sie.
forever $ do
    (conn,_) <- accept sock
           <- socketToHandle conn ReadWriteMode</pre>
    hSetBuffering hdl NoBuffering
    -- 'hdl' ist nun ein gewöhnlicher Handle, mit dem 'hGetLine'
    -- und 'hPutStrLn' verwendet werden können.
    -- Dupliziere den Kanal, um mehrere Zuhörer zu unterstützen.
    -- Schreibe gelesene Nachrichten in den Kanal.
    forkIO $ forever $ do
        msg <- hGetLine hdl
    -- Leite Nachrichten der anderen Verbindungen weiter.
    forkIO $ forever $ do
       hPutStrLn hdl msg
```

3.3 Shared Transactional Memory (STM)

In dem folgenden Programm kommt es zu einer Race Condition. Wiederholte Aufrufe des Programms werden verschiedene Ergebnisse liefern.

```
import Control.Concurrent
import Control.Monad
forkThread :: IO () -> IO (MVar ())
forkThread = {- siehe oben -}
go :: IORef Integer -> IORef Integer -> IO ()
go xRef yRef = do
    x <- readIORef xRef
   y <- readIORef yRef
    let x' = y + 1
        y' = x' + 1
    writeIORef xRef x'
    writeIORef yRef y'
main = do
    xRef <- newIORef 1
    yRef <- newIORef 2
    jobs <- replicateM 40000 $ forkThread $ go xRef yRef
   mapM_ takeMVar jobs
```

```
x <- readIORef xRef
y <- readIORef yRef
print (x, y)</pre>
```

Mit STM passiert das nicht. Statt IORef 's verwendet man dann TVar 's. Die Operationen übertragen sich wörtlich, spielen sich dann aber in der STM - statt der IO-Monade ab: newTVar :: a -> STM (TVar a) und so weiter. Man führt STM-Aktionen mit atomically :: STM a -> IO a aus.

Der angepasste Code sieht so aus:

```
import Control.Concurrent
import Control.Concurrent.STM
import Control.Monad
forkThread :: IO () -> IO (MVar ())
forkThread = {- siehe oben -}
go :: TVar Integer -> TVar Integer -> STM ()
go xRef yRef = do
    x <- readTVar xRef
    y <- readTVar yRef
    let x' = y + 2
        y' = x' + 2
    writeTVar xRef x'
    writeTVar yRef y'
main = do
    xRef <- newTVarIO 1
    yRef <- newTVarIO 2</pre>
    jobs <- replicateM 40000 $ forkThread $ atomically $ go xRef yRef
    mapM_ takeMVar jobs
    x <- readTVarIO xRef
    y <- readTVarIO yRef
   print (x, y)
```

Wie funktioniert STM? In erster Näherung so: Wird ein atomically -Block ausgeführt, so werden Änderungen an Tvar 's nicht sofort geschrieben. Stattdessen werden sie in einem Log notiert. Am Ende des Blocks prüft das Laufzeitsystem in einer atomaren Operation, ob sich seit Ablauf des Blocks seine Abhängigkeiten (zum Beispiel veränderliche Variablen, auf die lesend zugegriffen wurde) geändert haben. Wenn nein, macht es die Änderungen an den Tvar 's wirksam. Wenn ja, wird der Block einfach erneut ausgeführt. Da in der STM-Monade nicht beliebige Nebenwirkungen wie fireMissiles möglich sind, ist das ein fundiertes Vorgehen.