# EFFECTS FOR LESS

Alexis King ZuriHac 2020

- FEB 2016 Alexis starts writing Haskell.

  → Mostly a CRUD app.
  - → Performance not that important.

DEC 2017 Effect management is getting complicated.

- → Start exploring effect systems.
- → Clean up freer, release as freer-simple.

FEB 2018 O Break from writing Haskell to write Racket.

Jul 2019 O Back to writing Haskell.



"Realtime GraphQL on PostgreSQL"

Secretly: a GraphQL to SQL JIT compiler



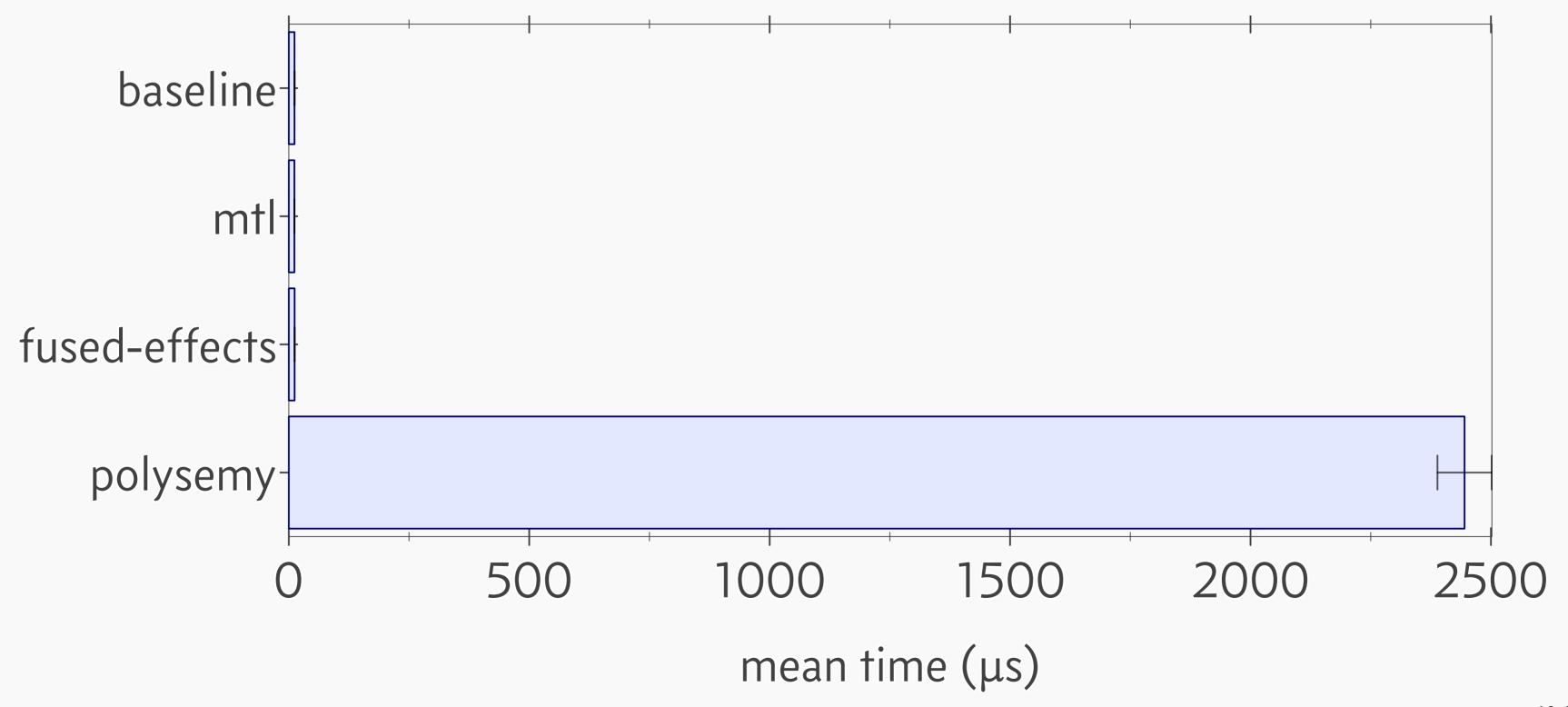
Performance is really important!

## Can we afford to use an effect system?

## BENCHMARKS

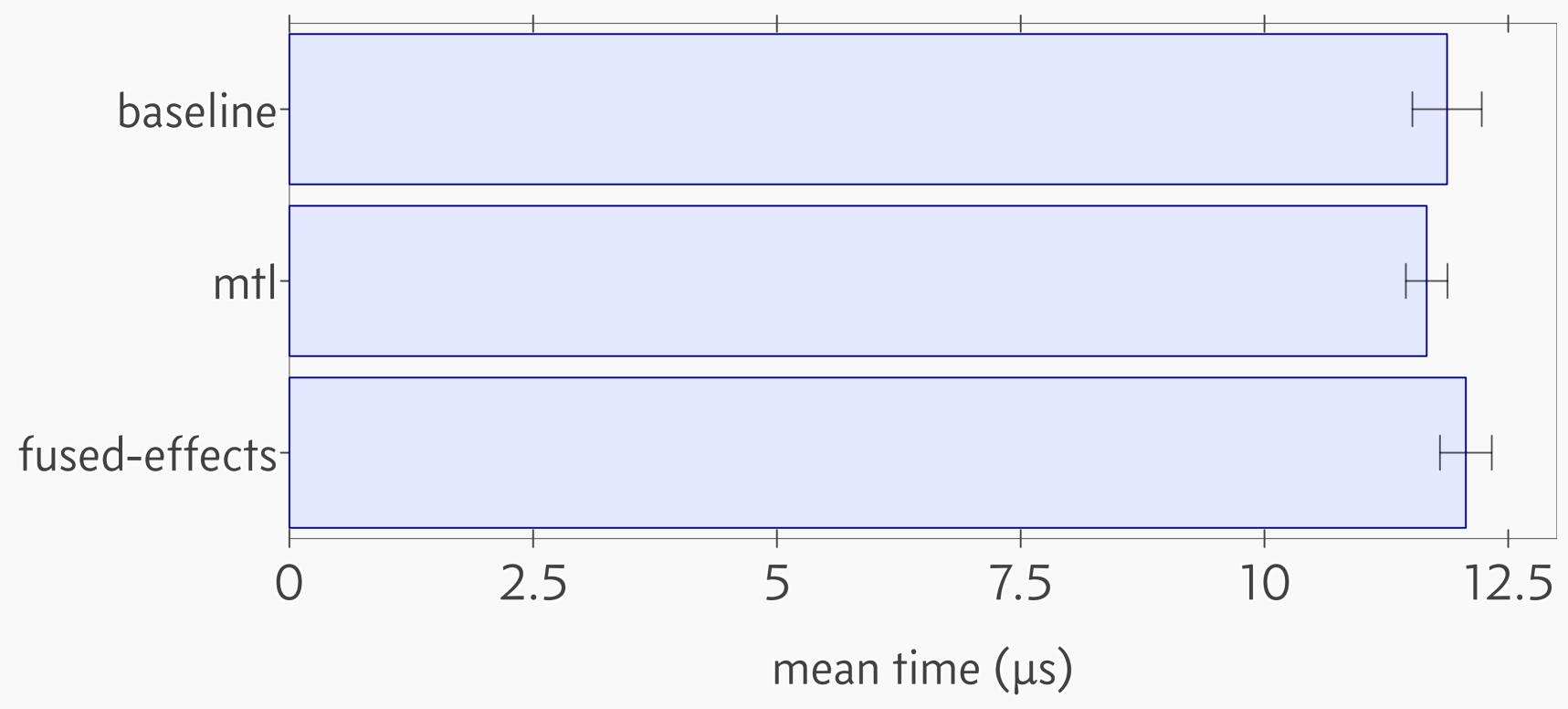
#### BENCHMARK: COUNTDOWN

### (lower is better)



#### BENCHMARK: COUNTDOWN

## (lower is better)



Can we afford to use an effect system?

Answer: yes.

Just don't pick polysemy.

# It's never that simple!

## QUESTIONS

- 1. What is the countdown benchmark?
- 2. What is the "baseline" implementation?
- 3. Are these differences even meaningful?
- 4. Why does polysemy do that much worse?

#### THE COUNTDOWN MICROBENCHMARK

```
countDown initial = runState program initial
program :: MonadState Int m ⇒ m Int
program = do n ← get
             if n \leq 0
               then return n
               else put (n - 1) >> program
```

countDown :: Int  $\rightarrow$  (Int, Int)

#### COUNTDOWN MICROBENCHMARK: BASELINE

```
countDown :: Int → (Int, Int)
countDown = program
```

# This is incredibly synthetic!

Is countdown a bad benchmark?

#### IN DEFENSE OF MICROBENCHMARKS

#### REAL-WORLD BENCHMARKS

#### MICROBENCHMARKS



- → Probably representative of something.
- → Unlikely to be a fluke/flaw of the benchmarking process.

- → Easy to isolate costs!
- → Small enough to completely understand.
- → If thoroughly understood, can provide a useful cost model.



- → Really big!
- → Difficult to isolate costs.
- → May be challenging to extrapolate results.
- → Easy to measure the wrong thing!
- → Crucial to understand the scope of results.
- → Costs are not considered in context.

Effects are particularly hard to measure with real-world benchmarks!



### Effects are particularly hard to measure with real-world benchmarks!



## What does countdown measure?

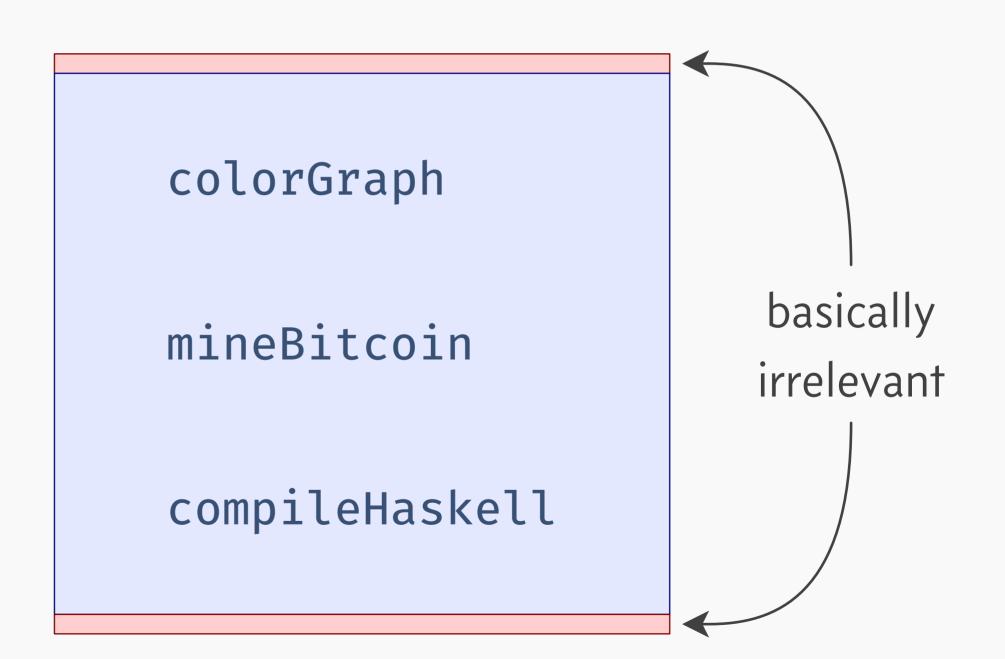
```
countDown :: Int → (Int, Int)
countDown initial = runState program initial
program :: MonadState Int m ⇒ m Int
program = do n ← get
             if n \leq 0
               then return n
               else put (n - 1) >> program
```

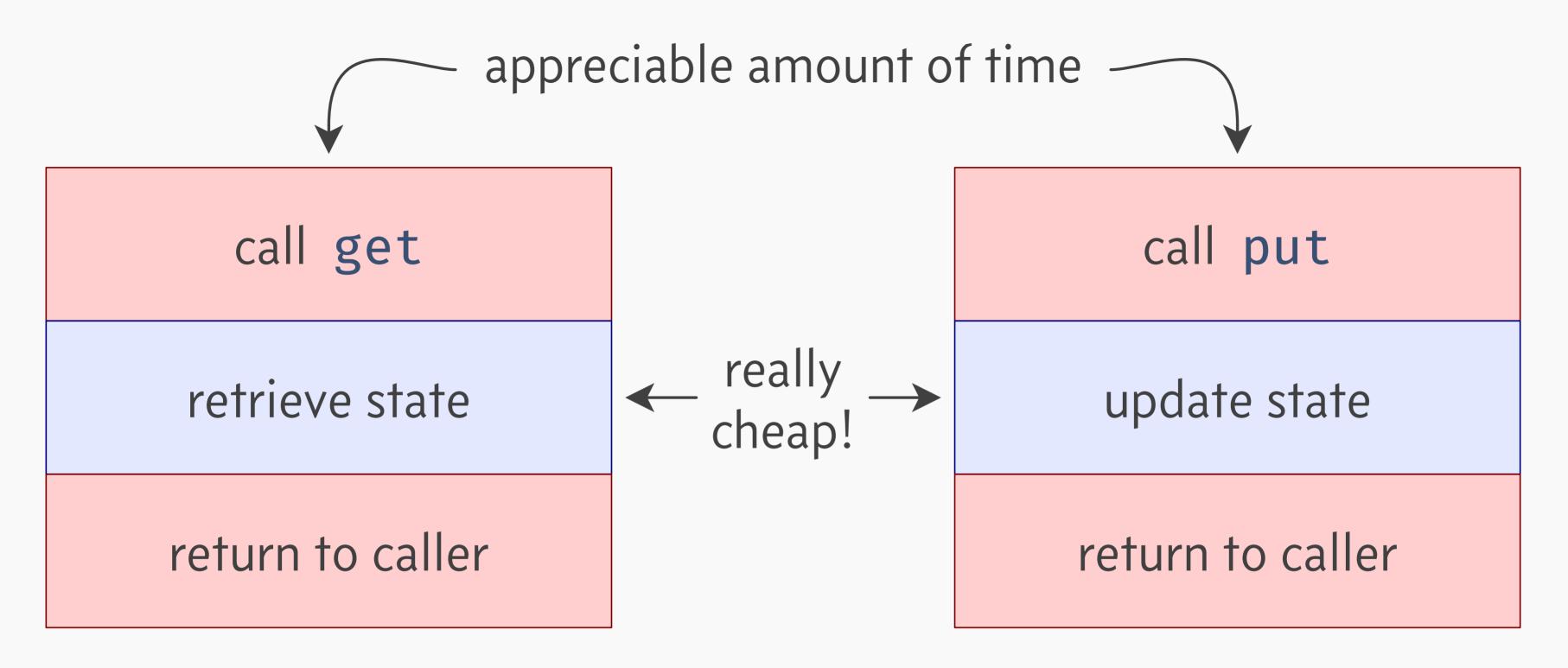
reallyExpensive :: MonadExpensive m ⇒ m Blah reallyExpensive = colorGraph >> mineBitcoin >> compileHaskell

call reallyExpensive

colorGraph
mineBitcoin
compileHaskell

return to caller





Countdown benchmarks effect dispatch.

```
countDown :: Int \rightarrow (Int, Int)
countDown initial = runState program initial
program :: MonadState Int m ⇒ m Int
program = get \gg \setminusn \rightarrow
               if n \leq 0
                 then return n
                 else put (n - 1) >> program
```

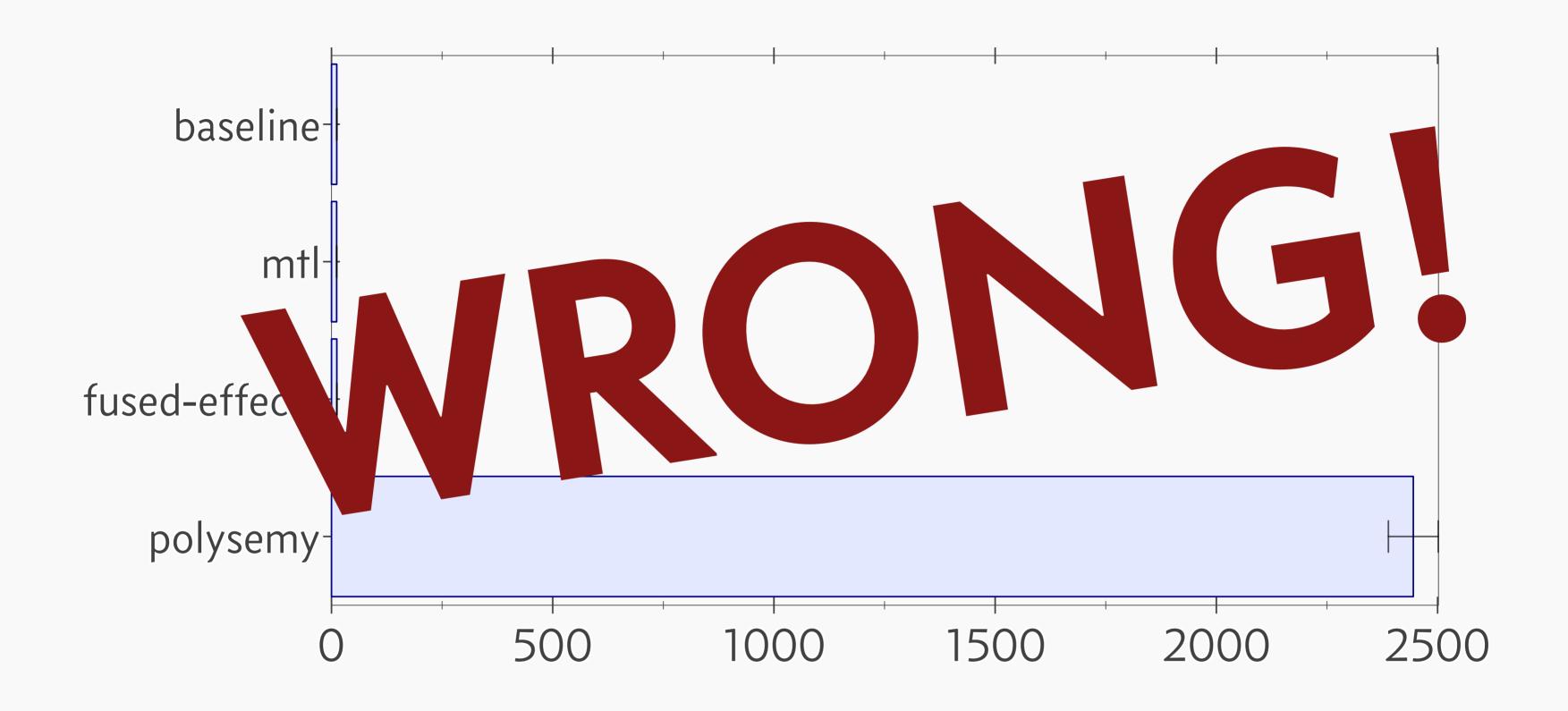
## is an exceptionally common operation.

mapM traverse sequence when unless replicateM

## RECAP

- 1. Countdown is a microbenchmark.
- 2. Theoretically, it's a valuable benchmark.
- 3. It measures two things:
  - → ...the cost of effect dispatch.
  - $\rightarrow$  ...the cost of  $\gg$  .

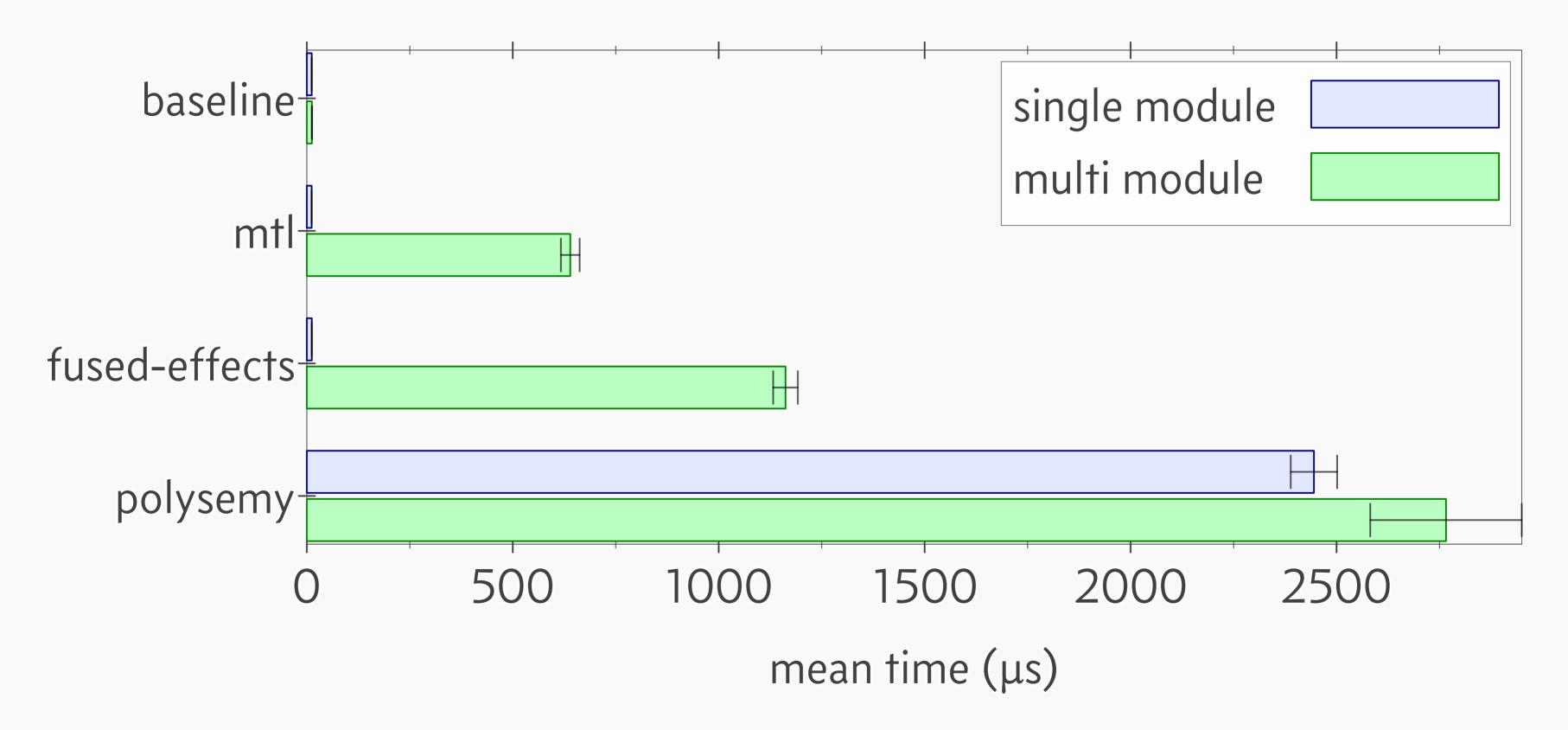
## Why am I belaboring this point?

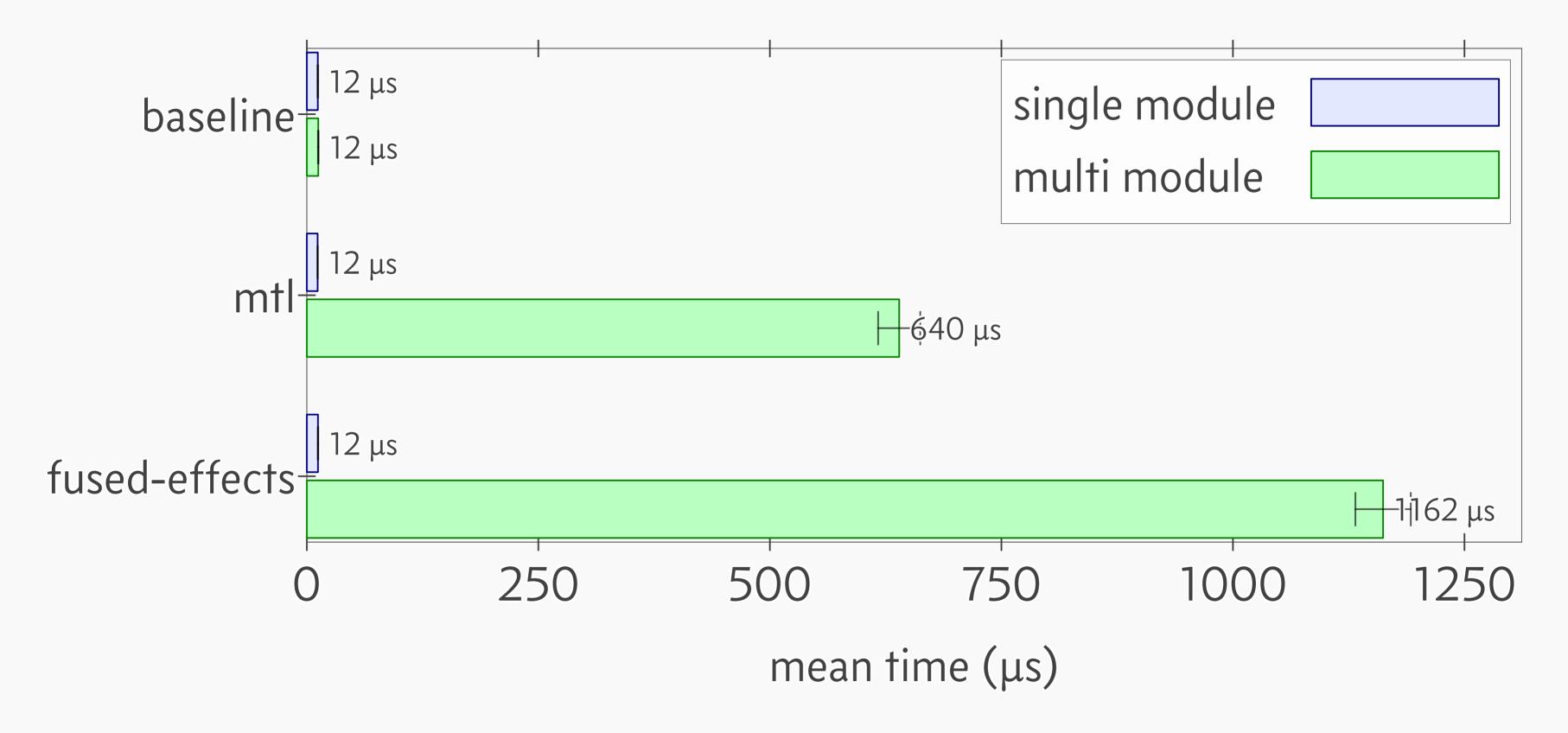


## ...or at least highly misleading.

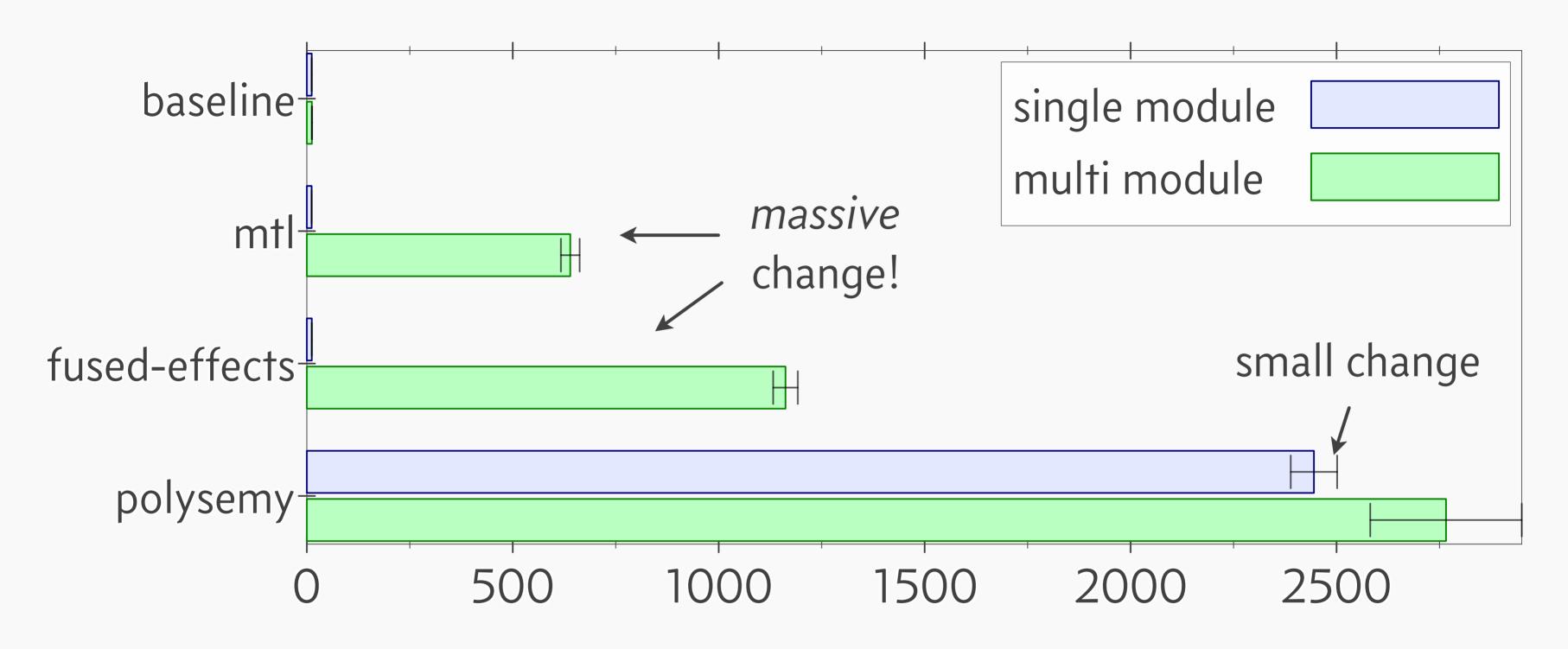
```
module CountDown where
 import Program
 countDown :: Int → (Int, Int)
 countDown initial = runState program initial
module Program where
 program :: MonadState Int m ⇒ m Int
 program = do n \leftarrow get
               if n \leq 0
                 then return n
                 else put (n - 1) >> program
```

Surely this shouldn't change anything?





# What happened?



mtl and fused-effects are victims of the optimizer.

### KEY TAKEAWAYS

- 1. mtl and fused-effects are faster than polysemy...
- 2. ...but are more reliant on compiler optimizations for best-case perf.
- 3. Tiny program changes can have huge perf diffs!

- ① Monad transformers + typeclasses.
  - → Tried and true; believed to be performant.
  - → Can require complex boilerplate.
  - Examples: mtl, fused-effects
- (2) Free-like monads.
  - → Highly flexible, can be simpler to use.
  - → Performance is a known limitation.
  - Examples: freer-simple, polysemy

```
if n \leq 0
             then return n
             else put (n - 1) >> program
program :: Eff (State Int) Int
program = Get Then \n →
           if n \leq 0
            then Return n
            else Put (n - 1) Then \setminus program
```

```
data Eff f a where
  Return :: a \rightarrow Eff a
  Then :: f a \rightarrow (a \rightarrow Eff b) \rightarrow Eff b
data State s a where
  Get :: State s s
  Put :: s \rightarrow State ()
runState :: s \rightarrow Eff (State s) a \rightarrow (s, a)
runState s (Return x) = (s, x)
runState s (Get `Then` k) = runState s (k s)
runState _ (Put s `Then` k) = runState s (k ())
```

### PROS

- → Beautifully simple.
- → Extremely flexible.

### CONS

- → Reifies the entire program as a tree.
- → Obscures structure to the optimizer.

```
newtype State s a = State { runState :: s \rightarrow (s, a) }
instance Monad (State s) where
  return x = State \$ \s \rightarrow (s, x)
  m \gg f = State $ \s \rightarrow case runState m s of
     (s', a) \rightarrow runState (f a) s'
get :: State s s
get = State \$ \s \rightarrow (s, s)
put :: s \rightarrow State s ()
put s = State \$ \setminus \rightarrow (s, ())
```

```
program :: State Int Int

program = State $ \s1 \rightarrow runState  get s1 of

(s2, n) \rightarrow if n \leqslant 0

then runState (return n) s2

else case runState (put (n - 1)) s2 of

(s3, _) \rightarrow runState program s3
```

```
program :: Int \rightarrow (Int, Int)

program s1 = case get s1 of

(s2, n) \rightarrow if n \leqslant 0

then (s2, n)

else case (n - 1, ()) of

(s3, _) \rightarrow program s3
```

```
program :: Int \rightarrow (Int, Int)

program s1 = case get s1 of

(s2, n) \rightarrow if n \leq 0

then (s2, n)

else program (n - 1)
```

```
program :: Int \rightarrow (Int, Int)
program s1 = case (s1, s1) of
(s2, n) \rightarrow if n \leq 0
then (s2, n)
else program (n - 1)
```

```
program :: Int \rightarrow (Int, Int)
program s1 = if s1 \leq 0
then (s1, s1)
else program (s1 - 1)
```

```
program :: Int \rightarrow (Int, Int)
program s1 = if s1 < 0
                 then (s1, s1)
                 else program (s1 - 1)
program :: Int \rightarrow (Int, Int)
program n = if n < 0
                 then (n, n)
                 else program (n - 1)
```

# This is great! But it's not an effect system.

```
program :: State Int Int \leftarrow coupled to implementation program = get \gg \setminus n \rightarrow if n \leqslant 0 then return n else put (n - 1) >> program
```

Can we get the flexibility of free monads with the performance of monad transformers?

$$1 + 2 + 3 + 4 + 5 + 0$$

```
program :: Eff (State Int) Int
program = Get `Then` \n →
            if n \leq 0
              then Return n
              else Put (n - 1) Then \setminus program
       producer –
  runState :: s \rightarrow Eff (State s) a \rightarrow (s, a)
  runState s (Return x) = (s, x)
  runState s (Get `Then` k) = runState s (k s)
  runState (Put s `Then` k) = runState s (k ())
```

# Can we get GHC to do this? Might mtl do better?

```
program :: MonadState Int m \Rightarrow m Int program = get \gg \n \rightarrow if n \leqslant 0 then return n else put (n - 1) \gg program
```

How is this compiled?

# How are typeclasses compiled?

# Non-Solution 1: Type Dispatch

```
show x = case typeOf x of
Bool → show_Bool x
Char → show_Char x
String → show_String x
```

Immediate problem: requires whole-program compilation.

### Deeper problem: full type erasure.

Java: if (obj instanceof Foo) { ... }

obj  $\longrightarrow$  fields

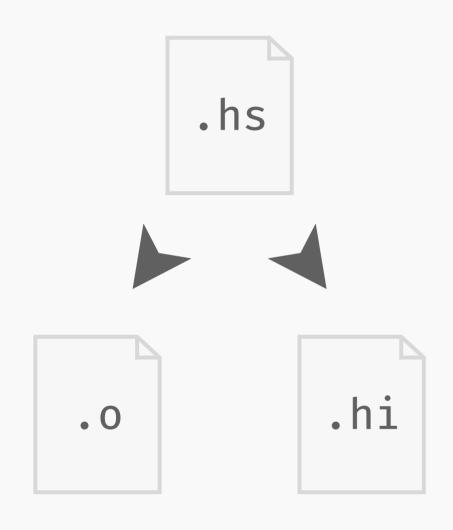
Foo  $\longrightarrow$  class methods



### Non-Solution 2: Monomorphization

```
exclaim :: Show a \Rightarrow a \rightarrow String exclaim x = Show x ++ "!"
```

- 1. Generate no code for exclaim.
- 2. Record exclaim's definition in the interface file.



- .hi "Haskell interface"
  - → type/class/instance declarations
  - → types of exported bindings
  - → source code of small bindings
  - → for monomorphization: source code of all overloaded bindings

### exclaim True

```
exclaim :: Show a \Rightarrow a \rightarrow String exclaim x = show x ++ "!"
```

```
exclaim_Bool :: Bool → String exclaim_Bool x = show_Bool x ++ "!"
```

exclaim True > exclaim\_Bool True

Overloading has no runtime cost!

### ...but it can create a lot of bloat.

```
exclaim True exclaim 42 exclaim "hello"
exclaim_Bool :: Bool → String
exclaim_Bool x = show_Bool x ++ "!"
exclaim_Int :: Int → String
exclaim Int x = show_Int x ++ "!"
```

```
exclaim_String :: String → String exclaim_String x = show_String x ++ "!"
```

# reallyBig :: (Foo a, Bar b, Baz c) ⇒ ... reallyBig = < really big RHS >



reallyBig	<b>∂Bool</b>	<b>∂Bool</b>	<b>@Bool</b>	reallyBig	<b>∂Bool</b>	<b>∂Bool</b>	@Int	reallyBig	<b>∂Bool</b>	<b>∂Bool</b>	<b>@String</b>
reallyBig	<b>∂Bool</b>	@Int	<b>@Bool</b>	reallyBig	<b>∂Bool</b>	@Int	@Int	reallyBig	@Bool	@Int	<b>@String</b>
reallyBig	<b>∂Bool</b>	@String	<b>@Bool</b>	reallyBig	<b>∂Bool</b>	<pre>aString</pre>	@Int	reallyBig	@Bool	<b>@String</b>	<b>@String</b>
reallyBig	@Int	<b>∂Bool</b>	<b>@Bool</b>	reallyBig	@Int	<b>∂Bool</b>	@Int	reallyBig	@Int	<b>∂Bool</b>	<b>@String</b>
reallyBig	@Int	@Int	<b>@Bool</b>	reallyBig	@Int	@Int	@Int	reallyBig	@Int	@Int	<b>@String</b>
reallyBig	<b>@Int</b>	<b>@String</b>	<b>@Bool</b>	reallyBig	@Int	<pre>aString</pre>	aInt	reallyBig	@Int	aString	aString
reallyBig	<b>@String</b>	<b>@Bool</b>	<b>∂Bool</b>	reallyBig	aString	<b>∂Bool</b>	aInt	reallyBig	<b>@String</b>	<b>∂Bool</b>	<b>@String</b>
reallyBig	<b>@String</b>	@Int	<b>@Bool</b>	reallyBig	<b>@String</b>	@Int	@Int	reallyBig	<pre>aString</pre>	@Int	<b>@String</b>
reallyBig	aString	@String	<b>∂Bool</b>	reallyBig	<b>@String</b>	<pre>aString</pre>	aInt	reallyBig	@String	<b>@String</b>	<b>@String</b>

# Monomorphization

- → Can be good for runtime performance.
- → Can be very bad for code size & compile times.
- → C++/Rust programmers have to worry about this!
- → Haskell programmers generally do not.

```
exclaim :: Show a \Rightarrow a \rightarrow String exclaim x = Show x ++ "!"
```



```
exclaim :: (a \rightarrow String) \rightarrow a \rightarrow String
exclaim show_a x = show_a x ++ "!"
```

exclaim True ➤ exclaim show\_Bool True exclaim 42 ➤ exclaim show\_Int 42

```
class Show a where
    show :: a \rightarrow String
    showsPrec :: Int \rightarrow a \rightarrow ShowS
    showList :: [a] → ShowS
data Show a = ShowDict
   \{ \text{ show } :: a \rightarrow \text{String} \}
   , showsPrec :: Int \rightarrow a \rightarrow ShowS
   , showList :: [a] → ShowS }
exclaim :: Show a \rightarrow a \rightarrow String
exclaim dict x = show dict x ++ "!"
```

# This is dictionary passing.

- → Elegantly simple.
- → Cheap to compile.
- → Does it have a runtime cost?

```
program :: MonadState Int m ⇒ m Int
    program = get \gg \backslash n \rightarrow
                  if n \leq 0
                     then return n
                     else put (n - 1) >> program
program :: MonadState Int m → m Int
program stateDict@(MonadStateDict monadDict ) =
  (>=) monadDict
         (get stateDict)
         (n \rightarrow if n \leq 0)
            then return monadDict n
            else (>>) monadDict
                       (put stateDict (n - 1))
                       (program stateDict)
```

# Is this performant?

Nope: we can't inline anything.

### Most calls are known calls.

```
fst ("hello", "world")

fst :: (a, b) \rightarrow a

fst (x, _) = x
```

## Most calls are known calls.

```
fst ("hello", "world")
(\x, _) \rightarrow x) ("hello", "world")
               "hello"
```

Higher-order functions make unknown calls.

```
foldr :: (a \rightarrow b \rightarrow b) \rightarrow b \rightarrow [a] \rightarrow b

foldr _ v [] = v

foldr f v (x:xs) = f x (foldr f v xs)

???
```

Unknown calls are hard stops for the optimizer.

#### KNOWN CALL

#### UNKNOWN CALL

- → Compiled to direct jump.
  - ect jump. → Compiled to indirect jump.
- → Exposes strictness info.

→ Assumed lazy in all args.

→ Args can be unboxed.

→ Args must have declared types.

→ Can have rewrite RULES.

→ Opaque to RULES.

→ Inlined if small.

→ Never inlined.

### This is the cost of AOT compilation!

# Typeclass overloading creates unknown calls. Overloading is **not** free!

#### Unknown calls are not a death sentence.

```
sumIndicies :: Eq a \Rightarrow a \rightarrow [a] \rightarrow [Int] sumIndicies v xs = zip [1..] xs & filter ((== v) . snd) & map fst & sum
```

List fusion can still happen!

Unknown calls to >= are a problem.

Unknown calls to >= are a problem.

$$\Rightarrow$$
 is glue.

Conclusion: not surprising at all that mtl has a cost!

But why is it sometimes fast?

## SPECIALIZATION

### GHC monomorphizes in limited circumstances.

- 1. GHC looks for calls to overloaded functions at known, concrete types.
- 2. The function must satisfy at least one of the following:
  - → It is defined in the current module.
  - → It was declared with an INLINABLE pragma.
  - → It is a class method and its unfolding (source code) is in the interface file.

### GHC monomorphizes in limited circumstances.

- 1. GHC looks for calls to overloaded functions at known, concrete types.
- 2. The function must satisfy at least one of the following:
  - → It is defined in the current module.
  - → It was declared with an INLINABLE pragma.
  - → It is a class method and its unfolding (source code) is in the interface file.

```
program :: MonadState Int m ⇒ m Int
program = do n \leftarrow get
            if n \leq 0
              then return n
              else put (n - 1) >> program
program :: State Int Int
if n \leq 0
              then return n
              else put (n - 1) >> program
```

```
program :: MonadState Int m ⇒ m Int
    program = do n ← get
                 if n \leq 0
                   then return n
                   else put (n - 1) >> program
program :: MonadState Int m → m Int
program stateDict@(MonadStateDict monadDict ) =
 (≫ ) monadDict
        (get stateDict)
        (n \rightarrow if n \leq 0)
           then return monadDict n
           else (>>) monadDict
                      (put stateDict (n - 1))
                      (program stateDict)
```

## Why bother explaining all of this?

## Because we're not done.

## Can we avoid the performance regression?

- → It is defined in the current module.
- → It was declared with an INLINABLE pragma.
- → It is a class method and its unfolding (source code) is in the interface file.

### This approach has many problems.

- 1. It would be annoying.
- 2. It requires whole-program specialization!
  - → Probably recompiles everything in Main.hs.
  - → Compilation may not terminate.
  - → This cost is incurred for each specialization.
  - → Can be defeated by existential quantification.

## Let's take a step back.

Why is specialization necessary here?

## Proposition 1: effect systems are fundamentally about dynamic dispatch.

```
\begin{array}{lll} \text{program} &=& \text{get} & & \longrightarrow & \text{interpretation not} \\ & \text{if } n & \leqslant & 0 & \text{yet decided!} \\ & \text{then return } n \\ & \text{else put } (n-1) & >> \text{program} \end{array}
```

Proposition 2: effect dispatch can be made perfectly affordable. The real problem is >= .

```
program = get \gg \n \rightarrow if n \leqslant 0 then return n else put (n - 1) >> program
```

## Passing >= via dictionary creates problems!

- 1.  $\gg$  gets called a *lot*.
- 2. >= is glue; it needs to be inlined to expose further optimizations.
- 3. Unknown calls to >= balloon closure allocation.

captured in closure used under a lambda –  $f \times y = foo \times \Rightarrow \langle z \rightarrow bar (+ y z)$ stored on stack  $f x \dot{y} = case foo x of$ Left  $e \rightarrow Left e$ Right  $z \rightarrow bar (+ y z)$ used in a case RHS —

If >= is an unknown call, the caller must allocate a closure for the continuation.

When >= is passed via dictionary, the program is reified as a tree of lambdas!

```
program :: Eff (State Int) Int program = Get Then \n \rightarrow if n \leq 0 then Return n else Put (n - 1) Then \_ \rightarrow program
```

Without specialization, these aren't that different!

## How do we escape?

#### ESCAPE PLAN

- 1) We need a monad with an inlinable >= .
- (2) >= itself must not allocate.
- 3 It must be able to handle all algebraic effects.
- 4 Effect dispatch can be dynamic (but must be fast).

This is not a small order.

Requirements (2) and (3) are especially hard.

## Monad transformers are out of the question. (Effect behavior lives in >= .)

Free-like monads are also disqualified. ( >= must allocate to construct the tree.)

What's left?

#### DELIMITED CONTINUATIONS

There is a deep, well-known connection between delimited continuations and algebraic effects.

(Well outside the scope of this talk!)

Big idea: equip a monad with a pair of super powerful control operators, prompt and control.

All effect handlers can be defined in terms of these operators.

>= does not need to know anything about effect behavior.

New problem: the only way to implement delimited continuations in Haskell is CPS.

In CPS, continuation is passed via closure:

$$m \gg f = \langle k \rightarrow m (\langle x \rightarrow f x k \rangle)$$

All calls to non-inlined monadic functions must allocate. (This isn't much better than free monads.)

Cost is okay for continuation-happy code, but effects that don't need them (very common!) still must pay for them.

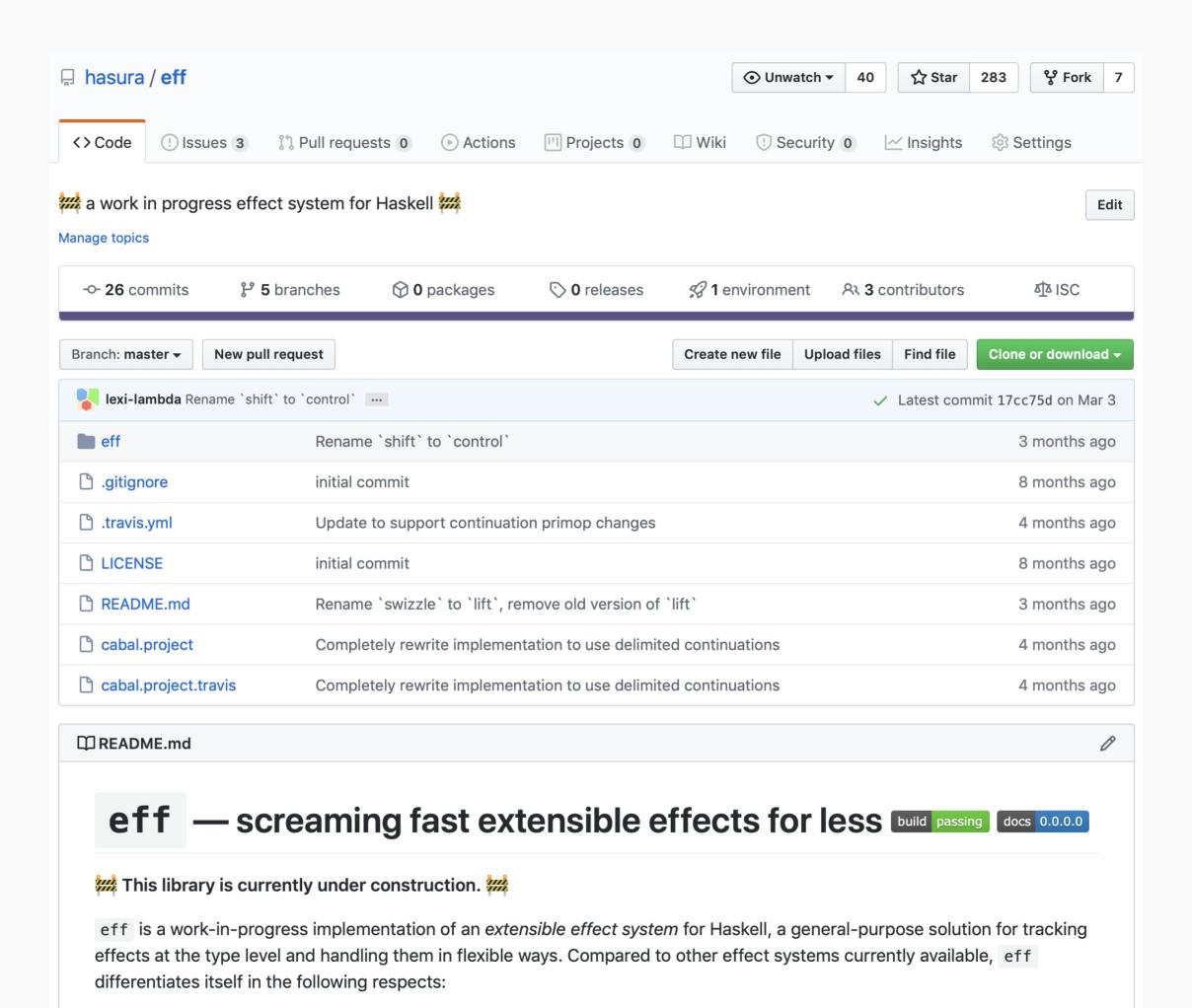
## But this is frustrating!

There are well-known, efficient implementation techniques for delimited continuations!

#### I give up. Let's just patch GHC.



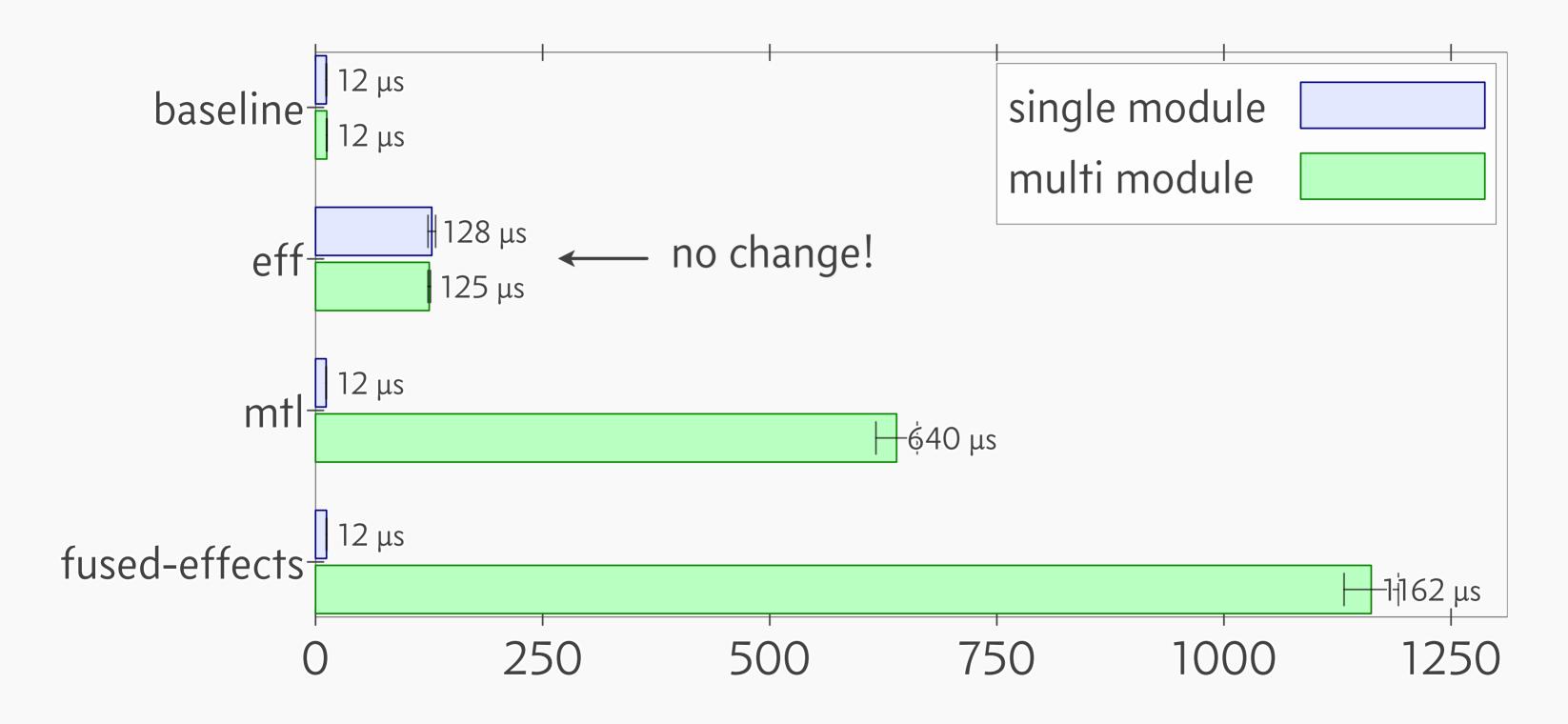
## EFF



#### EFF: KEY FEATURES

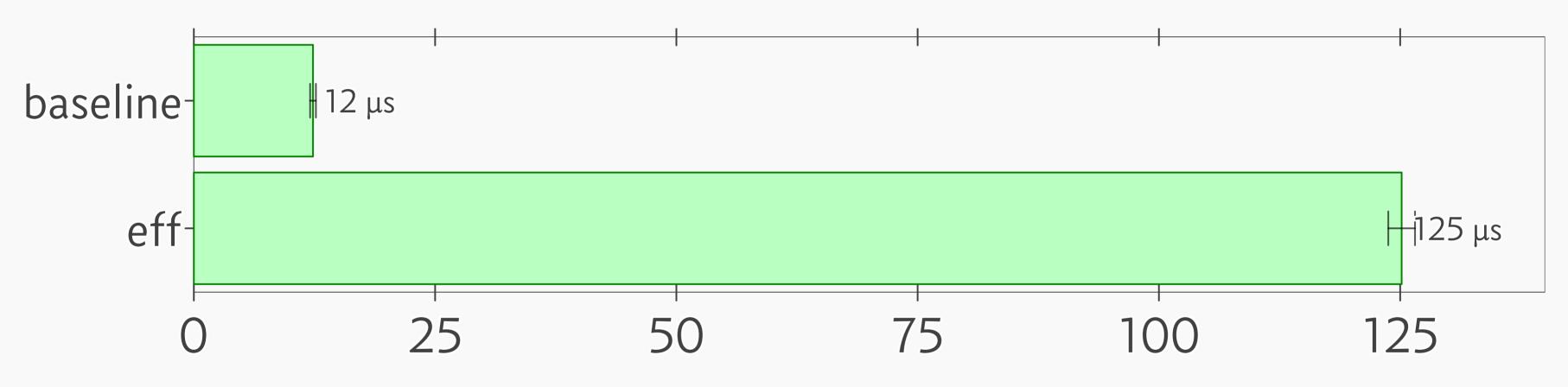
- → API is similar to free-like approaches (actually simpler!).
- → Provides a single monad, named Eff.
- → Basically ReaderT over ST under the hood.
  - → Reader context holds current handlers.
  - → Wraps the unsafe primops in a safe API.
- → Effect dispatch takes constant time (not amortized).
- → Faster for most use cases than unspecialized mtl.

#### EFF: THE NUMBERS



eff's performance does *not* depend on compiler optimizations!

#### EFF: THE NUMBERS



Should we be worried?

#### COUNTDOWN: THE REST OF THE STORY

Why so much faster?



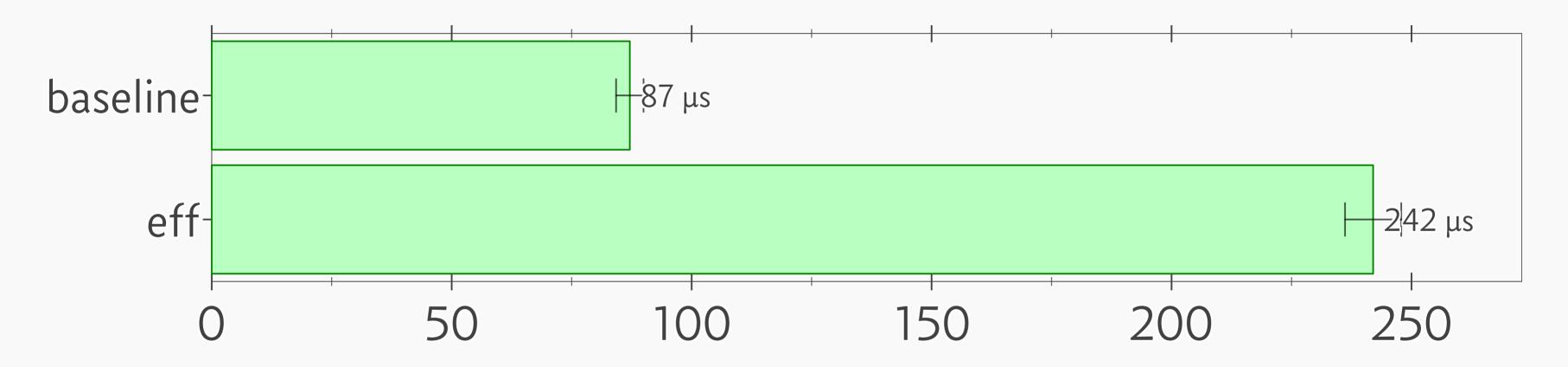
```
$wprogram :: Int# \rightarrow (# Int#, Int# #)
$wprogram n = if n \leq # 0# then (# n, n #)
else $wprogram (n -# 1#)
```

```
program :: Int# \rightarrow (# Int#, Int# #)
$wprogram n = if n \leq \# 0\# then (\# n, n \#)
                            else $wprogram (n -# 1#)
             _c4ap: addq $16,%r12
                     cmpq 856(%r13),%r12
                     ja c4aB
             _c4aA: testq %r14,%r14
                                          8 instructions!
                     jle _c4aw
             _c4av: addq $-16,%r12
                     decq %r14
                     jmp _c4ap
```



```
program :: Integer \rightarrow (Integer, Integer)
program n = if n \leq 0 then (n, n)
else program (n - 1)
```

#### COUNTDOWN: NO UNBOXING



Not quite so bad, after all!

# Phew.

#### RECAP

- → Countdown benchmarks > and effect dispatch.
- → mtl and fused-effects rely on compiler optimizations.
- → Those optimizations are very often not viable.
- → eff closes the gap by being inherently fast:
  - → It exposes more local transformations to the optimizer.
  - → Delimited continuation primops avoid CPSing.
  - → Result: eff handily wins the benchmark shootout.
- → Numbers are not enough. It is our responsibility to ask why.

#### CLOSING THOUGHTS

- → The eff story is not over.
  - → Need to plumb in support for IO exceptions.
  - → The GHC proposal needs to be accepted.
- → The design of eff itself could be its own talk!
- → mtl's performance is misunderstood and overhyped!
- → Effect systems are too foundational to ignore perf.
- → We really do need good real-world benchmarks.
- → Someday: effect specialization?

#### THE END

- → Countdown benchmarks > and effect dispatch.
- → mtl and fused-effects rely on compiler optimizations.
- → Those optimizations are very often not viable.
- → eff closes the gap by being inherently fast:
  - → It exposes more local transformations to the optimizer.
  - → Delimited continuation primops avoid CPSing.
  - → Result: eff handily wins the benchmark shootout.
- → Numbers are not enough. It is our responsibility to ask why.

- → The eff story is not over.
  - → Need to plumb in support for IO exceptions.
  - → The GHC proposal needs to be accepted.
- → The design of eff itself could be its own talk!
- → mtl's performance is misunderstood and overhyped!
- → Effect systems are too foundational to ignore perf.
- → We really do need good real-world benchmarks.
- → Someday: effect specialization?

eff: https://github.com/hasura/eff

benchmarks: https://github.com/ocharles/effect-zoo

proposal: https://github.com/ghc-proposals/ghc-proposals/pull/313

Hasura: https://hasura.io

me: https://lexi-lambda.github.io