

Contents

1	General Introduction	2
2	Short introduction to parallel Haskells	2
2.1	Multicore Haskell	2
2.2	ParMonad	2
2.3	Eden	2
2.4	HdpH	2
3	Arrows	2
4	Utility Functions	3
5	Parallel Arrows	4
5.1	The ArrowParallel interface	4
5.2	Multicore Haskell	5
5.3	ParMonad	5
5.4	Eden	6
5.5	HdpH	6
5.6	Benefits of parallel Arrows	6
6	Skeletons	7
6.1	parEvalNLazy	7
6.2	parEval2	7
6.3	parMap	8
6.4	parMapStream	8
6.5	farm	9
6.6	farmChunk	9
7	Syntactic Sugar	10
8	Benchmarks	10
9	Conclusion	12

1 General Introduction

Arrows were introduced in John Hughes paper as an alternative to Monads for API design [1]. In the paper Hughes describes that Arrows have one powerful additional property when compared to Monads for API design: Extensibility. In this paper we will show how this property can be used to add parallelism capabilities to different arrows.

Groß-/Kleinschreibung von Arrows, Monads, etc.

2 Short introduction to parallel Haskells

2.1 Multicore Haskell

2.2 ParMonad

2.3 Eden

2.4 HdpH

3 Arrows

John Hughes defined the Arrow typeclass as follows[1]:

```
1 class Arrow arr where
2   arr :: (a -> b) -> arr a b
3   (>>>) :: arr a b -> arr b c -> arr a c
4   first :: arr a b -> arr (a,c) (b,c)
```

`arr` is used to lift an ordinary function to an Arrow type. This can be thought of as analogous to the monadic `return`. The `>>>` operator, in a similar way, is analogous to the monadic composition operator `>>=`. And lastly, the `first` operator, which takes the input arrow from `b` to `c` and converts it into an arrow on pairs with the second argument untouched, is also needed for actual useful code as without it, we wouldn't have a way to save input across arrows.

The most prominent instances of this interface are regular functions (`->`),

```
1 instance Arrow (->) where
2   arr f = f
3   f >>> g = g . f
4   first f = \ (a, c) -> (f a, c)
```

and the Kleisli type.

```
1 instance Arrow (Kleisli m) where
2   arr f = Kleisli $ return . f
3   f >>> g = Kleisli $ \a -> f a >>= g
4   first f = Kleisli $ \ (a,c) -> f a >>= \b -> return (b,c)
```

With this typeclass in place, Hughes also defined some syntactic sugar: The mirrored version of `first`, called `second`,

```
1 second :: Arrow arr => arr a b -> arr (c, a) (c, b)
2 second f = arr swap >>> first f >>> arr swap
3 where swap (x, y) = (y, x)
```

the `***` combinator which combines `first` and `second` to handle two inputs in one arrow,

```
1 (***) :: Arrow arr => arr a b -> arr c d -> arr (a, c) (b, d)
2 f *** g = first f >>> second g
```

and the `&&&` combinator that constructs an arrow which outputs 2 different values like `***`, but takes only one input.

```
1 (&&&) :: Arrow arr => arr a b -> arr a c -> a a (b, c)
2 f &&& g = arr (\a -> (a, a)) >>> (f *** g)
```

A short example given by Hughes on how to use this is `add` over arrows:

```
1 add :: Arrow arr => arr a Int -> arr a Int -> arr a Int
2 add f g = (f &&& g) >>> arr (\(u, v) -> u + v)
```

The benefit of using the `Arrow` typeclass is that any type which is shown to be an `Arrow` can be used in conjunction with this newly created `add` combinator. Even though this example is quite simple, the power of the `Arrow` interface immediately is clear: If a type is an `Arrow`, it can immediately be used together with every library that works on `Arrows`. Compared to simple `Monads`, this enables us to write code that is more extensible, without touching the internals of the specific `Arrows`.

Note: In the definitions Hughes gave in his paper, the notation `a b c` for an arrow from `b` to `c` is used. We use the equivalent definition `arr a b` for an arrow from `a` to `b` instead, to make it easier to find the arrow type in type signatures. We kept the original notation `a b c` for this section to not change too much from Hughes' original definitions.

4 Utility Functions

Before we go into detail on parallel arrows, we introduce some utility combinators first, that will help us later: `map` and `zipWith` on arrows.

The `map` combinator lifts any arrow `arr a b` to an arrow `arr [a] [b]`,

```
1 -- from http://www.cse.chalmers.se/~rjmh/afp-arrows.pdf
2 mapArr :: ArrowChoice arr => arr a b -> arr [a] [b]
3 mapArr f =
4   arr listcase >>>
5   arr (const []) ||| (f *** mapArr f >>> arr (uncurry ()))
6   where
```

```

7 | listcase [] = Left ()
8 | listcase (x:xs) = Right (x,xs)

```

and zipWith lift any arrow `arr (a, b) c` to an arrow `arr ([a], [b]) [c]`.

```

1 | zipWithArr :: ArrowChoice arr => arr (a, b) c -> arr ([a], [b]) [c]
2 | zipWithArr f = (arr $ \ (as, bs) -> zipWith (,) as bs) >>> mapArr f

```

These two combinators make use of the `ArrowChoice` typeclass, which allows us to use the `|||` combinator. It takes two arrows `arr a c` and `arr b c` and combines them into a new arrow `arr (Either a b) c` which pipes all **Left** a's to the first arrow and all **Right** b's to the second arrow.

```

1 | (|||) :: ArrowChoice arr a c -> arr b c -> arr (Either a b) c

```

With the `zipWithArr` combinator we can also write a combinator `listApp`, which lifts a list of arrows `[arr a b]` to an arrow `arr [a] [b]`.

```

1 | listApp :: (ArrowChoice arr, ArrowApply arr) => [arr a b] -> arr [a] [b]
2 | listApp fs = (arr $ \as -> (fs, as)) >>> zipWithArr app

```

This combinator also makes use of the `ArrowApply` typeclass which allows us to evaluate arrows with `app :: arr (arr a b, a) c`.

5 Parallel Arrows

We have seen what Arrows are and how they can be seen as a general interface to computation. In the following section we will discuss how Arrows can also be seen as a general interface not only to computation, but to **parallel computation** as well. We start by introducing the interface and explaining the reasonings behind it. Then, we discuss some implementations using existing parallel Haskell. Finally, we explain why using Arrows for expressing parallelism is beneficial.

5.1 The ArrowParallel interface

In its purest form, parallel computation (on functions) can be looked at as the execution of some functions `a -> b` in parallel:

```

1 | parEvalN :: [a -> b] -> [a] -> [b]

```

Translating this into arrow terms gives us a new operator `parEvalN` that lifts a list of arrows `[arr a b]` to a parallel arrow `arr [a] [b]` (This combinator is similar to our utility function `listApp`, but does parallel instead of serial evaluation).

```

1 | parEvalN :: (Arrow arr) => [arr a b] -> arr [a] [b]

```

With this definition of `parEvalN`, parallel execution can be looked at as yet another arrow combinator. But as the implementation may differ depending on

the actual type of the arrow `arr` and we want this to be an interface for different backends, we have to introduce the new typeclass `ArrowParallel` to host this combinator.

```
1 class Arrow arr => ArrowParallel arr a b where
2   parEvalN :: [arr a b] -> arr [a] [b]
```

Sometimes parallel Haskells require additional configuration parameters for information about the execution environment. This is why we also introduce an additional `conf` parameter to the function. We also do not want `conf` to be of a fixed type, as the configuration parameters can differ for different instances of `ArrowParallel`, so we add it to the type signature of the typeclass as well.

```
1 class Arrow arr => ArrowParallel arr a b conf where
2   parEvalN :: conf -> [arr a b] -> arr [a] [b]
```

We don't require the `conf` parameter in every implementation. If it is not needed, we usually want to allow the `conf` type parameter to be of any type and don't even evaluate it by blanking it in the type signature of the implemented `parEvalN` as we will see in the implementation sections.

5.2 Multicore Haskell

The Multicore Haskell implementation of this class is straightforward using `listApp` from 4 combined with the `using` operator from Multicore Haskell.

```
1 instance (NFData b, ArrowApply arr, ArrowChoice arr) =>
2   ArrowParallel arr a b conf where
3   parEvalN _ fs = listApp fs >>> arr (flip using $ parList rdeepseq)
```

We hardcode the `parList rdeepseq` strategy here as in this context it is the only one making sense as we usually want the output list to be fully evaluated to its normal form.

5.3 ParMonad

The `ParMonad` implementation makes use of Haskell's laziness and `ParMonad`'s `spawnP :: a -> Par (IVar a)` function which forks away the computation of a value and returns an `IVar` containing the result in the `Par` monad.

We therefore apply each function to its corresponding input value with `app` and then fork the computation away with `arr spawnP` inside a `zipWithArr` call. This yields a list `[Par (IVar b)]`, which we then convert into `Par [IVar b]` with `arr sequenceA`. In order to wait for the computation to finish, we map over the `IVars` inside the `ParMonad` with `arr (>=> mapM get)`. The result of this operation is a `Par [b]` from which we can finally remove the monad again by running `arr runPar` to get our output of `[b]`.

```

1 instance (NFData b, ArrowApply arr, ArrowChoice arr) =>
2   ArrowParallel arr a b conf where
3     parEvalN _ fs =
4       (arr $ \as -> (fs, as)) >>>
5       zipWithArr (app >>> arr spawnP) >>>
6       arr sequenceA >>>
7       arr (>>= mapM get) >>>
8       arr runPar

```

5.4 Eden

For the Multicore and ParMonad implementation we could use general implementations that just required the ArrowApply and ArrowChoice interface. With Eden this is not the case as we can only spawn a list of functions and we cannot extract simple functions out of arrows. While this could be "fixed" by introducing a typeclass like

```

1 class (Arrow arr) => ArrowUnwrap arr where
2   arr a b -> (a -> b)

```

we don't do this in this paper, as this seems too hacky. For now, we just implement ArrowParallel for normal functions

```

1 instance (Trans a, Trans b) => ArrowParallel (->) a b conf where
2   parEvalN _ fs as = spawnF fs as

```

and the Kleisli type.

```

1 instance (Monad m, Trans a, Trans b, Trans (m b)) =>
2   ArrowParallel (Kleisli m) a b conf where
3   parEvalN conf fs =
4     (arr $ parEvalN conf (map (\(Kleisli f) -> f) fs)) >>>
5     (Kleisli $ sequence)

```

5.5 HdpH

5.6 Benefits of parallel Arrows

We have seen that we can wrap parallel Haskell's inside of the ArrowParallel interface, but why do we abstract parallelism this way and what does this approach do better than the other parallel Haskell's?

- **Arrow API benefits:** With the ArrowParallel typeclass we do not lose any benefits of using arrows as parEvalN is just yet another arrow combinator. The resulting arrow can be used in the same way its serial version can be used. This is a big advantage of this approach, especially compared to the monad solutions as all arrow combinators will continue to work instead of requiring special ones for each parallel monad.

- **Abstraction:** With the `ArrowParallel` typeclass, we abstracted all parallel implementation logic away from the business logic. This gives us the beautiful situation of being able to write our code against the interface the typeclass gives us without being bound to any parallel Haskell. So as an example, during development, we can run the code on the simple Multicore version and afterwards deploy it on a cluster using an Eden built, by just swapping out the actual `ArrowParallel` instance.

6 Skeletons

With the `ArrowParallel` typeclass in place and implemented, we can now implement some basic parallel skeletons.

6.1 `parEvalNLazy`

`parEvalN` is 100% strict, which means that it fully evaluates all passed arrows. Sometimes this might not be feasible, as it will not work on infinite lists of functions like e.g. `map (arr . (+)) [1..]` or just because we need the arrows evaluated in chunks. `parEvalNLazy` fixes this. It works by first chunking the input from `[a]` to `[[a]]` with the given `ChunkSize` in `(arr $ chunksOf chunkSize)`. These chunks are then fed into a list `[arr [a] [b]]` of parallel arrows created by feeding chunks of the passed `ChunkSize` into the regular `parEvalN` by using `listApp`. The resulting `[[b]]` is lastly converted into `[b]` with `arr concat`.

```

1 parEvalNLazy :: (ArrowParallel arr a b conf, ArrowChoice arr, ArrowApply arr) =>
2   conf -> ChunkSize -> [arr a b] -> (arr [a] [b])
3 parEvalNLazy conf chunkSize fs =
4   arr (chunksOf chunkSize) >>>
5   listApp fchunks >>>
6   arr concat
7   where
8     fchunks = map (parEvalN conf) $ chunksOf chunkSize fs

```

6.2 `parEval2`

We have only talked about parallelizing the computation of arrows of the same type up until now. But sometimes we want to parallelize heterogenous types as well. For this, we introduce a helper combinator `arrMaybe` first, that converts an arrow `arr a b` to an arrow `arr (Maybe a) (Maybe b)`.

```

1 arrMaybe :: (ArrowApply arr) => (arr a b) -> arr (Maybe a) (Maybe b)
2 arrMaybe fn = (arr $ go) >>> app
3   where
4     go Nothing = (arr $ \Nothing -> Nothing, Nothing)
5     go (Just a) = ((arr $ \(Just x) -> (fn, x)) >>> app >>> arr Just, (Just a))

```

With this, we can now easily write `parEval2` which combines two arrows `arr a b` and `arr c d` into a new parallel arrow `arr (a, c) (b, d)`. We start by converting both arrows with `arrMaybe`, combining them with `***` into a new arrow `arr (Maybe a, Maybe c) (Maybe b, Maybe d)`. This is then replicated twice and fed into `parEvalN` to get a `arr [(Maybe a, Maybe c)] [(Maybe b, Maybe d)]`. We can then apply this arrow to the input `[(Just a, Nothing), (Nothing, Just c)]` and then extract the resulting values with `fromJust` and the `!!` operator on lists in the last step.

```

1 parEval2 :: (ArrowParallel arr a b conf,
2   ArrowParallel arr (Maybe a, Maybe c) (Maybe b, Maybe d) conf,
3   ArrowApply arr) =>
4   conf -> arr a b -> arr c d -> (arr (a, c) (b, d))
5 parEval2 conf f g =
6   (arr $ \ (a, c) -> (f.g, [(Just a, Nothing), (Nothing, Just c)])) >>>
7   app >>>
8   (arr $ \comb -> (fromJust (fst (comb !! 0)), fromJust (snd (comb !! 1))))
9 where
10  f.g = parEvalN conf $ replicate 2 $ arrMaybe f *** arrMaybe g

```

6.3 parMap

`parMap` is probably the most common skeleton for parallel programs. We can implement it with `ArrowParallel` by repeating an arrow `arr a b` and then passing it into `parEvalN` to get an arrow `arr [a] [b]`. Just like `parEvalN`, `parMap` is 100 % strict.

```

1 parMap :: (ArrowParallel arr a b conf, ArrowApply arr) =>
2   conf -> (arr a b) -> (arr [a] [b])
3 parMap conf f =
4   (arr $ \as -> (f, as)) >>>
5   ( first $ arr repeat >>>
6     arr (parEvalN conf) ) >>>
7   app

```

6.4 parMapStream

As `parMap` is 100% strict it has the same restrictions as `parEvalN` compared to `parEvalNLazy`. So it makes sense to also have a `parMapStream` which behaves like `parMap`, but uses `parEvalNLazy` instead of `parEvalN`.

```

1 parMapStream :: (ArrowParallel arr a b conf, ArrowChoice arr, ArrowApply arr) =>
2   conf -> ChunkSize -> arr a b -> arr [a] [b]
3 parMapStream conf chunkSize f =
4   (arr $ \as -> (f, as)) >>>
5   ( first $ arr repeat >>>
6     arr (parEvalNLazy conf chunkSize) ) >>>
7   app

```


6.5 farm

`parMap` spawns every single computation in a new thread (at least for the instances of `ArrowParallel` we gave in this paper). This can be quite wasteful and a `farm` that equally distributes the workload over `numCores` workers (if `numCores` is greater than `actualProcessorCount`, the fastest processor(s) to finish will get more tasks) seems useful.

```
1 farm :: (ArrowParallel arr a b conf, ArrowParallel arr [a] [b] conf,  
2   ArrowChoice arr, ArrowApply arr) =>  
3   conf -> NumCores -> arr a b -> arr [a] [b]  
4 farm conf numCores f =  
5   (arr $ \as -> (f, as)) >>>  
6   ( first $ arr mapArr >>> arr repeat >>>  
7     arr (parEvalN conf)) >>>  
8   (second $ arr (unshuffle numCores)) >>>  
9   app >>>  
10  arr shuffle
```

The definition of `unshuffle` is

```
1 unshuffle :: Int  
2   -> [a]  
3   -> [[a]]  
4 unshuffle n xs = [takeEach n (drop i xs) | i <- [0..n-1]]
```

, while `shuffle` is defined as:

```
1 shuffle :: [[a]]  
2   -> [a]  
3 shuffle = concat . transpose
```

(These were taken from Eden's source code.)

6.6 farmChunk

As `farm` is basically just `parMap` with a different work distribution, it is, again, 100% strict. So we define `farmChunk` which uses `parEvalNLazy` instead of `parEvalN` like this:

```
1 farmChunk :: (ArrowParallel arr a b conf, ArrowParallel arr [a] [b] conf,  
2   ArrowChoice arr, ArrowApply arr) =>  
3   conf -> ChunkSize -> NumCores -> arr a b -> arr [a] [b]  
4 farmChunk conf chunkSize numCores f =  
5   (arr $ \as -> (f, as)) >>>  
6   ( first $ arr mapArr >>> arr repeat >>>  
7     arr (parEvalNLazy conf chunkSize)) >>>  
8   (second $ arr (unshuffle numCores)) >>>  
9   app >>>  
10  arr shuffle
```

7 Syntactic Sugar

To make using our new API a bit easier, we also introduce some syntactic sugar. We start with `|>>>|`, which is basically `>>>` on lists of arrows.

```
1 (|>>>|) :: (Arrow arr) => [arr a b] -> [arr b c] -> [arr a c]
2 (|>>>|) = zipWith (>>>)
```

For the basic Arrow case, we have the `***` combinator which allows us to combine two arrows `arr a b` and `arr c d` into an arrow `arr (a, c) (b, d)` which does both computations at once. This can easily be lifted into a parallel computation with `parEval2`, but requires a backend which has a implementation that does not require any configuration (hence the `()` as the `conf` parameter in the following code snippet).

```
1 (|***) :: (ArrowParallel arr a b (),
2   ArrowParallel arr (Maybe a, Maybe c) (Maybe b, Maybe d) ()),
3   ArrowApply arr) =>
4   arr a b -> arr c d -> arr (a, c) (b, d)
5 (|***) = parEval2 ()
```

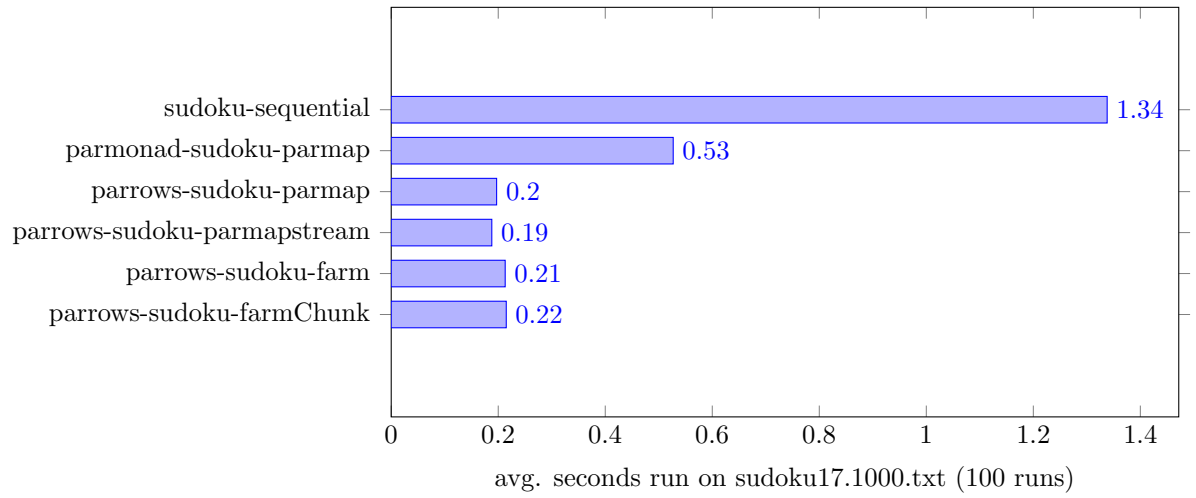
With this we can analogously to the serial `&&&` define the parallel `|&&&|`.

```
1 (|&&&|) :: (ArrowParallel arr a b (),
2   ArrowParallel arr (Maybe a, Maybe a) (Maybe b, Maybe c) ()),
3   ArrowApply arr) =>
4   arr a b -> arr a c -> arr a (b, c)
5 (|&&&|) f g = (arr $ \a -> (a, a)) >>> f |***) g
```

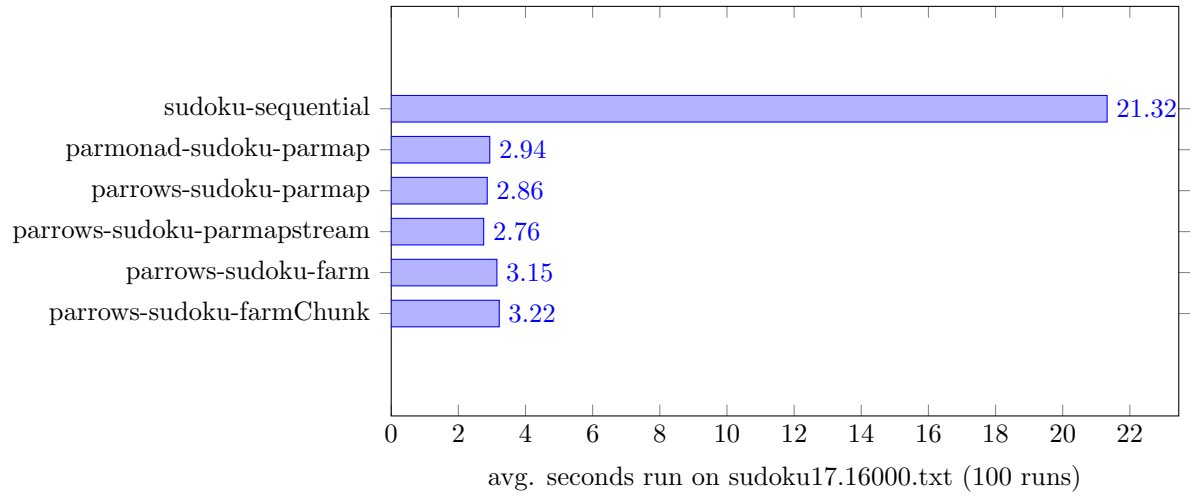
8 Benchmarks

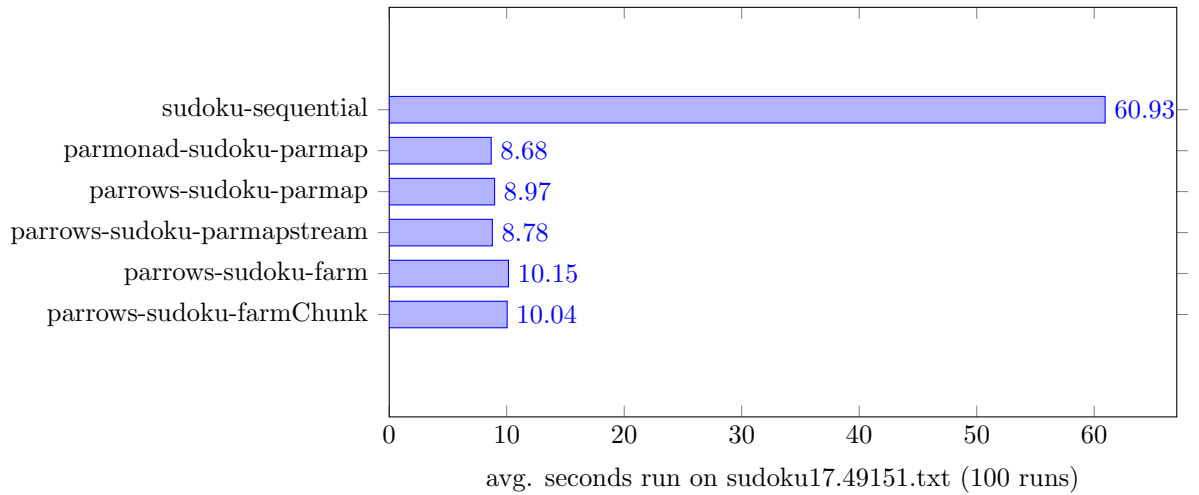
To check performance of our new parallel arrow API, we conducted some benchmarks. These are based on a sudoku solver from the examples for Simon Marlow's book "Parallel and Concurrent Programming in Haskell" which can be found on his GitHub [2]. The results are displayed in the following graphs which are ordered by problem size.

The Benchmarks were run on a Core i7-3970X CPU @ 3.5GHz with 6 cores and 12 threads. All versions were run 100 times. The resulting running-times were then averaged.



Note: the bad result for parmonad-sudoku-parmap seems to be a artefact as it performs much better in the other benchmarks





As we can see, the parallel arrow versions of the program all have around the same speedup as the ParMonad based version. This means that using the `ArrowParallel` typeclass doesn't add any real overhead. Also, the slightly worse results for "parrows-sudoku-farm" and "parrows-sudoku-farmChunk" can be explained by either imperfect parameters or by the fact that the benchmarks do not really require the scheduling introduced by the skeletons which in this case introduces unnecessary overhead. This would probably look different if the benchmark would use e.g. a divide and conquer approach to solve the sudoku puzzles.

9 Conclusion

References

- [1] Generalising Monads to Arrows <https://jcp.org/en/jsr/detail?id=220>, November 10, 1998
- [2] Sample code for Simon Marlow's book "Parallel and Concurrent Programming in Haskell" <https://github.com/simonmar/parconc-examples>, retrieved on 01/12/2017