Parallel Sequences

COS 326
Speaker: Andrew Appel
Princeton University

Credits:
Dan Grossman, UW
http://homes.cs.washington.edu/~djg/teachingMaterials/spac
Blelloch, Harper, Licata (CMU, Wesleyan)
Programming with shared mutable data is very hard!

With pure functional code and parallel futures, many error modes disappear

Are there more great abstractions like futures?
  – you betcha!
What if you had a really big job to do?

Example: Create an index of every web page on the planet.
- Google does that regularly!
- There are billions of them!

Example: Search Facebook for a friend or Twitter for a tweet.

To get big jobs done, we typically need 1000s of computers, but:
- how do we distribute work across all those computers?
- you definitely can't use shared-memory parallelism because the computers don't share memory!
- when you use 1 computer, you just hope it doesn't fail. If it does, you go to the store, buy a new one and restart the job.
- when you use 1000s of computers at a time, failures become the norm. What to do when 1 of 1000 computers fail? Start over?
Need high-level interfaces to shield application programmers from the complex details. Complex implementations solve the problems of distribution, fault tolerance and performance.

Common abstraction: Parallel collections

Example collections: sets, tables, dictionaries, sequences
Example bulk operations: create, map, reduce, join, filter
COMPLEXITY OF PARALLEL ALGORITHMS
let x = 1 + 2 in
3 + x
let \( x = 1 + 2 \) in

\( 3 + x \)

- \( x = 1 + 2 \) cost = 1
- \( 3 + x \) cost = 1
let x = 1 + 2 in
3 + x

 dependence:
x = 1 + 2 happens before 3 + x
 Execution of dependency diagrams: A processor can only begin executing the computation associated with a block when the computations of all of its predecessor blocks have been completed.
Visualizing Computational Costs

step 1: execute first block

Cost so far: 0
Visualizing Computational Costs

step 1:
execute first block

Cost so far: 1
Visualizing Computational Costs

step 2: execute second block because all of its predecessors have been completed

Cost so far: 1
step 2: execute second block because all of its predecessors have been completed

Cost so far: 1 + 1
let x = 1 + 2 in
3 + x

x = 1 + 2  
cost = 1

3 + x  
cost = 1

total cost
= 1 + 1
= 2
(1 + 2 || f 3)

parallel pair:
compute both left and right-hand sides independently
return pair of values
(easy to implement using futures)
Visualizing Computational Costs

\[(1 + 2 \ || \ f \ 3)\]

- **A**: cost = 1
- **B**: 1 + 2, cost = 1
- **C**: f 3, cost = 7
- **D**: ( ), cost = 1

The diagram shows the computational costs associated with the expression \((1 + 2 \ || \ f \ 3)\).
Suppose we have 1 processor. How much time does this computation take?
Visualizing Computational Costs

Suppose we have 1 processor. How much time does this computation take? Schedule A-B-C-D: \( 1 + 1 + 7 + 1 \)
Suppose we have 1 processor. How much time does this computation take?
Schedule A-C-B-D: $1 + 1 + 7 + 1$
Suppose we have 2 processors. How much time does this computation take?
Suppose we have 2 processors. How much time does this computation take?  
Cost so far: 1
Suppose we have 2 processors. How much time does this computation take? Cost so far: $1 + \max(1,7)$
Suppose we have 2 processors. How much time does this computation take? Cost so far: 1 + max(1,7) + 1
Suppose we have 2 processors. How much time does this computation take? Total cost: $1 + \max(1,7) + 1$. We say the schedule we used was: A-CB-D
Suppose we have 3 processors. How much time does this computation take?
Suppose we have 3 processors. How much time does this computation take? Schedule A-BC-D: $1 + \max(1,7) + 1 = 9$
Suppose we have infinite processors. How much time does this computation take?

Schedule A-BC-D: $1 + \max(1,7) + 1 = 9$
Understanding the complexity of a parallel program is a little more complex than a sequential program:

- the number of processors has a significant effect

One way to approximate the cost is to consider a parallel algorithm independently of the machine it runs on is to consider two metrics:

- **Work**: The cost of executing a program with just 1 processor.
- **Span**: The cost of executing a program with an infinite number of processors

Always good to minimize work:

- Every instruction executed consumes energy
- Minimize span as a second consideration
- Communication costs are also crucial (we are ignoring them)
Parallelism

The **parallelism** of an algorithm is an estimate of the maximum number of processors an algorithm can profit from.

- parallelism = \( \frac{\text{work}}{\text{span}} \)

If work = span then parallelism = 1.
- We can only use 1 processor
- It's a sequential algorithm

If span = \( \frac{1}{2} \) work then parallelism = 2
- We can use up to 2 processors

If work = 100, span = 1
- All operations are independent & can be executed in parallel
- We can use up to 100 processors
Series-parallel graphs arise from execution of functional programs with parallel pairs. Also known as well-structured, nested parallelism.
let both \( f \ x \ g \ y = \)
let \( ff = \text{future} \ f \ x \) in
let \( gv = g \ y \) in
(\text{force} \ ff, \ gv)
In general, a series-parallel graph has a source and a sink and is:

- a single node, or
- two series-parallel graphs in sequence, or
- two series-parallel graphs in parallel
However:
The results about greedy schedulers (next few slides) do apply to DAG schedules as well as series-parallel schedules!
Let's assume each node costs 1.

- **Work**: sum the nodes.
- **Span**: longest path from source to sink.
Let's assume each node costs 1.

**Work**: sum the nodes.

**Span**: longest path from source to sink.

work = 10
span = 5
Let's assume each node costs 1.

Let's assume we have 2 processors. How do we schedule computation?
Let's assume each node costs 1.

Let's assume we have 2 processors. How do we schedule computation?

Option 1:
A
B G
C D
Let's assume each node costs 1.

Let's assume we have 2 processors. How do we schedule computation?

Option 1:
A
B G
C D
E H

Option 2:
Let's assume each node costs 1.

Let's assume we have 2 processors. How do we schedule computation?

Option 1:
A
B G
C D
E H
I
J
F
Let's assume each node costs 1.

Let's assume we have 2 processors. How do we schedule computation?

Option 1:
A
B G
C D
E H
F
I J
Let's assume each node costs 1.

Let's assume we have 2 processors. How do we schedule computation?

Option 1:

A
B G
C D
E H
I
J
F
Let's assume each node costs 1.

Let's assume we have 2 processors. How do we schedule computation?

Option 1:

A
B G
C D
E H
H I
J
F
Let's assume each node costs 1.

Let's assume we have 2 processors. How do we schedule computation?

Option 1:
A
B G
C D
E H
H I
I
J
E J
J
F
Let's assume each node costs 1.

Let's assume we have 2 processors. How do we schedule computation?

Option 1:

A
B G
C D
E H
I
F J
J
F
Let's assume each node costs 1.

Let's assume we have 2 processors. How do we schedule computation?

Option 1:
A
B G
C D
E H
H I
I
J
F

Conclusion:
How you schedule jobs can have an impact on performance
Greedy Schedulers

Greedy schedulers will schedule some task to a processor as soon as that processor is free.

– Doesn't sound so smart!
Greedy Schedulers

Greedy schedulers will schedule some task to a processor as soon as that processor is free.

– Doesn't sound so smart!

Properties (for p processors):

– $T(p) < \frac{\text{work}}{p} + \text{span}$
  • won't be worse than dividing up the data perfectly between processors, except for the last little bit, which causes you to add the span on top of the perfect division

– $T(p) \geq \max(\frac{\text{work}}{p}, \text{span})$
  • can't do better than perfect division between processors ($\frac{\text{work}}{p}$)
  • can't be faster than span
Greedy Schedulers

Properties (for p processors):

\[ \max(\text{work}/p, \text{span}) \leq T(p) < \text{work}/p + \text{span} \]

Consequences:

– as span gets small relative to \( \text{work}/p \)
  - \( \text{work}/p + \text{span} \rightarrow \text{work}/p \)
  - \( \max(\text{work}/p, \text{span}) \rightarrow \text{work}/p \)
  - so \( T(p) \rightarrow \text{work}/p \) -- greedy schedulers converge to the optimum!

– if span approaches the work
  - \( \text{work}/p + \text{span} \rightarrow \text{span} \)
  - \( \max(\text{work}/p, \text{span}) \rightarrow \text{span} \)
  - so \( T(p) \rightarrow \text{span} \) – greedy schedulers converge to the optimum!
And therefore

Even though greedy schedulers are simple to implement,

they can be effective in building a parallel programming system.

and

This *supports* the idea that *work and span* are useful ways to reason about the cost of parallel programs.
PARALLEL SEQUENCES
Parallel Sequences

Parallel sequences

\[ \langle e_1, e_2, e_3, \ldots, e_n \rangle \]

Operations:

- creation (called \textit{tabulate})
- indexing an element in constant span
- map
- scan -- like a fold: \( \langle u, u + e_1, u + e_1 + e_2, \ldots \rangle \) \( \log n \) span!

Languages:

- Nesl [Blelloch]
- Data-parallel Haskell
tabulate : (int -> 'a) -> int -> 'a seq

\[ \text{tabulate } f \ n \ = \ <f \ 0, \ f \ 1, \ \ldots, \ f \ (n-1)> \]

work = \( O(n) \) \quad \text{span} = \( O(1) \)
Parallel Sequences: Selected Operations

\textbf{tabulate}: \((\text{int} \rightarrow \text{'a}) \rightarrow \text{int} \rightarrow \text{'a seq}\)

\text{tabulate } f \ n = \langle f \ 0, \ f \ 1, \ldots, \ f \ (n-1) \rangle

\text{work} = O(n) \quad \text{span} = O(1)

\textbf{nth}: \text{'a seq} \rightarrow \text{int} \rightarrow \text{'a}

\text{nth} <e_0, \ e_1, \ldots, \ e(n-1)> \ i = \ e_i

\text{work} = O(1) \quad \text{span} = O(1)
Parallel Sequences: Selected Operations

**tabulate** : (int -> 'a) -> int -> 'a seq

\[ \text{tabulate } f \ n \ = \ <f \ 0, \ f \ 1, \ ... , \ f \ (n-1)> \]

work = O(n) \hspace{1cm} \text{span} = O(1)

**nth** : 'a seq -> int -> 'a

\[ \text{nth } <e_0, \ e_1, \ ... , \ e_(n-1)> \ i \ = \ e_i \]

work = O(1) \hspace{1cm} \text{span} = O(1)

**length** : 'a seq -> int

\[ \text{length } <e_0, \ e_1, \ ... , \ e_(n-1)> \ = \ n \]

work = O(1) \hspace{1cm} \text{span} = O(1)
Example Problems

Write a function that creates the sequence \(<0, \ldots, n-1>\) with \(\text{Span} = O(1)\) and \(\text{Work} = O(n)\).
Example Problems

Write a function that creates the sequence \(<0, \ldots, n-1>\) with \(\text{Span} = O(1)\) and \(\text{Work} = O(n)\).

```plaintext
(* create n == \(<0, 1, \ldots, n-1>\) *)
let create n =
```

<table>
<thead>
<tr>
<th>Operations</th>
<th>Work</th>
<th>Span</th>
</tr>
</thead>
<tbody>
<tr>
<td>tabulate f n</td>
<td>n</td>
<td>1</td>
</tr>
<tr>
<td>nth i s</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>length s</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
Example Problems

Write a function that creates the sequence \(<0, \ldots, n-1>\) with \(\text{Span} = O(1)\) and \(\text{Work} = O(n)\).

(* create \(n == <0, 1, \ldots, n-1>\) *)
let create \(n =\)
    tabulate (fun \(i \rightarrow i\)) \(n\)

Operations:

<table>
<thead>
<tr>
<th>Operation</th>
<th>Work</th>
<th>Span</th>
</tr>
</thead>
<tbody>
<tr>
<td>tabulate (f\ n)</td>
<td>(n)</td>
<td>1</td>
</tr>
<tr>
<td>nth (i\ s)</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>length (s)</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
Example Problems

Write a function such that given a sequence \( <v_0, ..., v_{n-1}> \), maps \( f \) over each element of the sequence with \( \text{Span} = O(1) \) and \( \text{Work} = O(n) \), returning the new sequence (if \( f \) is constant work).

Operations:

<table>
<thead>
<tr>
<th>Operation</th>
<th>Work</th>
<th>Span</th>
</tr>
</thead>
<tbody>
<tr>
<td>tabulate ( f ) ( n )</td>
<td>( n )</td>
<td>1</td>
</tr>
<tr>
<td>nth ( i ) ( s )</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>length ( s )</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
Example Problems

Write a function such that given a sequence \(<v_0, \ldots, v_{n-1}>\), maps \(f\) over each element of the sequence with \(\text{Span} = O(1)\) and \(\text{Work} = O(n)\), returning the new sequence (if \(f\) is constant work)

\[
(* \text{ map } f <v_0, \ldots, v_{n-1}> == <f v_0, \ldots, f v_{n-1}> *)
let map f s =
\]

Operations:

<table>
<thead>
<tr>
<th>Operation</th>
<th>Work</th>
<th>Span</th>
</tr>
</thead>
<tbody>
<tr>
<td>tabulate (f, n)</td>
<td>(n)</td>
<td>1</td>
</tr>
<tr>
<td>(\text{nth } i\ s)</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>(\text{length } s)</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
Example Problems

Write a function such that given a sequence \(<v_0, \ldots, v_{n-1}>\), maps \(f\) over each element of the sequence with \(\text{Span} = O(1)\) and \(\text{Work} = O(n)\), returning the new sequence (if \(f\) is constant work)

\[
(* \: \text{map} \: f \: <v_0, \ldots, v_{n-1}> \: == \: <f \: v_0, \ldots, f \: v_{n-1}> *)
\]

\[
\text{let} \: \text{map} \: f \: s =
\]

\[
\text{tabulate} \: (\text{fun} \: i \rightarrow \: f \: (\text{nth} \: s \: i)) \: (\text{length} \: s)
\]

Operations:

<table>
<thead>
<tr>
<th>Operation</th>
<th>Work</th>
<th>Span</th>
</tr>
</thead>
<tbody>
<tr>
<td>tabulate (f) (n)</td>
<td>(n)</td>
<td>1</td>
</tr>
<tr>
<td>(\text{nth} : i : s)</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>(\text{length} : s)</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
Example Problems

Write a function such that given a sequence \(<v_0, ..., v_{n-1}>\), reverses the sequence. with Span = O(1) and Work = O(n)

<table>
<thead>
<tr>
<th>Operation</th>
<th>Work</th>
<th>Span</th>
</tr>
</thead>
<tbody>
<tr>
<td>tabulate f n</td>
<td>n</td>
<td>1</td>
</tr>
<tr>
<td>nth i s</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>length s</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
Example Problems

Write a function such that given a sequence \(<v_0, ..., v_{n-1}>\), reverses the sequence. with Span = O(1) and Work = O(n)

(* reverse \(<v_0, ..., v_{n-1}>\) == \(<v_{n-1}, ..., v_0>\) *)
let reverse s =

Operations:

<table>
<thead>
<tr>
<th>Operation</th>
<th>Work</th>
<th>Span</th>
</tr>
</thead>
<tbody>
<tr>
<td>tabulate f n</td>
<td>n</td>
<td>1</td>
</tr>
<tr>
<td>nth i s</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>length s</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
Write a function such that given a sequence \(<v_0, \ldots, v_{n-1}>\), reverses the sequence. with Span = O(1) and Work = O(n)

(* reverse \(<v_0, \ldots, v_{n-1}> == <v_{n-1}, \ldots, v_0> *)
let reverse s =
  let n = length s in
  tabulate (fun i -> nth s (n-i-1)) n

Operations:

<table>
<thead>
<tr>
<th>Operation</th>
<th>Work</th>
<th>Span</th>
</tr>
</thead>
<tbody>
<tr>
<td>tabulate f n</td>
<td>n</td>
<td>1</td>
</tr>
<tr>
<td>nth i s</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>length s</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
A Parallel Sequence API

**Type**

```plaintext
type 'a seq
```

**Functions**

<table>
<thead>
<tr>
<th>Function</th>
<th>Description</th>
<th>Work</th>
<th>Span</th>
</tr>
</thead>
<tbody>
<tr>
<td>tabulate</td>
<td>(int -&gt; 'a) -&gt; int -&gt; 'a seq</td>
<td>O(N)</td>
<td>O(1)</td>
</tr>
<tr>
<td>length</td>
<td>'a seq -&gt; int</td>
<td>O(1)</td>
<td>O(1)</td>
</tr>
<tr>
<td>nth</td>
<td>'a seq -&gt; int -&gt; 'a</td>
<td>O(1)</td>
<td>O(1)</td>
</tr>
<tr>
<td>append</td>
<td>'a seq -&gt; 'a seq -&gt; 'a seq</td>
<td>O(N+M)</td>
<td>O(1)</td>
</tr>
<tr>
<td></td>
<td>(can build this from tabulate, nth, length)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>split</td>
<td>'a seq -&gt; int -&gt; 'a seq * 'a seq</td>
<td>O(N)</td>
<td>O(1)</td>
</tr>
</tbody>
</table>

For efficient implementations, see Blelloch's NESL project:

http://www.cs.cmu.edu/~scandal/nesl.html
We have seen many sequential algorithms can be programmed succinctly using fold or reduce. Eg: sum all elements:

sum: 0

7 4 3 9 8
We have seen many sequential algorithms can be programmed succinctly using fold or reduce. Eg: sum all elements:

```
sum: 0 7
```

```
7 4 3 9 8
```

```
sum: 7
```
We have seen many sequential algorithms can be programmed succinctly using fold or reduce. Eg: sum all elements:

```
7 4 3 9 8
```

sum: 0 → 7 → 11 → 14 → 23 → 31
Fold and Reduce

We have seen many sequential algorithms can be programmed succinctly using fold or reduce. Eg: sum all elements:

```
let sum_all (l:int list) = reduce (+) 0 l
```

```
let sum_all [7; 4; 3; 9; 8] = reduce (+) 0 [7; 4; 3; 9; 8]
```
We have seen many sequential algorithms can be programmed succinctly using fold or reduce. Eg: sum all elements:

```
let sum_all (l:int list) = reduce (+) 0 l
```

**Key to parallelization:** Notice that because sum is an *associative* operator, we do not have to add the elements strictly left-to-right:

\[
(((init + v1) + v2) + v3) + v4) + v5) \equiv ((init + v1) + v2) + ((v3 + v4) + v5)
\]
The key is *associativity*:

\[
(((\text{init} + v1) + v2) + v3) + v4) + v5) = ((\text{init} + v1) + v2) + (v3 + v4) + v5
\]

*Commutativity* not needed!

*Commutativity* allows us to reorder the elements:

\[
v1 + v2 = v2 + v1
\]

But we don't have to reorder elements to obtain a significant speedup; we just have to reorder the execution of the operations.
Parallel Sum

2 7 4 3 9 8 2 1
Parallel Sum

2 7 4 3 9 8 2 1

2 7 4 3

9 8 2 1
Parallel Sum

2  7  4  3  9  8  2  1

2  7  4  3

2  7

2  7

2  7

2  7

2  7

2  7

2  7

2  7
Parallel Sum

\[
\begin{align*}
9 &+ 7 &+ 17 &+ 3 \\
2 &+ 7 &+ 4 &+ 3 &+ 9 &+ 8 &+ 2 &+ 1
\end{align*}
\]
let rec psum (s : int seq) : int =
  match length s with
  0 -> 0
 | 1 -> nth s 0
 | n ->
  let (s1, s2) = split (n/2) s in
  let (a1, a2) = both psum s1
  psum s2 in
  a1 + a2

let both f x g y =
  let ff = future f x in
  let gv = g y in
  (force ff, gv)
If \( \text{op} \) is associative and the base case has the properties:

\[
\text{op base } X == X \quad \text{and} \quad \text{op } X \text{ base } == X
\]

then the parallel reduce is equivalent to the sequential left-to-right fold.
let rec reduce (f:'a -> 'a -> 'a) (base:'a) (s:'a seq) =
  match length s with
  | 0 -> base
  | 1 -> nth s 0
  | n ->
    let (s1,s2) = split (n/2) s in
    let (n1, n2) = both (reduce f base) s1
                     (reduce f base) s2 in
    f n1 n2
let rec reduce (f:'a -> 'a -> 'a) (base:'a) (s:'a seq) =
  match length s with
  | 0 -> base
  | 1 -> nth s 0
  | n ->
    let (s1,s2) = split (n/2) s in
    let (n1, n2) = both (reduce f base) s1 (reduce f base) s2 in
    f n1 n2

let sum s = reduce (+) 0 s
let rec mapreduce (inject: 'a -> 'b)
        (combine:'b -> 'b -> 'b)
        (base:'b)
        (s:'a seq) =

  match length s with
  0 -> base
| 1 -> inject (nth s 0)
| n ->
    let (s1,s2) = split (n/2) s in
    let (n1, n2) = both
        (mapreduce inject combine base) s1
    (mapreduce inject combine base) s2 in
    combine n1 n2
let rec mapreduce (inject: 'a -> 'b) (combine:'b -> 'b -> 'b) (base:'b) (s:'a seq) =
match length s with
  0 -> base
| 1 -> inject (nth s 0)
| n ->
  let (s1,s2) = split (n/2) s in
  let (n1, n2) = both
    (mapreduce inject combine base) s1
    (mapreduce inject combine base) s2 in
  combine n1 n2

let average s =
  let (count, total) =
    mapreduce (fun x -> (1,x))
    (fun (c1,t1) (c2,t2) -> (c1+c2, t1 + t2))
    (0,0) s in
  if count = 0 then 0 else total / count
DON’T PARALLELIZE AT TOO FINE A GRAIN
Parallel Reduce with Sequential Cut-off

When data is small, the overhead of parallelization isn't worth it. Revert to the sequential version!

let SHORT = 1000

let rec reduce (f:'a -> 'a -> 'a) (base:'a) (s:'a seq) =
    if length s < SHORT
    then sequential_reduce f base s
    else let (s1,s2) = split ((length s)/2) s in
        let (n1, n2) = both (reduce f base) s1 (reduce f base) s2 in
        f n1 n2

let sequential_reduce f base (s:'a seq) =
    let rec g i x =
        if i<0 then x else g (i-1) (f (nth a i) x)
    in g (length s - 1)