Scan – Algorithm Effects on Parallelism and Memory Conflicts

Parallel Prefix Sum (Scan)

• Definition:
The all-prefix-sums operation takes a binary associative operator $\oplus$ with identity $I$, and an array of $n$ elements $[a_0, a_1, \ldots, a_{n-1}]$ and returns the ordered set $[I, a_0, (a_0 \oplus a_1), \ldots, (a_0 \oplus a_1 \oplus \ldots \oplus a_{n-2})]$.

• Example:
if $\oplus$ is addition, then scan on the set $[3 \ 7 \ 0 \ 4 \ 1 \ 6 \ 3]$ returns the set $[0 \ 3 \ 4 \ 11 \ 11 \ 15 \ 16 \ 22]$ (From Blelloch, 1990, “Prefix Sums and Their Applications”)

Exclusive scan: last input element is not included in the result
Applications of Scan

- Scan is a simple and useful parallel building block
  - Convert recurrences from sequential:
    \[
    \text{for}(j=1; j<n; j++) \\
    \text{out}[j] = \text{out}[j-1] + f(j);
    \]
  - into parallel:
    \[
    \text{forall}(j) \{ \text{temp}[j] = f(j) \}; \\
    \text{scan(out, temp)};
    \]
- Useful for many parallel algorithms:
  - radix sort
  - quicksort
  - String comparison
  - Lexical analysis
  - Stream compaction
  - Polynomial evaluation
  - Solving recurrences
  - Tree operations
  - Histograms
  - Etc.

Scan on the CPU

```c
void scan( float* scanned, float* input, int length) 
{
    scanned[0] = 0;
    for(int i = 1; i < length; ++i)
    {
        scanned[i] = input[i-1] + scanned[i-1];
    }
}
```

- Just add each element to the sum of the elements before it
- Trivial, but sequential
- Exactly \( n \) adds: optimal in terms of work efficiency
A First-Attempt Parallel Scan Algorithm

1. Read input from device memory to shared memory. Set first element to zero and shift others right by one.

Each thread reads one value from the input array in device memory into shared memory array T0. Thread 0 writes 0 into shared memory array.

0

In

T0

0

3

1

7

0

4

1

6

3

Iteration #1

Stride 1

T1

0

3

1

7

0

4

1

6

0

4

8

7

4

5

7

A First-Attempt Parallel Scan Algorithm

1. (previous slide)

2. Iterate log(n) times: Threads stride to n: Add pairs of elements stride elements apart. Double stride at each iteration. (note must double buffer shared mem arrays)

• Active threads: stride to n-1 (n-stride threads)
• Thread j adds elements j and j-stride from T0 and writes result into shared memory buffer T1 (ping-pong)
A First-Attempt Parallel Scan Algorithm

1. Read input from device memory to shared memory. Set first element to zero and shift others right by one.

2. Iterate log(n) times: Threads stride to n: Add pairs of elements stride elements apart. Double stride at each iteration. (note must double buffer shared mem arrays)

Iteration #2
Stride = 2

Iteration #3
Stride = 4
A First-Attempt Parallel Scan Algorithm

1. Read input from device memory to shared memory. Set first element to zero and shift others right by one.

2. Iterate log(n) times: Threads stride to n: Add pairs of elements stride elements apart. Double stride at each iteration. (note must double buffer shared mem arrays)

3. Write output to device memory.

Work Efficiency Considerations

- The first-attempt Scan executes log(n) parallel iterations
  - The steps do (n/2 + n/2-1), (n/4+ n/2-1), (n/8+n/2-1)...(1+ n/2-1) adds each
  - Total adds: n * (log(n) – 1) + 1 \( \rightarrow \) O(n*\log(n)) work

- This scan algorithm is not very work efficient
  - Sequential scan algorithm does \( n \) adds
  - A factor of \( \log(n) \) hurts: 20x for \( 10^6 \) elements!

- A parallel algorithm can be slow when execution resources are saturated due to low work efficiency
Improving Efficiency

- A common parallel algorithm pattern: 
  *Balanced Trees*
  - Build a balanced binary tree on the input data and sweep it to and from the root
  - Tree is not an actual data structure, but a concept to determine what each thread does at each step

- For scan:
  - Traverse down from leaves to root building partial sums at internal nodes in the tree
    - Root holds sum of all leaves
  - Traverse back up the tree building the scan from the partial sums

---

Build the Sum Tree

| T | 3 | 1 | 7 | 0 | 4 | 1 | 6 | 3 |

Assume array is already in shared memory
Build the Sum Tree

Stride 1
Iteration 1, \( n/2 \) threads

\[
\begin{array}{cccccccc}
T & 3 & 1 & 7 & 0 & 4 & 1 & 6 & 3 \\
\end{array}
\]

Each \( \bullet \) corresponds to a single thread.

Iterate log\( (n) \) times. Each thread adds value \( stride \) elements away to its own value.

Stride 2
Iteration 2, \( n/4 \) threads

\[
\begin{array}{cccccccc}
T & 3 & 4 & 7 & 7 & 4 & 5 & 6 & 9 \\
\end{array}
\]

Each \( \bullet \) corresponds to a single thread.

Iterate log\( (n) \) times. Each thread adds value \( stride \) elements away to its own value.
Build the Sum Tree

<table>
<thead>
<tr>
<th>T</th>
<th>3</th>
<th>1</th>
<th>7</th>
<th>0</th>
<th>4</th>
<th>1</th>
<th>6</th>
<th>3</th>
</tr>
</thead>
</table>

Stride 1

<table>
<thead>
<tr>
<th>T</th>
<th>3</th>
<th>4</th>
<th>7</th>
<th>7</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>9</th>
</tr>
</thead>
</table>

Stride 2

<table>
<thead>
<tr>
<th>T</th>
<th>3</th>
<th>4</th>
<th>7</th>
<th>11</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>14</th>
</tr>
</thead>
</table>

Stride 4

<table>
<thead>
<tr>
<th>T</th>
<th>3</th>
<th>4</th>
<th>7</th>
<th>11</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>25</th>
</tr>
</thead>
</table>

Iterate log(n) times. Each thread adds value \textit{stride} elements away to its own value.

Note that this algorithm operates in-place: no need for double buffering.

Zero the Last Element

<table>
<thead>
<tr>
<th>T</th>
<th>3</th>
<th>4</th>
<th>7</th>
<th>11</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>0</th>
</tr>
</thead>
</table>

We now have an array of partial sums. Since this is an exclusive scan, set the last element to zero. It will propagate back to the first element.
Build Scan From Partial Sums

Each thread corresponds to a single thread.

Iterate log(n) times. Each thread adds value \textit{stride} elements away to its own value, and sets the value \textit{.stride} elements away to its own \textit{previous} value.
Build Scan From Partial Sums

Iterate log(n) times. Each thread adds value \textit{stride} elements away to its own value, and sets the value \textit{stride} elements away to its own \textit{previous} value.

Done! We now have a completed scan that we can write out to device memory.

Total steps: $2 \cdot \log(n)$.
Total work: $2 \cdot (n-1)$ adds = $O(n)$ Work Efficient!