List Processor Example

- Design a circuit that forms the sum of all the 2's complements integers stored in a linked-list structure starting at memory address 0:

```
  0   X0   X1   X2   ...   Xn-1
```

- All integers and pointers are 8-bit. The link-list is stored in a memory block with an 8-bit address port and 8-bit data port, as shown below. The pointer from the last element in the list is 0.

I/Os:
- START resets to head of list and starts addition process.
- DONE signals completion
- R, Bus that holds the final result
Architecture #1

Direct implementation of RTL description:

Datapath

Controller

If (START==1) NEXT=0, SUM=0;
repeat {
SUM = SUM + Memory[NEXT+1];
NEXT = Memory[NEXT];
} until (NEXT==0);
R = SUM, DONE = 1;

T > 31ns, F < 32 MHz

Cycle Optimized Version

• Architecture #2:

If (START==1) NEXT=0, SUM=0, NUMA=1;
repeat {
SUM = SUM + Memory[NUMA];
NUMA = Memory[NEXT] + 1, NEXT = Memory[NEXT];
} until (NEXT==0);
R = SUM, DONE = 1;

T > 24ns, F < 43MHz

• Incremental cost: addition of another register and mux.
Cycle Optimized with reduced cost

- Incremental cost:
  - Addition of another mux and control. Removal of an 8-bit adder.
- Performance:
  - mux adds 1ns to cycle time. 24ns, 41.67MHz.
- Is the cost savings worth the performance degradation?

Resource Utilization Charts

- One way to visualize these (and other possible) optimizations is through the use of a resource utilization charts.
- These are used in high-level design to help schedule operations on shared resources.
- Resources are listed on the y-axis. Time (in cycles) on the x-axis.
- Example:

<table>
<thead>
<tr>
<th></th>
<th>cycle</th>
</tr>
</thead>
<tbody>
<tr>
<td>memory</td>
<td></td>
</tr>
<tr>
<td>bus</td>
<td></td>
</tr>
<tr>
<td>register-file</td>
<td></td>
</tr>
<tr>
<td>ALU</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>memory</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>register-file</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALU</td>
<td>A1+B1</td>
<td>A2+B2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Our list processor has two shared resources: memory and adder.
## List Example Resource Scheduling

- **Unoptimized solution:** 1. \( \text{SUM} \leftarrow \text{SUM} + \text{Memory}[\text{NEXT}+1] \); 2. \( \text{NEXT} \leftarrow \text{Memory}[\text{NEXT}] \).

<table>
<thead>
<tr>
<th>memory</th>
<th>fetch x</th>
<th>fetch next</th>
<th>fetch x</th>
<th>fetch next</th>
</tr>
</thead>
<tbody>
<tr>
<td>adder1</td>
<td>next+1</td>
<td></td>
<td></td>
<td>next+1</td>
</tr>
<tr>
<td>adder2</td>
<td>sum</td>
<td></td>
<td></td>
<td>sum</td>
</tr>
</tbody>
</table>

- **Optimized solution:** 1. \( \text{SUM} \leftarrow \text{SUM} + \text{Memory}[\text{NUMA}] \); 2. \( \text{NEXT} \leftarrow \text{Memory}[\text{NEXT}], \text{NUMA} \leftarrow \text{Memory}[\text{NEXT}]+1 \).

<table>
<thead>
<tr>
<th>memory</th>
<th>fetch x</th>
<th>fetch next</th>
<th>fetch x</th>
<th>fetch next</th>
</tr>
</thead>
<tbody>
<tr>
<td>adder</td>
<td></td>
<td>numa</td>
<td>numa</td>
<td>sum</td>
</tr>
</tbody>
</table>

- **How about the other combination:** \( \text{add } x \text{ register} \)

<table>
<thead>
<tr>
<th>memory</th>
<th>fetch x</th>
<th>fetch next</th>
<th>fetch x</th>
<th>fetch next</th>
</tr>
</thead>
<tbody>
<tr>
<td>adder</td>
<td>numa</td>
<td>sum</td>
<td>numa</td>
<td>sum</td>
</tr>
</tbody>
</table>

1. \( x \leftarrow \text{Memory}[\text{NUMA}], \text{NUMA} \leftarrow \text{NEXT}+1 \); 2. \( \text{NEXT} \leftarrow \text{Memory}[\text{NEXT}], \text{SUM} \leftarrow \text{SUM}+x \).

- **Does this work?** If so, a very short clock period. Each cycle could have *independent* fetch and add. \( T = \max(T_{\text{mem}}, T_{\text{add}}) \) instead of \( T_{\text{mem}} + T_{\text{add}} \).

---

## List Example Resource Scheduling

- **Schedule one loop iteration followed by the next:**

<table>
<thead>
<tr>
<th>Memory</th>
<th>next(_1)</th>
<th>x(_1)</th>
<th>x(_2)</th>
<th>next(_2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>adder</td>
<td>numa(_1)</td>
<td>sum(_1)</td>
<td>numa(_2)</td>
<td>sum(_2)</td>
</tr>
</tbody>
</table>

- **How can we overlap iterations?** \( \text{next}_2 \) depends on \( \text{next}_1 \).
  - “slide” second iteration into first (4 cycles per result):

<table>
<thead>
<tr>
<th>Memory</th>
<th>next(_1)</th>
<th>x(_1)</th>
<th>x(_2)</th>
<th>next(_2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>adder</td>
<td>numa(_1)</td>
<td>sum(_1)</td>
<td>numa(_2)</td>
<td>sum(_2)</td>
</tr>
</tbody>
</table>

  - or further:

<table>
<thead>
<tr>
<th>Memory</th>
<th>next(_1)</th>
<th>x(_1)</th>
<th>x(_2)</th>
<th>x(_3)</th>
<th>x(_4)</th>
<th>next(_2)</th>
<th>next(_3)</th>
<th>x(_3)</th>
<th>x(_4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>adder</td>
<td>numa(_1)</td>
<td>numa(_2)</td>
<td>sum(_1)</td>
<td>sum(_2)</td>
<td>numa(_3)</td>
<td>numa(_4)</td>
<td>sum(_3)</td>
<td>sum(_4)</td>
<td></td>
</tr>
</tbody>
</table>

The repeating pattern is 4 cycles. Not exactly the pattern what we were looking for. But does it work correctly?
List Example Resource Scheduling

- In this case, first spread out, then pack.

<table>
<thead>
<tr>
<th>Memory</th>
<th>next₁</th>
<th>x₁</th>
<th>numa₁</th>
<th>sum₁</th>
</tr>
</thead>
<tbody>
<tr>
<td>adder</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Memory</th>
<th>next₂</th>
<th>x₂</th>
<th>numa₂</th>
<th>sum₂</th>
</tr>
</thead>
<tbody>
<tr>
<td>adder</td>
<td>numa₁</td>
<td>sum₁</td>
<td>numa₂</td>
<td>sum₃</td>
</tr>
</tbody>
</table>

1. \( x \leftarrow \text{Memory[NUMA]), \quad \text{NUMA} \leftarrow \text{NEXT} + 1; \)
2. \( \text{NEXT} \leftarrow \text{Memory[NEXT]}, \quad \text{SUM} \leftarrow \text{SUM} + X; \)

- Three different loop iterations active at once.
- Short cycle time (no dependencies within a cycle)
- Full utilization (only 2 cycles per result)
- Initialization: \( x = 0, \quad \text{numa} = 1, \quad \text{sum} = 0, \quad \text{next} = \text{memory}[0] \)
- Extra control states (out of the loop)
  - One to initialize next
  - One to finish off. 2 cycles after next == 0.

5. Optimization, Architecture #4

- Datapath:

- Incremental cost:
  - Addition of another register & mux, adder mux, and control.

- Performance: find max time of the four actions
  1. \( x \leftarrow \text{Memory[NUMA]}, \quad \text{NUMA} \leftarrow \text{NEXT} + 1; \) \( 0.5 + 1 + 10 + 1 + 1 + 0.5 = 14\text{ns} \)
  2. \( \text{NEXT} \leftarrow \text{Memory[NEXT]}, \quad \text{SUM} \leftarrow \text{SUM} + X; \) \( \text{same for all} = T > 14\text{ns}, \quad F < 71\text{MHz} \)
Other Optimizations

• Node alignment restriction:
  – If the application of the list processor allows us to restrict the placement of nodes in memory so that they are aligned on even multiples of 2 bytes.
    • NUMA addition can be eliminated.
    • Controller supplies “0” for low-bit of memory address for NEXT, and “1” for X.
  – Furthermore, if we could use a memory with a 16-bit wide output, then could fetch entire node in one cycle:

\[
\{\text{NEXT, X}\} \leftarrow \text{Memory}[\text{NEXT}], \quad \text{SUM} \leftarrow \text{SUM} + X;
\]

⇒ execution time cut in half (half as many cycles)

List Processor Conclusions

• Through careful optimization:
  – clock frequency increased from 32MHz to 71MHz
  – little cost increase.
• “Scheduling” was used to overlap and to maximize use of resources.
• Questions:
  – Consider the design process we went through:

  – Could a computer program go from RTL description to circuits automatically?
  – Could a computer program derive the optimizations that we did?
Modulo Scheduling

- Review of list processor scheduling:

<table>
<thead>
<tr>
<th>Memory</th>
<th>next₁</th>
<th>x₁</th>
<th>sum₁</th>
</tr>
</thead>
<tbody>
<tr>
<td>adder</td>
<td>numa₁</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- We “spread” out the schedule of one iteration to allow efficient packing?

<table>
<thead>
<tr>
<th>Memory</th>
<th>next₁</th>
<th>next₂</th>
<th>x₁</th>
<th>next₃</th>
<th>x₂</th>
<th>next₄</th>
<th>x₃</th>
</tr>
</thead>
<tbody>
<tr>
<td>adder</td>
<td>numa₁</td>
<td>numa₂</td>
<td>sum₁</td>
<td>numa₃</td>
<td>sum₂</td>
<td>numa₄</td>
<td>sum₃</td>
</tr>
</tbody>
</table>

- The goal of *modulo scheduling* is to find the schedule for one characteristic section of the computation. This is the part the control loops over.

- The entire schedule can then be derived, by repeating the characteristic section or repeating it with some pieces omitted.

---

Modulo Scheduling Procedure

1. **Calculate minimal length of characteristic section.**
   The maximum number of cycles that any one resource is used during one iteration of the computation (assuming a resource can only be used once per cycle).

2. Schedule one iteration of the computation on the characteristic section wrapping around when necessary. Each time the computation wraps around, decrease the iteration subscript by one.

3. If iteration will not fit on minimal length section, increase section by one and try again.
Modulo Scheduling List Processor

- Assuming a single adder and a single ported memory. Minimal schedule section length = 2.

Because both memory and adder are used for 2 cycles during one iteration.

- Finished schedule for 4 iterations:

<table>
<thead>
<tr>
<th>Memory</th>
<th>next₁</th>
<th>next₂</th>
<th>x₁</th>
<th>next₃</th>
<th>x₂</th>
<th>next₄</th>
<th>x₃</th>
</tr>
</thead>
<tbody>
<tr>
<td>adder</td>
<td>numa₁</td>
<td>numa₂</td>
<td>sum₁</td>
<td>numa₃</td>
<td>sum₂</td>
<td>numa₄</td>
<td>sum₃</td>
</tr>
</tbody>
</table>

Another Scheduling Example

- Assume A, B, C, D, E stored in a dual port memory.
- Assume a single adder.
- Minimal schedule section length = 3. (Both memory and adder are used for 3 cycles during one iteration.)

Compute Graph (one iteration of a repeating calculation)

Repeating schedule: