15-745

Instruction Scheduling
Software Pipelining

Copyright © Seth Copen Goldstein 2000-5
© Tim Callahan 2007

(some slides borrowed from M. Voss)

Instruction-level Parallelism

- Most modern processors have the ability to execute several adjacent instructions simultaneously.
  - Pipelined machines.
  - Very-long-instruction-word machines (VLIW).
  - Superscalar machines.
  - Dynamic scheduling/out-of-order machines.
- ILP is limited by several kinds of *execution constraints*:
  - Data dependence constraints.
  - Resource constraints ("hazards")
  - Control hazards

Execution Constraints

- Data-dependence constraints:
  - If instruction A computes a value that is read by instruction B, then B cannot execute before A is completed.
- Resource hazards:
  - Limited # of function[al units.]
    - If there are *n* functional u[nits] multipliers), then only *n* ins[tructions] of unit can execute at once.
  - Limited instruction issue.
    - If the instruction-issue unit can issue only *n* instructions at a time, then this limits ILP.
  - Limited register set.
    - Any schedule of instructions must have a valid register allocation.

For example:

```
ld   [%fp-28], %o1
add  %o1, %l2, %l3
```


Instruction Scheduling

- The purpose of instruction scheduling (IS) is to order the instructions for maximum ILP.
  - Keep all resources busy every cycle.
  - If necessary, eliminate data dependences and resource hazards to accomplish this.
- The IS problem is NP-complete (and bad in practice).
  - So heuristic methods are necessary.

Instruction Scheduling

- There are *many* different techniques for IS.
  - Still an open area of research.
- Most optimizing compilers perform good local IS, and only simple global IS.
- The biggest opportunities are in scheduling the code for loops.....

Should the Compiler Do IS?

- Many modern machines perform dynamic reordering of instructions.
  - Also called "out-of-order execution" (OOOE).
  - Not yet clear whether this is a good idea.
  - Pro:
    - OOOE can use additional registers and register renaming to eliminate data dependences that no amount of static IS can accomplish.
    - No need to recompile programs when hardware changes.
  - Con:
    - OOOE means more complex hardware (and thus longer cycle times and more wattage).
    - And can't be optimal since IS is NP-complete.

What we will cover

- Scheduling basic blocks
  - List scheduling
  - Long-latency operations
  - Delay slots
- Software Pipelining

- What we need to know
  - pipeline structure
  - data dependencies
  - register renaming
  - scalar replacement

Defining Dependencies

- Flow Dependence W → R δ^f } true
- Anti-Dependence R → W δ^a
- Output Dependence W → W δ^o
- Input Dependence R → R δ^i

(Anti- and Output Dependence: } false)

```
S1) a=0;
S2) b=a;
S3) c=a+d+e;
S4) d=b;
S5) b=5+e;
```

Not generally defined

Example Dependencies

```
S1) a=0;
S2) b=a;
S3) c=a+d+e;
S4) d=b;
S5) b=5+e;
```

S1 δ^f S2	due to a
S1 δ^f S3	due to a
S2 δ^f S4	due to b
S3 δ^a S4	due to d
S4 δ^a S5	due to b
S2 δ^o S5	due to b
S3 δ^i S5	due to a

Renaming of Variables

- Sometimes constraints are not "real," in the sense that a simple renaming of variables/registers can eliminate them.
  - Output dependence (WW): A and B write to the same variable.
  - Anti dependence (RW): A reads from a variable to which B writes.
- In such cases, the order of A and B cannot be changed unless variables are renamed.
  - Can sometimes be done by the hardware, to a limited extent.

Register Renaming Example

r1	← r2 + 1	r7	← r2 + 1	r7	← r2 + 1
[fp+8]	← r1	[fp+8]	← r7	r1	← r3 + 2
r1	← r3 + 2	r1	← r3 + 2	[fp+8]	← r7
[fp+12]	← r1	[fp+12]	← r1	[fp+12]	← r1

Phase ordering problem

- Can perform register renaming after register allocation
  - Constrained by available registers
  - Constrained by live on entry/exit
- Instead, do scheduling before register allocation

The Scheduler

- Given:
  - Code to schedule
  - Resources available (FU and # of Reg)
  - Latencies of instructions
- Goal:
  - Correct code
  - Better code [fewer cycles, less power, fewer registers, …]
  - Do it quickly

More Abstractly

- Given a graph G = (V,E) where
  - nodes are operations
    - Each operation has an associated delay and type
  - edges between nodes represent dependencies
  - The number of resources of type t, R(t)
- A schedule assigns to each node a cycle number:
  - $\sigma(n) \geq 0$
  - If $(n,m) \in G$, $\sigma(m) \geq \sigma(n) + \text{delay}(n)$
  - $|\{ n \mid \sigma(n) = x \text{ and type}(n) = t\}| <= R(t)$
- Goal is shortest length schedule, where length
  - $L(S)$ = max over n, $\sigma(n)+\text{delay}(n)$

List Scheduling

- Keep a list of **ready** instructions, I.e.,
  - If we are at cycle k, then all predecessors, p, in graph have all been scheduled so that $\sigma(p)+\text{delay}(p) \leq k$
  - Alternate?
- Pick some instruction, n, from ready queue such that there are **available resources** for type(n)
  - move n from "ready" to "scheduled"
- Update lists and continue
  - maybe move some from "not ready" to "ready"

Lots of Heuristics

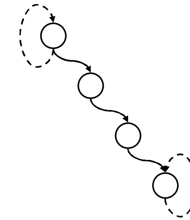
- forward or backward
- choose instructions on critical path
- ASAP or ALAP
- Balanced paths
- depth in schedule graph

Slack

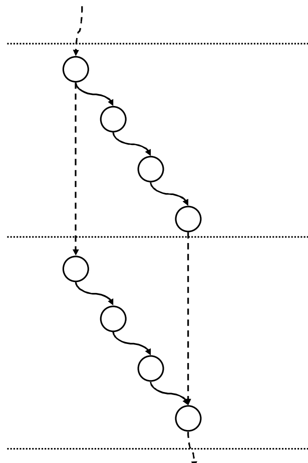
Software Pipelining

- Software pipelining is an IS technique that reorders the instructions in a loop.
  - Possibly moving instructions from one iteration to the previous or the next iteration.
  - Very large improvements in running time are possible.
- The first serious approach to software pipelining was presented by Aiken & Nicolau.
  - Aiken's 1988 Ph.D. thesis.
  - Impractical as it ignores resource hazards (focusing only on data-dependence constraints).
    - But sparked a large amount of follow-on research.

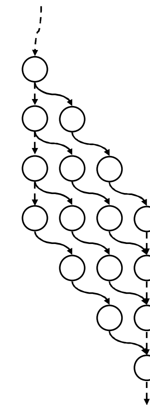
Software Pipelining



Software Pipelining



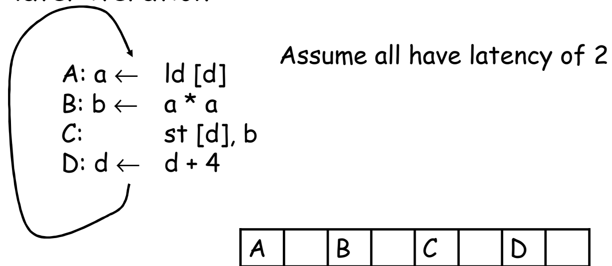
Software Pipelining



Goal of SP

- Increase distance between dependent operations by moving destination operation to a later iteration

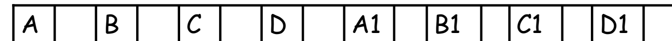
Assume all have latency of 2



Can we decrease the latency?

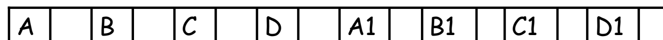
- Lets unroll

```
A:  a ← ld [d]
B:  b ← a * a
C:      st [d], b
D:  d ← d + 4
A1: a ← ld [d]
B1: b ← a * a
C1:     st [d], b
D1: d ← d + 4
```



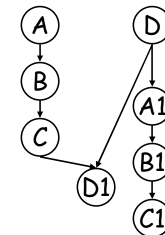
Rename variables

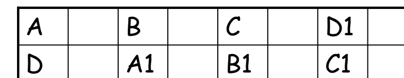
```
A:  a ←  ld [d]
B:  b ←  a * a
C:       st [d], b
D:  d1 ← d + 4
A1: a1 ← ld [d1]
B1: b1 ← a1 * a1
C1:      st [d1], b1
D1: d ←  d1 + 4
```



Schedule

```
A:  a ←  ld [d]
B:  b ←  a * a
C:       st [d], b
D:  d1 ← d + 4
A1: a1 ← ld [d1]
B1: b1 ← a1 * a1
C1:      st [d1], b1
D1: d ←  d1 + 4
```





A		B		C		D1	
D		A1		B1		C1	

Unroll Some More

```
A:  a ← ld [d]
B:  b ← a * a
C:      st [d], b
D:  d1 ← d + 4
A1: a1 ← ld [d1]
B1: b1 ← a1 * a1
C1:     st [d1], b1
D1: d2 ← d1 + 4
A2: a2 ← ld [d2]
B2: b2 ← a2 * a2
C2:     st [d2], b2
D2: d ← d2 + 4
```

A		B		C		D2	
D		A1		B1		C1	
	D1		A2		B2		C2

Unroll Some More

```
A:  a ←  ld [d]
B:  b ←  a * a
C:       st [d], b
D:  d1 ← d + 4
A1: a1 ← ld [d1]
B1: b1 ← a1 * a1
C1:      st [d1], b1
D1: d2 ← d1 + 4
A2: a2 ← ld [d2]
B2: b2 ← a2 * a2
C2:      st [d2], b2
D2: d ←  d2 + 4
```

A		B		C		D3		
D		A1		B1		C1		
	D1		A2		B2		C2	
		D2		A3		B3		C3

One More Time

A		B		C		D4			
D		A1		B1		C1			
	D1		A2		B2		C2		
		D2		A3		B3		C3	
			D3		A4		B4		C4

Can Rearrange

A		B		C		D4			
D		A1		B1		C1			
	D1 →		A2		B2		C2		
		D2 →		A3		B3		C3	
			D3		A4		B4		C4

Rearrange

```
A:   a ←    ld [d]
B:   b ←    a * a
C:          st [d], b
D:   d1 ←   d + 4
A1:  a1 ←   ld [d1]
B1:  b1 ←   a1 * a1
C1:         st [d1], b1
D1:  d2 ←   d1 + 4
A2:  a2 ←   ld [d2]
B2:  b2 ←   a2 * a2
C2:         st [d2], b2
D2:  d ←    d2 + 4
```

A		B		C		D3				
D		A1		B1		C1				
		D1		A2		B2		C2		
				D2		A3		B3		C3

Rearrange

```
A:   a ←    ld [d]
B:   b ←    a * a
C:          st [d], b
D:   d1 ←   d + 4
A1:  a1 ←   ld [d1]
B1:  b1 ←   a1 * a1
C1:         st [d1], b1
D1:  d2 ←   d1 + 4
A2:  a2 ←   ld [d2]
B2:  b2 ←   a2 * a2
C2:         st [d2], b2
D2:  d ←    d2 + 4
```

A		B		C		D3				
D		A1		B1		C1				
		D1		A2		B2		C2		
				D2		A3		B3		C3

SP Loop

Prolog

```
A:   a ←    ld [d]
B:   b ←    a * a
D:   d1 ←   d + 4
A1:  a1 ←   ld [d1]
D1:  d2 ←   d1 + 4
```

Body

```
C:          st [d], b
B1:  b1 ←   a1 * a1
A2:  a2 ←   ld [d2]
D2:  d ←    d2 + 4
```

Epilog

```
B2:  b2 ←   a2 * a2
C1:         st [d1], b1
D3:  d2 ←   d1 + 4
C2:         st [d2], b2
```

A		B		C	C	C	D3		
D		A1		B1	B1	B1	C1		
		D1		A2	A2	A2	B2		C2
				D2	D2	D2			

Goal of SP

- Increase distance between dependent operations by moving destination operation to a later iteration

A, B, C

after SP

iteration i i+1 i+2

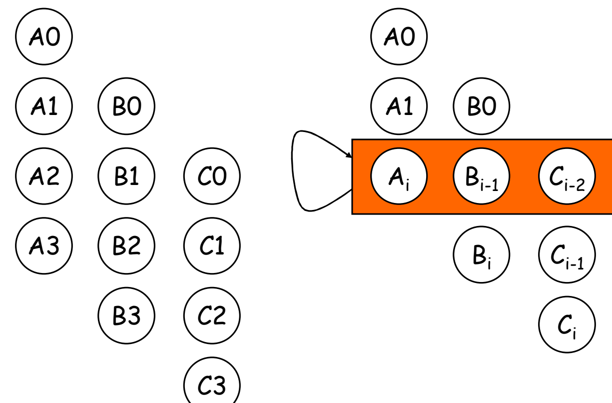
dependencies in initial loop

Goal of SP

- Increase distance between dependent operations by moving destination operation to a later iteration
- But also, to uncover ILP across iteration boundaries!

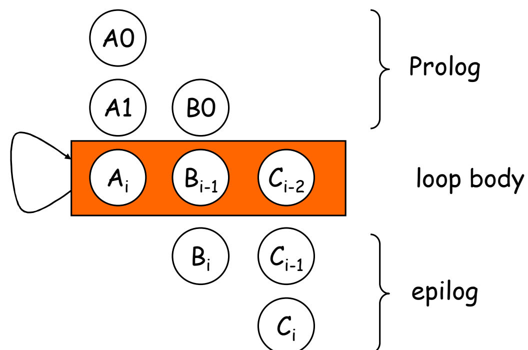
Example

Assume operating on a infinite wide machine



Example

Assume operating on a infinite wide machine



Dealing with exit conditions

```
for (i=0; i<N; i++)
{
    Ai
    Bi
    Ci
}
```

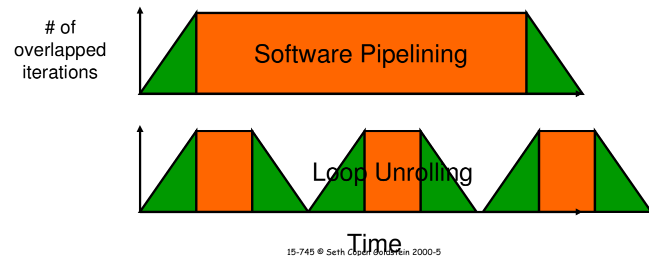
```
i=0
if (i >= N) goto done
A0
B0
if (i+1 == N) goto last
i=1
A1
if (i+2 == N) goto epilog
i=2
```

```
loop:
    Ai
    Bi-1
    Ci-2
    i++
    if (i < N) goto loop
epilog:
    Bi
    Ci-1
last:
    ci
done:
```

Loop Unrolling V. SP

For SuperScalar or VLIW

- Loop Unrolling reduces loop overhead
- Software Pipelining reduces fill/drain
- Best is if you combine them



Aiken/Nicolau Scheduling
Step 1

Perform *scalar replacement* to eliminate memory references where possible.

```
for i:=1 to N do
    a  := j ⊕ V[i-1]
    b  := a ⊕ f
    c  := e ⊕ j
    d  := f ⊕ c
    e  := b ⊕ d
    f  := U[i]
 g: V[i]  := b
 h: W[i]  := d
    j  := X[i]
```

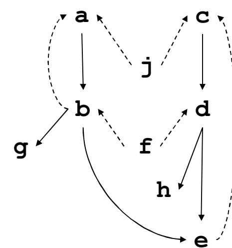
```
for i:=1 to N do
    a  := j ⊕ b
    b  := a ⊕ f
    c  := e ⊕ j
    d  := f ⊕ c
    e  := b ⊕ d
    f  := U[i]
 g: V[i]  := b
 h: W[i]  := d
    j  := X[i]
```


Aiken/Nicolau Scheduling
Step 2

Unroll the loop and compute the data-dependence graph (DDG).

DDG for rolled loop:

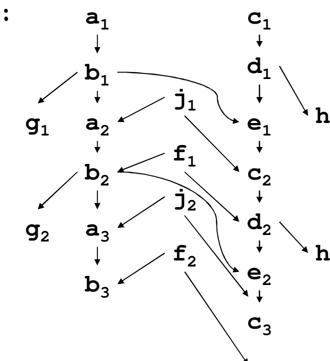
```
for i:=1 to N do
    a  := j ⊕ b
    b  := a ⊕ f
    c  := e ⊕ j
    d  := f ⊕ c
    e  := b ⊕ d
    f  := U[i]
 g: V[i]  := b
 h: W[i]  := d
    j  := X[i]
```



Aiken/Nicolau Scheduling
Step 2, cont'd

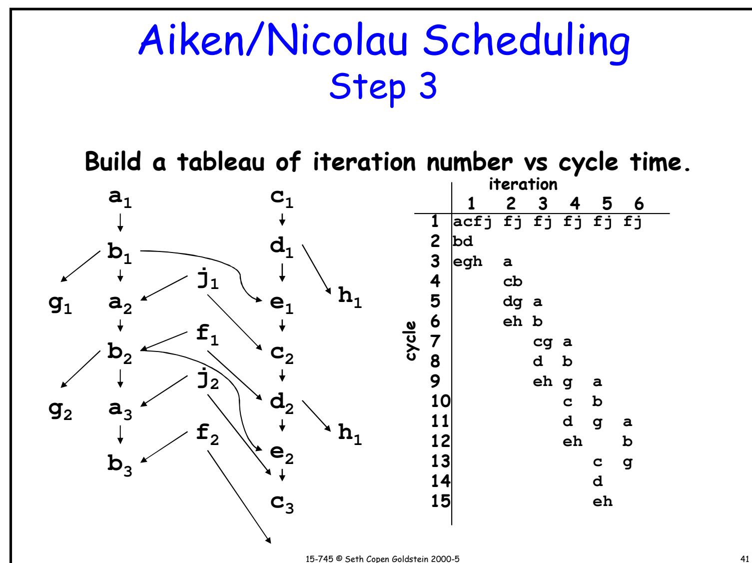
DDG for unrolled loop:

```
for i:=1 to N do
    a  := j ⊕ b
    b  := a ⊕ f
    c  := e ⊕ j
    d  := f ⊕ c
    e  := b ⊕ d
    f  := U[i]
 g: V[i]  := b
 h: W[i]  := d
    j  := X[i]
```



Aiken/Nicolau Scheduling
Step 3

Build a tableau of iteration number vs cycle time.



cycle	1	2	3	4	5	6
1	acfj	fj	fj	fj	fj	fj
2	bd					
3	egh	a				
4		cb				
5		dg	a			
6		eh	b			
7			cg	a		
8			d	b		
9			eh	g	a	
10				c	b	
11				d	g	a
12				eh		b
13					c	g
14					d	
15					eh	

Aiken/Nicolau Scheduling
Step 3

basically, you're emulating a superscalar with infinite resources, infinite register renaming, always predicting the loop-back branch: thus, just pure data dependency

on number vs cycle time.

cycle	1	2	3	4	5	6
1	acfj	fj	fj	fj	fj	fj
2	bd					
3	egh	a				
4		cb				
5		dg	a			
6		eh	b			
7			cg	a		
8			d	b		
9			eh	g	a	
10				c	b	
11				d	g	a
12				eh		b
13					c	g
14					d	
15					eh	

Aiken/Nicolau Scheduling
Step 4

Find repeating patterns of instructions.

cycle	1	2	3	4	5	6
1	acfj	fj	fj	fj	fj	fj
2	bd					
3	egh	a				
4		cb				
5		dg	a			
6		eh	b			
7			cg	a		
8			d	b		
9			eh	g	a	
10				c	b	
11				d	g	a
12				eh		b
13					c	g
14					d	
15					eh	

Aiken/Nicolau Scheduling
Step 4

Find repeating patterns of instructions.

cycle	1	2	3	4	5	6
1	acfj	fj	fj	fj	fj	fj
2	bd					
3	egh	a				
4		cb				
5		dg	a			
6		eh	b			
7			cg	a		
8			d	b		
9			eh	g	a	
10				c	b	
11				d	g	a
12				eh		b
13					c	g
14					d	
15					eh	

Aiken/Nicolau Scheduling Step 4

Find repeating patterns of instructions.

cycle	iteration 1	2	3	4	5	6
1	acfj	fj	fj	fj	fj	fj
2	bd					
3	egh	a				
4		cb				
5		dg	a			
6		eh	b			
7			cg	a		
8			d	b		
9			eh	g	a	
10				c	b	
11				d	g	a
12				eh		b
13					c	g
14					d	
15					eh	

Go back and relate slopes to DDG

Aiken/Nicolau Scheduling Step 5

"Coalesce" the slopes.

cycle	iteration 1	2	3	4	5	6
1	acfj	fj	fj	fj	fj	fj
2	bd					
3	egh	a				
4		cb				
5		dg	a			
6		eh	b			
7			cg	a		
8			d	b		
9			eh	g	a	
10				c	b	
11				d	g	a
12				eh		b
13					c	g
14					d	
15					eh	

cycle	iteration 1	2	3	4	5	6
1	acfj					
2	bd	fj				
3	egh	a				
4		cb	fj			
5		dg	a			
6		eh	b	fj		
7			cg	a		
8			d	b		
9			eh	g	fj	
10				c	a	
11				d	b	
12				eh	g	
13					c	
14					d	
15					eh	

Aiken/Nicolau Scheduling Step 6

Find the loop body and "reroll" the loop.

cycle	iteration 1	2	3	4	5	6
1	acfj					
2	bd	fj				
3	egh	a				
4		cb	fj			
5		dg	a			
6		eh	b	fj		
7			cg	a		
8			d	b		
9			eh	g	fj	
10				c	a	
11				d	b	
12				eh	g	
13					c	
14					d	
15					eh	

Aiken/Nicolau Scheduling Step 6

Find the loop body and "reroll" the loop.

cycle	iteration 1	2	3	4	5	6	
1	acfj						
2	bd	fj					
3	egh	a					← Prologue/entry code
4		cb	fj				
5		dg	a				
6		eh	b	fj			
7			cg	a			
8			d	b			
9			eh	g	fj		← Loop body
10				c	a		
11				d	b		
12				eh	g		
13					c		← Epilogue/exit code
14					d		
15					eh		

Aiken/Nicolau Scheduling Step 7

Generate code.

(Assume VLIW-like machine for this example. The instructions on each line should be issued in parallel.)

```
   a1  := j0 ⊕ b0    c1  := e0 ⊕ j0    f1  := U[1]      j1  := X[1]
   b1  := a1 ⊕ f0    d1  := f0 ⊕ c1    f2  := U[2]      j2  := X[2]
   e1  := b1 ⊕ d1    V[1]  := b1       W[1]  := d1      a2  := j1 ⊕ b1
   c2  := e1 ⊕ j1    b2  := a2 ⊕ f1    f3  := U[3]      j3  := X[3]
   d2  := f1 ⊕ c2    V[2]  := b2       a3  := j2 ⊕ b2
   e2  := b2 ⊕ d2    W[2]  := d2       b3  := a3 ⊕ f2   f4  := U[4]    j4 := X[4]
   c3  := e2 ⊕ j2    V[3]  := b3       a4  := j3 ⊕ b3    i  := 3
L:
   d_i  := f_{i-1} ⊕ c_i    b_{i+1}  := a_i ⊕ f_i
   e_i  := b_i ⊕ d_i        W[i]  := d_i          V[i+1]  := b_{i+1}  f_{i+2}  := U[I+2]  j_{i+2}  := X[i+2]
   c_{i+1}  := e_i ⊕ j_i    a_{i+2}  := j_{i+1} ⊕ b_{i+1}  i := i+1  if i<N-2 goto L

   d_{N-1}  := f_{N-2} ⊕ c_{N-1}  b_N  := a_N ⊕ f_{N-1}
   e_{N-1}  := b_{N-1} ⊕ d_{N-1}  W[N-1]  := d_{N-1}   v[N]  := b_N
   c_N  := e_{N-1} ⊕ j_{N-1}
   d_N  := f_{N-1} + c_N
   e_N  := b_N ⊕ d_N       w[N]  := d_N
```


Aiken/Nicolau Scheduling Step 8

- Since several versions of a variable (e.g., j_i and j_{i+1}) might be live simultaneously, we need to add new temps and moves

```
   a1  := j0 ⊕ b0    c1  := e0 ⊕ j0    f1  := U[1]      j1  := X[1]
   b1  := a1 ⊕ f0    d1  := f0 ⊕ c1    f2  := U[2]      j2  := X[2]
   e1  := b1 ⊕ d1    V[1]  := b1       W[1]  := d1      a2  := j1 ⊕ b1
   c2  := e1 ⊕ j1    b2  := a2 ⊕ f1    f3  := U[3]      j3  := X[3]
   d2  := f1 ⊕ c2    V[2]  := b2       a3  := j2 ⊕ b2
   e2  := b2 ⊕ d2    W[2]  := d2       b3  := a3 ⊕ f2   f4  := U[4]    j4 := X[4]
   c3  := e2 ⊕ j2    V[3]  := b3       a4  := j3 ⊕ b3    i  := 3
L:
   d_i  := f_{i-1} ⊕ c_i    b_{i+1}  := a_i ⊕ f_i
   e_i  := b_i ⊕ d_i        W[i]  := d_i          V[i+1]  := b_{i+1}  f_{i+2}  := U[I+2]  j_{i+2}  := X[i+2]
   c_{i+1}  := e_i ⊕ j_i    a_{i+2}  := j_{i+1} ⊕ b_{i+1}  i := i+1  if i<N-2 goto L

   d_{N-1}  := f_{N-2} ⊕ c_{N-1}  b_N  := a_N ⊕ f_{N-1}
   e_{N-1}  := b_{N-1} ⊕ d_{N-1}  W[N-1]  := d_{N-1}   v[N]  := b_N
   c_N  := e_{N-1} ⊕ j_{N-1}
   d_N  := f_{N-1} + c_N
   e_N  := b_N ⊕ d_N       w[N]  := d_N
```


Aiken/Nicolau Scheduling Step 8

- Since several versions of a variable (e.g., j_i and j_{i+1}) might be live simultaneously, we need to add new temps and moves

```
   a1  := j0 ⊕ b0    c1  := e0 ⊕ j0    f1  := U[1]      j1  := X[1]
   b1  := a1 ⊕ f0    d1  := f0 ⊕ c1    f''  := U[2]      j2  := X[2]
   e1  := b1 ⊕ d1    V[1]  := b1       W[1]  := d1      a2  := j1 ⊕ b1
   c2  := e1 ⊕ j1    b2  := a2 ⊕ f1    f'  := U[3]      j'  := X[3]
   d2  := f1 ⊕ c2    V[2]  := b2       a3  := j2 ⊕ b2
   e2  := b2 ⊕ d2    W[2]  := d2       b3  := a3 ⊕ f''   f4  := U[4]    j4 := X[4]
   c3  := e2 ⊕ j2    V[3]  := b3       a4  := j' ⊕ b3    i  := 3
L:
   d_i  := f'' ⊕ c_i    b_{i+1}  := a' ⊕ f'   b' := b; a'=a; f''=f'; f'=f; j''=j'; j'=j
   e_i  := b' ⊕ d_i     W[i]  := d_i          V[i+1]  := b_{i+1}  f_{i+2}  := U[I+2]  j_{i+2}  := X[i+2]
   c_{i+1}  := e_i ⊕ j'    a_{i+2}  := j'' ⊕ b_{i+1}  i := i+1  if i<N-2 goto L

   d_{N-1}  := f_{N-2} ⊕ c_{N-1}  b_N  := a_N ⊕ f_{N-1}
   e_{N-1}  := b_{N-1} ⊕ d_{N-1}  W[N-1]  := d_{N-1}   v[N]  := b_N
   c_N  := e_{N-1} ⊕ j_{N-1}
   d_N  := f_{N-1} + c_N
   e_N  := b_N ⊕ d_N       w[N]  := d_N
```


Next Step in SP

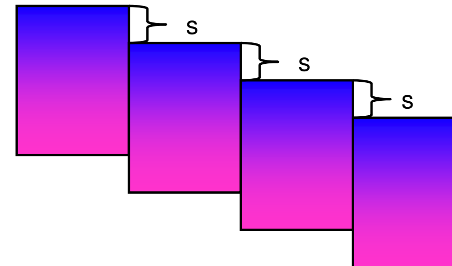
- AN88 did not deal with resource constraints.
- Modulo Scheduling is a SP algorithm that does.
- It schedules the loop based on
  - resource constraints
  - precedence constraints
- Basically, it's list scheduling that takes into account resource conflicts from overlapping iterations

Resource Constraints

- Minimally indivisible sequences, *i* and *j*, can execute together if combined resources in a step do not exceed available resources.
- R(i) is a resource configuration vector
  R(i) is the number of units of resource i
- r(i) is a resource usage vector s.t.
  $0 \leq r(i) \leq R(i)$
- Each node in G has an associated r(i)

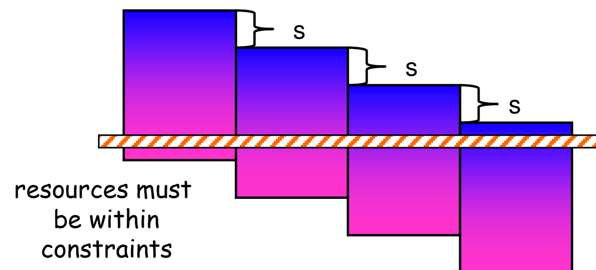
Software Pipelining Goal

- Find the same schedule for each iteration.
- Stagger by iteration initiation interval, *s*
- Goal: minimize *s*.



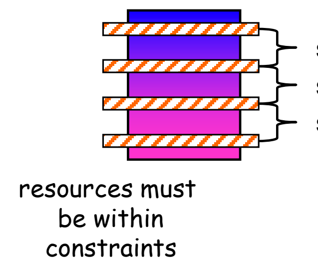
Software Pipelining Goal

- Find the same schedule for each iteration.
- Stagger by iteration <u>initiation interval</u>, *s*
- Goal: minimize *s*.

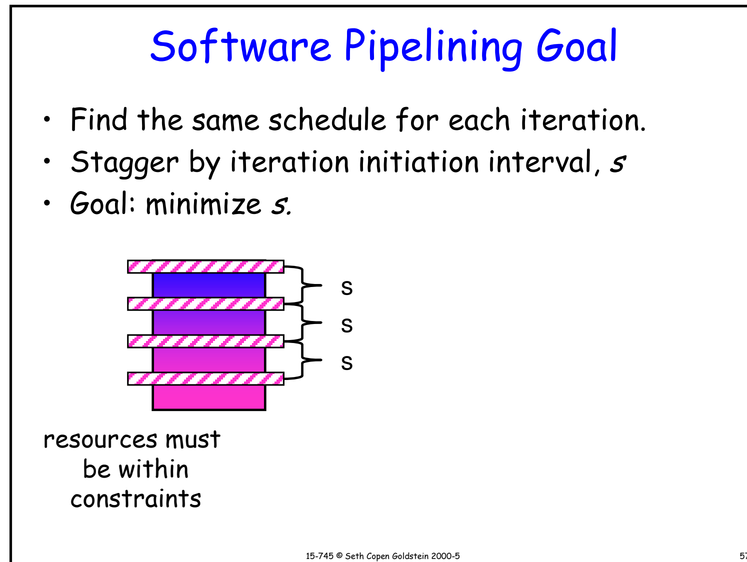


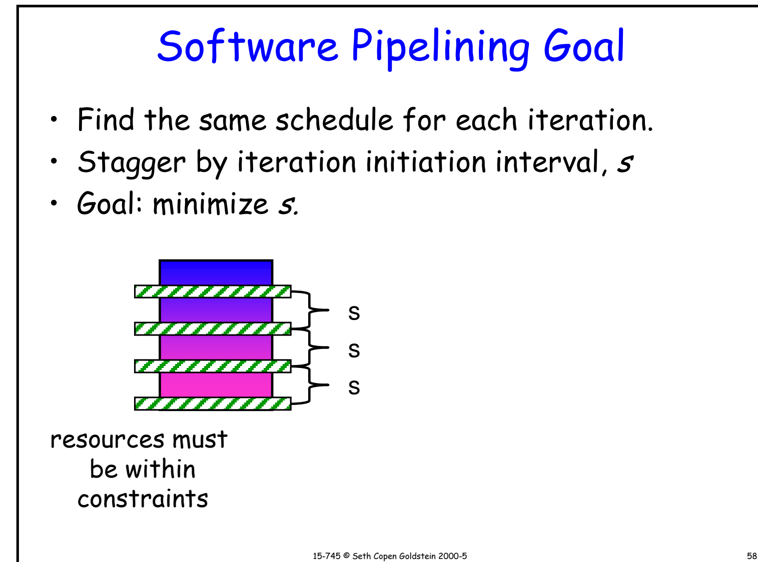
Software Pipelining Goal

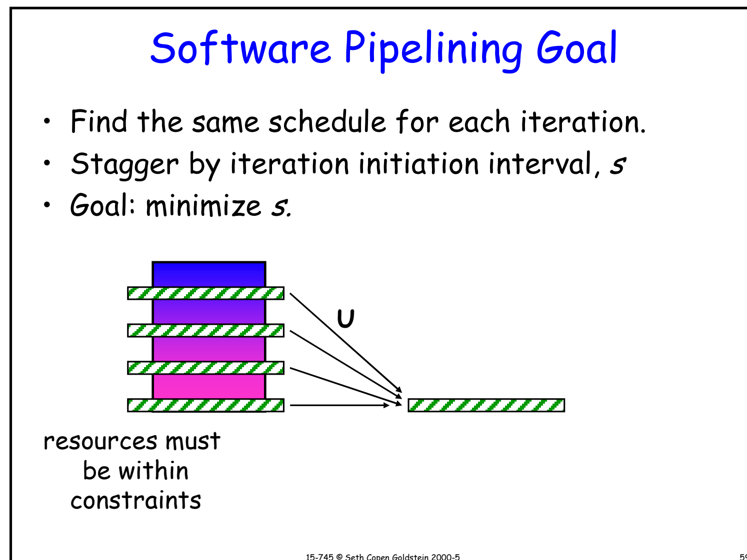
- Find the same schedule for each iteration.
- Stagger by iteration initiation interval, *s*
- Goal: minimize *s*.

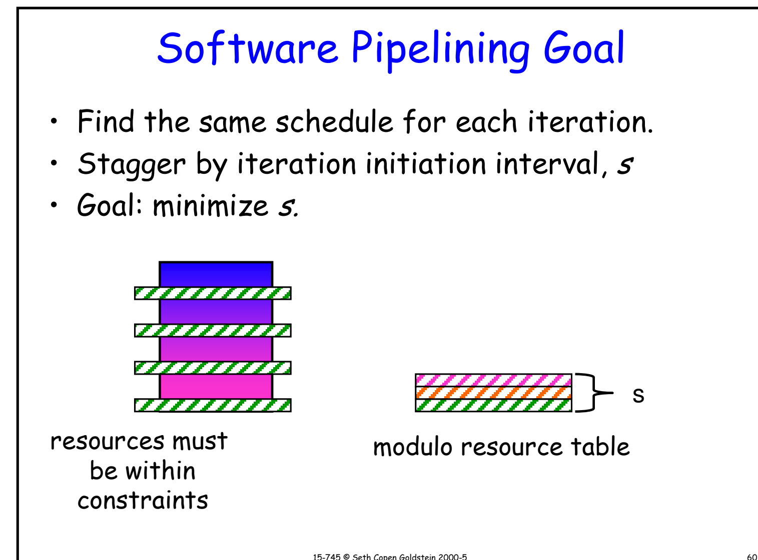


resources must be within constraints


Software Pipelining Goal
• Find the same schedule for each iteration.
• Stagger by iteration initiation interval, s
• Goal: minimize s.
s
s
s
resources must be within constraints
15-745 © Seth Copen Goldstein 2000-5
57


Software Pipelining Goal
• Find the same schedule for each iteration.
• Stagger by iteration initiation interval, s
• Goal: minimize s.
s
s
s
resources must be within constraints
15-745 © Seth Copen Goldstein 2000-5
58


Software Pipelining Goal
• Find the same schedule for each iteration.
• Stagger by iteration initiation interval, s
• Goal: minimize s.
U
resources must be within constraints
15-745 © Seth Copen Goldstein 2000-5
59


Software Pipelining Goal
• Find the same schedule for each iteration.
• Stagger by iteration initiation interval, s
• Goal: minimize s.
s
resources must be within constraints
modulo resource table
15-745 © Seth Copen Goldstein 2000-5
60

Precedence Constraints

- Review: for acyclic scheduling, constraint is just the required delay between two ops u, v: <d(u,v)>
- For an edge, $u \rightarrow v$, we must have
  $\sigma(v)-\sigma(u) \geq d(u,v)$

Precedence Constraints

- Cyclic: constraint becomes a tuple: <p,d>
  - p is the minimum iteration delay (or the loop carried dependence distance)
  - d is the delay
- For an edge, $u \rightarrow v$, we must have
  $\sigma(v)-\sigma(u) \geq d(u,v)-s*p(u,v)$
- $p \geq 0$
- If data dependence is
  - within an iteration, p=0
  - loop-carried across p iter boundaries, p>0

Iterative Approach

- Finding minimum S that satisfies the constraints is NP-Complete.
- Heuristic:
  - Find lower and upper bounds for S
  - foreach s from lower to upper bound?
    - Schedule graph.
    - If succeed, done
    - Otherwise try again (with next higher s)
- Thus: "Iterative Modulo Scheduling" Rau, et.al.

Iterative Approach

- Heuristic:
  - Find lower and upper bounds for S
  - foreach s from lower to upper bound
    - Schedule graph.
    - If succeed, done
    - Otherwise try again (with next higher s)
- So the key difference:
  - AN88 does not assume S when scheduling
  - IMS must assume an S for each scheduling attempt to understand resource conflicts

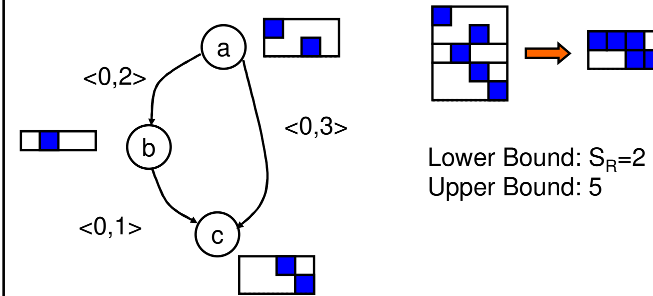
Lower Bounds

- Resource Constraints: S_R (also called II_{res}) maximum over all resources of # of uses divided by # available… rounded up or down?
- Precedence Constraints: S_E (also called II_{rec}) max over all cycles: d(c)/p(c)

In practice, one is easy, other is hard.

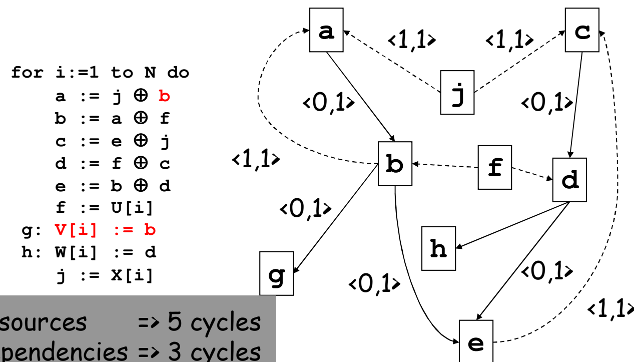
Tim's secret approach: just use S_R as lower bound, then do binary search for best S

Acyclic Example



Lower Bound on s

- Assume 1 ALU and 1 MU
- Assume latency Op or load is 1 cycle



Resources => 5 cycles
Dependencies => 3 cycles

Scheduling data structures

To schedule for initiation interval s:

- Create a resource table with s rows and R columns
- Create a vector, σ, of length N for n instructions in the loop
  - σ[n] = the time at which n is scheduled, or NONE
- Prioritize instructions by some heuristic
  - critical path (or cycle)
  - resource critical

Scheduling algorithm

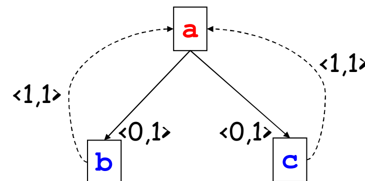
- Pick an instruction, n
- Calculate earliest time due to dependence constraints
  For all x=pred(n),
      earliest = max(earliest, σ(x)+d(x,n)-s·p(x,n))
- try and schedule n from earliest to (earliest+s-1) s.t. resource constraints are obeyed.
  - possible twist: deschedule a conflicting node to make way for n, maybe randomly, like sim anneal
- If we fail, then this schedule is faulty (i.e. give up on this s)

Scheduling algorithm – cont.

- We now schedule n at earliest, I.e., σ(n) = earliest
- Fix up schedule
  - Successors, x, of n must be scheduled s.t. σ(x) >= σ(n)+d(n,x)-s·p(n,x), otherwise they are removed (descheduled) and put back on worklist.
- repeat this some number of times until either
  - succeed, then register allocate
  - fail, then increase s

Simplest Example

```
for () {
  a = b+c
  b = a*a
  c = a*194
}
```

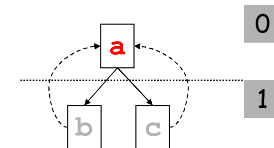


Resources: 1 1

What is IIres?
What is IIrec?

Simplest Example

```
for () {
  a = b+c
  b = a*a
  c = a*194
}
```



0
1

Try II = 2

Modulo Resource Table:

0	1	
1		

Simplest Example

```
for () {
  a = b+c
  b = a*a
  c = a*194
}
```

Try II = 2

Modulo Resource Table:

0	1	
1		1

Simplest Example

```
for () {
  a = b+c
  b = a*a
  c = a*194
}
```

Try II = 2

Modulo Resource Table:

0	1	1
1		1

Simplest Example

```
for () {
  a = b+c
  b = a*a
  c = a*194
}
```

Try II = 2

Modulo Resource Table:

0		1
1		1

earliest a: sigma(c) + delay(c) - 2
= 2+1-2 = 1

Simplest Example

```
for () {
  a = b+c
  b = a*a
  c = a*194
}
```

Try II = 2

Modulo Resource Table:

0		1
1	1	

earliest b?
scheduled b?
what next?

Simplest Example

```
for () {
  a = b+c
  b = a*a
  c = a*194
}
```

Try II = 2

Modulo Resource Table:

0		1
1		1

Lesson: lower bound may not be achievable

Example

```
for i:=1 to N do
    a := j ⊕ b
    b := a ⊕ f
    c := e ⊕ j
    d := f ⊕ c
    e := b ⊕ d
    f := U[i]
 g: V[i] := b
 h: W[i] := d
    j := X[i]
```

Priorities: ?

Example

```
for i:=1 to N do
    a := j ⊕ b
    b := a ⊕ f
    c := e ⊕ j
    d := f ⊕ c
    e := b ⊕ d
    f := U[i]
 g: V[i] := b
 h: W[i] := d
    j := X[i]
```

Priorities: c,d,e,a,b,f,j,g,h


```
for i:=1 to N do
    a := j ⊕ b
    b := a ⊕ f
    c := e ⊕ j
    d := f ⊕ c
    e := b ⊕ d
    f := U[i]
 g: V[i] := b
 h: W[i] := d
    j := X[i]
```

s=5

Priorities: c,d,e,a,b,f,j,g,h

ALU	MU

instr	σ
a	
b	
c	
d	
e	
f	
g	
h	
j	

```
for i:=1 to N do
    a := j ⊕ b
    b := a ⊕ f
    c := e ⊕ j
    d := f ⊕ c
    e := b ⊕ d
    f := U[i]
 g: V[i] := b
 h: W[i] := d
    j := X[i]
```

s=5

Priorities: a,b,f,j,g,h

ALU	MU
c	
d	
e	

instr	σ
a	
b	
c	0
d	1
e	2
f	
g	
h	
j	

15-745 © Seth Copen Goldstein 2000-5

```
for i:=1 to N do
    a := j ⊕ b
    b := a ⊕ f
    c := e ⊕ j
    d := f ⊕ c
    e := b ⊕ d
    f := U[i]
 g: V[i] := b
 h: W[i] := d
    j := X[i]
```

s=5

Priorities: b,f,j,g,h

ALU	MU
c	
d	
e	
a	

instr	σ
a	3
b	
c	0
d	1
e	2
f	
g	
h	
j	

15-745 © Seth Copen Goldstein 2000-5

```
for i:=1 to N do
    a := j ⊕ b
    b := a ⊕ f
    c := e ⊕ j
    d := f ⊕ c
    e := b ⊕ d
    f := U[i]
 g: V[i] := b
 h: W[i] := d
    j := X[i]
```

s=5

Priorities: b,f,j,g,h

ALU	MU
c	
d	
e	
a	
b	

instr	σ
a	3
b	4
c	0
d	1
e	2
f	
g	
h	
j	

15-745 © Seth Copen Goldstein 2000-5

```
for i:=1 to N do
    a := j ⊕ b
    b := a ⊕ f
    c := e ⊕ j
    d := f ⊕ c
    e := b ⊕ d
    f := U[i]
 g: V[i] := b
 h: W[i] := d
    j := X[i]
```

s=5

Priorities: e,f,j,g,h

ALU	MU
c	
d	
a	
b	

instr	σ
a	3
b	4
c	0
d	1
e	
f	
g	
h	
j	

b causes b->e edge violation

15-745 © Seth Copen Goldstein 2000-5

```
for i:=1 to N do
    a := j ⊕ b
    b := a ⊕ f
    c := e ⊕ j
    d := f ⊕ c
    e := b ⊕ d
    f := U[i]
 g: V[i] := b
 h: W[i] := d
    j := X[i]
```

s=5

Priorities: e,f,j,g,h

ALU	MU
c	
d	
e	
a	
b	

instr	σ
a	3
b	4
c	0
d	1
e	7
f	
g	
h	
j	

e causes e->c edge violation


```
for i:=1 to N do
    a := j ⊕ b
    b := a ⊕ f
    c := e ⊕ j
    d := f ⊕ c
    e := b ⊕ d
    f := U[i]
 g: V[i] := b
 h: W[i] := d
    j := X[i]
```

s=5

Priorities: f,j,g,h

ALU	MU
c	f
d	
e	
a	
b	

instr	σ
a	3
b	4
c	5
d	6
e	7
f	0
g	
h	
j	


```
for i:=1 to N do
    a := j ⊕ b
    b := a ⊕ f
    c := e ⊕ j
    d := f ⊕ c
    e := b ⊕ d
    f := U[i]
 g: V[i] := b
 h: W[i] := d
    j := X[i]
```

s=5

Priorities:j,g,h

ALU	MU
c	f
d	j
e	
a	
b	

instr	σ
a	3
b	4
c	5
d	6
e	7
f	0
g	
h	
j	1


```
for i:=1 to N do
    a := j ⊕ b
    b := a ⊕ f
    c := e ⊕ j
    d := f ⊕ c
    e := b ⊕ d
    f := U[i]
 g: V[i] := b
 h: W[i] := d
    j := X[i]
```

s=5

Priorities:g,h

ALU	MU
c	f
d	j
e	g
a	h
b	

instr	σ
a	3
b	4
c	5
d	6
e	7
f	0
g	7
h	8
j	1

Creating the Loop

- Create the body from the schedule.
- Determine which iteration an instruction falls into
  - Mark its sources and dest as belonging to that iteration.
  - Add Moves to update registers
- Prolog fills in gaps at beginning
  - For each move we will have an instruction in prolog, and we fill in dependent instructions
- Epilog fills in gaps at end

instr	σ
a	3
b	4
c	5
d	6
e	7
f	0
g	7
h	8
j	1


```
f0 = U[0];
j0 = X[0];

FOR i = 0 to N
   f1 := U[i+1]
   j1 := X[i+1]
   nop
   a := j0 ? b
   b := a ? f0
   c := e ? j0
   d := f0 ? c
   e := b ? d                  g: V[i] := b
 h: W[i] := d
   f0 = f1
   j0 = j1
```


Conditionals

- What about internal control structure, I.e., conditionals
- Three approaches
  - Schedule both sides and use conditional moves
  - Schedule each side, then make the body of the conditional a macro op with appropriate resource vector
  - Trace schedule the loop

What to take away

- Dependence analysis is very important
- Software pipelining is cool
- Registers are a key resource