Introduction to Pipelining

Lecture Goals

 Conclude the discussion on multi-cycle computations from last lecture

Lecture Goals

- Conclude the discussion on multi-cycle computations from last lecture
- Introduce pipelining, a technique that uses registers to improve the throughput of a circuit

Lecture Goals

- Conclude the discussion on multi-cycle computations from last lecture
- Introduce pipelining, a technique that uses registers to improve the throughput of a circuit
- Examine the tradeoffs of different design styles (combinational, pipelined, multi-cycle) through a multiplier case study

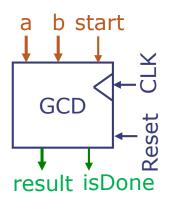
Reminder: Multi-Cycle Computations

- Sequential circuits can implement more computations than combinational circuits
 - Variable amount of input/output
 - Variable number of steps

Reminder: Multi-Cycle Computations

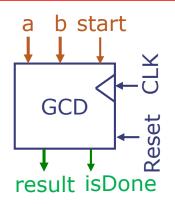
- Sequential circuits can implement more computations than combinational circuits
 - Variable amount of input/output
 - Variable number of steps
- Multi-cycle circuits implement a computation over multiple cycles
 - New computation is started by setting inputs at a particular cycle
 - Circuit takes several cycles (possibly a variable number) to finish computation, then makes output available
 - Circuit performs only one computation at a time; a new computation cannot begin until previous one has finished

```
def gcd(a, b):
    x = a
    y = b
    while x != 0:
        if x >= y:
            x = x - y
        else:
            (x, y) = (y, x)
    return y
```

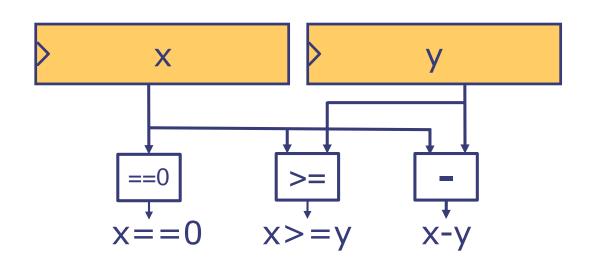


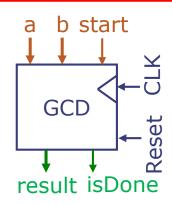
```
def gcd(a, b):
    x = a
    y = b
    while x != 0:
        if x >= y:
            x = x - y
        else:
            (x, y) = (y, x)
    return y
```



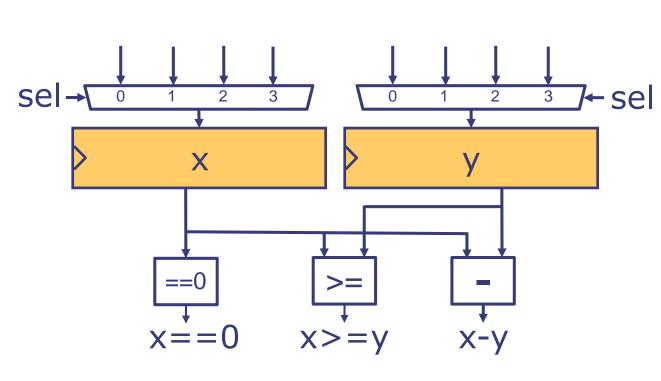


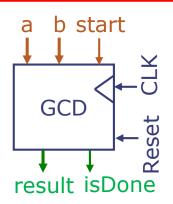
```
def gcd(a, b):
    x = a
    y = b
    while x != 0:
        if x >= y:
            x = x - y
        else:
            (x, y) = (y, x)
    return y
```



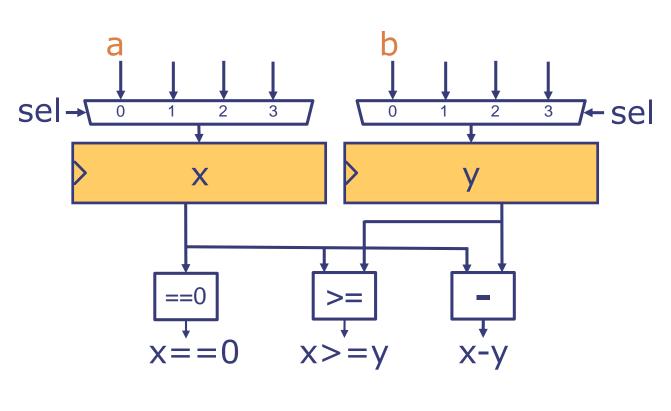


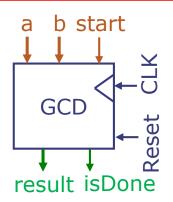
```
def gcd(a, b):
    x = a
    y = b
    while x != 0:
        if x >= y:
            x = x - y
        else:
            (x, y) = (y, x)
    return y
```



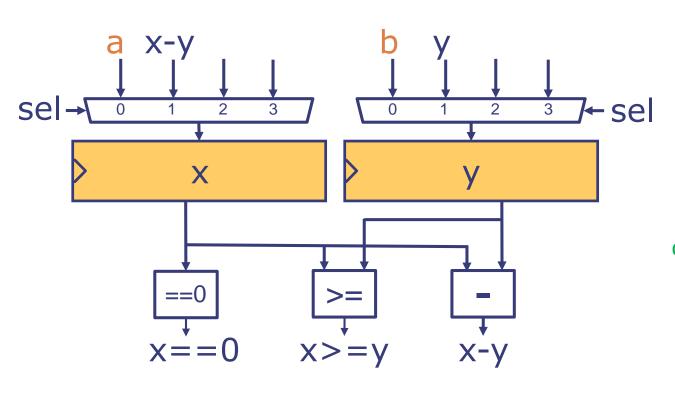


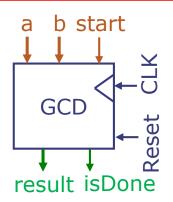
```
def gcd(a, b):
    x = a
    y = b
    while x != 0:
        if x >= y:
            x = x - y
        else:
            (x, y) = (y, x)
    return y
```



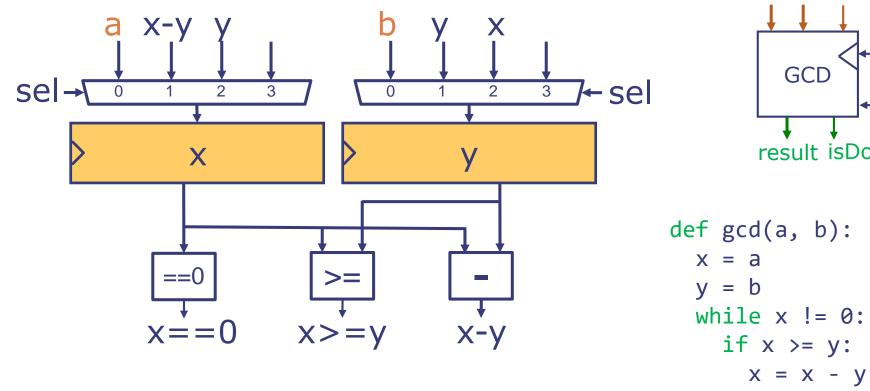


```
def gcd(a, b):
    x = a
    y = b
    while x != 0:
        if x >= y:
            x = x - y
        else:
            (x, y) = (y, x)
    return y
```





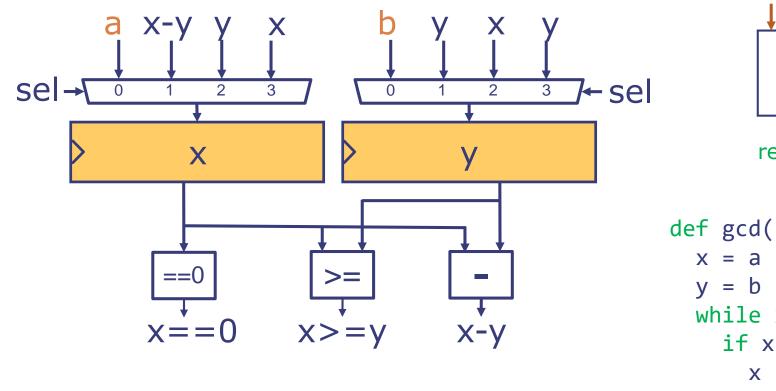
```
def gcd(a, b):
    x = a
    y = b
    while x != 0:
        if x >= y:
            x = x - y
        else:
            (x, y) = (y, x)
    return y
```

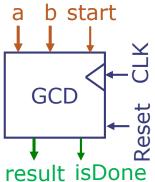


(x, y) = (y, x)

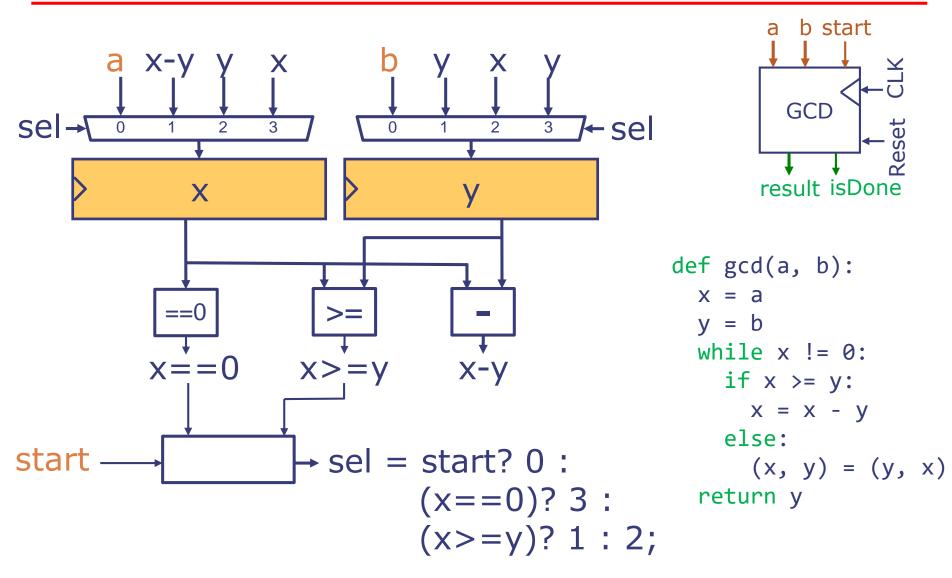
else:

return y

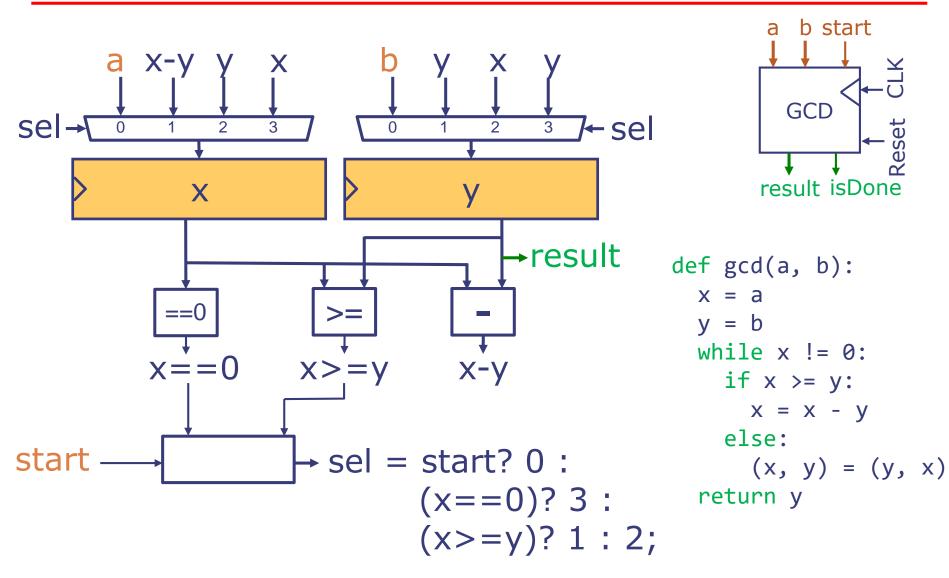




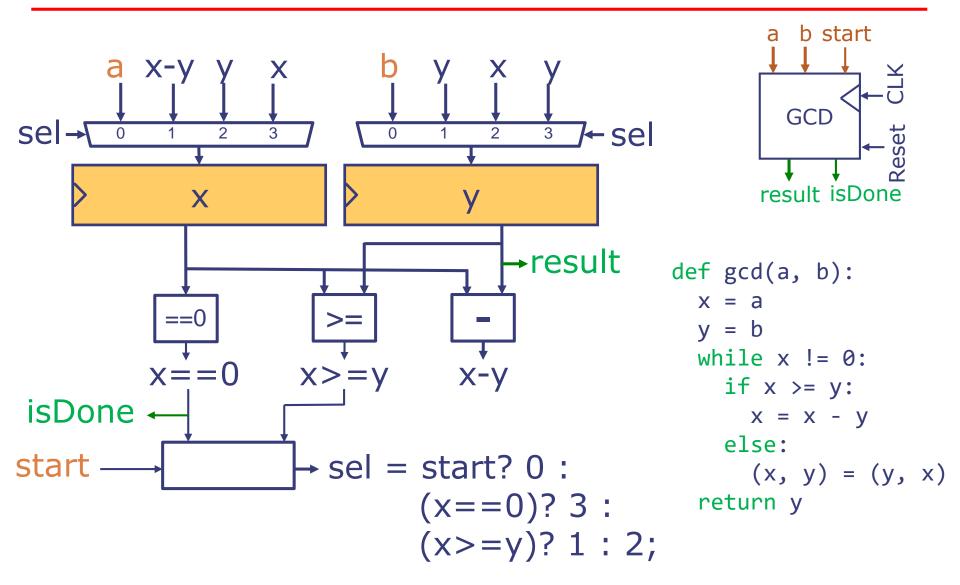
```
def gcd(a, b):
    x = a
    y = b
    while x != 0:
        if x >= y:
            x = x - y
        else:
            (x, y) = (y, x)
    return y
```



March 17, 2022 MIT 6.004 Spring 2022 L11-4



March 17, 2022 MIT 6.004 Spring 2022 L11-4



March 17, 2022 MIT 6.004 Spring 2022 L11-4

```
typedef Bit#(32) Word;
module GCD;
```

```
typedef Bit#(32) Word;
module GCD;
  Reg#(Word) x(1);
  Reg#(Word) y(0);
```

```
typedef Bit#(32) Word;
module GCD;
  Reg#(Word) x(1);
  Reg#(Word) y(0);
  input Bool start;
  input Word a;
  input Word b;
```

```
typedef Bit#(32) Word;
module GCD;
  Reg#(Word) x(1);
  Reg#(Word) y(0);
  input Bool start;
  input Word a;
  input Word b;
  rule gcd;
    if (start) begin
      x \leftarrow a; y \leftarrow b;
    end else if (x != 0) begin
      if (x \ge y) begin // subtract
         X \le X - Y;
      end else begin // swap
         x \leftarrow y; y \leftarrow x;
      end
    end
  endrule
```

endmodule

```
typedef Bit#(32) Word;
module GCD;
  Reg#(Word) x(1);
  Reg#(Word) y(0);
  input Bool start;
  input Word a;
  input Word b;
  rule gcd;
    if (start) begin
      x <= a; y <= b;
    end else if (x != 0) begin
      if (x \ge y) begin // subtract
        X \le X - Y;
      end else begin // swap
        x \leftarrow y; y \leftarrow x;
      end
    end
  endrule
  method Word result = y;
  method Bool isDone = (x == 0);
endmodule
```

```
typedef Bit#(32) Word;
module GCD;
  Reg#(Word) x(1);
  Reg#(Word) y(0);
  input Bool start;
  input Word a;
  input Word b;
  rule gcd;
    if (start) begin
      x <= a; y <= b;
    end else if (x != 0) begin
      if (x \ge y) begin // subtract
        X \le X - Y;
      end else begin // swap
        x \leftarrow y; y \leftarrow x;
      end
    end
  endrule
  method Word result = y;
  method Bool isDone = (x == 0);
endmodule
```

Poor interface: Several inputs and outputs are closely coupled

```
typedef Bit#(32) Word;
module GCD;
 Reg#(Word) x(1);
 Reg#(Word) y(0);
 input Bool start;
 input Word a;
 input Word b;
 rule gcd;
    if (start) begin
      x <= a; y <= b;
    end else if (x != 0) begin
      if (x \ge y) begin // subtract
        X \leq X - Y;
      end else begin // swap
        x \le y; y \le x;
      end
   end
 endrule
 method Word result = y;
 method Bool isDone = (x == 0);
endmodule
```

Poor interface: Several inputs and outputs are closely coupled

 New GCD computation is started by setting start input to True and passing arguments through inputs a and b

```
typedef Bit#(32) Word;
module GCD;
 Reg#(Word) x(1);
 Reg#(Word) y(0);
 input Bool start;
 input Word a;
 input Word b;
 rule gcd;
   if (start) begin
      x <= a; y <= b;
   end else if (x != 0) begin
      if (x >= y) begin // subtract
        X \le X - Y;
      end else begin // swap
        x \le y; y \le x;
      end
   end
 endrule
 method Word result = y;
 method Bool isDone = (x == 0);
endmodule
```

Poor interface: Several inputs and outputs are closely coupled

- New GCD computation is started by setting start input to True and passing arguments through inputs a and b
- Several cycles later, the module signals that it has finished by having isDone return True; only then, the result method returns the correct result for gcd(a,b)

The previous GCD module has a poor interface

- The previous GCD module has a poor interface
 - Easy to misuse. Why?

- The previous GCD module has a poor interface
 - Easy to misuse. Why?
 - e.g., may forget to check isDone and read wrong result!

- The previous GCD module has a poor interface
 - Easy to misuse. Why?
 - e.g., may forget to check isDone and read wrong result!
 - Tedious to use. Why?

- The previous GCD module has a poor interface
 - Easy to misuse. Why?
 - e.g., may forget to check isDone and read wrong result!
 - Tedious to use. Why?
 - e.g., if start is False, we still have to set the a and b inputs, even though they are not used!

- The previous GCD module has a poor interface
 - Easy to misuse. Why?
 - e.g., may forget to check isDone and read wrong result!
 - Tedious to use. Why?
 - e.g., if start is False, we still have to set the a and b inputs, even though they are not used!
- To design good interfaces, group related inputs and outputs

- The previous GCD module has a poor interface
 - Easy to misuse. Why?
 - e.g., may forget to check isDone and read wrong result!
 - Tedious to use. Why?
 - e.g., if start is False, we still have to set the a and b inputs, even though they are not used!
- To design good interfaces, group related inputs and outputs
 - In our case, GCD should have:
 - A single output that is either invalid or a valid result
 - A single input that is either no arguments or arguments

- The previous GCD module has a poor interface
 - Easy to misuse. Why?
 - e.g., may forget to check isDone and read wrong result!
 - Tedious to use. Why?
 - e.g., if start is False, we still have to set the a and b inputs, even though they are not used!
- To design good interfaces, group related inputs and outputs
 - In our case, GCD should have:
 - A single output that is either invalid or a valid result
 - A single input that is either no arguments or arguments
 - This requires we learn about one last type...

The Maybe Type

- Maybe#(T) represents an optional value of type T
 - Either Invalid and no value, or Valid and a value

The Maybe Type

- Maybe#(T) represents an optional value of type T
 - Either Invalid and no value, or Valid and a value
- Possible implementation: A value + a valid bit

```
typedef struct { Bool valid; T value; } Maybe#(type T);
```

The Maybe Type

- Maybe#(T) represents an optional value of type T
 - Either Invalid and no value, or Valid and a value
- Possible implementation: A value + a valid bit typedef struct { Bool valid; T value; } Maybe#(type T);
 - Although we could implement our own, optional values are so common that Maybe#(T) has a few built-in operations

The Maybe Type

- Maybe#(T) represents an optional value of type T
 - Either Invalid and no value, or Valid and a value
- Possible implementation: A value + a valid bit

```
typedef struct { Bool valid; T value; } Maybe#(type T);
```

 Although we could implement our own, optional values are so common that Maybe#(T) has a few built-in operations

```
Maybe#(Word) x = Invalid; // no need to give value!
Maybe#(Word) y = Valid(42); // must specify a value
```

The Maybe Type

- Maybe#(T) represents an optional value of type T
 - Either Invalid and no value, or Valid and a value
- Possible implementation: A value + a valid bit

```
typedef struct { Bool valid; T value; } Maybe#(type T);
```

 Although we could implement our own, optional values are so common that Maybe#(T) has a few built-in operations

```
typedef struct {Word a; Word b;} GCDArgs;
module GCD;
   Reg#(Word) x(1);
   Reg#(Word) y(0);
```

endmodule

```
typedef struct {Word a; Word b;} GCDArgs;
module GCD;
  Reg#(Word) x(1);
  Reg#(Word) y(0);
  input Maybe#(GCDArgs) in;
```

endmodule

```
typedef struct {Word a; Word b;} GCDArgs;
module GCD;
  Reg#(Word) x(1);
  Reg#(Word) y(0);
  input Maybe#(GCDArgs) in;
  rule gcd;
    if (isValid(in)) begin
      let args = fromMaybe(?, in);
      x \leftarrow args.a; y \leftarrow args.b;
    end else if (x != 0) begin
      if (x \ge y) begin // subtract
        X \le X - Y;
      end else begin // swap
        x \leftarrow y; y \leftarrow x;
      end
    end
  endrule
```

endmodule

```
typedef struct {Word a; Word b;} GCDArgs;
module GCD;
  Reg#(Word) x(1);
  Reg#(Word) y(0);
  input Maybe#(GCDArgs) in;
  rule gcd;
    if (isValid(in)) begin
      let args = fromMaybe(?, in);
      x \leftarrow args.a; y \leftarrow args.b;
    end else if (x != 0) begin
      if (x \ge y) begin // subtract
        X \le X - Y;
      end else begin // swap
        x \le y; y \le x;
      end
    end
  endrule
  method Maybe#(Word) result =
    (x == 0)? Valid(y): Invalid;
endmodule
```

```
typedef struct {Word a; Word b;} GCDArgs;
module GCD;
  Reg#(Word) x(1);
  Reg#(Word) y(0);
  input Maybe#(GCDArgs) in;
  rule gcd;
    if (isValid(in)) begin
      let args = fromMaybe(?, in);
      x \leftarrow args.a; y \leftarrow args.b;
    end else if (x != 0) begin
      if (x \ge y) begin // subtract
        X \le X - Y;
      end else begin // swap
        x \le y; y \le x;
      end
    end
  endrule
  method Maybe#(Word) result =
    (x == 0)? Valid(y): Invalid;
endmodule
```

Single input and output:

- New GCD computation is started by setting a Valid input in (which always includes a and b)
- When GCD computation finishes, result becomes a Valid output

Two timing metrics of interest when designing a system:

- Two timing metrics of interest when designing a system:
- 1. Latency: The *delay* from when an input enters the system until its associated output is produced

- Two timing metrics of interest when designing a system:
- 1. Latency: The *delay* from when an input enters the system until its associated output is produced
- 2. Throughput: The *rate* at which inputs or outputs are processed

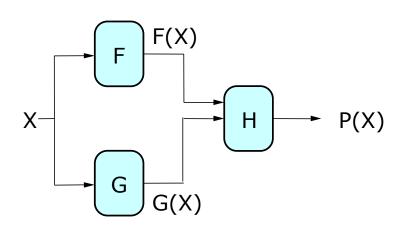
- Two timing metrics of interest when designing a system:
- 1. Latency: The *delay* from when an input enters the system until its associated output is produced
- 2. Throughput: The *rate* at which inputs or outputs are processed
- The metric to prioritize depends on the application

- Two timing metrics of interest when designing a system:
- 1. Latency: The *delay* from when an input enters the system until its associated output is produced
- 2. Throughput: The *rate* at which inputs or outputs are processed
- The metric to prioritize depends on the application
 - Airbag deployment system?

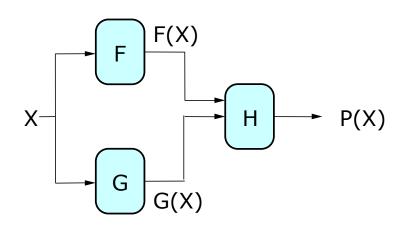
- Two timing metrics of interest when designing a system:
- 1. Latency: The *delay* from when an input enters the system until its associated output is produced
- 2. Throughput: The *rate* at which inputs or outputs are processed
- The metric to prioritize depends on the application
 - Airbag deployment system? Latency

- Two timing metrics of interest when designing a system:
- 1. Latency: The *delay* from when an input enters the system until its associated output is produced
- 2. Throughput: The *rate* at which inputs or outputs are processed
- The metric to prioritize depends on the application
 - Airbag deployment system? Latency
 - General-purpose processor?

- Two timing metrics of interest when designing a system:
- 1. Latency: The *delay* from when an input enters the system until its associated output is produced
- 2. Throughput: The *rate* at which inputs or outputs are processed
- The metric to prioritize depends on the application
 - Airbag deployment system? Latency
 - General-purpose processor? Throughput (maximize instructions/second)

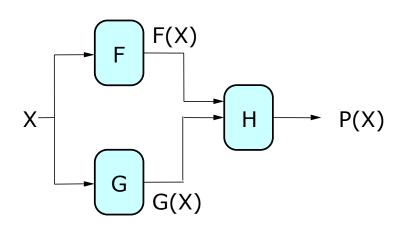


For combinational logic: $latency = t_{PD}$ $throughput = 1/t_{PD}$



For combinational logic: $latency = t_{PD}$ $throughput = 1/t_{PD}$

We can't get the answer any faster, but are we making effective use of our hardware at all times?

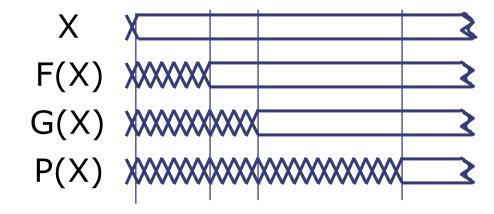


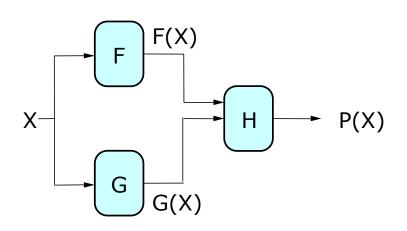
For combinational logic:

latency =
$$t_{PD}$$

throughput = $1/t_{PD}$

We can't get the answer any faster, but are we making effective use of our hardware at all times?



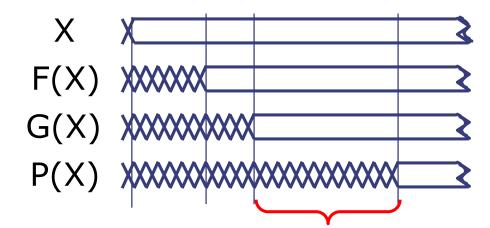


For combinational logic:

latency =
$$t_{PD}$$

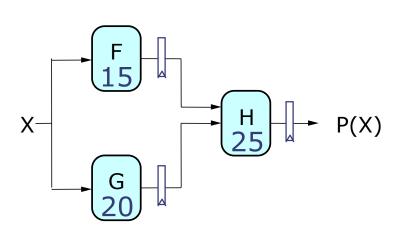
throughput = $1/t_{PD}$

We can't get the answer any faster, but are we making effective use of our hardware at all times?



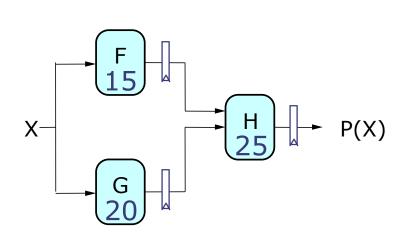
F & G are "idle", just holding their outputs stable while H performs its computation

Use registers to hold H's input stable!



Now F & G can be working on input X_{i+1} while H is performing its computation on X_i . We've created a 2-stage *pipeline*: if we have a valid input X during clock cycle j, P(X) is valid during clock j+2.

Use registers to hold H's input stable!

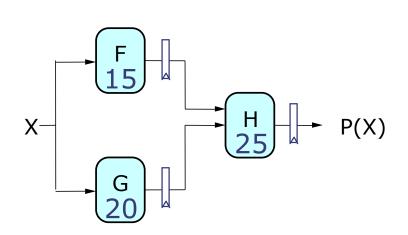


Now F & G can be working on input X_{i+1} while H is performing its computation on X_i . We've created a 2-stage *pipeline*: if we have a valid input X during clock cycle j, P(X) is valid during clock j+2.

Suppose F, G, H have propagation delays of 15, 20, 25 ns and we are using ideal registers ($t_{PD} = 0$, $t_{SETUP} = 0$):

unpipelined 45 ns 1/(45 ns)
2-stage pipeline

Use registers to hold H's input stable!

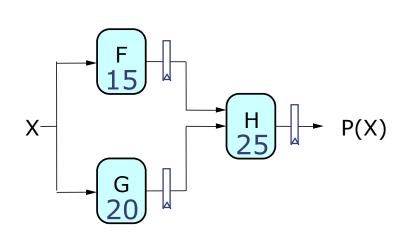


Now F & G can be working on input X_{i+1} while H is performing its computation on X_i . We've created a 2-stage *pipeline*: if we have a valid input X during clock cycle j, P(X) is valid during clock j+2.

Suppose F, G, H have propagation delays of 15, 20, 25 ns and we are using ideal registers ($t_{PD} = 0$, $t_{SETUP} = 0$):

	<u>latency</u>	throughput
unpipelined	45 ns	1/(45 ns)
2-stage pipeline	50 ns	

Use registers to hold H's input stable!



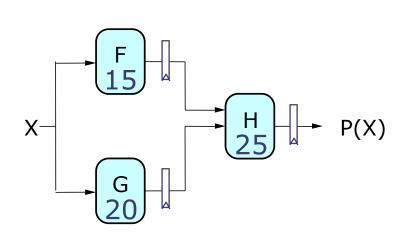
Now F & G can be working on input X_{i+1} while H is performing its computation on X_i . We've created a 2-stage *pipeline*: if we have a valid input X during clock cycle j, P(X) is valid during clock j+2.

Suppose F, G, H have propagation delays of 15, 20, 25 ns and we are using ideal registers ($t_{PD} = 0$, $t_{SETUP} = 0$):

	<u>latency</u>	throughput
unpipelined	45 ns	1/(45 ns)
2-stage pipeline	50 ns	
	worse	

March 17, 2022 MIT 6.004 Spring 2022 L12-12

Use registers to hold H's input stable!



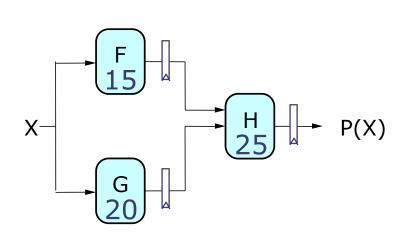
Now F & G can be working on input X_{i+1} while H is performing its computation on X_i . We've created a 2-stage *pipeline*: if we have a valid input X during clock cycle j, P(X) is valid during clock j+2.

Suppose F, G, H have propagation delays of 15, 20, 25 ns and we are using ideal registers ($t_{PD} = 0$, $t_{SETUP} = 0$):

	<u>latency</u>	throughput
unpipelined	45 ns	1/(45 ns)
2-stage pipeline	50 ns	1/(25 ns)
	worse	

March 17, 2022 MIT 6.004 Spring 2022 L12-12

Use registers to hold H's input stable!

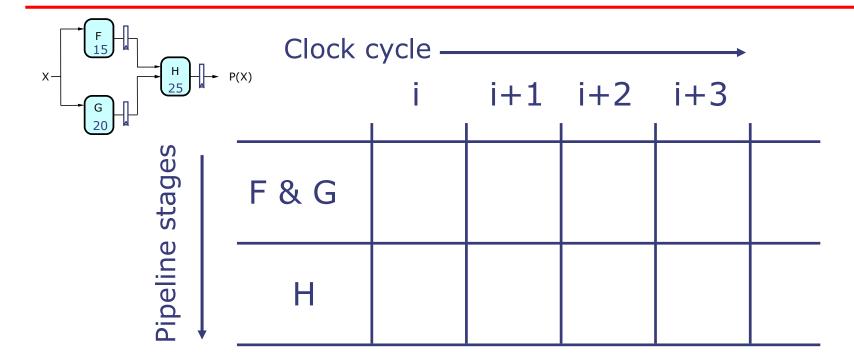


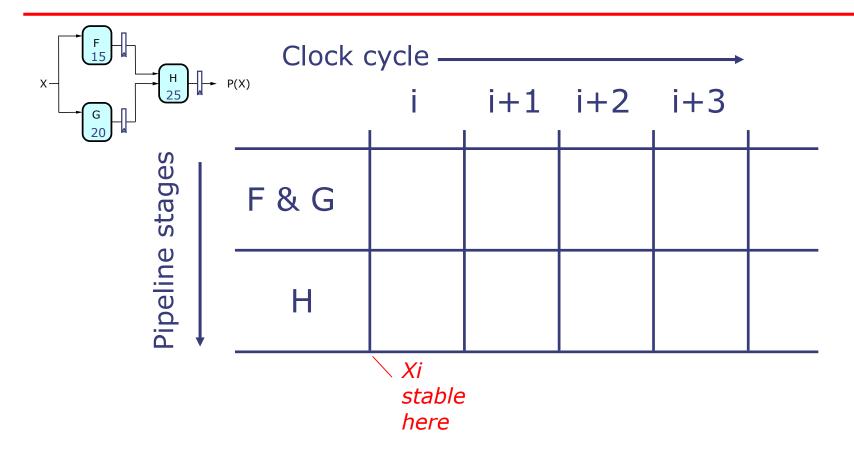
Now F & G can be working on input X_{i+1} while H is performing its computation on X_i . We've created a 2-stage *pipeline*: if we have a valid input X during clock cycle j, P(X) is valid during clock j+2.

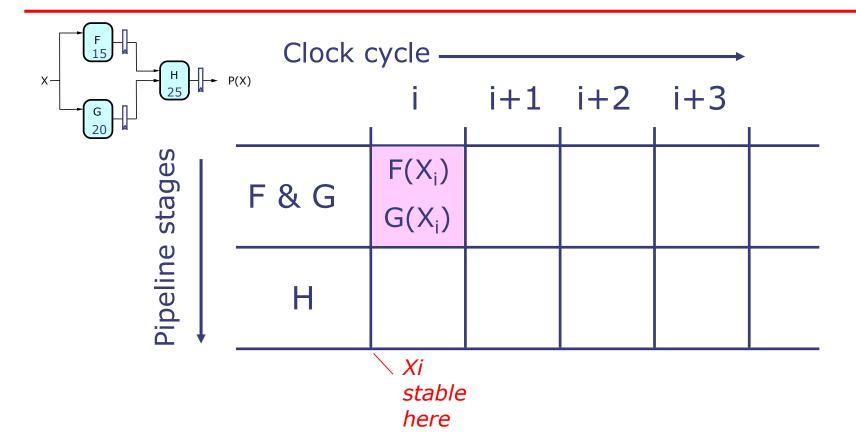
Suppose F, G, H have propagation delays of 15, 20, 25 ns and we are using ideal registers ($t_{PD} = 0$, $t_{SETUP} = 0$):

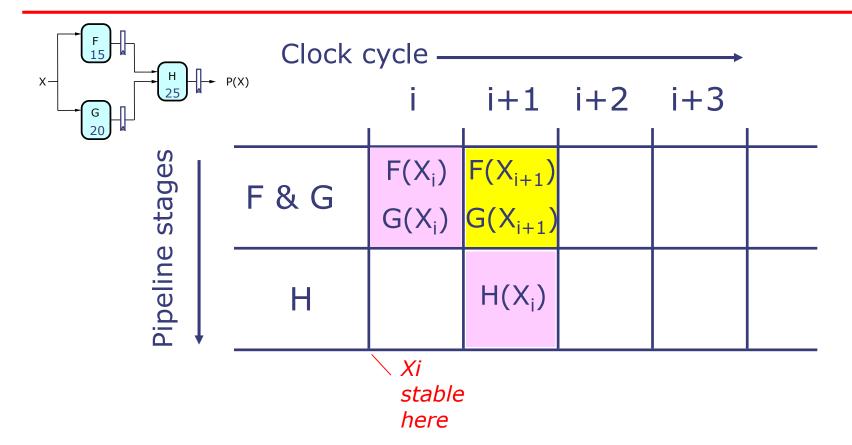
	<u>latency</u>	throughput
unpipelined	45 ns	1/(45 ns)
2-stage pipeline	50 ns	1/(25 ns)
	worse	better!

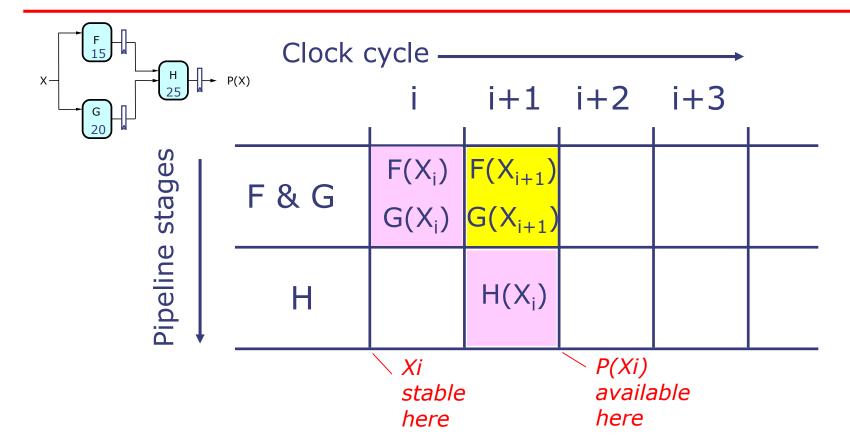
March 17, 2022 MIT 6.004 Spring 2022 L12-12

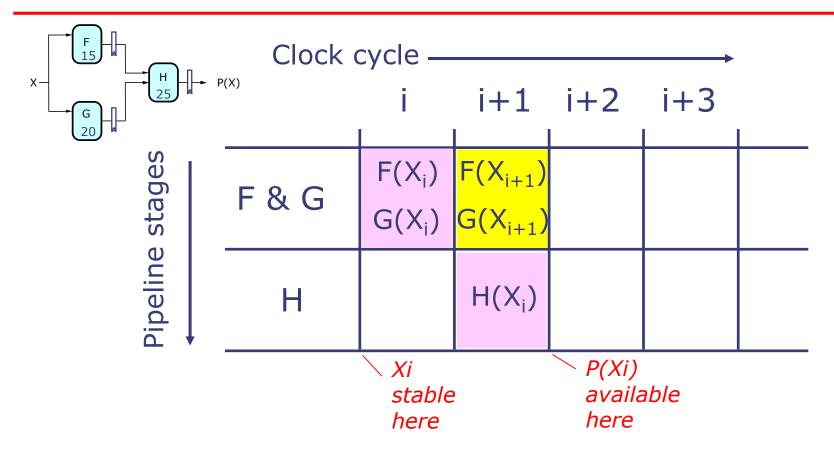




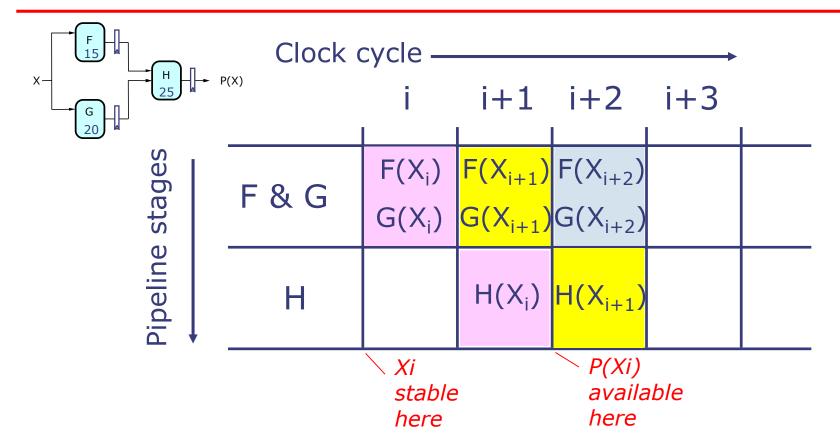




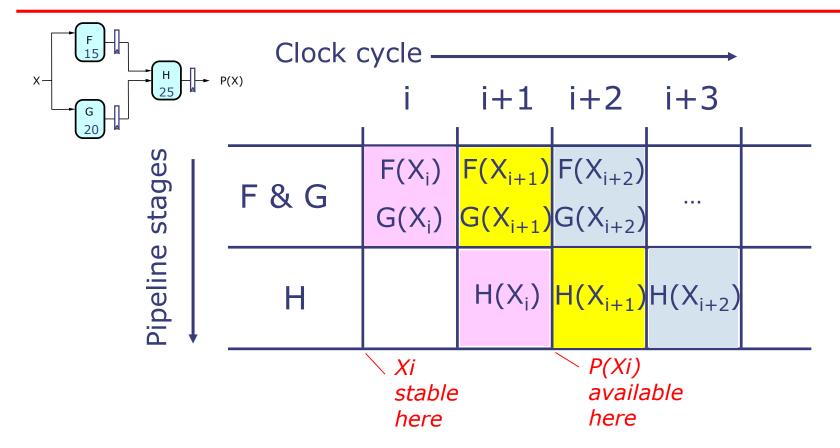




The results associated with a particular set of input data moves *diagonally* through the diagram, progressing through one pipeline stage each clock cycle.



The results associated with a particular set of input data moves *diagonally* through the diagram, progressing through one pipeline stage each clock cycle.



The results associated with a particular set of input data moves *diagonally* through the diagram, progressing through one pipeline stage each clock cycle.

Pipeline Conventions

Definition:

A well-formed *K-Stage Pipeline* ("K-pipeline") is an acyclic circuit having exactly K registers on *every* path from an input to an output.

A combinational circuit is thus a 0-stage pipeline.

Pipeline Conventions

Definition:

A well-formed *K-Stage Pipeline* ("K-pipeline") is an acyclic circuit having exactly K registers on *every* path from an input to an output.

A combinational circuit is thus a 0-stage pipeline.

Composition convention:

Every pipeline stage, hence every K-Stage pipeline, has a register on its *output* (not on its input).

Pipeline Conventions

Definition:

A well-formed *K-Stage Pipeline* ("K-pipeline") is an acyclic circuit having exactly K registers on *every* path from an input to an output.

A combinational circuit is thus a 0-stage pipeline.

Composition convention:

Every pipeline stage, hence every K-Stage pipeline, has a register on its *output* (not on its input).

Clock period:

The clock must have a period t_{CLK} sufficient to cover the longest register to register propagation delay plus setup time.

Pipeline Conventions

Definition:

A well-formed *K-Stage Pipeline* ("K-pipeline") is an acyclic circuit having exactly K registers on *every* path from an input to an output.

A combinational circuit is thus a 0-stage pipeline.

Composition convention:

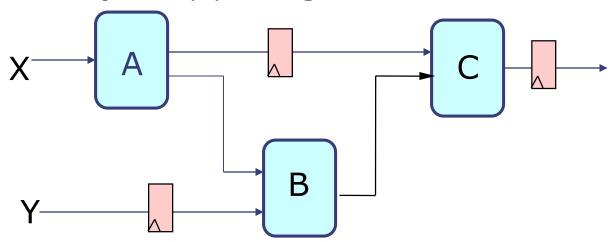
Every pipeline stage, hence every K-Stage pipeline, has a register on its *output* (not on its input).

Clock period:

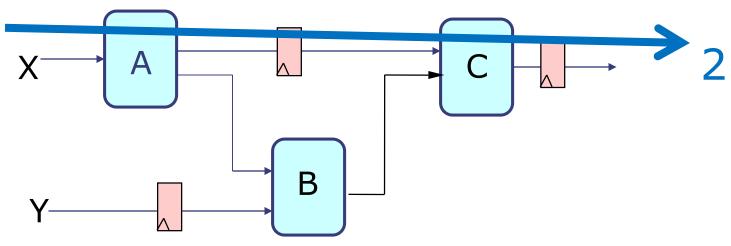
The clock must have a period t_{CLK} sufficient to cover the longest register to register propagation delay plus setup time.

K-pipeline latency $L = K * t_{CLK}$ K-pipeline throughput $T = 1 / t_{CLK}$

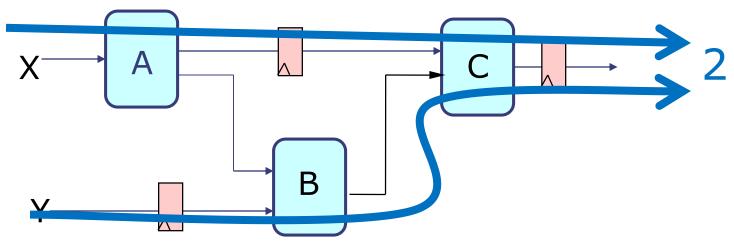
Consider a BAD job of pipelining:



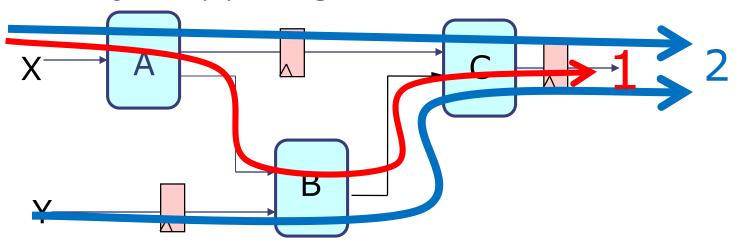
Consider a BAD job of pipelining:



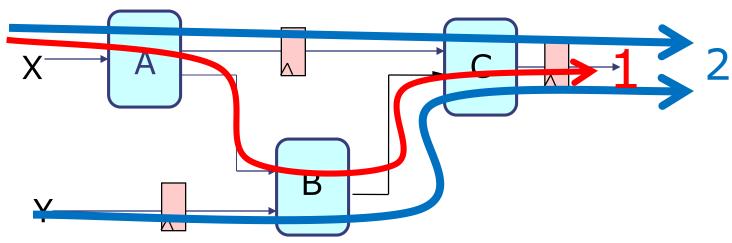
Consider a BAD job of pipelining:



Consider a BAD job of pipelining:



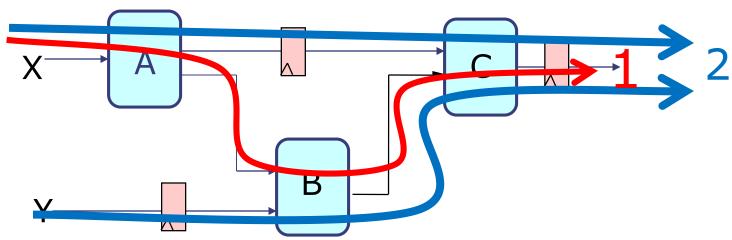
Consider a BAD job of pipelining:



For what value of K is the following circuit a K-Pipeline?

none

Consider a BAD job of pipelining:

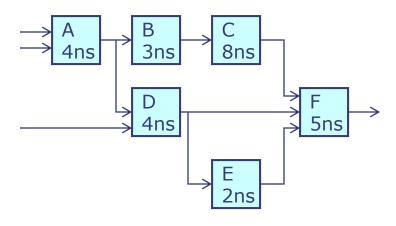


For what value of K is the following circuit a K-Pipeline? none

Problem:

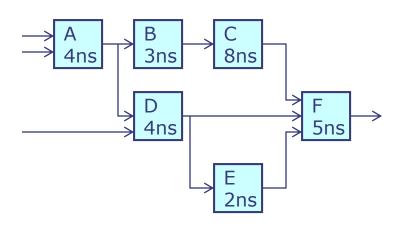
Successive inputs get mixed: e.g., $B(A(X_{i+1}), Y_i)$. This happens because some paths from inputs to outputs have 2 registers, and some have only 1!

This can't happen in a well-formed K pipeline!



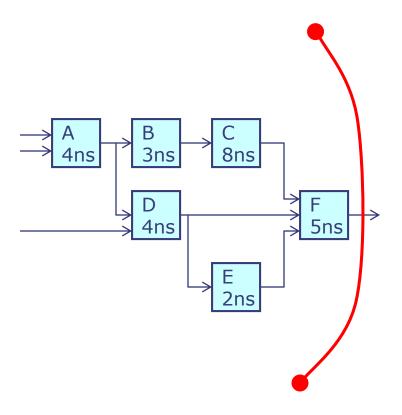
Step 1:

Draw a line that crosses every output in the circuit, and mark the endpoints as terminal points.



Step 1:

Draw a line that crosses every output in the circuit, and mark the endpoints as terminal points.

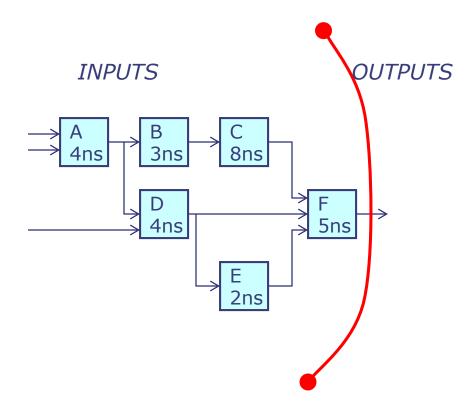


Step 1:

Draw a line that crosses every output in the circuit, and mark the endpoints as terminal points.

Step 2:

Continue to draw new lines between the terminal points across various circuit connections, ensuring that every connection crosses each line in the same direction. These lines demarcate *pipeline stages*.

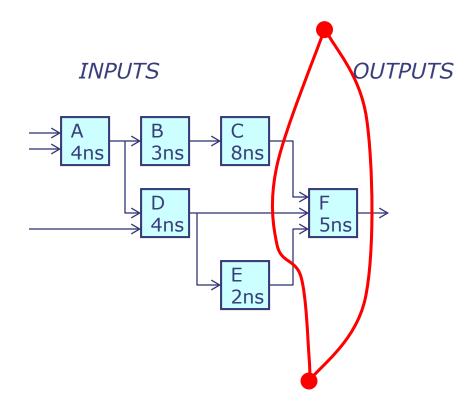


Step 1:

Draw a line that crosses every output in the circuit, and mark the endpoints as terminal points.

Step 2:

Continue to draw new lines between the terminal points across various circuit connections, ensuring that every connection crosses each line in the same direction. These lines demarcate *pipeline stages*.

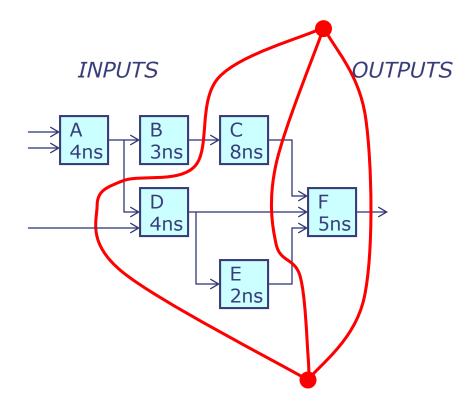


Step 1:

Draw a line that crosses every output in the circuit, and mark the endpoints as terminal points.

Step 2:

Continue to draw new lines between the terminal points across various circuit connections, ensuring that every connection crosses each line in the same direction. These lines demarcate *pipeline stages*.

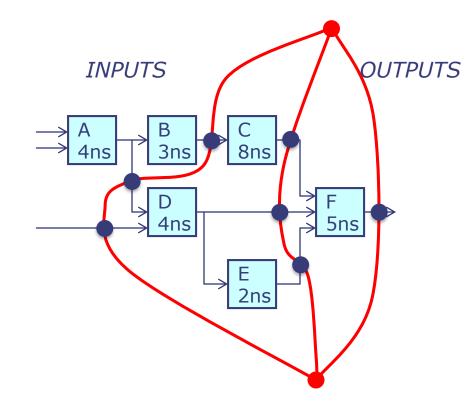


Step 1:

Draw a line that crosses every output in the circuit, and mark the endpoints as terminal points.

Step 2:

Continue to draw new lines between the terminal points across various circuit connections, ensuring that every connection crosses each line in the same direction. These lines demarcate *pipeline stages*.

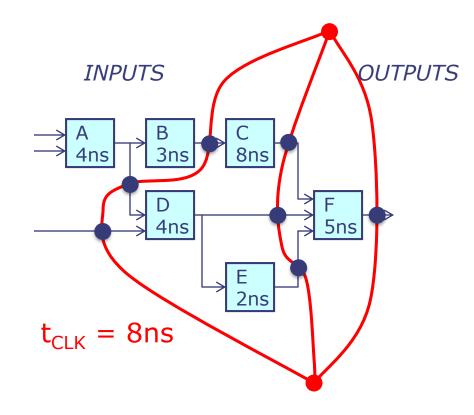


Step 1:

Draw a line that crosses every output in the circuit, and mark the endpoints as terminal points.

Step 2:

Continue to draw new lines between the terminal points across various circuit connections, ensuring that every connection crosses each line in the same direction. These lines demarcate *pipeline stages*.

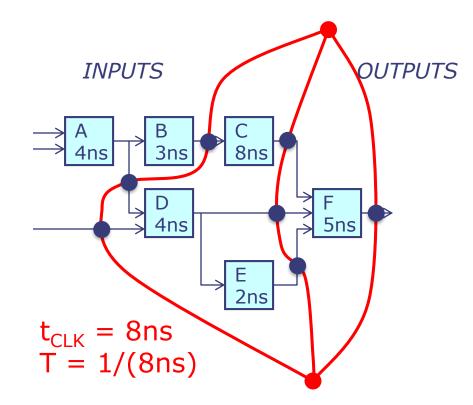


Step 1:

Draw a line that crosses every output in the circuit, and mark the endpoints as terminal points.

Step 2:

Continue to draw new lines between the terminal points across various circuit connections, ensuring that every connection crosses each line in the same direction. These lines demarcate *pipeline stages*.

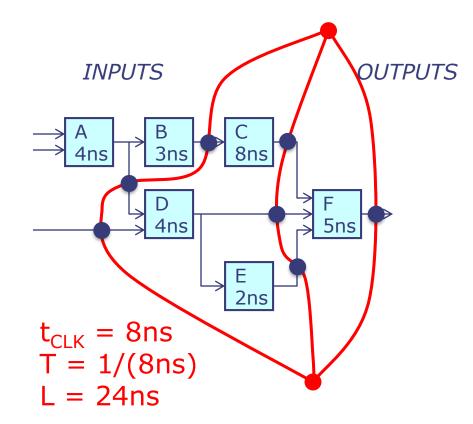


Step 1:

Draw a line that crosses every output in the circuit, and mark the endpoints as terminal points.

Step 2:

Continue to draw new lines between the terminal points across various circuit connections, ensuring that every connection crosses each line in the same direction. These lines demarcate *pipeline stages*.



Step 1:

Draw a line that crosses every output in the circuit, and mark the endpoints as terminal points.

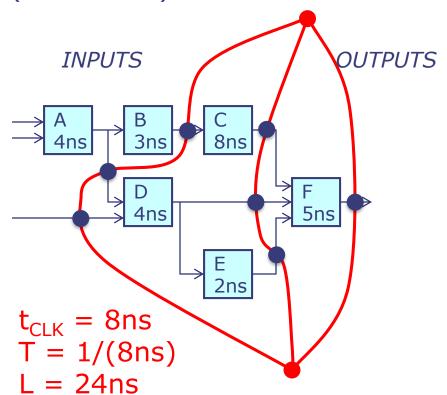
Step 2:

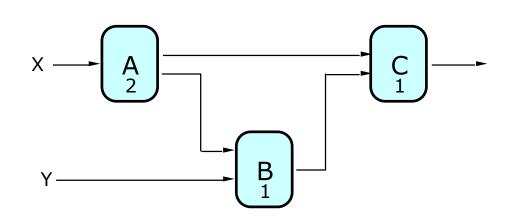
Continue to draw new lines between the terminal points across various circuit connections, ensuring that every connection crosses each line in the same direction. These lines demarcate *pipeline stages*.

Adding a pipeline register at every point where a separating line crosses a connection will always generate a valid pipeline.

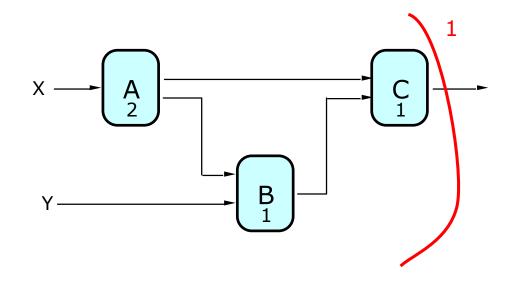
Strategy:

Focus your attention on placing pipelining registers around the slowest circuit elements (bottlenecks).





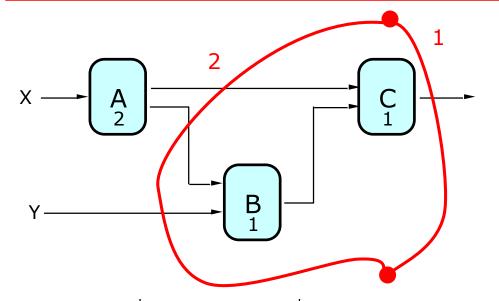
	LATENCY	THROUGHPU
0-pipe:	4	1/4
1-pipe:		
2-pipe:		
3-pipe:		



OBSERVATIONS:

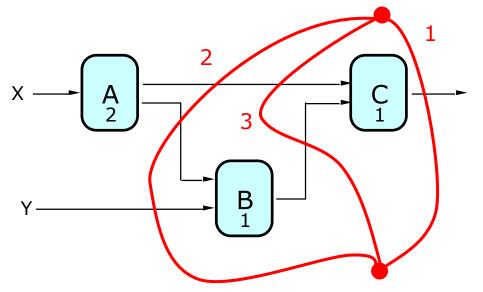
• 1-pipeline improves neither L nor T.

	LATENCY	THROUGHPU
0-pipe:	4	1/4
1-pipe:	4	1/4
2-pipe:		
3-pipe:		



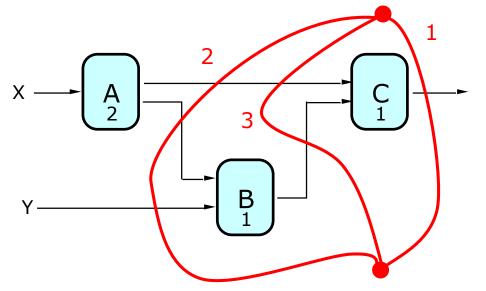
	LATENCY	THROUGHPUT
0-pipe:	4	1/4
1-pipe:	4	1/4
2-pipe:	4	1/2
3-pipe:		

- 1-pipeline improves neither L nor T.
- T improved by breaking long combinational paths, allowing faster clock.



	LATENCY	THROUGHPUT
0-pipe:	4	1/4
1-pipe:	4	1/4
2-pipe:	4	1/2
3-pipe:	6	1/2

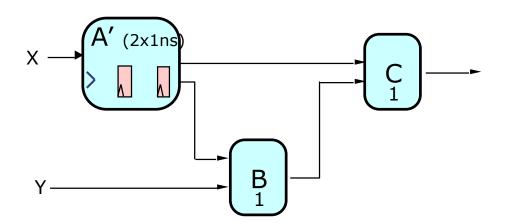
- 1-pipeline improves neither L nor T.
- T improved by breaking long combinational paths, allowing faster clock.
- Too many stages cost L, don't improve T.



	LATENCY	THROUGHPUT
0-pipe:	4	1/4
1-pipe:	4	1/4
2-pipe:	4	1/2
3-pipe:	6	1/2

- 1-pipeline improves neither L nor T.
- T improved by breaking long combinational paths, allowing faster clock.
- Too many stages cost L, don't improve T.
- Back-to-back registers are sometimes needed to keep pipeline wellformed.

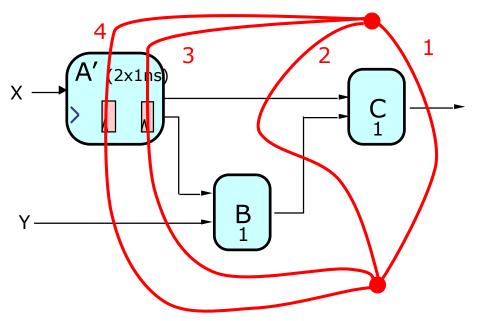
Pipelined Components



Pipelined systems can be hierarchical:

 Replacing a slow combinational component with a kpipe version may let us decrease the clock period

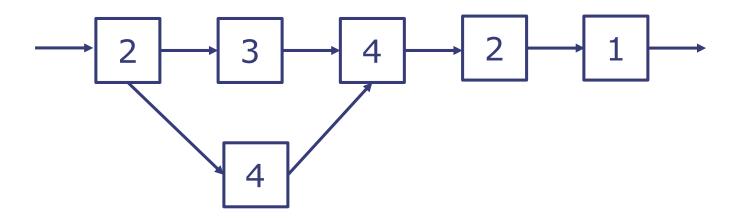
Pipelined Components

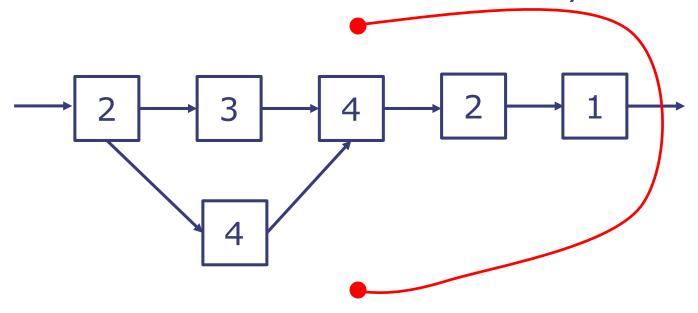


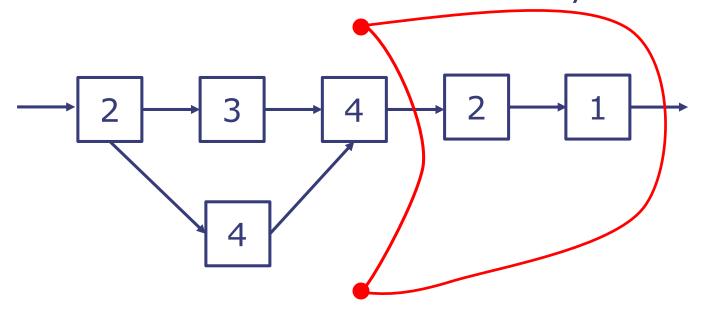
4-stage pipeline, throughput=1

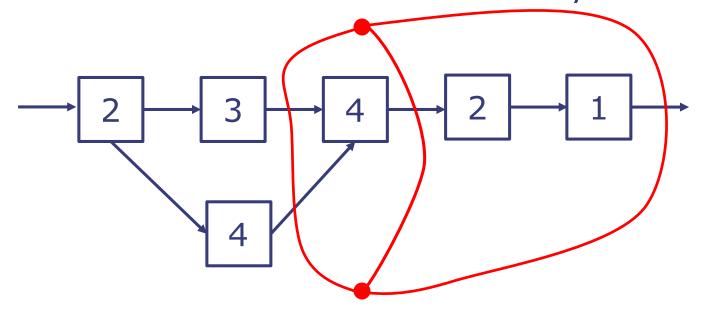
Pipelined systems can be hierarchical:

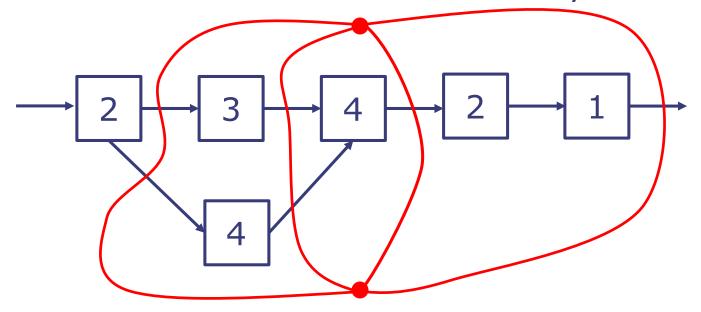
- Replacing a slow combinational component with a kpipe version may let us decrease the clock period
- Must account for new pipeline stages in our plan

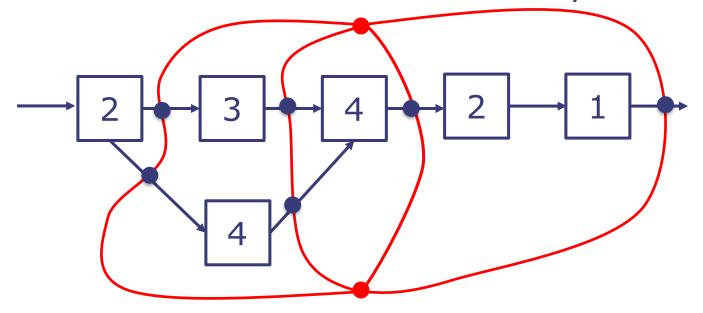




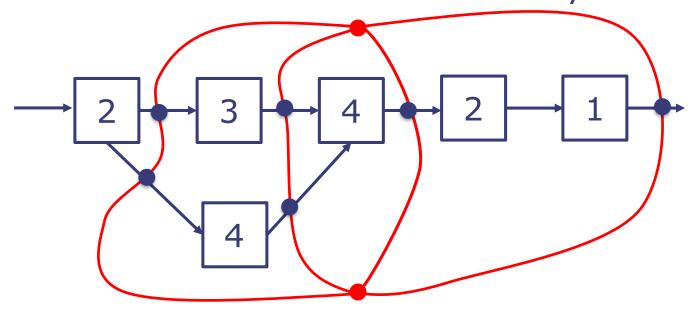






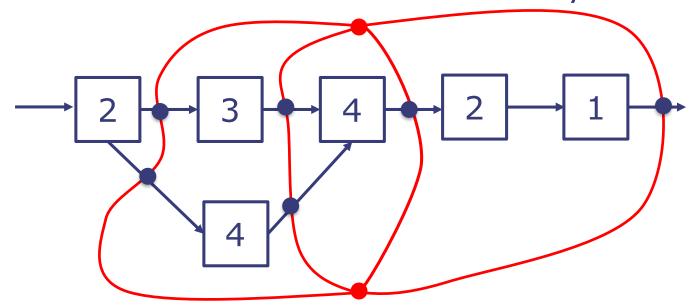


 Pipeline the following circuit for maximum throughput while minimizing latency. The number in each module is the module's latency.



What is the latency and throughput of your pipelined circuit?

 Pipeline the following circuit for maximum throughput while minimizing latency. The number in each module is the module's latency.



• What is the latency and throughput of your pipelined circuit?

• $t_{CLK} = 4$

$$T = 1/(4)$$

L = 4*4 = 16

Design Tradeoffs Introduction: Multiplier Case Study

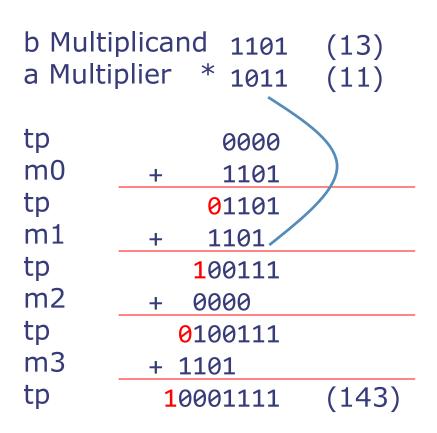
b Multip a Multip		1101 1011	(13) (11)
tp		0000	
m0	+	1101	
tp	6	1101	
m1	+ 1	L101	
tp	16	00111	
m2	+ 06	900	
tp	016	90111	
m3	+ 116	91	
tp	1000	1111	(143)

b Multip a Multip		1101 1011	(13)(11)
tp		0000	
m0	+	1101	
tp	(1101	
m1	+ 1	L101	
tp	16	00111	
m2	+ 00	900	
tp	016	00111	
m3	+ 116	91	
tp	1000	91111	(143)

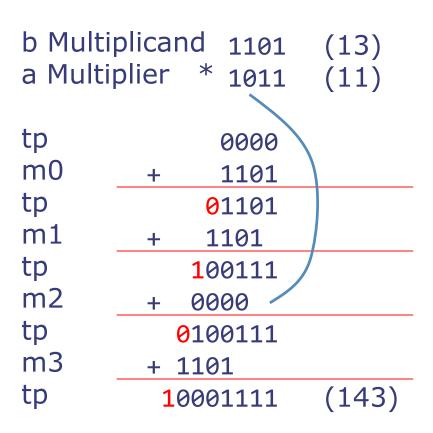
$$mi = (a[i]==0)? 0 : b;$$

b Multip a Multip		1101 1011	(13)(11)
tp		0000	
m0	+	1101	
tp	(1101	
m1	+ 1	1101	
tp	16	90111	
m2	+ 00	900	
tp	016	90111	
m3	+ 110	21	
tp	1000	91111	(143)

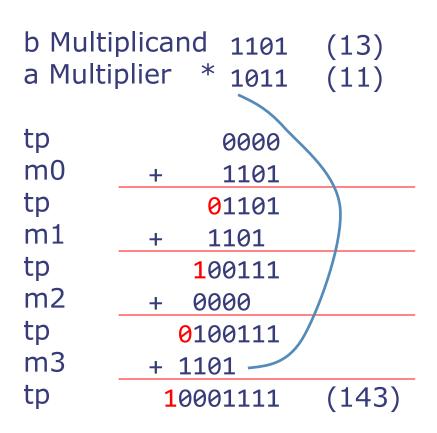
$$mi = (a[i]==0)? 0 : b;$$



$$mi = (a[i]==0)? 0 : b;$$



$$mi = (a[i]==0)? 0 : b;$$



$$mi = (a[i]==0)? 0 : b;$$

b Multi a Multi	•	1101 1011	(13)(11)
tp		0000	
m0	+	1101	
tp	6	1101	
m1	+ 1	101	
tp	10	00111	
m2	+ 00	000	
tp	010	00111	
m3	+ 110	1	
tp	1000	1111	(143)

At each step we add either b (1101) or 0 to the result depending upon a bit in the multiplier

$$mi = (a[i]==0)? 0 : b;$$

We also shift the result by one position at every step

b Multi	plicand	1101	(13)
a Multi	plier *	1011	(11)
tp		0000	
m0	+	1101	
tp	6	1101	
m1	+ 1	L101	
tp	16	00111	
m2	+ 06	900	
tp	016	00111	
m3	+ 116	91	
tp	1 000	91111	(143)

At each step we add either b (1101) or 0 to the result depending upon a bit in the multiplier

$$mi = (a[i]==0)? 0 : b;$$

We also shift the result by one position at every step

Note the first addition is unnecessary because it simply yields m0

b Multipa Multip		1101	(13)
a Multip	Jilei	1011	(11)
tp		0000	
m0	+	1101	
tp	(1101	
m1	+ 1	1101	
tp	16	90111	
m2	+ 00	900	
tp	016	90111	
m3	+ 116	21	
tp	1000	91111	(143)

At each step we add either b (1101) or 0 to the result depending upon a bit in the multiplier

$$mi = (a[i]==0)? 0 : b;$$

We also shift the result by one position at every step

Note the first addition is unnecessary because it simply yields m0

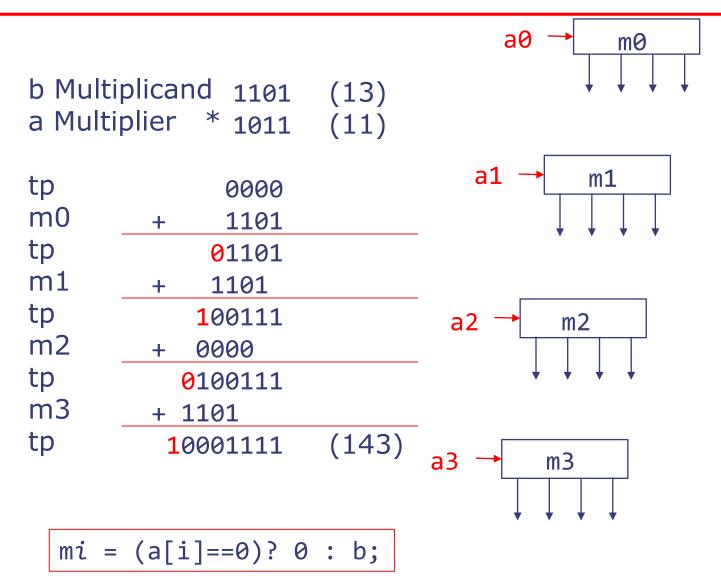
Also note that these are unsigned binary numbers.

Multiplication by repeated addition Combinational circuit

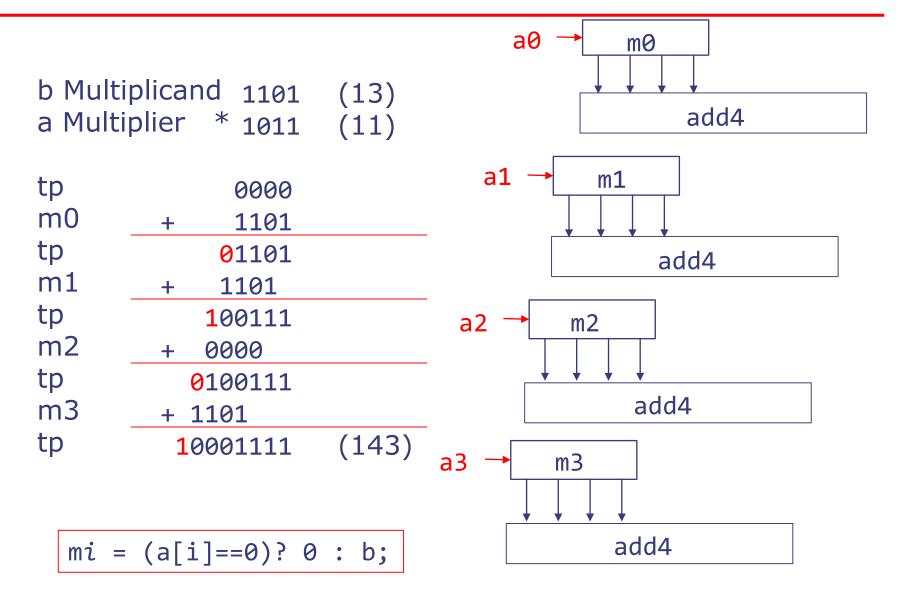
```
b Multiplicand 1101
                       (13)
a Multiplier
             * 1011
                       (11)
tp
               0000
m0
               1101
         +
tp
              01101
m1
              1101
tp
             100111
m2
             0000
tp
            0100111
m3
         + 1101
tp
                       (143)
          10001111
```

```
mi = (a[i]==0)? 0 : b;
```

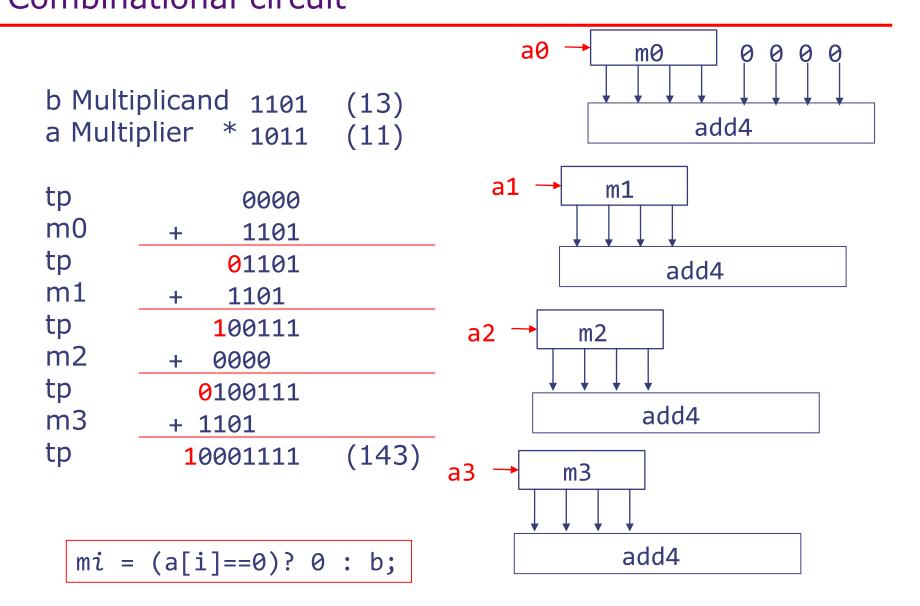
Combinational circuit



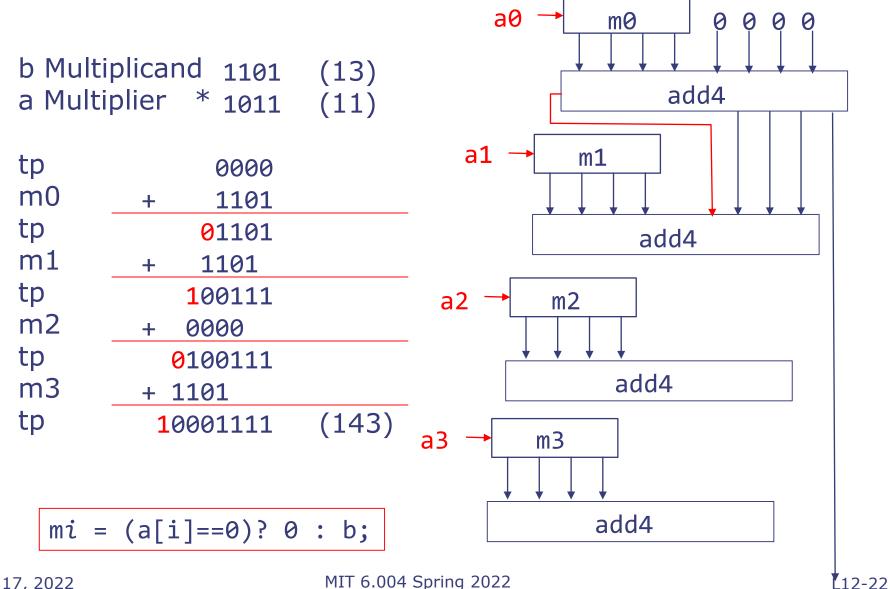
Combinational circuit



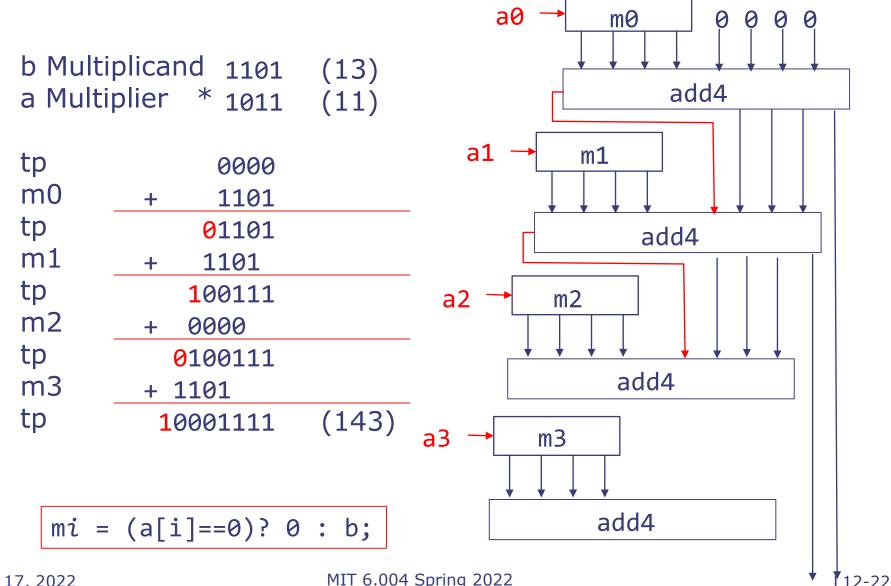
Multiplication by repeated addition Combinational circuit



Combinational circuit

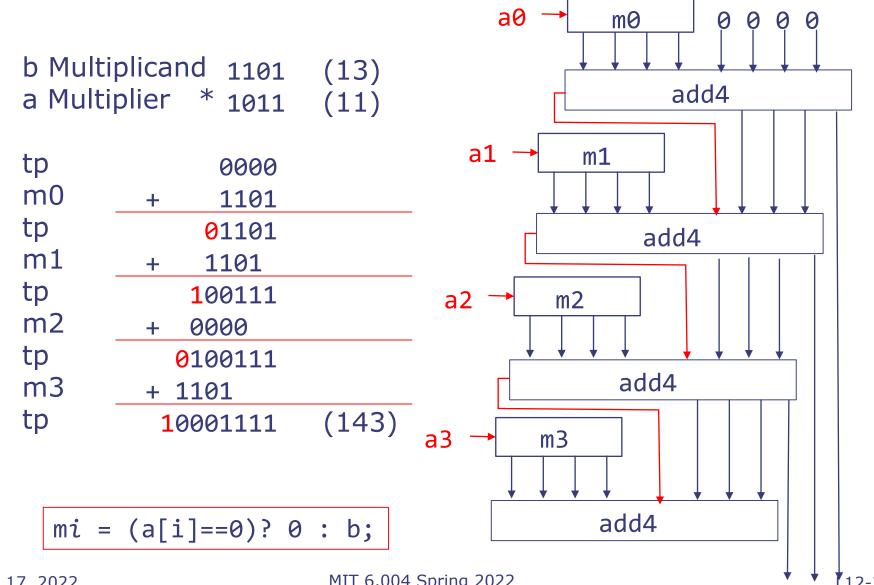


Combinational circuit



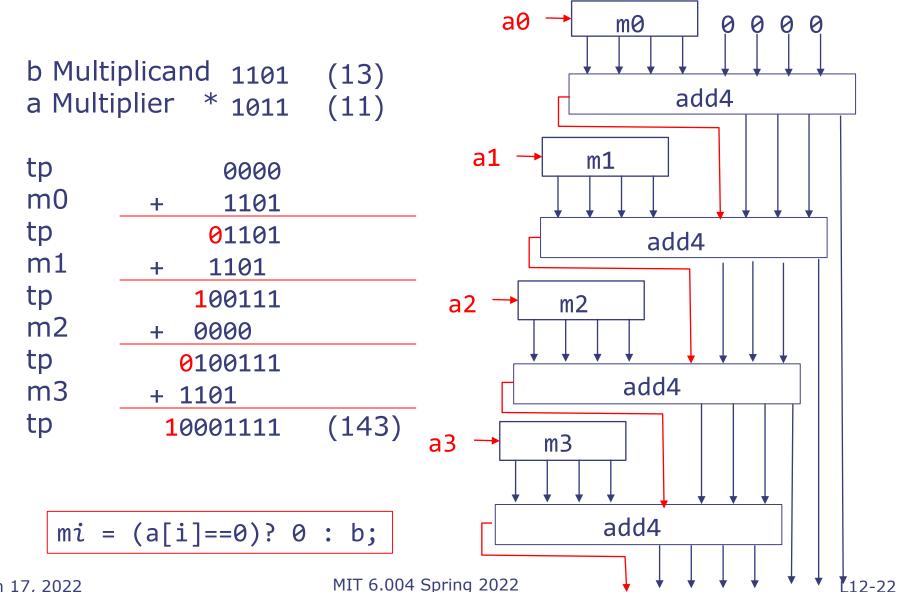
March 17, 2022

Combinational circuit



March 17, 2022

Combinational circuit

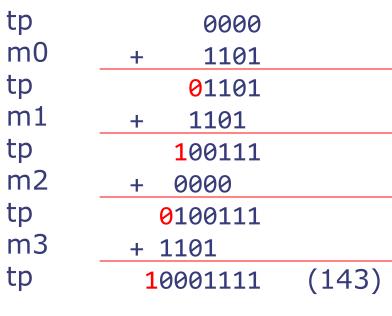


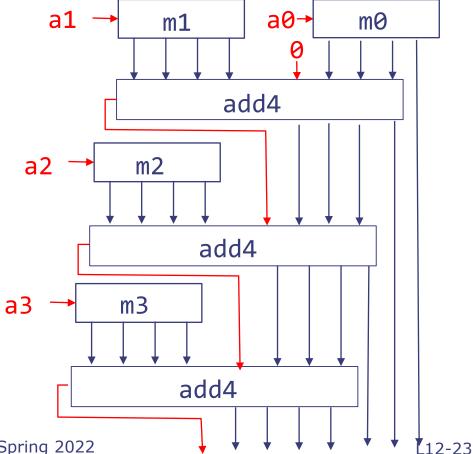
March 17, 2022

Combinational circuit

b Multiplicand a Multiplier *	

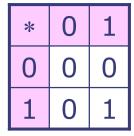
First adder is unnecessary (always yields m0)





mi = (a[i]==0)? 0 : b;

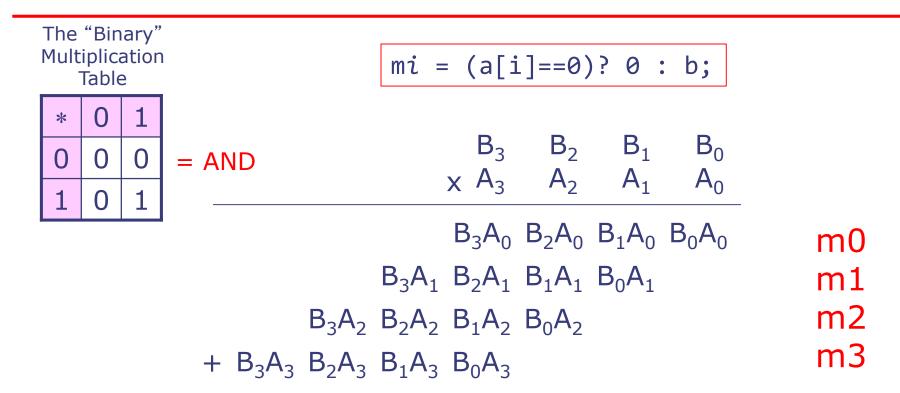
The "Binary" Multiplication Table

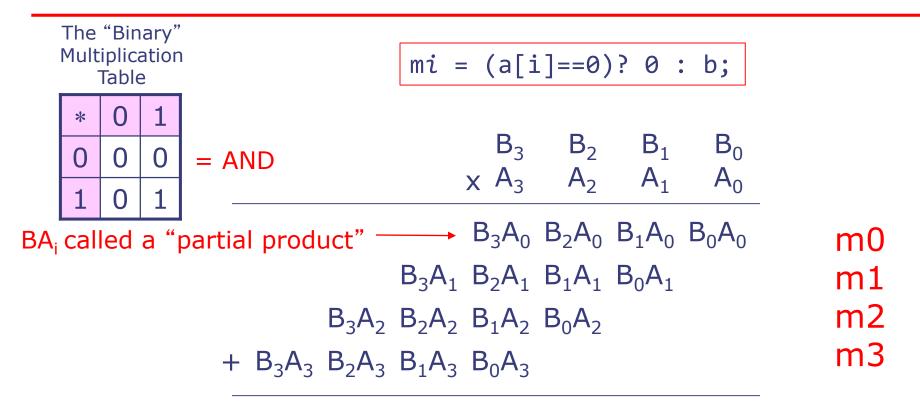


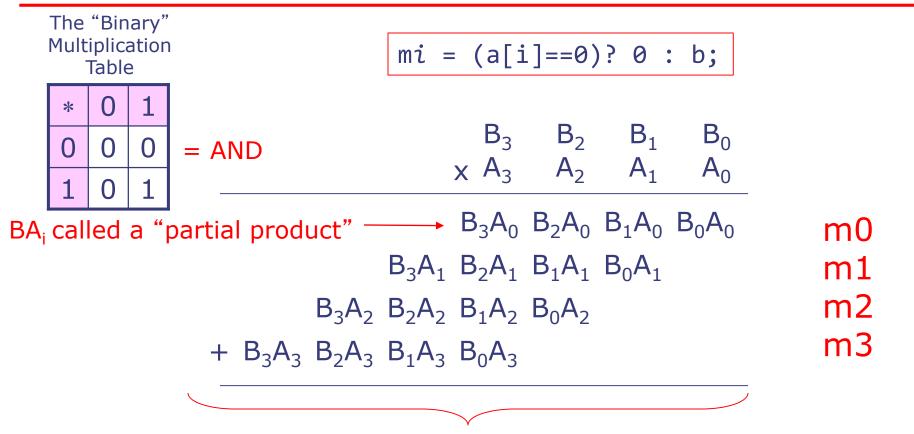
$$mi = (a[i]==0)? 0 : b;$$

$$B_{3}A_{0}$$
 $B_{2}A_{0}$ $B_{1}A_{0}$ $B_{0}A_{0}$
 $B_{3}A_{1}$ $B_{2}A_{1}$ $B_{1}A_{1}$ $B_{0}A_{1}$
 $B_{3}A_{2}$ $B_{2}A_{2}$ $B_{1}A_{2}$ $B_{0}A_{2}$
 $B_{3}A_{3}$ $B_{2}A_{3}$ $B_{1}A_{3}$ $B_{0}A_{3}$

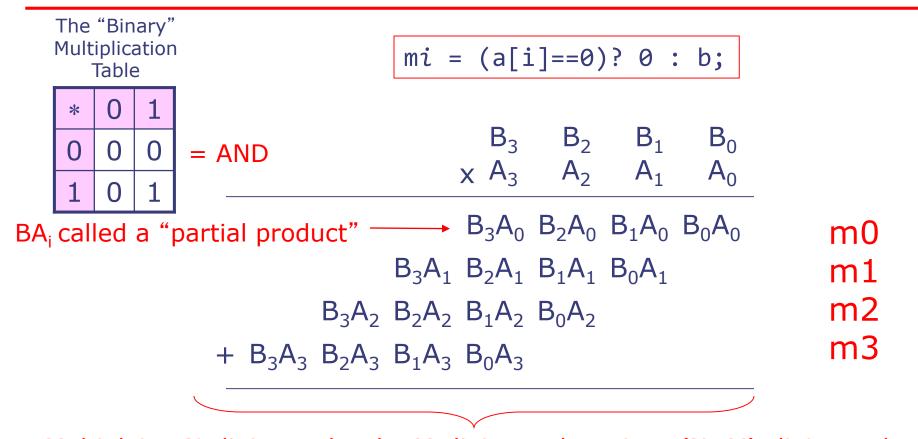
m0 m1 m2 m3







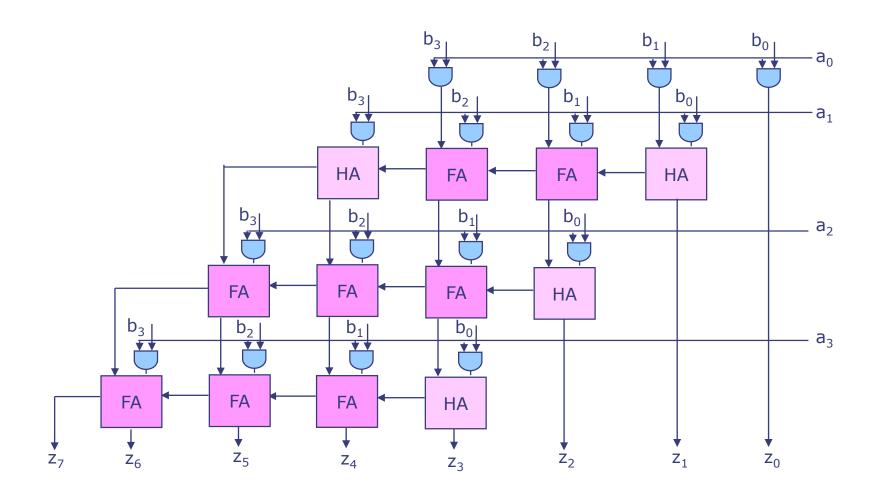
Multiplying N-digit number by M-digit number gives (N+M)-digit result

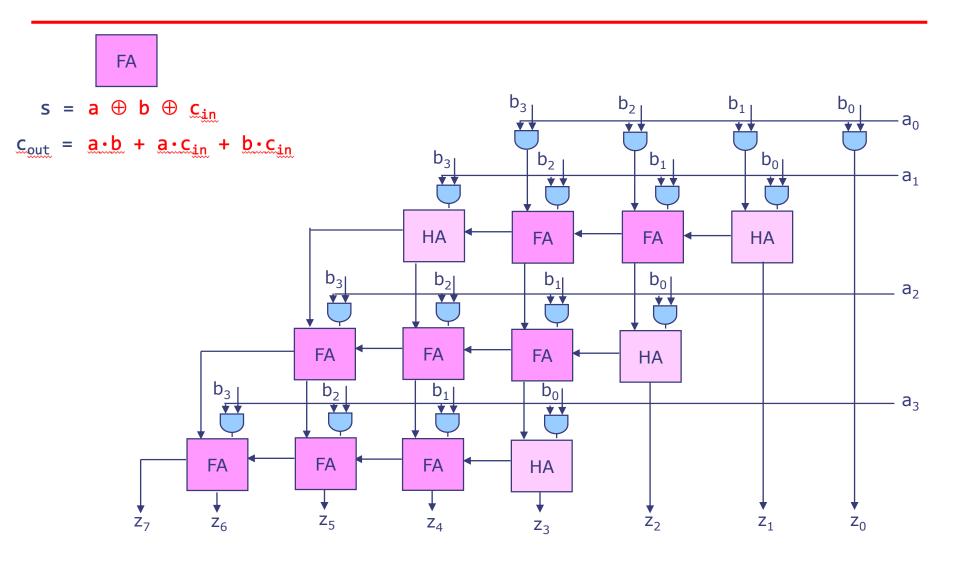


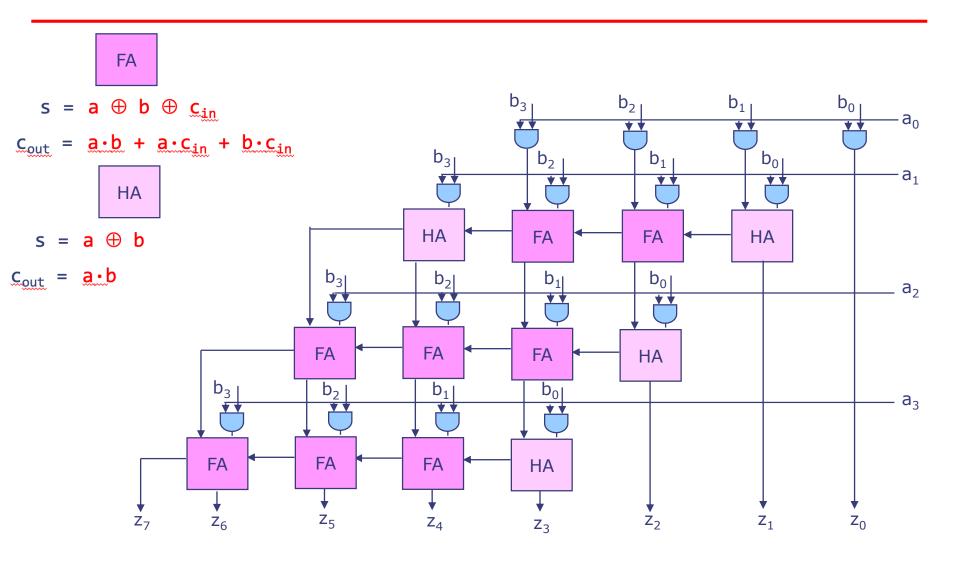
Multiplying N-digit number by M-digit number gives (N+M)-digit result

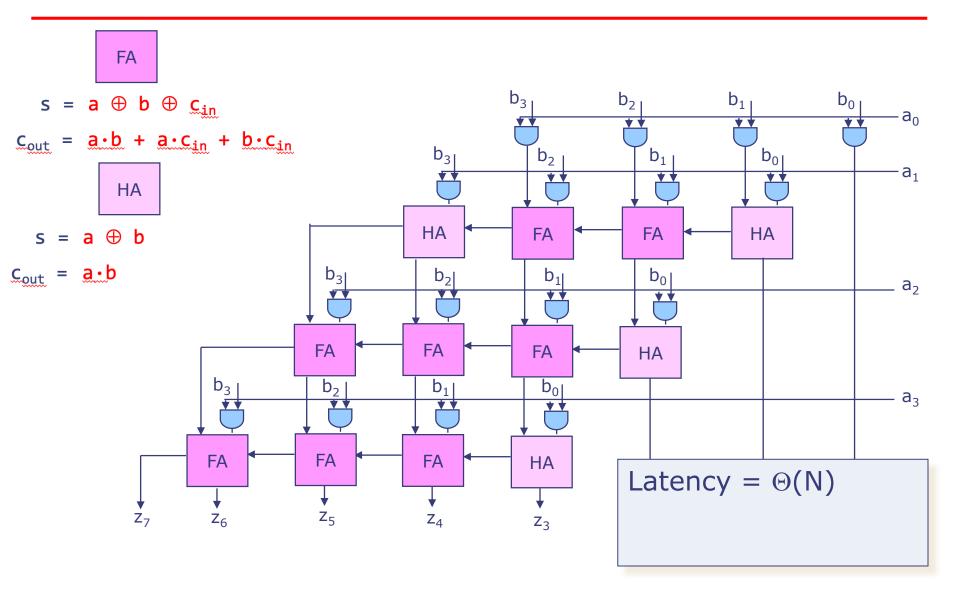
Easy part: forming partial products (bunch of AND gates)

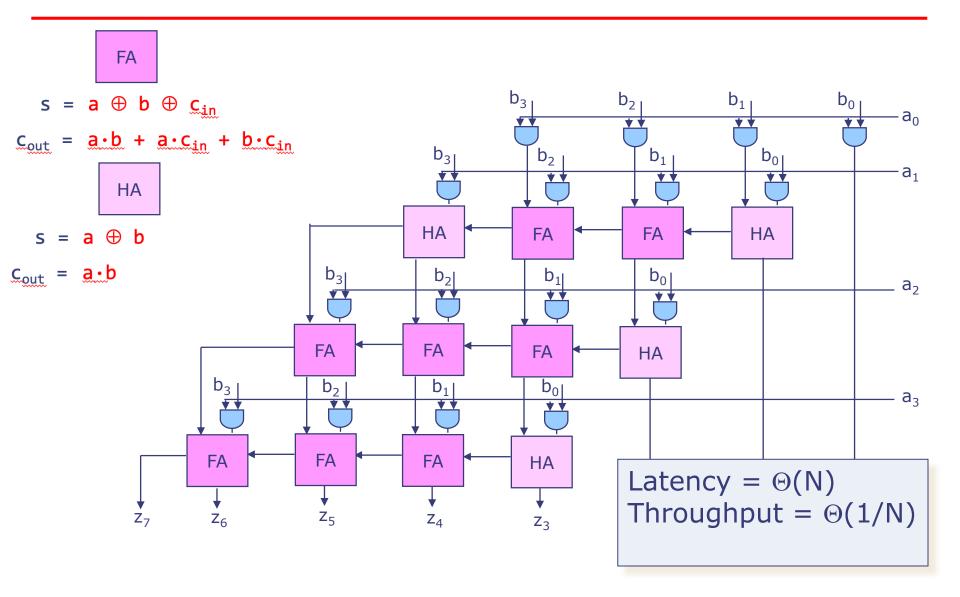
Hard part: adding M N-bit partial products

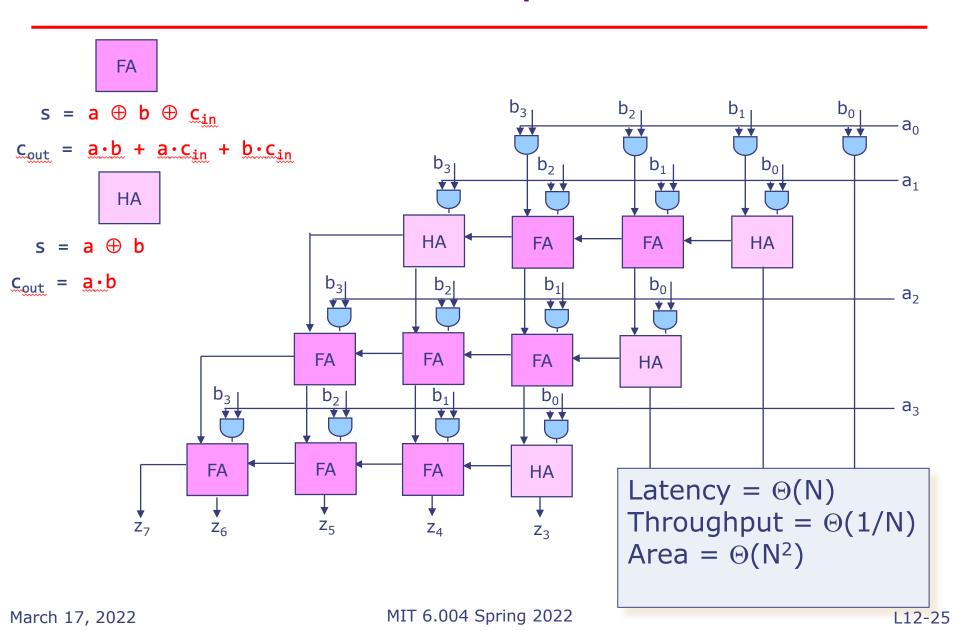


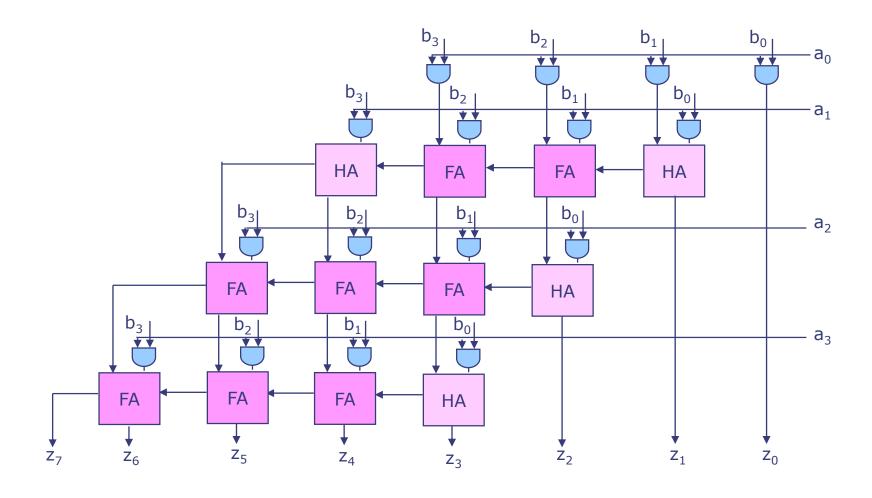


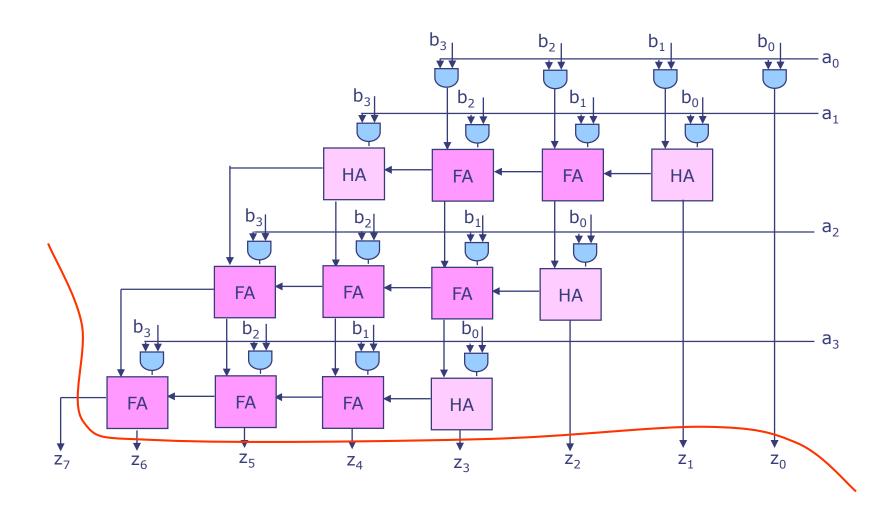


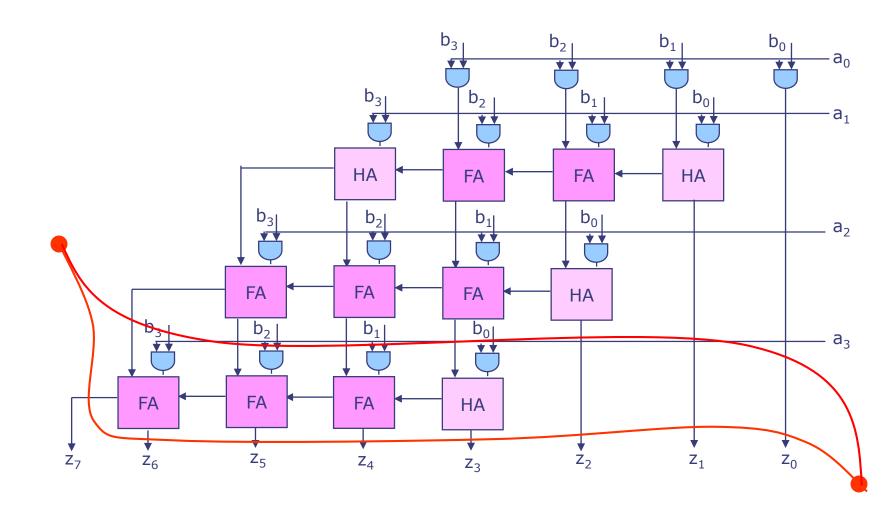


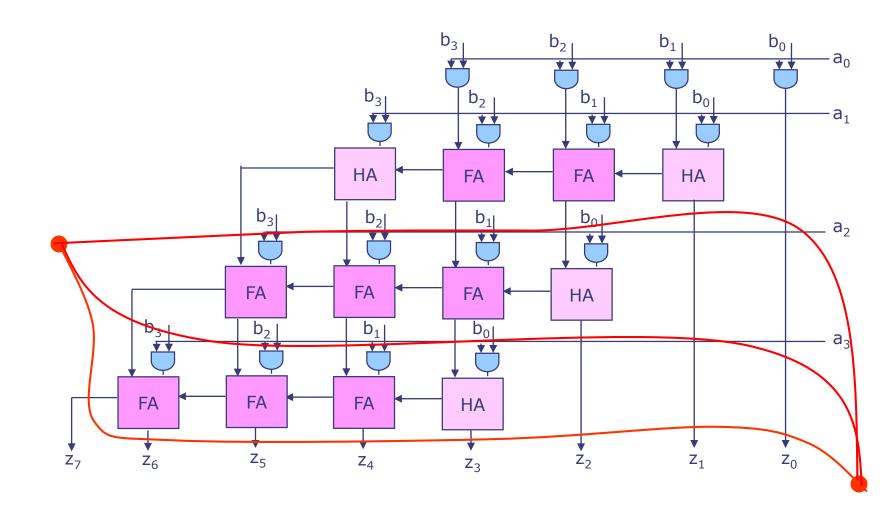


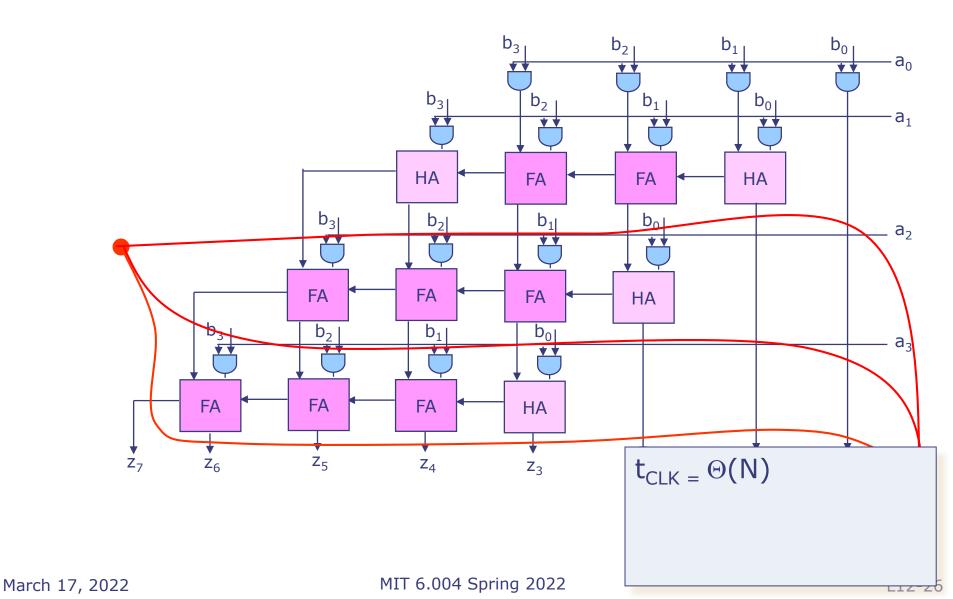


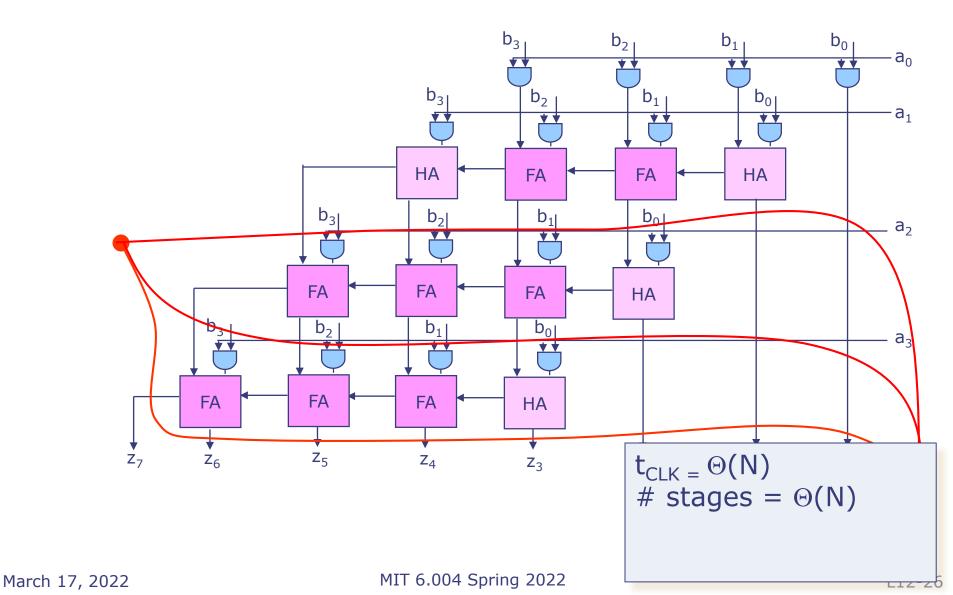


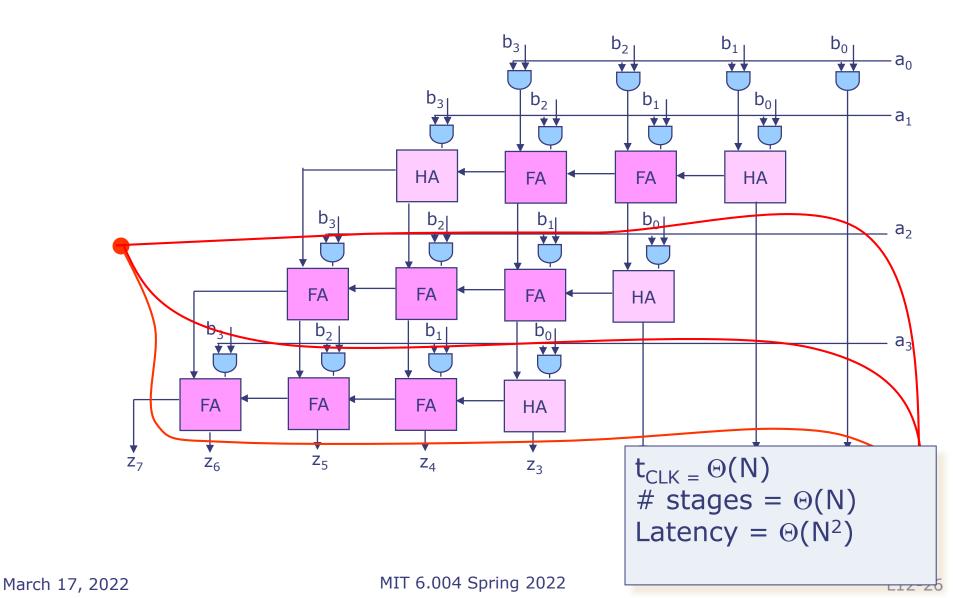


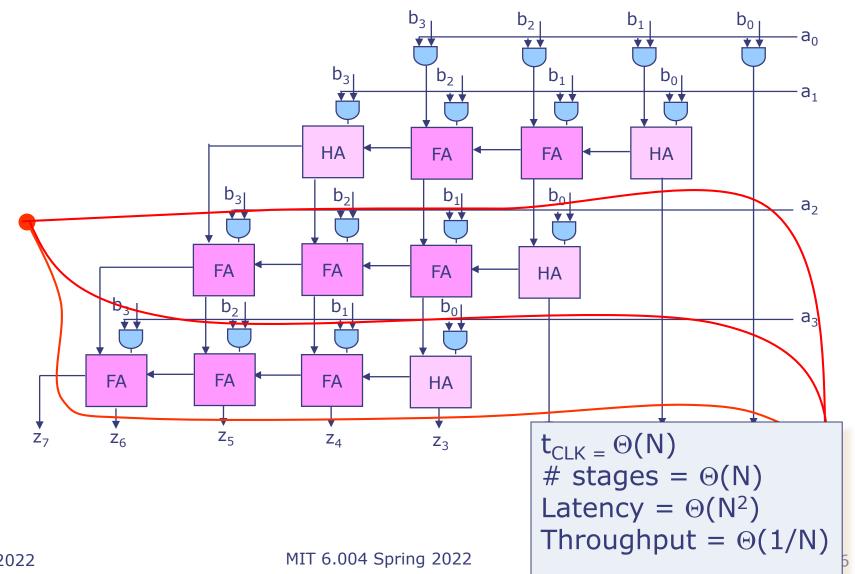




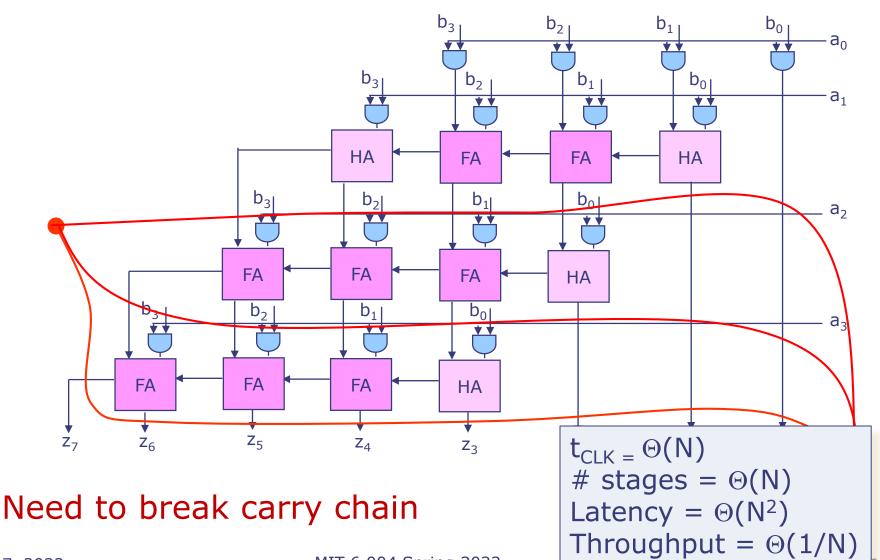




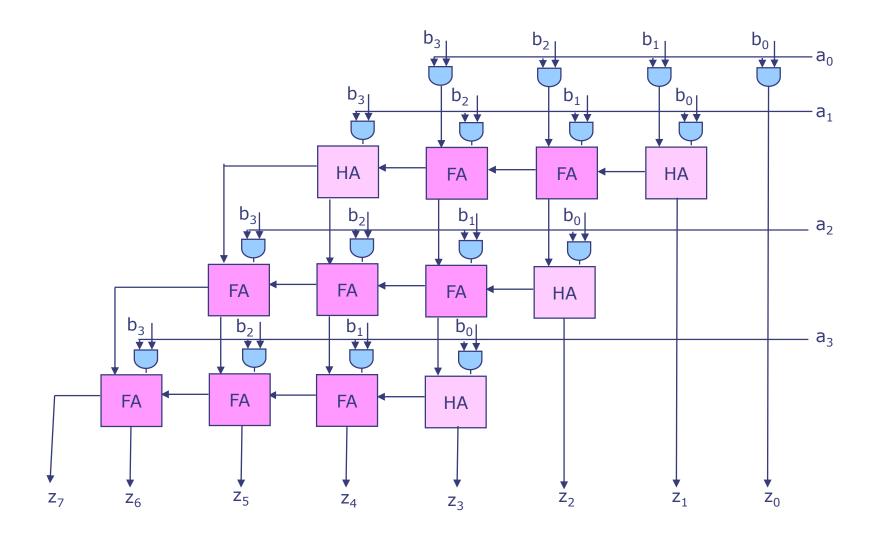


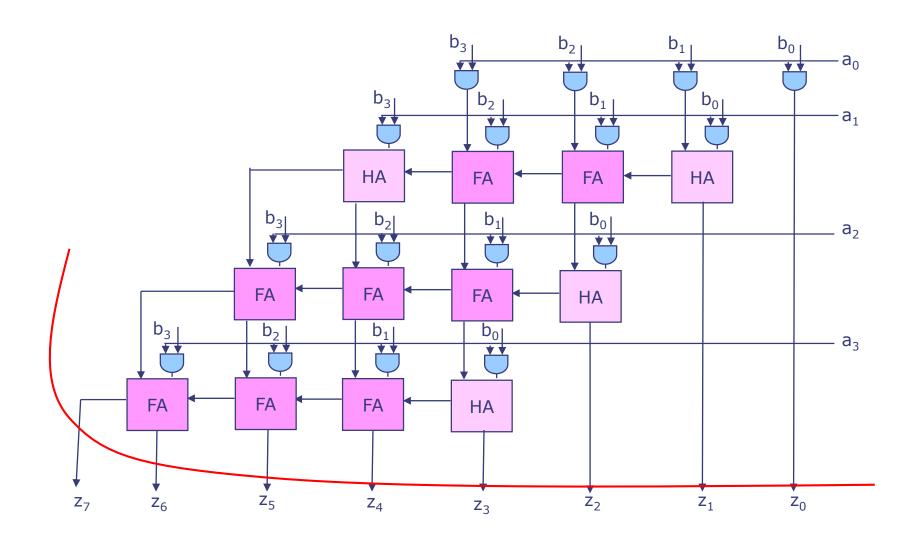


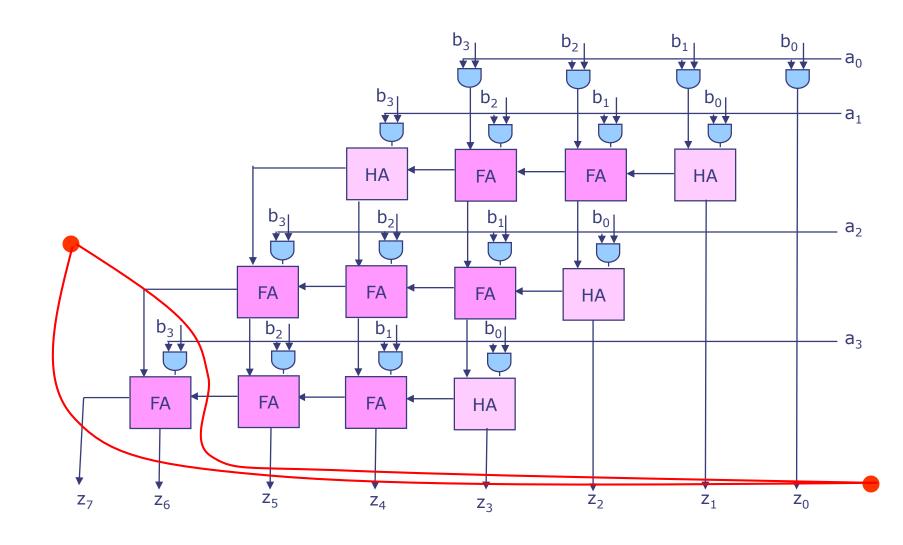
March 17, 2022

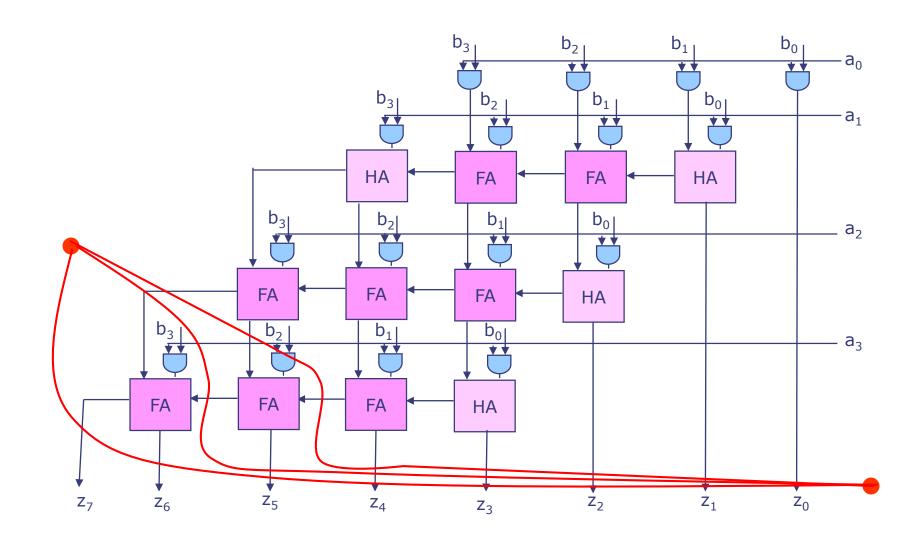


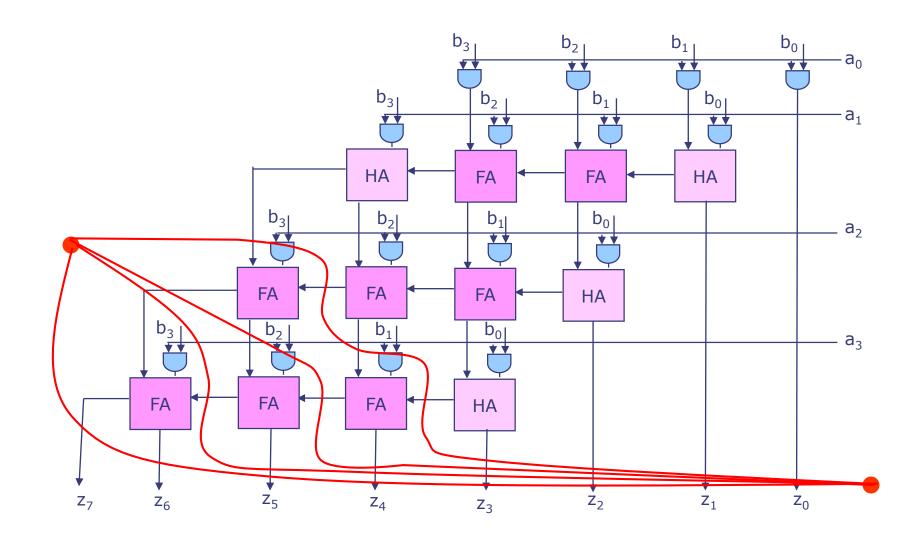
MIT 6.004 Spring 2022

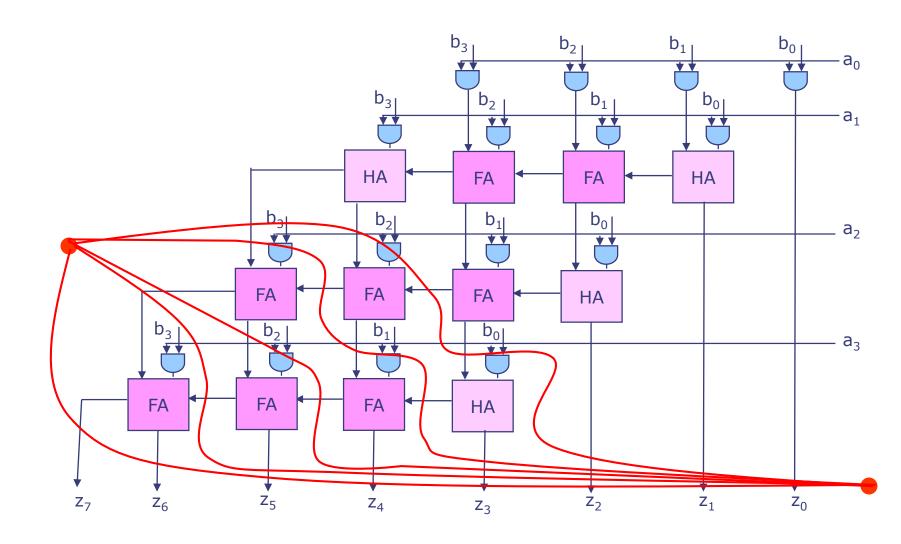


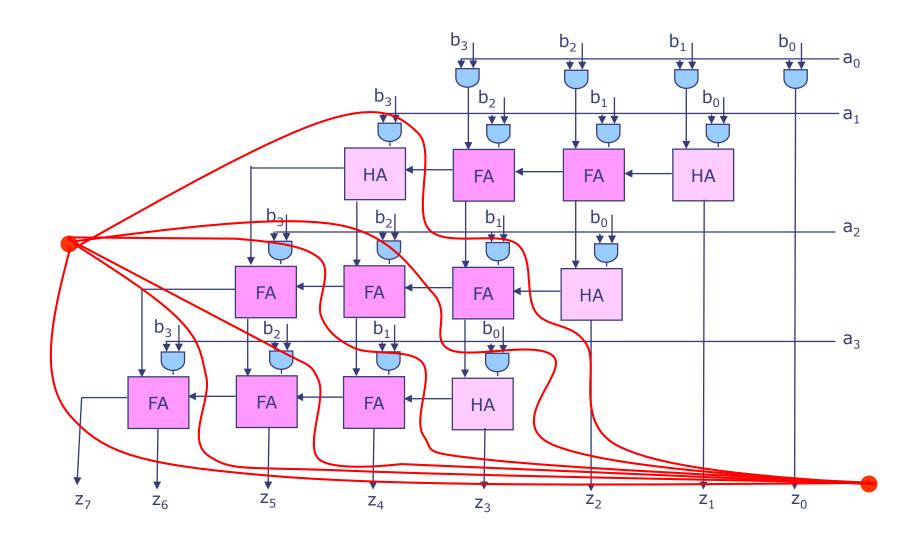


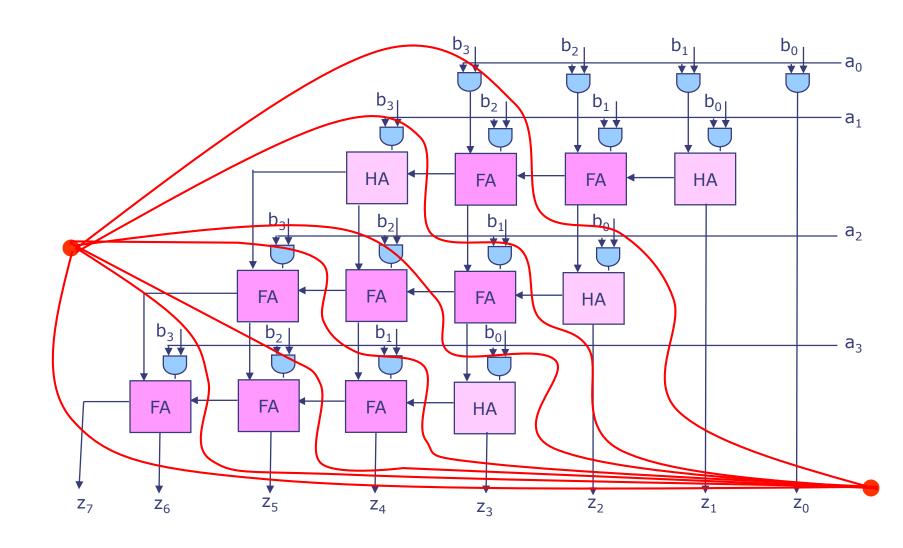


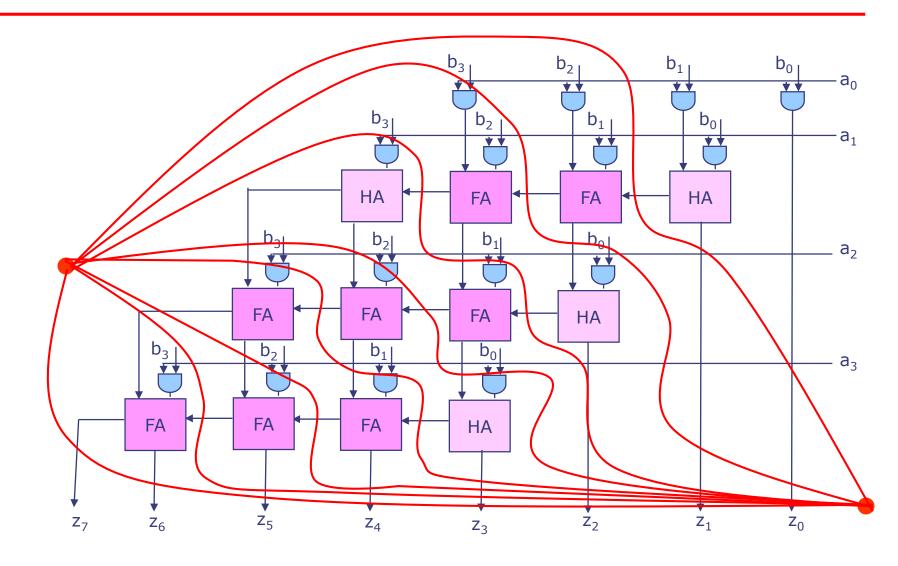


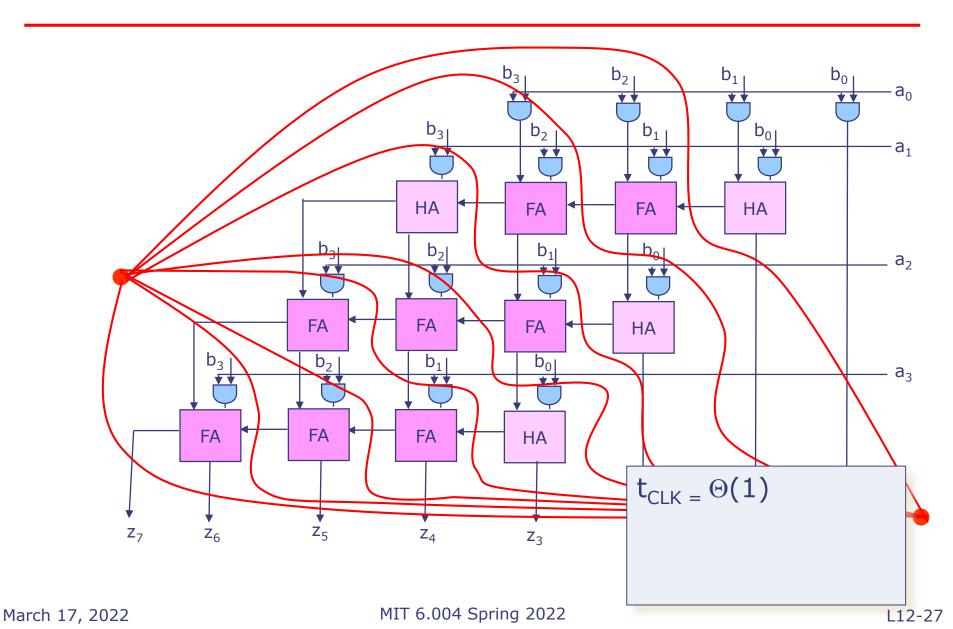


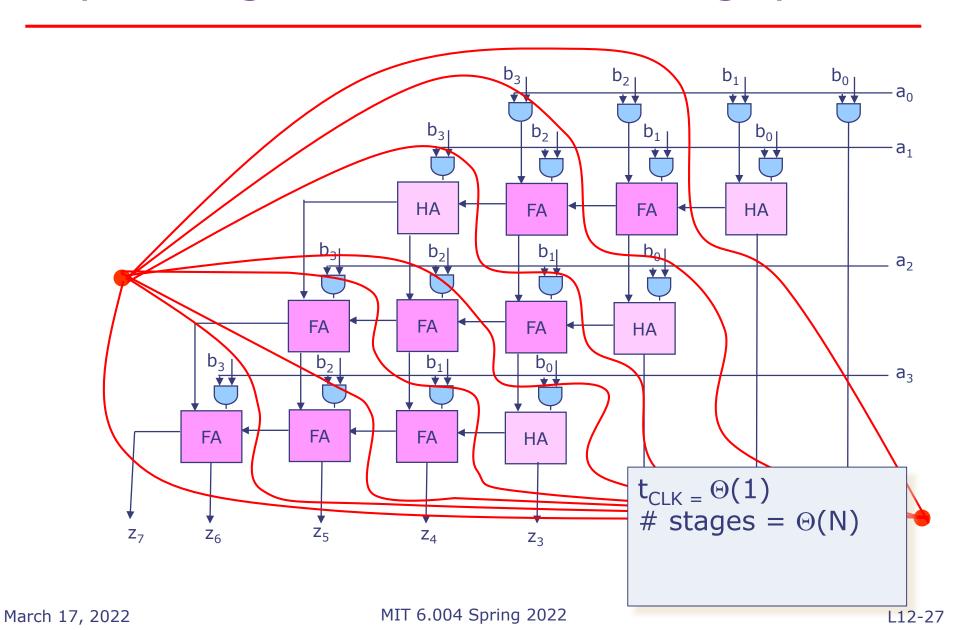


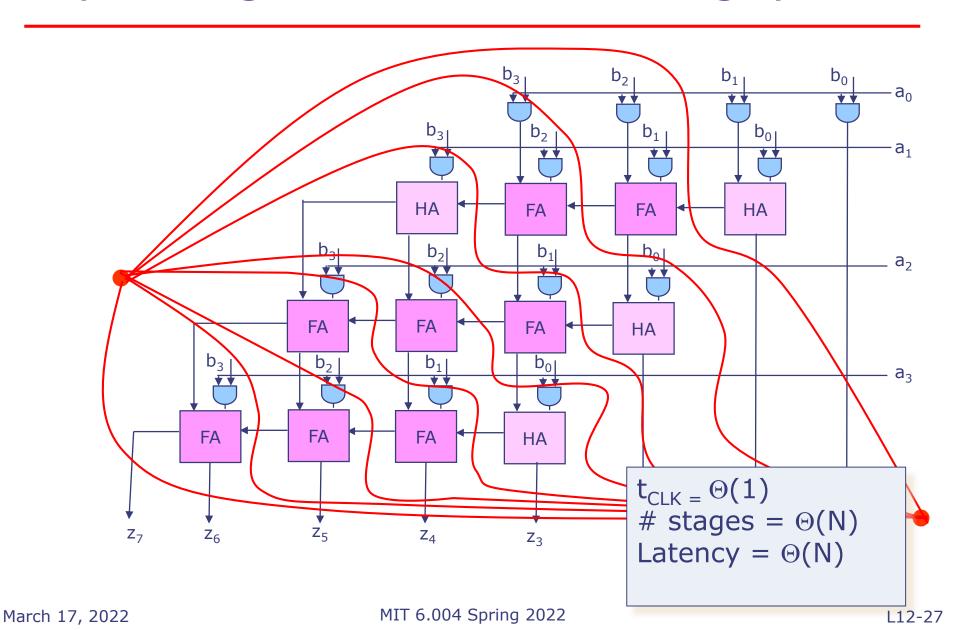


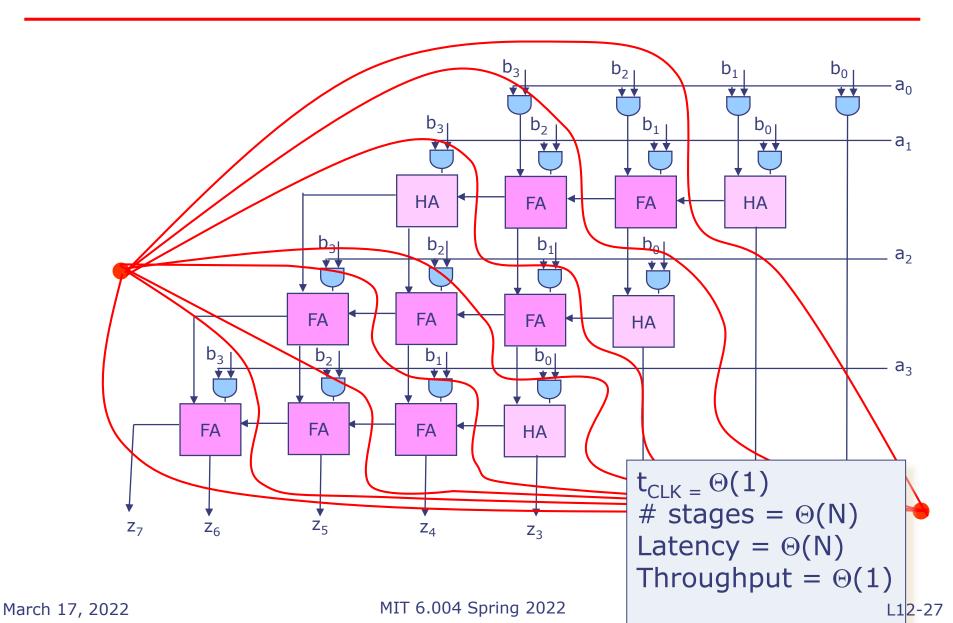








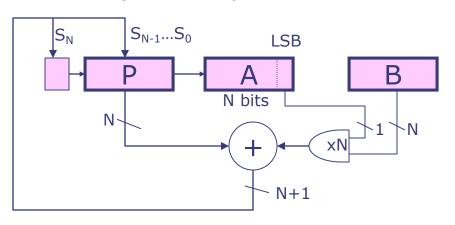




- Combinational circuits often have repetitive logic
 - Example: N-bit multiplier has N-1 adders

- Combinational circuits often have repetitive logic
 - Example: N-bit multiplier has N-1 adders
- Folded circuits use less combinational logic, reuse it over multiple cycles
 - Example: Implement multiplication with one adder, taking
 N cycles to perform the additions

- Combinational circuits often have repetitive logic
 - Example: N-bit multiplier has N-1 adders
- Folded circuits use less combinational logic, reuse it over multiple cycles
 - Example: Implement multiplication with one adder, taking
 N cycles to perform the additions

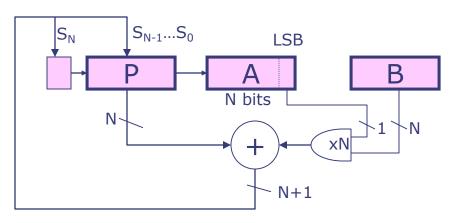


```
Init: P←0, load A&B

Repeat N times {
   P ← P + (A<sub>LSB</sub>==1 ? B : 0)
   shift S<sub>N</sub>,P,A right one bit
}

Done: 2N-bit result in P,A
```

- Combinational circuits often have repetitive logic
 - Example: N-bit multiplier has N-1 adders
- Folded circuits use less combinational logic, reuse it over multiple cycles
 - Example: Implement multiplication with one adder, taking
 N cycles to perform the additions



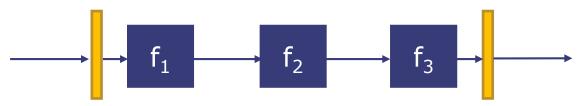
```
Init: P←0, load A&B

Repeat N times {
   P ← P + (A<sub>LSB</sub>==1 ? B : 0)
   shift S<sub>N</sub>,P,A right one bit
}

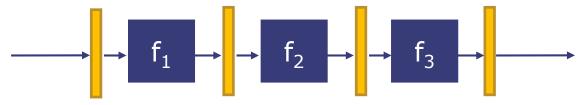
Done: 2N-bit result in P,A
```

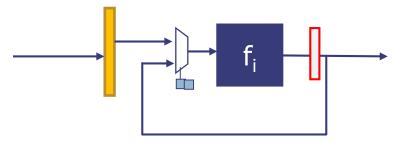
Tradeoff: reduced area, but lower throughput

Several combinational modules in one pipeline stage (A)

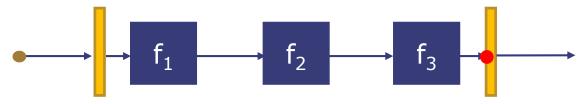


One module per pipeline stage (B)

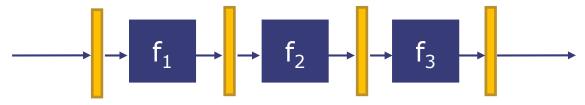


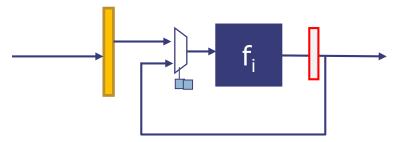


Several combinational modules in one pipeline stage (A)

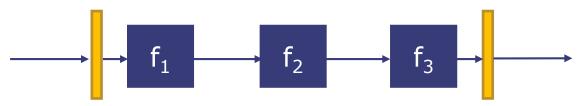


One module per pipeline stage (B)

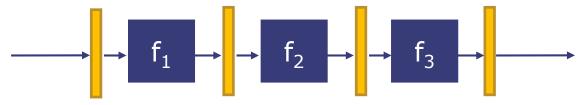


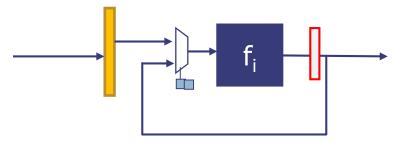


Several combinational modules in one pipeline stage (A)

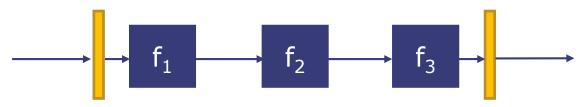


One module per pipeline stage (B)

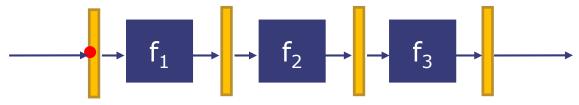


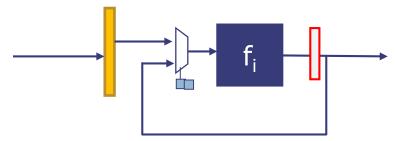


Several combinational modules in one pipeline stage (A)

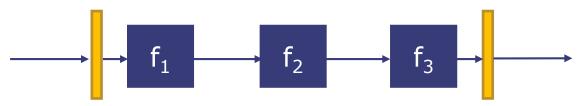


One module per pipeline stage (B)

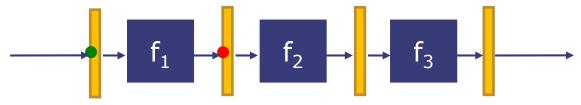


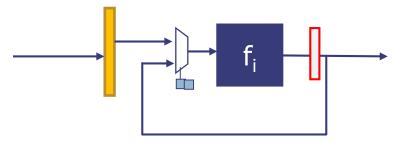


Several combinational modules in one pipeline stage (A)

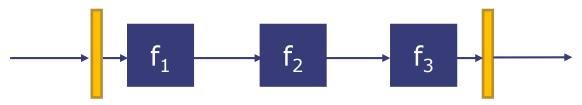


One module per pipeline stage (B)

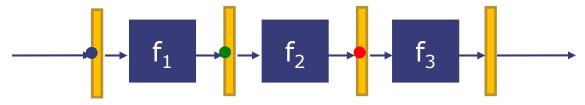


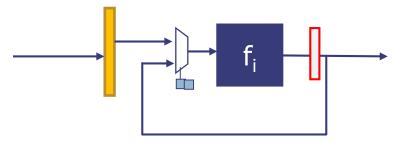


Several combinational modules in one pipeline stage (A)

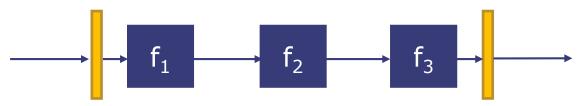


One module per pipeline stage (B)

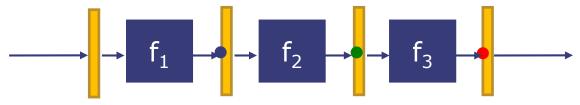


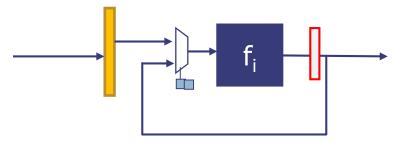


Several combinational modules in one pipeline stage (A)

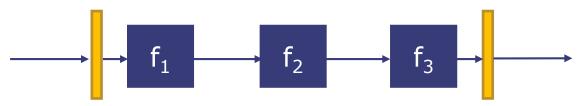


One module per pipeline stage (B)

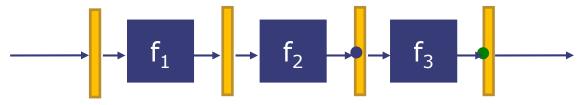


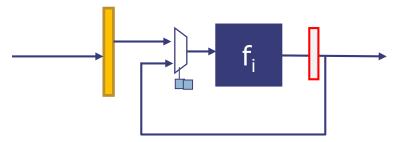


Several combinational modules in one pipeline stage (A)

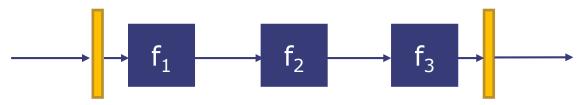


One module per pipeline stage (B)

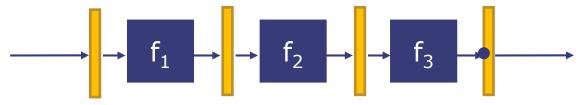


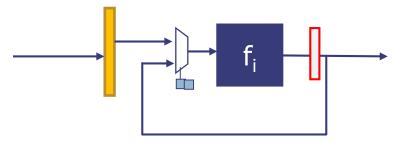


Several combinational modules in one pipeline stage (A)

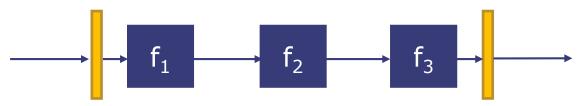


One module per pipeline stage (B)

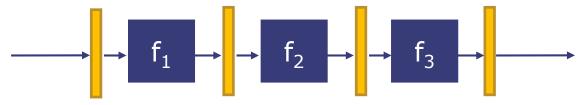


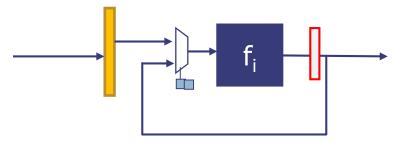


Several combinational modules in one pipeline stage (A)

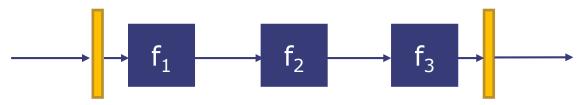


One module per pipeline stage (B)

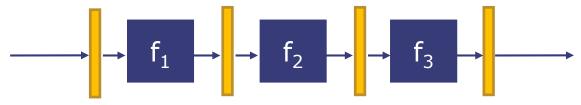


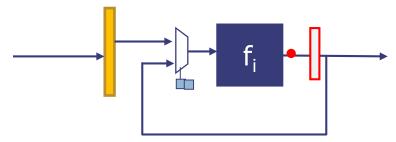


Several combinational modules in one pipeline stage (A)

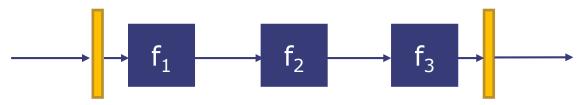


One module per pipeline stage (B)

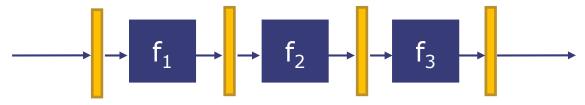


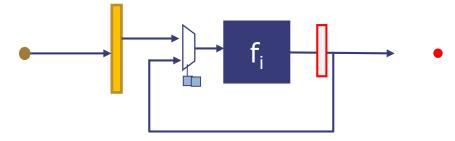


Several combinational modules in one pipeline stage (A)

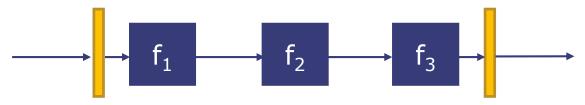


One module per pipeline stage (B)

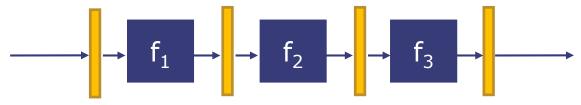


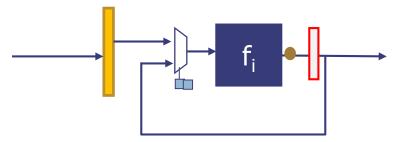


Several combinational modules in one pipeline stage (A)

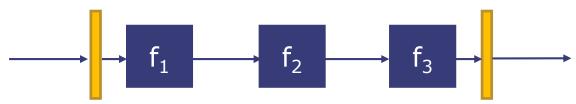


One module per pipeline stage (B)

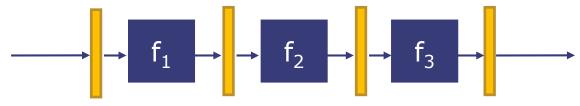




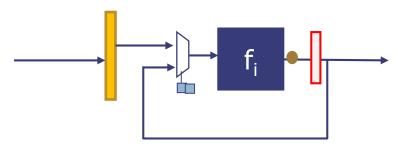
Several combinational modules in one pipeline stage (A)



One module per pipeline stage (B)



Folded reuse a block, multi-cycle (C)

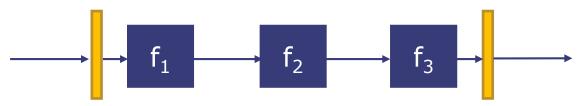


Clock?

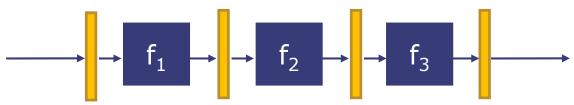
Latency?

Area?

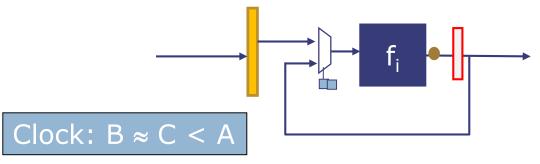
Several combinational modules in one pipeline stage (A)



One module per pipeline stage (B)



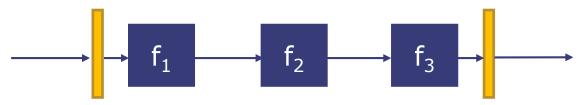
Folded reuse a block, multi-cycle (C)



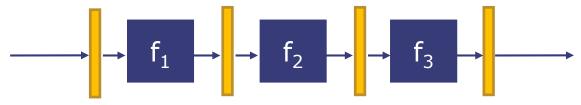
Latency?

Area?

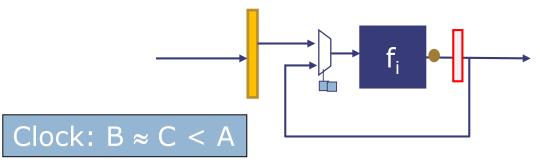
Several combinational modules in one pipeline stage (A)



One module per pipeline stage (B)



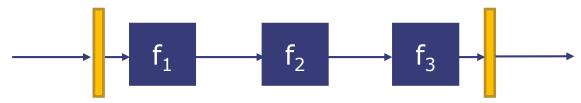
Folded reuse a block, multi-cycle (C)



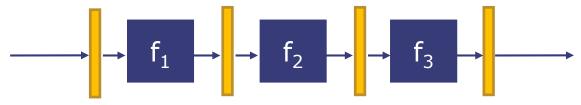
Latency?

Area: C < A < B

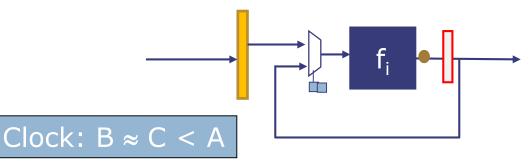
Several combinational modules in one pipeline stage (A)



One module per pipeline stage (B)



Folded reuse a block, multi-cycle (C)

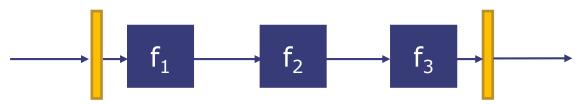


Area: C < A < B

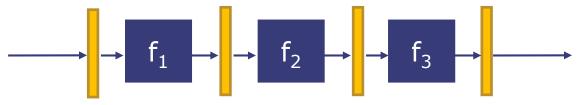
Latency: A < B < C

L12-29

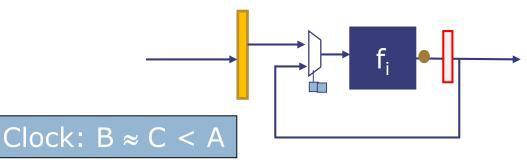
Several combinational modules in one pipeline stage (A)



One module per pipeline stage (B)



Folded reuse a block, multi-cycle (C)



Latency: A < B < C

Throughput: C < A < B

Area: C < A < B

MIT 6.004 Spring 2022

Thank You!

Next Lecture:
Design Tradeoffs in Sequential Circuits