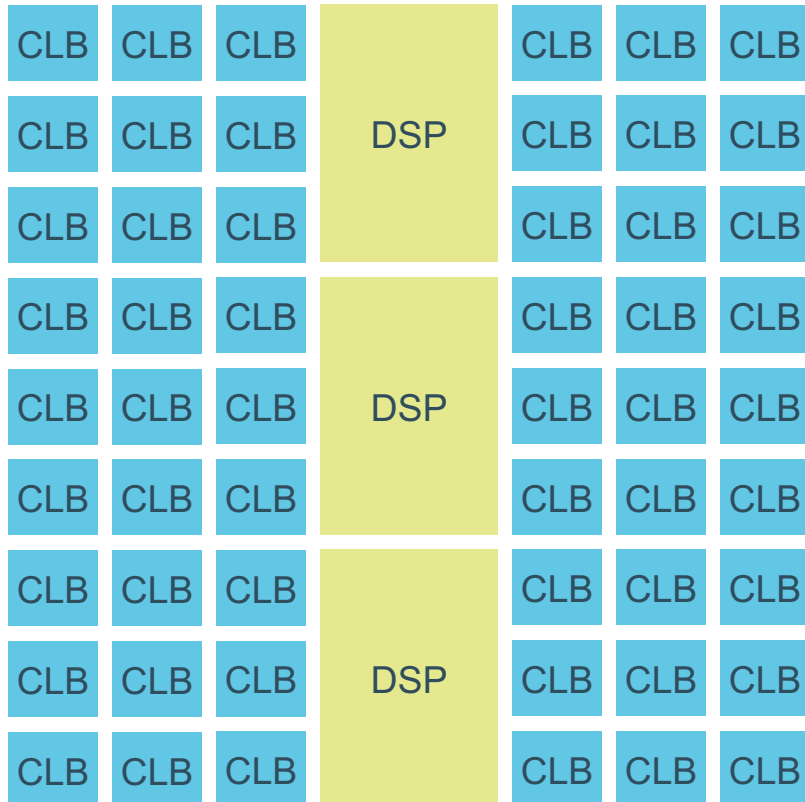


# Area-Efficient LUT-Based Multipliers for AMD Versal FPGAs

Zetao Miao, Xander Pottier, Jonas Bertels, Wouter Legiest and Ingrid Verbauwhede

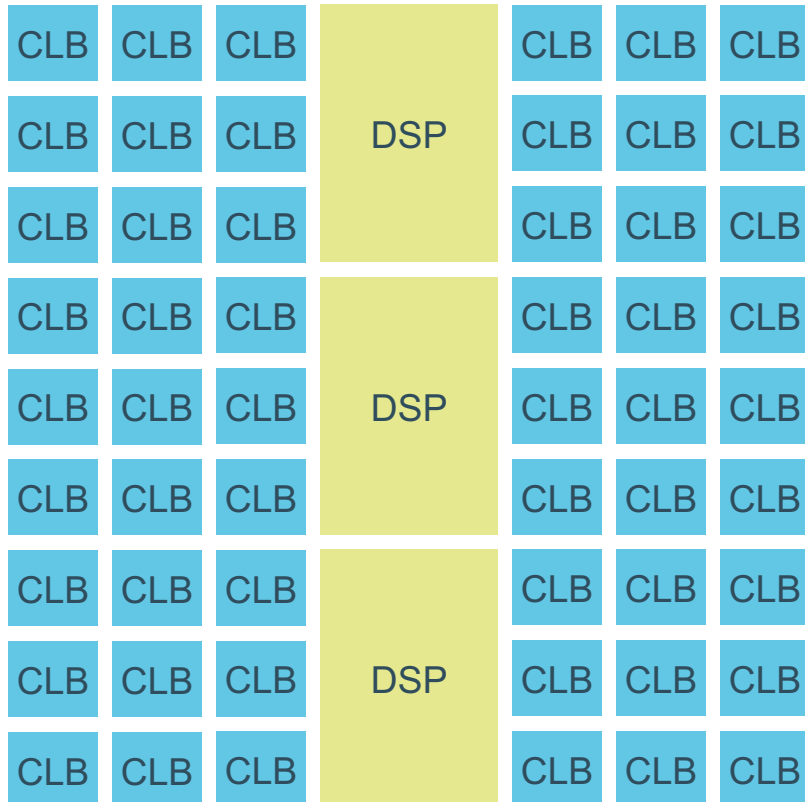
*COSIC, KU Leuven, Belgium*

# Integer Multipliers on FPGAs





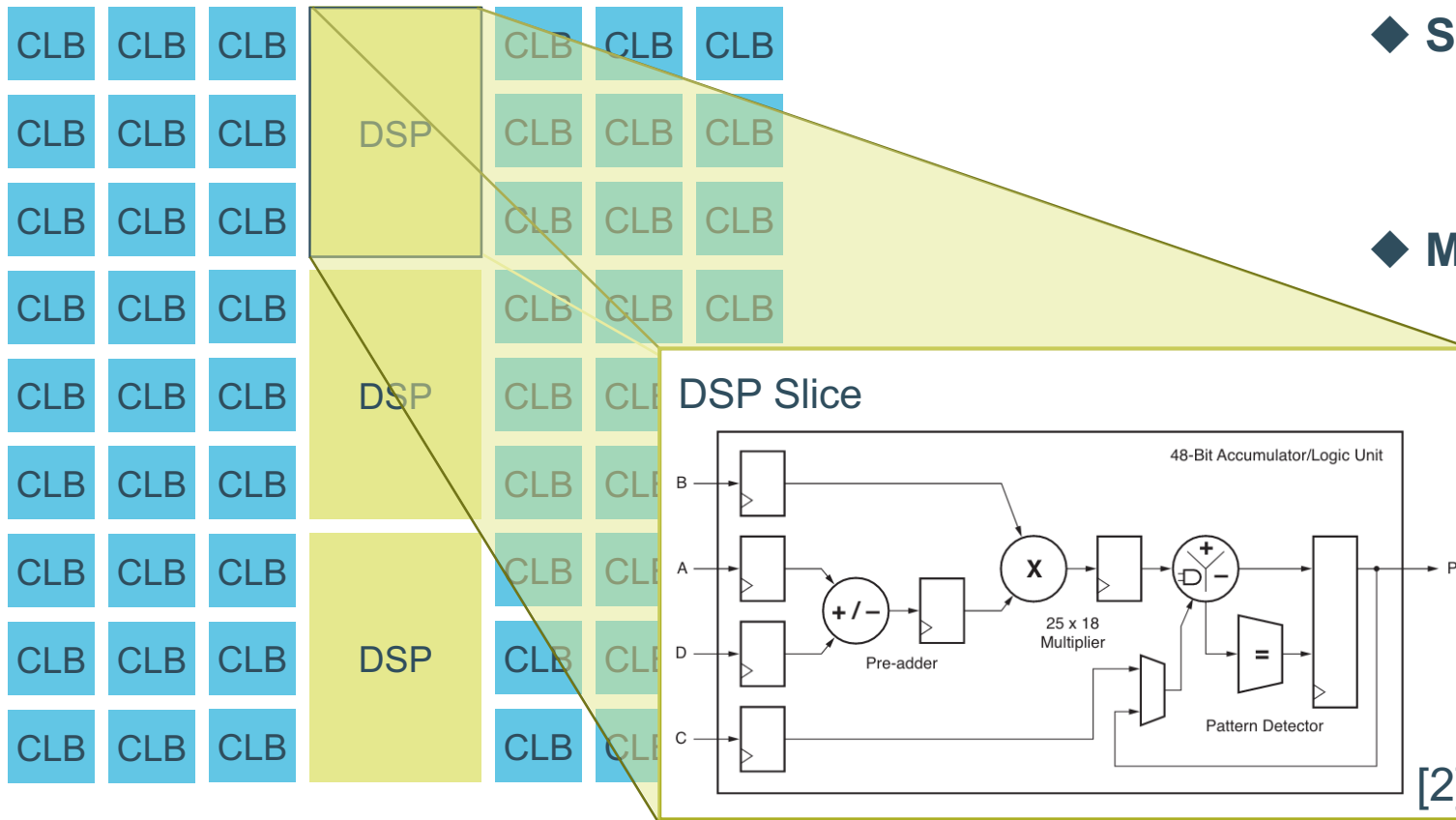
# Integer Multipliers on FPGAs



## ◆ Small Multipliers

LUTs & Carry Logic

# Integer Multipliers on FPGAs



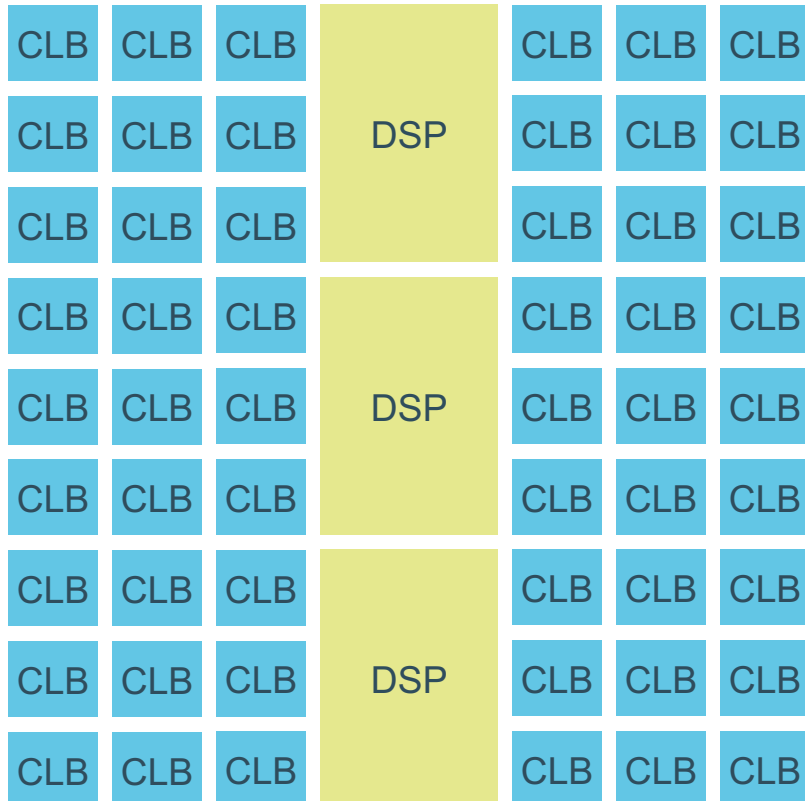
## ◆ Small Multipliers

LUTs & Carry Logic

## ◆ Medium Multipliers

DSP blocks

# Integer Multipliers on FPGAs



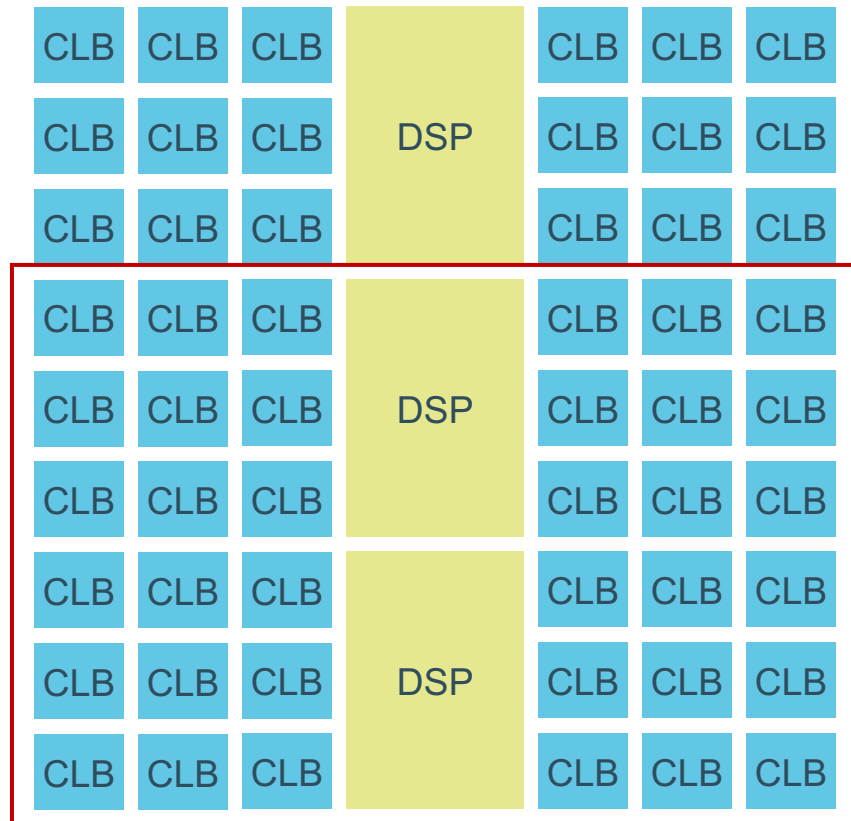
## ◆ Small Multipliers

LUTs & Carry Logic

## ◆ Medium Multipliers

DSP blocks

# Integer Multipliers on FPGAs



## ◆ Small Multipliers

LUTs & Carry Logic

## ◆ Medium Multipliers

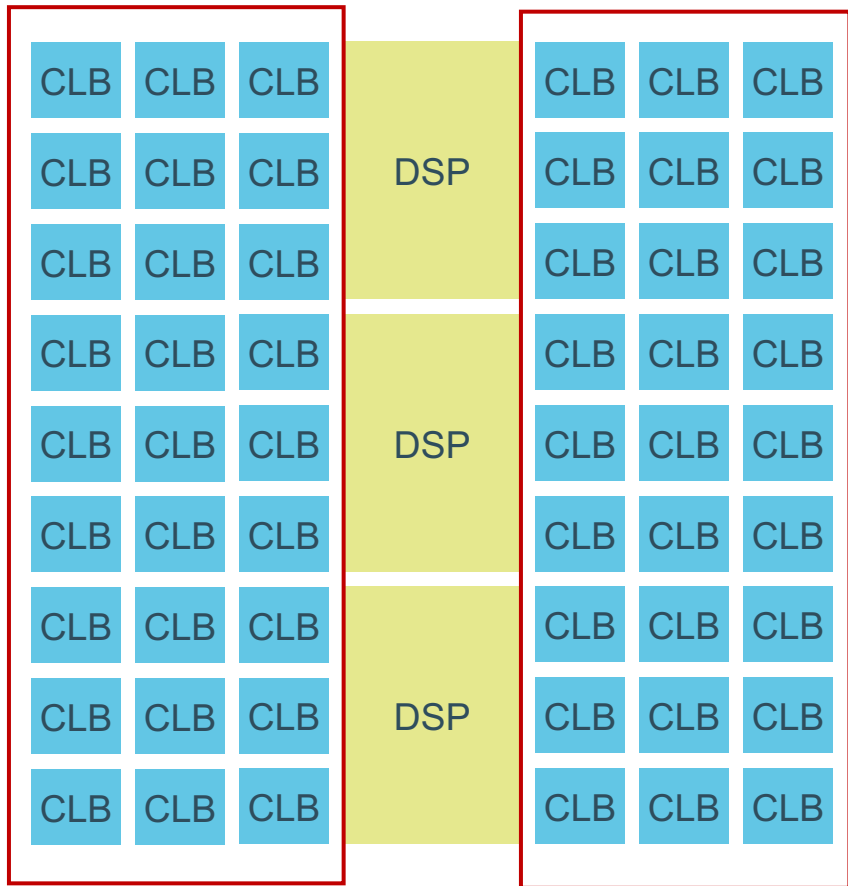
DSP blocks

## ◆ Large Multipliers

DSP blocks

LUTs & Carry Logic

# Integer Multipliers on FPGAs



This Work

## ◆ Small Multipliers

LUTs & Carry Logic

## ◆ Medium Multipliers

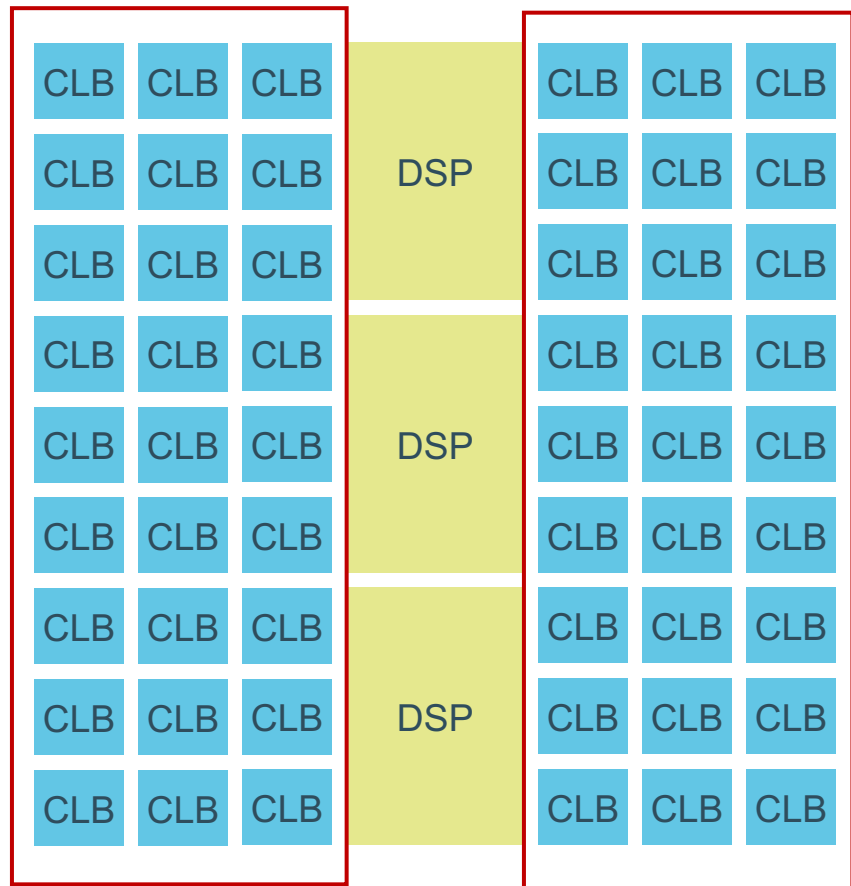
DSP blocks

## ◆ Large Multipliers

DSP blocks

LUTs & Carry Logic

# Integer Multipliers on FPGAs



This Work

## ◆ Small Multipliers

LUTs & Carry Logic

## ◆ Medium Multipliers

DSP blocks

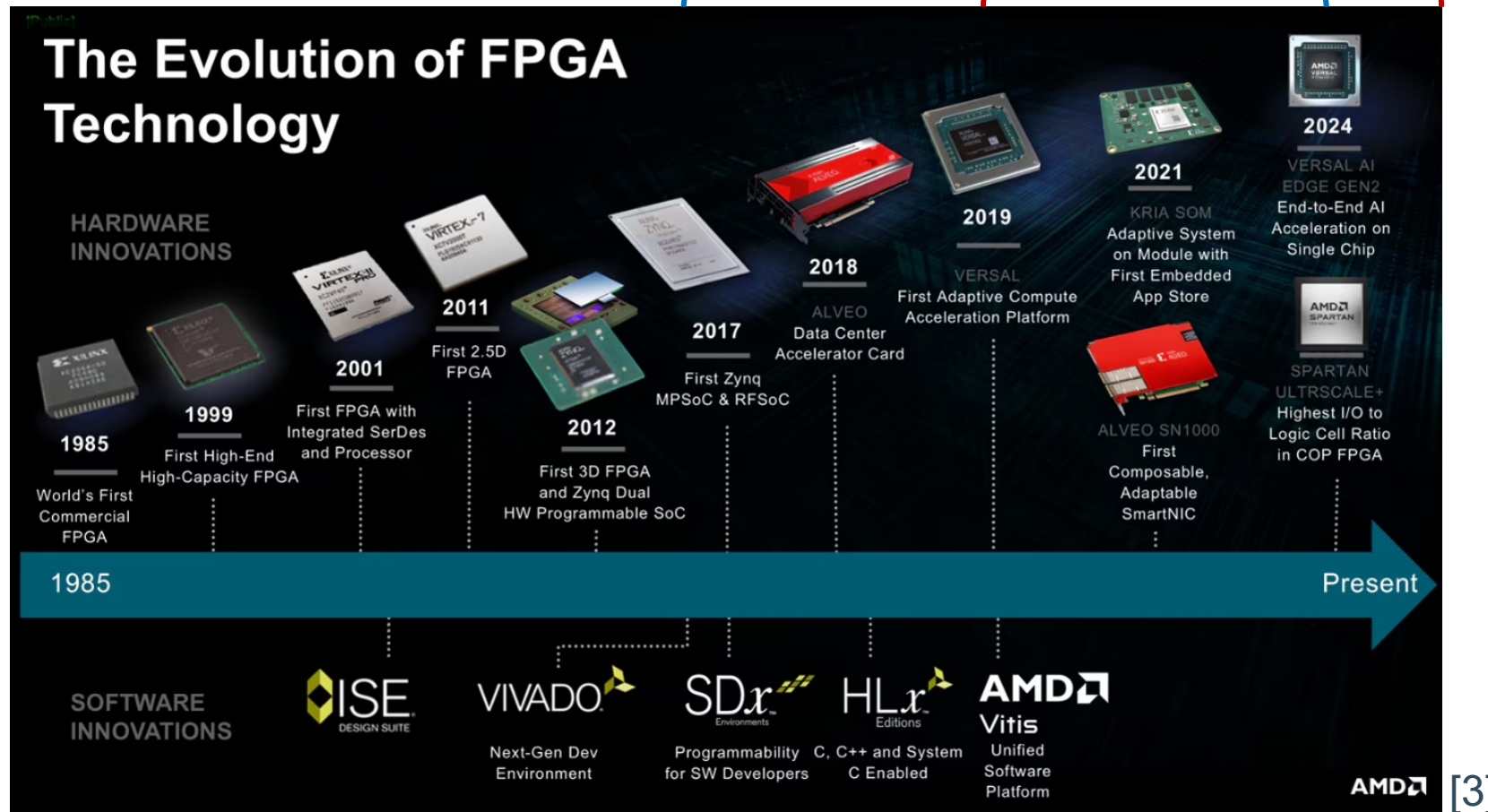
## ◆ Large Multipliers

DSP blocks

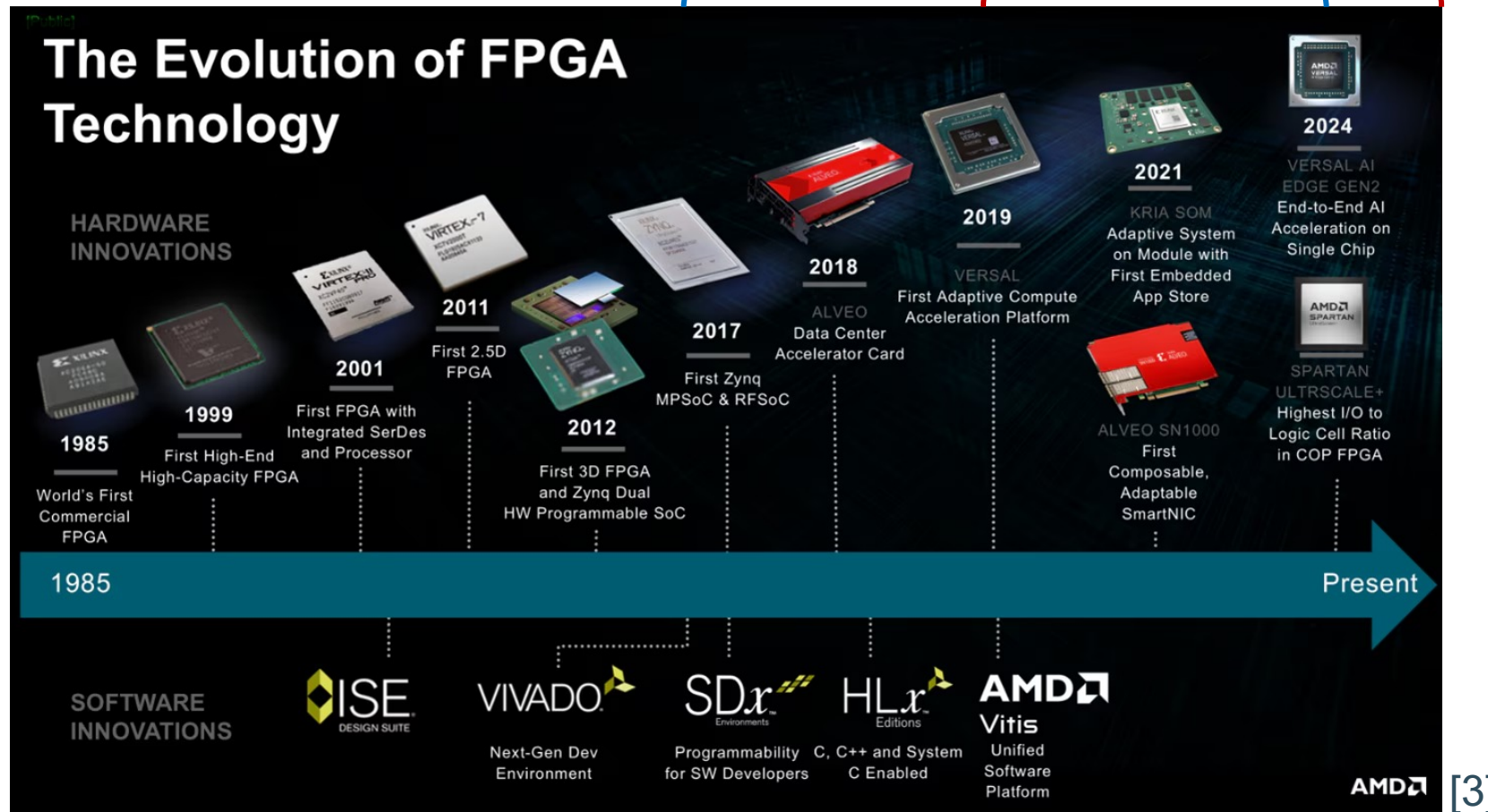
LUTs & Carry Logic

**Do more with less => increase logic density on FPGA primitives**

# AMD FPGA: from *UltraScale* to *Versal*



# AMD FPGA: from *UltraScale* to *Versal*



Micro-architecture changes

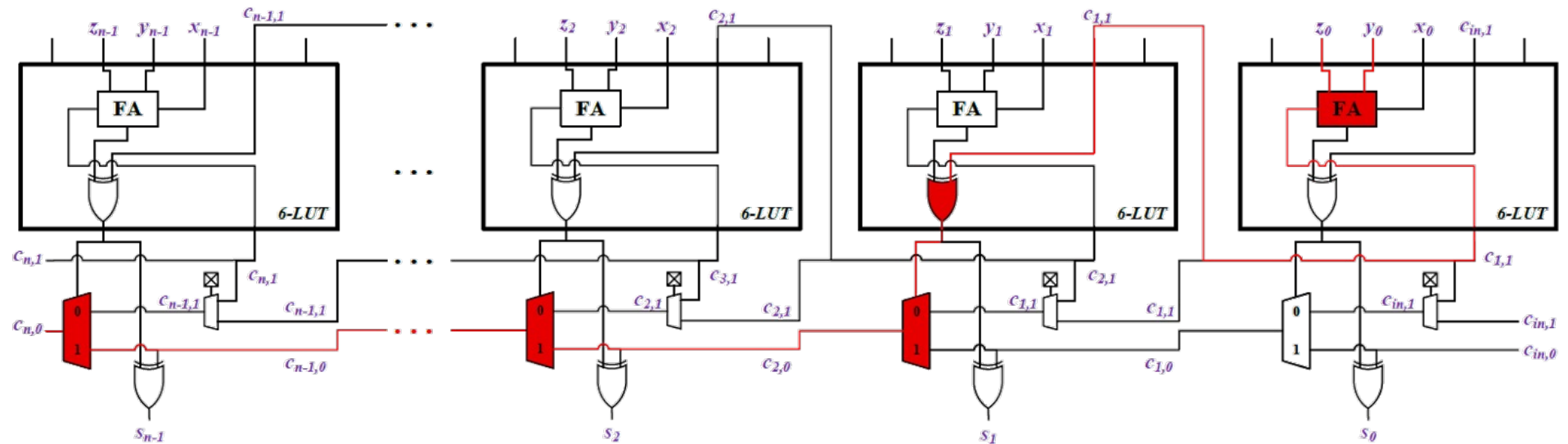
Look-up Tables  
Carry Logic  
DSPs  
...



Implementation changes

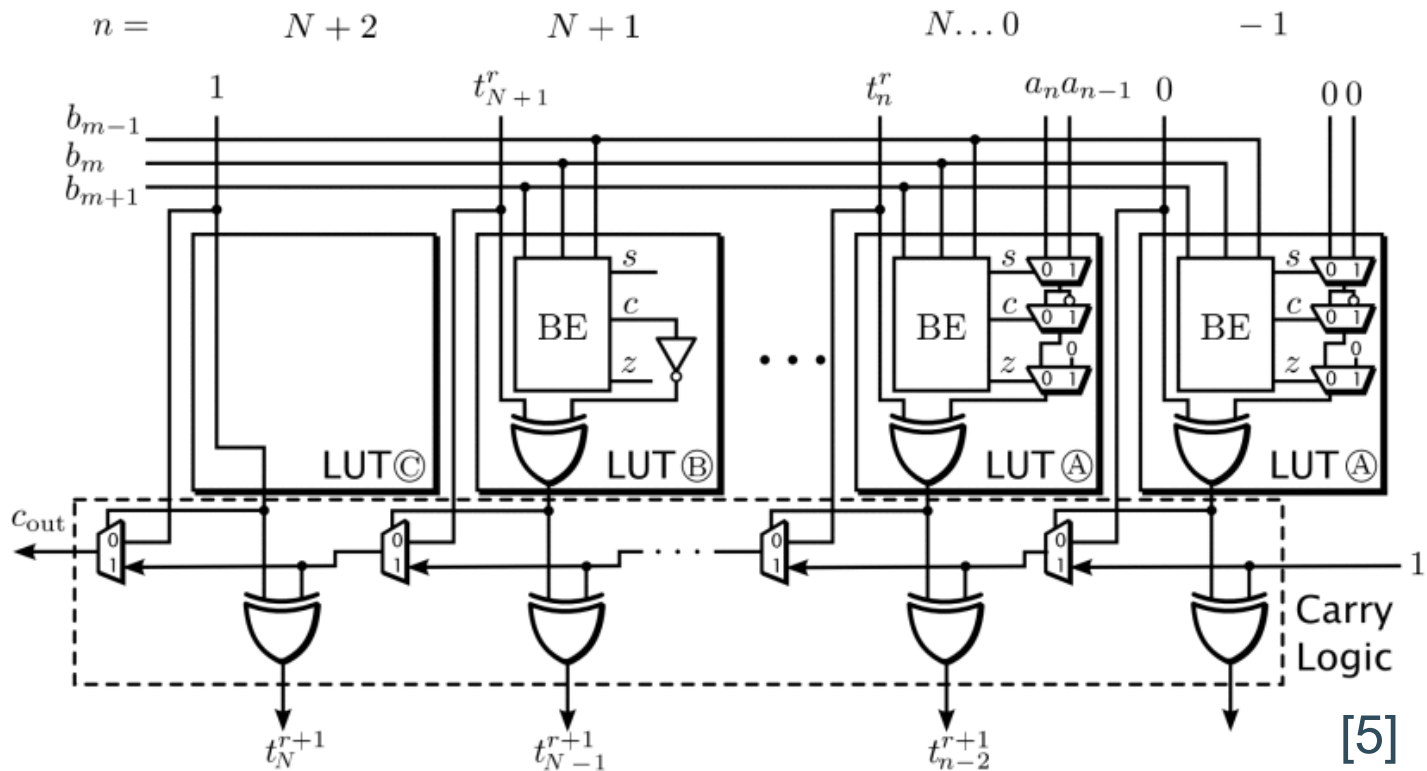
[3]

# Ternary Adder [4] on *UltraScale*

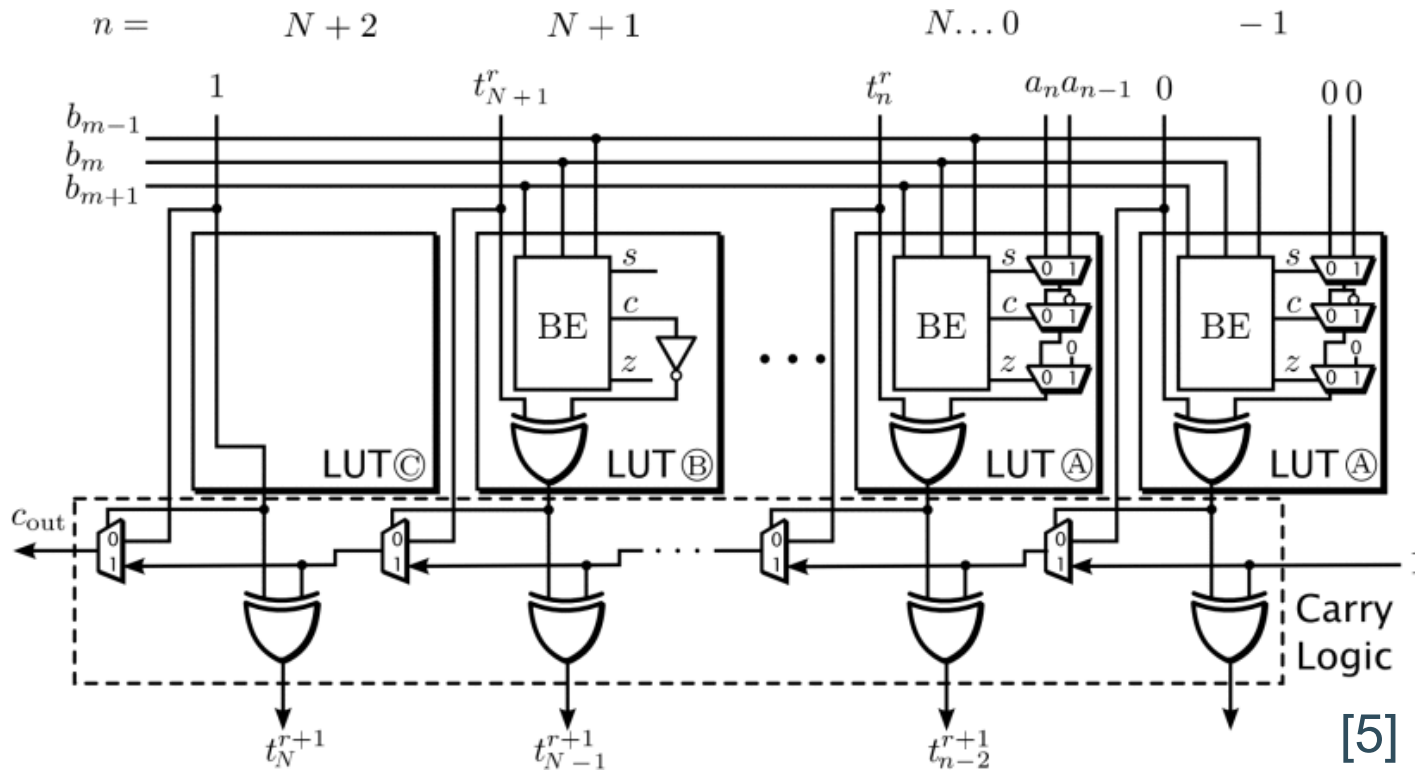




# LUT-Based Multiplier on *UltraScale*



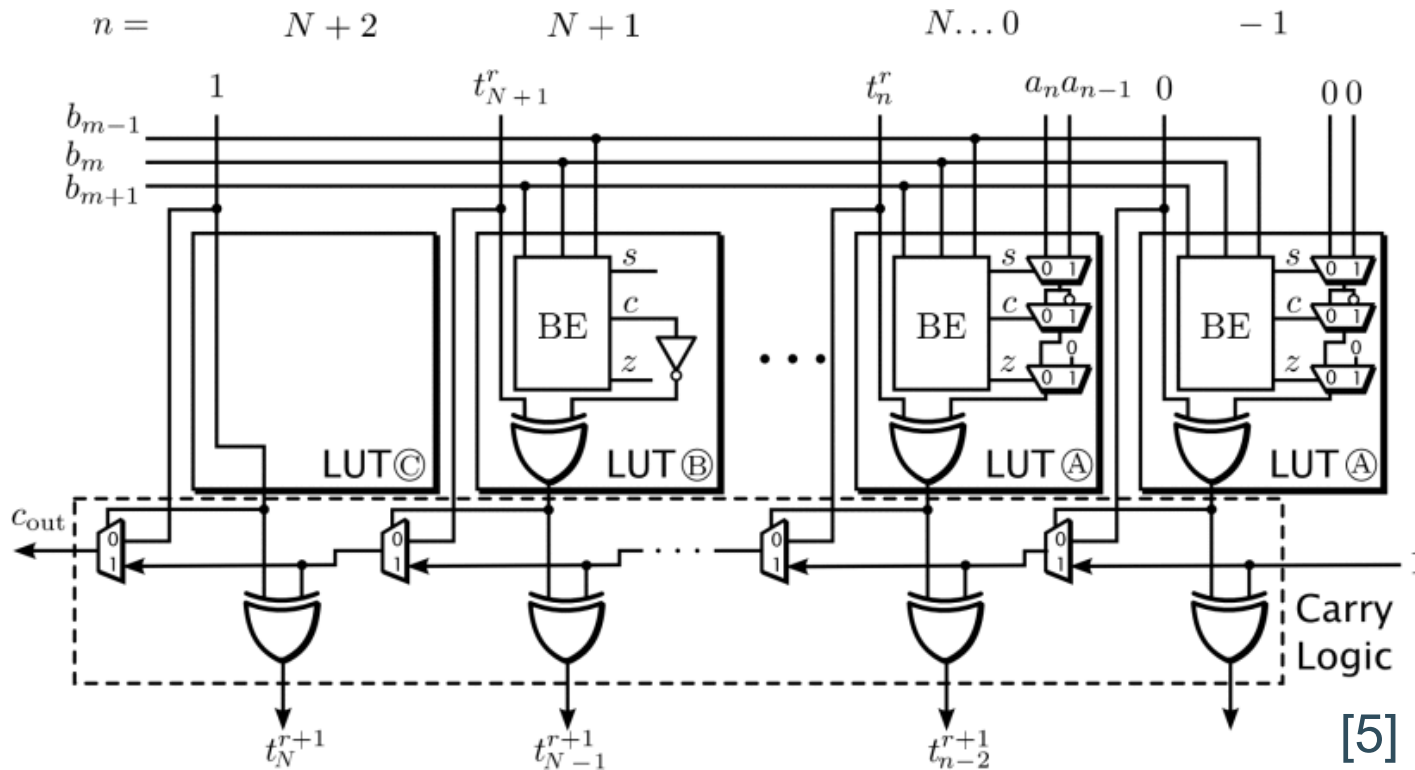
# LUT-Based Multiplier on *UltraScale*



Such an 8-bit multiplier:

Platform	LUT usage
<i>UltraScale</i>	36
<i>Versal</i>	76

# LUT-Based Multiplier on *UltraScale*



Such an 8-bit multiplier:

Platform	LUT usage
<i>UltraScale</i>	36
<i>Versal</i>	76

Naïve \* inferred by Vivado:

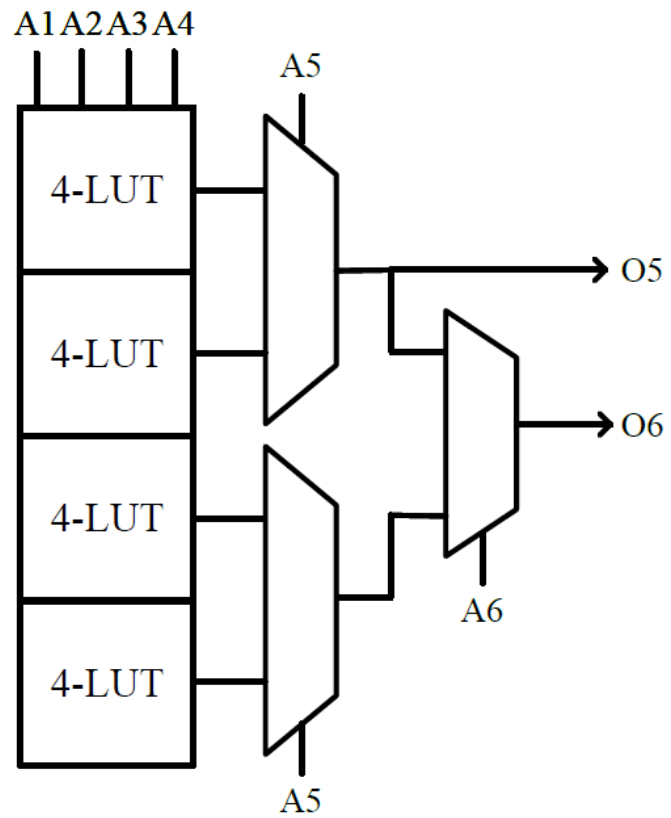
Platform	LUT usage
<i>UltraScale</i>	69
<i>Versal</i>	73



*Mapped poorly on Versal CLBs*

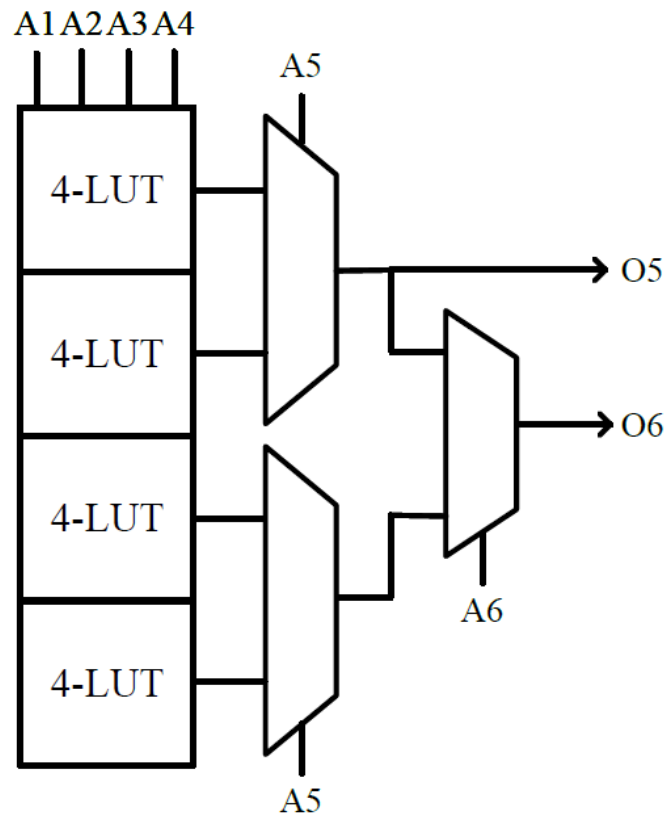
# Look-up Tables (LUTs)

UltraScale



# Look-up Tables (LUTs)

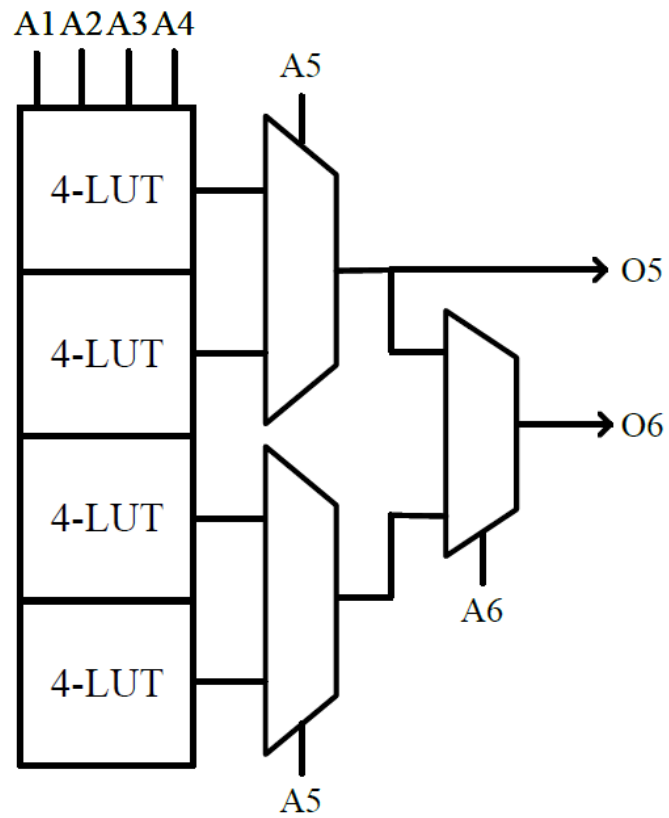
UltraScale



- 6-LUT mode:  
 $f(A_1, A_2, A_3, A_4, A_5, A_6)$

# Look-up Tables (LUTs)

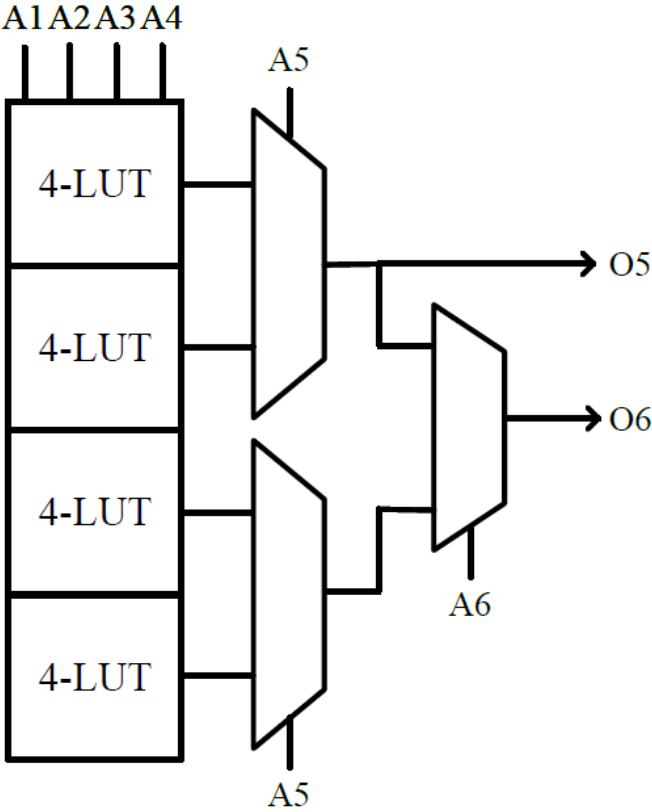
UltraScale



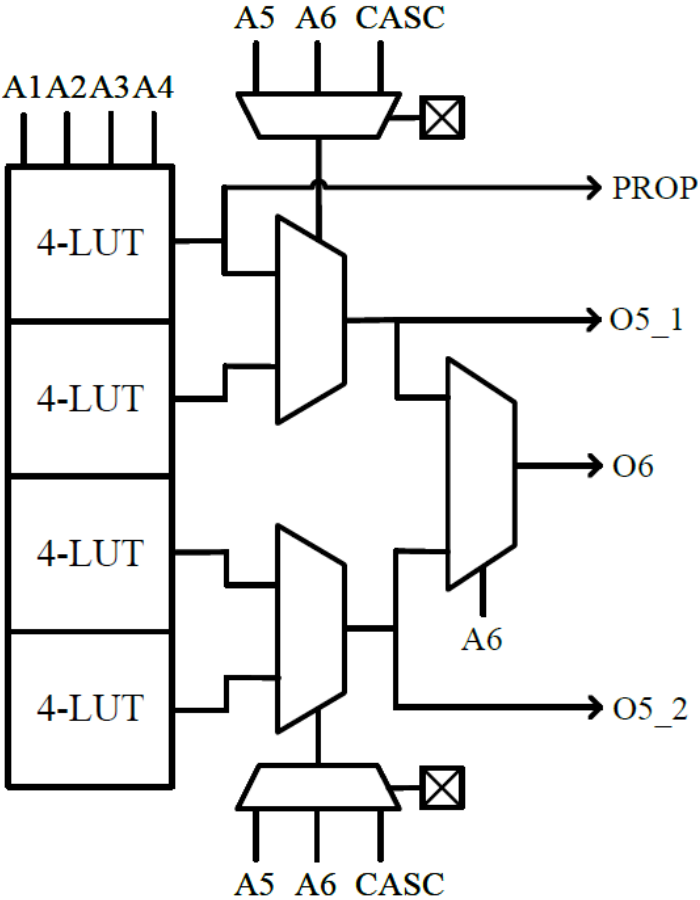
- 6-LUT mode:  
 $f(A_1, A_2, A_3, A_4, A_5, A_6)$
- Dual 5-LUT mode:  
 $f(A_1, A_2, A_3, A_4, A_5)$   
 $g(A_1, A_2, A_3, A_4, A_5)$

# Look-up Tables (LUTs)

UltraScale

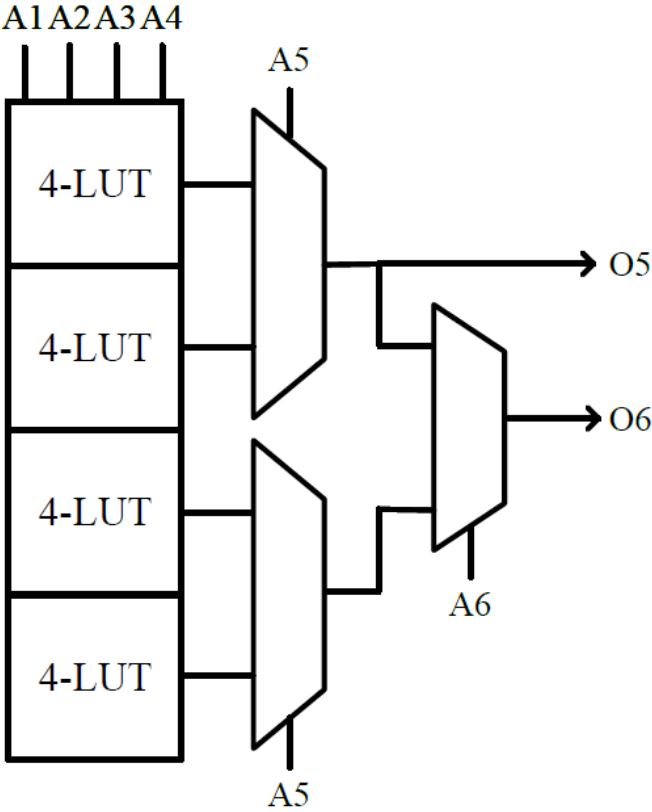


Versal

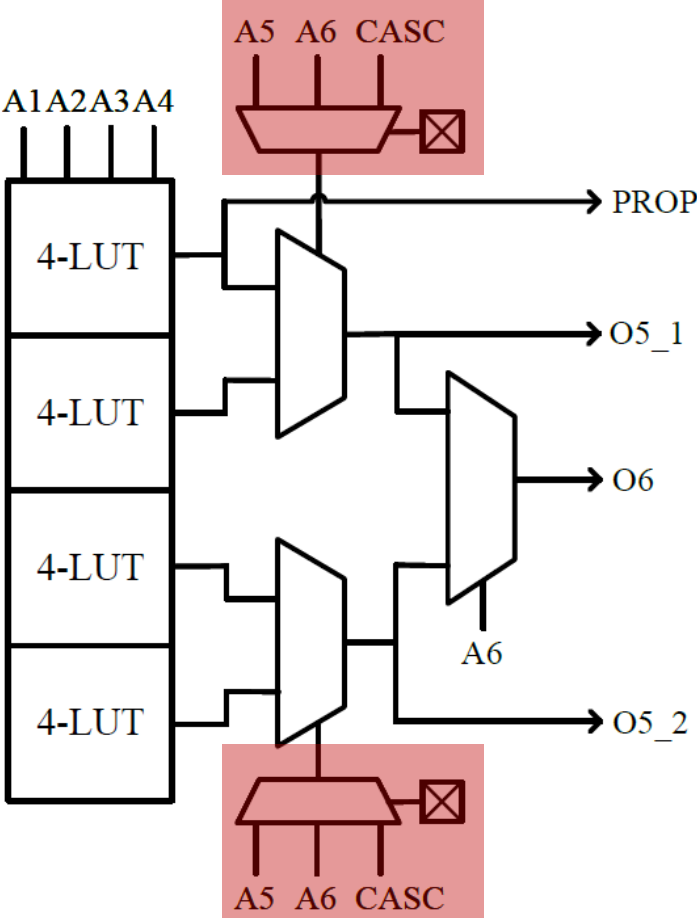


# Look-up Tables (LUTs)

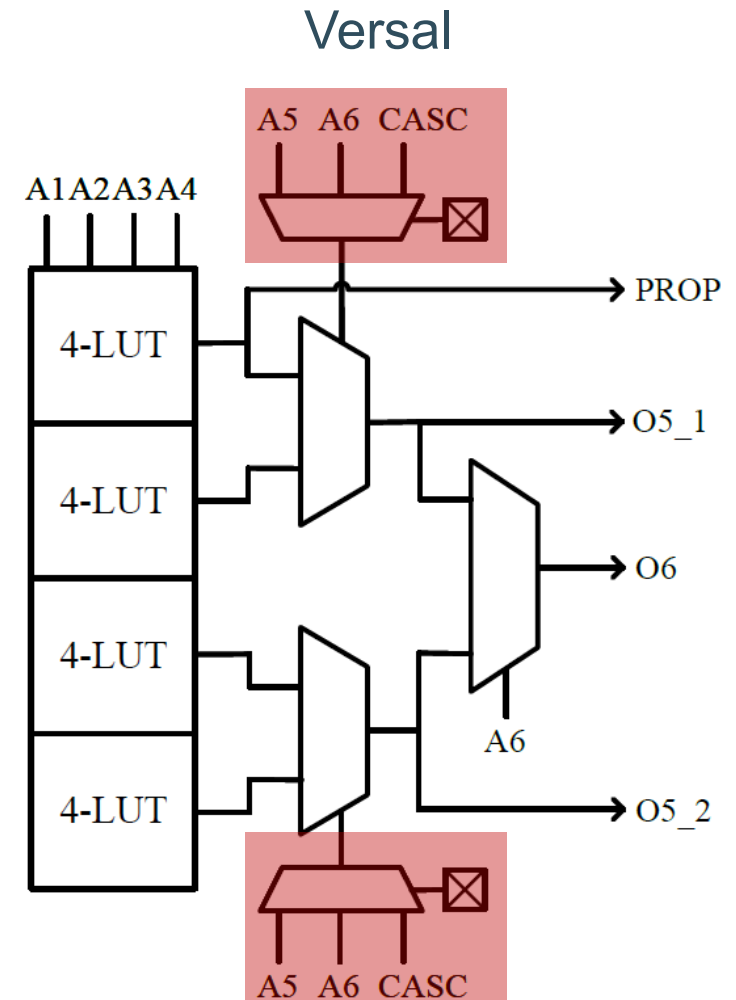
UltraScale



Versal



# Look-up Tables (LUTs)



# Look-up Tables (LUTs)

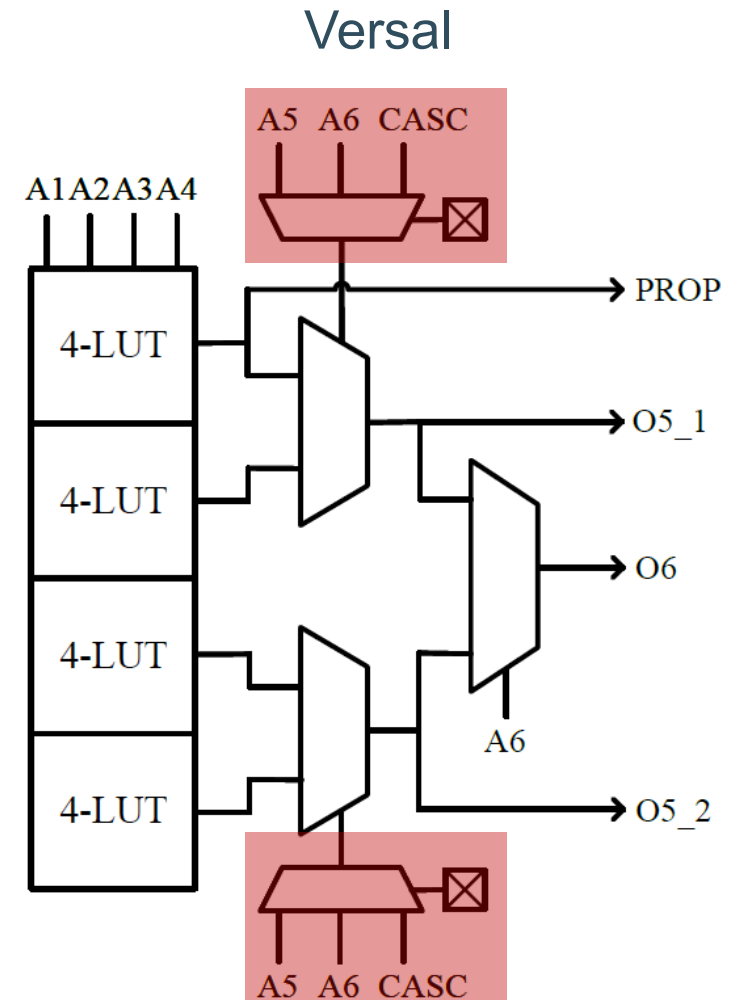
- 6-LUT mode:

$$f(A_1, A_2, A_3, A_4, A_5, A_6)$$

- Dual 5-LUT mode:

$$f(A_1, A_2, A_3, A_4, A_5)$$

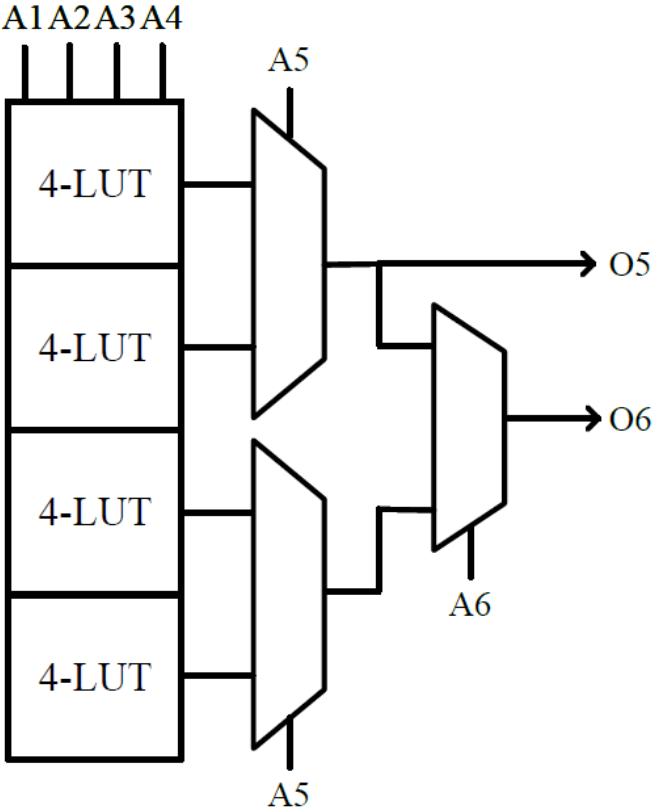
$$g(A_1, A_2, A_3, A_4, A_5)$$



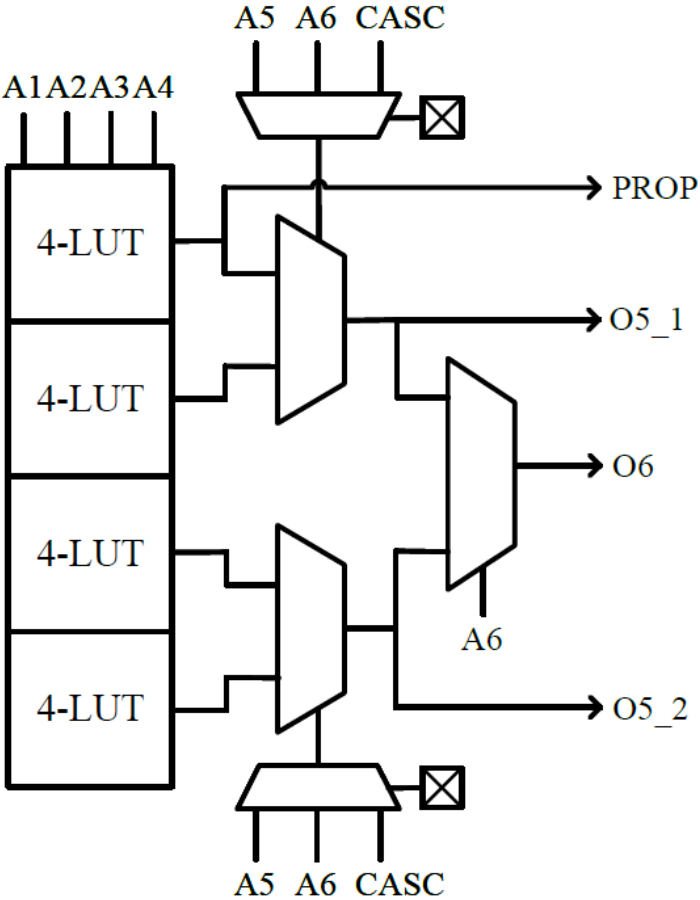


# Look-up Tables (LUTs)

UltraScale

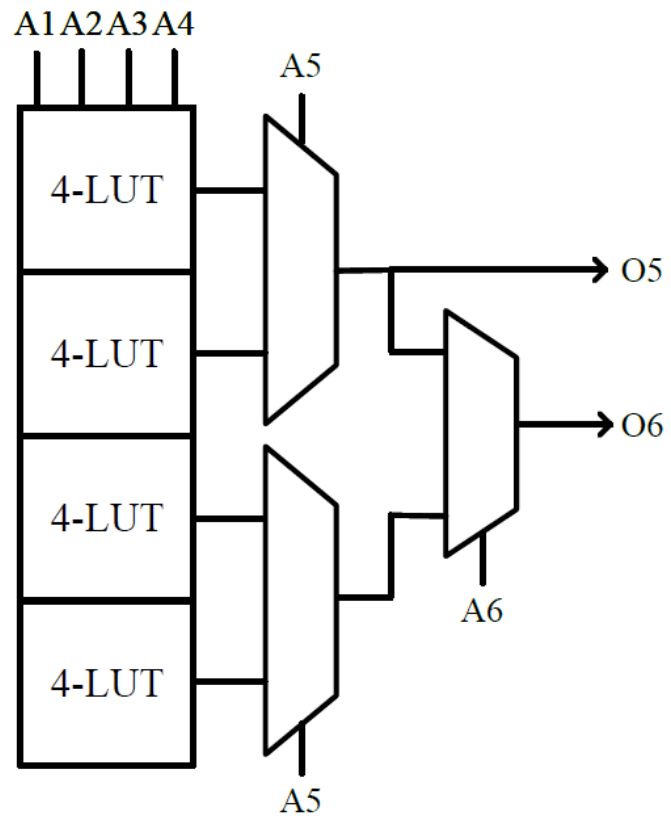


Versal

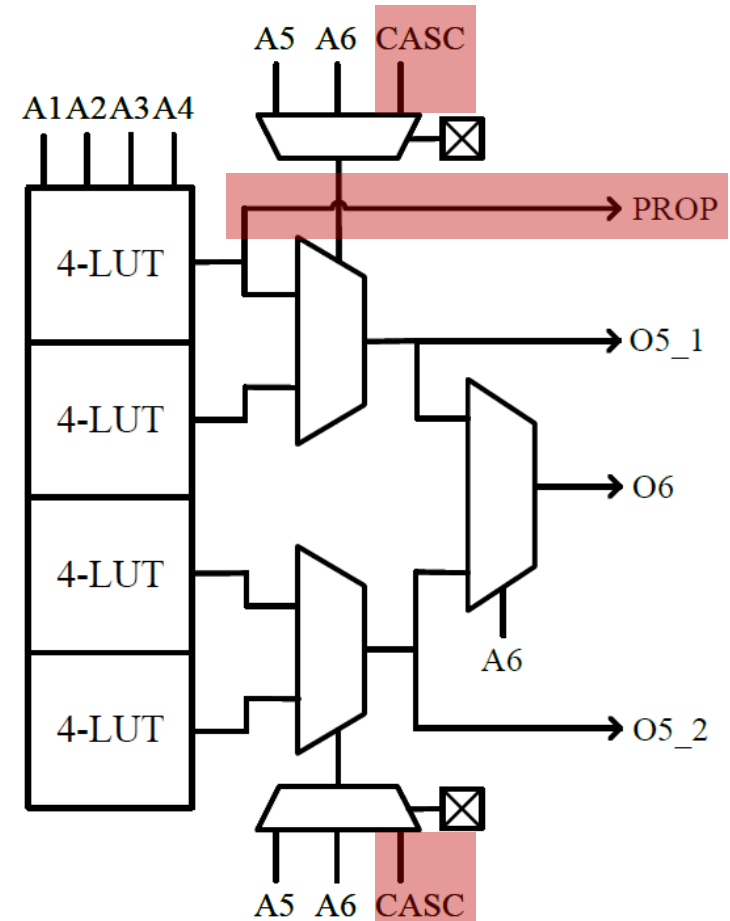


# Look-up Tables (LUTs)

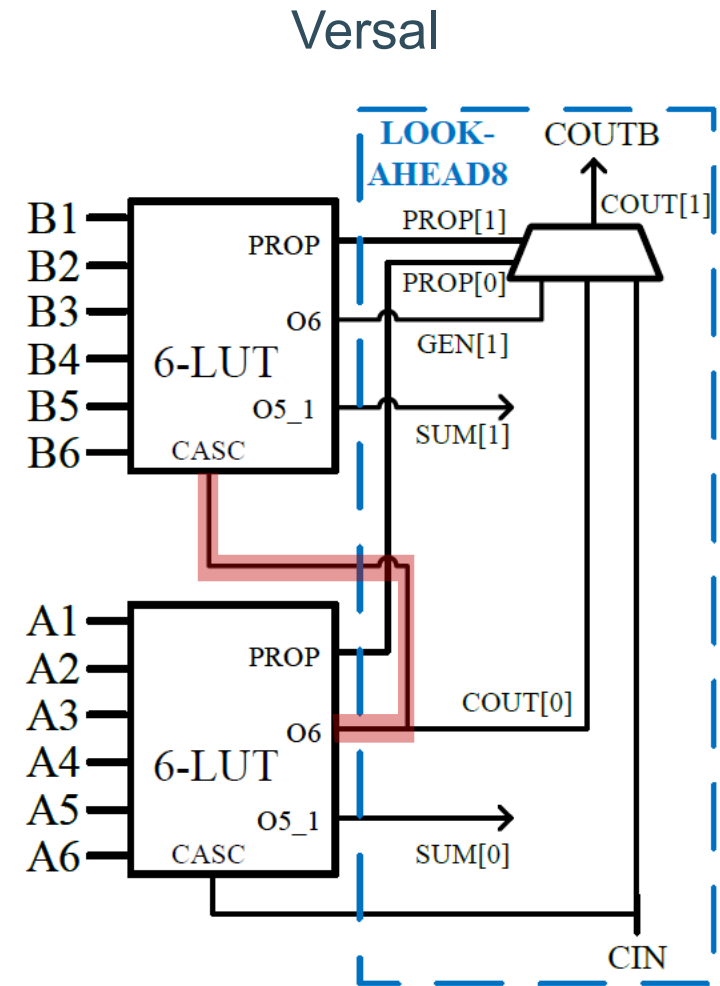
UltraScale



Versal

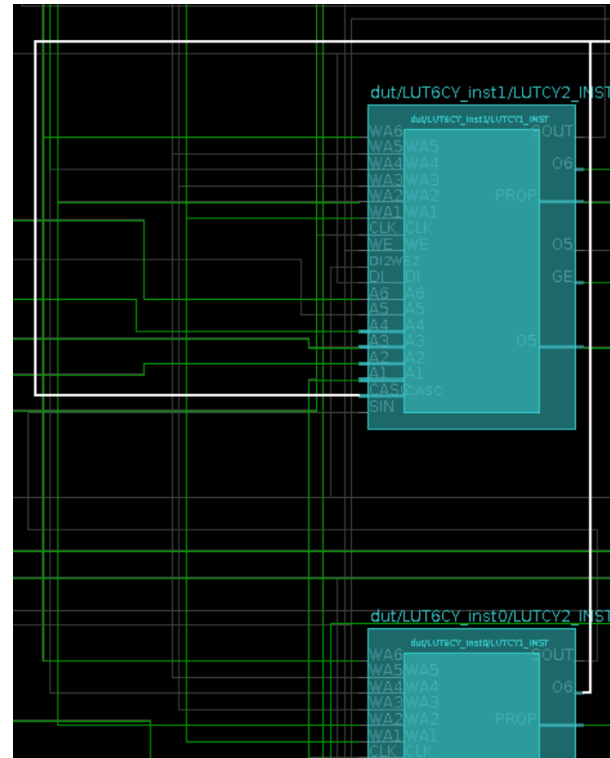


# Look-up Tables (LUTs)

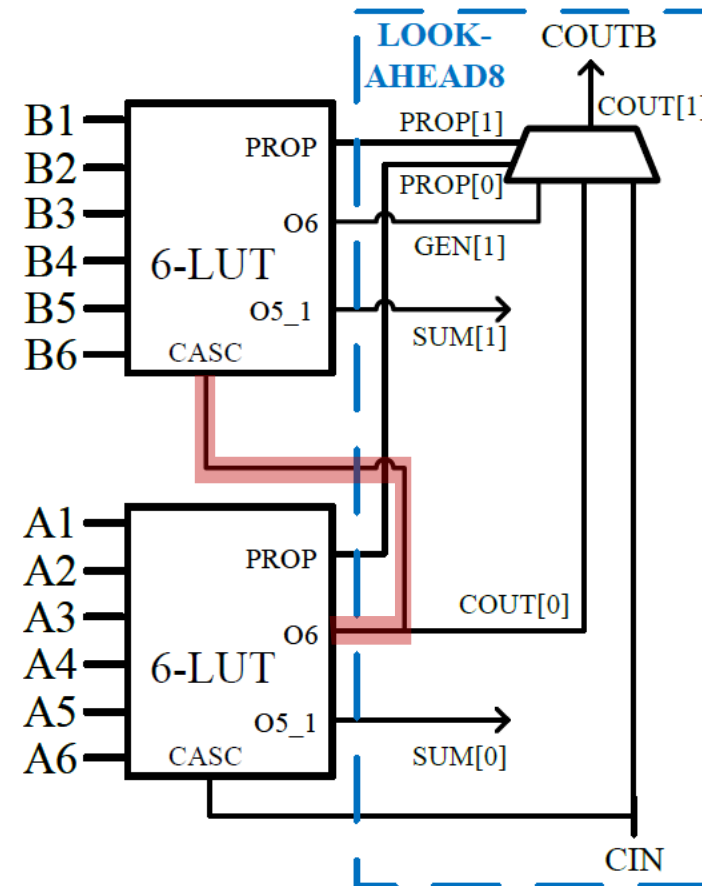


# Look-up Tables (LUTs)

Routing	Delay (ns)
LUT CASC	0.021
General Routing	0.25~0.45



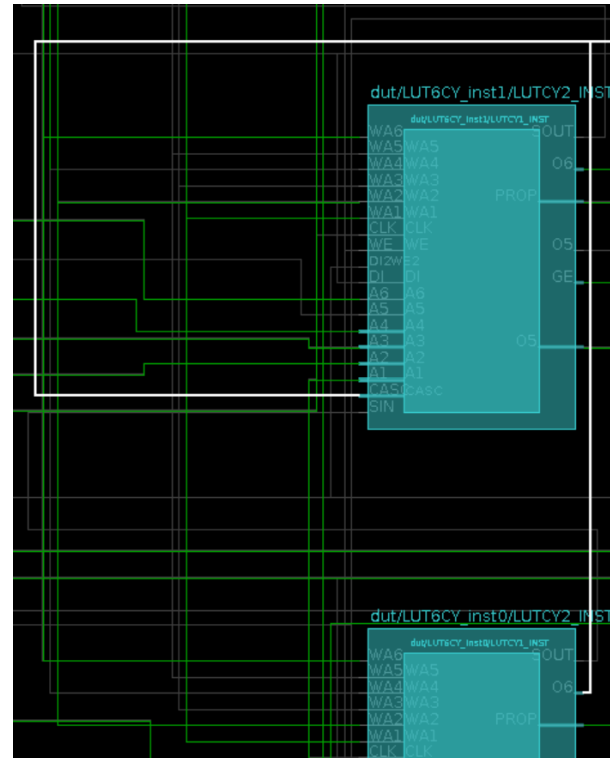
Versal



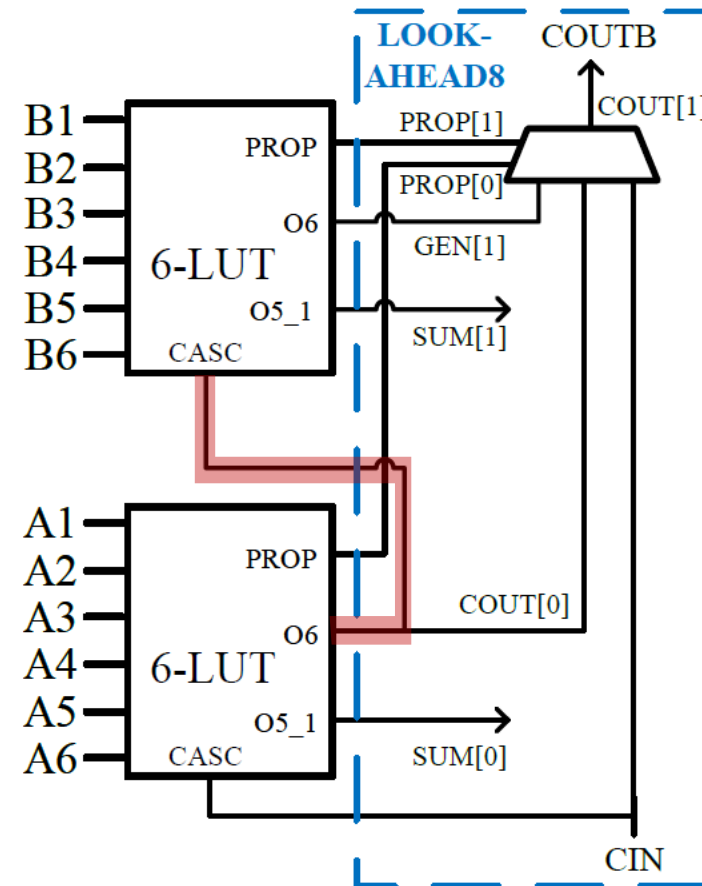


# Look-up Tables (LUTs)

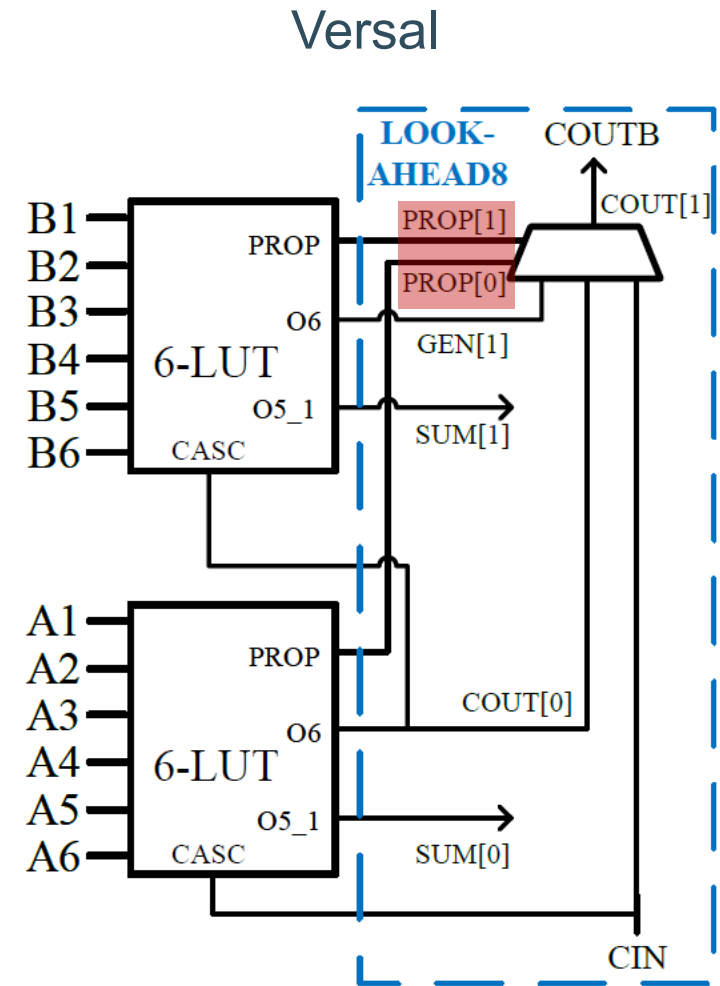
Routing	Delay (ns)
LUT CASC	0.021
General Routing	0.25~0.45



Versal

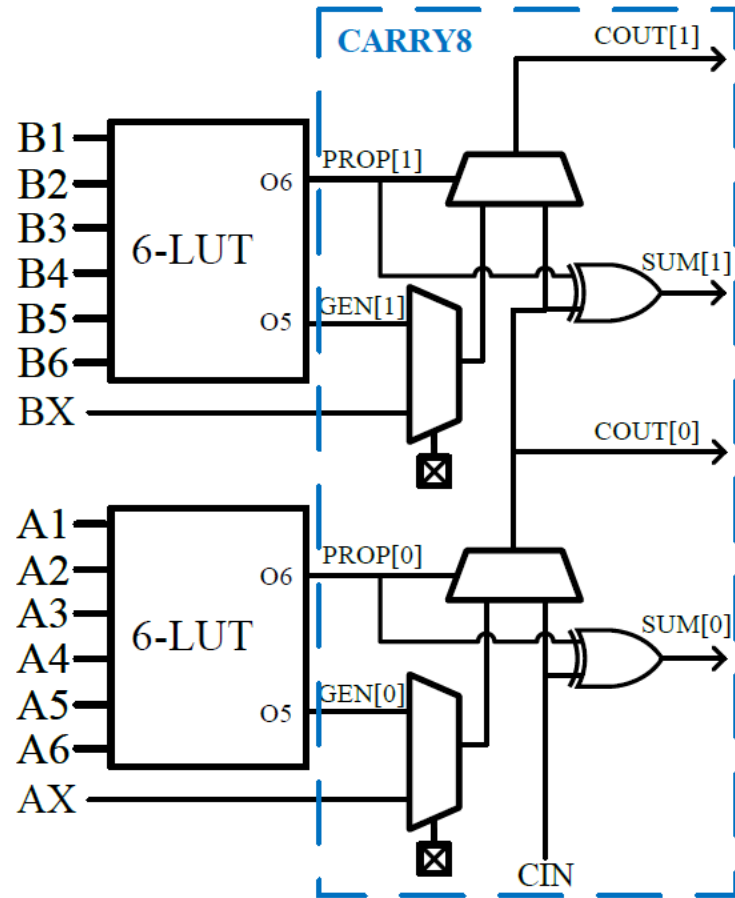


# Look-up Tables (LUTs)

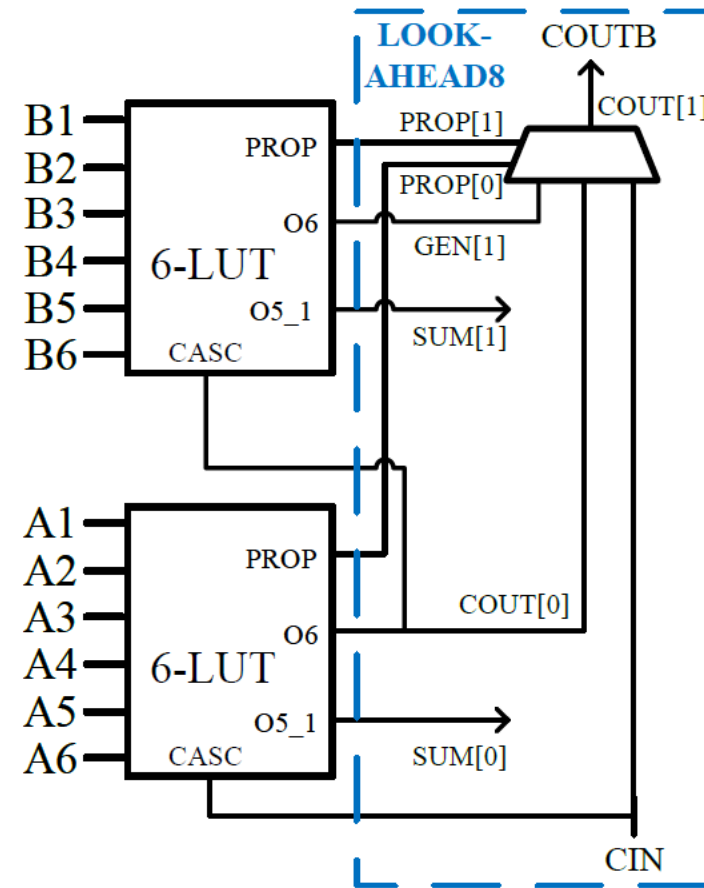


# Carry Logic

UltraScale

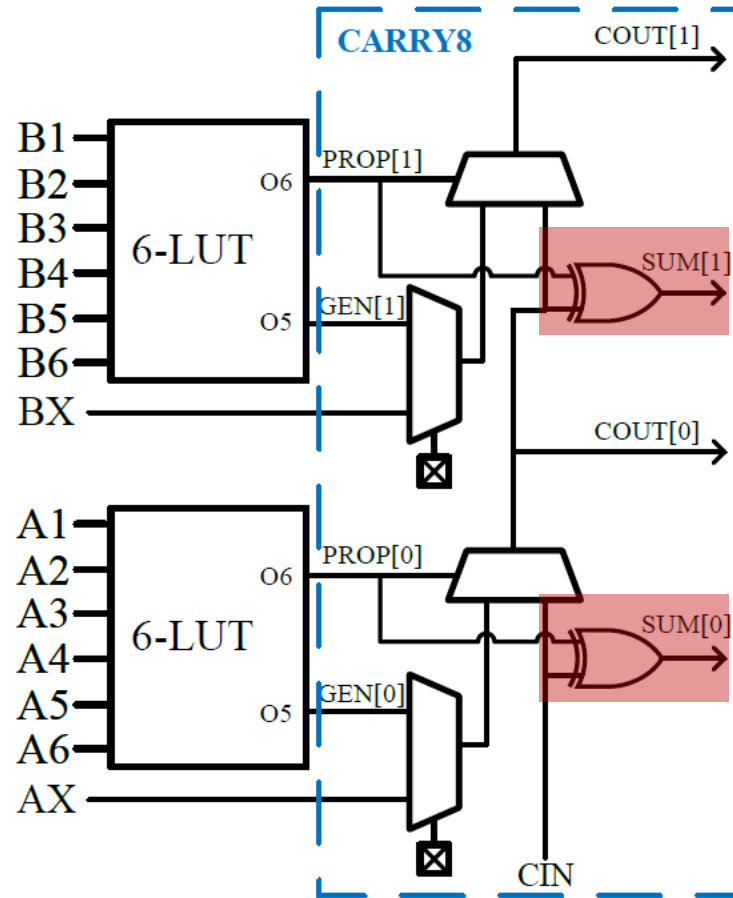


Versal

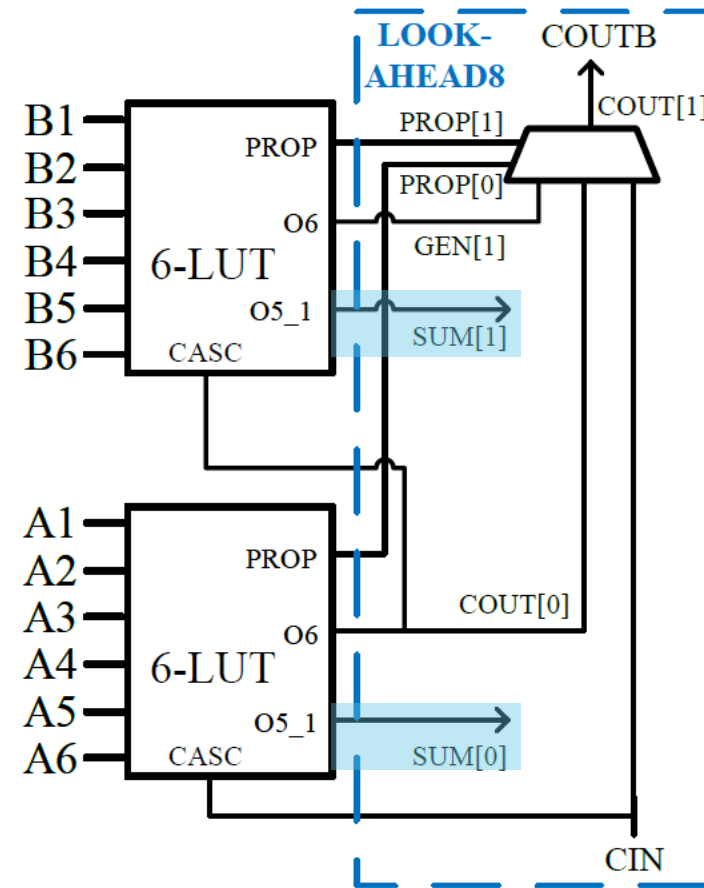


# Carry Logic

UltraScale

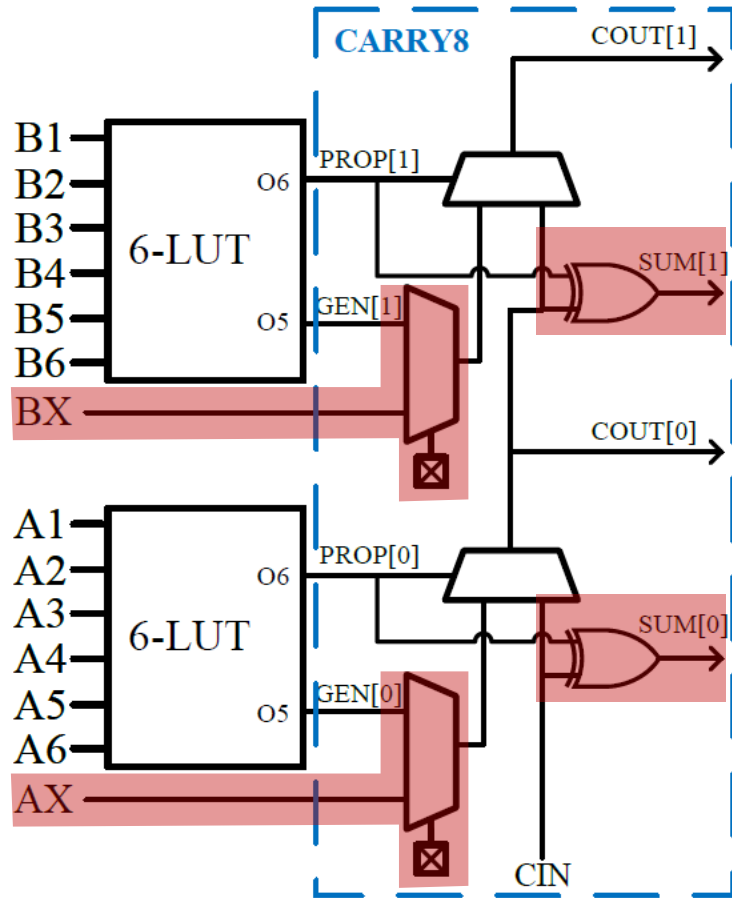


Versal

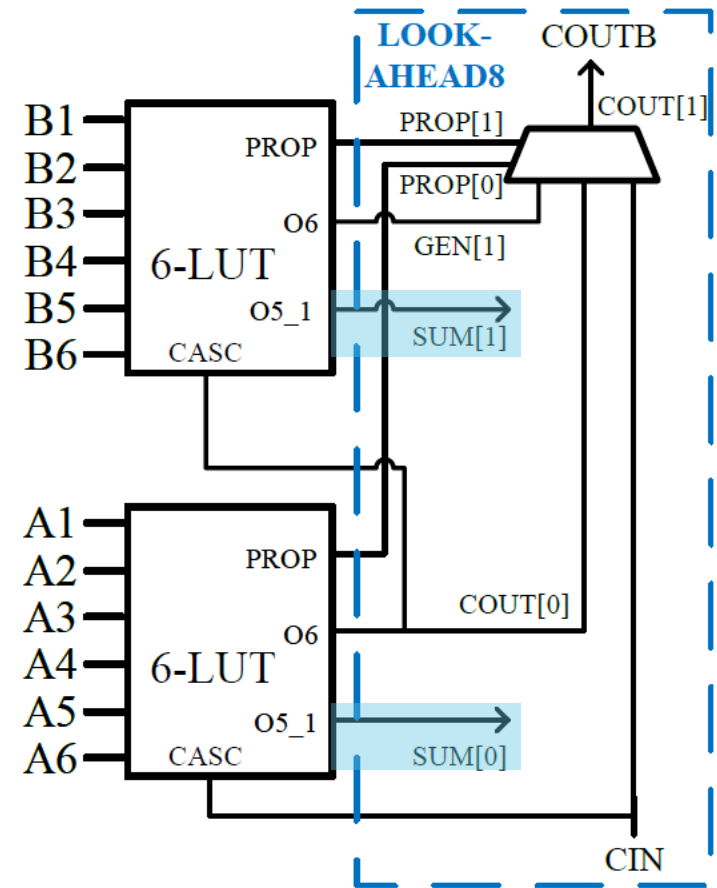


# Carry Logic

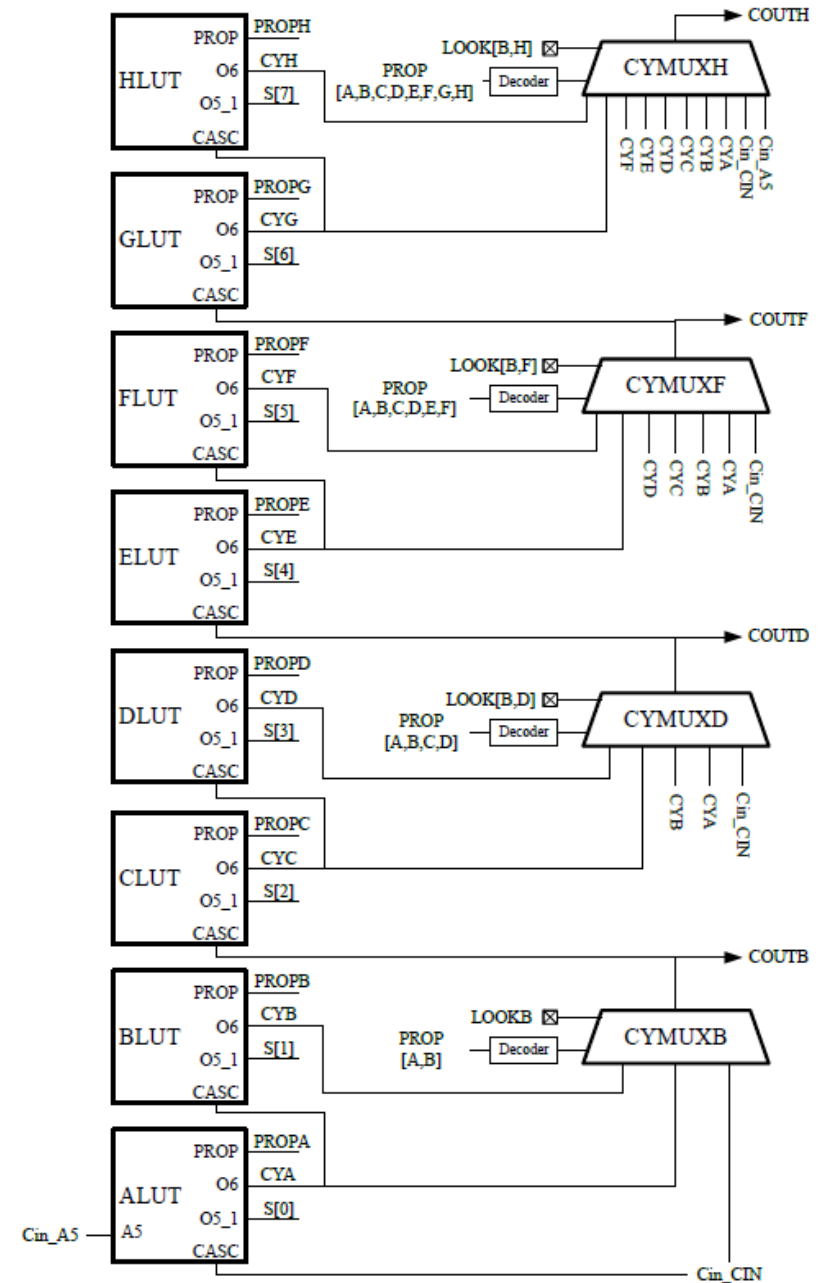
UltraScale



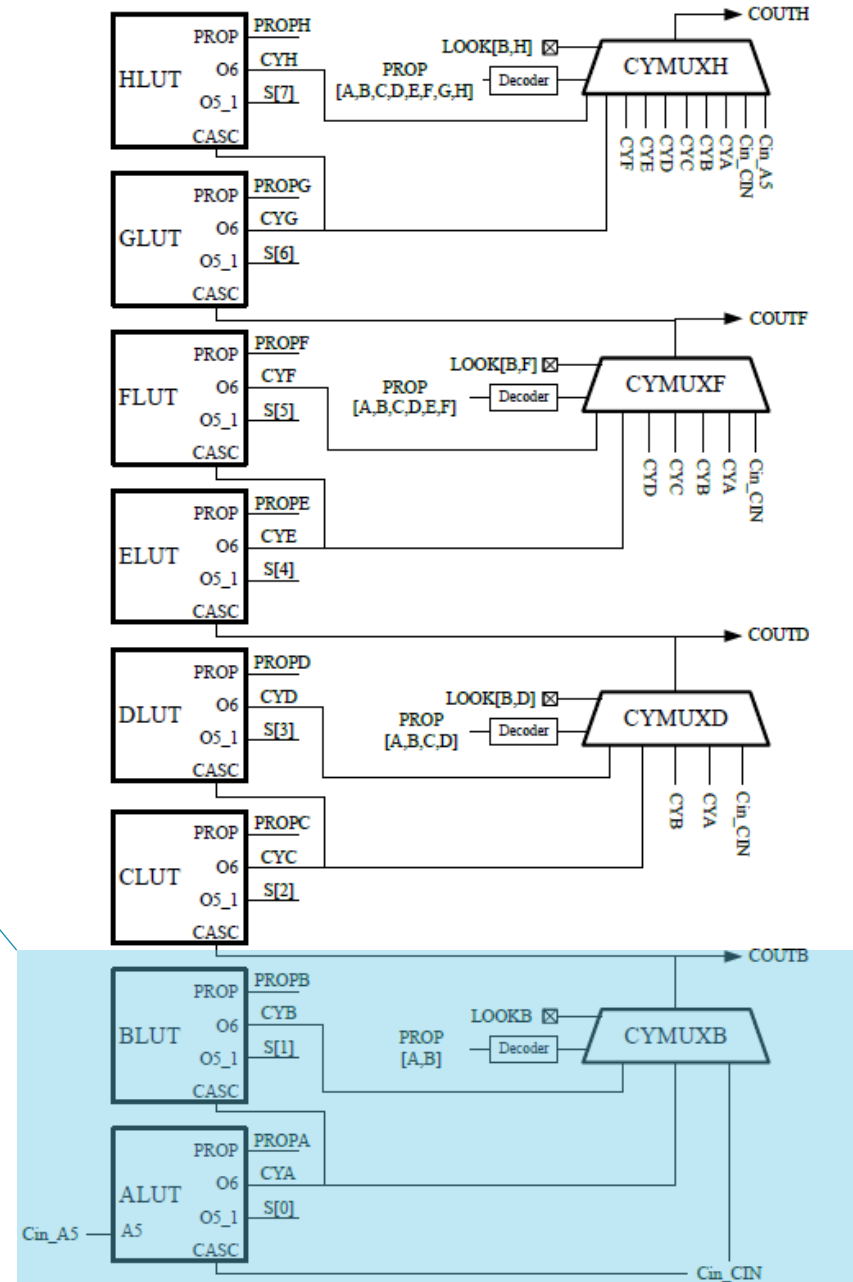
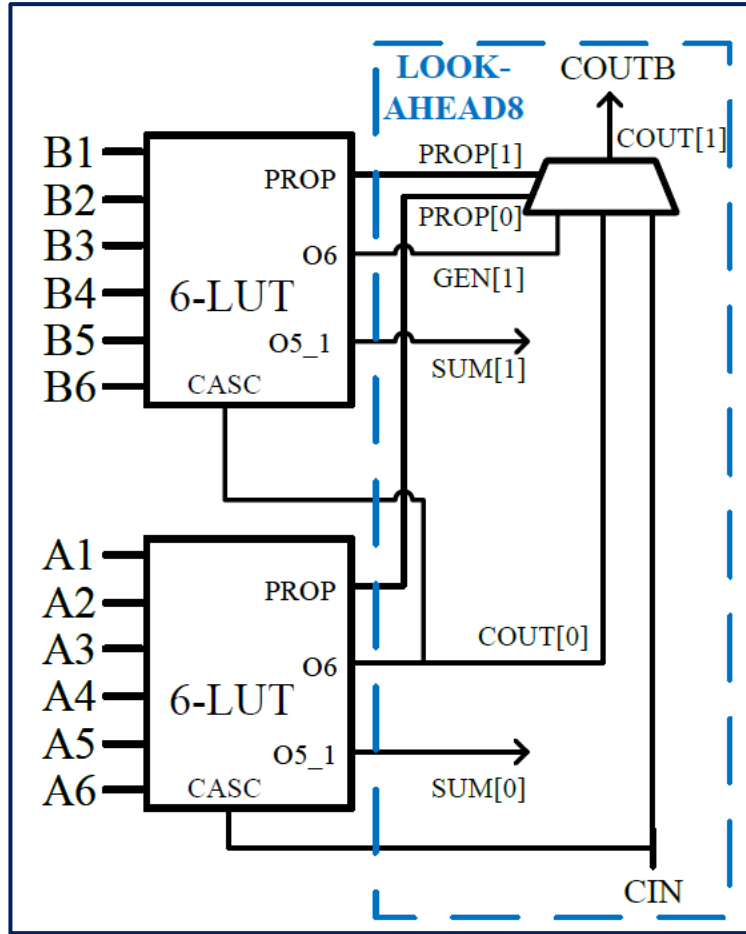
Versal



# Carry Logic



# Carry Logic



# Mapping Multiplication onto *Versal* CLBs

- Partial Product Generation
- Partial Product Addition/Compression
- Evaluation Results

# Partial Product Generation

# Partial Product Generation

➤ **Radix-2 Partial Product:**

$n$ -bit  $\times$   $n$ -bit multiplication  $\Rightarrow n^2$  partial product bits  $\Rightarrow \left\lceil \frac{n^2}{2} \right\rceil$  LUTs in dual 5-LUT mode

# Partial Product Generation

## ➤ Radix-2 Partial Product:

$n$ -bit  $\times$   $n$ -bit multiplication  $\Rightarrow n^2$  partial product bits  $\Rightarrow \left\lceil \frac{n^2}{2} \right\rceil$  LUTs in dual 5-LUT mode

## ➤ Radix-4 Booth Recoding:

$$A = a_3a_2a_1a_0 = a_3 \times (-2^3) + a_2 \times 2^2 + a_1 \times 2^1 + a_0 \times 2^0$$

$$B = b_3b_2b_1b_0 = b_3 \times (-2^3) + b_2 \times 2^2 + b_1 \times 2^1 + b_0 \times 2^0 \\ = (-2b_3 + b_2 + b_1) \times 2^2 + (-2b_1 + b_0 + b_{-1}) \times 2^0$$

# Partial Product Generation

## ➤ Radix-2 Partial Product:

$n$ -bit  $\times$   $n$ -bit multiplication  $\Rightarrow n^2$  partial product bits  $\Rightarrow \left\lceil \frac{n^2}{2} \right\rceil$  LUTs in dual 5-LUT mode

## ➤ Radix-4 Booth Recoding:

$$A = a_3a_2a_1a_0 = a_3 \times (-2^3) + a_2 \times 2^2 + a_1 \times 2^1 + a_0 \times 2^0$$

$$B = b_3b_2b_1b_0 = b_3 \times (-2^3) + b_2 \times 2^2 + b_1 \times 2^1 + b_0 \times 2^0$$
$$= \underbrace{(-2b_3 + b_2 + b_1)}_{\text{New Digit}} \times 2^2 + \underbrace{(-2b_1 + b_0 + b_{-1})}_{\text{New Digit}} \times 2^0$$

$$\text{New Digit} \quad b_i' = -2b_{2i+1} + b_{2i} + b_{2i-1}$$

# Partial Product Generation

## ➤ Radix-2 Partial Product:

$n$ -bit  $\times$   $n$ -bit multiplication  $\Rightarrow n^2$  partial product bits  $\Rightarrow \left\lceil \frac{n^2}{2} \right\rceil$  LUTs in dual 5-LUT mode

## ➤ Radix-4 Booth Recoding:

$$A = a_3a_2a_1a_0 = a_3 \times (-2^3) + a_2 \times 2^2 + a_1 \times 2^1 + a_0 \times 2^0$$

$$\begin{aligned} B = b_3b_2b_1b_0 &= b_3 \times (-2^3) + b_2 \times 2^2 + b_1 \times 2^1 + b_0 \times 2^0 \\ &= \underbrace{(-2b_3 + b_2 + b_1)}_{\text{New Digit}} \times 2^2 + \underbrace{(-2b_1 + b_0 + b_{-1})}_{\text{New Digit}} \times 2^0 \end{aligned}$$

$$\text{New Digit } b_i' = -2b_{2i+1} + b_{2i} + b_{2i-1}$$

$n$ -bit  $\times$   $n$ -bit multiplication  $\Rightarrow n \times \left\lceil \frac{n}{2} \right\rceil$  partial product bits

# Partial Product Generation

➤ **Radix-2 Partial Product:**

$n$ -bit  $\times$   $n$ -bit multiplication  $\Rightarrow n^2$  partial product bits  $\Rightarrow \lceil \frac{n^2}{2} \rceil$  LUTs in dual 5-LUT mode

➤ **Radix-4 Booth Recoding:**

$$A = a_3a_2a_1a_0 = a_3 \times (-2^3) + a_2 \times 2^2 + a_1 \times 2^1 + a_0 \times 2^0$$

$$B = b_3b_2b_1b_0 = b_3 \times (-2^3) + b_2 \times 2^2 + b_1 \times 2^1 + b_0 \times 2^0$$

$$= \underbrace{(-2b_3 + b_2 + b_1)}_{\text{New Digit}} \times 2^2 + \underbrace{(-2b_1 + b_0 + b_{-1})}_{\text{New Digit}} \times 2^0$$

$$\text{New Digit } b_i' = -2b_{2i+1} + b_{2i} + b_{2i-1}$$

$b_{2i+1}$	$b_{2i}$	$b_{2i-1}$	$b_i'$
0	0	0	0
0	0	1	1
0	1	0	1
0	1	1	2
1	0	0	-2
1	0	1	-1
1	1	0	-1
1	1	1	0

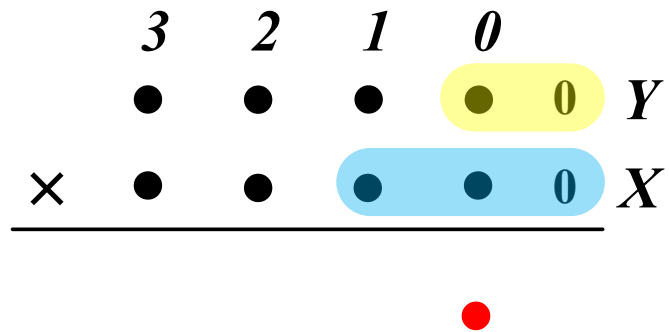
$n$ -bit  $\times$   $n$ -bit multiplication  $\Rightarrow n \times \lceil \frac{n}{2} \rceil$  partial product bits

# Partial Product Generation

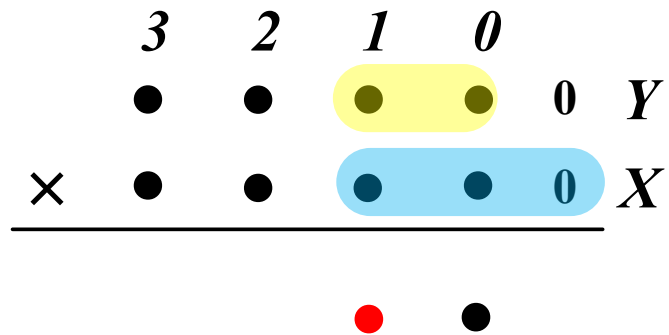
$$\begin{array}{rcccccc} & 3 & 2 & 1 & 0 & & \\ & \bullet & \bullet & \bullet & \bullet & 0 & Y \\ \times & \bullet & \bullet & \bullet & \bullet & 0 & X \\ \hline \end{array}$$

# Partial Product Generation

$$P'_{0,0}: x_1, x_0, 0, y_0, 0$$



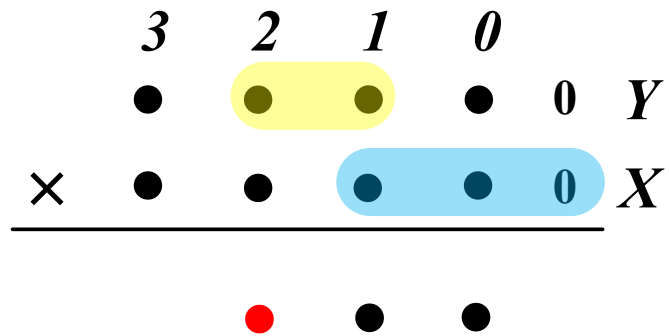
# Partial Product Generation



$$P'_{0,0}: x_1, x_0, 0, y_0, 0$$

$$P'_{0,1}: x_1, x_0, 0, y_1, y_0$$

# Partial Product Generation

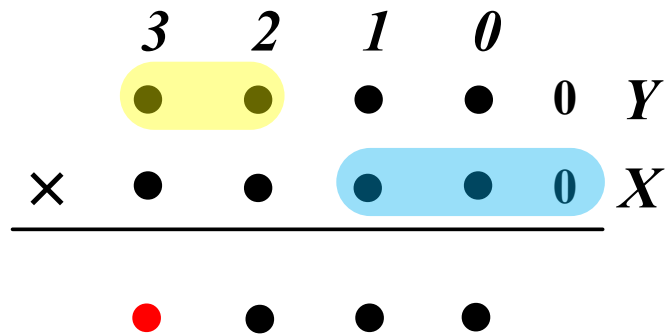


$$P'_{0,0}: x_1, x_0, 0, y_0, 0$$

$$P'_{0,1}: x_1, x_0, 0, y_1, y_0$$

$$P'_{0,2}: x_1, x_0, 0, y_2, y_1$$

# Partial Product Generation



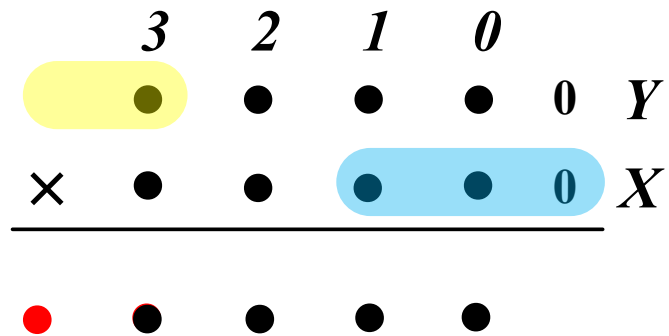
$$P'_{0,0}: x_1, x_0, 0, y_0, 0$$

$$P'_{0,1}: x_1, x_0, 0, y_1, y_0$$

$$P'_{0,2}: x_1, x_0, 0, y_2, y_1$$

$$P'_{0,3}: x_1, x_0, 0, y_3, y_2$$

# Partial Product Generation



$$P'_{0,0}: x_1, x_0, 0, y_0, 0$$

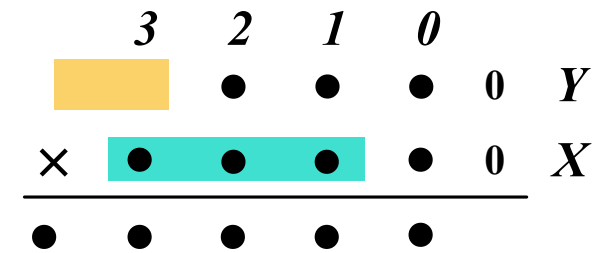
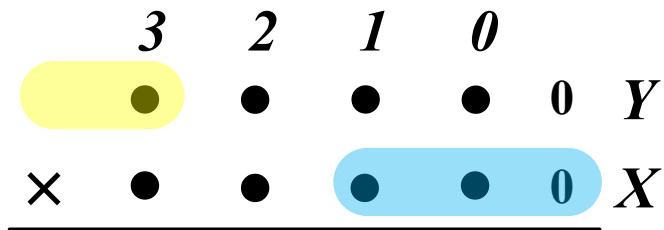
$$P'_{0,1}: x_1, x_0, 0, y_1, y_0$$

$$P'_{0,2}: x_1, x_0, 0, y_2, y_1$$

$$P'_{0,3}: x_1, x_0, 0, y_3, y_2$$

$$P'_{0,4}: x_1, x_0, 0, y_3$$

# Partial Product Generation



$$P'_{0,0}: x_1, x_0, 0, y_0, 0$$

$$P'_{0,1}: x_1, x_0, 0, y_1, y_0$$

$$P'_{0,2}: x_1, x_0, 0, y_2, y_1$$

$$P'_{0,3}: x_1, x_0, 0, y_3, y_2$$

$$P'_{0,4}: x_1, x_0, 0, y_3$$

$$P'_{1,0}: x_3, x_2, x_1, y_0, 0$$

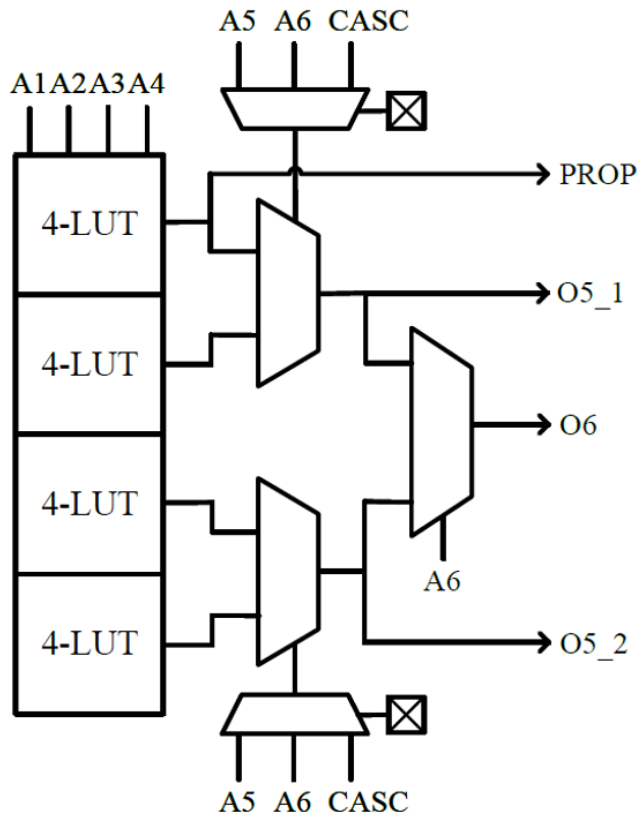
$$P'_{1,1}: x_3, x_2, x_1, y_1, y_0$$

$$P'_{1,2}: x_3, x_2, x_1, y_2, y_1$$

$$P'_{1,3}: x_3, x_2, x_1, y_3, y_2$$

$$P'_{1,4}: x_3, x_2, x_1, y_3$$

# Partial Product Generation



$$P'_{0,0}: x_1, x_0, 0, y_0, 0$$

$$P'_{0,1}: x_1, x_0, 0, y_1, y_0$$

$$P'_{0,2}: x_1, x_0, 0, y_2, y_1$$

$$P'_{0,3}: x_1, x_0, 0, y_3, y_2$$

$$P'_{0,4}: x_1, x_0, 0, y_3$$

$$P'_{1,0}: x_3, x_2, x_1, y_0, 0$$

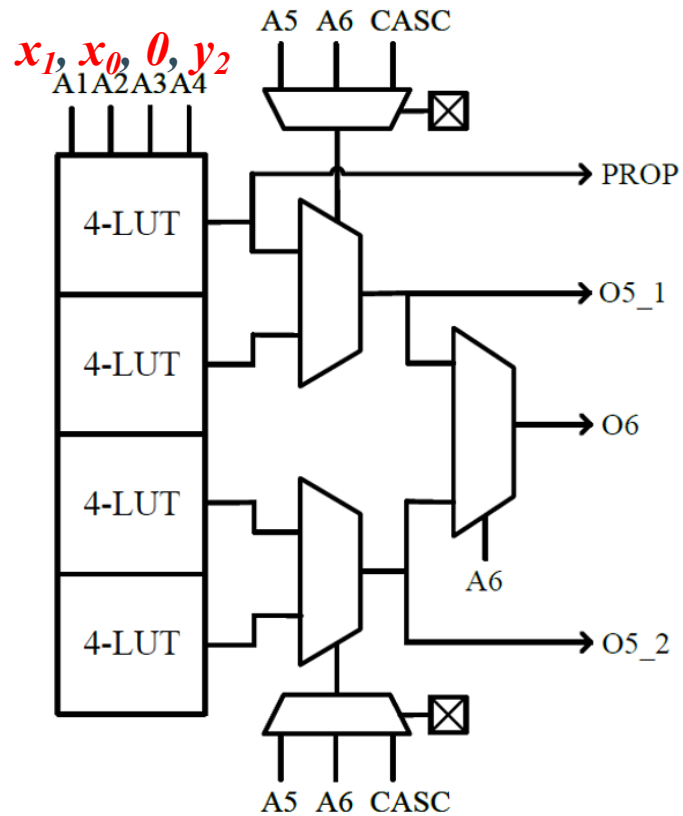
$$P'_{1,1}: x_3, x_2, x_1, y_1, y_0$$

$$P'_{1,2}: x_3, x_2, x_1, y_2, y_1$$

$$P'_{1,3}: x_3, x_2, x_1, y_3, y_2$$

$$P'_{1,4}: x_3, x_2, x_1, y_3$$

# Partial Product Generation



$$P'_{0,0}: x_1, x_0, 0, y_0, 0$$

$$P'_{0,1}: x_1, x_0, 0, y_1, y_0$$

$$P'_{0,2}: x_1, x_0, 0, y_2, y_1$$

$$P'_{0,3}: x_1, x_0, 0, y_3, y_2$$

$$P'_{0,4}: x_1, x_0, 0, y_3$$

$$P'_{1,0}: x_3, x_2, x_1, y_0, 0$$

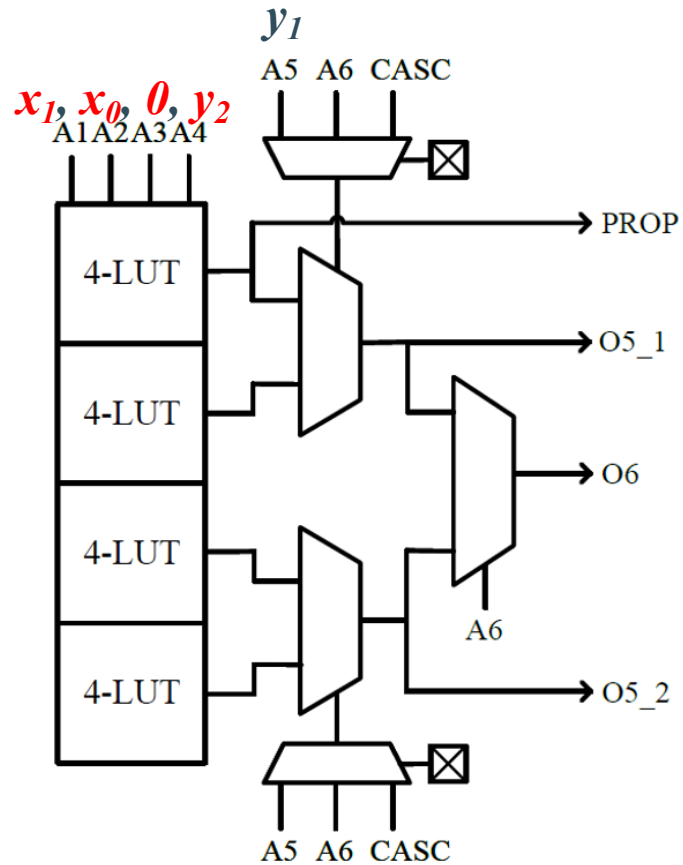
$$P'_{1,1}: x_3, x_2, x_1, y_1, y_0$$

$$P'_{1,2}: x_3, x_2, x_1, y_2, y_1$$

$$P'_{1,3}: x_3, x_2, x_1, y_3, y_2$$

$$P'_{1,4}: x_3, x_2, x_1, y_3$$

# Partial Product Generation



$$P'_{0,0}: x_1, x_0, 0, y_0, 0$$

$$P'_{0,1}: x_1, x_0, 0, y_1, y_0$$

$$P'_{0,2}: x_1, x_0, 0, y_2, y_1$$

$$P'_{0,3}: x_1, x_0, 0, y_3, y_2$$

$$P'_{0,4}: x_1, x_0, 0, y_3$$

$$P'_{1,0}: x_3, x_2, x_1, y_0, 0$$

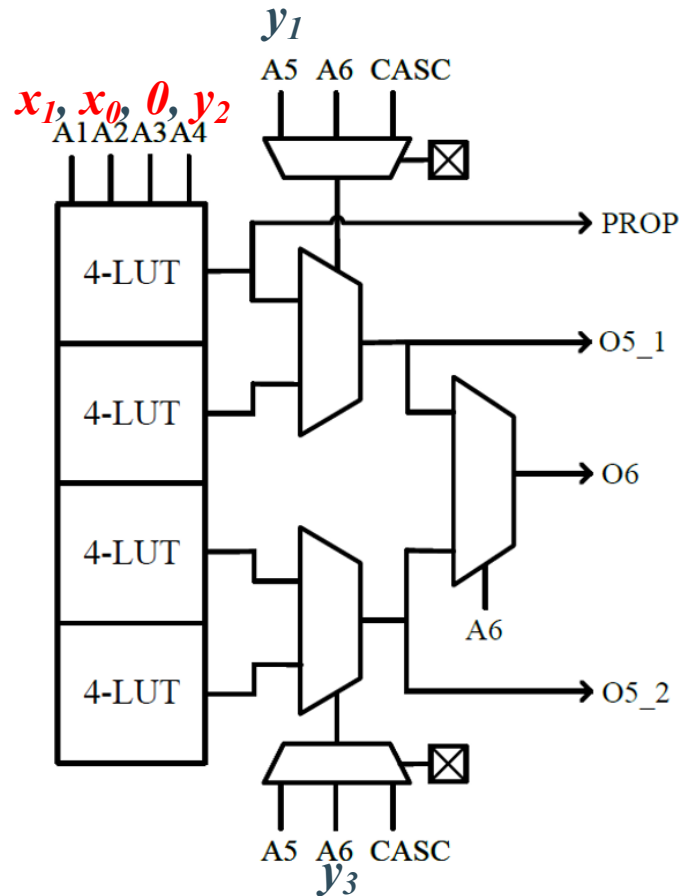
$$P'_{1,1}: x_3, x_2, x_1, y_1, y_0$$

$$P'_{1,2}: x_3, x_2, x_1, y_2, y_1$$

$$P'_{1,3}: x_3, x_2, x_1, y_3, y_2$$

$$P'_{1,4}: x_3, x_2, x_1, y_3$$

# Partial Product Generation



$$P'_{0,0}: x_1, x_0, 0, y_0, 0$$

$$P'_{0,1}: x_1, x_0, 0, y_1, y_0$$

$$P'_{0,2}: x_1, x_0, 0, y_2, y_1$$

$$P'_{0,3}: x_1, x_0, 0, y_3, y_2$$

$$P'_{0,4}: x_1, x_0, 0, y_3$$

$$P'_{1,0}: x_3, x_2, x_1, y_0, 0$$

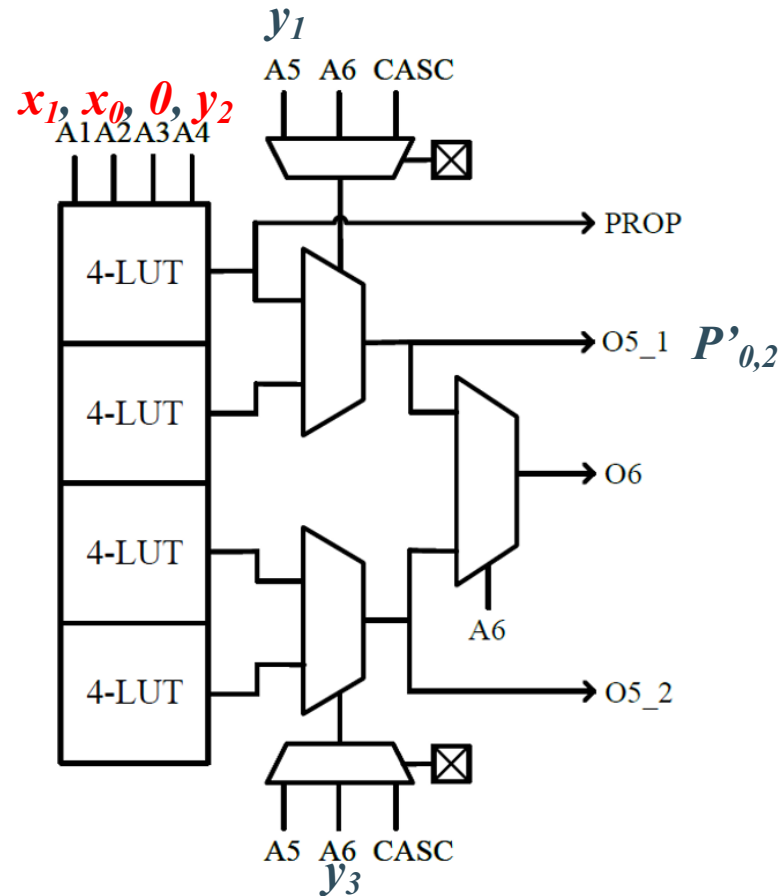
$$P'_{1,1}: x_3, x_2, x_1, y_1, y_0$$

$$P'_{1,2}: x_3, x_2, x_1, y_2, y_1$$

$$P'_{1,3}: x_3, x_2, x_1, y_3, y_2$$

$$P'_{1,4}: x_3, x_2, x_1, y_3$$

# Partial Product Generation



$$P'_{0,0}: x_1, x_0, 0, y_0, 0$$

$$P'_{0,1}: x_1, x_0, 0, y_1, y_0$$

$$P'_{0,2}: x_1, x_0, 0, y_2, y_1$$

$$P'_{0,3}: x_1, x_0, 0, y_3, y_2$$

$$P'_{0,4}: x_1, x_0, 0, y_3$$

$$P'_{1,0}: x_3, x_2, x_1, y_0, 0$$

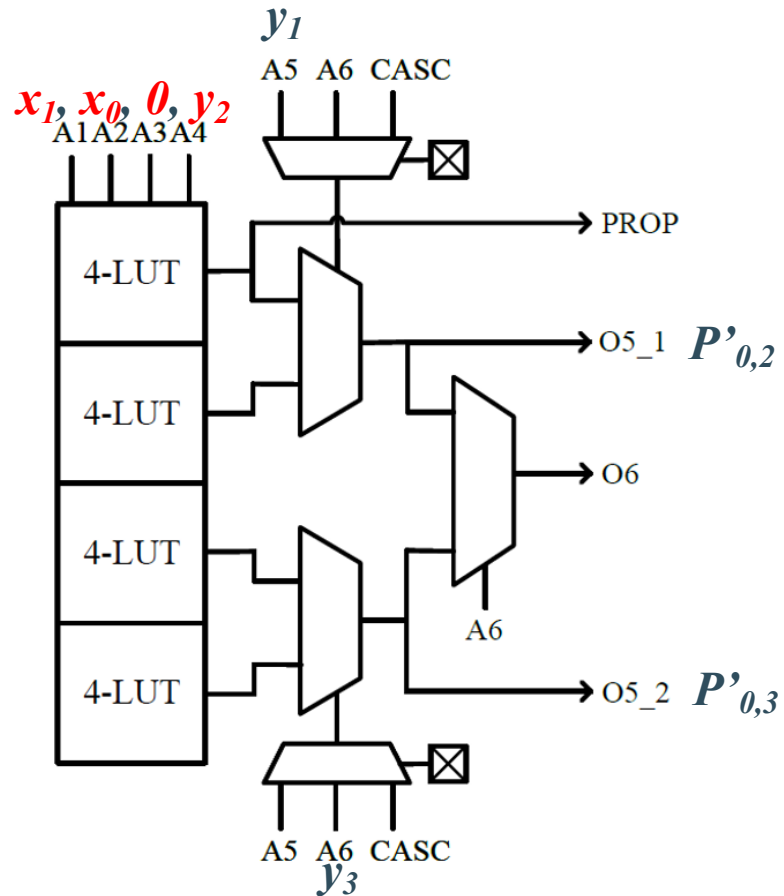
$$P'_{1,1}: x_3, x_2, x_1, y_1, y_0$$

$$P'_{1,2}: x_3, x_2, x_1, y_2, y_1$$

$$P'_{1,3}: x_3, x_2, x_1, y_3, y_2$$

$$P'_{1,4}: x_3, x_2, x_1, y_3$$

# Partial Product Generation



$$P'_{0,0}: x_1, x_0, 0, y_0, 0$$

$$P'_{0,1}: x_1, x_0, 0, y_1, y_0$$

$$P'_{0,2}: x_1, x_0, 0, y_2, y_1$$

$$P'_{0,3}: x_1, x_0, 0, y_3, y_2$$

$$P'_{0,4}: x_1, x_0, 0, y_3$$

$$P'_{1,0}: x_3, x_2, x_1, y_0, 0$$

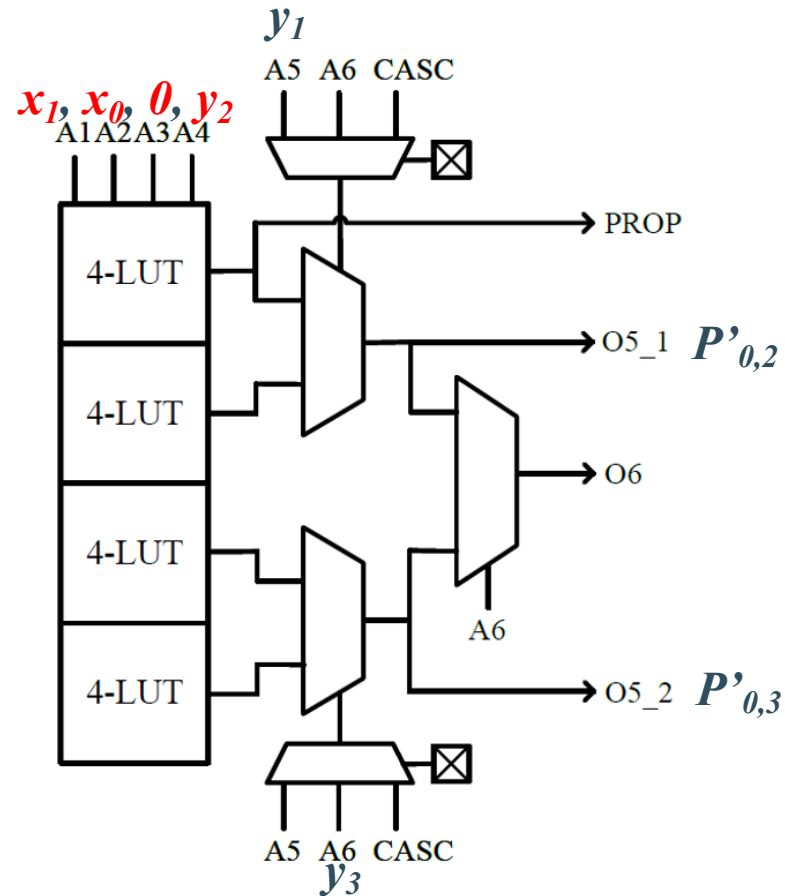
$$P'_{1,1}: x_3, x_2, x_1, y_1, y_0$$

$$P'_{1,2}: x_3, x_2, x_1, y_2, y_1$$

$$P'_{1,3}: x_3, x_2, x_1, y_3, y_2$$

$$P'_{1,4}: x_3, x_2, x_1, y_3$$

# Partial Product Generation



$$P'_{0,0}: x_1, x_0, 0, y_0, 0$$

$$P'_{0,1}: x_1, x_0, 0, y_1, y_0$$

$$P'_{0,2}: x_1, x_0, 0, y_2, y_1$$

$$P'_{0,3}: x_1, x_0, 0, y_3, y_2$$

$$P'_{0,4}: x_1, x_0, 0, y_3$$

$$P'_{1,0}: x_3, x_2, x_1, y_0, 0$$

$$P'_{1,1}: x_3, x_2, x_1, y_1, y_0$$

$$P'_{1,2}: x_3, x_2, x_1, y_2, y_1$$

$$P'_{1,3}: x_3, x_2, x_1, y_3, y_2$$

$$P'_{1,4}: x_3, x_2, x_1, y_3$$

1 LUT generates 2 partial product bits

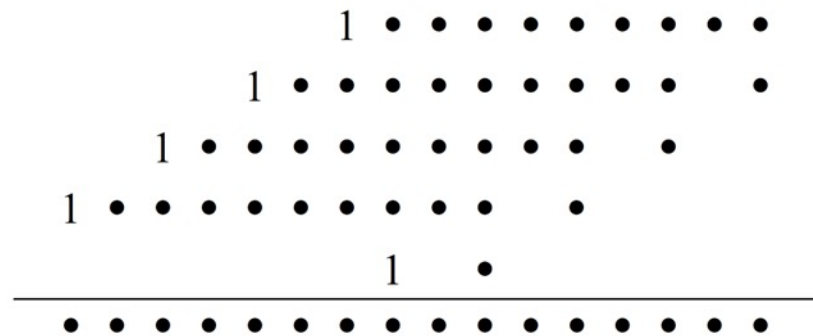
$\lceil \frac{n^2}{4} \rceil$  LUTs for all partial product bits of a n-bit multiplication

# Mapping Multiplication onto *Versal* CLBs

- Partial Product Generation
- Partial Product Addition/Compression
- Evaluation Results

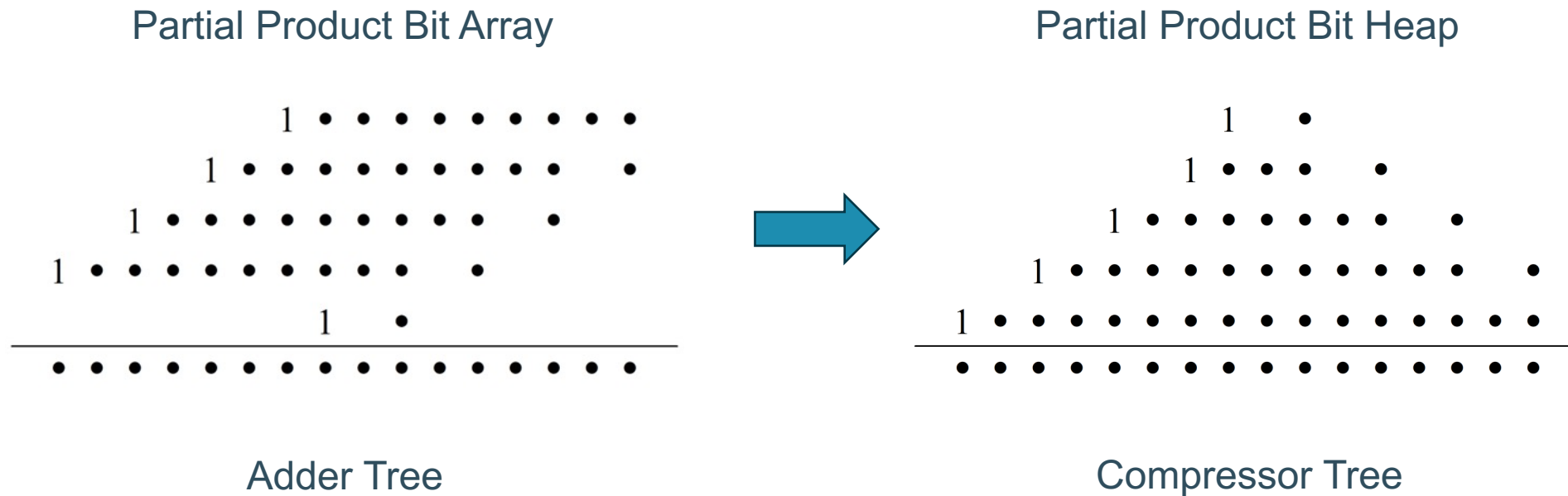
# Partial Product Compression

Partial Product Bit Array



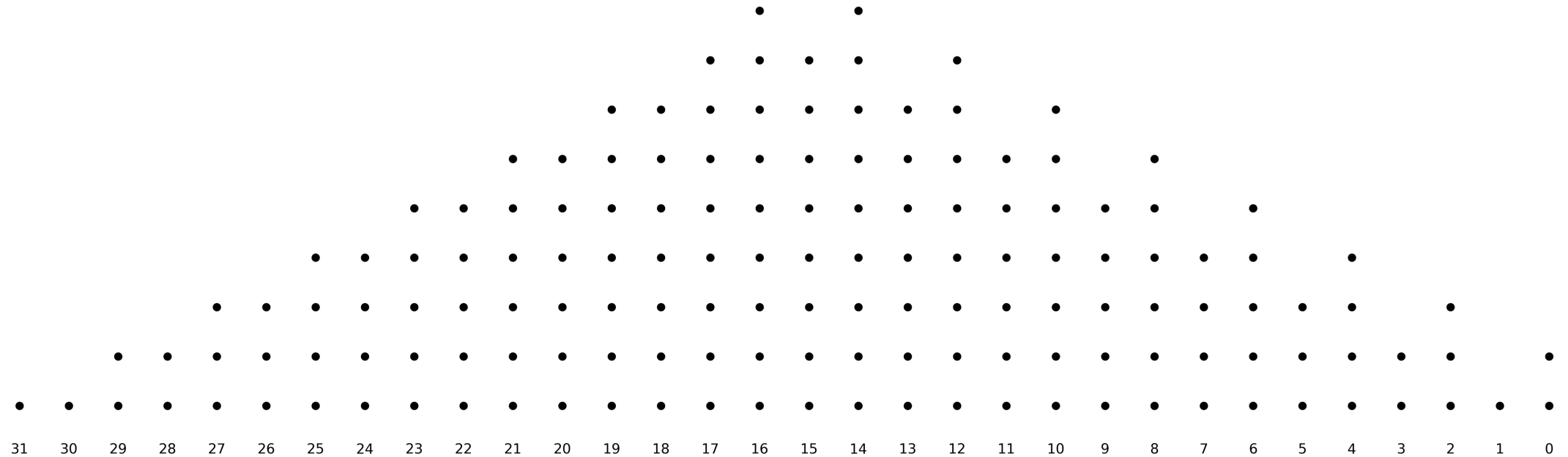
Adder Tree

# Partial Product Compression



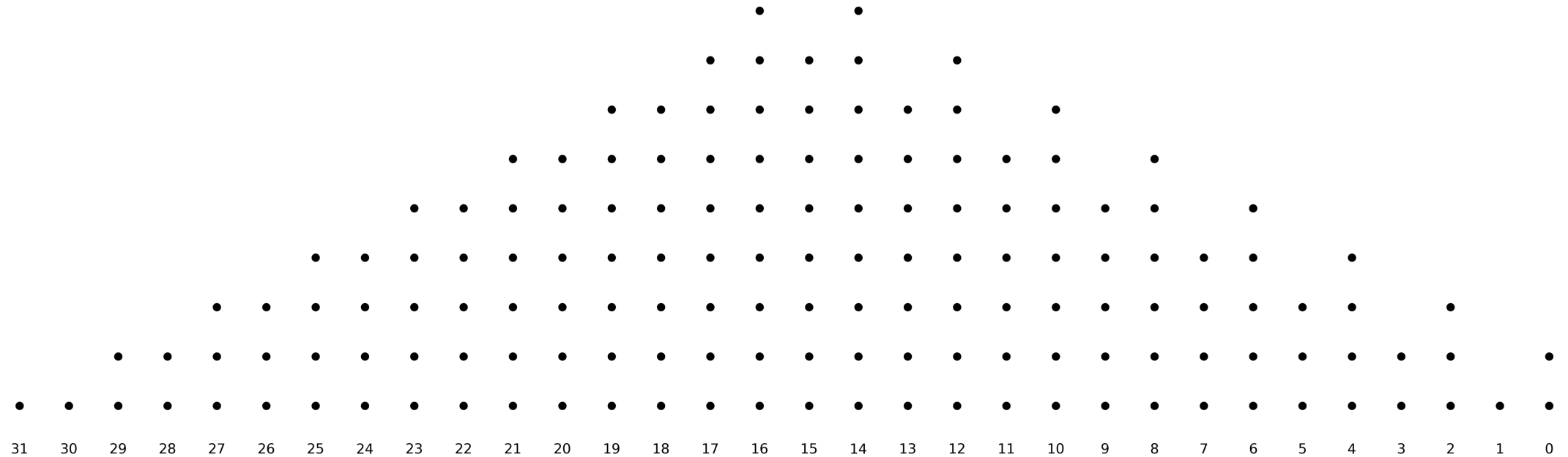
# Partial Product Compression

# Partial Product Compression



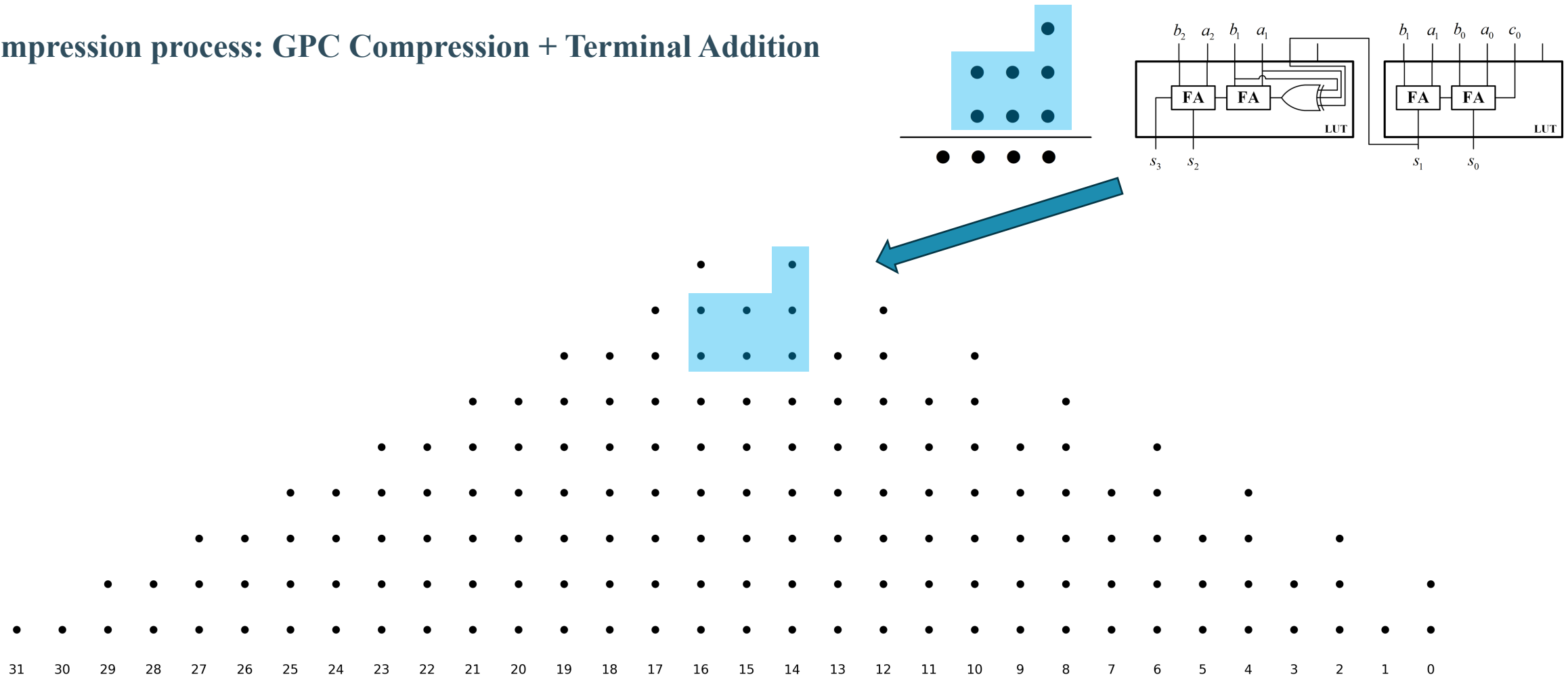
# Partial Product Compression

Compression process: GPC Compression + Terminal Addition



# Partial Product Compression

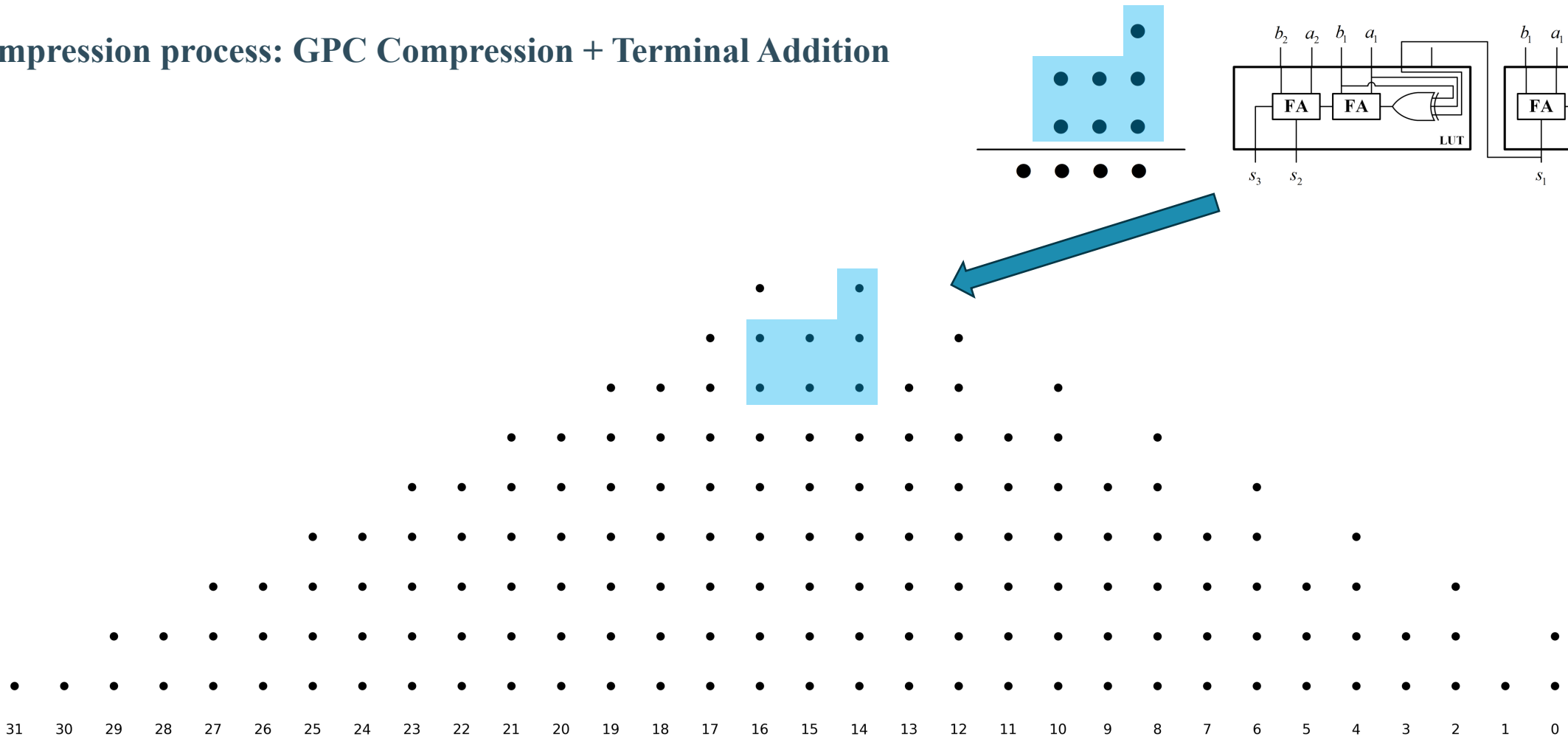
Compression process: GPC Compression + Terminal Addition



# Partial Product Compression

Compression process: GPC Compression + Terminal Addition

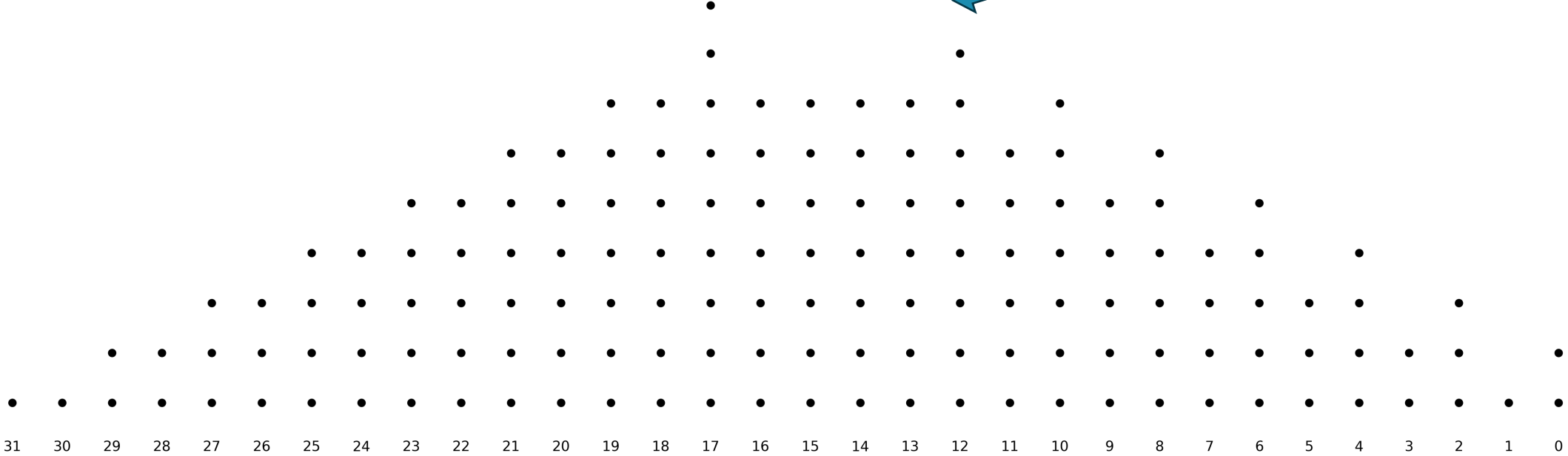
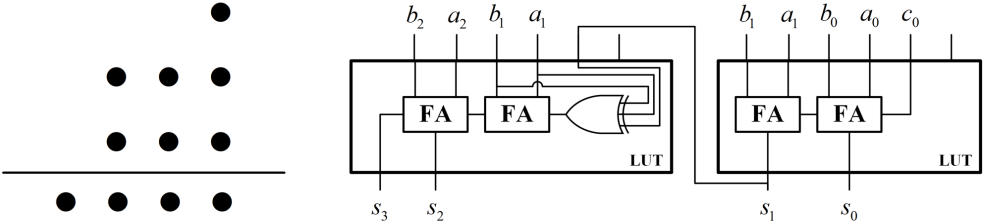
(2, 2, 3 : 4] counter



# Partial Product Compression

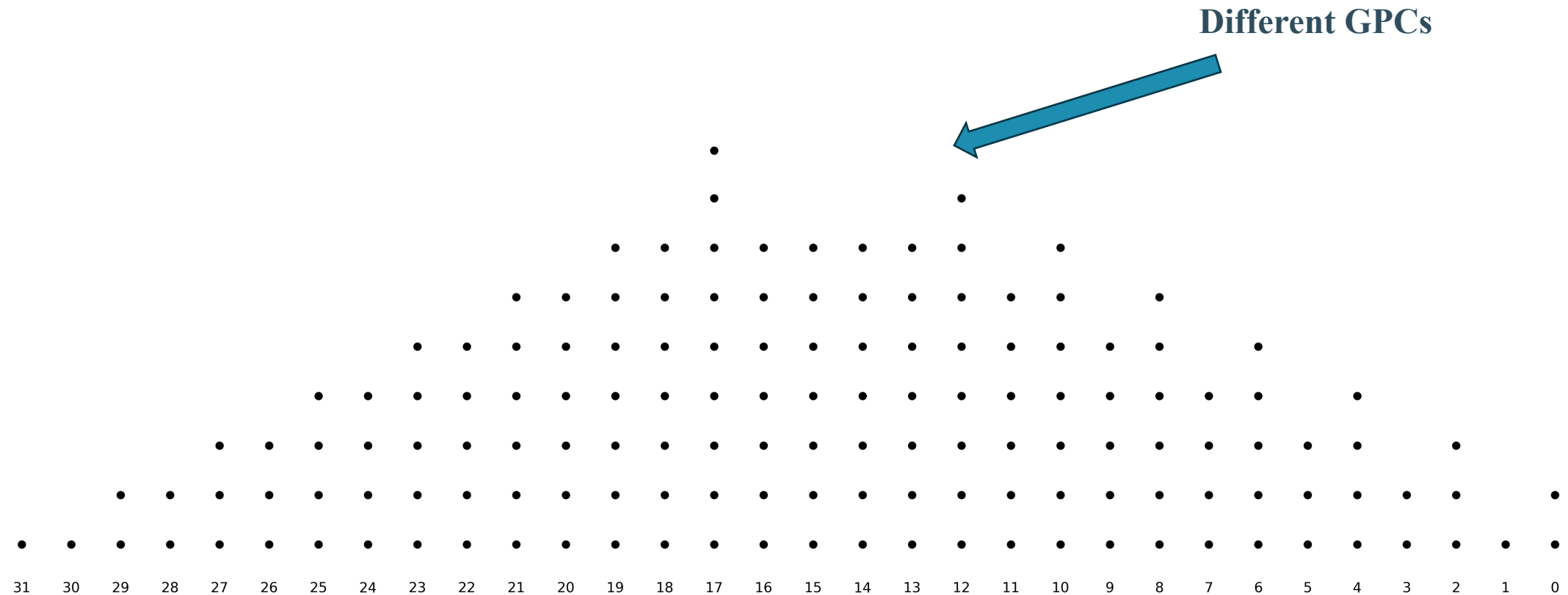
Compression process: GPC Compression + Terminal Addition

(2, 2, 3 : 4] counter



# Partial Product Compression

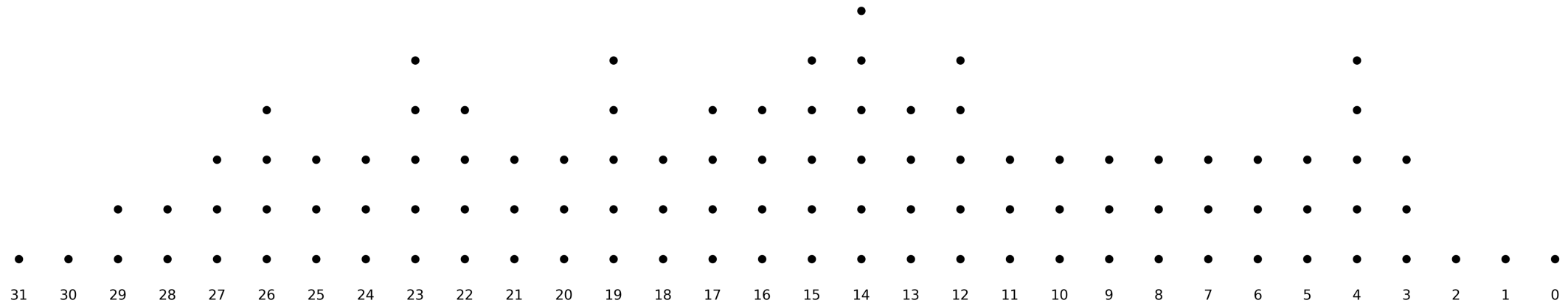
Compression process: GPC Compression + Terminal Addition



# Partial Product Compression

Compression process: GPC Compression + Terminal Addition

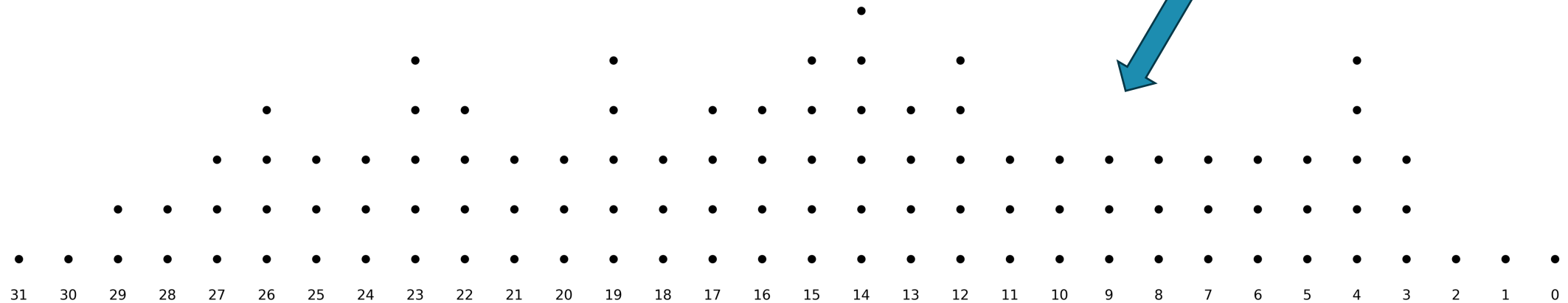
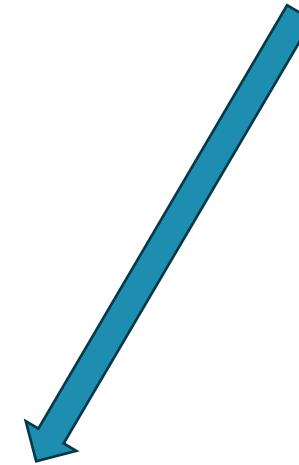
Different GPCs



# Partial Product Compression

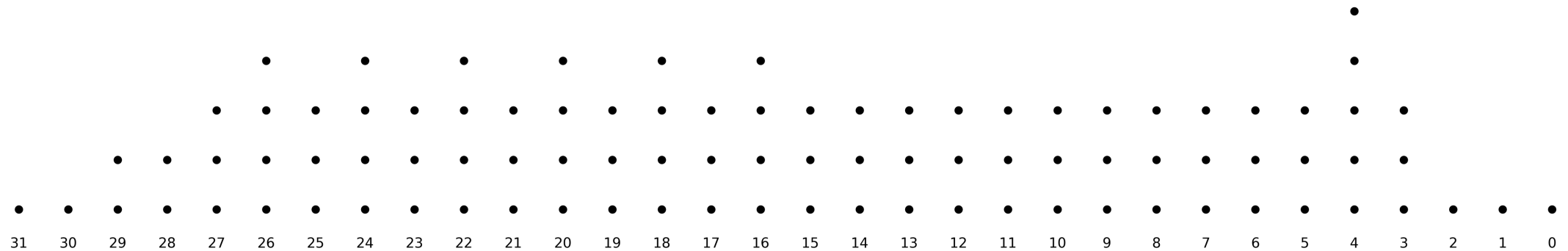
Compression process: GPC Compression + Terminal Addition

Different GPCs



# Partial Product Compression

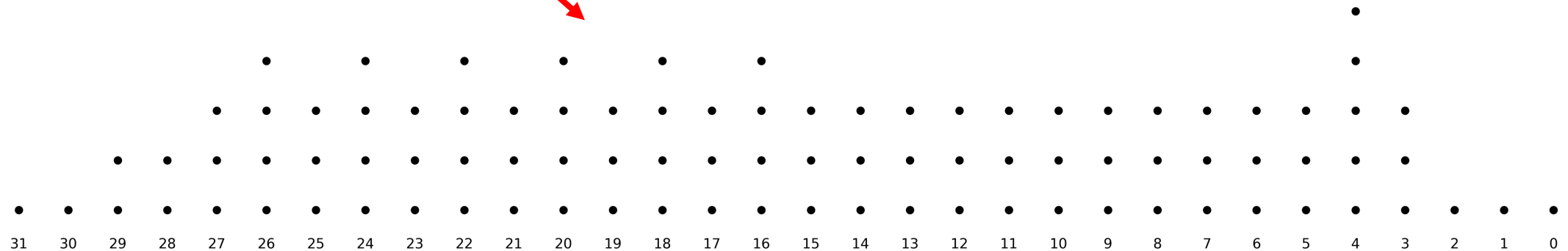
Compression process: GPC Compression + Terminal Addition



# Partial Product Compression

Compression process: GPC Compression + Terminal Addition

Terminal adder







# Versal GPCs and Compressor Trees [6]

Counter	#LUTs	E	S	LO.AH. <sup>1</sup>	Row/Column <sup>2</sup>
(3 : 2]	1	1.0	1.5	✓	Both
(6 : 3]	3	1.0	2.0	×	Neither
(10 : 4, 2)	3	1.33	1.67	×	Neither
(2, 5 : 1, 2, 1)	2	1.5	1.75	×	Column
(1, 5 : 3]	2	1.5	2.0	✓	Row
(2, 2, 3 : 4]	2	1.5	1.75	×	Row

<sup>1</sup>if the GPC is compatible with carry-lookahead structure

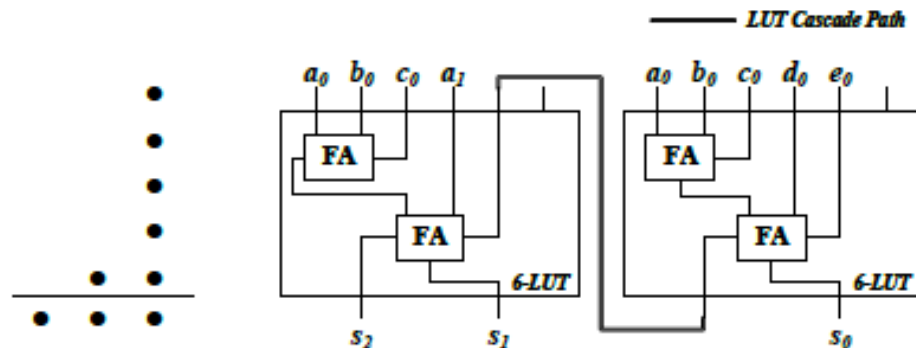
<sup>2</sup>if the GPC is used to construct row/column counters

# Versal GPCs and Compressor Trees [6]

Counter	#LUTs	E	S	LO.AH. <sup>1</sup>	Row/Column <sup>2</sup>
(3 : 2]	1	1.0	1.5	✓	Both
(6 : 3]	3	1.0	2.0	×	Neither
(10 : 4, 2)	3	1.33	1.67	×	Neither
(2, 5 : 1, 2, 1)	2	1.5	1.75	×	Column
(1, 5 : 3]	2	1.5	2.0	✓	Row
(2, 2, 3 : 4]	2	1.5	1.75	×	Row

<sup>1</sup>if the GPC is compatible with carry-lookahead structure

<sup>2</sup>if the GPC is used to construct row/column counters



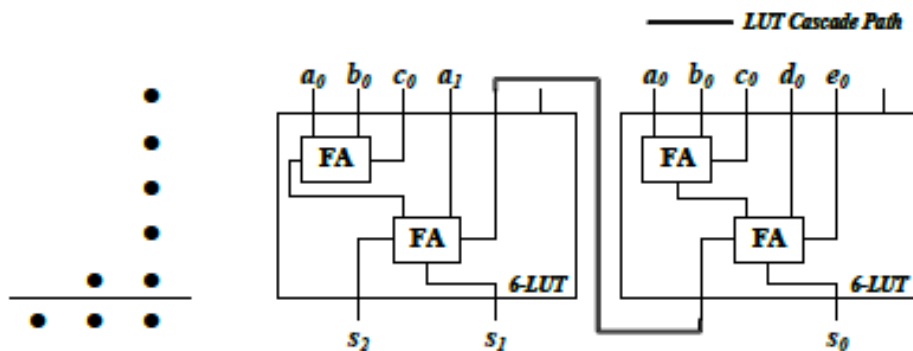
(1, 5 : 3] counter

# Versal GPCs and Compressor Trees [6]

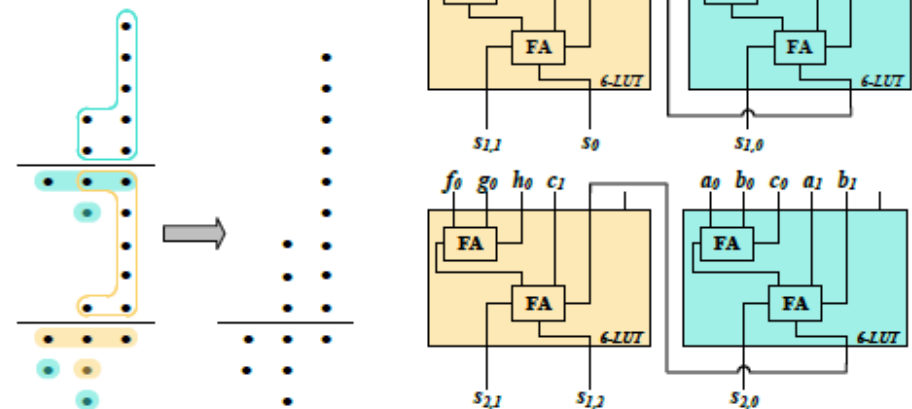
Counter	#LUTs	E	S	LO.AH. <sup>1</sup>	Row/Column <sup>2</sup>
(3 : 2]	1	1.0	1.5	✓	Both
(6 : 3]	3	1.0	2.0	×	Neither
(10 : 4, 2)	3	1.33	1.67	×	Neither
(2, 5 : 1, 2, 1)	2	1.5	1.75	×	Column
(1, 5 : 3]	2	1.5	2.0	✓	Row
(2, 2, 3 : 4]	2	1.5	1.75	×	Row

<sup>1</sup>if the GPC is compatible with carry-lookahead structure

<sup>2</sup>if the GPC is used to construct row/column counters



(1, 5 : 3] counter



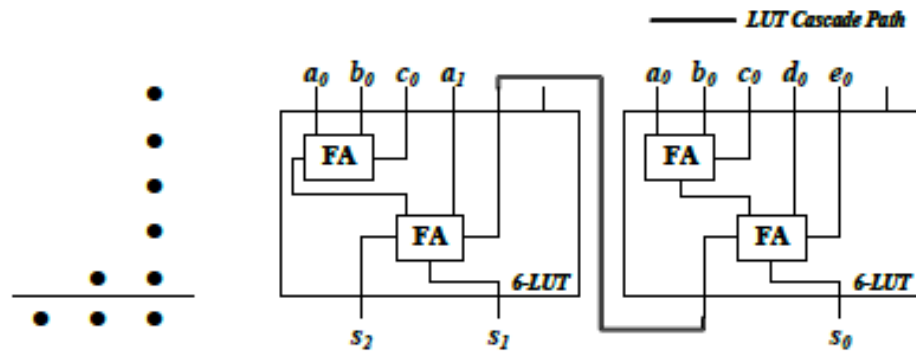
(3, 9 : 2, 3, 1) counter  
composed of two (2, 5 : 1, 2, 1) counter

# Versal GPCs and Compressor Trees [6]

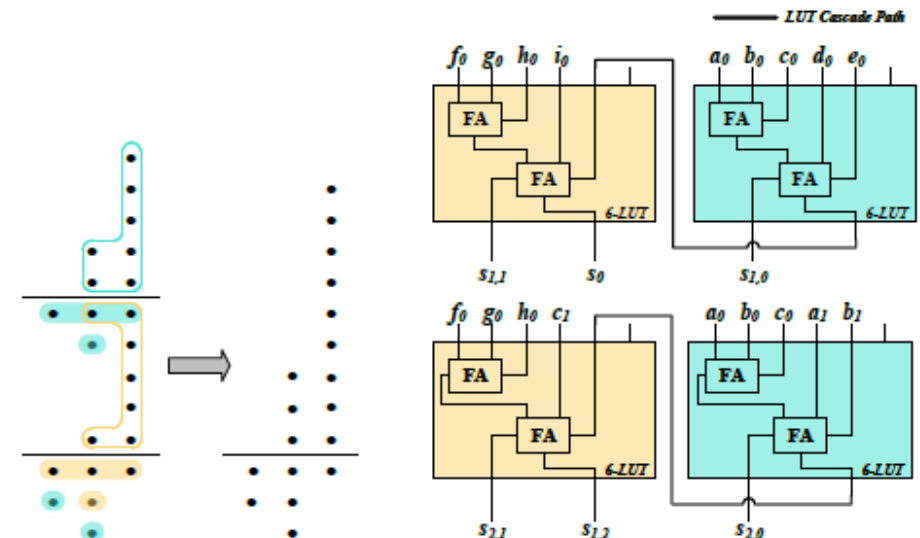
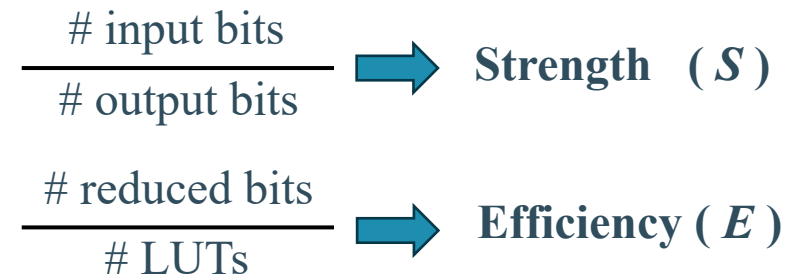
Counter	#LUTs	E	S	LO.AH. <sup>1</sup>	Row/Column <sup>2</sup>
(3 : 2]	1	1.0	1.5	✓	Both
(6 : 3]	3	1.0	2.0	×	Neither
(10 : 4, 2)	3	1.33	1.67	×	Neither
(2, 5 : 1, 2, 1)	2	1.5	1.75	×	Column
(1, 5 : 3]	2	1.5	2.0	✓	Row
(2, 2, 3 : 4]	2	1.5	1.75	×	Row

<sup>1</sup>if the GPC is compatible with carry-lookahead structure

<sup>2</sup>if the GPC is used to construct row/column counters



(1, 5 : 3] counter



(3, 9 : 2, 3, 1) counter  
composed of two (2, 5 : 1, 2, 1) counter

# Our *Versal* GPCs and Compressor Trees

## Basic GPCs in our work:

Counter	#LUTs	E	S	LO.AH. <sup>1</sup>	Row <sup>2</sup>
(5, 17 : 4, 5, 1)	8	1.5	2.2	×	×
(4, 13 : 3, 4, 1)	6	1.5	2.125	×	×
(3, 9 : 2, 3, 1)	4	1.5	2.0	×	✓
(9 : 4, 1) <sup>3</sup>	3	1.33	1.8	×	✓
(6 : 3]	3	1.0	2.0	×	✓
(2, 2, 3 : 4]	2	1.5	1.75	×	✓
(3 : 2]	1	1.0	1.5	✓	✓
(1, 5 : 3]	2	1.5	2.0	✓	✓

<sup>1</sup>if the GPC is compatible with carry-lookahead structure

<sup>2</sup>if the GPC is used to construct row counters

<sup>3</sup>new GPC proposed by this work

# Our *Versal* GPCs and Compressor Trees

## Basic GPCs in our work:

Counter	#LUTs	E	S	LO.AH. <sup>1</sup>	Row <sup>2</sup>
(5, 17 : 4, 5, 1)	8	1.5	2.2	×	×
(4, 13 : 3, 4, 1)	6	1.5	2.125	×	×
(3, 9 : 2, 3, 1)	4	1.5	2.0	×	✓
(9 : 4, 1) <sup>3</sup>	3	1.33	1.8	×	✓
(6 : 3]	3	1.0	2.0	×	✓
(2, 2, 3 : 4]	2	1.5	1.75	×	✓
(3 : 2]	1	1.0	1.5	✓	✓
(1, 5 : 3]	2	1.5	2.0	✓	✓

<sup>1</sup>if the GPC is compatible with carry-lookahead structure

<sup>2</sup>if the GPC is used to construct row counters

<sup>3</sup>new GPC proposed by this work

- New (9 : 4, 1) counter

# Our *Versal* GPCs and Compressor Trees

## Basic GPCs in our work:

Counter	#LUTs	E	S	LO.AH. <sup>1</sup>	Row <sup>2</sup>
(5, 17 : 4, 5, 1)	8	1.5	2.2	×	×
(4, 13 : 3, 4, 1)	6	1.5	2.125	×	×
(3, 9 : 2, 3, 1)	4	1.5	2.0	×	✓
(9 : 4, 1) <sup>3</sup>	3	1.33	1.8	×	✓
(6 : 3]	3	1.0	2.0	×	✓
(2, 2, 3 : 4]	2	1.5	1.75	×	✓
(3 : 2]	1	1.0	1.5	✓	✓
(1, 5 : 3]	2	1.5	2.0	✓	✓

<sup>1</sup>if the GPC is compatible with carry-lookahead structure

<sup>2</sup>if the GPC is used to construct row counters

<sup>3</sup>new GPC proposed by this work

- New (9 : 4, 1) counter
- Unrolled efficient column counters from [6]

# Our *Versal* GPCs and Compressor Trees

## Basic GPCs in our work:

Counter	#LUTs	E	S	LO.AH. <sup>1</sup>	Row <sup>2</sup>
(5, 17 : 4, 5, 1)	8	1.5	2.2	×	×
(4, 13 : 3, 4, 1)	6	1.5	2.125	×	×
(3, 9 : 2, 3, 1)	4	1.5	2.0	×	✓
(9 : 4, 1) <sup>3</sup>	3	1.33	1.8	×	✓
(6 : 3]	3	1.0	2.0	×	✓
(2, 2, 3 : 4]	2	1.5	1.75	×	✓
(3 : 2]	1	1.0	1.5	✓	✓
(1, 5 : 3]	2	1.5	2.0	✓	✓

<sup>1</sup>if the GPC is compatible with carry-lookahead structure

<sup>2</sup>if the GPC is used to construct row counters

<sup>3</sup>new GPC proposed by this work

- New (9 : 4, 1) counter
- Unrolled efficient column counters from [6]
- Extend the eligible row counter candidates
- New way to build row counters using LOOKAHEAD8

# Our *Versal* GPCs and Compressor Trees

## Basic GPCs in our work:

Counter	#LUTs	E	S	LO.AH. <sup>1</sup>	Row <sup>2</sup>
(5, 17 : 4, 5, 1)	8	1.5	2.2	×	×
(4, 13 : 3, 4, 1)	6	1.5	2.125	×	×
(3, 9 : 2, 3, 1)	4	1.5	2.0	×	✓
(9 : 4, 1) <sup>3</sup>	3	1.33	1.8	×	✓
(6 : 3]	3	1.0	2.0	×	✓
(2, 2, 3 : 4]	2	1.5	1.75	×	✓
(3 : 2]	1	1.0	1.5	✓	✓
(1, 5 : 3]	2	1.5	2.0	✓	✓

<sup>1</sup>if the GPC is compatible with carry-lookahead structure

<sup>2</sup>if the GPC is used to construct row counters

<sup>3</sup>new GPC proposed by this work

- New (9 : 4, 1) counter
- Unrolled efficient column counters from [6]
- Extend the eligible row counter candidates
- New way to build row counters using LOOKAHEAD8
- New implementation of quaternary terminal adder from [6]

# Our *Versal* GPCs and Compressor Trees

## Basic GPCs in our work:

Counter	#LUTs	E	S	LO.AH. <sup>1</sup>	Row <sup>2</sup>
(5, 17 : 4, 5, 1)	8	1.5	2.2	×	×
(4, 13 : 3, 4, 1)	6	1.5	2.125	×	×
(3, 9 : 2, 3, 1)	4	1.5	2.0	×	✓
(9 : 4, 1) <sup>3</sup>	3	1.33	1.8	×	✓
(6 : 3]	3	1.0	2.0	×	✓
(2, 2, 3 : 4]	2	1.5	1.75	×	✓
(3 : 2]	1	1.0	1.5	✓	✓
(1, 5 : 3]	2	1.5	2.0	✓	✓

<sup>1</sup>if the GPC is compatible with carry-lookahead structure

<sup>2</sup>if the GPC is used to construct row counters

<sup>3</sup>new GPC proposed by this work

- New (9 : 4, 1) counter
- Unrolled efficient column counters from [6]
- Extend the eligible row counter candidates
- New way to build row counters using LOOKAHEAD8
- New implementation of quaternary terminal adder from [6]
- New heuristics for compressor tree synthesis

# Our *Versal* GPCs and Compressor Trees

## Basic GPCs in our work:

Counter	#LUTs	E	S	LO.AH. <sup>1</sup>	Row <sup>2</sup>
(5, 17 : 4, 5, 1)	8	1.5	2.2	×	×
(4, 13 : 3, 4, 1)	6	1.5	2.125	×	×
(3, 9 : 2, 3, 1)	4	1.5	2.0	×	✓
(9 : 4, 1) <sup>3</sup>	3	1.33	1.8	×	✓
(6 : 3]	3	1.0	2.0	×	✓
(2, 2, 3 : 4]	2	1.5	1.75	×	✓
(3 : 2]	1	1.0	1.5	✓	✓
(1, 5 : 3]	2	1.5	2.0	✓	✓

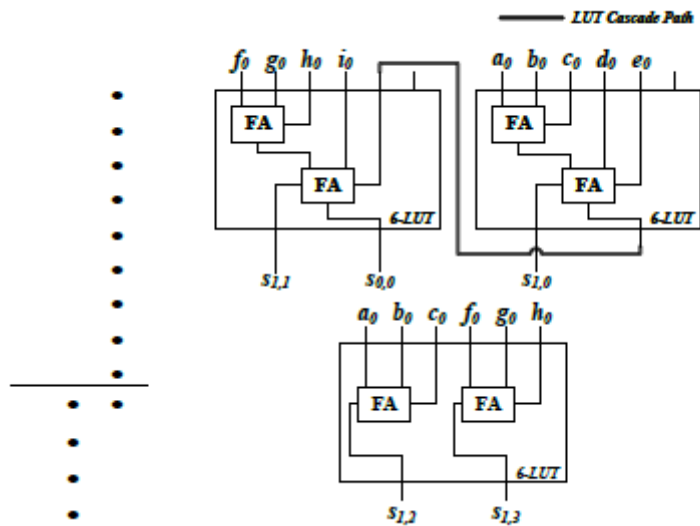
<sup>1</sup>if the GPC is compatible with carry-lookahead structure

<sup>2</sup>if the GPC is used to construct row counters

<sup>3</sup>new GPC proposed by this work

- New (9 : 4, 1) counter
- Unrolled efficient column counters from [6]
- Extend the eligible row counter candidates
- New way to build row counters using LOOKAHEAD8
- New implementation of quaternary terminal adder from [6]
- New heuristics for compressor tree synthesis

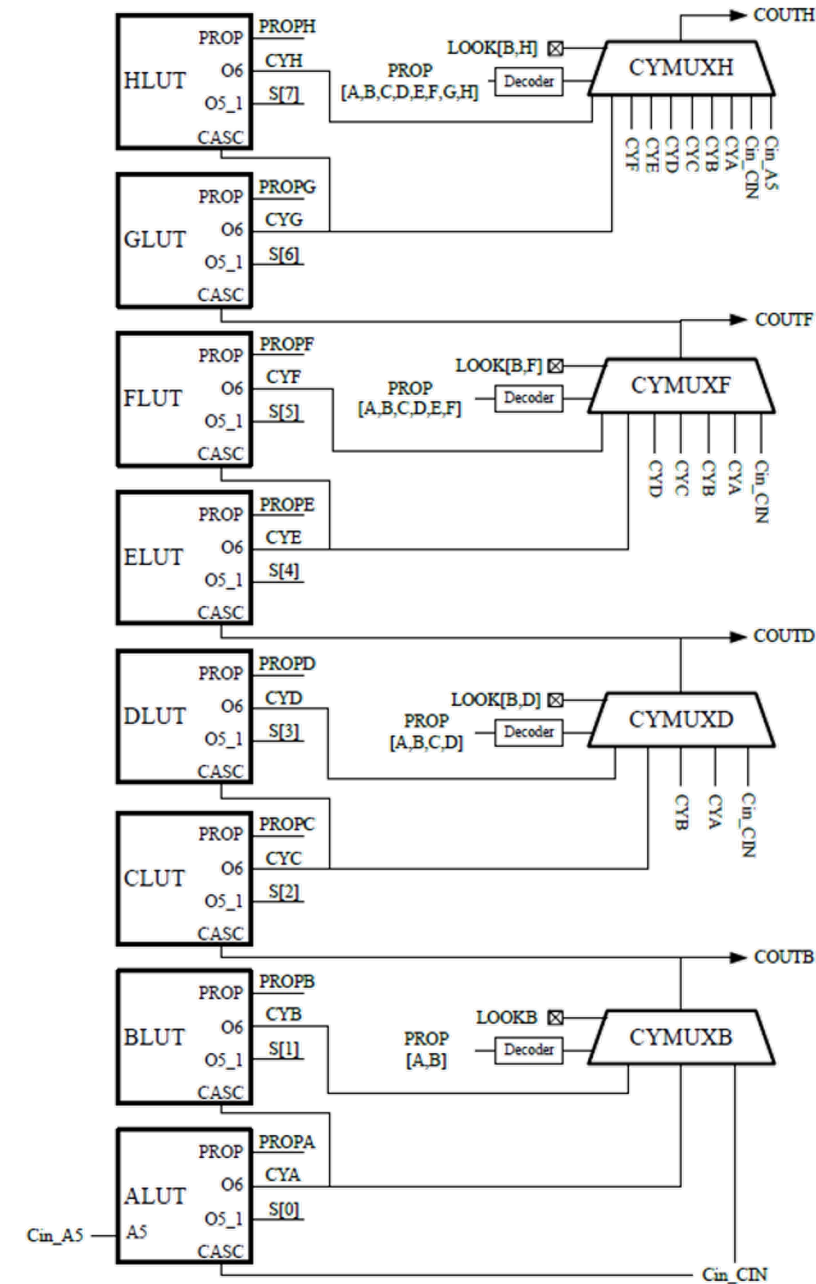
# GPCs on *Versal* CLBs



Proposed (9 : 4, 1) counter

counter	LUT cost	E	S
Proposed (9 : 4,1)	3	1.33	1.8
column counter (9 : 4,1) [6]	4	1	1.8
(10 : 4,2) [6]	3	1.33	1.67

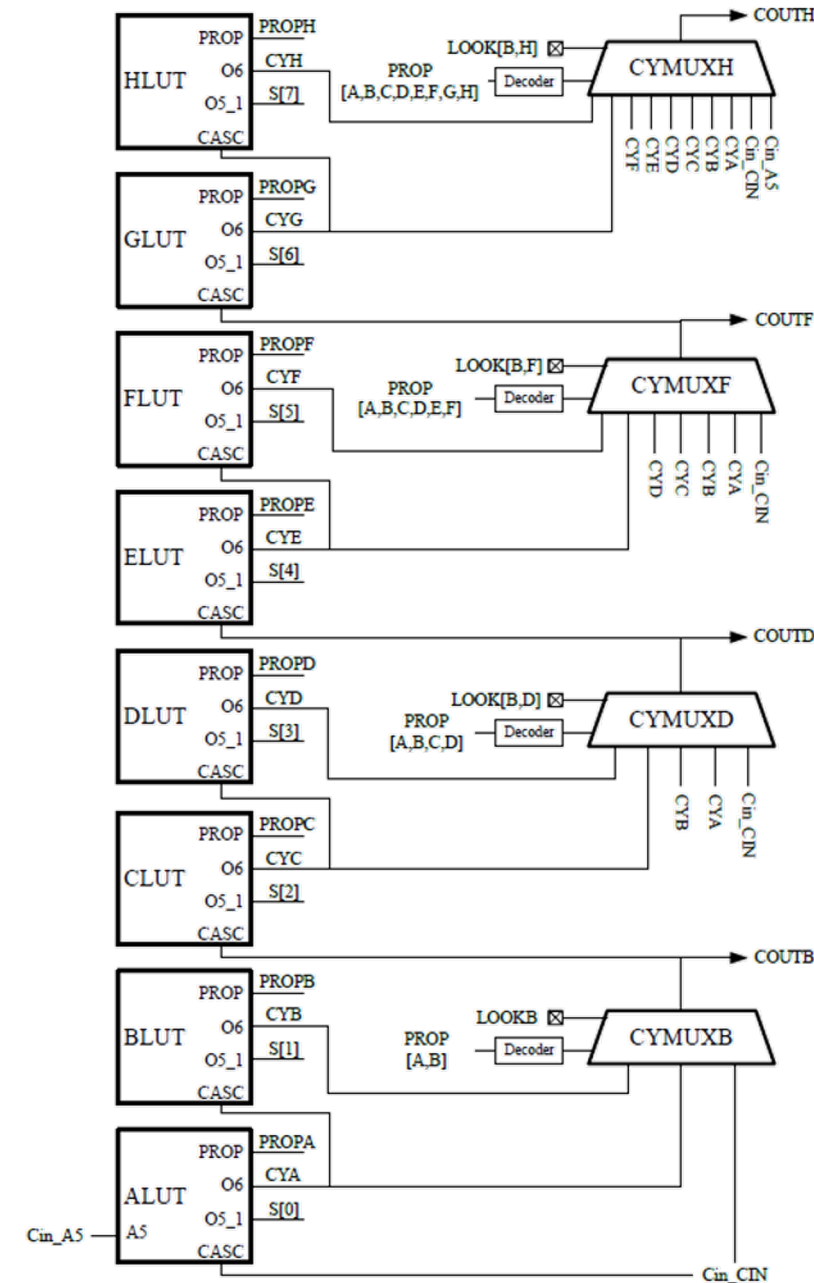
# Row Counter Design



# Row Counter Design

## Goal:

- confine the signal propagation of row counters within the LOOKAHEAD8 blocks
- No use of general-purpose routing

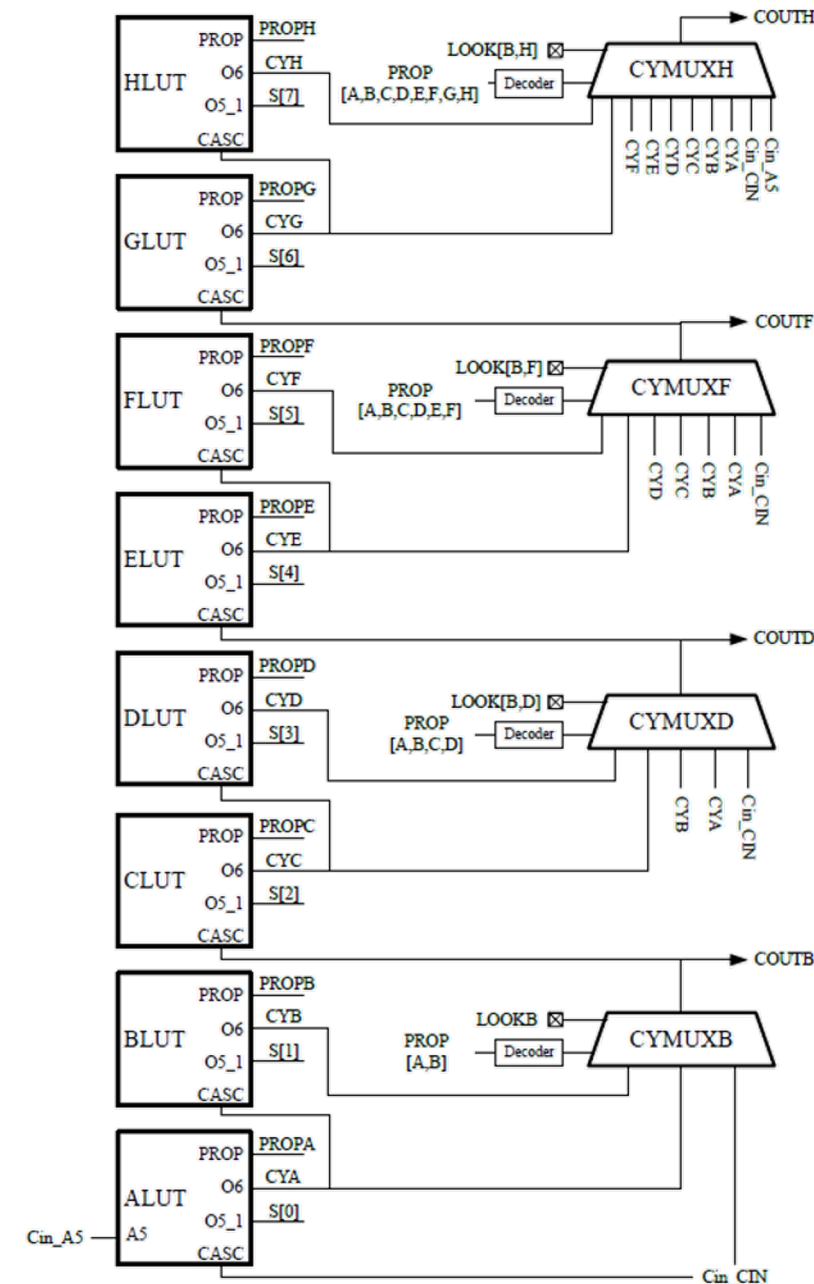


# Row Counter Design

## Goal:

- confine the signal propagation of row counters within the LOOKAHEAD8 blocks
- No use of general-purpose routing

Design under LOOKAHEAD8 constraints:



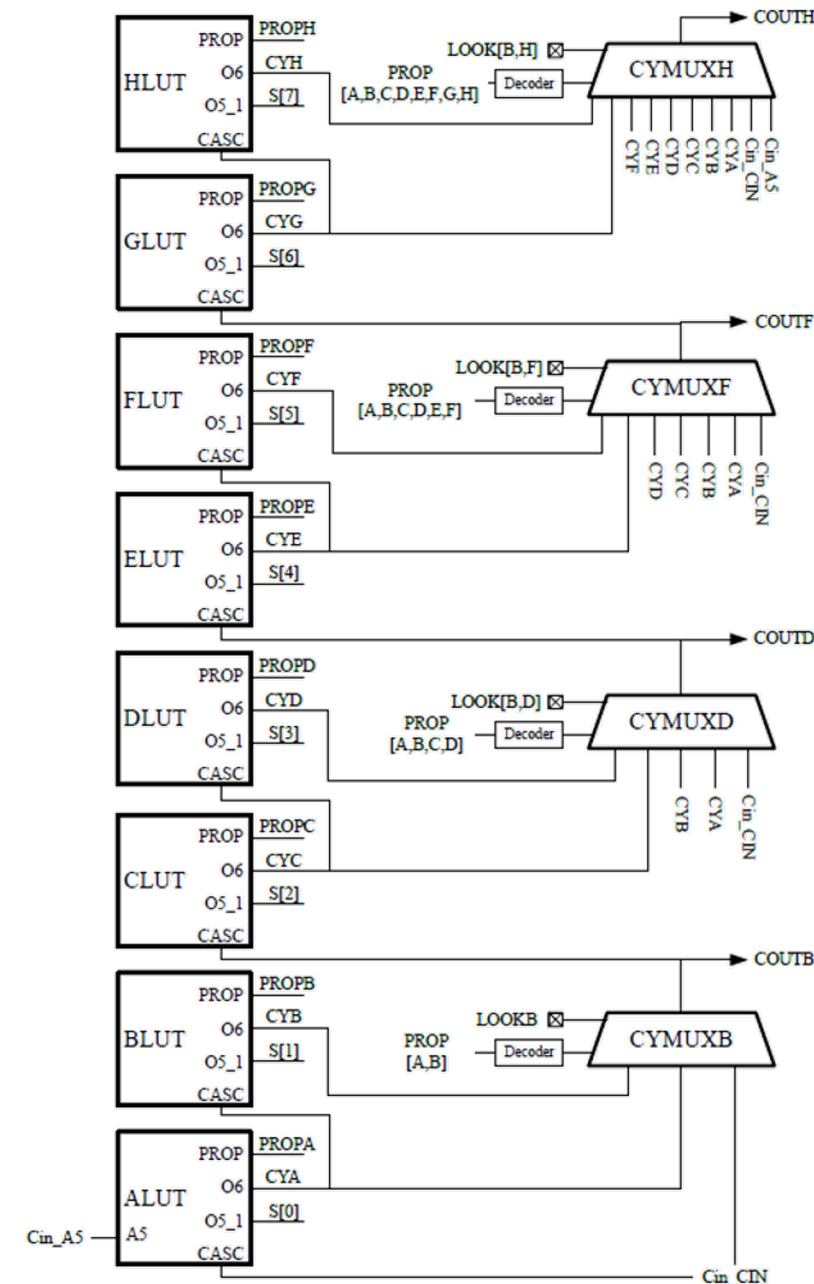
# Row Counter Design

## Goal:

- confine the signal propagation of row counters within the LOOKAHEAD8 blocks
- No use of general-purpose routing

## Design under LOOKAHEAD8 constraints:

- LOOKx Attributes



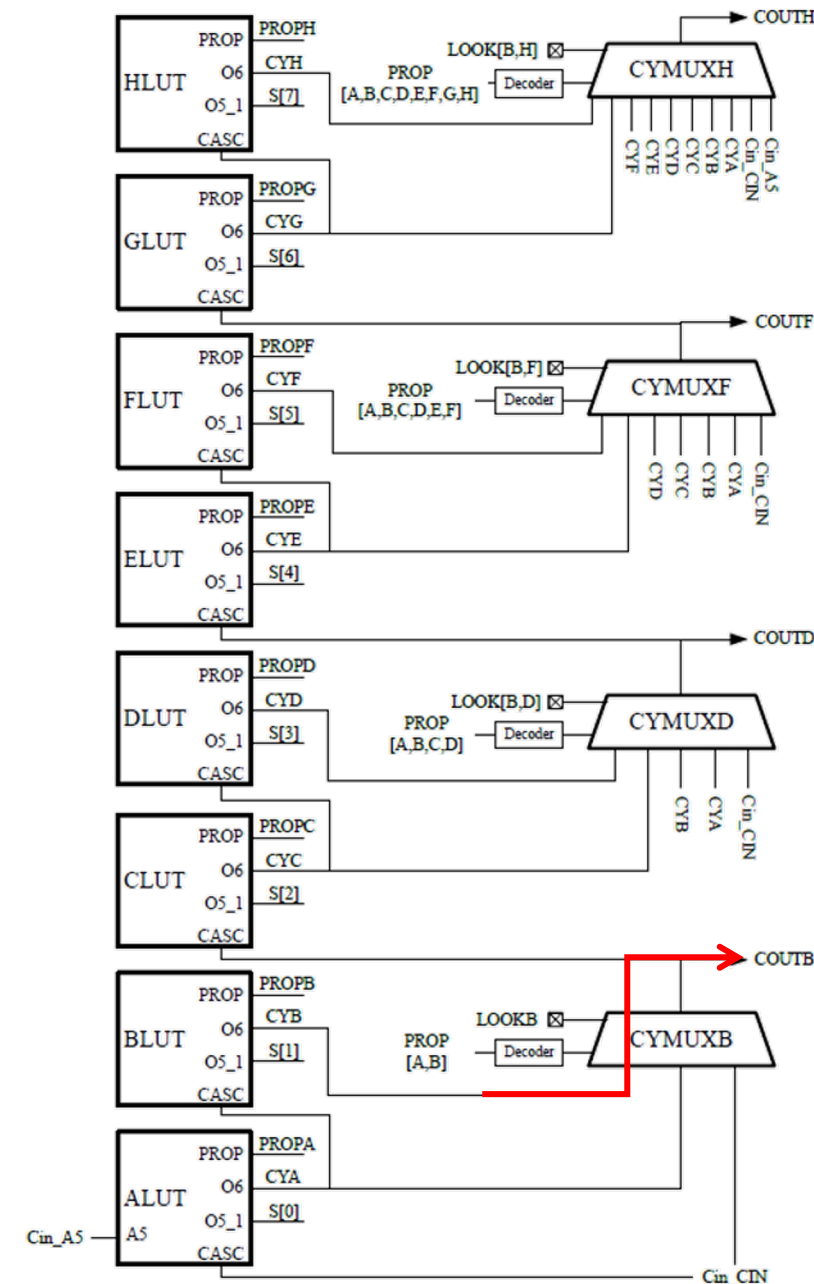
# Row Counter Design

## Goal:

- confine the signal propagation of row counters within the LOOKAHEAD8 blocks
- No use of general-purpose routing

## Design under LOOKAHEAD8 constraints:

- LOOKx Attributes



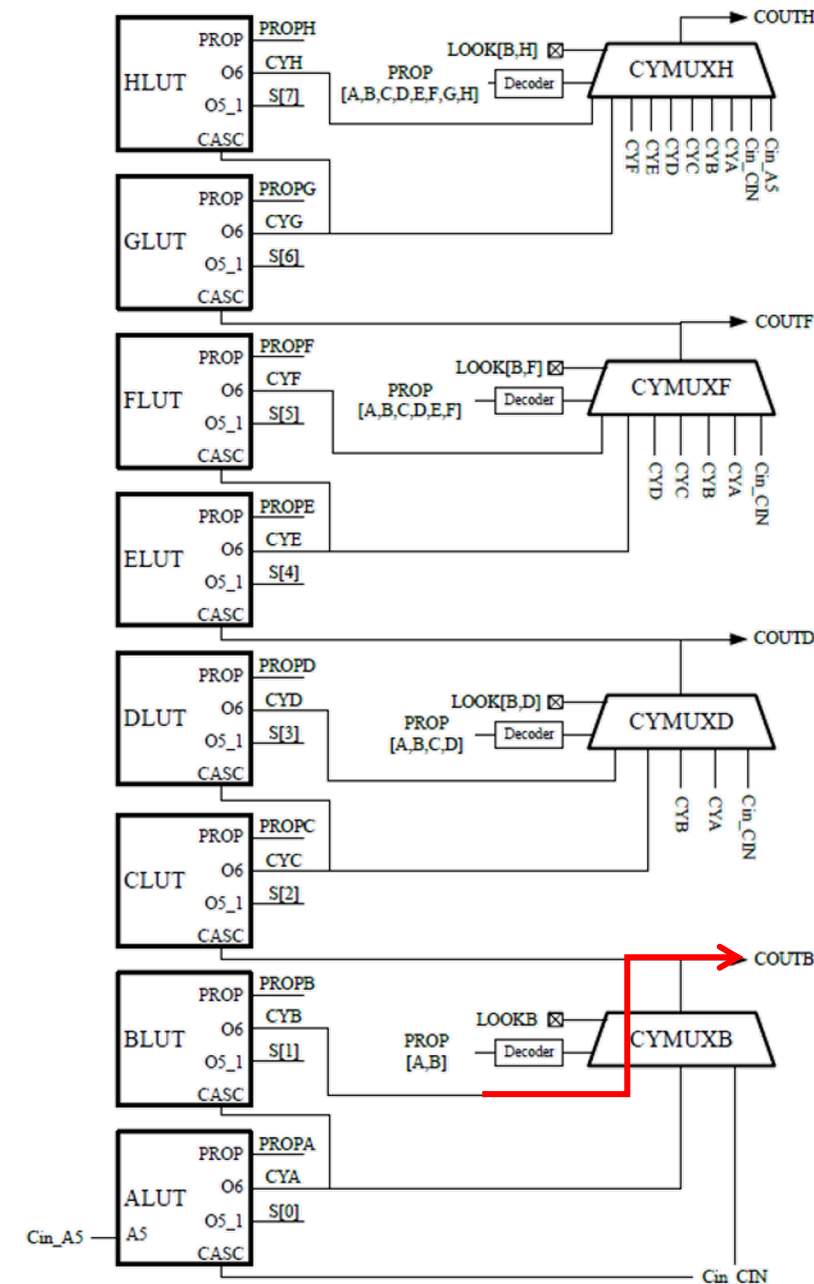
# Row Counter Design

## Goal:

- confine the signal propagation of row counters within the LOOKAHEAD8 blocks
- No use of general-purpose routing

## Design under LOOKAHEAD8 constraints:

- LOOKx Attributes
- Use the dual-rail property of the (3,9 : 2,3,1) counter to mitigate the row counter length limit brought by LOOKAHEAD8



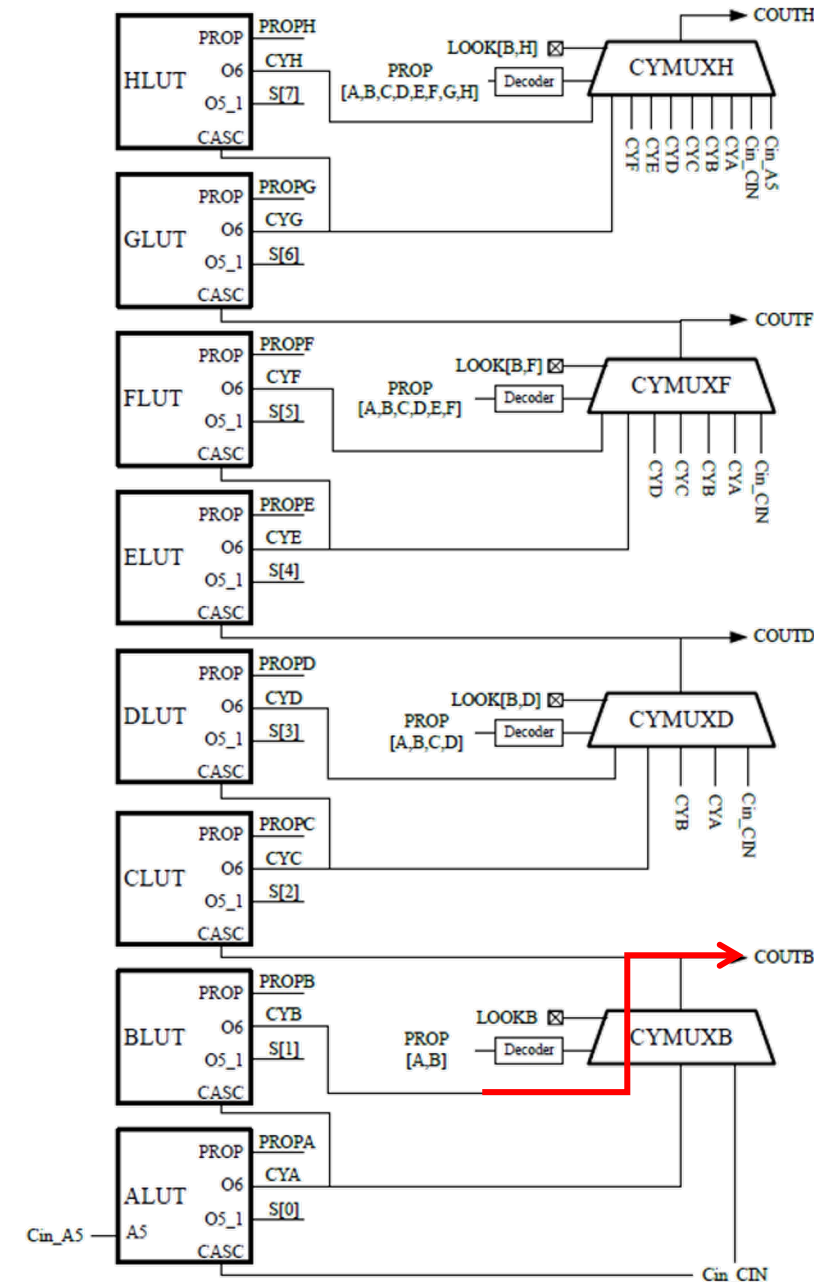
# Row Counter Design

## Goal:

- confine the signal propagation of row counters within the LOOKAHEAD8 blocks
- No use of general-purpose routing

## Design under LOOKAHEAD8 constraints:

- LOOKx Attributes
- Use the dual-rail property of the (3,9 : 2,3,1) counter to mitigate the row counter length limit brought by LOOKAHEAD8
- .....



# New heuristic for compressor tree synthesis

A counter will be selected and placed when it is both *applicable* and *necessary*

- *Applicable*: there are enough input bits to fully populate the counter

# New heuristic for compressor tree synthesis

A counter will be selected and placed when it is both *applicable* and *necessary*

- *Applicable*: there are enough input bits to fully populate the counter
- *Necessary*: it is the best counter on the list giving the local shape of the bit heap

Counter	Necessity Conditions
(5, 17 : 4, 5, 1)	Always Necessary
(4, 13 : 3, 4, 1)	$H_c \geq 16$
(3, 9 : 2, 3, 1)	$H_c \geq 12$
(9 : 4, 1)	$H_c \geq 12, 5 H_c > 17 H_{c+1}$
(6 : 3]	$H_c = 9, H_{c+1} \leq 3, H_{c+2} \leq 3$
(2, 2, 3 : 4]	$5 \leq H_c \leq 6, 4 \leq H_{c+1} \leq 5, 4 \leq H_{c+2} \leq 5$
(3 : 2]	$5 \leq H_c \leq 6$
(1, 5 : 3]	Always Necessary

# New heuristic for compressor tree synthesis

A counter will be selected and placed when it is both applicable and necessary

- Applicable: there are enough input bits to fully populate the counter
- Necessary: it is the best counter on the list giving the local shape of the bit heap

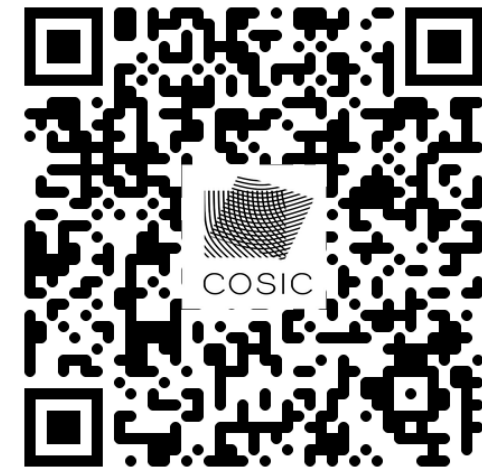
Counter	Necessity Conditions
(5, 17 : 4, 5, 1)	Always Necessary
(4, 13 : 3, 4, 1)	$H_c \geq 16$
(3, 9 : 2, 3, 1)	$H_c \geq 12$
(9 : 4, 1)	$H_c \geq 12, 5 H_c > 17 H_{c+1}$
(6 : 3]	$H_c = 9, H_{c+1} \leq 3, H_{c+2} \leq 3$
(2, 2, 3 : 4]	$5 \leq H_c \leq 6, 4 \leq H_{c+1} \leq 5, 4 \leq H_{c+2} \leq 5$
(3 : 2]	$5 \leq H_c \leq 6$
(1, 5 : 3]	Always Necessary

- Row counter first: a counter will start/continue a row counter if possible

# Open-Source\* RTL-code Generator



# Open-Source\* RTL-code Generator



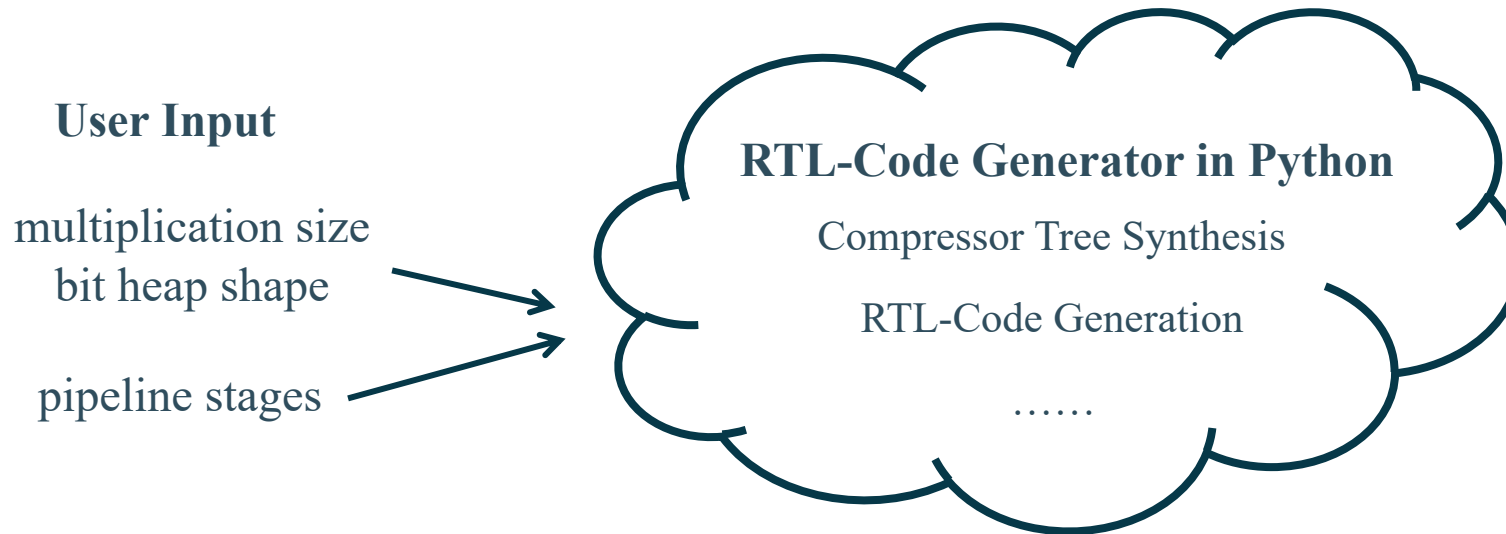
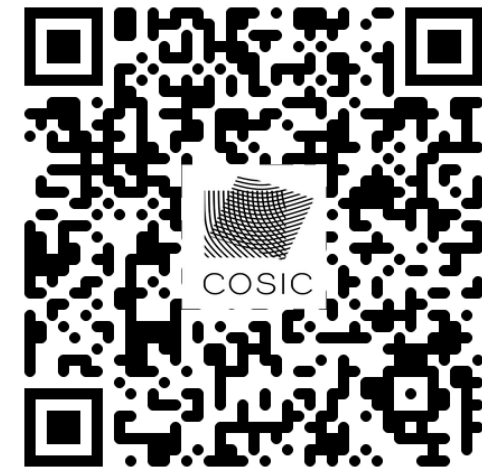
## RTL-Code Generator in Python

Compressor Tree Synthesis

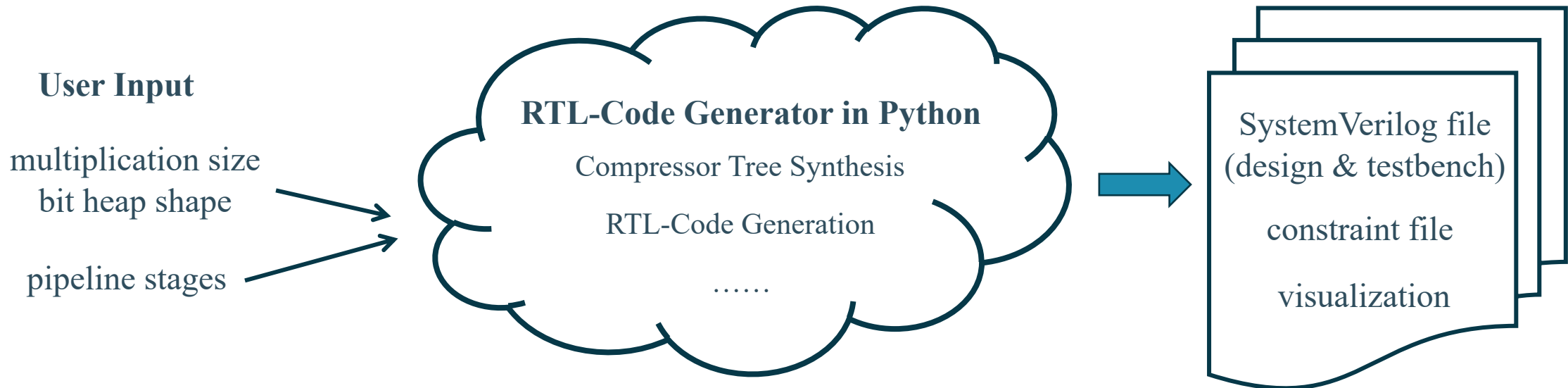
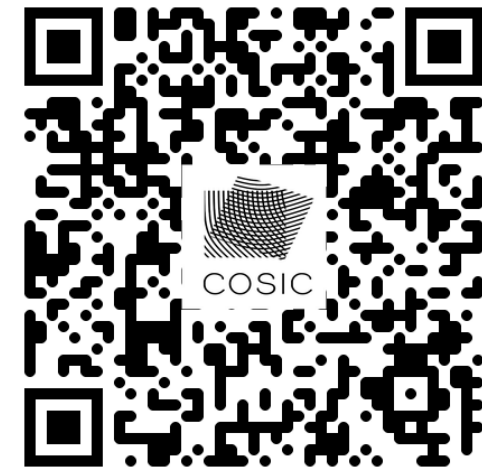
RTL-Code Generation

.....

# Open-Source\* RTL-code Generator



# Open-Source\* RTL-code Generator



# Open-Source\* RTL-code Generator

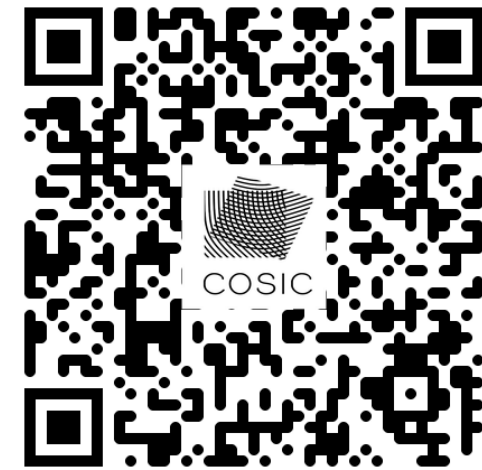
The generator is very easy to use!



# Open-Source\* RTL-code Generator

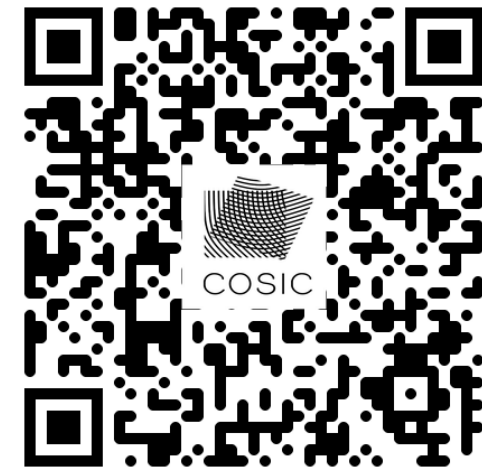
The generator is very easy to use!

```
(base) zetaomiao@Zetaos-MacBook-Air versal_arith % python cli.py -operator      bmult \  
> -width_a      16 \  
> -width_b      10 \  
> -pipeline_stages 2 \  
> -test_size    2000 \  
> -visualization True  
The LUT Usage without terminal addition is 23  
Maximum Number of Pipeline Stages of the Compressor Tree: 2  
Generated compressor tree is pipelined into 2 stages  
  
Output written to: versal_arith_generated/Bmult16x10/
```



# Open-Source\* RTL-code Generator

The generator is very easy to use!

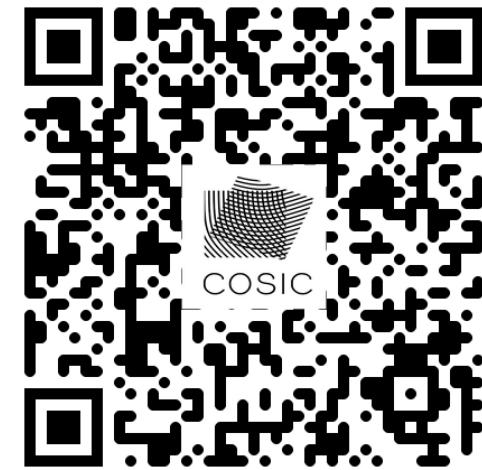


```
(base) zetaomiao@Zetaos-MacBook-Air versal_arith % python cli.py -operator      bmult \  
> -width_a      16 \  
> -width_b      10 \  
> -pipeline_stages 2 \  
> -test_size    2000 \  
> -visualization True  
The LUT Usage without terminal addition is 23  
Maximum Number of Pipeline Stages of the Compressor Tree: 2  
Generated compressor tree is pipelined into 2 stages  
Output written to: versal_arith_generated/Bmult16x10/
```

```
▼ versal_arith_generated/Bmult16x10  
  > bitheap_visualization  
  ▼ RTL_generated  
    ≡ Bmult16x10_bitheap_cmp_tb.sv  
    ≡ Bmult16x10_bitheap_cmp.sv  
    ≡ Bmult16x10_bitheap_gen.sv  
    ≡ Bmult16x10_tb.sv  
    ≡ Bmult16x10.sv  
  > testvectors  
  > xdc_generated  
    ≡ bitheap.txt
```

# Open-Source\* RTL-code Generator

The generator is very easy to use!



```
(base) zetaomiao@Zetaos-MacBook-Air versal_arith % python cli.py -operator      bmult \  
> -width_a      16 \  
> -width_b      10 \  
> -pipeline_stages 2 \  
> -test_size    2000 \  
> -visualization True  
The LUT Usage without terminal addition is 23  
Maximum Number of Pipeline Stages of the Compressor Tree: 2  
Generated compressor tree is pipelined into 2 stages  
Output written to: versal_arith_generated/Bmult16x10/
```

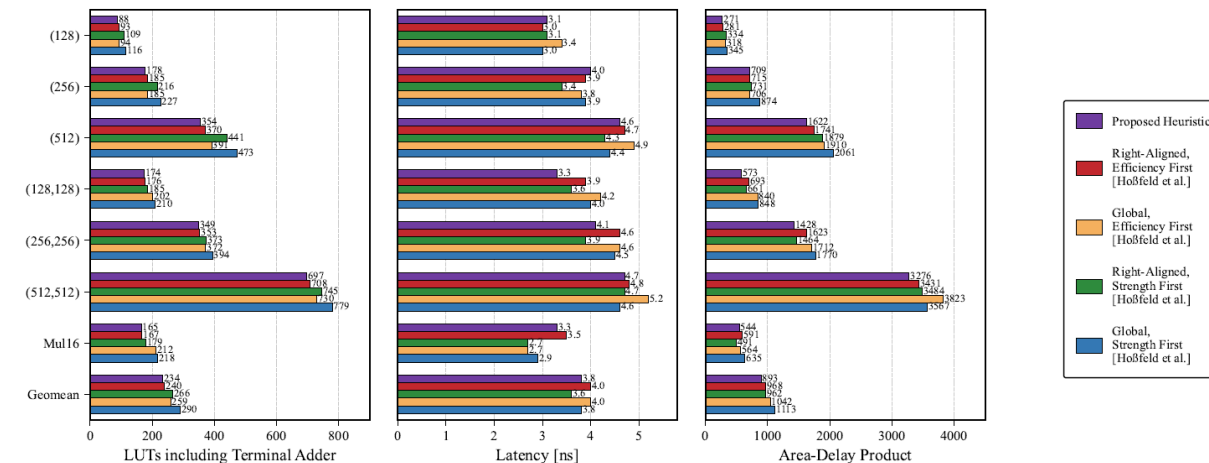
```
(base) zetaomiao@Zetaos-MacBook-Air versal_arith % python cli.py -operator      cmp \  
-txt_file_name  bitheap.txt\  
-pipeline_stages 2 \  
-test_size      2000 \  
-visualization  True  
The LUT Usage without terminal addition is 108  
Maximum Number of Pipeline Stages of the Compressor Tree: 4  
Generated compressor tree is pipelined into 2 stages  
Output written to: versal_arith_generated/bitheap cmp/
```

```
▼ versal_arith_generated/Bmult16x10  
  > bitheap_visualization  
  ▼ RTL_generated  
    ≡ Bmult16x10_bitheap_cmp_tb.sv  
    ≡ Bmult16x10_bitheap_cmp.sv  
    ≡ Bmult16x10_bitheap_gen.sv  
    ≡ Bmult16x10_tb.sv  
    ≡ Bmult16x10.sv  
  > testvectors  
  > xdc_generated  
    ≡ bitheap.txt
```

# Mapping Multiplication onto *Versal* CLBs

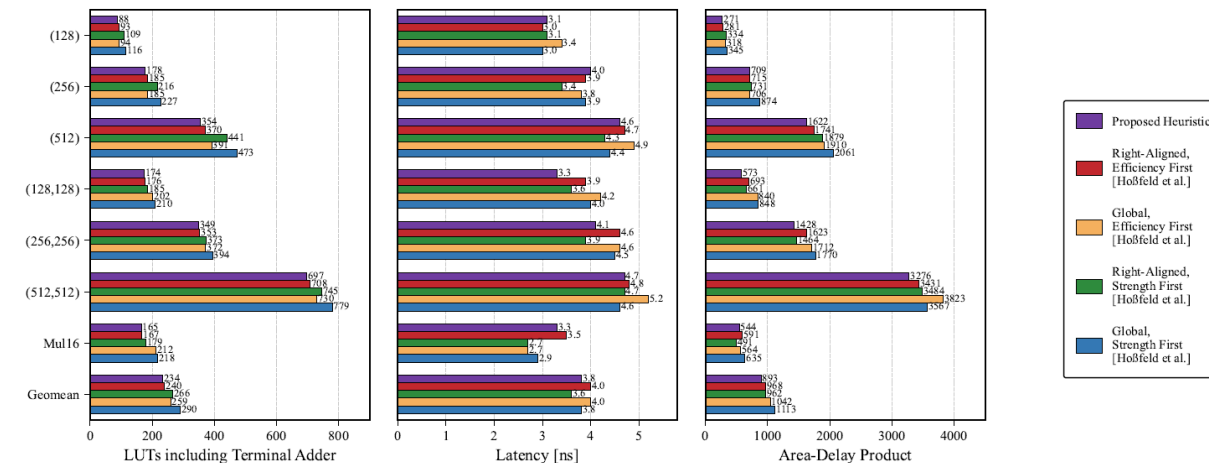
- Partial Product Generation
- Partial Product Addition/Compression
- Evaluation Results

# Evaluation Results – Compressor Trees



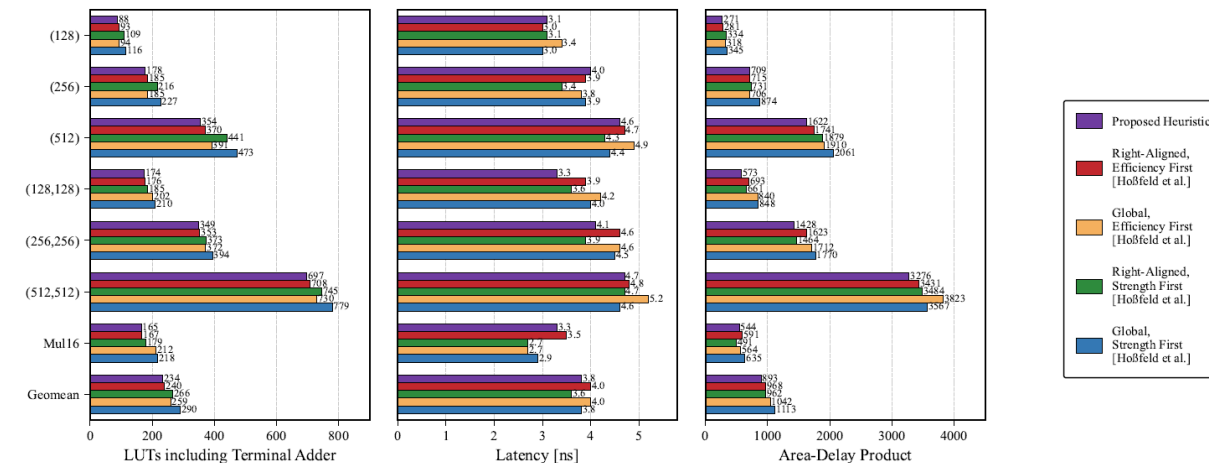
# Evaluation Results – Compressor Trees

- Highest Area Efficiency  
new GPCs & new heuristics



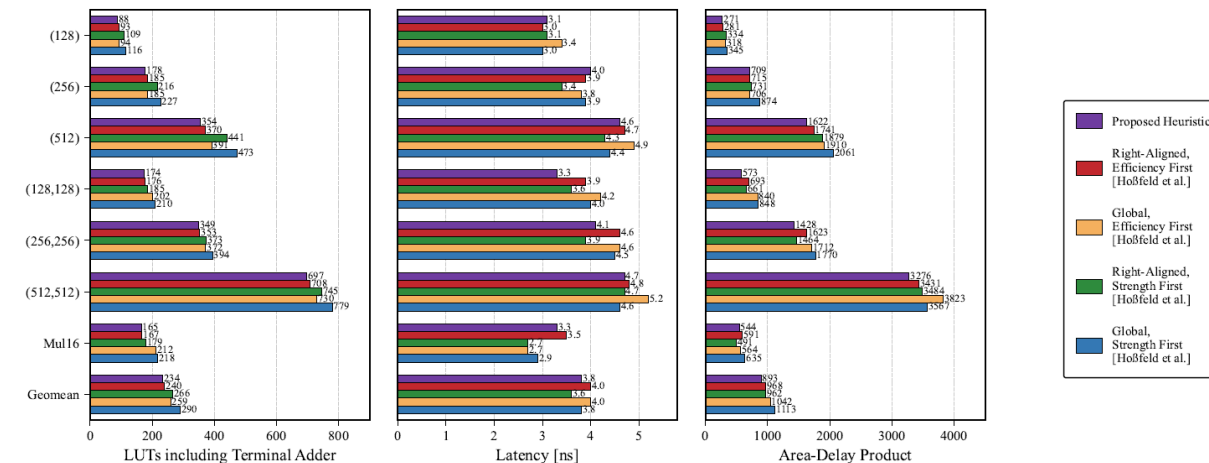
# Evaluation Results – Compressor Trees

- Highest Area Efficiency  
new GPCs & new heuristics
- Better Timing than Efficiency-First Heuristic in [6]  
row counter confines signal propagation in local LOOKAHEAD8 blocks



# Evaluation Results – Compressor Trees

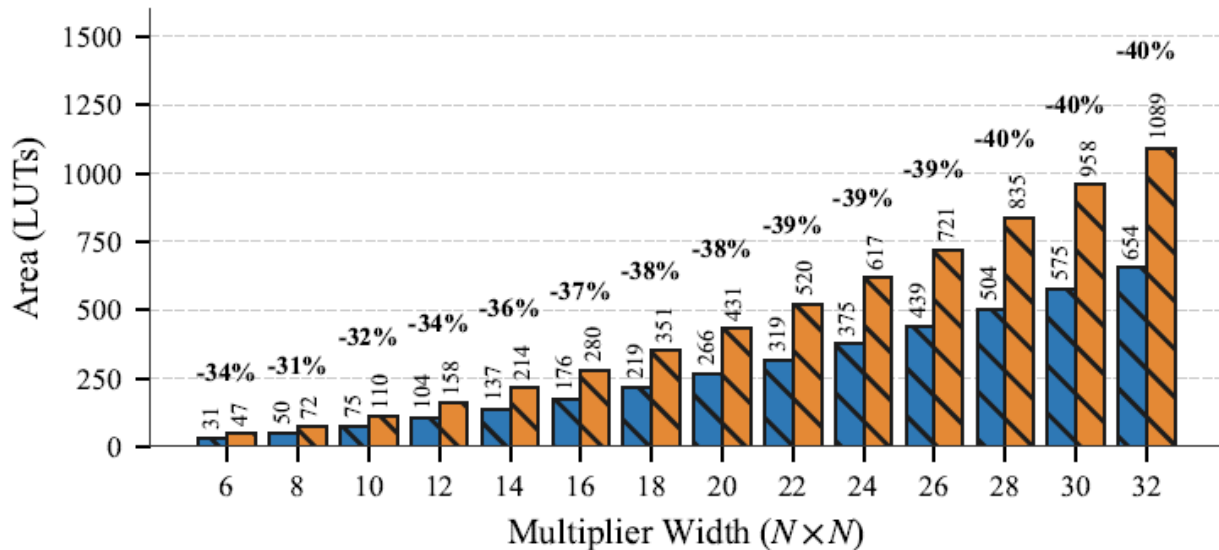
- Highest Area Efficiency  
new GPCs & new heuristics
- Better Timing than Efficiency-First Heuristic in [6]  
row counter confines signal propagation in local LOOKAHEAD8 blocks
- ADP Improvement: 8% ~ 20%



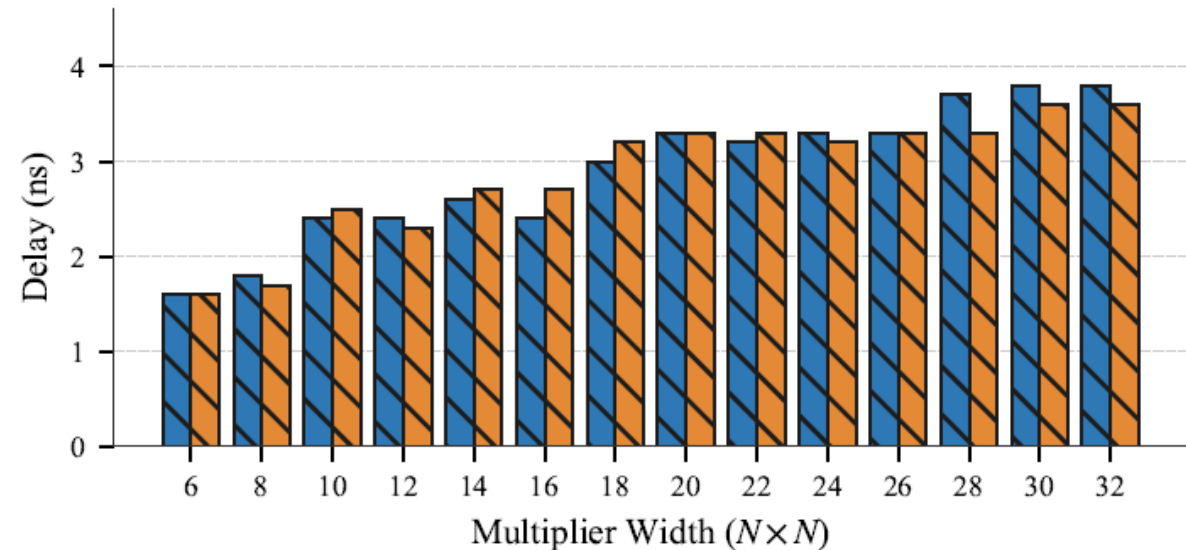
# Evaluation Results - Multipliers

(a) Area comparison

Proposed LogiCORE (Speed Opt.)



(b) Delay comparison

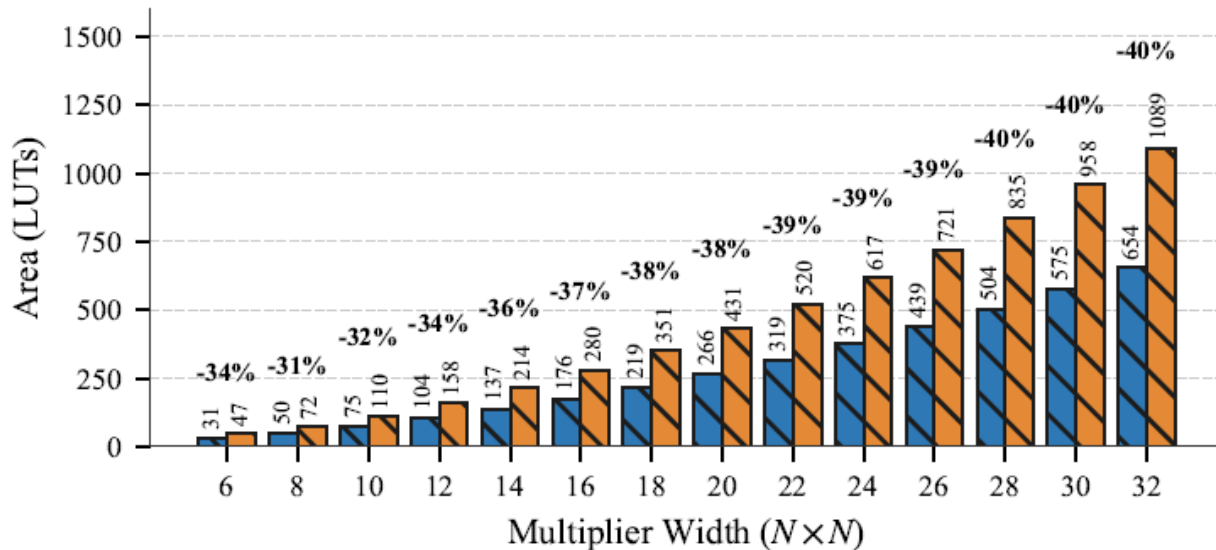


# Evaluation Results - Multipliers

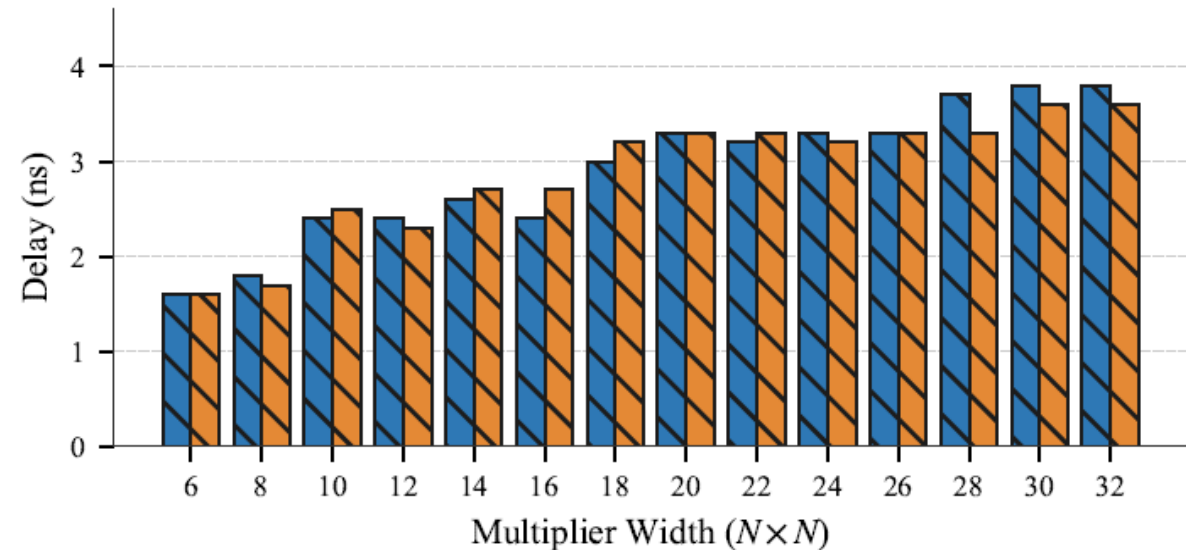
- LUT usage reduction up to 40%

(a) Area comparison

Proposed LogiCORE (Speed Opt.)

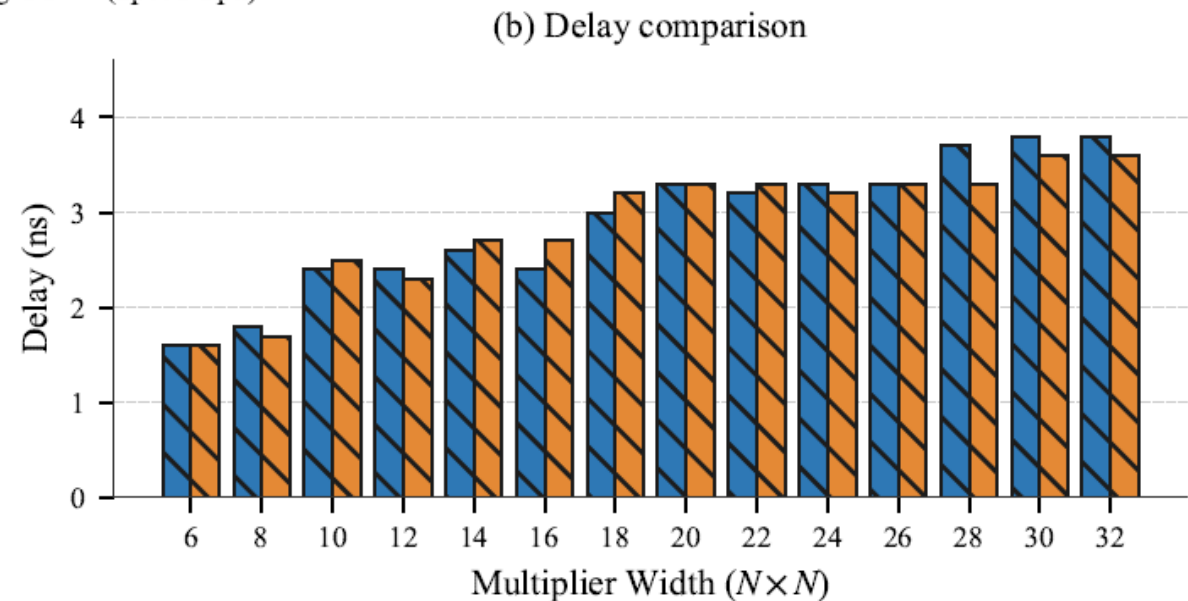
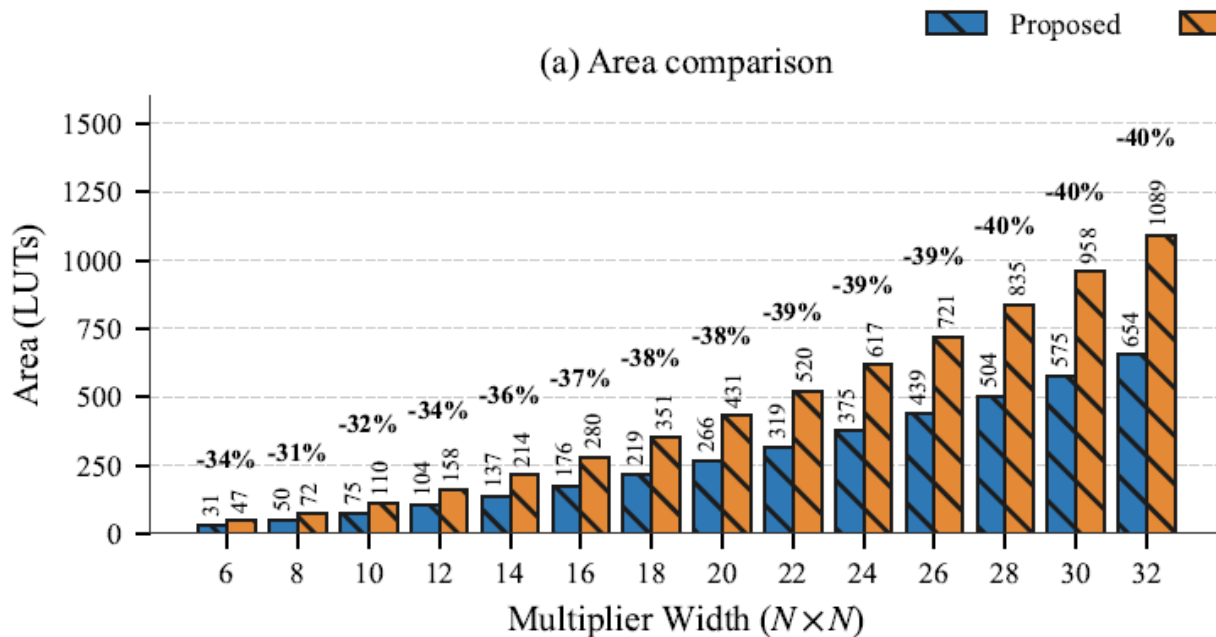


(b) Delay comparison



# Evaluation Results - Multipliers

- LUT usage reduction up to 40%
- Comparable critical path delay



# Conclusion

# Conclusion

- Area-Efficient LUT-Based Multipliers on *Versal* FPGAs

  - Efficient partial product generation using the new *Versal* LUT feature

  - Efficient *Versal* compressor tree synthesis for partial product compression

# Conclusion

- Area-Efficient LUT-Based Multipliers on *Versal* FPGAs

  - Efficient partial product generation using the new *Versal* LUT feature

  - Efficient *Versal* compressor tree synthesis for partial product compression

- Open-Source RTL-code generator for both compressors and multipliers



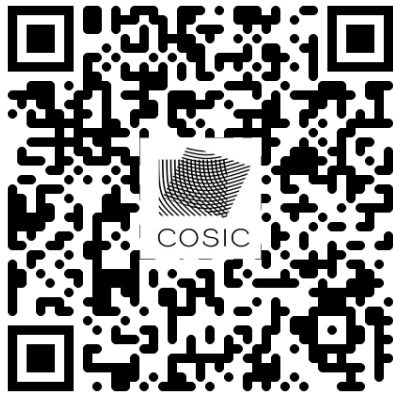
# Conclusion

- Area-Efficient LUT-Based Multipliers on *Versal* FPGAs

  - Efficient partial product generation using the new *Versal* LUT feature

  - Efficient *Versal* compressor tree synthesis for partial product compression

- Open-Source RTL-code generator for both compressors and multipliers



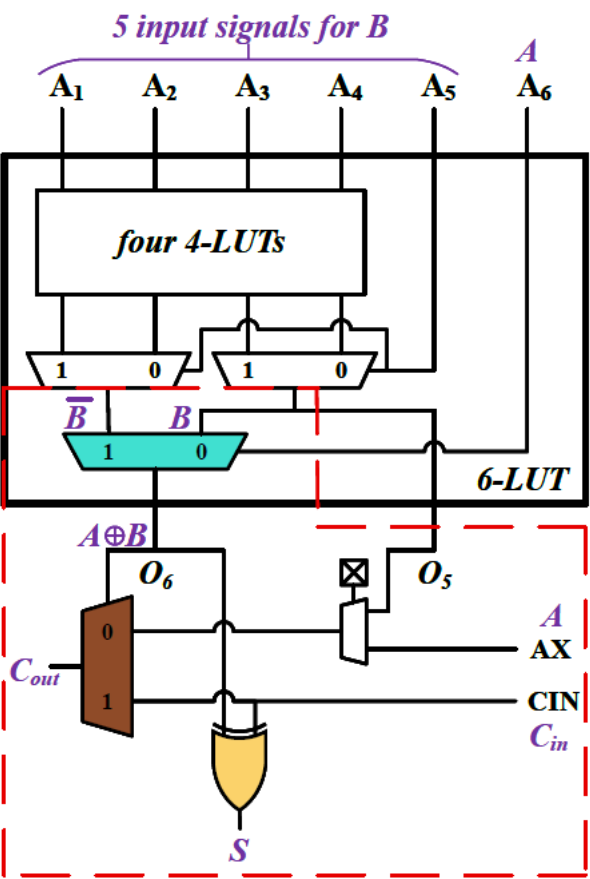
- More *Versal* (cryptographic) designs to come  
constant multipliers, NTTs, ...

# Area-Efficient LUT-Based Multipliers for AMD Versal FPGAs

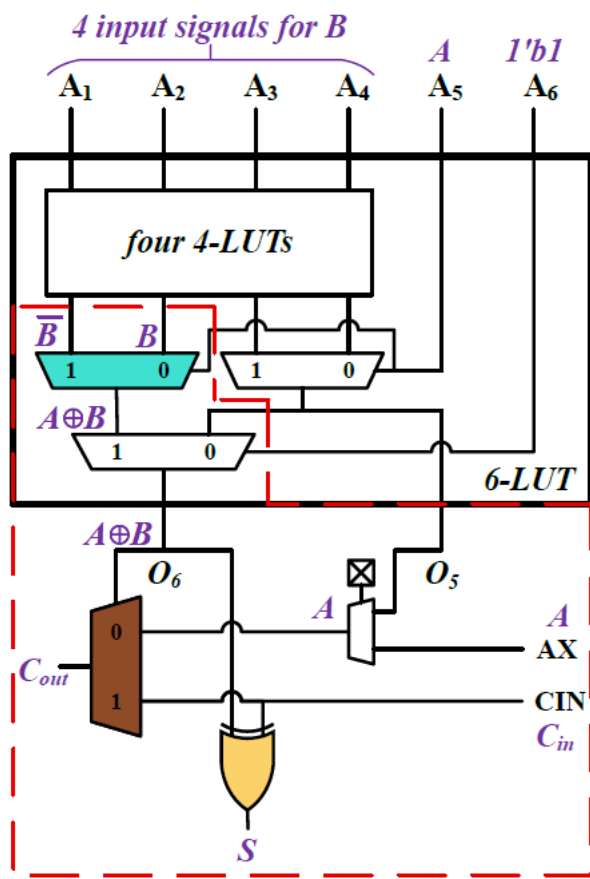
Zetao Miao, Xander Pottier, Jonas Bertels, Wouter Legiest and Ingrid Verbauwhede

*COSIC, KU Leuven, Belgium*

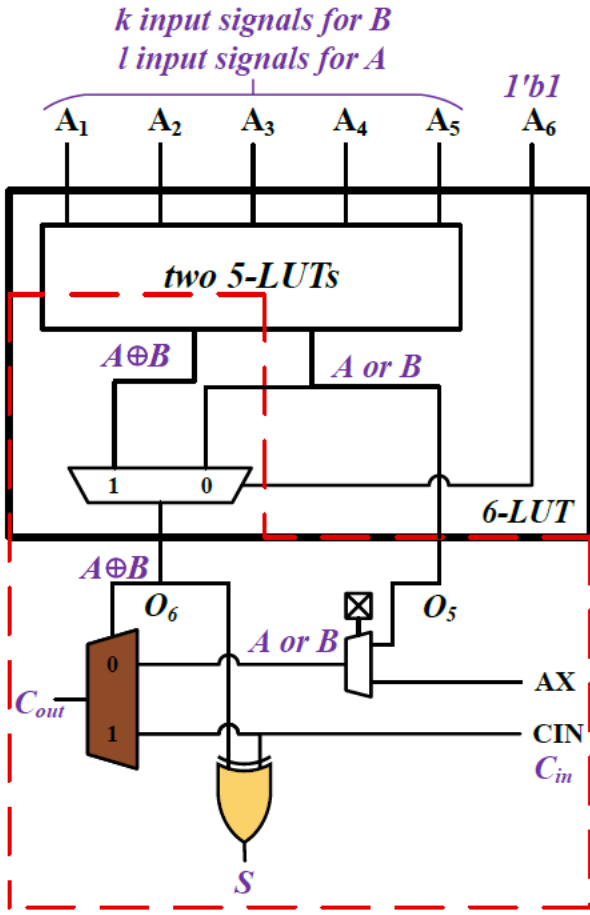
# Efficient LUT-Based Designs on *UltraScale*



(A)

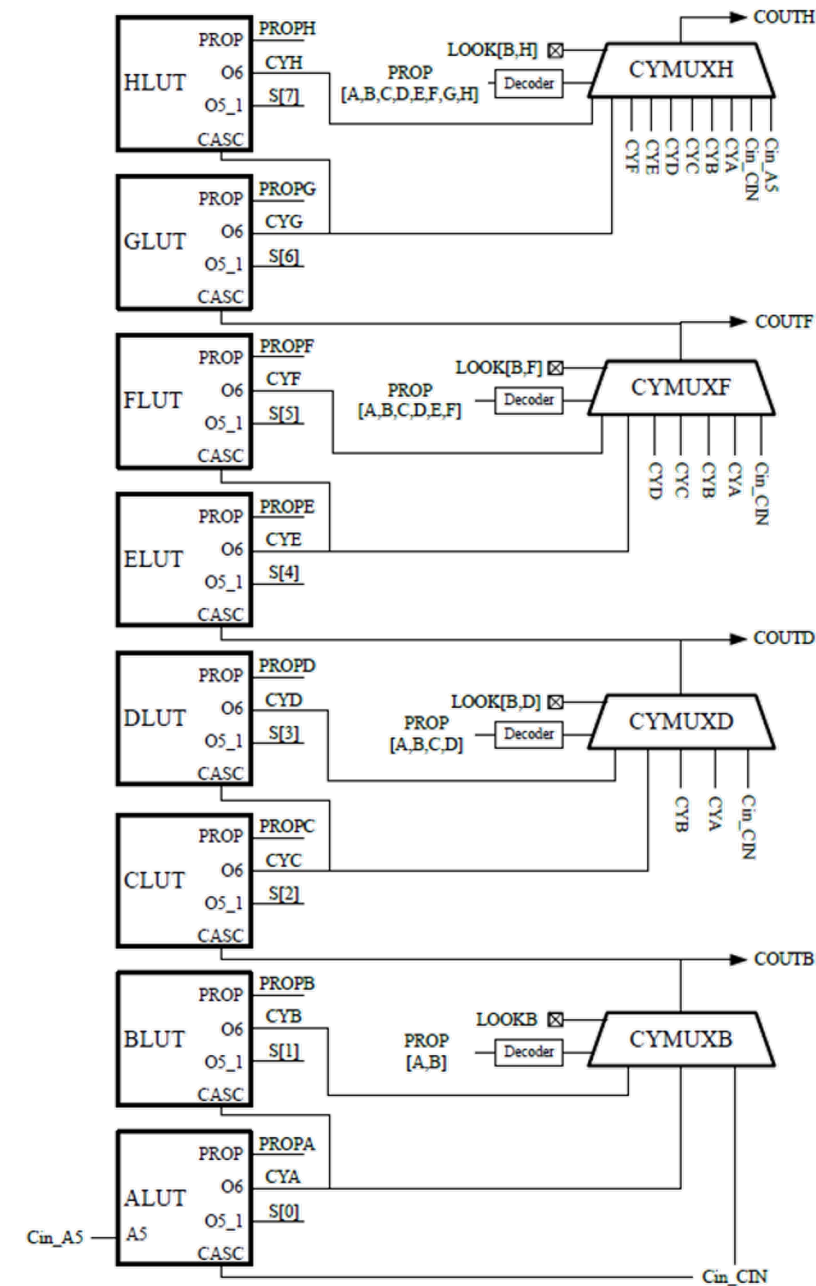


(B)



(C)

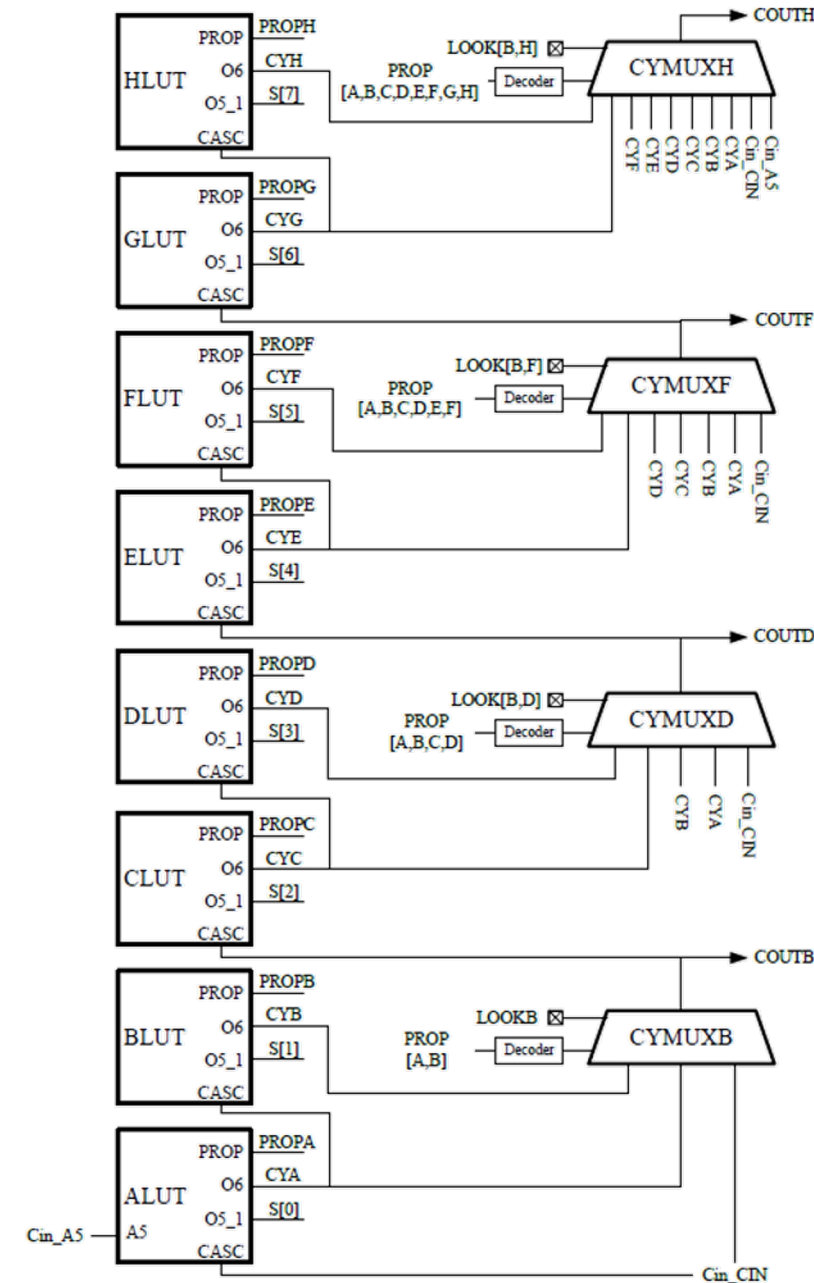
# Row Counter Design



# Row Counter Design

Goal:

- confine the carry propagation of row counters



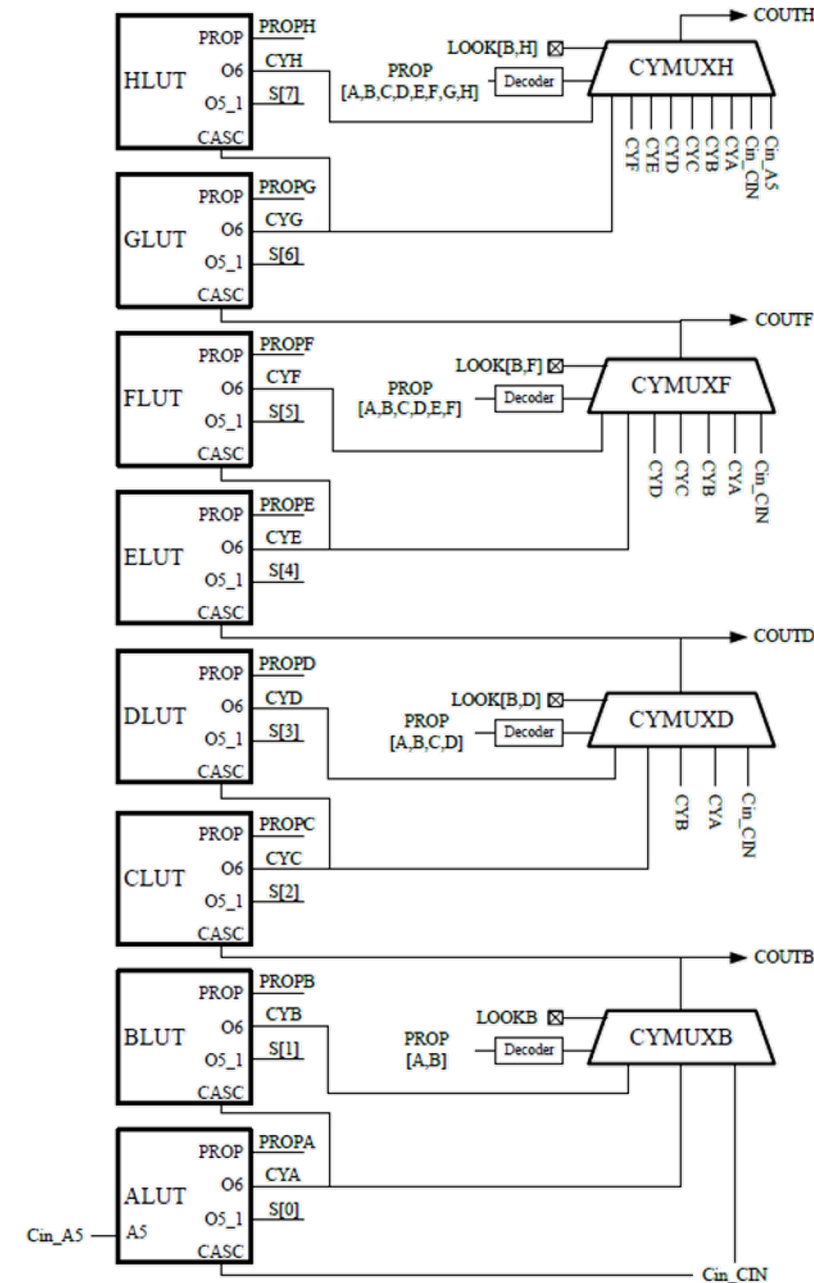
# Row Counter Design

Goal:

- confine the carry propagation of row counters

Observations:

- $LOOKB = FALSE \rightarrow COUTB = CYB$



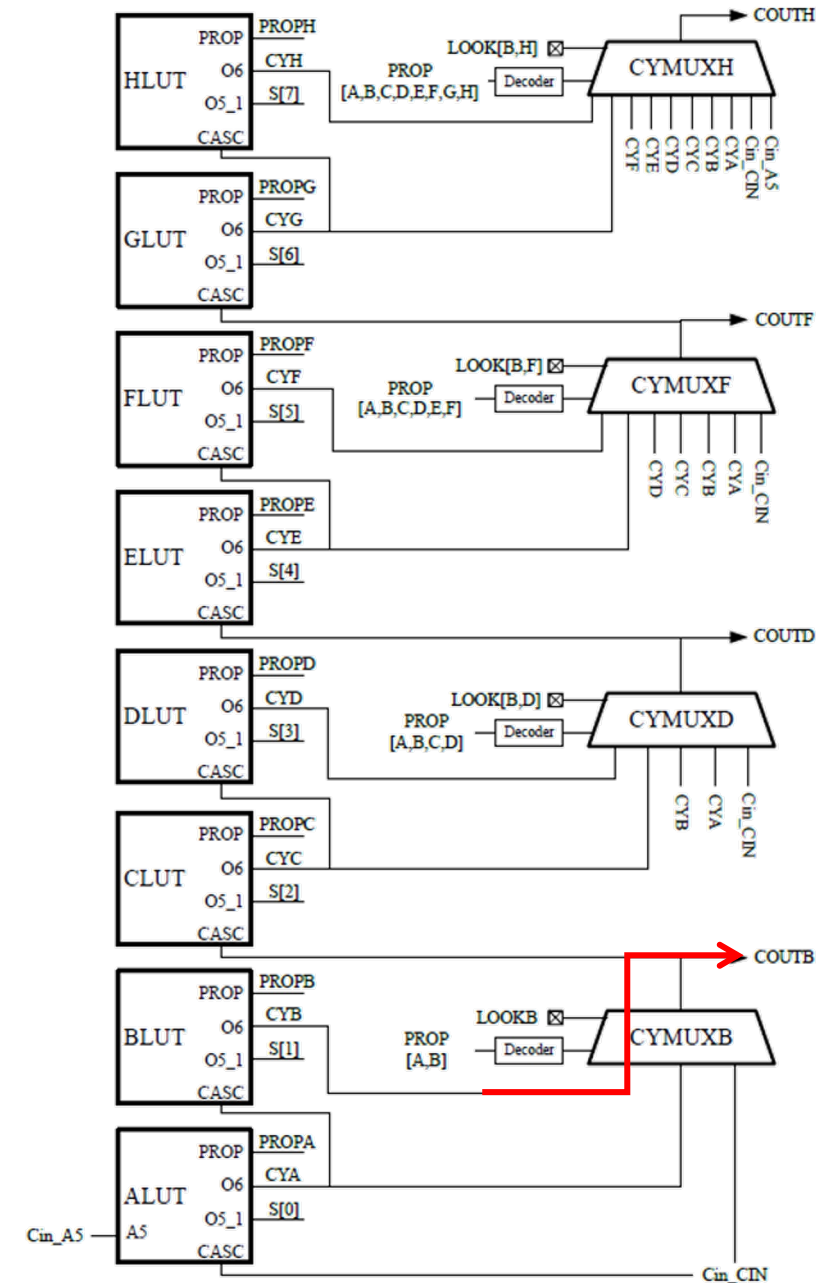
# Row Counter Design

Goal:

- confine the carry propagation of row counters

Observations:

- $LOOKB = FALSE \rightarrow COUTB = CYB$



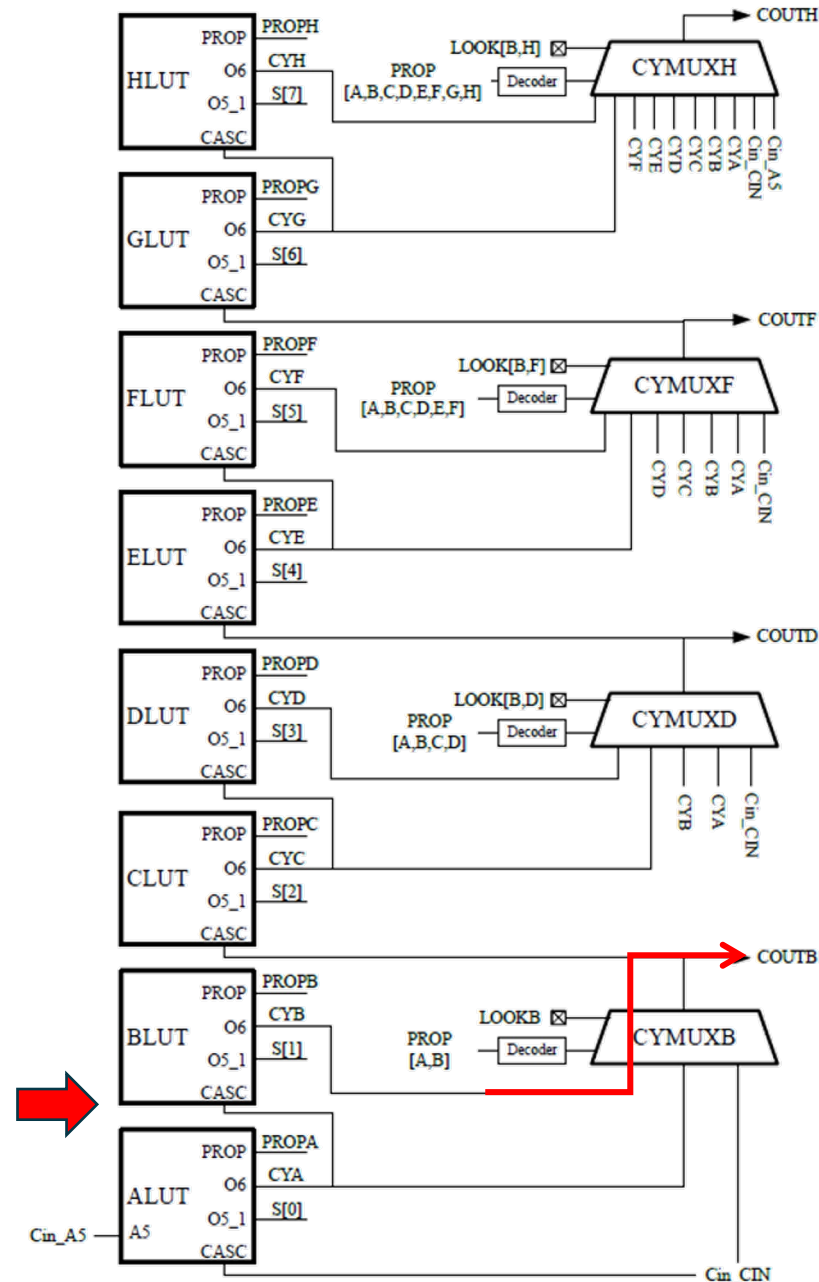
# Row Counter Design

Goal:

- confine the carry propagation of row counters

Observations:

- $LOOKB = FALSE \rightarrow COUTB = CYB$





# Row Counter Design

Goal:

- confine the carry propagation of row counters

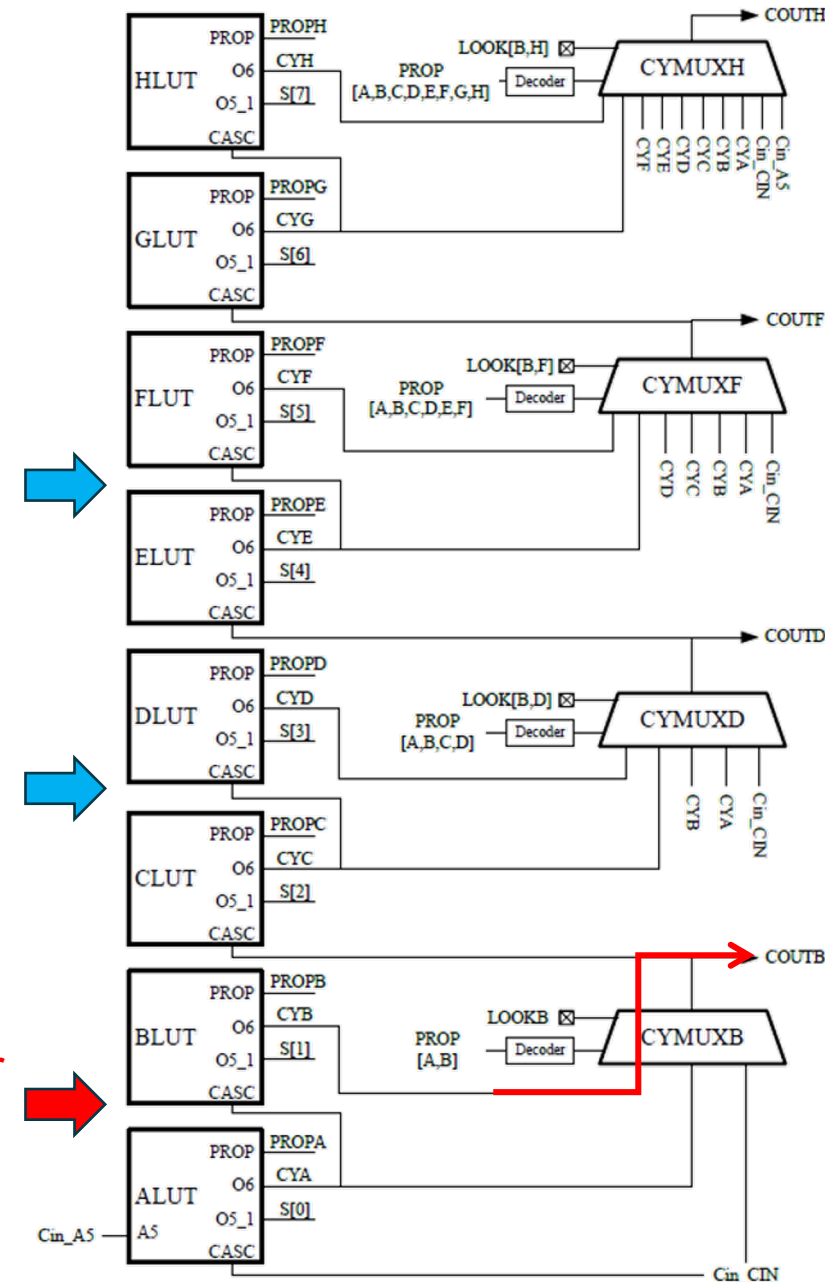
Observations:

- LOOKB = FALSE  $\rightarrow$  COUTB = CYB

LOOKAHEAD compatible candidate

LOOKAHEAD compatible candidate

any row counter candidate



# Row Counter Design

Goal:

- confine the carry propagation of row counters

Observations:

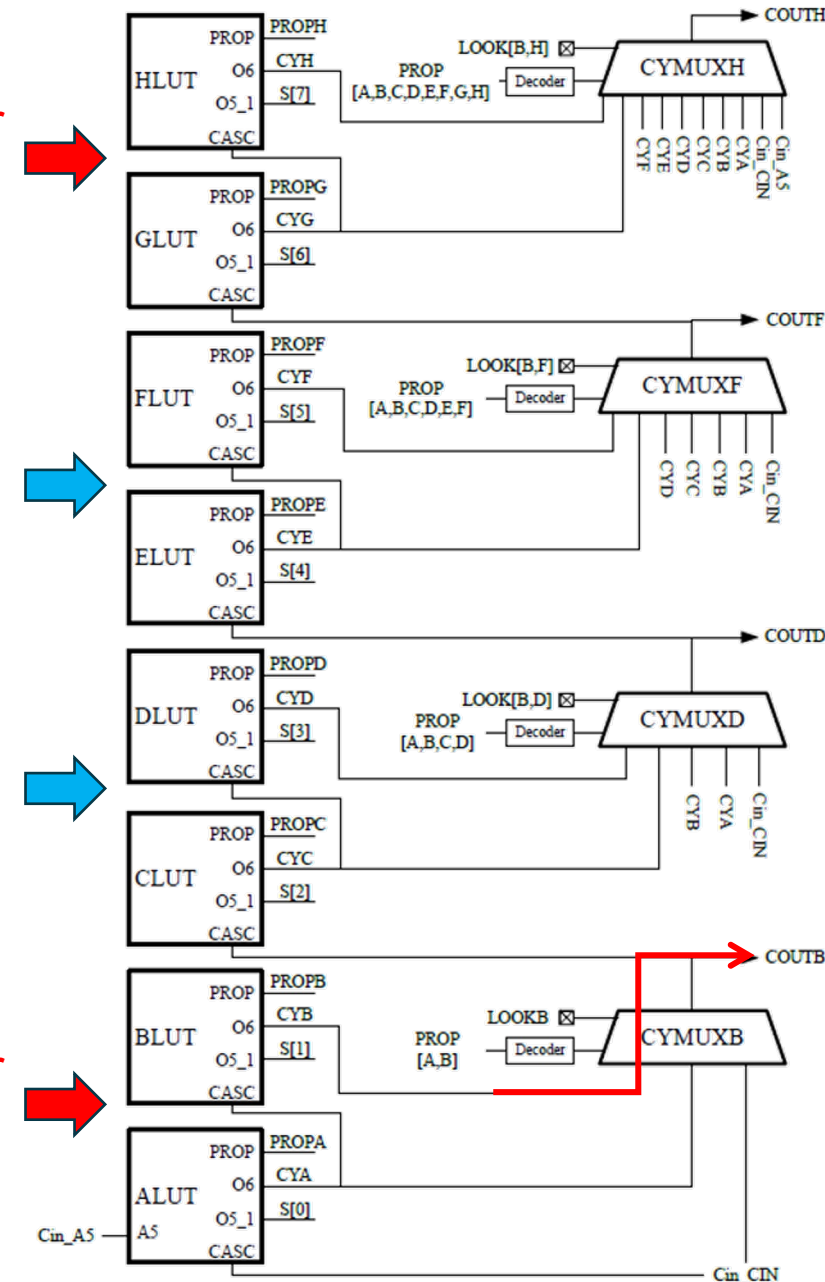
- LOOKB = FALSE  $\rightarrow$  COUTB = CYB
- Last GPC in row counter doesn't need to be compatible if it doesn't rely on CYMUX

any row counter candidate

LOOKAHEAD compatible candidate

LOOKAHEAD compatible candidate

any row counter candidate



# Row Counter Design

Goal:

- confine the carry propagation of row counters

Observations:

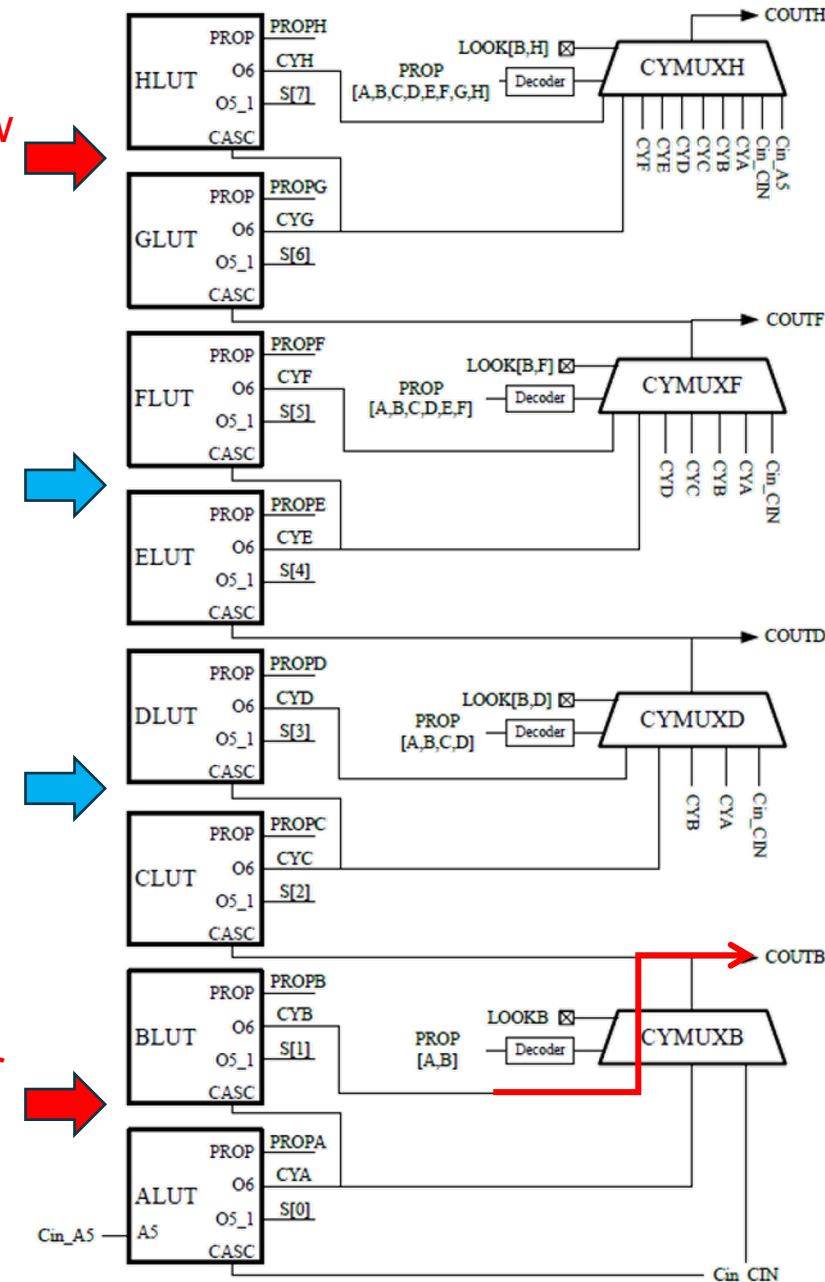
- LOOKB = FALSE  $\rightarrow$  COUTB = CYB
- Last GPC in row counter doesn't need to be compatible if it doesn't rely on CYMUX
- Length limit of 8 if the first GPC is incompatible

Last GPC in row counter

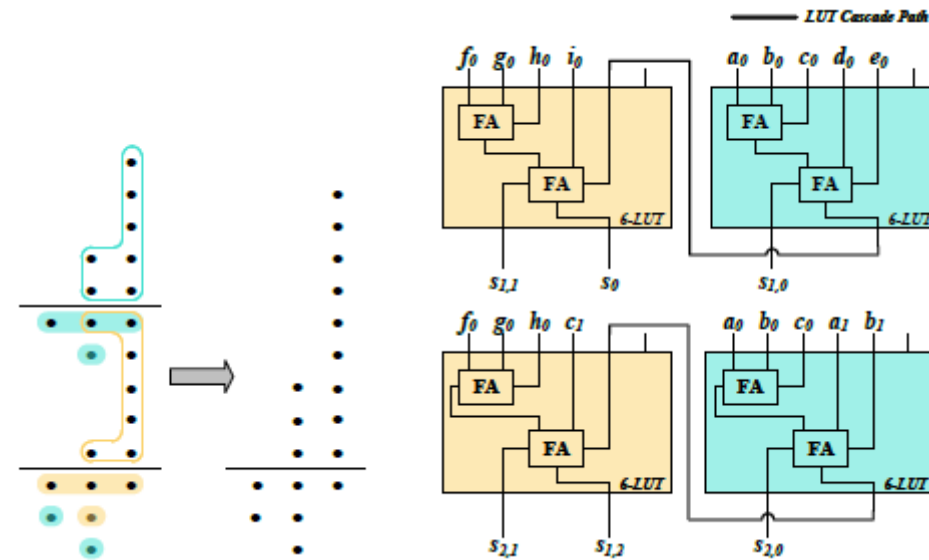
LOOKAHEAD compatible candidate

LOOKAHEAD compatible candidate

any row counter candidate

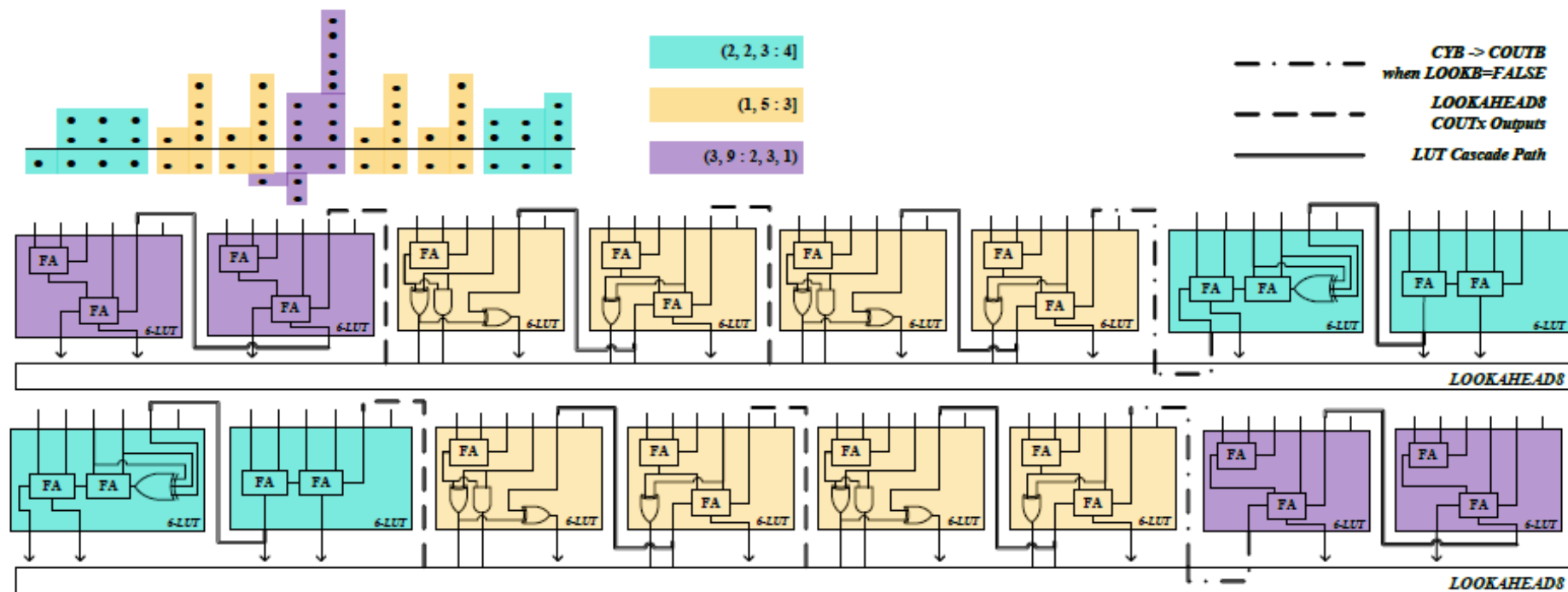


# (3,9 : 2,3,1) counter in row counter design

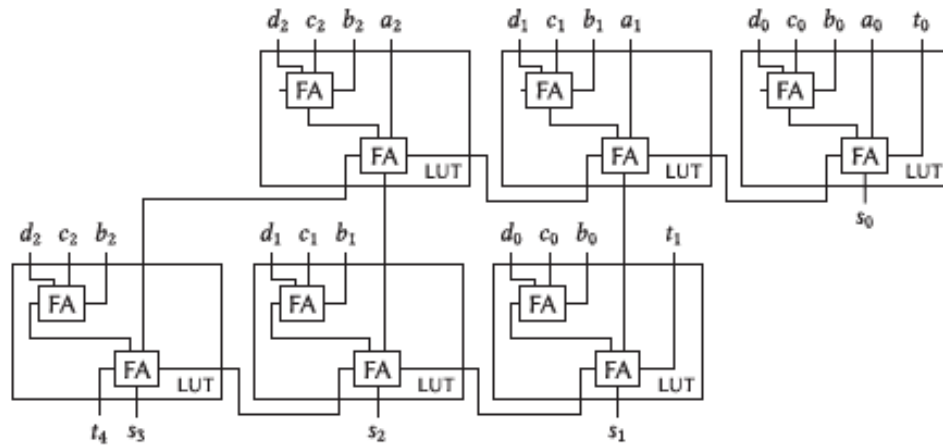


(3,9 : 2,3,1) counter

# (3,9 : 2,3,1) counter in row counter design

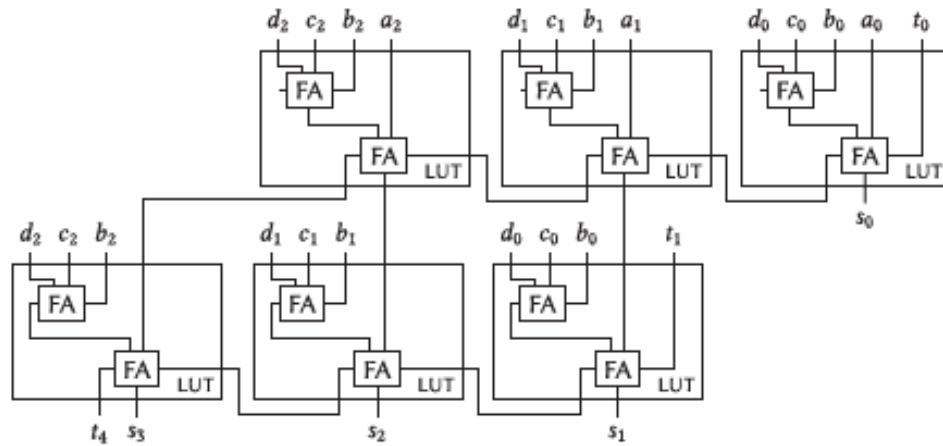


# Quaternary Terminal Addition



Quaternary Adder proposed by [6]

# Quaternary Terminal Addition

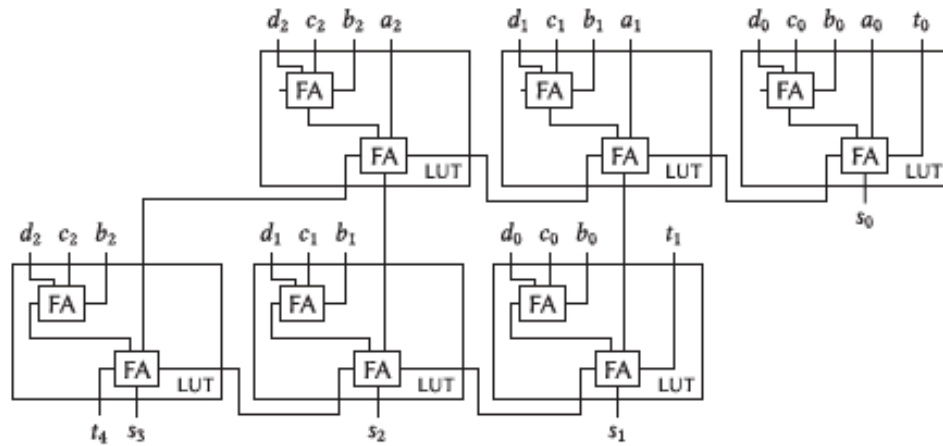


Quaternary Adder proposed by [6]

Adding 4 operands using 2 LUTs per bit

33% less LUT usage compared to Naïve +

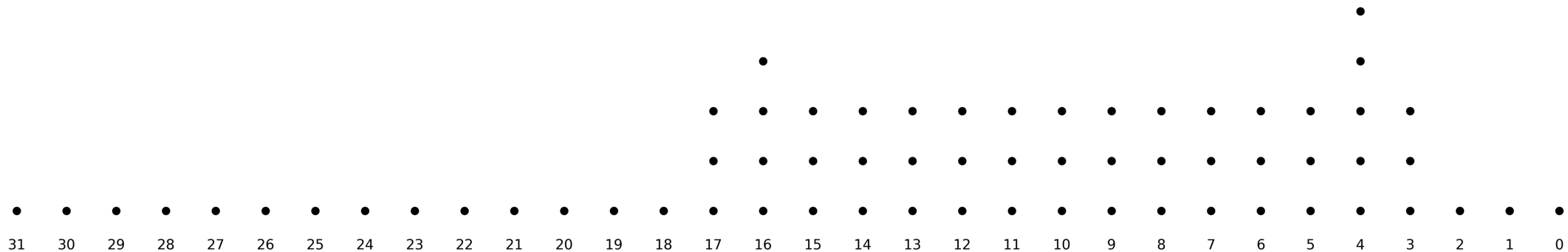
# Quaternary Terminal Addition



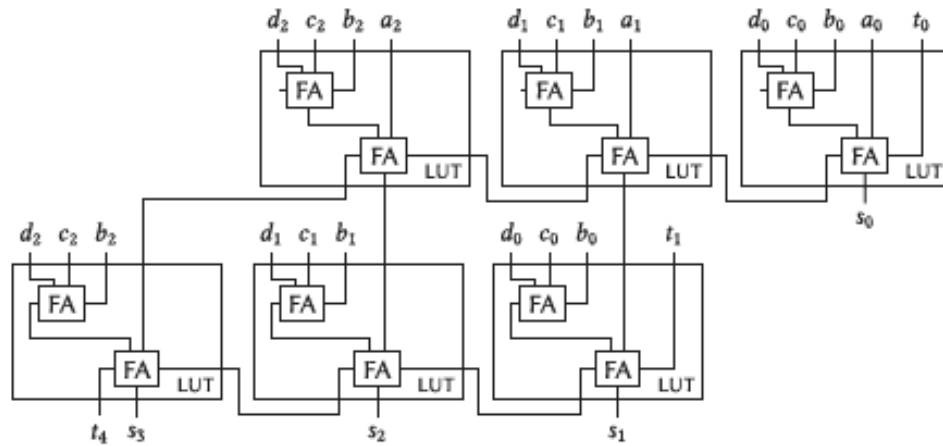
Quaternary Adder proposed by [6]

Adding 4 operands using 2 LUTs per bit

33% less LUT usage compared to Naïve +



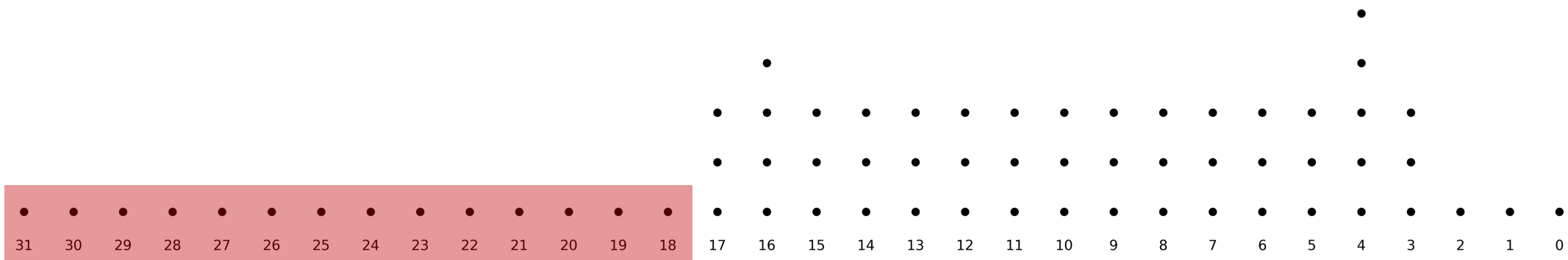
# Quaternary Terminal Addition



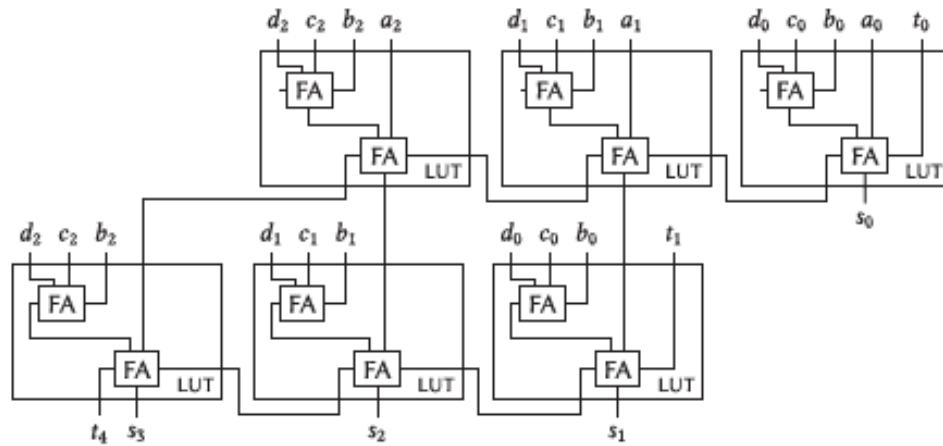
Quaternary Adder proposed by [6]

Adding 4 operands using 2 LUTs per bit

33% less LUT usage compared to Naïve +



# Quaternary Terminal Addition

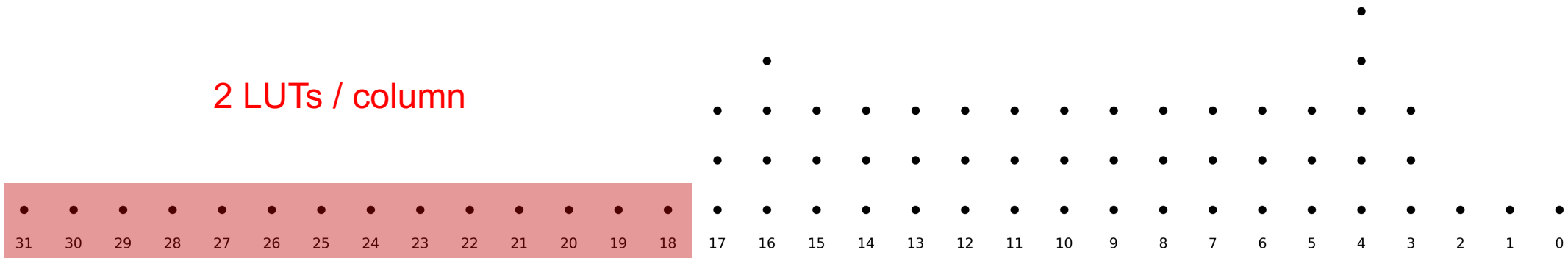


Quaternary Adder proposed by [6]

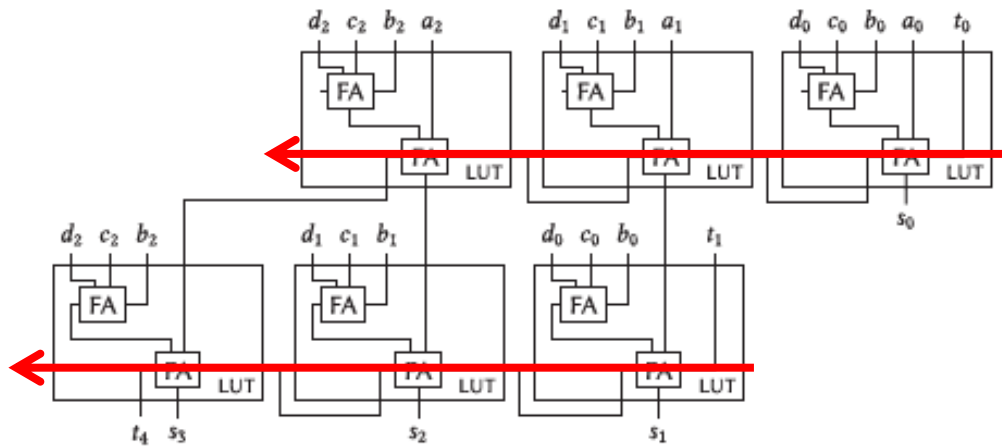
Adding 4 operands using 2 LUTs per bit  
 33% less LUT usage compared to Naïve +

2 LUTs / column

2 LUTs / column



# Quaternary Terminal Addition



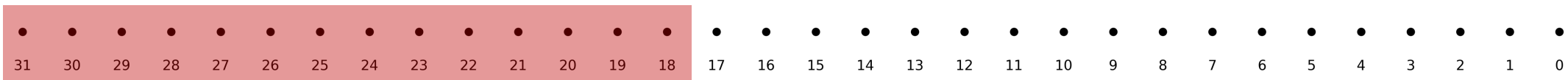
Quaternary Adder proposed by [6]

Adding 4 operands using 2 LUTs per bit

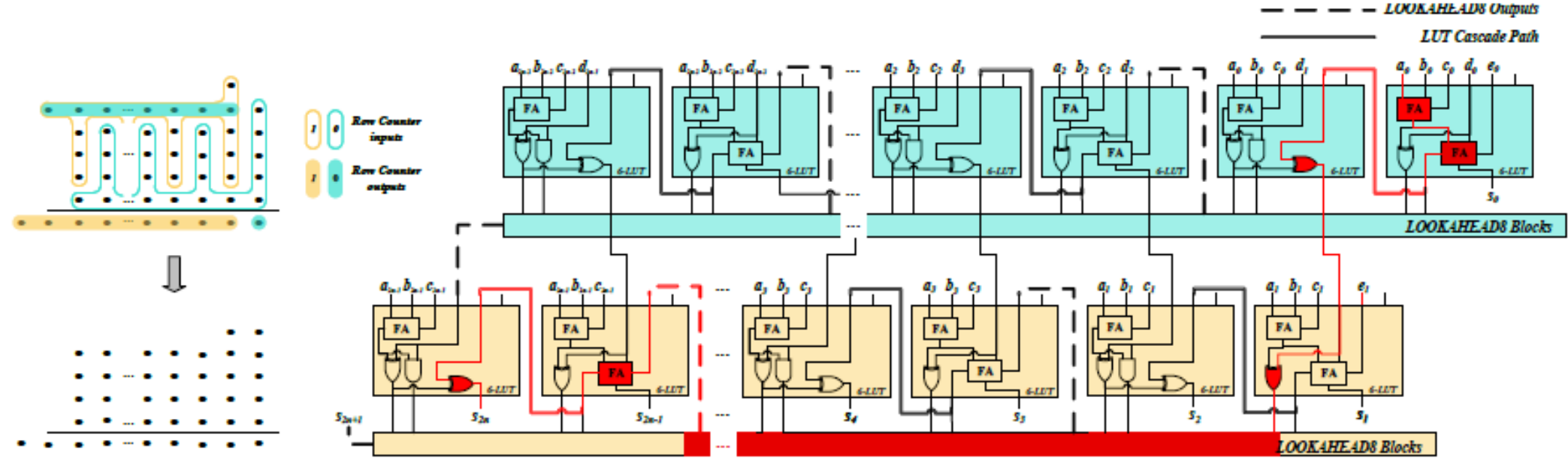
33% less LUT usage compared to Naïve +

2 LUTs / column

2 LUTs / column

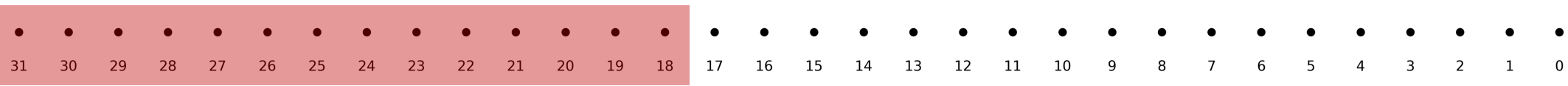


# New Implementation of Quaternary Adder

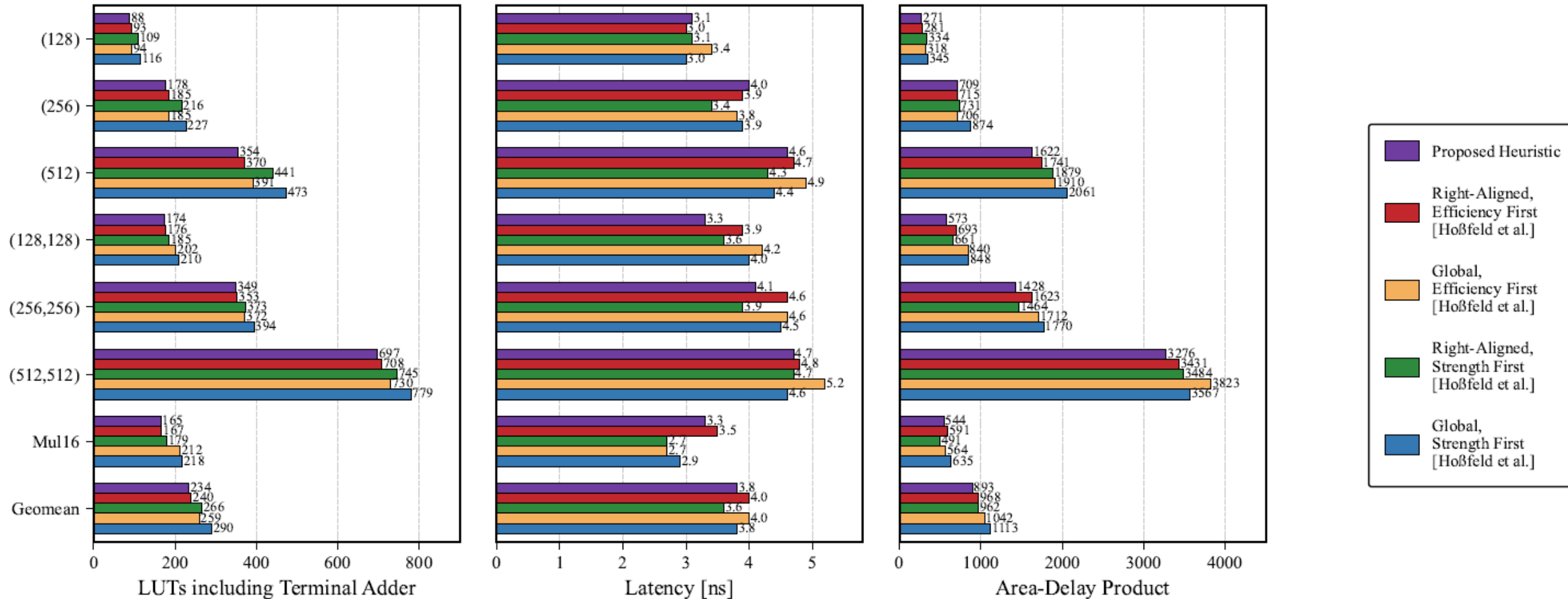


(1,5 : 3] GPCs : 2 LUTs / column

(3 : 2] GPCs: 1 LUTs / column



# Evaluation Results – Compressor Trees

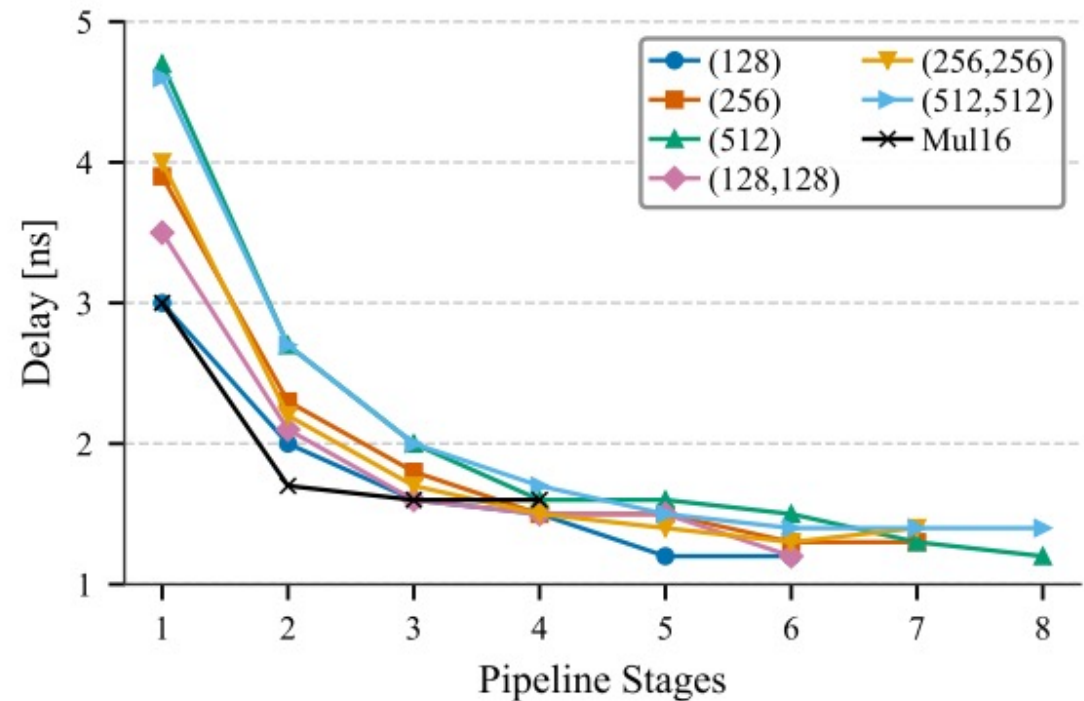
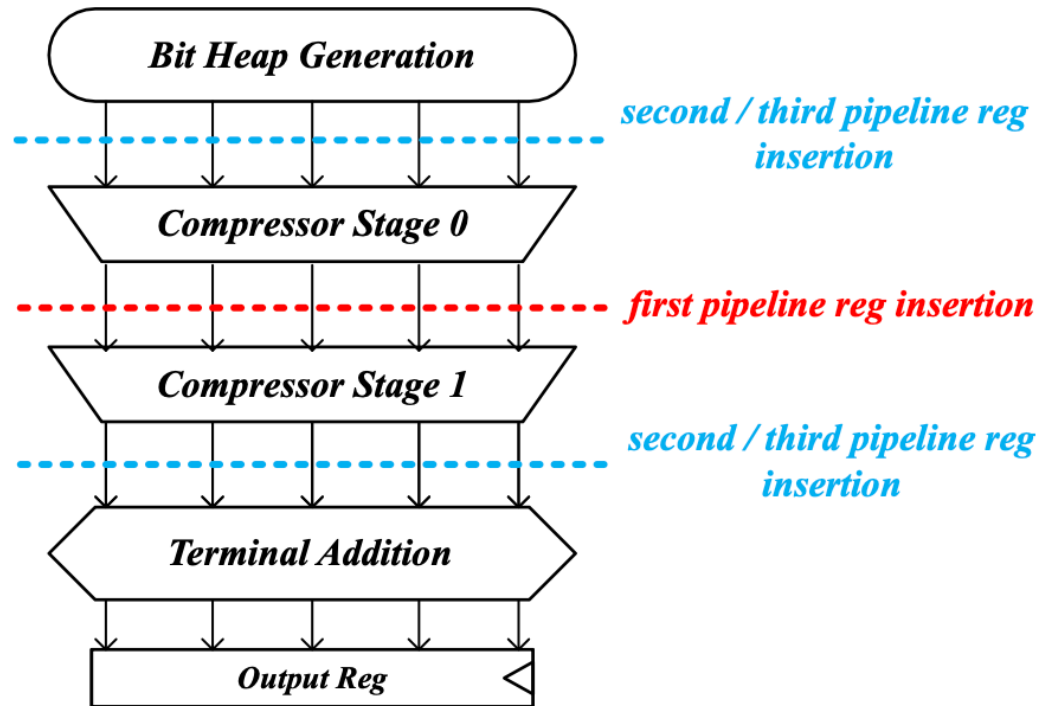


# What about larger multipliers?

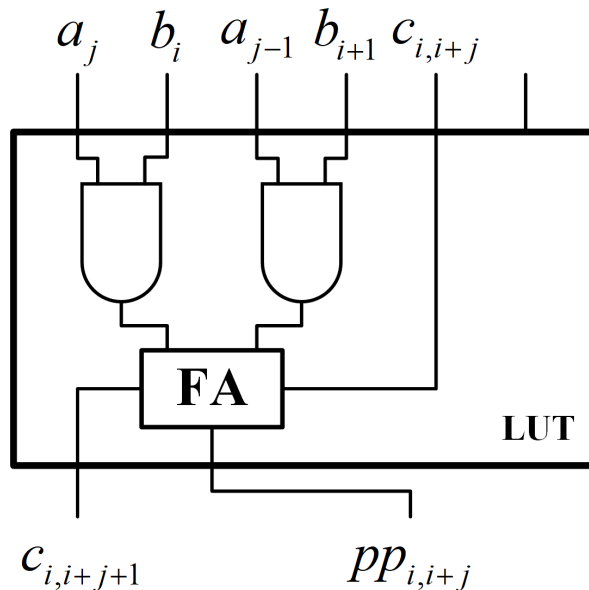
64-bit multiplier:

multiplier	LUT	CP (ns)
proposed	<b>2499</b>	4.6
LogiCore IP	4261	4.6
* Inferred (up to $f_{\max}$ )	6303	<b>3.8</b>
* Inferred (100MHz)	5362	/

# How does the generator pipeline the design?



# Comparison to Gate Absorption in [6]



Multiplier	LUTs
Unsigned 16-bit using gate absorption [6]	245
Proposed signed 16-bit	176
Proposed signed 18-bit	219