


# Combinational Building Blocks



# Index

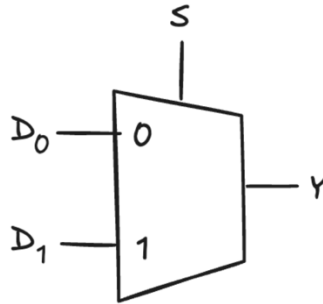
- Building Block definition
- Multiplexer
- Decoders
- Adders
- Subtractor
- Comparators
- Arithmetic/Logical Unit (ALU)
- Shifters and Rotators
- Multiplier and divider
- Floating-Point Addition

# Building Block Definition

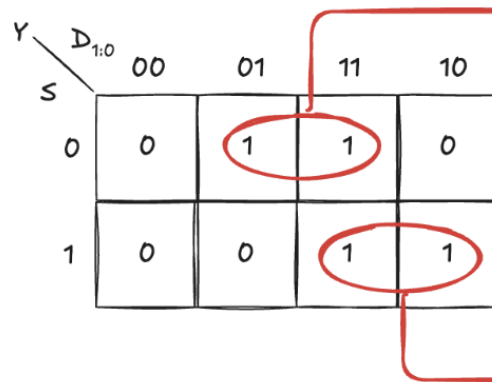
- Combinational logic can be grouped in **larger building blocks** to build more complex systems
- **Abstraction, hierarchy, modularity, and regularity**
  - **abstract** (hide) gate-level details to **emphasize the function** of the **module** (block)
  - **hierarchically** assembled from simpler components
  - **well-defined interfaces** in order to use them as a black boxes
- Examples:
  - seven-segment display decoders (already considered)
  - multiplexers
  - decoders
  - arithmetic circuits
- We will use many of these building blocks to build a **microprocessor**

# Multiplexer (mux)

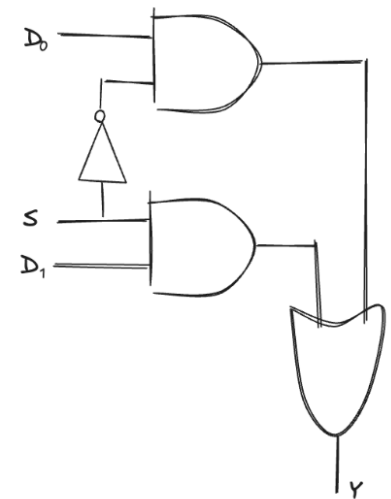
- A device that **choose an output among several possible inputs**, based on the value of a **select signal**
- A 2:1 multiplexer has two data inputs D and one output Y and a select signals S:
  - if  $S=0$ ,  $Y=D_0$
  - if  $S=1$ ,  $Y=D_1$



$S$	$D_0$	$D_1$	$Y$
0	0	0	0
0	0	1	0
0	1	0	1
0	1	1	1
1	0	0	0
1	0	1	1
1	1	0	0
1	1	1	1

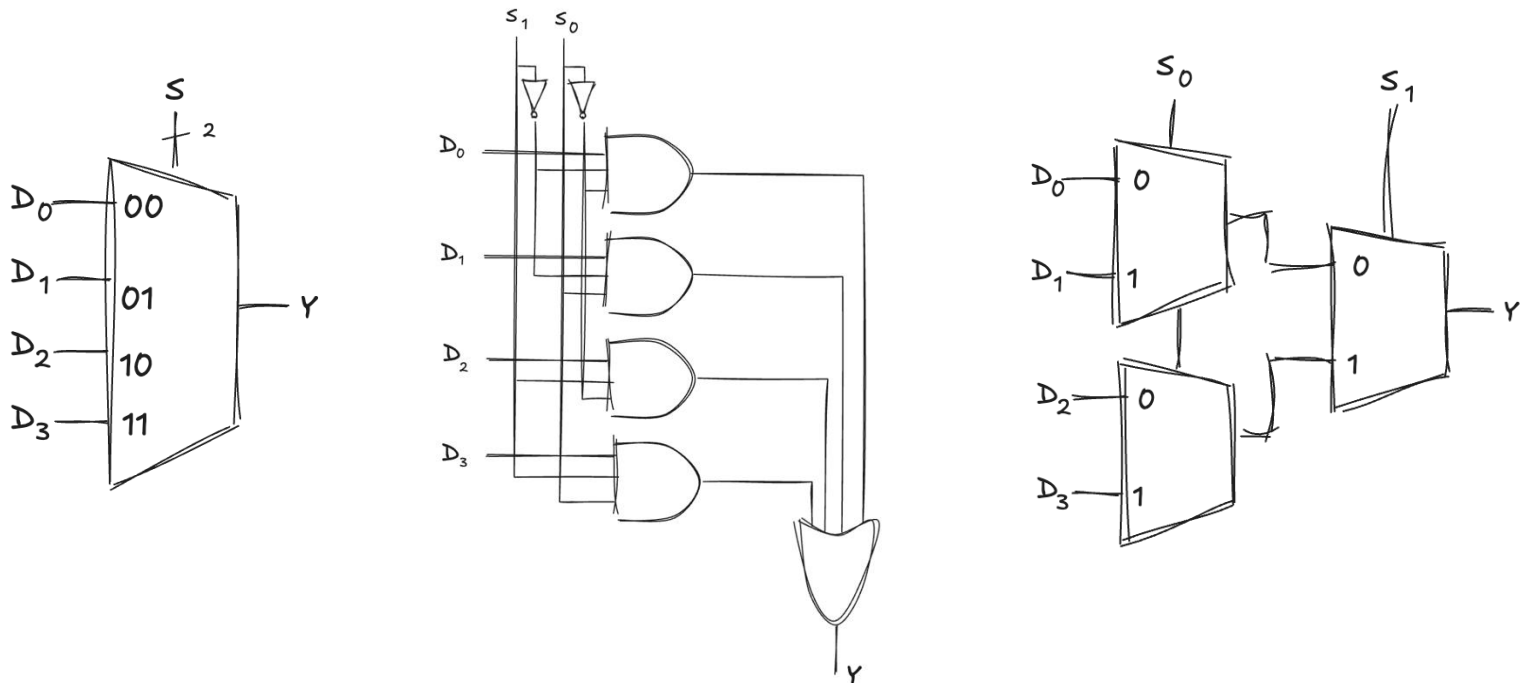


$$Y = D_0 \bar{S} + D_1 S$$



# Wider multiplexer

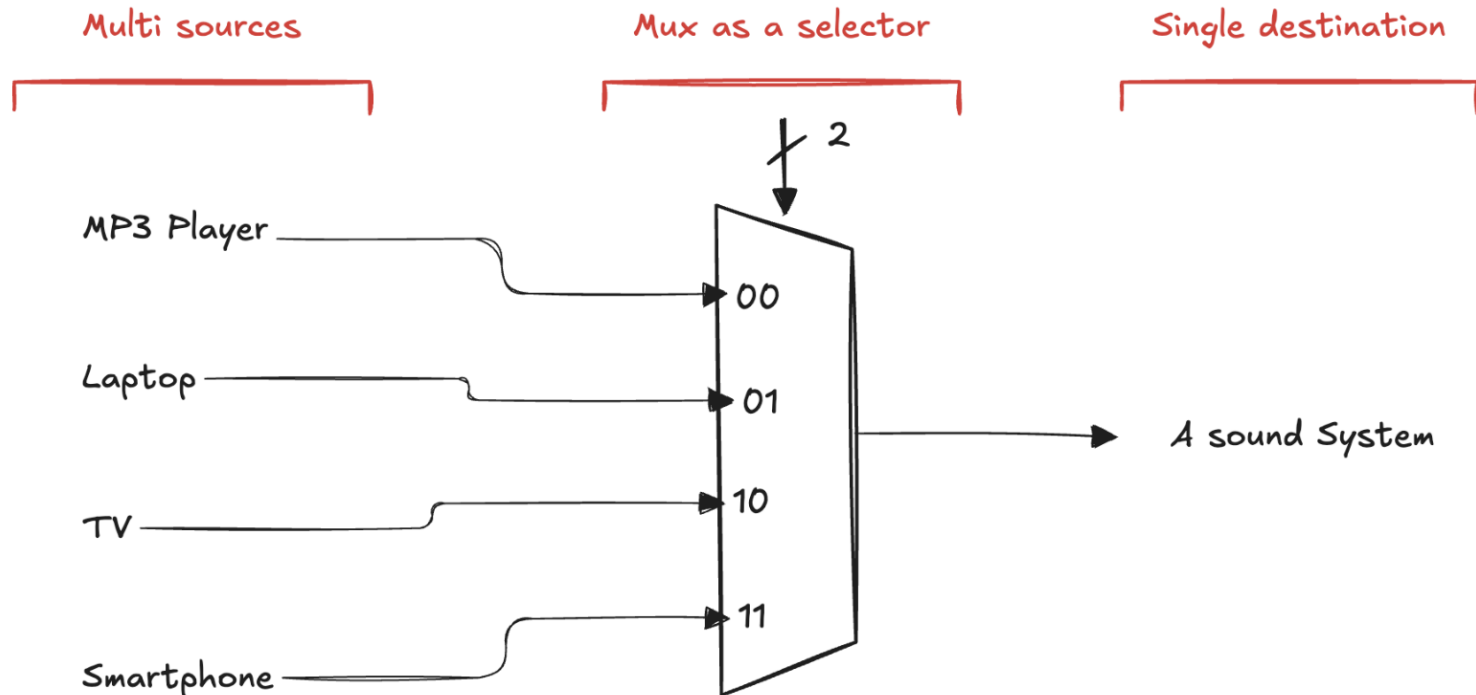
- A 4:1 multiplexer has four data inputs and one output and **two select signals** are needed to choose among the four data inputs
  - it can be built using sum-of-product logic, or multiple 2:1 multiplexers



- Wider multiplexers (8:1 and 16:1) can be built by expanding the method
  - in general, a **N:1 multiplexer needs  $\log_2 N$  select lines**

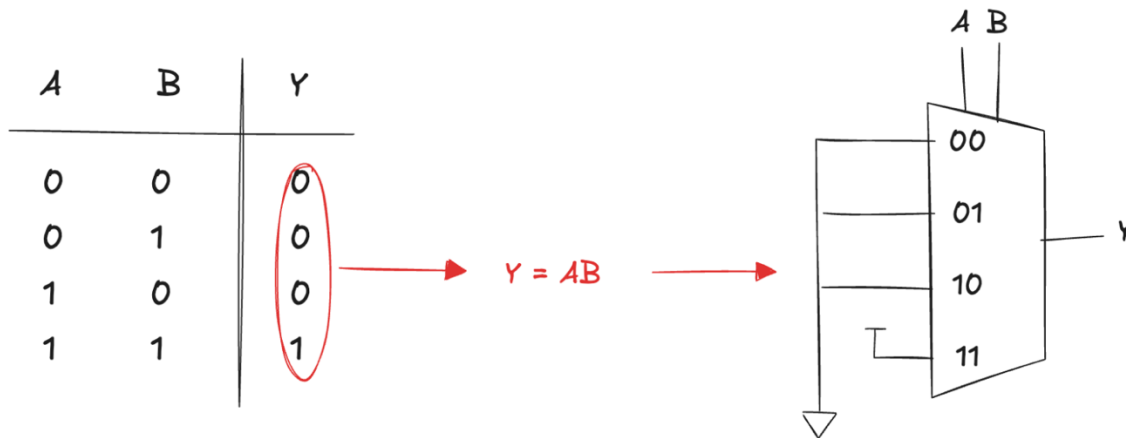
# Multiplexer example

- A multiplexer is like a **data router**: connects only one input at a time, **allowing different data sources to share the same output line efficiently**
- This makes it possible to **handle many inputs in a controlled, sequential way** without needing separate circuits for each one



# Multiplexer Logic (1)

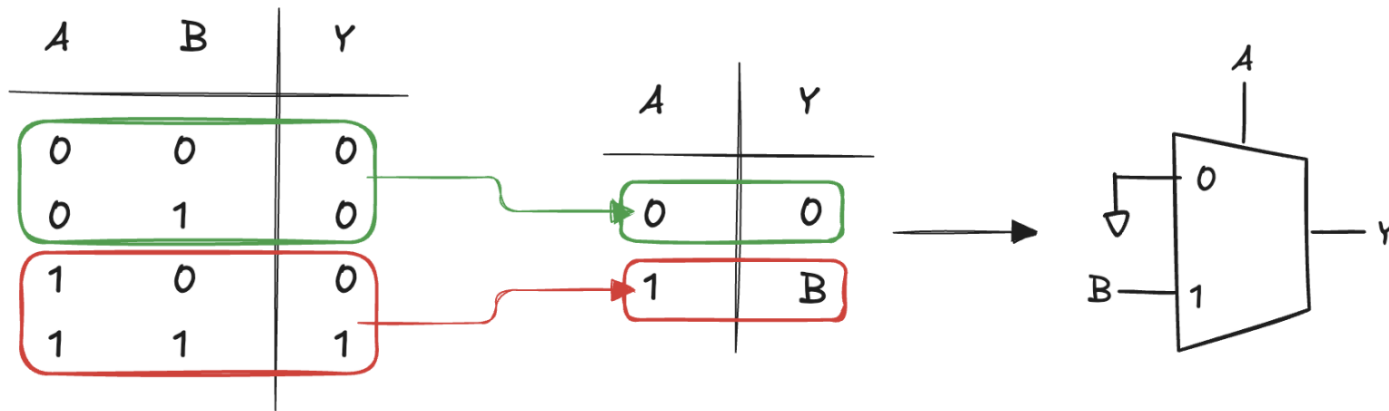
- Multiplexers can be used as **lookup tables** to **perform logic functions**
- For example, we can implement a two-input AND gate:



- A and B serve as select lines
  - inputs are connected to 0 or 1, **according to the truth table**
  - changing the data inputs, the multiplexer **can be reprogrammed** to perform a different function
- In general, a  **$2^N$ -input multiplexer can be programmed to perform any N-input logic function** by applying 0 and 1 to the appropriate data inputs

# Multiplexer Logic (2)

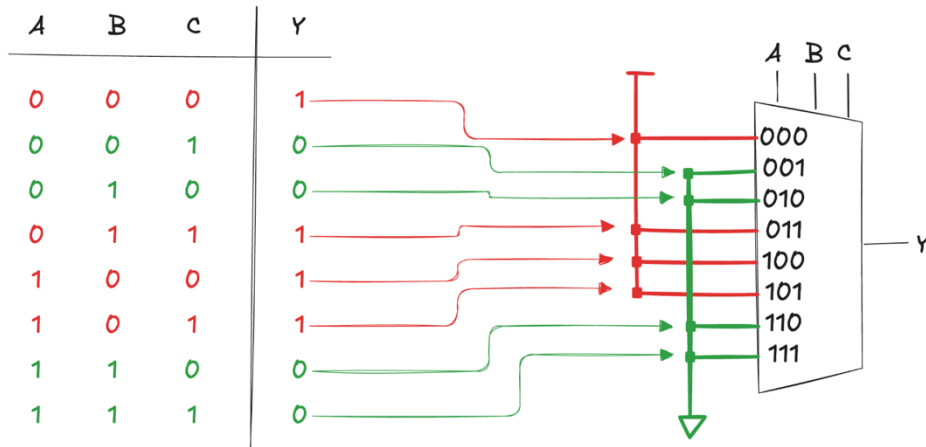
- We can **cut the multiplexer size in half**, using only a  $2^{N-1}$  input multiplexer to perform any N-input logic function
- Provide **one of the literals to the multiplexer inputs:**



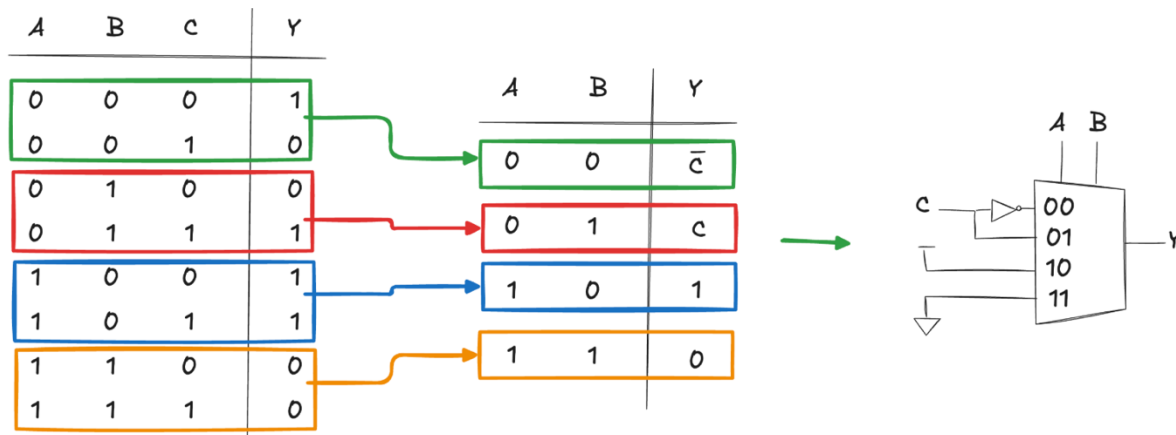


# Multiplexer Logic (3)

- As an example, implement the following function  $Y = A \bar{B} + \bar{B} \bar{C} + \bar{A} B C$ 
  - using a 8:1 multiplexer:



- and a 4:1 multiplexer:



# Multiplexer VHDL (1)

```
entity mux_4_1 is
    port( s0, s1: in std_logic;
          a, b, c, d: in std_logic;
          y: out std_logic
    );
end mux_4_1;

architecture rtl of mux_4_1 is
begin
    y <= a when s0='0' and s1='0' else
        b when s0='0' and s1='1' else
        c when s0='1' and s1='0' else
        d when s0='1' and s1='1' else
        '-';
end rtl;
```

# Multiplexer VHDL (2)

```
entity mux_4_1_8bit is
    port( s0, s1: in std_logic;
          a, b, c, d: in std_logic_vector(0 to 7);
          y: out std_logic_vector(0 to 7)
    );
end mux_4_1_8bit;

architecture rtl of mux_4_1_8bit is
    signal sel: std_logic_vector(0 to 1);
begin
    with sel select
        y <= a when "00",
            b when "01",
            c when "10",
            d when "11",
            "-----" when others;
end rtl;
```

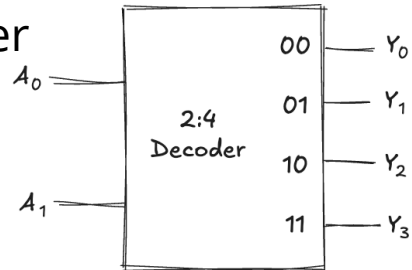
# Multiplexer VHDL (3)

```
entity mux_4_1_Nbit is
    generic( N: integer );
    port( sel: in std_logic_vector(0 to 1);
          a, b, c, d: in std_logic_vector(0 to N-1);
          y: out std_logic_vector(0 to N-1)
    );
end mux_4_1_Nbit;
```

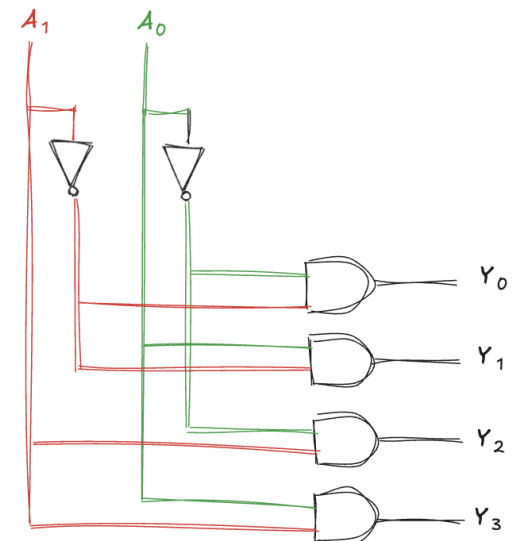
```
architecture rtl of mux_4_1_Nbit is
begin
    with sel select
        y <= a when "00",
            b when "01",
            c when "10",
            d when "11",
            (others => '-') when others;
end rtl;
```

# Decoder

- A device with **N inputs** and  **$2^N$  outputs** that asserts **exactly one** of its outputs depending on the input combination
- For example, consider a 2:4 decoder
  - when  $A_{1:0} = 00$ ,  $Y_0$  is 1
  - when  $A_{1:0} = 01$ ,  $Y_1$  is 1
  - and so forth
- The outputs are called **one-hot**, because just one is “hot” (HIGH) at a given time
- In the implementation, each gate depends on either the true or the complementary form of each input
  - an **N:2<sup>N</sup> decoder** can be constructed from **2<sup>N</sup> N-input AND gates** that accept the various combinations of true or complementary inputs
  - each output in a decoder represents a single minterm
    - for example,  $Y_3$  represents the minterm  $A_1 A_0$ .



$A_1$	$A_0$	$Y_3$	$Y_2$	$Y_1$	$Y_0$
0	0	0	0	0	1
0	1	0	0	1	0
1	0	0	1	0	0
1	1	1	0	0	0



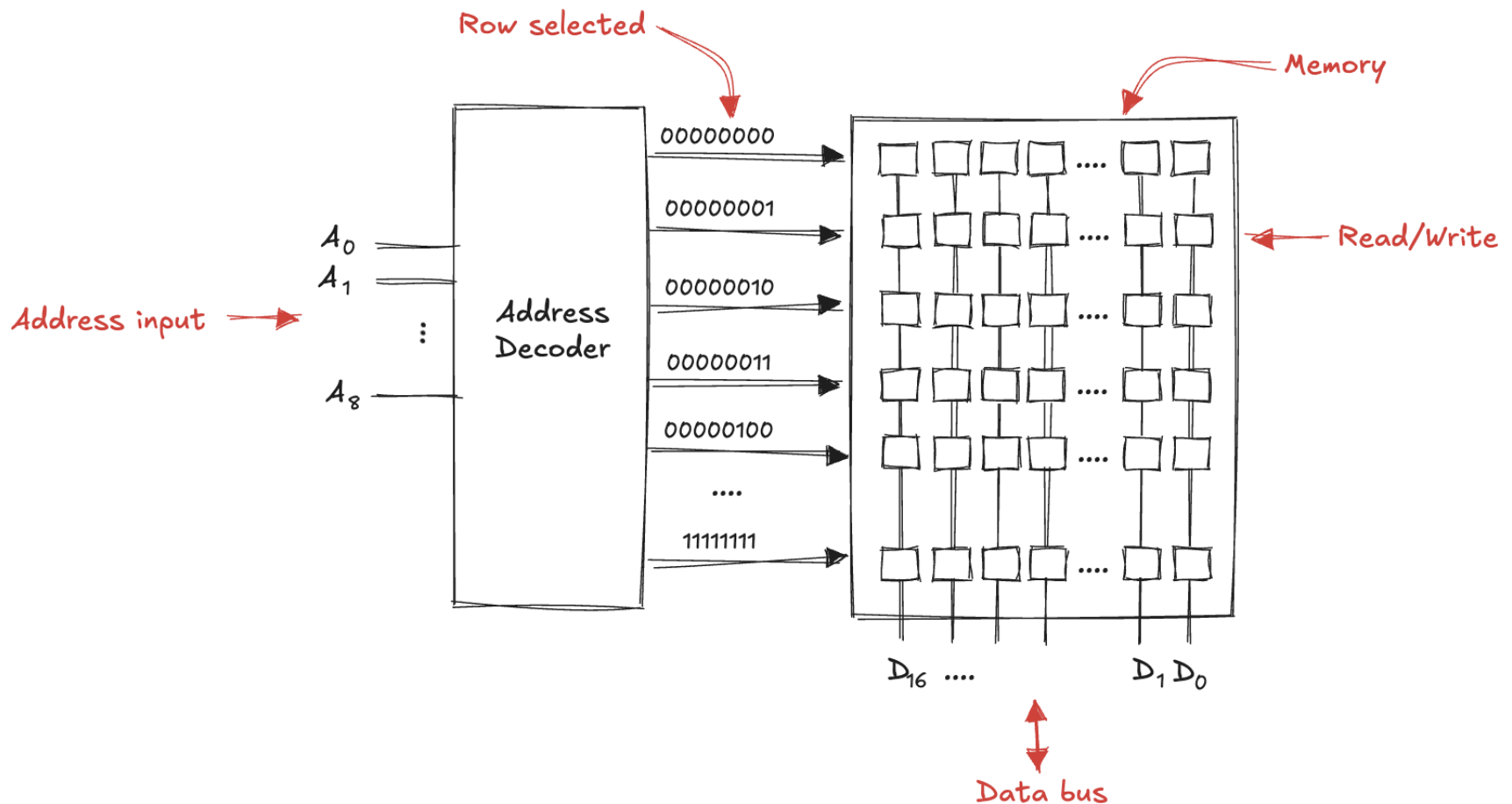
# Decoder VHDL

```
entity decoder_2_4 is
    port( dec_in: in std_logic_vector(0 to 1);
          dec_out: out std_logic_vector(0 to 3)
    );
end decoder_2_4;
```

```
architecture rtl of decoder_2_4 is
begin
    with dec_in select
    dec_out <= "0001" when "00",
               "0010" when "01",
               "0100" when "10",
               "1000" when "11",
               "----" when others;
end rtl;
```

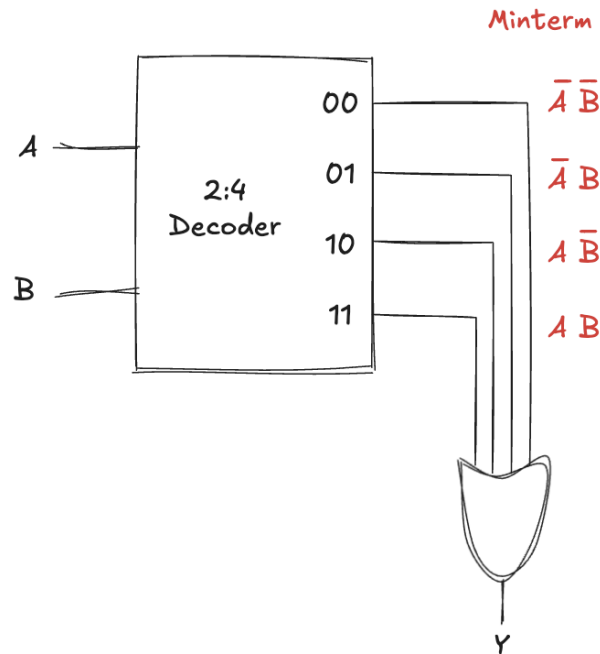
# Decoder example

- A decoder can be used to control which of several lines is “on” based on the binary value of the input. It can be used in **memory address selection**:



# Decoder Logic

- Decoders can be combined with OR gates **to build logic functions**
  - because each output of a decoder represents a single minterm, the function is built as the OR of all of the minterms in the function
  - example: a two-input XNOR function using a 2:4 decoder and a single OR gate



- An N-input function with M 1s in the truth table can be built with an N:2N decoder and an M-input OR.

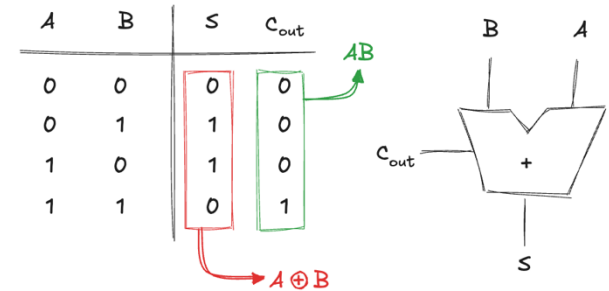


# Half and Full Adders

- Addition is one of the most common operations in digital systems

- **Half adder**

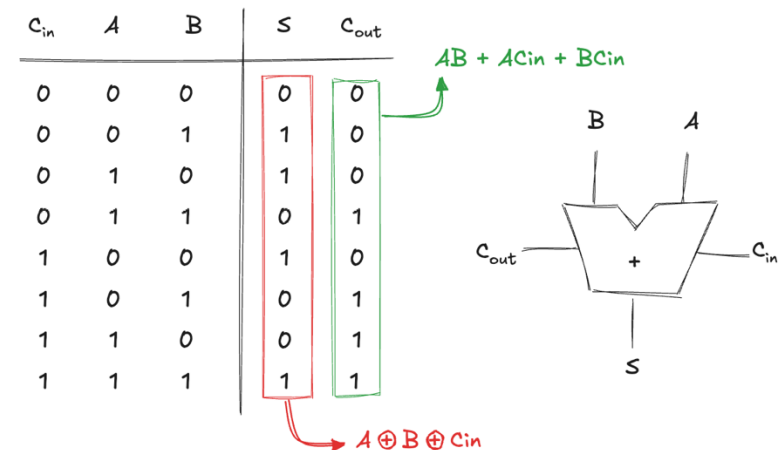
- two inputs, A and B, and two outputs, S and  $C_{out}$
  - S is the sum of A and B
  - $C_{out}$  is the eventual carry out



- In a multi-bit adder,  $C_{out}$  is added (carried) in to the next most significant bit
  - half adder lacks a  $C_{in}$  input to accept the  $C_{out}$  of the previous column

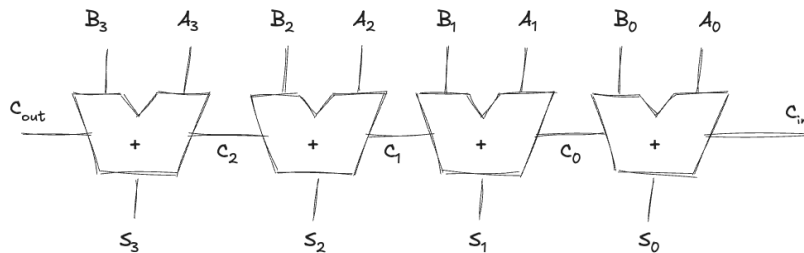
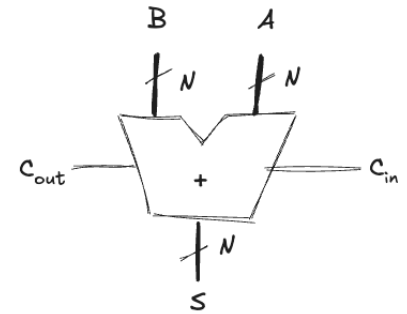
- **Full Adder**

- accepts also the carry in  $C_{in}$



# Carry Propagate Adder

- A **N-bit adder** that sums two N-bit inputs and a carry-in to produce an N-bit result and a carry-out
  - the carry bit propagates into the adder
  - drawn like a full adder except that input/output are **busses** rather than single bits
- Simplest way to build: chain together N full adders (**Ripple-Carry Adder RCA**)
  - $C_{out}$  of one stage acts as the  $C_{in}$  of the next stage



- A good application of modularity and regularity
  - full adder module is reused many times to form a larger system
- It is **slow when N is large**.
  - S<sub>3</sub> depends on C<sub>2</sub>, which depends on C<sub>1</sub>, and so forth
  - the delay of the adder  $t_{ripple} = N * t_{FA}$  grows directly with the number of bits

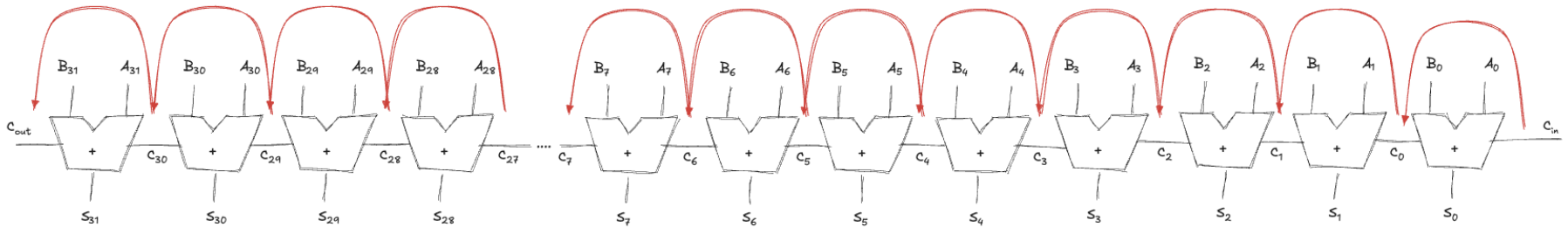
# CPA VHDL

```
entity adder is
|
|   generic(N: integer := 8);
|
|   port(A, B: in std_logic_vector(N-1 downto 0);
|       |   C_in: in std_logic;
|       |   S: out std_logic_vector(N-1 downto 0);
|       |   C_out: out std_logic);
|
end;

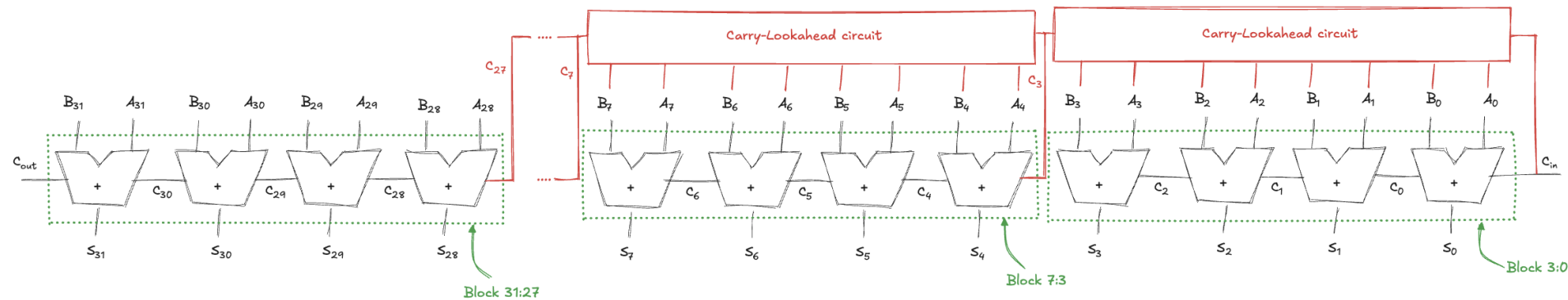
architecture synth of adder is
|   signal result: std_logic_vector(N downto 0);
begin
|   result <= ("0" & A) + ("0" & B) + C_in;
|   S <= result(N-1 downto 0);
|   Cout <= result(N);
end;
```

# Carry-Lookahead Adder (1)

- RCA are slow because the carry must propagate:

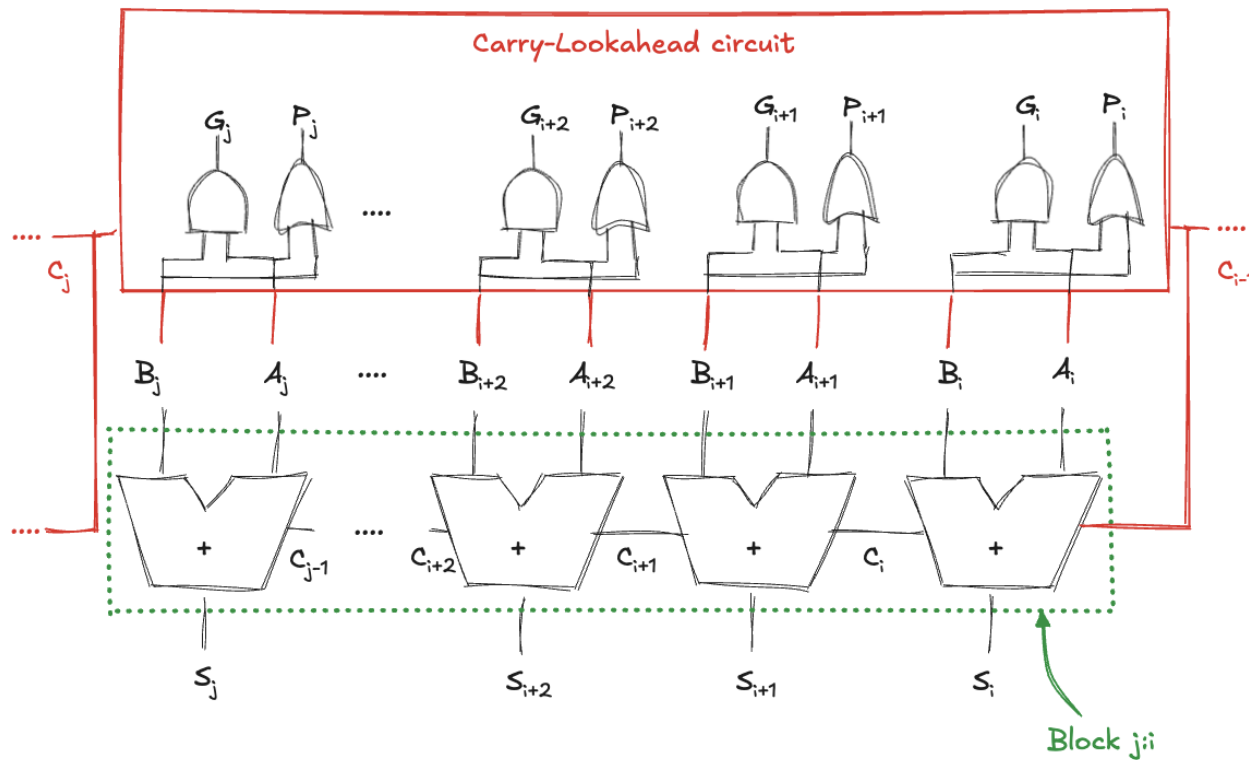


- The idea is to **divide the adder into blocks** and determine the carry out of a block **as soon as** the carry in is known
  - it looks ahead across the blocks rather than wait to ripple through all the full adders



# Carry-Lookahead Adder (2)

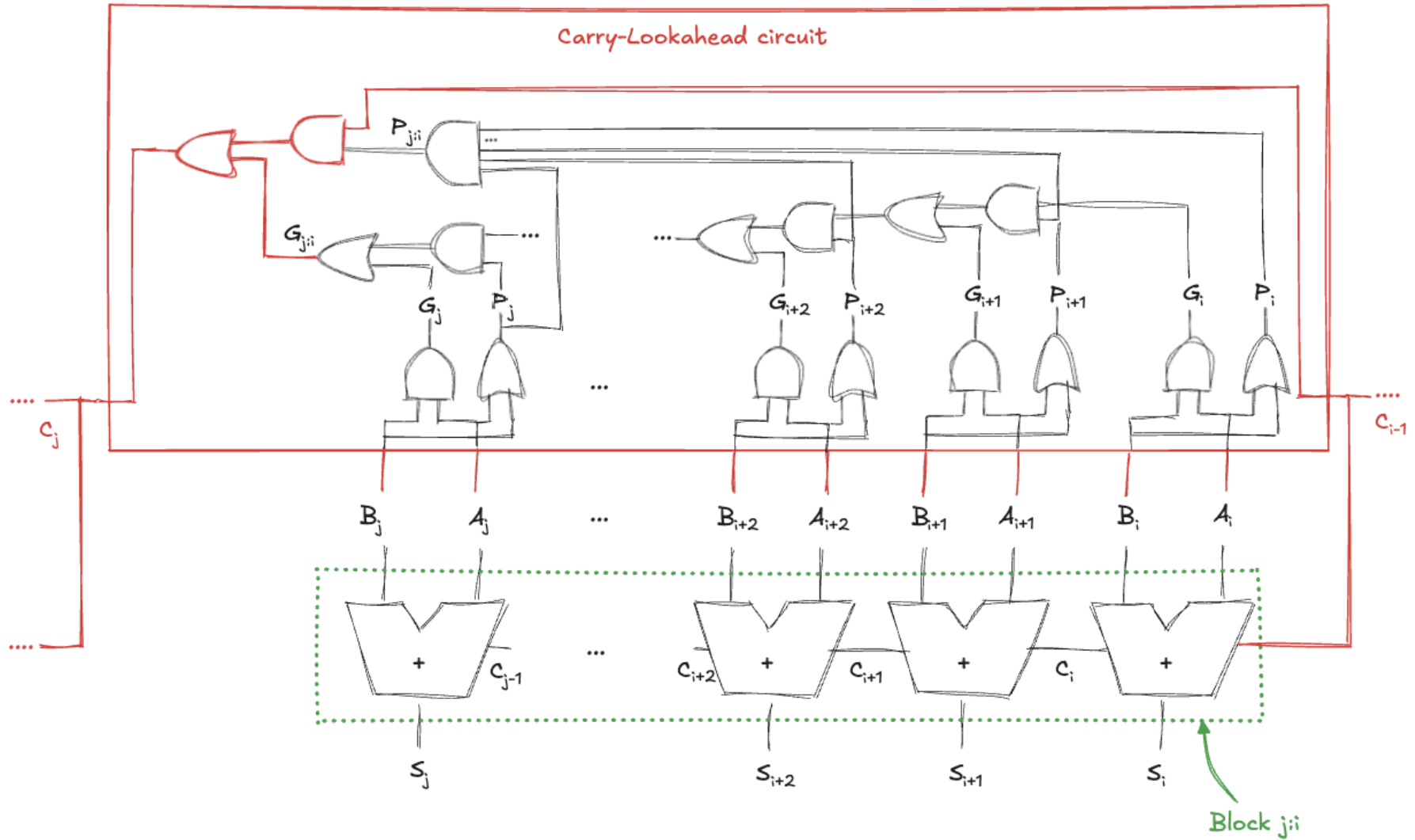
- **Carry-Lookahead Adder CLA** use **generate** (G) and **propagate** (P) signals
- A column **“generates”** a carry if it produces a carry independent of its carry-in
  - both input are 1 ->  $G_i = A_i \text{ AND } B_i$
- A column **“propagates”** a carry if it produces a carry whenever there is a carry-in
  - either one of the input is 1 ->  $P_i = A_i \text{ OR } B_i$



# Carry-Lookahead Adder (2)

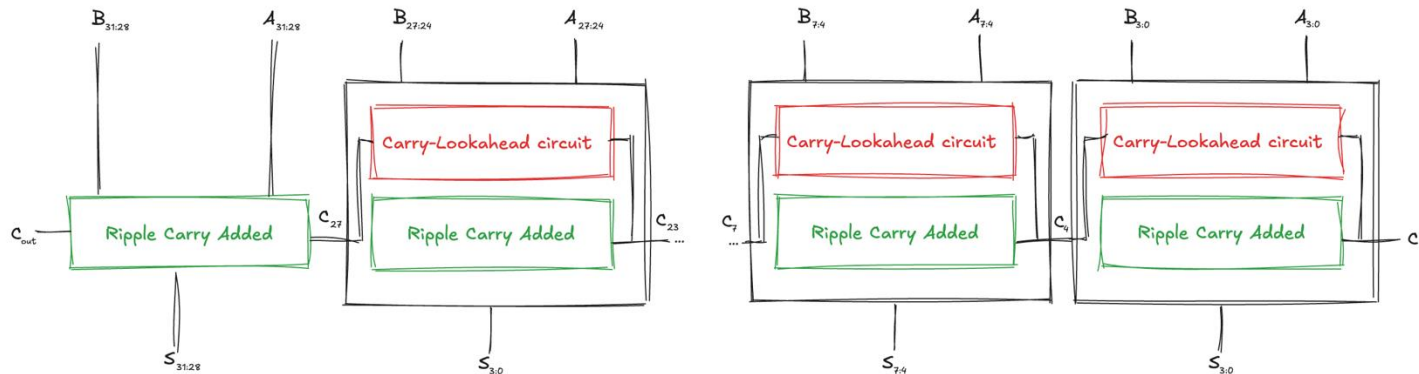
- We can extend to multiple-bit blocks
  - a block “generates” a carry if it produces a carry out independent of the carry-in
  - a block “propagate” a carry if it produces a carry out whenever there is a carry-in
- A block  $j:i$  (spanning columns  $i$  through  $j$ ) **generates** a carry
  - if the most significant column generates a carry
  - if the previous column generated a carry and the most significant column propagates it
  - and so forth
  - $G_{j:i} = G_j + P_j(G_{j-1} + P_{j-1}(G_{j-2} + P_{j-2}(\dots(G_{i+2} + P_{i+2}(G_{i+1} + P_{i+1}G_i)))$ 
    - e.g.  $G_{7:4} = G_7 + P_7(G_6 + P_6(G_5 + P_5G_4))$
- A block  $i:j$  **propagates** a carry
  - if all the columns in the block propagate the carry
  - $P_{j:i} = P_j P_{j-1} P_{j-2} \dots P_{i+2} P_{i+1} P_i$ 
    - e.g.  $P_{3:0} = P_3 P_2 P_1 P_0$
- So, we can quickly compute the carry out of the block
  - $C_j = G_{j:i} + P_{j:i}C_{i-1}$

# Carry-Lookahead Adder (3)



# Carry-Lookahead Adder (4)

- A 32-bit carry-lookahead adder composed of eight 4-bit blocks:

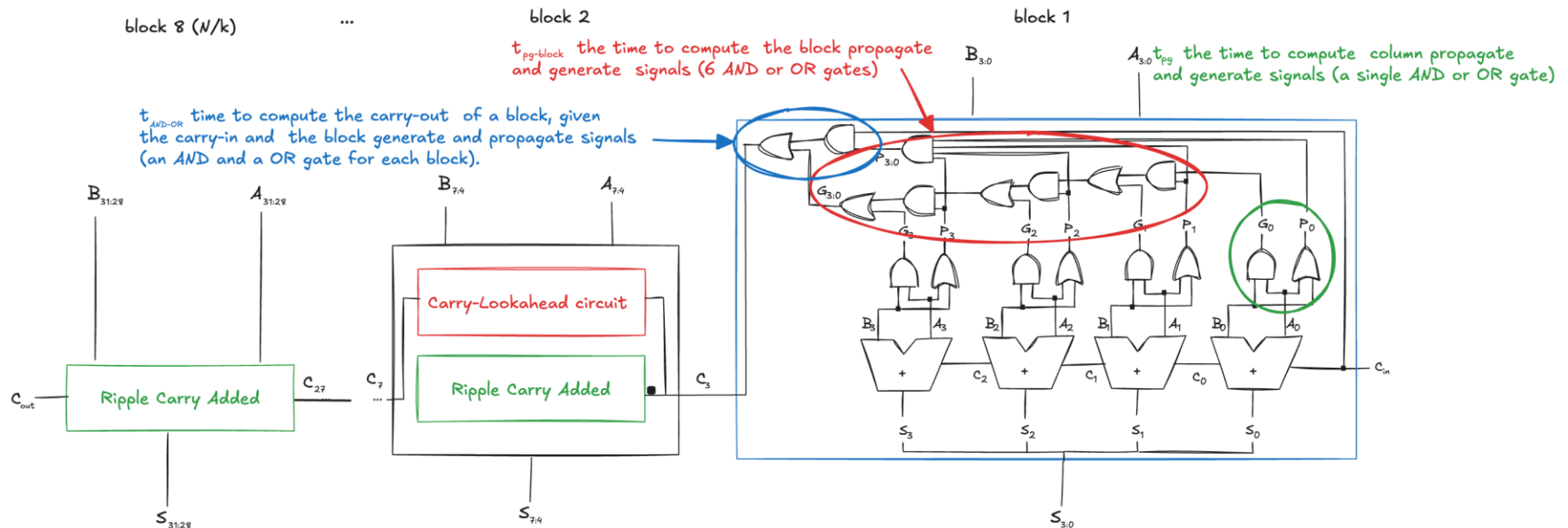


- Each block contains a 4-bit ripple-carry adder and the **lookahead** circuit to compute the carry out of the block given the carry-in
- All blocks compute column generate and propagate signals **simultaneously**
- Then  $C_{in}$  advances to  $C_{out}$  through each block until the last
  - $C_{in}$  proceeds through the lookahead gates to produce  $C_3$
  - $C_3$  proceeds through its lookahead block to produce  $C_7$
  - $C_7$  proceeds through its lookahead block to produce  $C_{11}$
  - and so on until  $C_{27}$ , the carry-in to the last block
- The last block contains a short ripple-carry adder (no more the lookahead logic)



# Carry-Lookahead Adder delay

- An N-bit adder divided into k-bit blocks has a delay
  - $t_{pg}$  is the delay to generate signals  $P_i$  and  $G_i$
  - $t_{pg\_block}$  is the delay to generate signals  $P_{j:i}$  and  $G_{j:i}$  for a k-bit block
  - $t_{lookahead}$  is the delay from  $C_{in}$  to  $C_{out}$  through the logic of the k-bit CLA block



- $t_{CLA} = t_{pg} + t_{pg\_block} + (N/k - 1) t_{lookahead} + k t_{FA}$ 
  - For  $N > 16$ , the carry-lookahead adder is much faster than the ripple-carry adder
  - however, the adder delay still increases linearly with N
- **Faster adders: more hardware, more expensive and power-hungry**
  - **trade-offs** must be considered when choosing an appropriate adder for a design

# Ripple-Carry Adder vs Carry-Lookahead Adder

- Compare the delays of a 32-bit ripple-carry adder and a 32-bit carry-lookahead adder with 4-bit blocks
- Assume that each gate delay is 100ps and that a full adder delay is 300ps
- The propagation delay of the 32-bit ripple-carry adder is
  - $t_{\text{ripple}} = N * T_{\text{FA}} = 32 * 300\text{ps} = 9.6 \text{ ns}$
- The carry-lookahead adder has
  - $t_{\text{pg}} = 100\text{ps}$  (one AND or one OR)
  - $t_{\text{pg\_block}} = 6 * 100 \text{ ps} = 600\text{ps}$  (six AND/OR gates)
  - $t_{\text{lookahead}} = 2 * 100 \text{ ps} = 200\text{ps}$  (one OR and one AND)
  - $t_{\text{CLA}} = t_{\text{pg}} + t_{\text{pg\_block}} + (N/k - 1)t_{\text{lookahead}} + kt_{\text{FA}} = 100 + 600 + (32/4 - 1)*200 + 4*300 = 3.3\text{ns}$
- The carry-lookahead adder is almost 3 times faster than the ripple-carry adder

# CLA VHDL

```
entity CLA4 is
    port (A, B : in std_logic_vector(3 downto 0);
          C_in : in std_logic;
          S : out std_logic_vector(3 downto 0);
          C_out : out std_logic);
end CLA4;

architecture behavioral of CLA4 is
    signal G, P : std_logic_vector(3 downto 0);
    signal C : std_logic_vector(4 downto 0);

begin
    -- Initial carry
    C(0) <= C_in;

    -- Generate and Propagate
    G <= A and B;
    P <= A xor B;

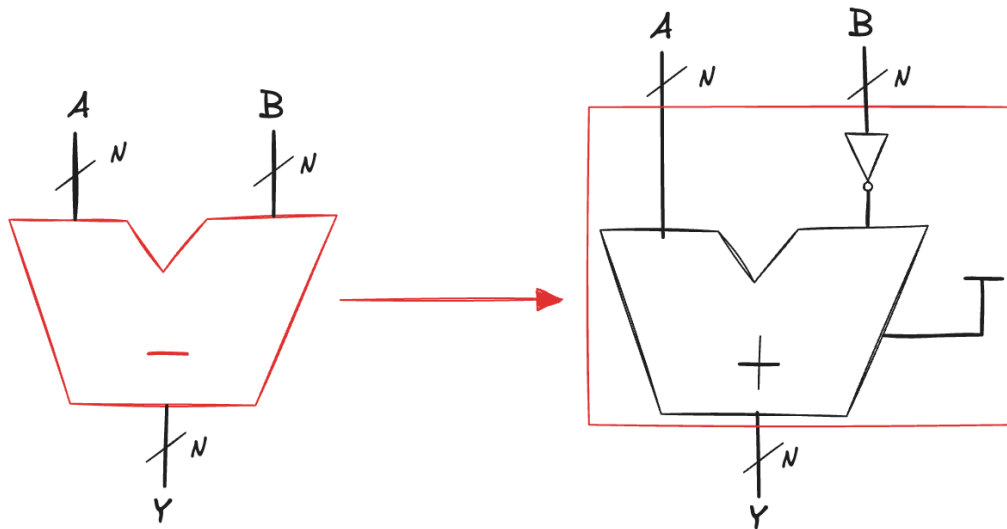
    -- Carry Lookahead Logic
    C(1) <= G(0) or (P(0) and C(0));
    C(2) <= G(1) or (P(1) and G(0)) or (P(1) and P(0) and C(0));
    C(3) <= G(2) or (P(2) and G(1)) or (P(2) and P(1) and G(0)) or
        (P(2) and P(1) and P(0) and C(0));
    C(4) <= G(3) or (P(3) and G(2)) or (P(3) and P(2) and G(1)) or
        (P(3) and P(2) and P(1) and G(0)) or
        (P(3) and P(2) and P(1) and P(0) and C(0));

    -- Sum
    S <= P xor C(3 downto 0);

    -- Final carry-out
    C_out <= C(4);
end behavioral;
```

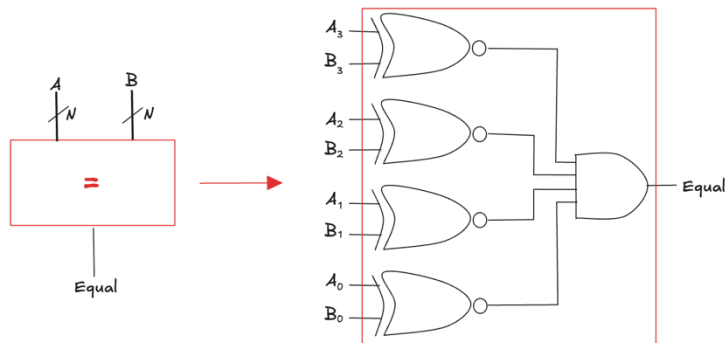
# Subtractor

- Performed by taking the two's complement of the second number, then adding
- To compute  $Y = A - B$ 
  - first create the two's complement of B
    - invert the bits of B to obtain B
    - add 1 to get  $-B = B + 1$  —
  - add this quantity to A—
    - $Y = A + B + 1 = A - B$
- We can use a single carry-lookahead adder by adding  $A + B$  with  $C_{in} = 1$



# Comparators (1)

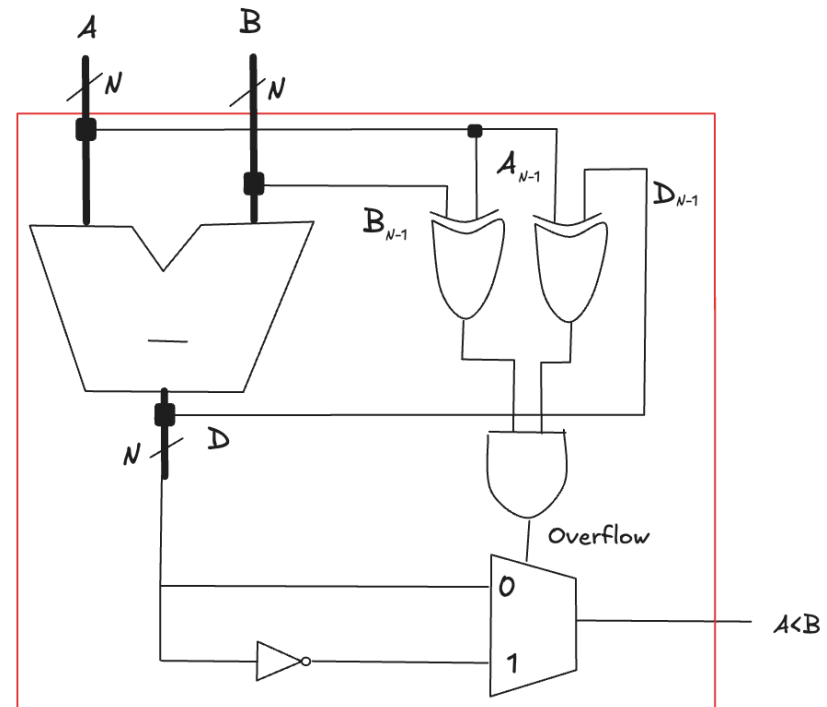
- Determines whether two binary numbers are equal or if one is greater or less than the other
  - receives two N-bit binary numbers A and B
  - an **equality comparator** produces a single output, indicating A is equal to B
  - a **magnitude comparator** produces one or more outputs, indicating the relative values of A and B
- The equality comparator is the simpler piece of hardware



- it checks whether the corresponding bits in each column of A and B are equal
  - using XNOR gates
- the numbers are equal if all the columns are equal

# Comparators (2)

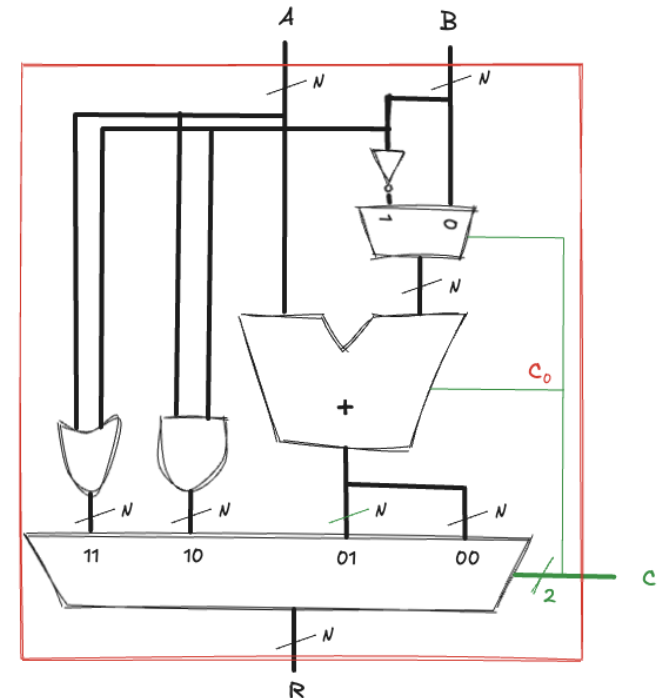
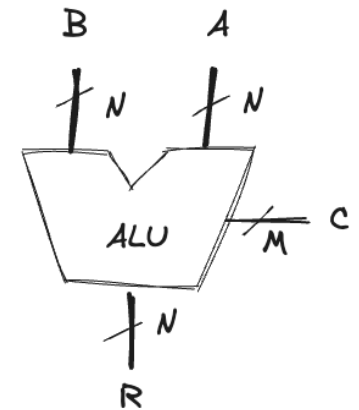
- Magnitude comparison of signed numbers is usually done by computing  $A - B$  and looking at the sign (most significant bit) of the result
  - if the result is negative (i.e., the sign bit is 1), then  $A$  is less than  $B$
  - otherwise,  $A$  is greater than or equal to  $B$
- This comparator functions **incorrectly upon overflow**
  - overflow occurs when
    - the two inputs have different signs
    - and the sign of the subtraction result has a different sign than the  $A$  input



# Arithmetic/Logical Unit (1)

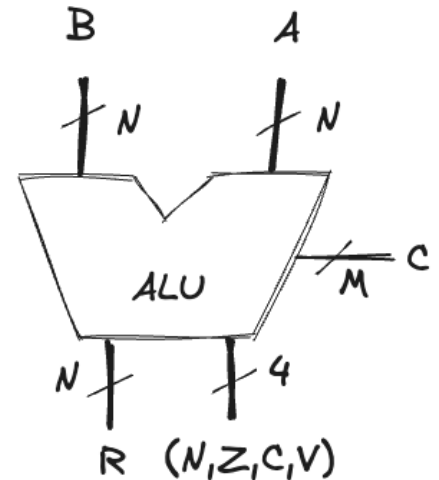
- ALU **combines** a variety of **mathematical** and **logical** operations into a single unit
  - addition, subtraction, AND, and OR operations
  - **control signal** to specifies the function to perform
  - the heart of most computer systems
- The following implementation contains
  - an N-bit adder, N 2-input AND and OR gates

Control		Funtion
$C_1$	$C_0$	
0	0	Add
0	1	Subtract
1	0	AND
1	1	OR



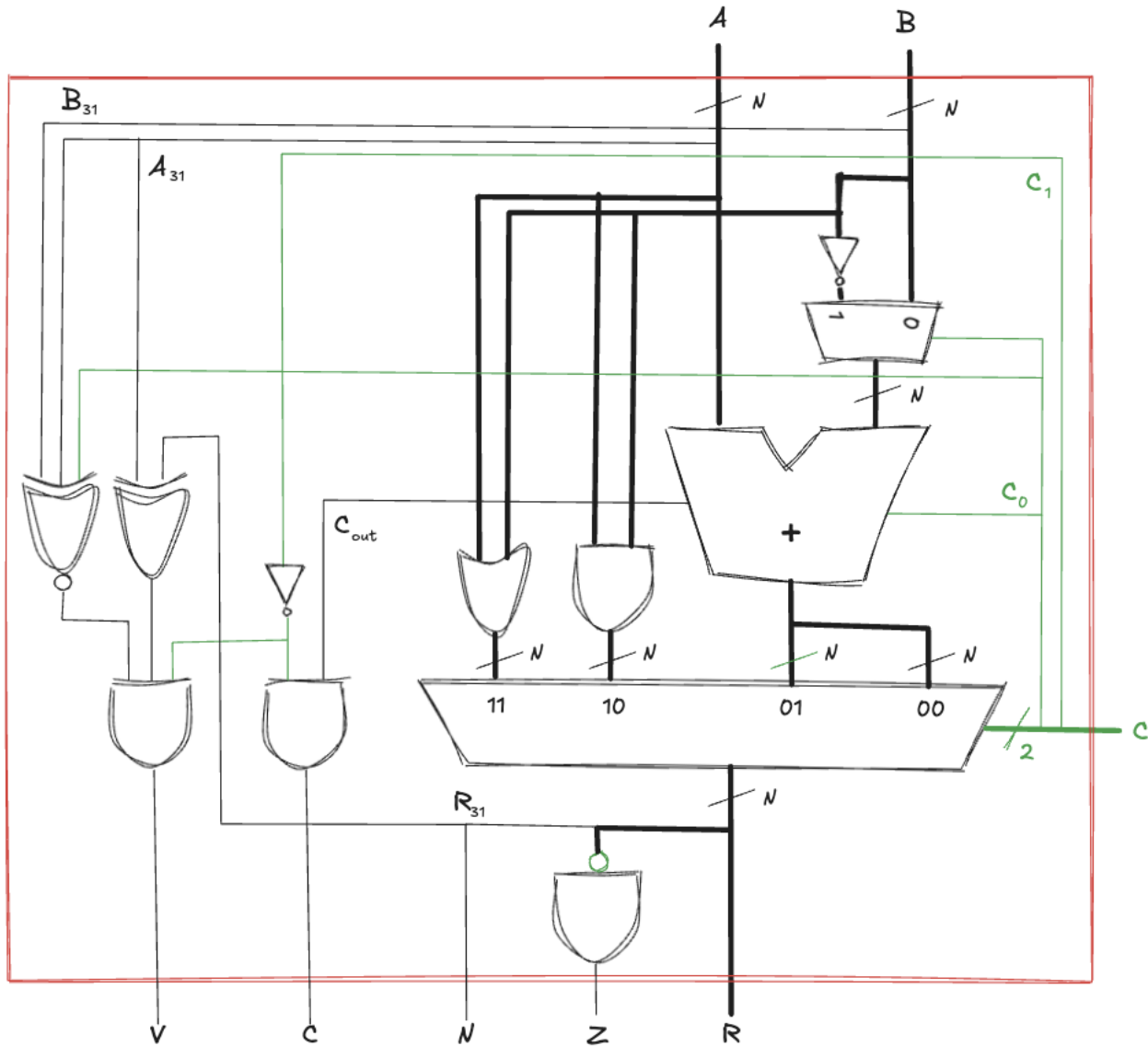
# Arithmetic/Logical Unit (2)

- ALU can produce **flags** to provide information about the result
- **Z (zero)**
  - all the bits of the output are 0
- **N (negative)**
  - the most significant bit of the output
- **C (carry)**
  - adder produces a carry out and the ALU is performing addition or subtraction
- **V (overflow)**
  - overflow occurs when the addition of two same signed numbers produces a result with the opposite sign
    - ALU is performing addition or subtraction ( $C_1=0$ )
    - A and Sum have opposite signs
    - A and B have the same sign, and the adder is performing addition ( $C_0=0$ )
    - or A and B have opposite sign, and the adder is performing subtraction ( $C_0=1$ )





# Arithmetic/Logical Unit (3)



# Arithmetic/Logical Unit (4)

- The ALU flags can also be used for **comparisons**:

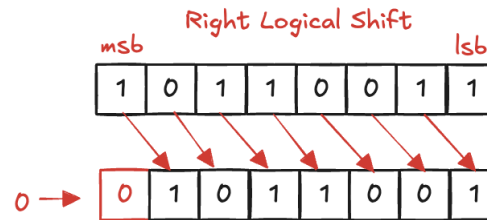
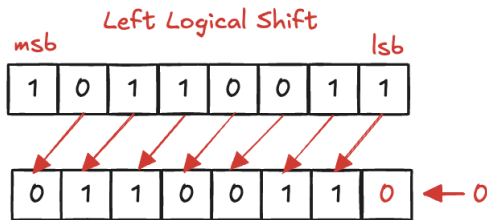
Condition	Expression	Meaning
$A = B$	$Z = 1$	The subtraction result is zero
$A \neq B$	$Z = 0$	The subtraction result is nonzero
$A < B$	$N \text{ xor } V = 1$	Negative and overflow flags differ
$A \leq B$	$Z \text{ or } (N \text{ xor } V) = 1$	Either equal or less than
$A > B$	$Z \text{ and } (N \text{ xor } V) = 0$	Not equal and not less than
$A \geq B$	$N \text{ xor } V = 0$	Negative and overflow flags are the same

- Processors can use combinations of status bits for **conditional branching** (e.g., checking if a value is less than zero or within bounds)
- Many variations** on this basic ALU exist that support other functions

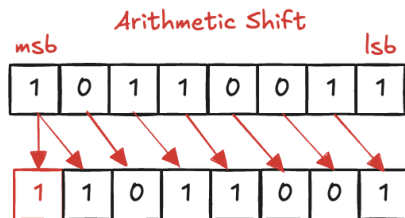
# Shifters (1)

- Shifters **moves binary data**, crucial in arithmetic operations, logical operations, and data manipulation. The main types are:

- logical shifter**: shifts all bits to the left (or right), inserting a 0 into lsb (or msb) and discarding msb (or lsb)



- arithmetic shifter**: like a logical shift with sign bit preserved if possible



- A left shift by N bits multiplies the number by  $2^N$ 
  - $000011_2 \ll 4 = 110000_2$  is equivalent to  $3_{10} \times 2^4 = 48_{10}$
- A right shift by N bits multiplies the number by  $2^{-N}$ 
  - $11100_2 \gg 2 = 11111_2$  is equivalent to  $-4_{10}/2^2 = -1_{10}$

# Shifters VHDL

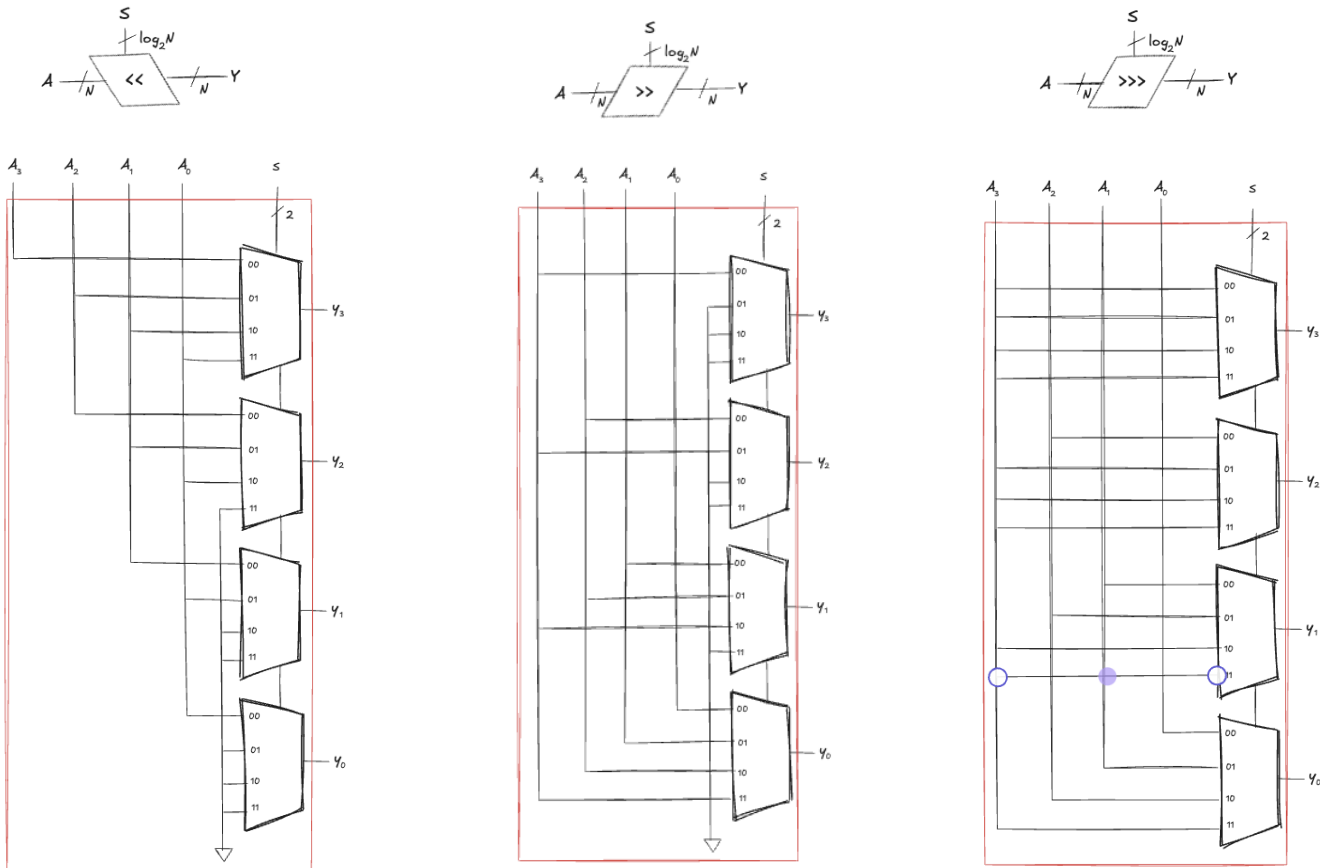
```
library IEEE;
use IEEE.std_logic_1164.all;
use IEEE.numeric_std.all;

entity Shifter is
    generic (
        N : integer := 8 -- width of the data word
    );
    port (
        A      : in  std_logic_vector(N-1 downto 0); -- input data
        SHIFT  : in  integer range 0 to N-1;         -- shift amount
        DIR     : in  std_logic;                      -- 0 = left, 1 = right
        ARITH   : in  std_logic;                      -- 0 = logical, 1 = arithmetic
        Y      : out std_logic_vector(N-1 downto 0) -- output data
    );
end Shifter;

architecture Behavioral of Shifter is
begin
    process (A, SHIFT, DIR, ARITH)
        variable temp : std_logic_vector(N-1 downto 0);
    begin
        if DIR = '0' then -- LEFT SHIFT
            if SHIFT < N then
                temp := std_logic_vector(shift_left(unsigned(A), SHIFT));
            else
                temp := (others => '0');
            end if;
        else -- RIGHT SHIFT
            if ARITH = '1' then -- Arithmetic right shift (preserve sign bit)
                temp := std_logic_vector(shift_right(signed(A), SHIFT));
            else -- Logical right shift (fill with zeros)
                temp := std_logic_vector(shift_right(unsigned(A), SHIFT));
            end if;
        end if;
        Y <= temp;
    end process;
end Behavioral;
```

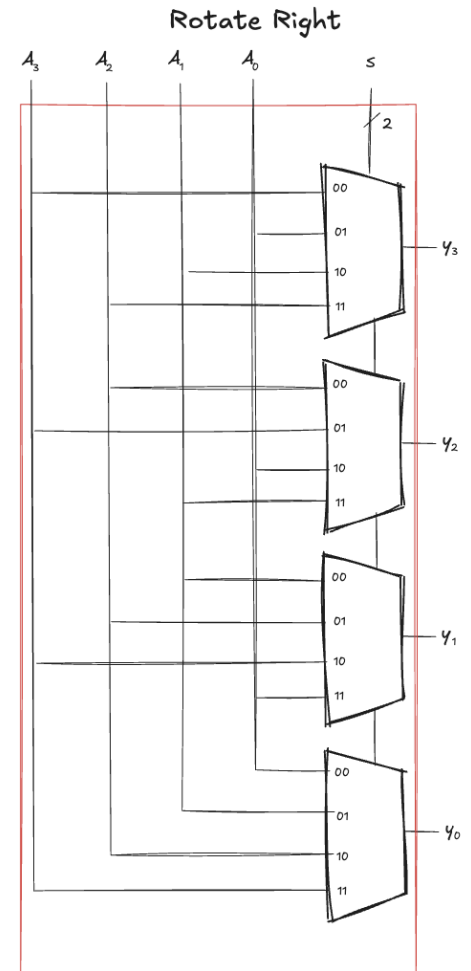
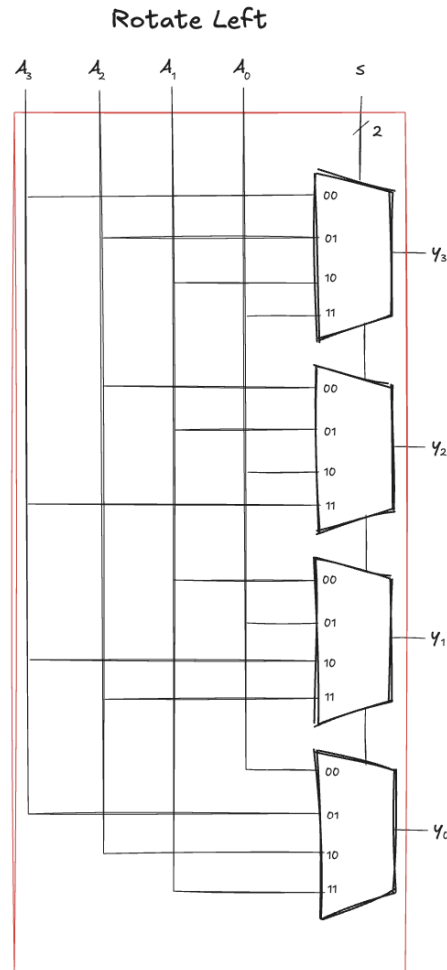
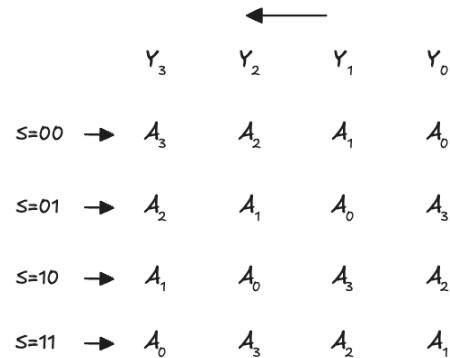
# Shifters as mux

- An N-bit shifter can be built from N N:1 multiplexers
- Depending on the value of a  $\log_2(N)$ -bit shift amount control signal, the output receives the input shifted by 0 to N-1 bits



# Rotators

- Rotator rotates (left or right) a number **in a circle** such that empty spots are filled with bits shifted off the other end



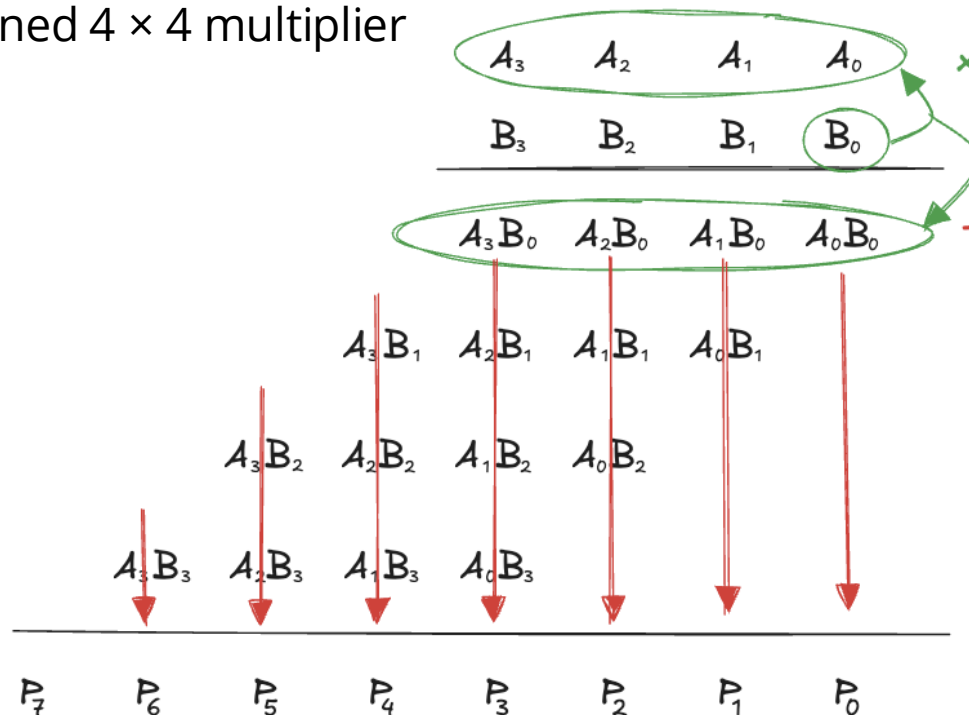
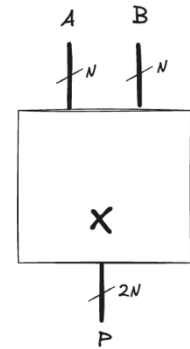
# Rotators VHDL

```
entity Rotator is
  generic (
    N : integer := 8 -- width of the data word
  );
  port (
    A      : in  std_logic_vector(N-1 downto 0); -- input data
    SHIFT  : in  integer range 0 to N-1;         -- rotation amount
    DIR    : in  std_logic;                       -- 0 = left, 1 = right
    Y      : out std_logic_vector(N-1 downto 0)  -- output data
  );
end Rotator;

architecture Behavioral of Rotator is
begin
  process (A, SHIFT, DIR)
    variable temp : std_logic_vector(N-1 downto 0);
  begin
    if DIR = '0' then
      -- Rotate left
      temp := A(N-1-SHIFT downto 0) & A(N-1 downto N-SHIFT);
    else
      -- Rotate right
      temp := A(SHIFT-1 downto 0) & A(N-1 downto SHIFT);
    end if;
    Y <= temp;
  end process;
end Behavioral;
```

# Multiplier (1)

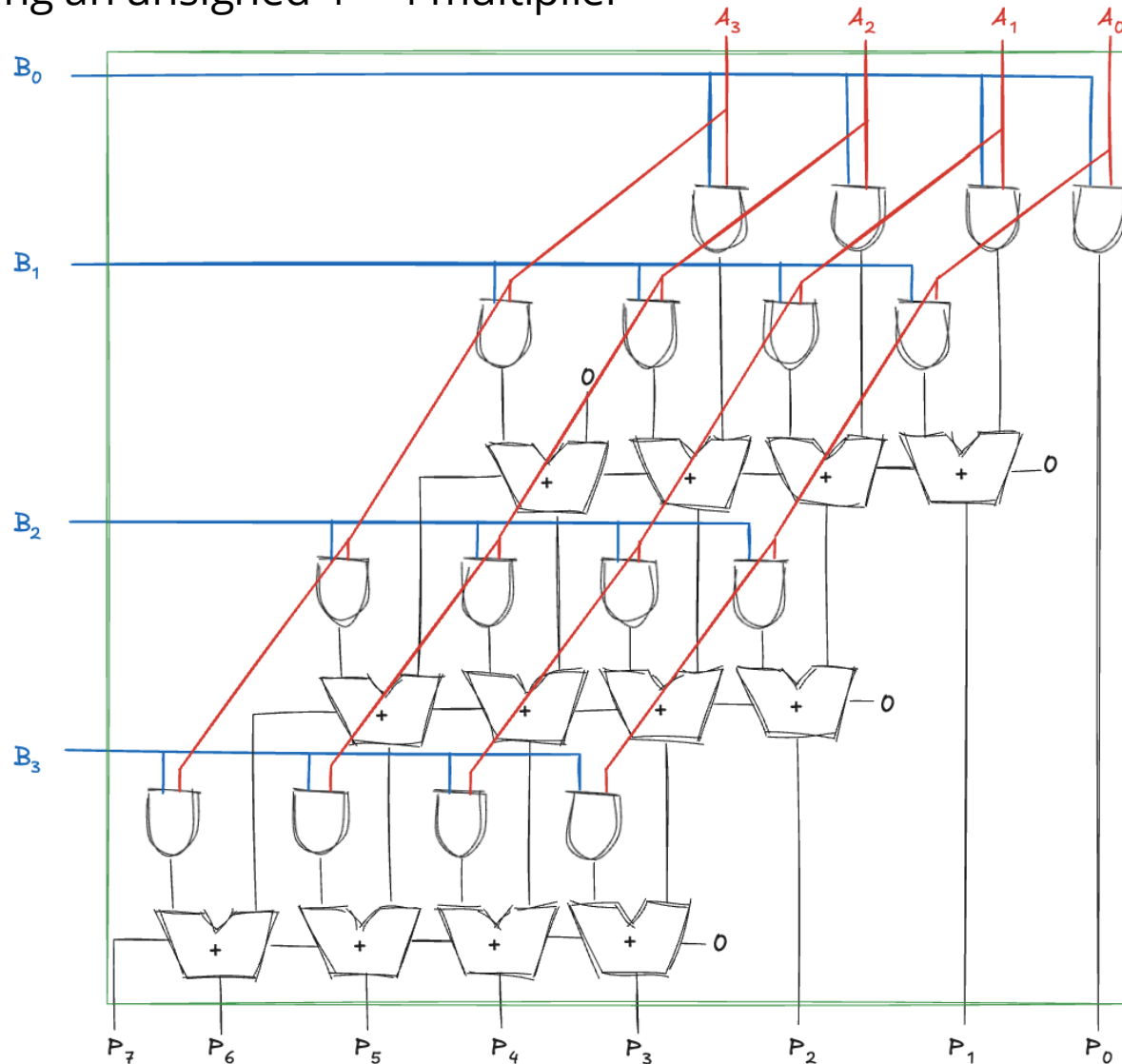
- Multiplication of 1-bit numbers is equivalent to the AND operation
- A  $N \times N$  multiplier produces a  $2N$ -bit result
  - partial product is the AND of a single multiplier bit ( $B_3, B_2, \dots$ ) with the multiplicand bits ( $A_3, A_2, \dots$ )
  - each partial products is added to the shifted next partial product
    - $B_0$  AND ( $A_3, A_2, A_1, A_0$ ) is added to  $B_1$  AND ( $A_3, A_2, A_1, A_0$ )
- Considering an unsigned  $4 \times 4$  multiplier





# Multiplier (2)

- Considering an unsigned  $4 \times 4$  multiplier



# Multiplier VHDL

```
entity Multiplier is
  generic (
    N : integer := 4 -- width of operands
  );
  port (
    A : in  std_logic_vector(N-1 downto 0);
    B : in  std_logic_vector(N-1 downto 0);
    P : out std_logic_vector(2*N-1 downto 0)
  );
end Multiplier;

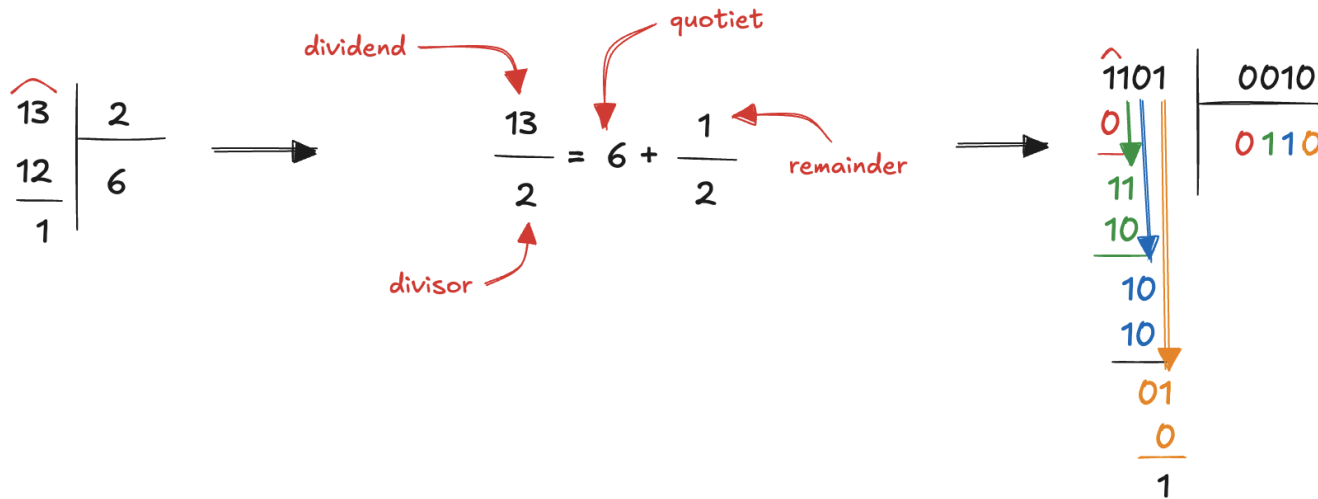
architecture Behavioral of Multiplier is
begin
  process (A, B)
    variable A_u : unsigned(N-1 downto 0);
    variable B_u : unsigned(N-1 downto 0);
    variable P_u : unsigned(2*N-1 downto 0);
  begin
    -- convert to unsigned
    A_u := unsigned(A);
    B_u := unsigned(B);

    -- perform multiplication
    P_u := A_u * B_u;

    -- assign result
    P <= std_logic_vector(P_u);
  end process;
end Behavioral;
```

# Division (1)

- The most complex operation to perform among all the arithmetic operations  $\frac{A}{B} = Q + \frac{R}{B}$
- The basic idea is the same as long division in decimal
  - how many times the divisor can fit into sections of the dividend without exceeding it?
  - each time it fits, we add a 1 to the quotient
  - if it doesn't fit, we add a 0



- This is the **restoring division algorithm**

# Division (2)

$R' = 0$

for  $i=N-1$  to 0:

$R = R' \ll 1, A_i$

$D = R - B$

if  $D < 0$ :

$Q_i = 0$

$R' = R$

else:

$Q_i = 1$

$R' = D$

$R = R'$

$A = 1101, B = 0010$

$R' = 0000$

$i = N-1 = 3$

$R = R' \ll 1, A_3 = 0000 \ll 1, 1 = 0001$

$D = R - B = 0001 - 0010 < 0$

$Q_3 = 0$  and  $R' = R = 0001$

$i = 2$

$R = R' \ll 1, A_2 = 0001 \ll 1, 1 = 0011$

$D = R - B = 0011 - 0010 = 0001 \geq 0$

$Q_2 = 1$  and  $R' = D = 0001$

$i = 1$

$R = R' \ll 1, A_1 = 0001 \ll 1, 0 = 0010$

$D = R - B = 0010 - 0010 = 0000 \geq 0$

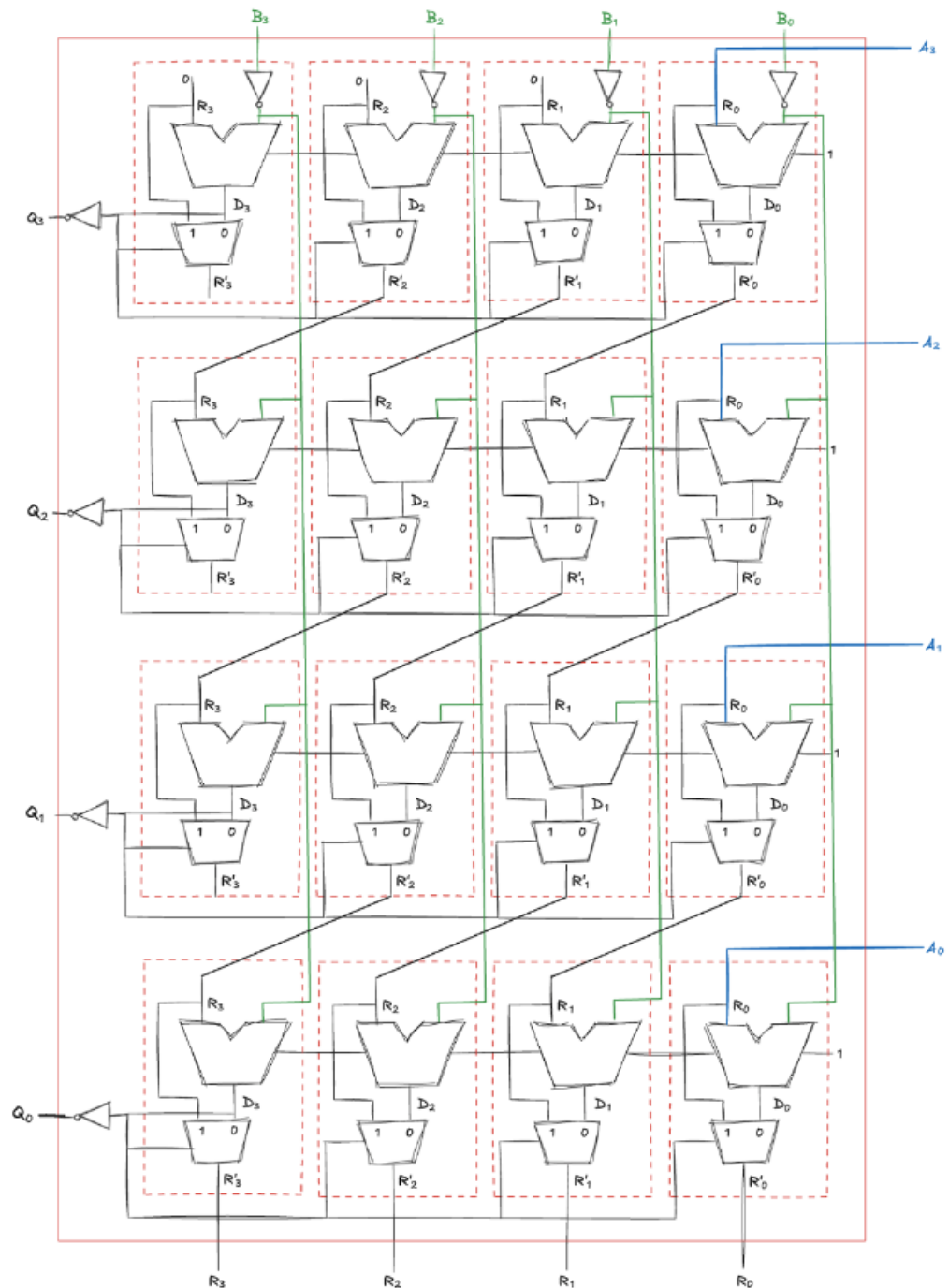
$Q_1 = 1$  and  $R' = D = 0000$

$i = 0$

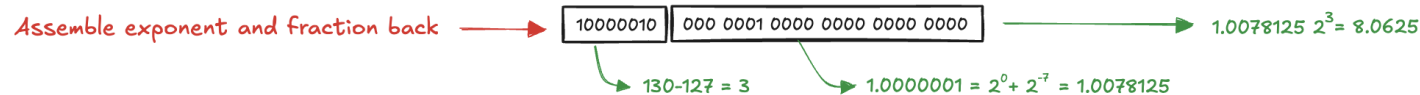
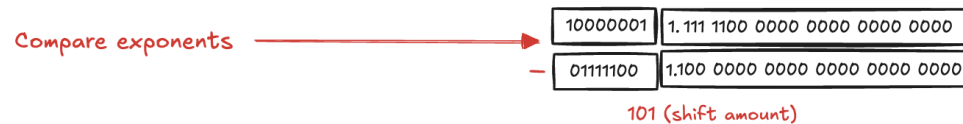
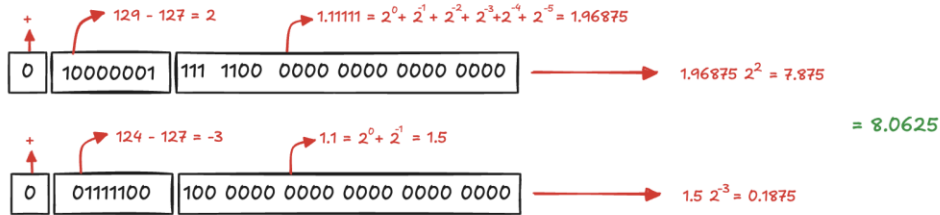
$R = R' \ll 1, A_0 = 0000 \ll 1, 1 = 0001$

$D = R - B = 0001 - 0010 < 0$

$Q_0 = 0$  and  $R' = R = 0001$



# Floating-Point Addition (1)



# Floating-Point Addition (2)

- Floating-point arithmetic is usually done in hardware to make it fast
  - **floating-point unit (FPU)**
  - typically, distinct from the **central processing unit (CPU)**
- The **infamous floating-point division bug (1994)**
  - the processor returns incorrect binary floating point results when dividing certain pairs of high-precision numbers...
  - missing values in a lookup table used by the FPU's division algorithm led to calculations acquiring small errors
  - these errors only occur rarely producing small deviations from the correct output
  - in certain circumstances the errors can occur frequently and lead to more significant deviations
  - **Byte magazine** estimated that **1 in 9 billion floating point divides with random parameters would produce inaccurate results**
  - **December 1994**, Intel recalled the defective processors in what was **the first full recall of a computer chip**: it cost \$475 million

