COMPUTER ARCHITECTURE

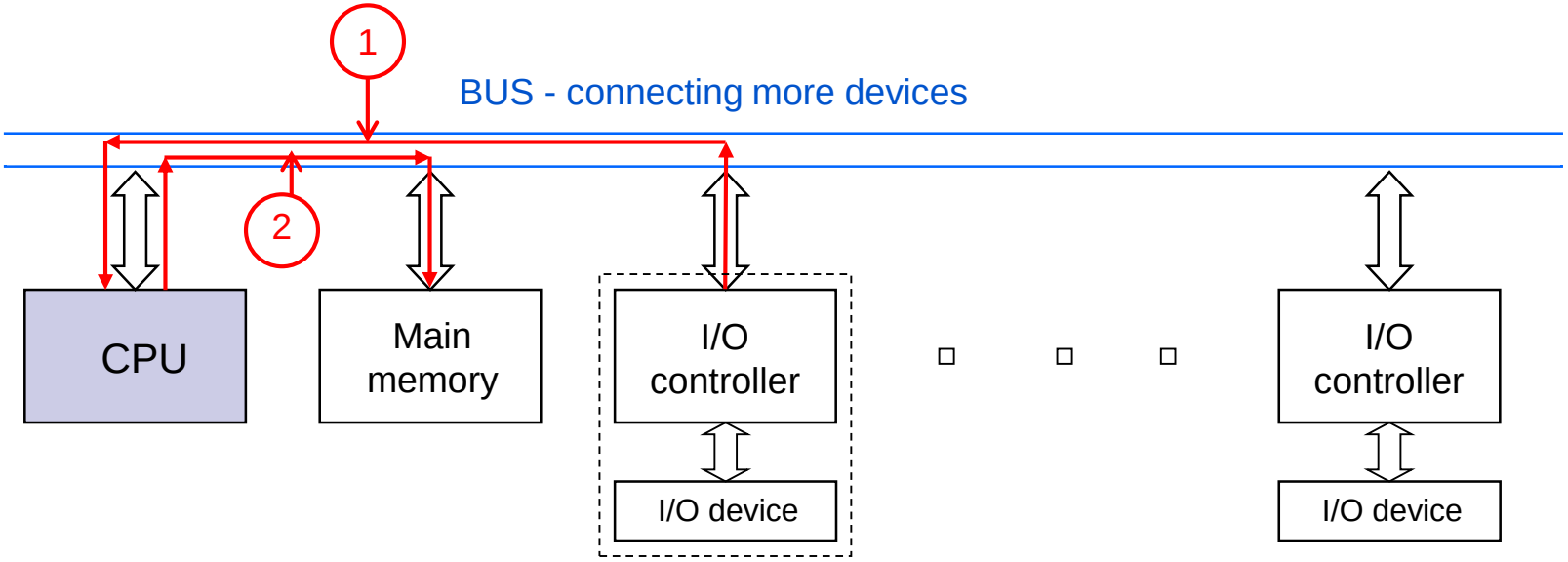
10 Input and output in von Neumann‘s Computer – selected themes

10. Input and output in von Neumann‘s computer

- I/O devices are used to convert information from one form to another (keyboard, mouse, monitor, printer, ...)
- I/O devices for storing information - auxiliary storage devices (hard disk drives, solid state drives, magnetic tapes, DVDs,...)
- Basic mode of I/O system operation: data transfer between the I/O system and the main memory

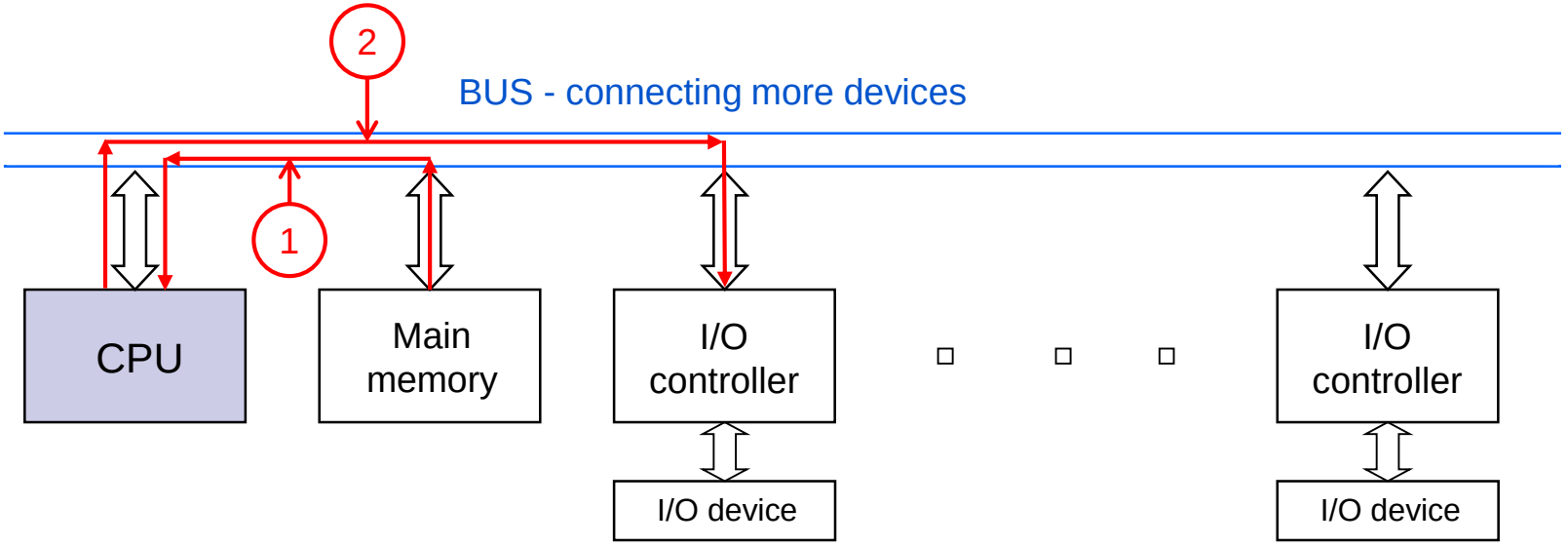
Types of I/O transfers

- Programmed Input/Output (Programmed I/O - PIO)
  - Transfer is controlled by a program executed on the CPU.
  - CPU communicates with I/O device.
  - Transfer progress (read from I/O device)
    - I/O → (1) CPU → (2) memory

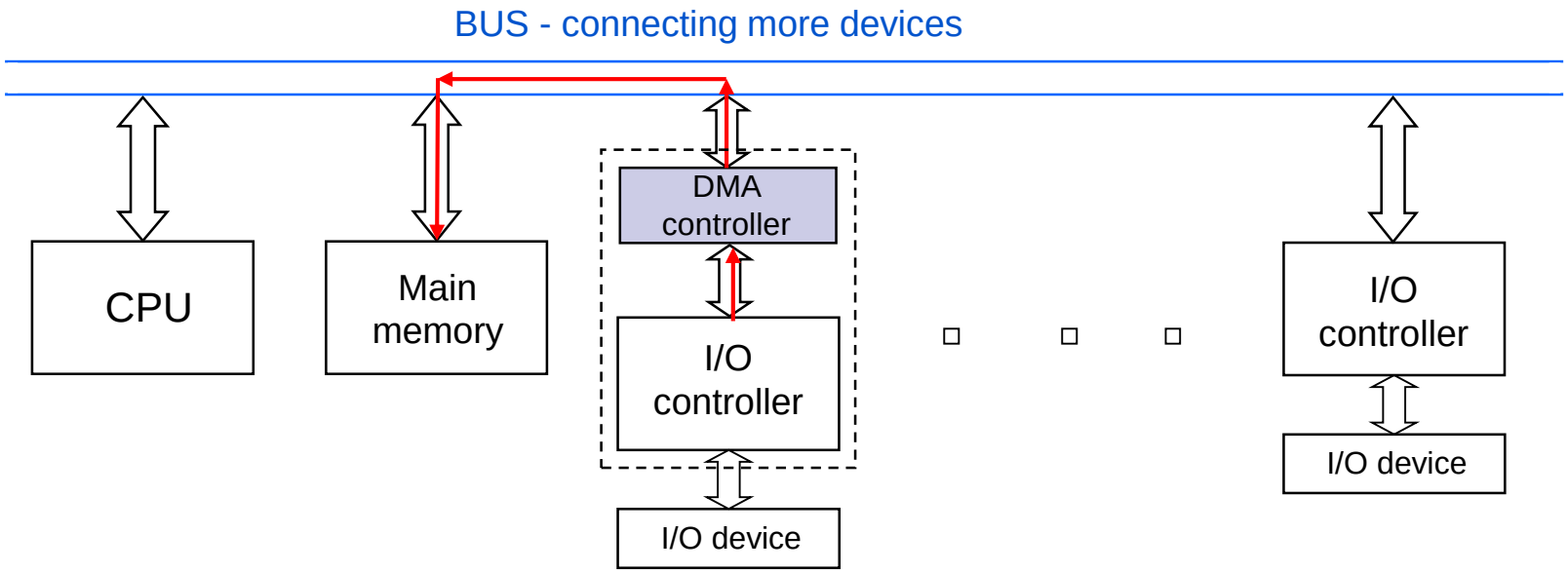


Programmed Input/Output (Programmed I/O - PIO)

- Transfer progress (write to I/O device)
  - memory → CPU → I/O device

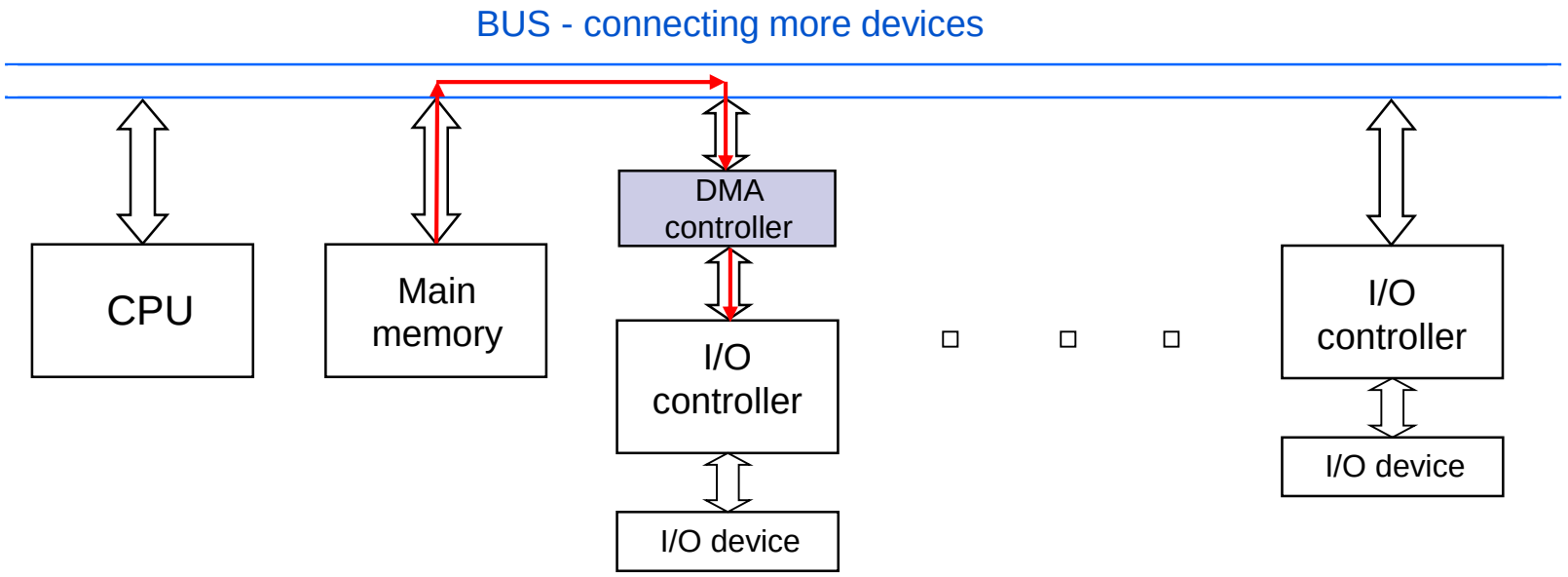


- Direct memory access (Direct memory access - DMA)
  - CPE begins and ends transfer, transfer is controlled by DMA controller.
  - Requires DMA controller.
  - Transfer progress (read from I/O)
    - I/O → DMA controller → memory



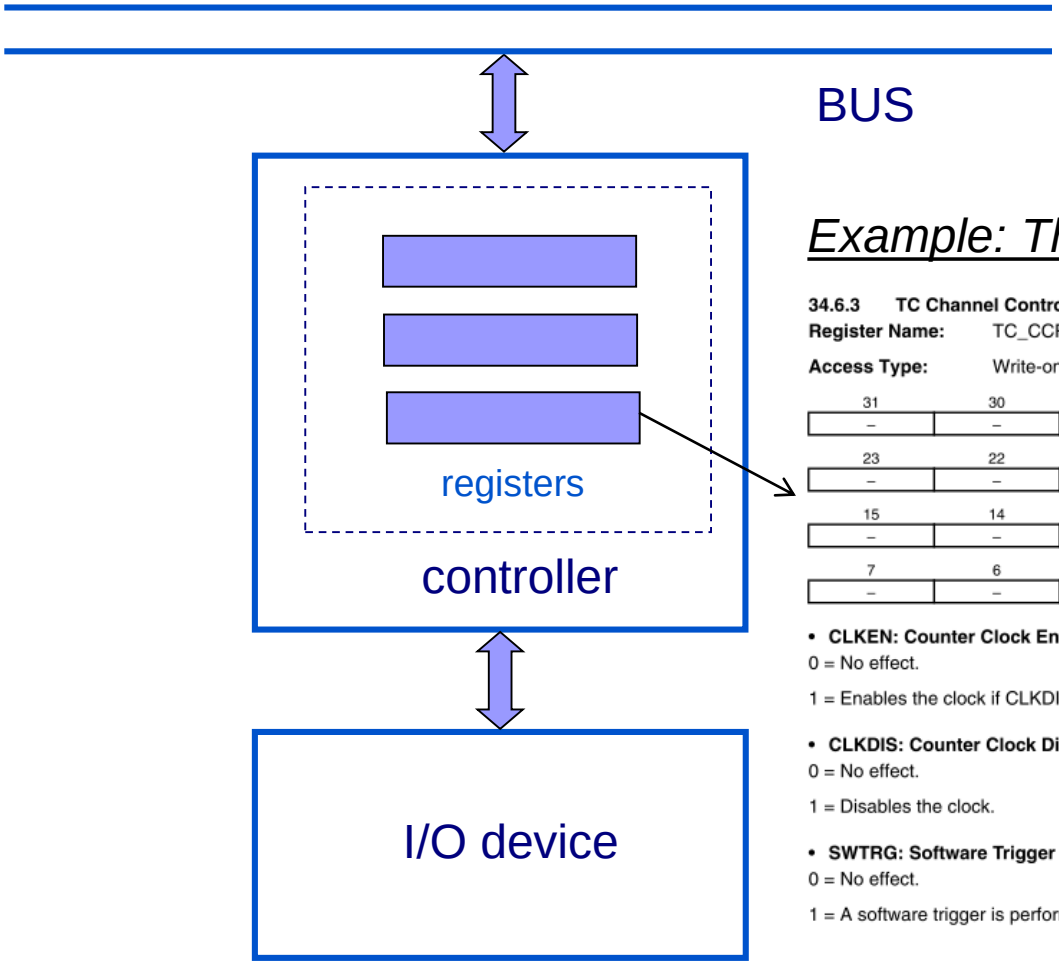
Direct memory access (Direct memory access - DMA)

- Transfer progress (write to I/O)
  - Memory → DMA controller → I/O device



Controller of I/O devices

- Each I/O device is connected through a controller device.
- CPU sees controller like a certain number of registers to be written or read from them.
- To write (or read) in a particular register (command register) can trigger the operation of the I/O device (device command).
- Read from a particular register (status register) reflects the status of the device after the I/O operation (device status).
- Even the transfer of information is carried out by reading from the specific controller register or by writing to a controller register (data register).



Example: The Timer control register on ARM9

34.6.3 TC Channel Control Register

Register Name: TC_CCRx [x=0..2]

Access Type: Write-only

31	30	29	28	27	26	25	24
–	–	–	–	–	–	–	–
23	**22**	**21**	**20**	**19**	**18**	**17**	**16**
–	–	–	–	–	–	–	–
15	**14**	**13**	**12**	**11**	**10**	**9**	**8**
–	–	–	–	–	–	–	–
7	**6**	**5**	**4**	**3**	**2**	**1**	**0**
–	–	–	–	–	SWTRG	CLKDIS	CLKEN

- **CLKEN: Counter Clock Enable Command**

0 = No effect.

1 = Enables the clock if CLKDIS is not 1.

- **CLKDIS: Counter Clock Disable Command**

0 = No effect.

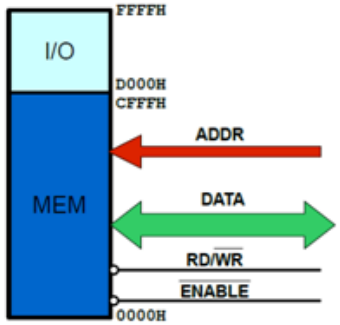
1 = Disables the clock.

- **SWTRG: Software Trigger Command**

0 = No effect.

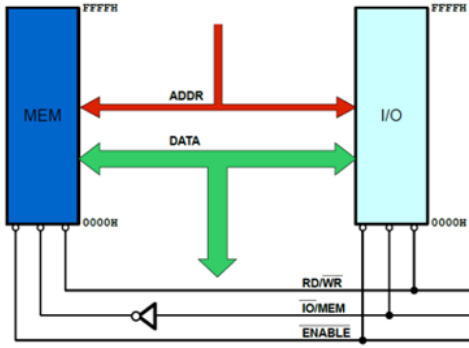
1 = A software trigger is performed: the counter is reset and the clock is started.

Access to registers in I/O controller



- **The memory mapped input/output** (memory mapped I/O) - registers the controls from the CPU will look the same as the memory locations. Headlines registers occupy memory address space.

I/O Mapped I/O (Port I/O)



- **Separate input/output address space**- need for specific input/output instructions to access the registers in controllers. Register in controller and memory location can have the same address.

- **Indirect addressing through I/O processors** - registers on controllers are in a separate address space to which only I/O processors have access .

Transmission (transfer) paths in von Neumann‘s computer

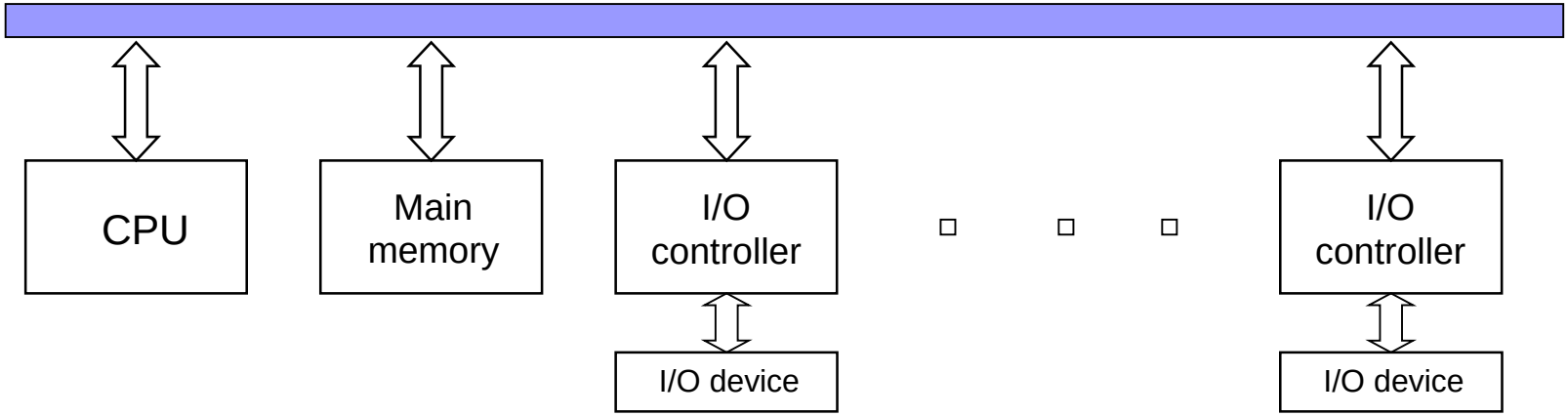
- They form the link between the CPU, main memory and I/O system.
- Fast computer requires fast links that allow you to transfer as much data per time unit.

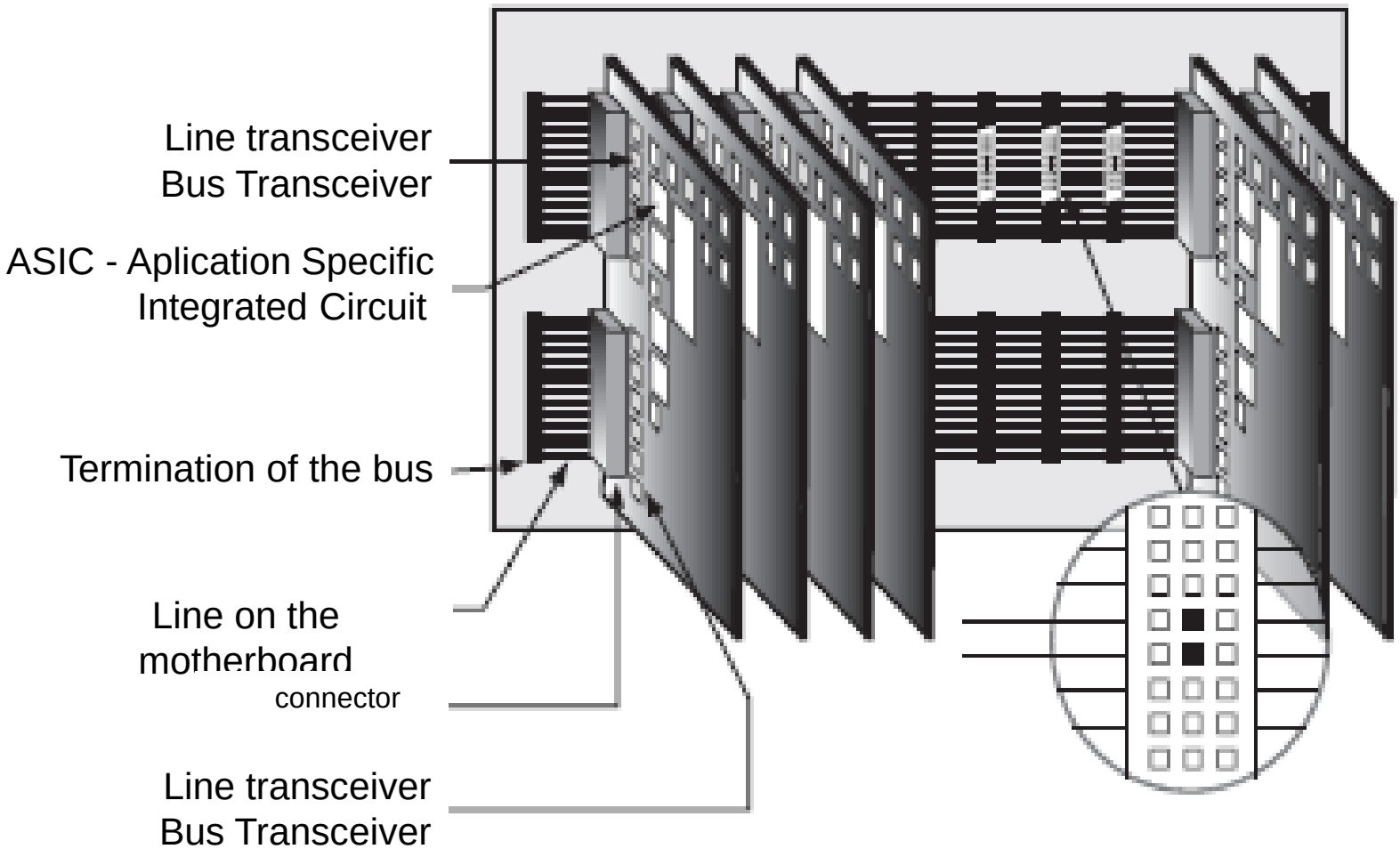
- One transmission path allows only single transfer at the same moment.
- The amount of information that can be transferred over a certain path, can be increased:
  - with shortening the time required for one transfer,
  - increasing the number of bits transferred in one transfer (= by increasing the width of the transmission path).

- The speed is also affected by the number of transmission paths – we talk about connection structures.
- Types of transmission paths:
  - Bus
  - Point to point

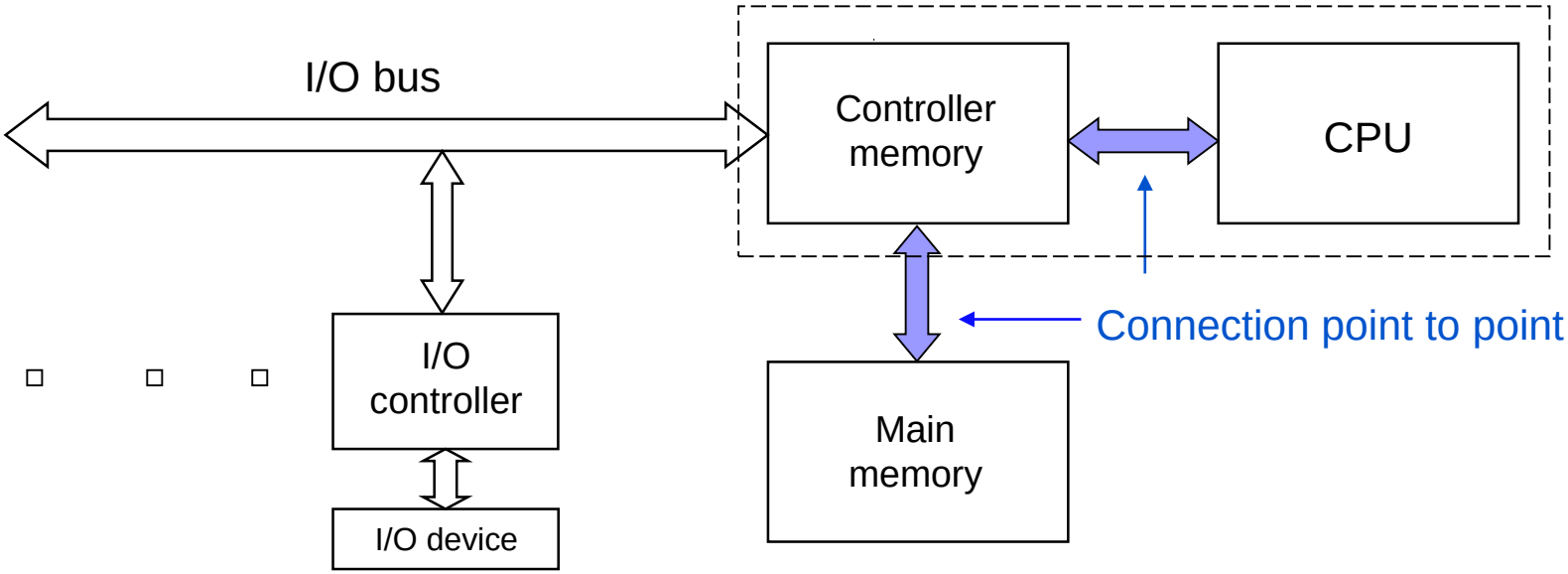
- Bus
  - Bus is shared between all units connected to it.
  - Physically is bus a set of parallel electrical lines (wires), that carry the electrical signals.
  - Lines have branches, through which the units are connected to the bus.

BUS - connecting more devices





- Point to point connection
  - Transmission path which connects only two devices (PCIe, USB3, SATA, ...)



Signals transmitted through transmission paths

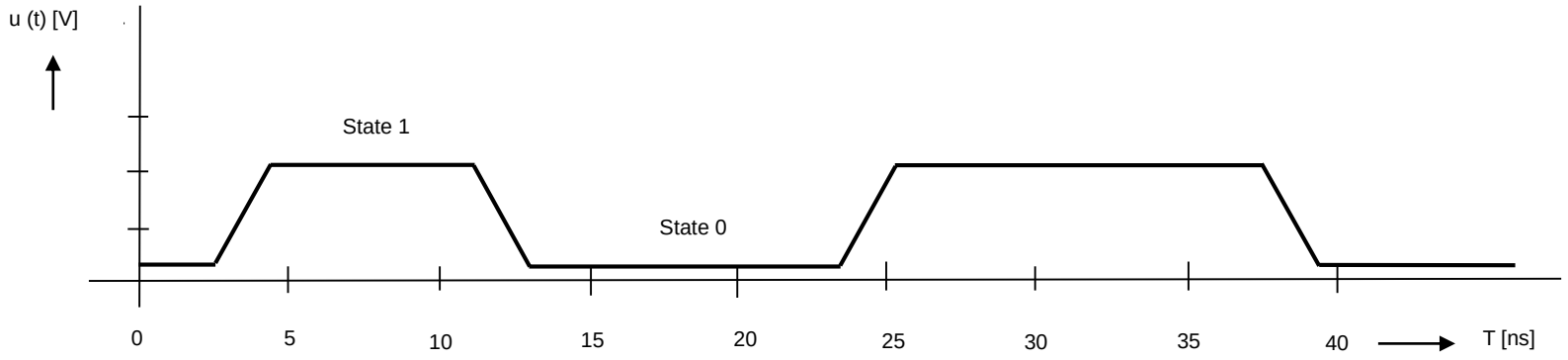
- On the lines that make up the transmission path, three types of signals are transmitted :
  - address signals
  - control signals
  - data signals

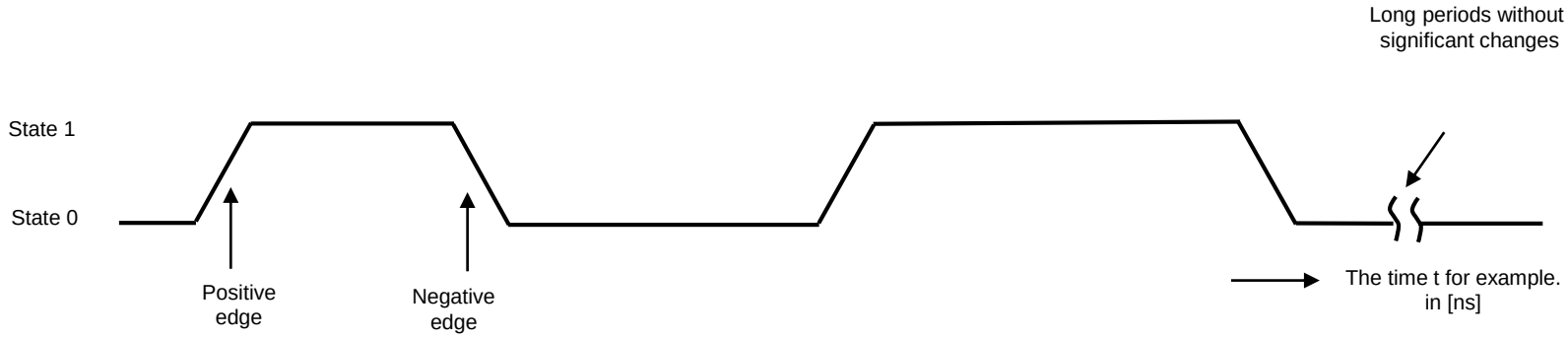
- Address signals
  - Determine the memory address of a word or I/O devices (register in I/O device controller) to which the transmission relates.
  - Number of address signals (bits) determines the size of the address space.
  - Conventional labels are e.g. for the 32-bit address A0 - A31

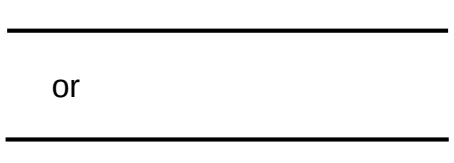
- Control signals
  - Determine the direction of transfer (read or write), the number of bits transferred and the chronology of events in the transfer etc.
  - Some control signals:
    - clock signal (a common designation is CLK),
    - random access signal (designation R/W);
    - signal to reset the processor (designation RESET)
    - interrupt request (designation INT or IRQ)
    - . . .

- Data signals
  - Number of data signals (lines) is equal to the number of bits that are at once transmitted through a transmission path - the width of the transmission path.
  - With the width of 64 bits, 8-, 16-, 32- and 64-bit transfers are possible.
  - Conventional labels are e.g. for 64-bit width D0 - D63

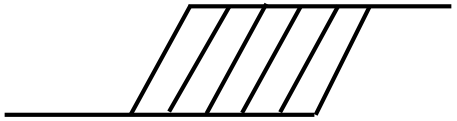
Introduction to Signal timing diagrams



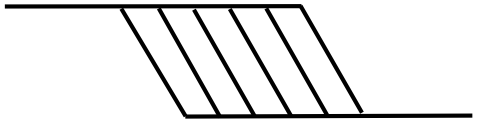




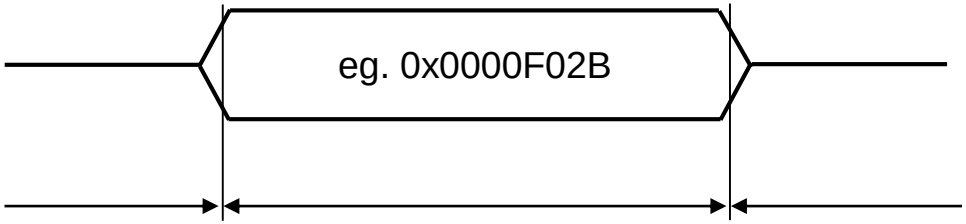
stable (constant) signal.



Signal changes from
state 0 to the state 1



Signal changes from
statet 1 to the state 1

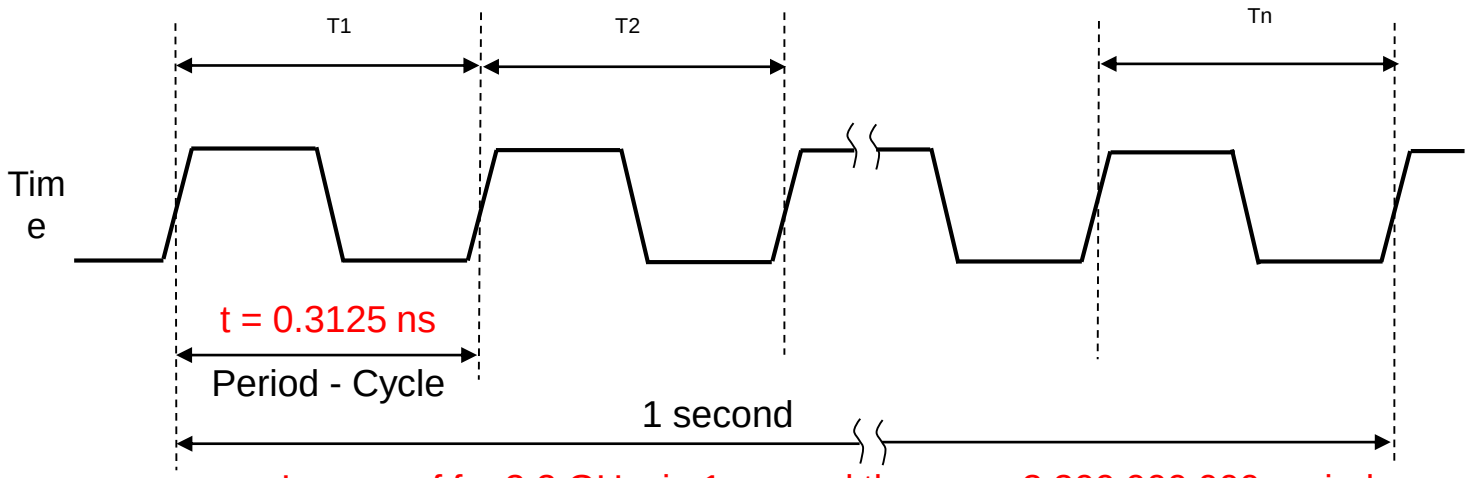


Presentation of signals
eg. on data lines

Each of the lines can be
in the state 0 or 1,

All lines are disconnected
- the third state (Hi-Z output)

Clock signal - a periodic rectangular signal



In case of f = 3.2 GHz, in 1 second there are 3 200 000 000 periods

The frequency of the periodic signal f = is the number of periods (cycles) in 1 second

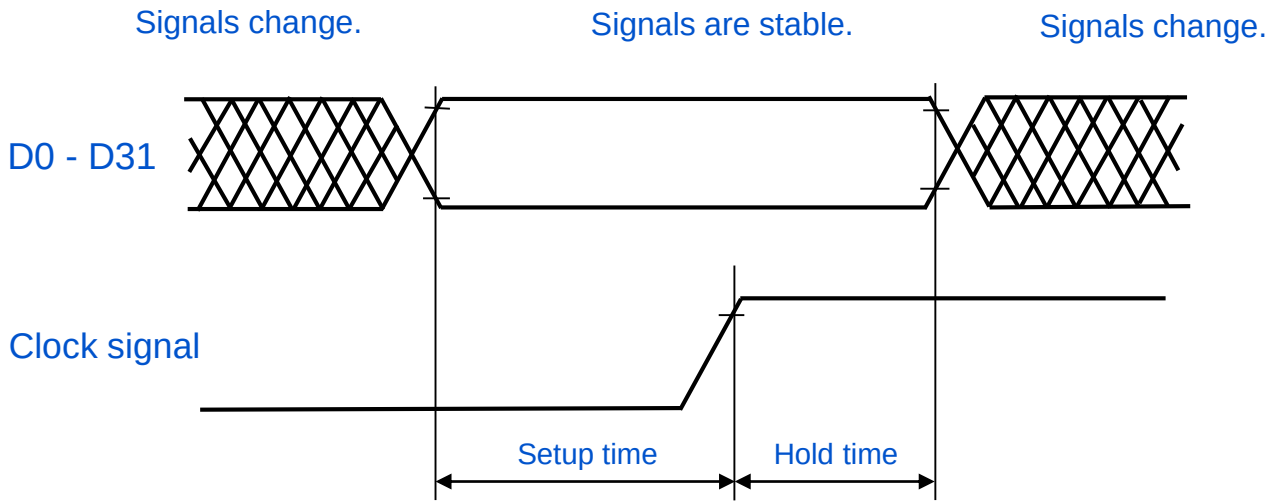
The unit of frequency is Hertz (Hz): 1 Hz = 1 period/s = 1 s^{-1}

The duration of one period T = 1 / f

Example: f = 3.2 GHz

$$t = \frac{1}{f} = \frac{1}{3{,}2 \cdot GHz} = \frac{1}{3{,}2 \cdot 10^9 \frac{1}{s}} = 0{,}3125 \cdot 10^{-9}\, s = 0{,}3125 ns$$

Using the clock signal



Role of devices in transfer

- Each transfer takes place only between two devices.
- Roles of devices at transfer
  - master – controls the transfer
  - slave

- On the bus, there could several potential masters - only one can be active at each moment.
- In connection point to point, there is only one master.
- When there are more masters on the bus (co-processor, DMA controller, etc.), the arbitration is required.
- Arbitration - a mechanism to determine, who from the masters will have the control of the transfer and the bus.

Sequence of events in the transfer

- Address signals
  - Master sends the address of the memory location on transmission path (address signals are placed in a state which corresponds to the target address).
- Control signals
  - Master determines the direction of the transfer:
    - read transfer
    - write transfer
  - Number of bits to be transferred (if transfers of different widths are possible)

- Data signals
  - When the transfer direction is read data, the master expects data on incoming transmission path (data signals).
  - When the transfer direction is write data, the master sends also data on the transmission path (data signals).
- Start and end of transfer
  - Master device uses the control signals also to dertermine start and end of the transfer.

- All units connected to the transmission path (slaves), compare their addresses to the address on the address lines.
- Unit (slave), which determines equality in address, performs the transfer required by the master.
- **! Transfer:** a single parallel transfer of as many bits as the width of a data transmission path.

Transmission (transfer) bandwidth (capacity)

- Capacity of transmission (transfer) path B (bandwidth, Throughput) is the maximum number of:
  - transfers in second (T/s)
    (MT/s = 10^6 T/s, GT/s = 10^9 T/s)
  - transferred bits per second (b/s, Mb/s, Gb/s)
  - transferred bytes per second (B/s, MB/s, GB/s)
    - MB/s = 10^6 B/s; GB/s = 10^9 B/s

$$B = f_{bus} \cdot \frac{1}{periods\_per\_transfer} \cdot Bus\_width$$

- B = Capacity (bandwidth) of the transmission path in bits/second (or bytes/second, if the width is in bytes)
- f_{bus} = Frequency of the clock signal on the bus in [Hz] = [1/s] (= number of clock periods in 1 second)
- *Periods_per_transfer* = Duration of a transfer in clock periods
- *Bus_width* = the number of transferred bits (or bytes) n one transfer

- Example: PCI bus bandwidth
  - Width of the bus = 32 bits (= 4 bytes = 4 B)
  - Frequency of clock signal on the bus f_{bus}= 33 MHz
    (33MHz = 33 * 10^6 Hz = 33 * 10^6 s^{-1} !)
  - One transfer takes one clock period.

$$B\,[b/s] = 33 \cdot 10^6 [1/s] \cdot 1\,[period\ /transfer] \cdot 32[bits] = 1056 \cdot 10^6\ [b/s]$$

$$B\,[B/s] = \frac{1056 \cdot 10^6 [b/s]}{8[b/B]} = 132 \cdot 10^6\ [B/s] = 132\,[MB/s]$$

  - ali

$$B\,[B/s] = 33 \cdot 10^6 [1/s] \cdot 1[\,perioda\ /\ prenos] \cdot 4[B] =$$

$$= 132 \cdot 10^6 [B/s] = 132[MB/s]$$

Types of transmission paths	Bandwidth in bits/s	Bandwidth in bytes/s
ISA (8 bits, 4.77 MHz)	9.6 Mb / s	1.2 MB / s
PCI (32 bit, 33 MHz)	1056 Mb / s	133 MB / s
PCI (64 bit, 66 MHz)	4224 Mb / s	528 MB / s
PCI Express 2.0 (x16)	64 Gb / s	8 GB / s
HT (AMD 3.2 GHz)	409.6 Gb / s	51.2 GB / s
QPI (Intel 3.2 GHz)	204.8 Gb / s	25.6 GB / s