



Ministerstwo Nauki  
i Szkolnictwa Wyższego

**Regionalna Inicjatywa Doskonałości w Dyscyplinach  
Informatyki, Elektrotechniki, Elektroniki, Automatyki i Robotyki  
na Politechnice Częstochowskiej**



**Projekt w ramach programu Regionalna Inicjatywa Doskonałości, decyzja nr 020/RID/2018/19**

# **Architektura Systemów Komputerowych**

## **Magistrale szeregowe i równoległe systemów komputerowych**

dr hab. inż. Łukasz Szustak, prof. PCz  
lszustak@icis.pcz.pl

Katedra Informatyki  
Wydział Inżynierii Mechanicznej i Informatyki  
Politechnika Częstochowska



Wydział Inżynierii  
Mechanicznej  
i Informatyki  
The Faculty of Mechanical Engineering  
and Computer Science



Ministerstwo Nauki  
i Szkolnictwa Wyższego

# Regionalna Inicjatywa Doskonałości w Dyscyplinach Informatyki, Elektrotechniki, Elektroniki, Automatyki i Robotyki na Politechnice Częstochowskiej



Projekt w ramach programu Regionalna Inicjatywa Doskonałości, decyzja nr 020/RID/2018/19

**Zamieszczane materiały służą wyłącznie do celów indywidualnego kształcenia. Nie wyrażam zgody na ich utrwalanie, przekazywanie osobom trzecim ani rozpowszechnianie.**

dr hab. inż. Łukasz Szustak, prof. PCz  
lszustak@icis.pcz.pl

Katedra Informatyki

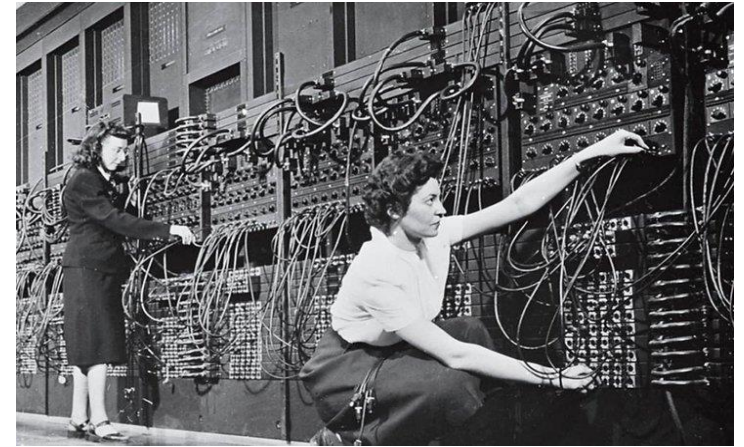
Wydział Inżynierii Mechanicznej i Informatyki  
Politechnika Częstochowska



Wydział Inżynierii  
Mechanicznej  
i Informatyki  
The Faculty of Mechanical Engineering  
and Computer Science

# Magistrala komunikacyjna

- W architekturze systemów komputerowych magistrala komunikacyjna to system, pozwalający na przesyłanie danych między komponentami komputera lub między komputerami
- Pierwsze komputery nie przypominały sprzętu, który znamy z czasów obecnych. Poszczególne elementy z racji swojej wielkości, nie były montowane na jednej płycie, jak obecnie
- Każdy z nich, np. procesor, pamięć, znajdowały się w różnych szafach połączonych ze sobą wiązkami przewodów
- Początkowo magistrale komunikacyjne były więc równoległymi przewodami elektrycznymi, łączącymi poszczególne komponenty komputera

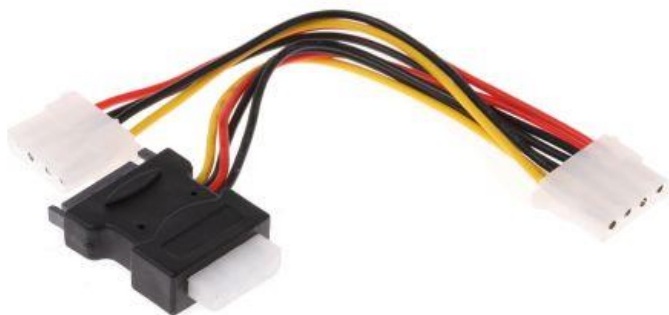


*Komputer Eniac, 1945*



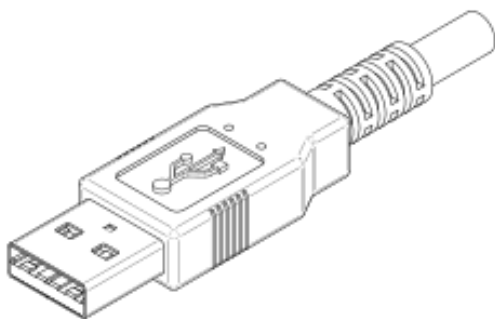
*Współczesne magistrale komunikacyjne*

# Klasyfikacja magistral komunikacyjnych



- **Magistrale wewnętrzne**

Umożliwiają wymianę danych pomiędzy komponentami znajdującymi się wewnątrz komputera. Dotyczą więc m.in. połączeń pomiędzy płytą główną, procesorem, pamięcią



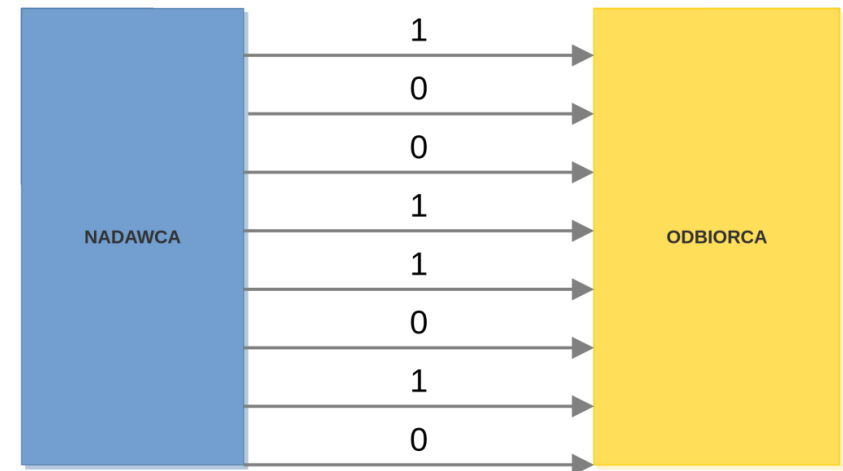
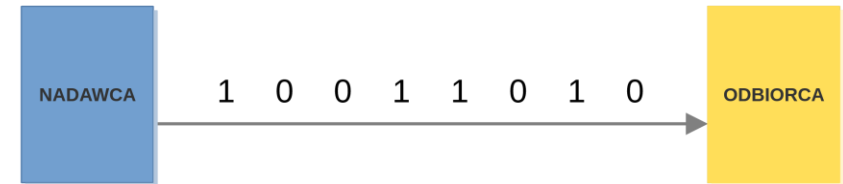
- **Magistrale zewnętrzne**

Zapewniają transfer danych pomiędzy komputerem a urządzeniami peryferyjnymi, np. monitorem, klawiaturą, myszą, pamięcią przenośną. Charakteryzują się mniejszą szybkością niż magistrale wewnętrzne

# Typ transmisji

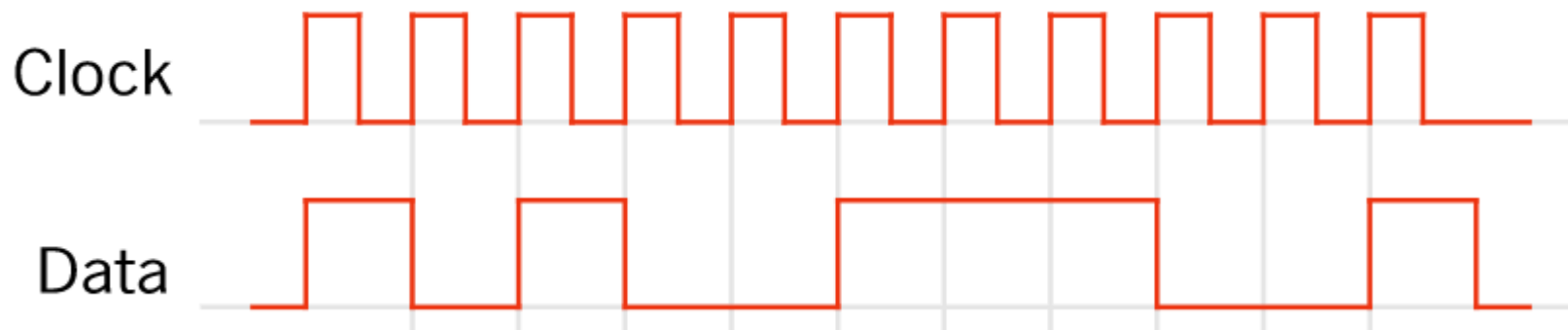
Ze względu na typ transmisji danych, magistrale komunikacyjne dzieli się na:

- **Magistrale szeregowe**  
Dane przesyłane są sekwencyjnie, bit po bicie, za pomocą jednego kanału transmisyjnego
- **Magistrale równoległe**  
Dane przesyłane są jednocześnie za pomocą wielu kanałów transmisyjnych



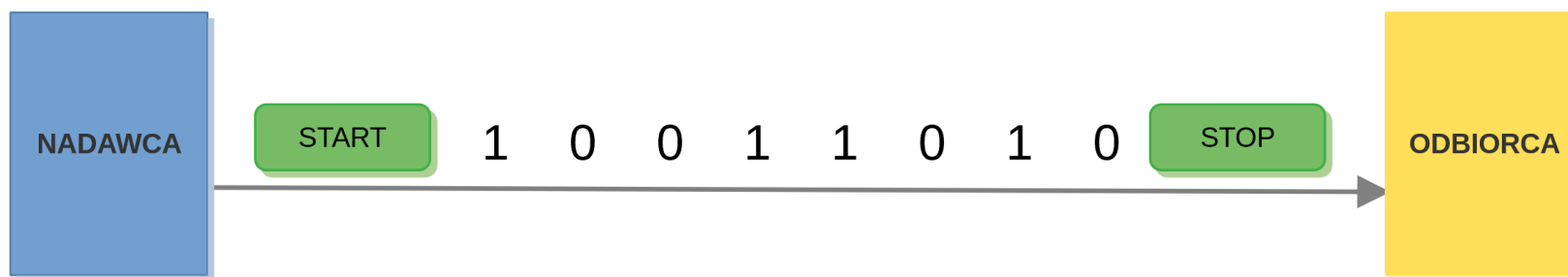
# Transmisja szeregową - synchroniczna

- Transmisja szeregową synchroniczna charakteryzuje się występowaniem sygnału zegarowego
- W transmisji synchronicznej nadajnik oraz odbiorca są taktowane taką samą częstotliwością
- Odbiornik odczytuje stan poszczególnych bitów w określonym momencie, którym jest np. narastające zbocze sygnału zegarowego
- Dane przesyłane są w postaci większych bloków, zawierających informacje kontrolne

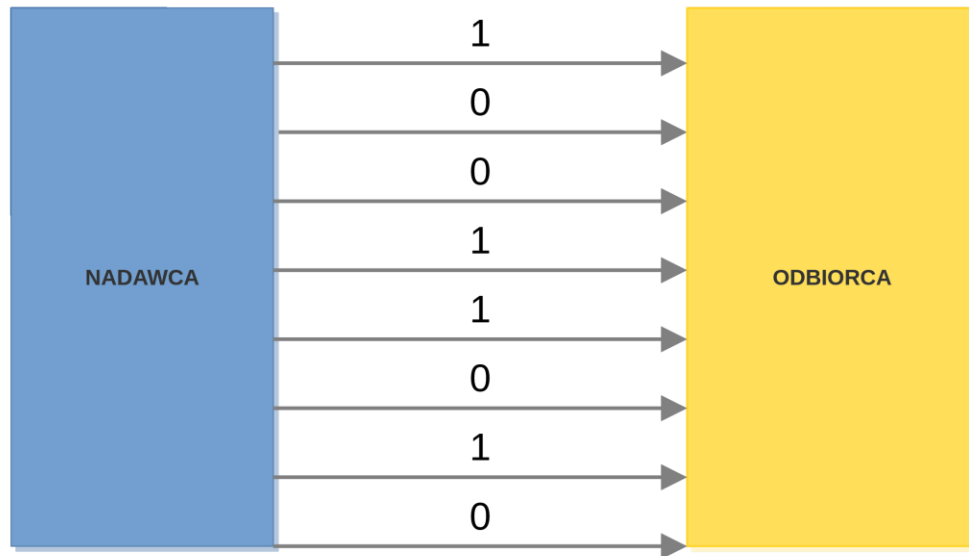


# Transmisja szeregową - asynchroniczna

- Transmisja szeregową asynchroniczna pozwala na przesyłanie danych w dowolnym momencie
- Dane przesyłane są w pakietach o określonej długości, np. 1 bajt
- Początek pakietu stanowią bity startu, służące synchronizacji nadawcy oraz odbiorcy
- Oprócz bitów danych, często przekazywany jest również bit parzystości, umożliwiając wykrycie zakłóceń w transmisji
- Koniec pakietu oznaczany jest za pomocą bitów stopu



# Transmisja równoległa



- Transmisja równoległa polega na przesyłaniu danych za pomocą wielu kanałów informacyjnych
- Teoretycznie pozwala to na zwielokrotnienie ilości przesyłanych danych na cykl zegara – zakładając, że każdy kanał pracuje z taką samą częstotliwością jak w przypadku transmisji szeregowej
- W praktyce bywa jednak różnie. Transmisja równoległa niesie ze sobą różne ograniczenia, które powodują, że w rozwiązaniach często stosuje się transmisję szeregową



# Zalety transmisji szeregowej

- Wymaga małej liczby przewodów, co powoduje, że ze względów ekonomicznych transmisja szeregowa jest zdecydowanie dominującym sposobem przesyłania danych na duże odległości
- Nowe technologie oferują coraz szybszą transmisję danych, co pozwala upowszechnić zastosowanie szeregowych magistral komputerowych także na krótszych odległościach
- Oszczędność miejsca, dzięki małej liczbie przewodów oraz mniejszym wtyczkom, złączom i gniazdom
- Większa odporność na zakłócenia
- Niski koszt wykonania – cena układu scalonego zależy od liczby pinów
- Łącza szeregowe często mogą pracować ze znacznie wyższą częstotliwością niż łącza równoległe

# Wady transmisji szeregowej

- Przesyłanie danych jednym kanałem ogranicza prędkość transmisji
- Aby osiągnąć odpowiednią prędkość transmisji, należy stosować dużą częstotliwość taktowania
- Problemy związane z przesunięciami czasowymi impulsów zegarowych (clock skew)
- Wymagane układy do zamiany danych na postać szeregową oraz równoległą

# Zalety transmisji równoległej

- Zastosowanie transmisji równoległej pozwala na przesłanie dużej ilości danych jednocześnie
- Transmisja równoległa jest najlepszym wyborem w przypadku komunikacji na małej odległości
- Implementacja sprzętowa transmisji równoległej jest prostsza niż transmisji szeregowej, gdzie w większości przypadków, dane w pierwszej kolejności muszą zostać przekonwertowane do postaci równoległej

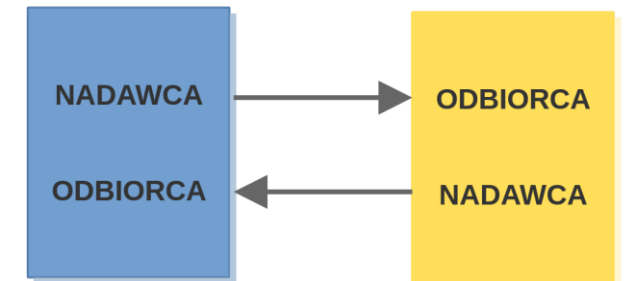
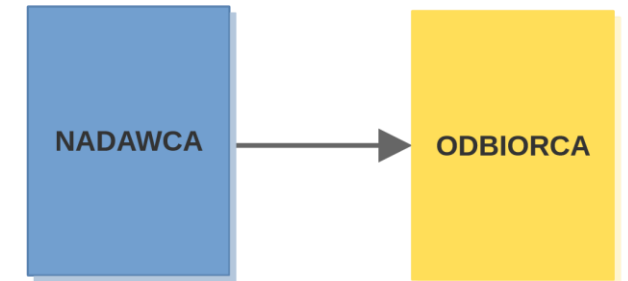
# Wady transmisji równoległej

- Wymaga dużej liczby przewodów, co powoduje, że jest rozwiązaniem drogim w porównaniu do transmisji synchronicznej
- Duża liczba przewodów wymaga dużego zapotrzebowanie na miejsce
- Sygnały na równoległych liniach wzajemnie się zakłócają
- Aby przesyłać dane na duże odległości, należy zwiększyć grubość przewodów, aby zmniejszyć zakłócenia sygnału
- Problemy z synchronizacją danych z wielu kanałów transmisyjnych

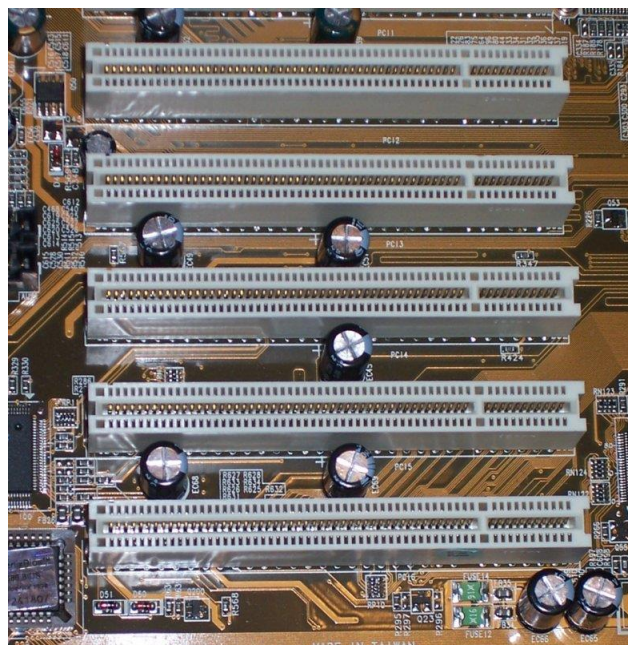
# Sposób transmisji

Magistrale komunikacyjne można również podzielić ze względu na kierunek przepływu danych:

- **magistrale jednokierunkowe (simplex)**  
dane przesyłane tylko w jednym kierunku – od nadawcy do odbiorcy
- **magistrale dwukierunkowe (duplex)**  
dane mogą przepływać w obu kierunkach, z wyszczególnieniem dwóch przypadków:
  - **pełny dupleks (full duplex)**  
dane mogą być przesyłane jednocześnie w obu kierunkach
  - **pół dupleks (half duplex)**  
dane mogą być przesyłane w obu kierunkach, ale nie jednocześnie



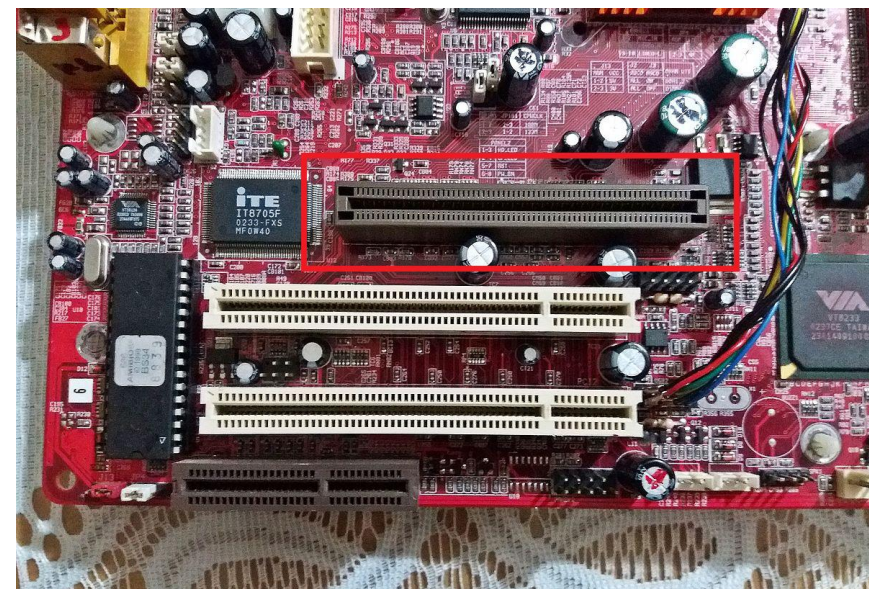
# Magistrala PCI



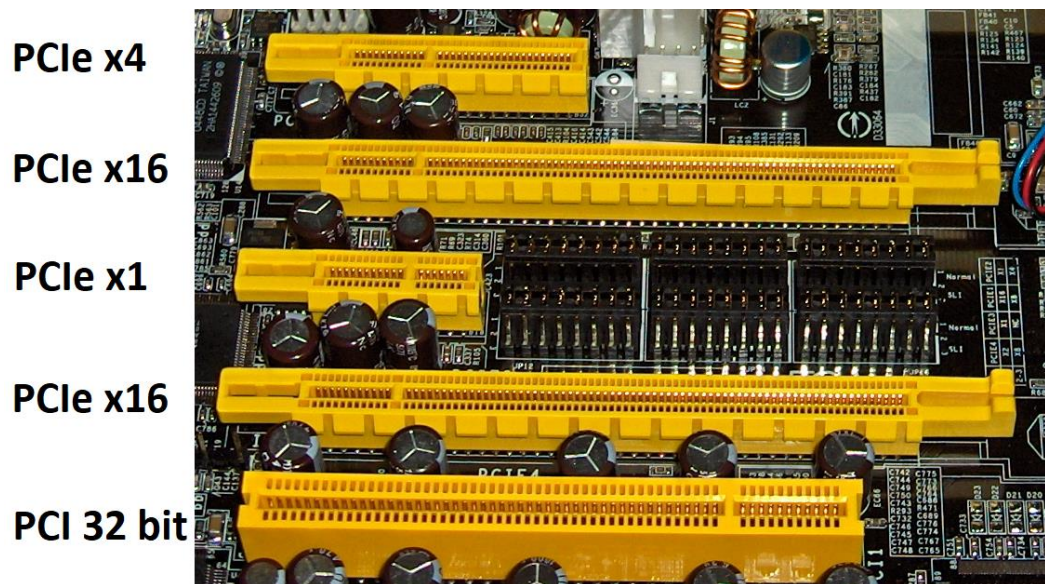
- **Peripheral Component Interconnect (PCI)** – magistrala równoległa umożliwiająca przyłączenie kart rozszerzeń do płyty głównej. Pracę nad magistralą PCI rozpoczęto w latach 90. XX wieku
- Magistrala PCI występowała w dwóch wersjach: 32 i 64 – bitowej
- Maksymalna przepustowość przy szerokości szyny danych 64 bit oraz częstotliwości taktowania 66 MHz to 532 MB/s
- Łatwo skalowalna – w komputerze może znajdować się kilka magistral PCI połączonych szeregowo lub równolegle
- Kompatybilna z rozwiązaniami różnych producentów
- Szybka instalacja karty za pomocą (Plug and Play)
- Została wyparta przez magistralę szeregową PCI Express

# Magistrala AGP

- **Accelerated Graphics Port (AGP)** – modyfikacja magistrali PCI zaprojektowana pod koniec XX wieku przez firmę Intel w celu obsługi kart graficznych
- 32 bitowa magistrala równoległa, której celem było jak najszybsze przesyłanie danych pomiędzy pamięcią operacyjną a kartą graficzną
- Maksymalna przepustowość osiągnięta przy częstotliwości taktowania 66 MHz w wersji AGP 8x to 2133 MB/s.
- Przez kilka lat stanowiła najbardziej wydajny interfejs kart graficznych.
- Została wyparta przez magistralę szeregową PCI Express



# Magistrala PCI Express



- **Peripheral Component Interconnect Express (PCIe)** magistrala szeregową, opracowaną wspólnie przez wiele firm w 2004 r, umożliwiającą instalację kart rozszerzeń na płycie głównej
- Zastąpiła starsze równoległe standardy PCI i AGP
- W odróżnieniu do magistrali PCI, PCIe działa w trybie *point to point*. W podejściu wykorzystanym w PCI, wszystkie urządzenia współdzielą pasmo przenoszenia danych, co powoduje, że prędkość transmisji spada wraz z liczbą podłączonych do szyny komponentów. Technologia *point to point* zakłada, że każde urządzenie jest podłączone bezpośrednio do kontrolera magistrali, co wyklucza potrzebę współdzielenia pasma



# Magistrala PCI Express

- Liczba linii pomiędzy dwoma urządzeniami to od 1 do 16. Specyfikacja zakłada również możliwość wykorzystania 32 linii, jednak takie rozwiązania nie są stosowane.
- Dane mogą być przekazywane pomiędzy urządzeniami w obu kierunkach (full duplex)
- Wielokanałowa konstrukcja szeregową zwiększa elastyczność dzięki możliwości przydzielania mniejszej liczby pasów dla wolniejszych urządzeń
- Obecnie na rynku implementowany jest standard 4.0 zakładający transfer 16 GT/s oraz przepustowość 31.5 GB/s w wersji x16.
- W 2021 r. ma się ukazać specyfikacja standardu 6.0, który powinien oferować transfer 64 GT/s oraz przepustowość 126 GB/s w wersji x16.

# Magistrala PCI Express

- Aby obliczyć przepustowość magistrali PCI Express należy pomnożyć liczbę transferów przez liczbę linii oraz uwzględnić sposób kodowania.
- Sposób kodowania zmieniał się w kolejnych wersjach standardu.
- Wersja 1.0 oraz 2.0 korzystały z kodowania 8b/10b, co oznacza, że pakiety 8-bitowe, ze względu na potrzeby transmisji szeregowej przesyłane są z wykorzystaniem 10 bitów. Strata przepustowości wynosi więc 20%.
- Wersja 3.0 i nowsze korzystają z kodowania 128b/130b, dzięki czemu strata wynosi około 1.5 %/.

# Magistrala PCI Express - przykłady

- Obliczyć przepustowość PCIe 1.0 x4.
  - Standard PCIe w wersji 1.0 zapewnia transfer 2.5 GT/s czyli 2.5 Gb/s
  - Liczba linii wynosi 4
  - Kodowanie 8b/10b

$$B = 2.5 \cdot 4 \cdot \frac{8}{10} \approx 8 \text{ Gb/s} \approx 1 \text{ GB/s}$$

# Magistrala PCI Express - przykłady

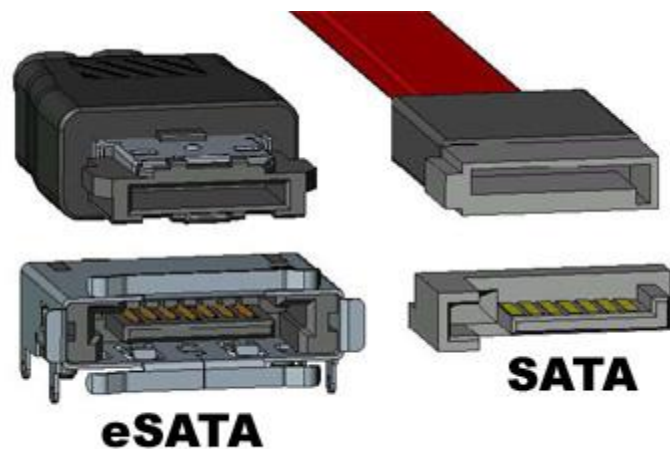
- Obliczyć przepustowość PCIe 3.0 x16.
  - Standard PCIe w wersji 3.0 zapewnia transfer 8 GT/s czyli 8 Gb/s
  - Liczba linii wynosi 16
  - Kodowanie 128b/130b

$$B = 8 \cdot 16 \cdot \frac{128}{130} \approx 126 \text{ Gb/s} \approx 15.75 \text{ GB/s}$$

# Magistrala PCI Express - przepustowość

Wersja	Kodowanie	Transfer [GT/s]	Przepustowość [GB/s]				
			1x	2x	4x	8x	16x
1.0	8b/10b	2.5	0,25	0,50	1,00	2,00	4,00
2.0	8b/10b	5.0	0,50	1,00	2,00	4,00	8,00
3.0	128b/130b	8.0	0,98	1,97	3,94	7,88	15,75
4.0	128b/130b	16.0	1,97	3,94	7,88	15,75	31,51
5.0	128b/130b	32.0	3,94	7,88	15,75	31,51	63,02
6.0	128b/130b	64.0	7,88	15,75	31,51	63,02	126,03

# Magistrala SATA

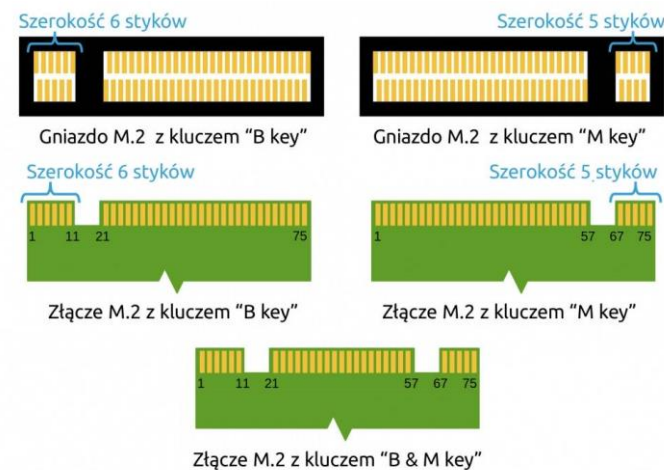
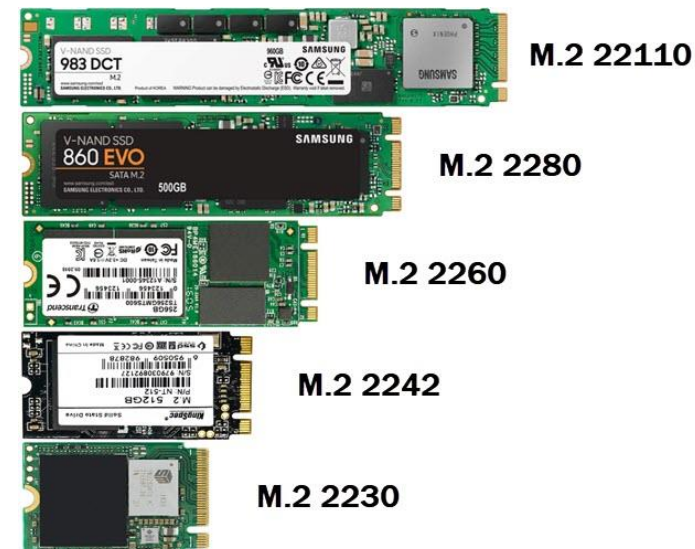


Dysk SSD mSATA

- **Serial Advanced Technology Attachment (SATA)** – szeregową magistralą, będącą następcą magistrali równoległej ATA
- Wykorzystywana do komunikacji między adapterami magistrali hosta a urządzeniami pamięci masowej (dysk twardy, SSD, napęd optyczny, napęd taśmowy)
- Od 2009 r. występuje w trzeciej generacji, która umożliwia przepustowość około 6 Gb/s
- Zewnętrzny port pozwalający na podłączanie pamięci masowych nosi nazwę eSATA. Wykorzystuje on solidniejsze złącze oraz dłuższe ekranowane kable o maksymalnej długości 2m (SATA – 1 m)
- Do niewielkich urządzeń, takich jak dyski SSD czy notebooki zaprojektowano złącze mini-SATA (mSATA).

# Standard M.2

- Następcą mSATA jest standard M.2
- Standard M.2 przewiduje kilka szerokości oraz długości dysków. Obecnie dominują dyski o szerokości 22mm i długości 30-110 mm.
- Rozmiary oznaczane są za pomocą kodu, np. 2260 oznacza dysk o szerokości 22mm i długości 60mm.
- Złącze M.2 występuje w różnych wersjach oznaczanych za pomocą kluczy „M” oraz „B”, określających możliwość podłączenia do magistral SATA oraz PCIe.



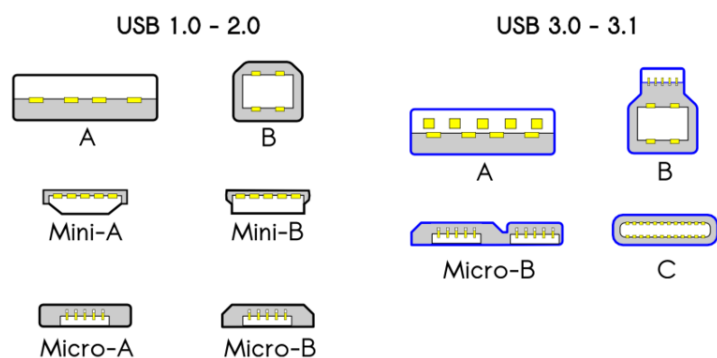
# Standard M.2

Interfejs dysku M.2	Przepustowość [GB/s]
SATA III	0.60
PCIe 2.0 x2	1.00
PCIe 2.0 x4	2.00
PCIe 3.0 x4	3.94
PCIe 4.0 x4	7.88

- Standard M.2 współpracuje z magistralami SATA, PCIe oraz USB.
- Zastosowanie magistrali PCIe pozwala na znaczne zwiększenie prędkości transmisji.
- W przypadku SATA III maksymalna prędkość wynosi 600 MB/s (6 Gb = 750 MB przy kodowaniu 8b/10b).
- Standard M.2 pozwala na skorzystanie z 4 linii PCIe, co w wersji 4.0 daje przepustowość około 8GB (kodowanie 128b/130b).



# Magistrala USB

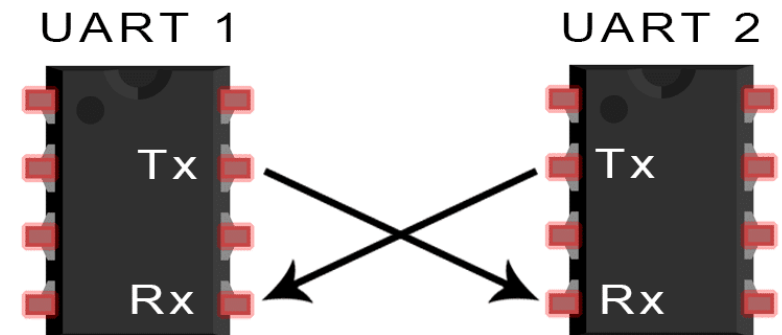


IEEE1284

- **Universal Serial Bus (USB)** – magistrala szeregową opracowana przez kilka firm w 1996 r.
- Umożliwia podłączenie maksymalnie 127 urządzeń
- Mocno rozpowszechniona, używana w komunikacji z wieloma urządzeniami, np. mysz, klawiatura, smartfon, skaner, pendrive itp.)
- Umożliwia dynamiczne podłączanie i rozłączanie urządzeń, w trakcie pracy komputera
- W 2019 r. ogłoszono specyfikację USB w wersji 4.0, która zakłada przepustowość 40 Gb/s
- Dane mogą być przekazywane w obu kierunkach (full duplex)
- Port USB zastąpił wiele starych portów równoległych. Przykładem może być IEEE1284, zwany LPT, który służył do podłączania urządzeń peryferyjnych, np. drukarek

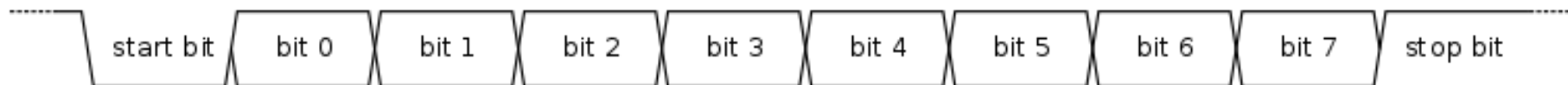
# UART

- **Universal Asynchronous Receiver and Transmitter (UART)** to układ scalony wykorzystujący transmisję szeregową w wersji asynchronicznej, znajdujący swoje zastosowanie w przeważającej większości mikrokontrolerów.
- Do komunikacji używa portu RS-232, który wraz z upowszechnieniem USB przestał być standardowym wyposażeniem komputerów. Możliwa jest jednak komunikacja przez USB za pomocą konwerterów wbudowanych w układ scalony.
- UART zawiera również konwertery równoległo-szeregowo oraz szeregowo-równoległe, pozwalające na konwersję danych odbieranych i wysyłanych do komputera.
- Urządzenia wspierające UART wykorzystują 2 piny: TX oraz RX. Pin TX służy do nadawania (transmit), natomiast pin RX do odbierania (receive).



# UART

- Transmisja asynchroniczna charakteryzuje się brakiem zegara, więc synchronizacja nadajnika oraz odbiornika polega na szybkość transmisji - jej parametry muszą być jednakowe dla nadajnika oraz odbiornika. Dane mogą być przesyłane z różną szybkością określoną w bitach na sekundę (baud – rate). Najczęściej spotykane warianty to 9600, 56700 i 115200.
- UART może korzystać z zarówno z transmisji jednostronnej (simplex) jak i dwustronnej (duplex) w oby dwu wariantach: pół dupleks (half duplex), pełny dupleks (full duplex)
- Bit startu to logiczne zero. Jego zadaniem jest zasygnalizowanie odbiornikowi rozpoczęcia transmisji. W zależności od konfiguracji po bicie startu następuje od 5 do 9 bitów danych. Transmisja kończy się bitem stopu, będącym logiczną jedynką.

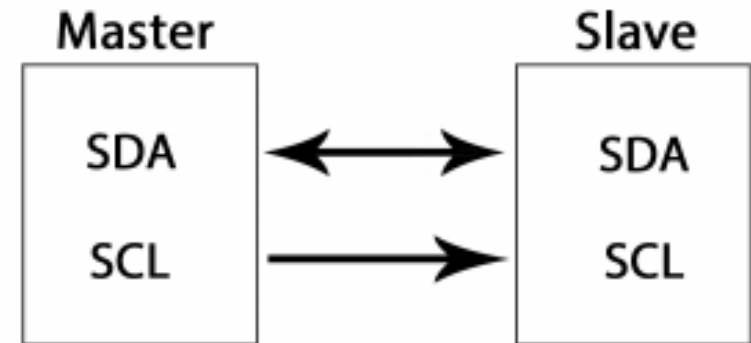


# I<sup>2</sup>C

- **Inter-Integrated Circuit (I<sup>2</sup>C)** to opracowana przez firmę Philips dwukierunkowa magistrala szeregowo synchroniczna, umożliwiająca komunikację między mikrokontrolerami a urządzeniami peryferyjnymi.
- Charakteryzuje się niewielką prędkością transmisji (w zależności od trybu osiąga od 0.1 do maksymalnie 5 Mb/s) oraz niskim kosztem.
- Stosowana jest więc w prostych rozwiązaniach, m.in. w systemach wbudowanych, np. do komunikacji z czujnikami i elementami wykonawczymi, przetwornikami cyfrowo-analogowymi i analogowo-cyfrowymi czy sterowaniem diodami LED.
- I<sup>2</sup>C jest magistralą adresowalną - siedmio- lub dziesięciobitowe pole adresowe umożliwia wysyłanie wiadomości do wybranego urządzenia przez urządzenie nadrzędne.

# I<sup>2</sup>C

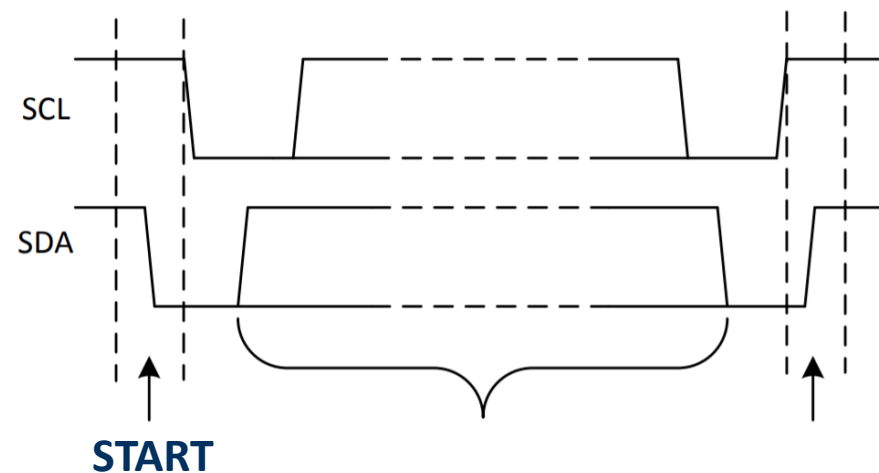
- Liczba węzłów podłączonych do magistrali I<sup>2</sup>C ograniczona jest przestrzenią adresową a także całkowitą pojemnością magistrali wynoszącą 400 pF, co obniża maksymalne odległości komunikacyjne do kilku metrów.
- W wersji podstawowej tylko jedno urządzenie (master) może inicjować transmisję. Wersja rozszerzona umożliwia pracę w trybie multi-master, dzięki mechanizmowi wykrywania kolizji.
- W przypadku jednoczesnego nadawania, pierwszeństwo ma urządzenie o niższym numerze.
- Do transmisji wykorzystuje dwie dwukierunkowe linie: linię danych SDA (Serial Data Line) oraz linię zegara SCL (Serial Clock Line).



# I<sup>2</sup>C

Proces przesyłania danych między urządzeniem głównym „master” i podrzędnym „slave”:

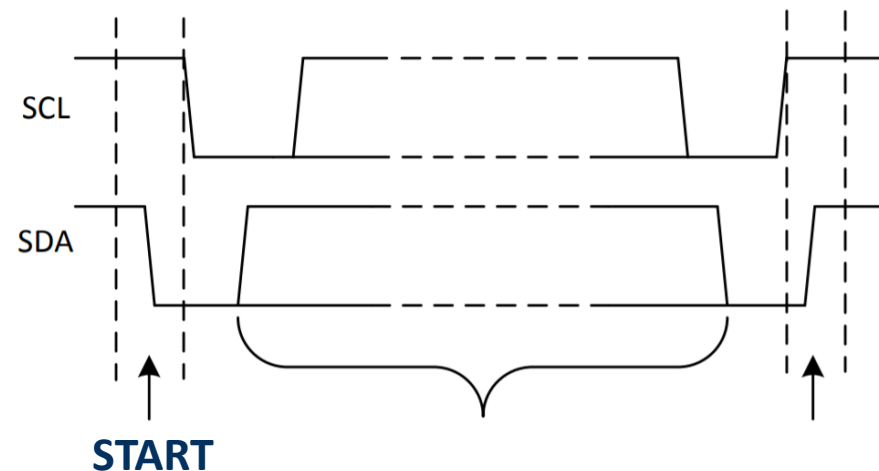
1. Urządzenie master wysyła sygnał transmisji do każdego podłączonego urządzenia slave zmieniając stan linii danych SDA z wysokiego na niski. Po wykonaniu tej operacji, stan niski ustawiany jest również na linii SCL.



# I<sup>2</sup>C

Proces przesyłania danych między urządzeniem głównym „master” i podrzędnym „slave”:

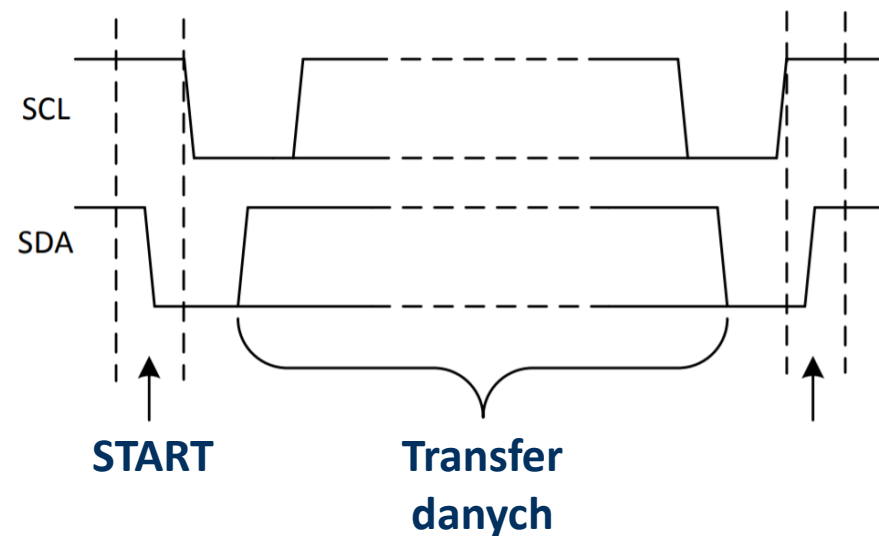
2. Urządzenie master przesyła adres docelowego urządzenia slave, do wszystkich podłączonych urządzeń slave. Każde urządzenie slave sprawdza przesłany adres. Urządzenia slave, które nie są odbiorcami zostają odłączone od linii SCL oraz SDA. Urządzenie, które jest odbiorcą (adres się zgadza) przesyła bit ACK, który zmienia stan linii SDA w wysoki, czym rozpoczyna transmisję danych.



# I<sup>2</sup>C

Proces przesyłania danych między urządzeniem głównym „master” i podrzędnym „slave”:

3. Dane przesyłane są w postaci ramek, każda udana transmisja kończy się przestaniem przez urządzenie slave bitu ACK do urządzenia master.

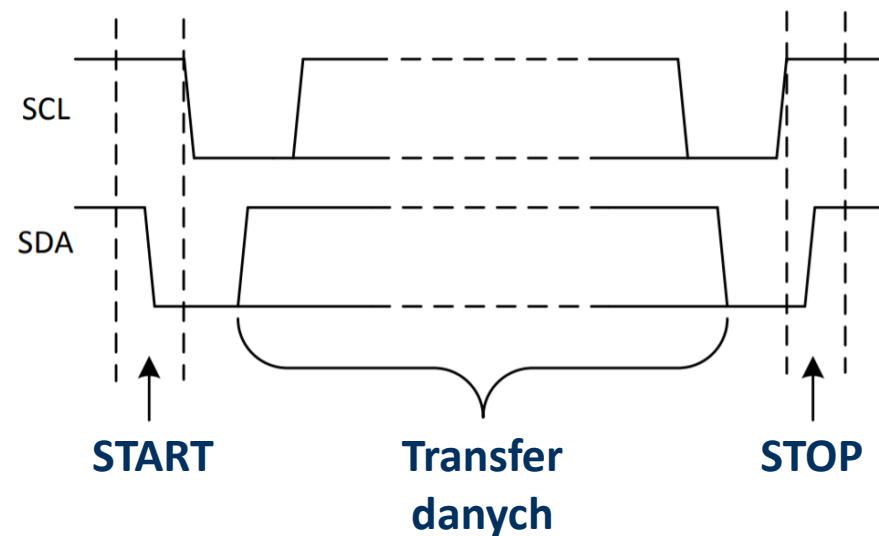




# I<sup>2</sup>C

Proces przesyłania danych między urządzeniem głównym „master” i podrzędnym „slave”:

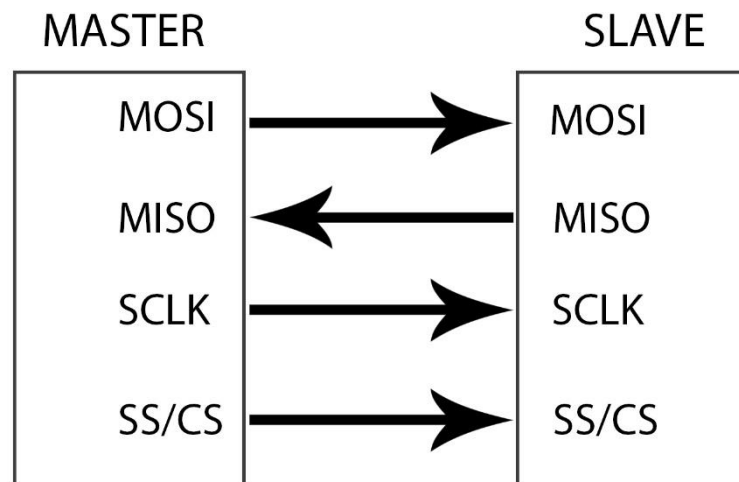
4. Zakończenie transmisji następuje, gdy urządzenie master wysyła do urządzenia slave sygnał stopu, zmieniając kolejno stany linii SCL oraz SDA na wysokie.



# SPI

- **Serial Peripheral Interface (SPI)** to protokół komunikacji szeregowej w wersji synchronicznej.
- Jego zaletą jest brak sygnału startu i stopu, co pozwala na ciągłe przesyłanie danych bez przerw. 1 bajt danych wysyłany jest co 8 cykli zegarowych generatora.
- W SPI dane mogą być przesyłane jednocześnie w obu kierunkach (full duplex).
- Podobnie jak I<sup>2</sup>C, działa w trybie master-slave, umożliwiając podłączenie wielu urządzeń typu slave.
- Pozwala na przesyłanie danych między mikrokontrolerem a urządzeniami peryferyjnymi, np. czujniki, rejestry przesuwne, przetworniki.

# SPI

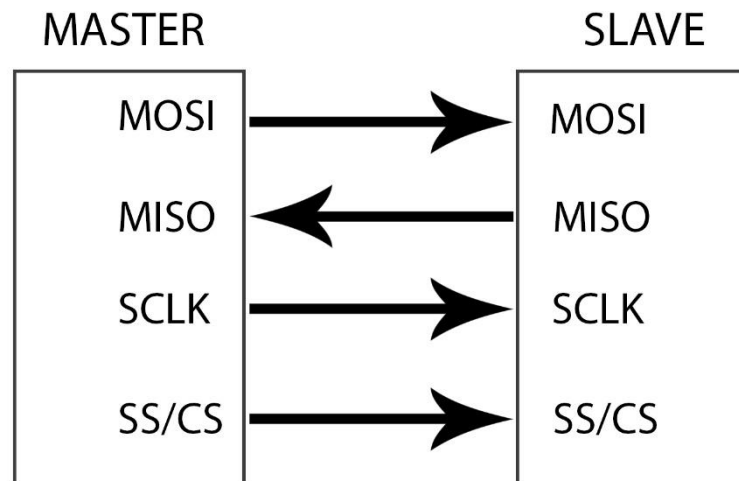


Komunikacja w SPI odbywa się za pomocą 4 linii:

- MOSI (Master-Out Slave-In) – dane przesyłane od urządzenia master do urządzenia slave
- MISO (Master-In Slave-Out), dane przesyłane od urządzenia slave do urządzenia master
- SCLK (Serial Clock) – linia sygnału zegarowego
- SS/CS (Slave Select / Chip Select) – linia służąca do wyboru urządzenia slave

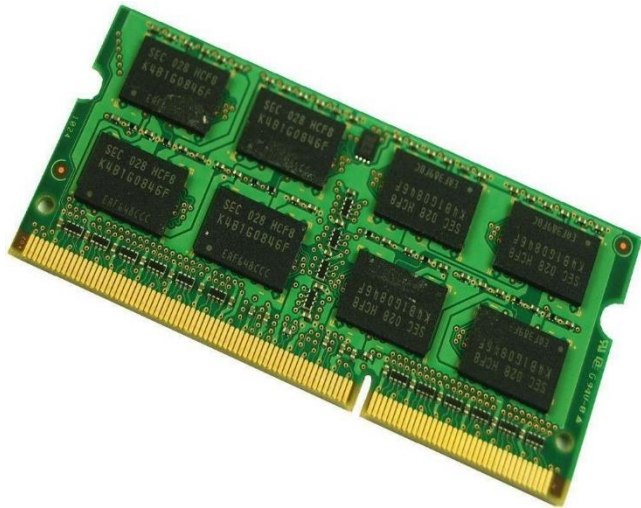
# SPI

Proces przesyłania danych między urządzeniem głównym „master” i podrzędnym „slave”:



1. Urządzenie master zarządza przesyłaniem danych, ustawiając sygnał zegarowy SCLK na częstotliwość mniejszą lub równą częstotliwości dozwolonej przez urządzenie slave.
2. Urządzenie master wybiera urządzenie slave będące odbiorcą, ustawiając stan niski na linii SS/CS.
3. Urządzenie master przesyła dane bit po bicie za pomocą linii MOSI.
4. Jeżeli wymagana jest odpowiedź, urządzenie slave przesyła dane bit po bicie za pomocą linii MISO.

# Pamięć RAM



- RAM (Random Access Memory) to pamięć o dostępie swobodnym
- Oznacza to, że kolejność dostępu do poszczególnych komórek pamięci jest dowolna
- Stanowi pamięć operacyjną komputera
- Informacja przechowywana jest w pamięci RAM w postaci bitów umieszczanych w komórkach
- Każda komórka pamięci ma unikalny adres i przechowuje dane w postaci grupy bitów (najczęściej jest ich 8 - czyli 1 bajt)

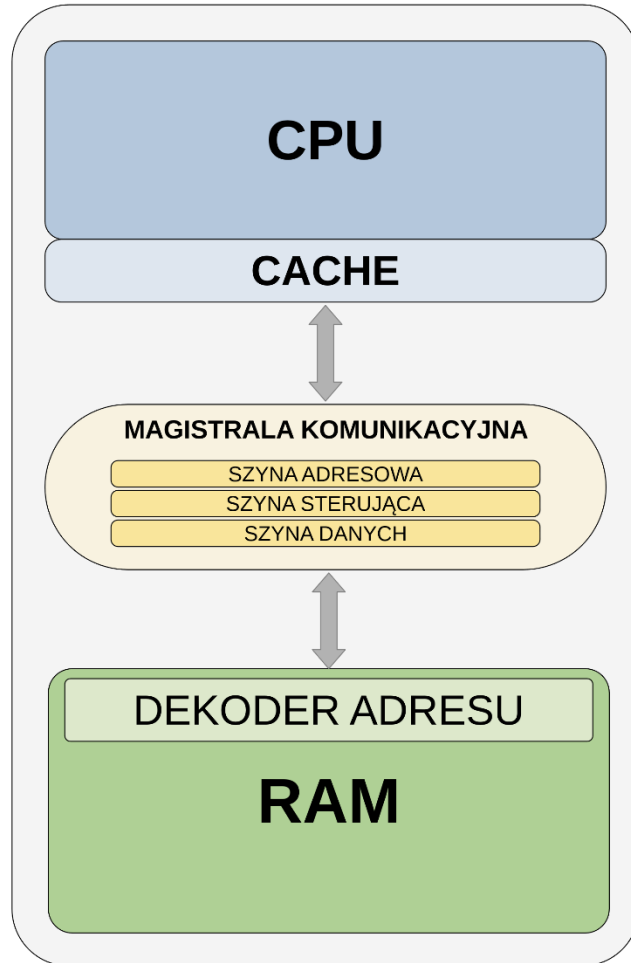
Pamięć	
Adres	Zawartość komórki
0	11000110
1	00001111
2	11000011
3	11111110
4	00000001
5	11100111
...	...

# Pamięć RAM

- Komputer steruje pamięcią przy pomocy „trzech” magistral
  - Magistrala adresowa
  - Magistrala danych
  - Magistrala sterująca
- Magistrale zbudowane są z linii, którymi transmituje się sygnały (poszczególne linie przesyłają tylko sygnały dwustanowe, czyli bity)



# Działanie RAM



## Zapis

- CPU przesyła adres komórki na szynę adresową
- CPU przesyła dane za pomocą szyny danych
- Żądanie zapisu danych zostaje wysłane do pamięci za pomocą szyny sterującej
- Dekoder adresu identyfikuje adres komórki
- Dane z szyny danych zostają zapisane w odpowiednich komórkach

## Odczyt

- CPU przesyła adres komórki na szynę adresową
- Żądanie odczytu danych zostaje wysłane do pamięci za pomocą szyny sterującej
- Dekoder adresu identyfikuje adres komórki
- Pamięć przesyła dane za pomocą szyny danych
- CPU odbiera dane z szyny danych

# Pamięć RAM: magistrala adresowa

- Magistrala adresowa przekazuje adres komórki w postaci binarnej, natomiast liczba linii na magistrali adresowej określa zakres dostępnych adresów,
- W efekcie maksymalny obsługiwany rozmiar pamięci komputera wynosi  $2^n$  gdzie parametr n określa liczbę linii na magistrali adresowej, a każda linia przyjmuje dwa stany 0 lub 1
- Przykładowo
  - dla magistrali adresowej składającej się z 16 linii maksymalny możliwy do zaadresowania rozmiar pamięci wynosił  $2^{16} = 65536$  komórek (64KB)
  - dla magistrali adresowej składającej się z 32 linii maksymalny możliwy do zaadresowania rozmiar pamięci wynosił  $2^{32} = 4294967296$  komórek (4GB)
- Oczywiście w systemie może być mniej pamięci, w takim przypadku część adresów nie jest wykorzystywana, gdyż nie stoją za nimi żadne komórki
- Ilość możliwych do zaadresowania komórek nosi nazwę przestrzeni adresowej (address space)
- Natomiast pamięć fizyczna (physical memory, physical storage) określa ilość pamięci rzeczywiście zainstalowanej w systemie komputerowym



# Pamięć RAM: magistrala danych

- Umożliwia komputerowi przekazywanie danych do pamięci oraz odczyt przechowywanych przez pamięć informacji z komórek
- Magistrala danych zbudowana jest z linii sygnałowych, po których przekazywane są bity
- Liczba linii na magistrali danych zależy od architektury komputera, przykładowo:
  - w systemach 64-bitowych magistrala danych może zawierać 64 linie, co pozwala w jednym cyklu dostępu do pamięci przesłać porcję 64 bitów
  - w systemach 32-bitowych magistrala danych może zawierać 32 linie, co pozwala w jednym cyklu dostępu do pamięci przesłać porcję 32 bitów
- Obecne systemy oferują wielokanałową konstrukcję dostępu do pamięci, zwielokrotniając tym szerokość magistrali,
  - W efekcie szerokość magistrali danych jest równa iloczynowi liczby kanałów i liczby linii
  - Przykładowo, szerokość magistrali danych dla systemów 64 bitowych mających 6 kanałów wynosi 384 bitów ( $64 \cdot 6$ )

# Pamięć RAM: magistrala danych

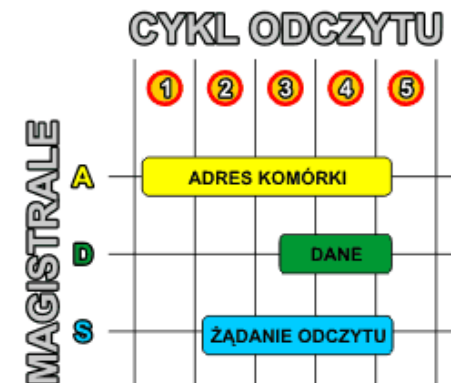
- Pamięć komputera jest podzielona na dedykowaną dla danej architektury liczbę banków
- Banki są podłączone do wspólnej magistrali adresowej
- Dla architektury 32 bitowej (32 linie na magistralę na kanał) pamięć komputera podzielona jest na 4 obszary nazywane bankami, a każdy bank współpracuje z 8 liniami magistrali danych i zarządza ćwiartką rozmiaru pamięci podłączonej do kanału
- Analogicznie, dla architektury 64 bitowej (64 linie na magistralę na kanał) pamięć komputera podzielona jest na 8 banków
- Dostęp do danych odbywa się poprzez banki, a w konsekwencji efektywne wykorzystanie magistrali wiąże się z dostępem do danych zapisanych w różnych bankach
- Pobieranie wielu danych z jednego banku zwiększa liczę niezbędnych transferów

Bank 3		Bank 2		Bank 1		Bank 0	
Adres	Zawartość	Adres	Zawartość	Adres	Zawartość	Adres	Zawartość
3	00000000	2	11111111	1	11110000	0	00001111
7	11001100	6	10101010	5	01010101	4	11000011
11	11100111	10	10000001	9	01111110	8	11010011
15	11010110	14	00101100	13	00111010	12	11010100
19	11010010	18	00010100	17	00100100	16	11011110
		...		...		...	

Magistrala danych 32 bitowe – dostęp do pamięci realizowany jest przez 4 banki

# Pamięć RAM: magistrala sterująca

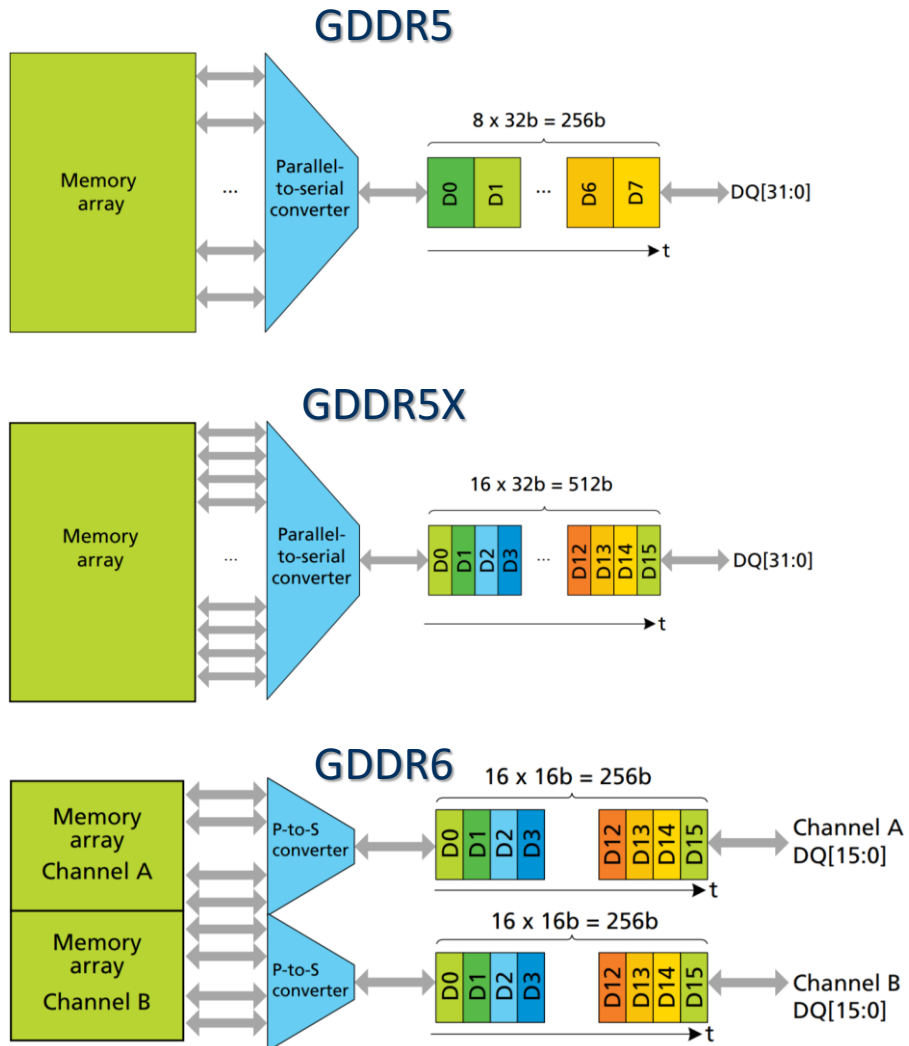
- Magistrala sterująca pozwala na kierowanie pracą pamięci
- Magistrala ta zawiera kilka linii, które określają rodzaj wykonywanej przez pamięć operacji (zapis lub odczyt) oraz uaktywniają odpowiednie banki pamięci w systemach 16-, 32- i 64-bitowych
- Uproszczony cykl odczytu:
  - Na magistrali adresowej A komputer umieszcza adres komórki pamięci, z której chce odczytać dane
  - Magistralą sterującą S przesłane zostaje do pamięci żądanie odczytu danych
  - W odpowiedzi pamięć wyszukuje pożądaną komórkę i umieszcza na magistrali danych D jej zawartość
  - Komputer odczytuje z magistrali danych zawartość zaadresowanej komórki
  - Cykl zostaje zakończony, sygnały wracają do stanu neutralnego



# GDDR

- **Graphic Double Data Rate (GDDR)** to typ pamięci RAM przeznaczony dla kart graficznych.
- W odróżnieniu od pamięci DDR stosowanej w CPU, której celem jest jak najszybsze dostarczenie małych porcji danych, pamięć GDDR została zoptymalizowana pod kątem przesyłania jak największej ilości danych w tym samym cyklu zegara (szybkość vs przepustowość).
- Opóźnienia pamięci GDDR są większe niż w przypadku DDR. Priorytetem jest jednak dostarczenie dużej ilości danych potrzebnych do renderowania obrazu.
- Pamięć GDDR może w jednym cyklu zegara zażądać oraz otrzymać dane, co nie jest możliwe w przypadku DDR.
- W przeciwieństwie do GDDR5 i pomimo swojej nazwy, GDDR6 wykorzystuje tryb QDR (quad data rate – 4 sygnały w jednym cyklu zegarowym).

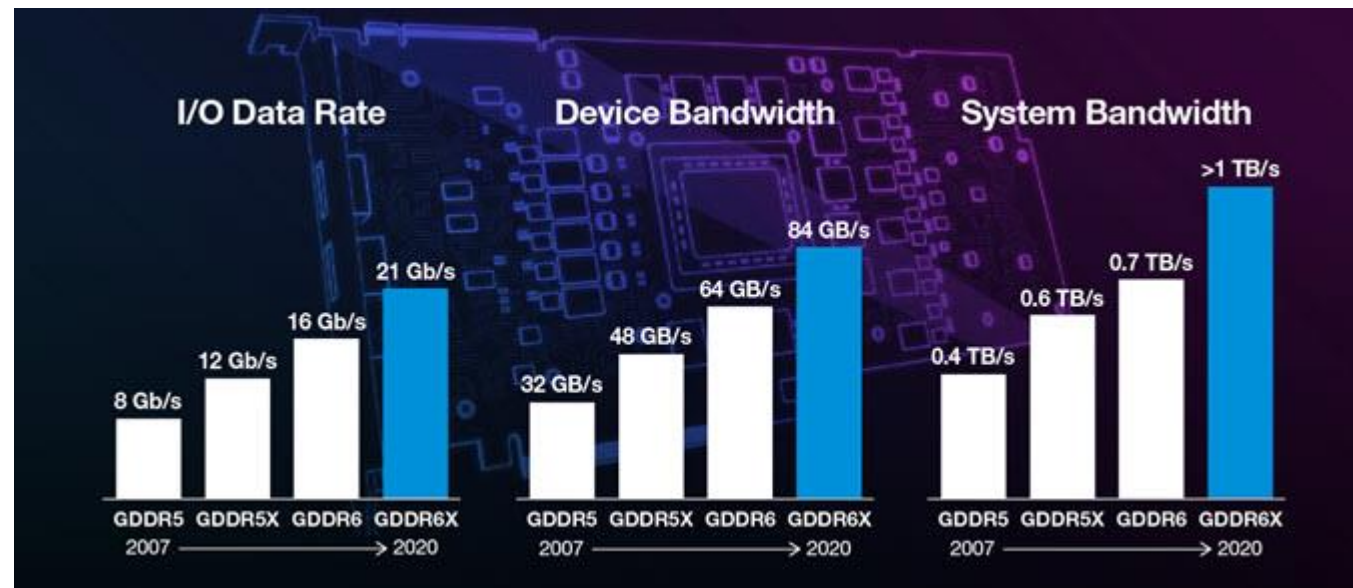
# GDDR



- GDDR5 korzysta z 8 magistral o szerokości 32 bitów, co pozwala na odczyt 256 bitów. Konwerter równoległo-szeregowy przekształca każdy 256 bitowy pakiet danych na osiem 32-bitowych słów danych, przesyłanych przez 32-bitową magistralę szeregową.
- W GDDR5X jednorazowo odczytywane jest 512 bitów, które przekształcane są na szesnaście 32-bitowych słów danych, przesyłanych przez 32-bitową magistralę szeregową.
- W GDDR6 zastosowano dwa całkowicie niezależne od siebie kanały. W ramach każdego kanału następuje odczyt 256 bitów, które konwertowane są na szesnaście 16-bitowych słów danych, przesyłanych 16-bitową magistralą szeregową.

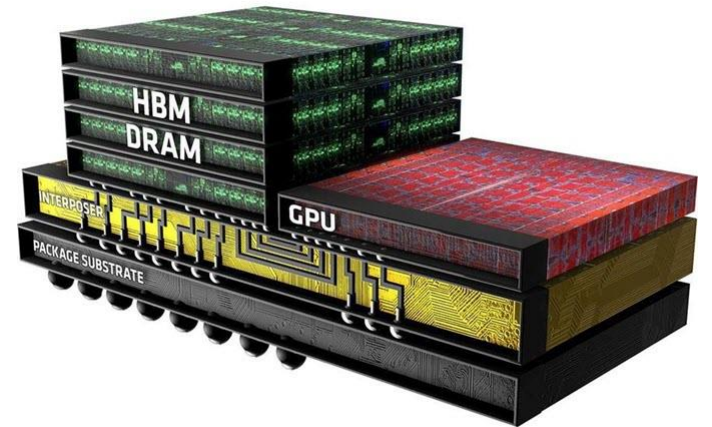
# GDDR

- Kolejne wersje GDDR charakteryzują się przede wszystkim zwiększoną wydajnością oraz coraz mniejszym zapotrzebowaniem na energię.
- Porównanie kolejnych wersji GDDR pod kątem szybkości transmisji danych, przepustowości urządzenia oraz systemu:

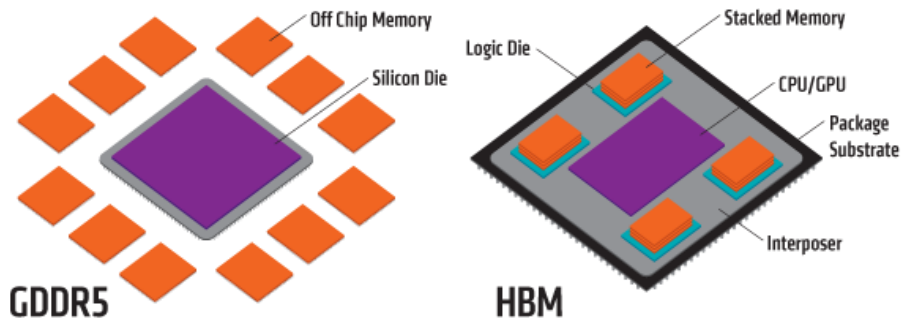


# HBM

- **High Bandwith Memory** (HBM) to interfejs pamięci RAM, przeznaczony głównie dla wysokowydajnościowych akceleratorów grafiki oraz urządzeń sieciowych.
- HBM ma trójwymiarową strukturę projektową. Wiele układów ułożonych jest jeden na drugim, co pozwala na zmniejszenie potrzebnego miejsca na płycie drukowanej.
- Magistrala w HBM jest znacznie szersza niż w przypadku GDDR. Matryce DRAM są zestawiane w stos. Typowa konfiguracja składa się z 4 modułów DRAM na matrycy bazowej, mających po dwa 128-bitowe kanały. W sumie daje to magistralę o szerokości 1024 ( $4 \cdot 2 \cdot 128$ ) bitów.
- W kartach graficznych można zastosować kilka stosów – zastosowanie czterech zwiększa szerokość szyny pamięci do 4096 bitów.



# HBM



- W porównaniu do GDDR, HBM osiąga większą wydajność przy niższym zużyciu energii.
- Wadą HBM jest stosunkowo wysoki koszt, w porównaniu do GDDR.
- HBM2 oferuje do 8 modułów DRAM w ramach jednego stosu. Pozwala to na podwojenie przepustowości z 128 GB/s na 256 GB/s przy jednoczesnym zachowaniu 1024 bitów szerokości magistrali.
- W niedalekiej przyszłości pojawi się HBM3, zapowiedziany w 2016 roku. Ma on umożliwić większą przepustowość pamięci – 512 GB/s na stos, przy niższym napięciu i zużyciu energii.



# Dziękuję za uwagę



Kontakt:

**dr hab. inż. Łukasz Szustak, prof. PCz**

[lszustak@icis.pcz.pl](mailto:lszustak@icis.pcz.pl)

Katedra Informatyki

Wydział Inżynierii Mechanicznej i Informatyki

Politechnika Częstochowska