

Architektura komputerów

Wykład 8

Karty rozszerzeń

Wojciech Kordecki

Collegium Witelona
Wydział Nauk Technicznych i Ekonomicznych
Zakład Informatyki

Semestr letni 2023/24

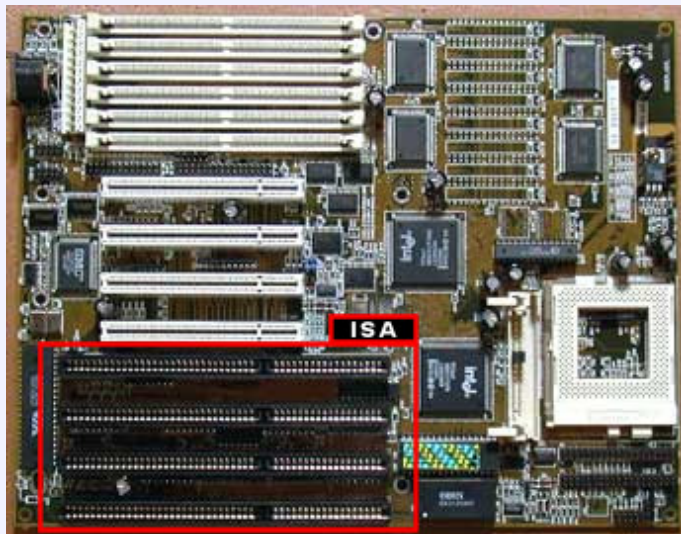


Źródła

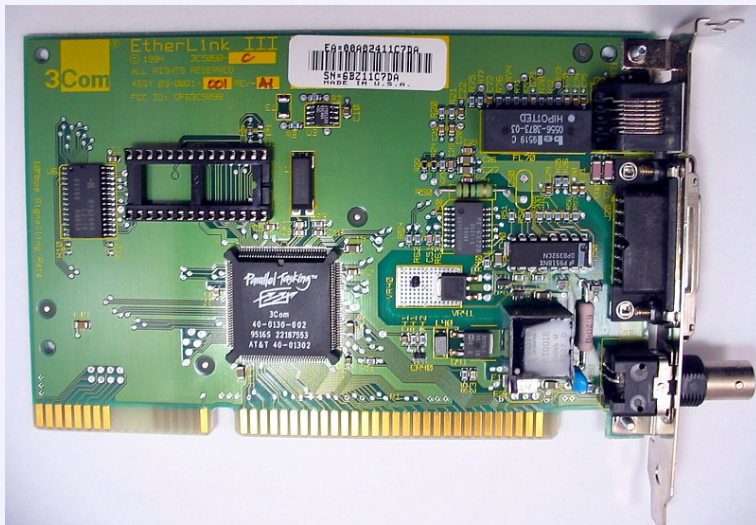
Ze względu na niezwykle szybki postęp, większość informacji jest zaczerpnięta z internetu, głównie z Wikipedii (konfrontowanych z książką Metzgera *Anatomia PC*), materiałów firmowych Intel'a i AMD, a także z wyimków artykułów umieszczanych w czasopismach komputerowych.



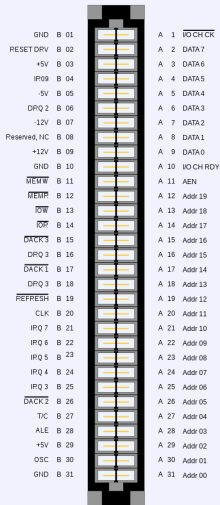
ISA – złącze na płycie głównej



ISA – karta



ISA – opis złącz



Parametry ISA

- szyna danych - 16-bitowa,
- częstotliwość pracy - 8,33 MHz,
- przepustowość 8 MB/s (efektywna w granicach od 1,6 MB/s do 1,8 MB/s)
- szyna adresowa - 24-bitowa,
- brak obsługi Plug and Play

Pod koniec lat dziewięćdziesiątych XX wieku znaczenie tej magistrali zaczęło maleć, a jej funkcje przejmował standard PCI.



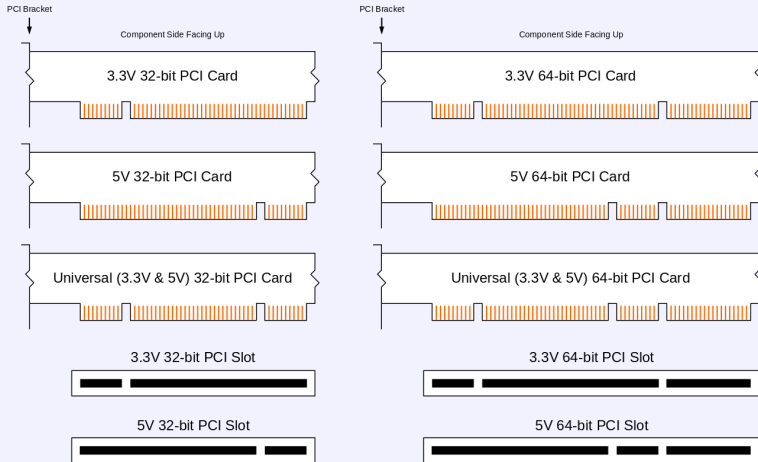
PCI – krótki opis

PCI (ang. Peripheral Component Interconnect) – magistrala komunikacyjna służąca do przyłączania kart rozszerzeń do płyty głównej w komputerach klasy PC.

Zaprezentowana w roku 1992 jako rozwiązanie umożliwiające szybszą komunikację pomiędzy procesorem i kartami niż stosowane dawniej ISA. Dodatkową zaletą PCI jest to, że nie ma znaczenia czy w gnieździe jest karta sterownika dysków (np. SCSI), sieciowa czy graficzna. Każda karta, pasująca do gniazda PCI, funkcjonuje bez jakichkolwiek problemów, gdyż nie tylko sygnały, ale i przeznaczenie poszczególnych styków gniazda są znormalizowane.



Rozmieszczenie kluczy w gniazdach



Przepustowość PCI

Szyna PCI przyspiesza współpracę z dowolnym urządzeniem zewnętrznym. Przy częstotliwości taktowania 33 MHz i szerokości 32 bitów magistrala PCI ma szybkość transmisji 133 MB/s. Szerokość szyny adresowej i danych nowych procesorów 64 bitowych zmiany nie wpływają na architekturę PCI a jedynie podwaja się przepustowość do 266 MB/s.

Karty dołączone do szyny PCI mogą się komunikować nawet bez udziału mikroprocesora, co zwiększa efektywność jego użytkowania. Dla każdej karty zdefiniowane są tzw. rejestry konfiguracyjne. Przy ładowaniu systemu procesor odczytuje zapisane w nich dane i rozpoznaje, jaka karta jest umieszczona w gnieździe. Instalacja i inicjacja karty następuje potem w pełni automatycznie.



Wersje PCI

Wersja	PCI 2.0	PCI 2.1	PCI 2.2	PCI 2.3
Rok wprowadzenia	1993	1994	1997	2002
Maksymalna szerokość szyny danych	32 bity	64 bity	32 bity	64 bity
Maksymalna częstotliwość taktowania	33 MHz	33 MHz	66 MHz	66 MHz
Maksymalna przepustowość	133 MB/s	266 MB/s	266 MB/s	533 MB/s
Napięcie	5 V	5 V	5 V / 3,3 V	3,3 V

Specyfikacja PCI rev.3.0

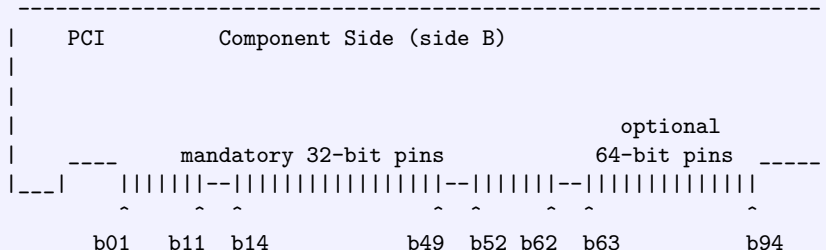
<http://people.na.infn.it/~garufi/didattica/CorsoAcq/PCI.Local.Bus.Specification.Revision.3.0.pdf>

Specyfikacja



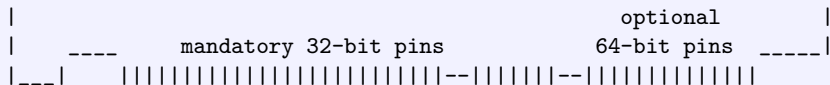
Schemat styków (1)

PCI Universal Card 32/64 bit



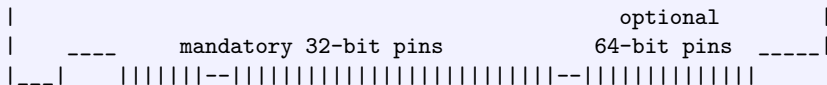
Schemat styków (2)

PCI 5V Card 32/64 bit



Schemat styków (3)

PCI 3.3V Card 32/64 bit



PCI Express

PCI Express, PCIe, (ang. Peripheral Component Interconnect Express), – połączenie Point-to-Point (podobnie jak HyperTransport), służące do instalacji kart rozszerzeń na płycie głównej.

Zastąpiła ona magistrale PCI oraz AGP. Istnieje możliwość wyprowadzenia interfejsu PCIe na zewnątrz, co zostało zastosowane m.in. w komputerach IBM 2827 EC12.

Standard został opracowany przez firmy Compaq, Dell, Hewlett-Packard, IBM, Intel i Microsoft.



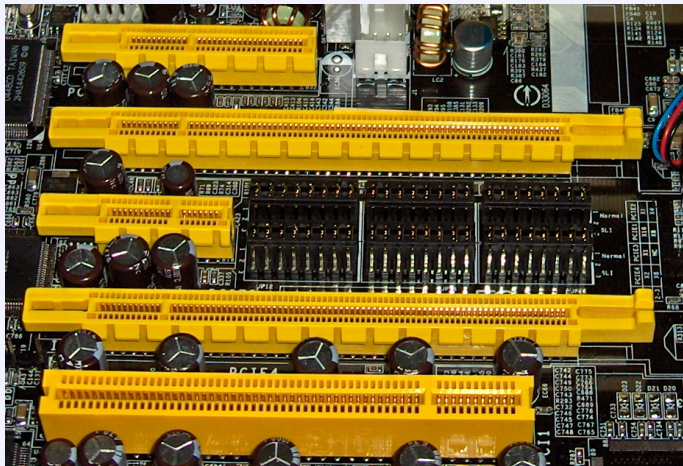
PCI Express – warianty

Warianty z 1, 2, 4, 8, 12, 16 lub 32 liniami. Wraz ze wzrostem liczby linii wydłużeniu ulega złącze, jego konstrukcja (poprzez wspólną część początkową i jedynie dodawanie na końcu nowych linii) umożliwia działanie kart wolniejszych niż te, które maksymalnie obsługuje gniazdo. Sytuacja odwrotna (karta przygotowana na większą liczbę linii w gnieździe o mniejszej ich liczbie) może nie być możliwa (ze względu na ograniczenie przepustowości) lub wymagać fizycznej modyfikacji karty lub złącza.

Gniazdo x1 ma 18 pinów z każdej strony, gniazdo x4 – 32, gniazdo x8 – 49, zaś gniazdo x16 – 82 piny z każdej strony.



Gniazda PCI



Gniazda PCI-E od góry: x4, x16, x1 i x16 w porównaniu ze złączem PCI (na dole).



Pobierana moc

Maksymalna moc, jaką karta PCIe (zazwyczaj karta graficzna) może pobierać przez złącze PCI Express, to 75 W. Jeśli karta wymaga większej mocy, musi ona zostać zasilona z zasilacza dodatkowym kablem, standardowo sześć-, ośmiożyłowym lub obydwoma (6+8). Dzięki temu karta może pobrać maksymalną moc 300 W – 75 W przez złącze PCIe, kolejne 75 W przez złącze sześciostykowe oraz 150 W przez złącze ośmiostykowe.



Karta graficzna PCIe



Karta graficzna przystosowana do pracy w porcie x16.



Przepustowości

Wariant PCIe	Przepustowość (w każdym kierunku)
x1 v1.0	250 MB/s
x2 v1.0	500 MB/s
x4 v1.0	1000 MB/s
x8 v1.0	2000 MB/s
x16 v1.0	4000 MB/s (4 GB/s)
x16 v2.0	8000 MB/s (8 GB/s)
x16 v3.0	16000 MB/s (16 GB/s)



Karta?

Karta dźwiękowa (ang. sound card, audio card) – komputerowa karta rozszerzeń, umożliwiająca rejestrację, przetwarzanie i odtwarzanie dźwięku. Poprawnym jest też równie często stosowany termin karta muzyczna.

Najbardziej znaną grupą kart dźwiękowych jest seria Sound Blaster firmy Creative Labs.

Obecnie układy dźwiękowe wystarczające do zastosowań amatorskich są zazwyczaj wbudowywane w płytę główną komputera, a nie stanowią karty rozszerzenia. Z powodów historycznych są jednak określane mianem zintegrowana karta dźwiękowa. Pojawiły się również zewnętrzne karty dźwiękowe podłączane do komputera przez port USB.



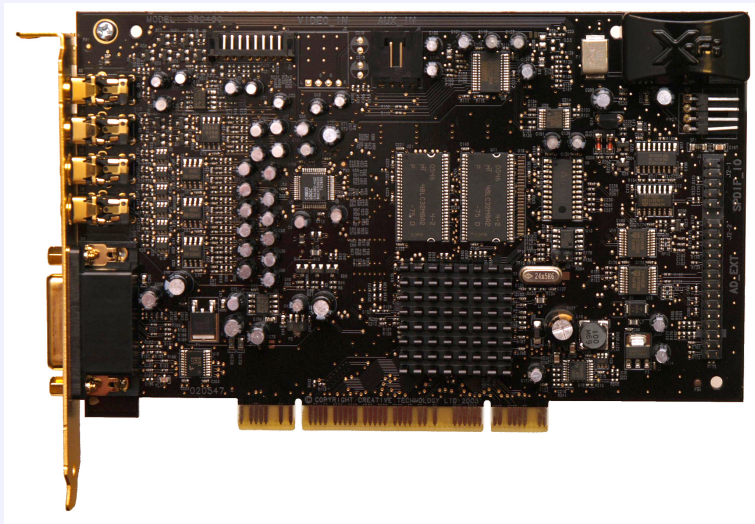
DAC i ADC

Przetworniki

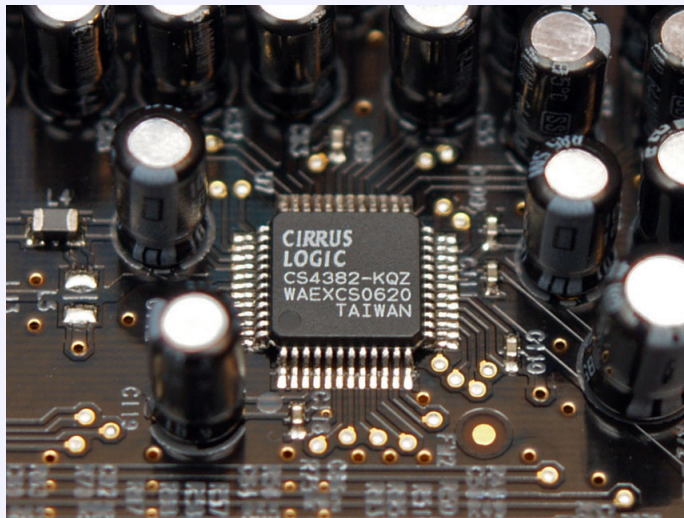
- DAC – Digital Analog Converter, przetwornik cyfrowo-analogowy C/A,
- ADC – Analog Digital Converter, przetwornik analogowo-cyfrowy A/C,



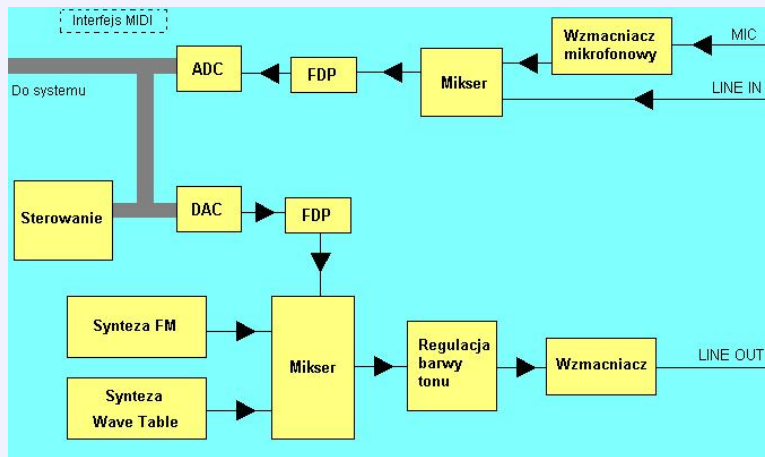
Sound Blaster X-Fi XtremeGamer Fatal1ty Pro



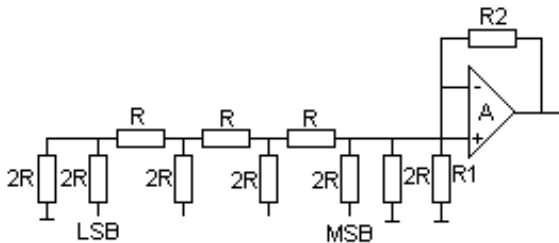
8-kanałowy DAC Cirrus Logic CS4382



Schemat karty dźwiękowej



DAC



$$U_{wy} = U_{odn} \left(\alpha_0/2^0 + \alpha_1/2^1 + \alpha_2/2^2 + \dots + \alpha_n/2^n \right).$$



ADC

Zasada działania.

Przetwornik porównuje U_{we} z kolejnymi wartościami referencyjnymi U_{ref} podawanymi na DAC.

Na początku $U_{ref} = U_{max}$.

Jeśli $U_{we} < U_{ref}$ to zmniejszamy U_{ref} , a jeśli $U_{we} > U_{ref}$, to zwiększamy. Zawsze wystarczy co najwyżej n kroków dla przetwornika n bitowego.



Karty pomiarowe

Wielofunkcyjna karta pomiarowa – Data Acquisition DAQ.
Wielofunkcyjna karta DAQ umożliwia bezpośrednie doprowadzenie do jej zacisków wejściowych sygnałów napięciowych oraz generowanie sygnałów napięciowych.

W prostych zastosowaniach można do pomiarów i analizy sygnałów użyć dobrej karty dźwiękowej, na przykład 24-bitowej.



Generowanie dźwięku – interferencja

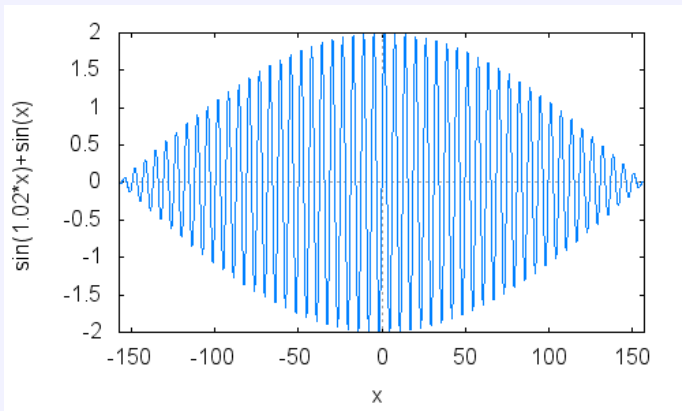
$$f(t) = \sin(t) + \sin(\alpha t).$$

Na kolejnych slajdach:

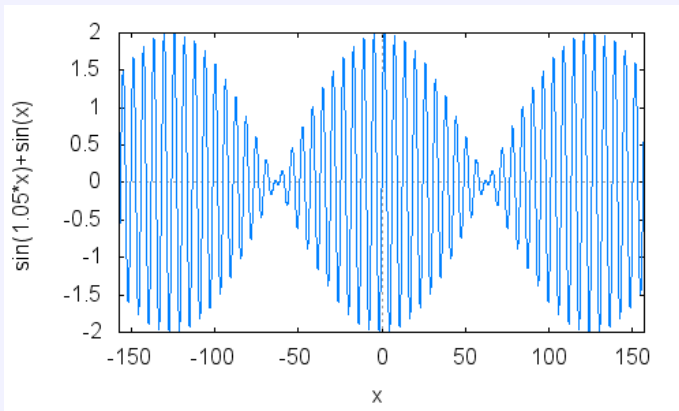
- $\alpha = 1.02$,
- $\alpha = 1.05$,
- $\alpha = 1.1$,
- $\alpha = 1.2$.



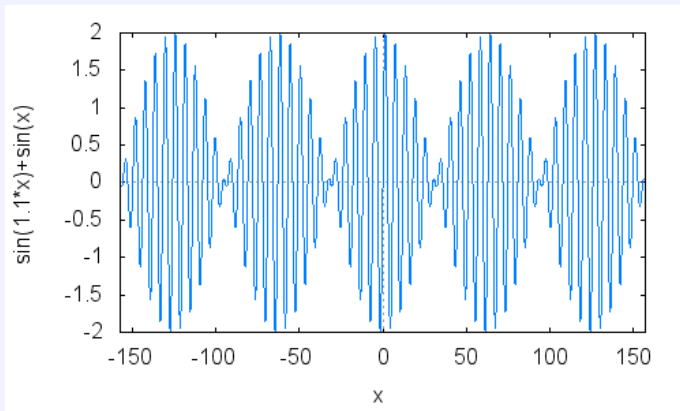
$$\alpha = 1.02$$



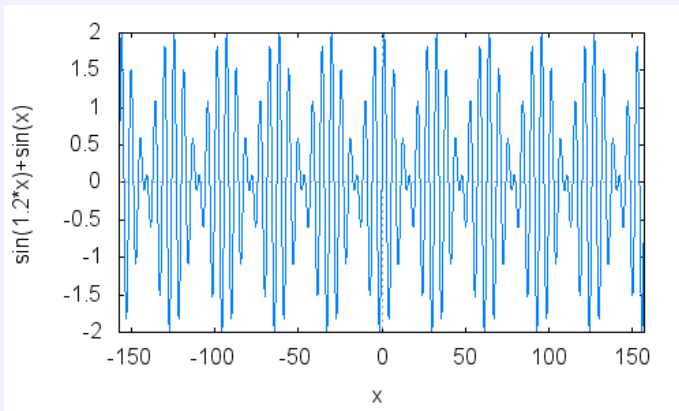
$\alpha = 1.05$



$$\alpha = 1.1$$



$$\alpha = 1.2$$



Synteza FM

Metoda modulacji częstotliwości (ang. Frequency modulation)
Wykorzystuje metodę stosowaną w telekomunikacji (transmisja sygnałów).

1973 – John Chowning publikuje pracę: „The Synthesis of Complex Audio Spectra by Means of Frequency Modulation”
podstawy teoretyczne zastosowania metody FM w syntezie dźwięku. Patent w latach 1975-1995.

Wyniki tych prac zostały wykorzystane przez firmę Yamaha do stworzenia cyfrowych syntezatorów.

$$FM(t) = A_C \sin(2\pi f_C t + A_M \sin(2\pi f_M t)).$$



Modulacja FM i sumowanie – wzory

- f_C – częstotliwość fali nośnej,
- f_M – częstotliwość przebiegu modulującego,
- A_C – amplituda fali nośnej,
- A_M – amplituda przebiegu modulującego.

Modulacja FM:

$$FM(t) = A_C \sin(2\pi f_C t + A_M \sin(2\pi f_M t)).$$

Sumowanie

$$FM(t) = A_C \sin(2\pi f_C t) + A_M \sin(2\pi f_M t).$$



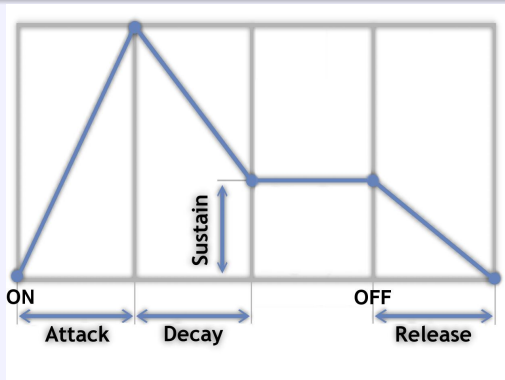
Efekty

Tremolo: modulacja amplitudy przebiegiem o niskiej częstotliwości, np. 3.7 Hz.

Wibracja: modulacja częstotliwości falą o niskiej częstotliwości, np. 6.4 Hz.



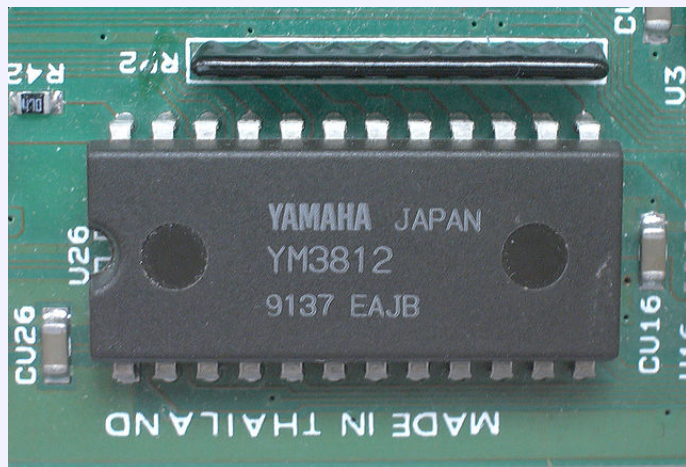
Fazy obwiedni



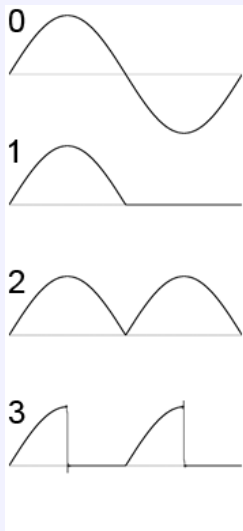
- attack (narastanie),
- decay (opadanie),
- sustain (podtrzymanie),
- release (wybrzmiewanie).



Yamaha YM3812 – OPL2

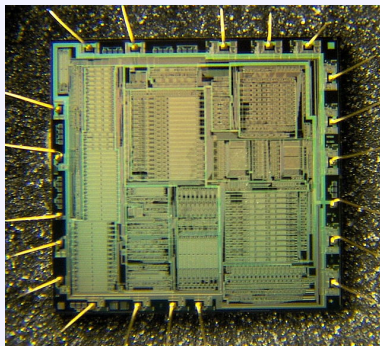


Yamaha YM3812 – kształty fal



OPLx decapsulated

https://docs.google.com/document/d/18IGx18NQY_Q1PJVZ-bHywao9bhsDoAqoIn1rIm42nwo/edit?pli=1



Strona OPL

<http://www.oplx.com/>

<http://chipmusic.org/forums/topic/12535/ym3812-opl2-vst-instrument/>

<http://www.malinov.com/Home/sergeys-projects/isa-opl2-card>



Synteza Wave Table

Synteza tablicowa, alternatywny w stosunku do syntezy FM sposób generowania dźwięków.

Krótkie próbki (ang. sample) danego instrumentu muzycznego są przetwarzane na dźwięk o odpowiedniej wysokości i czasie trwania. Technika ta opracowana została pod koniec lat 70. przez Wolfganga Palma.



Powstawanie dźwięku

Metoda tablicowa – dźwięk powstaje w przetworniku cyfrowo-analogowym do którego dostarczane są próbki odczytywane z tablicy próbek dla pojedynczego okresu sygnału i powtarzane w pętli.

Zmianę wysokości sygnału można uzyskać poprzez zmianę szybkości konwersji w DAC lub poprzez pomijanie niektórych próbek z tablicy.

Uwaga. Dwa dźwięki są odległe od siebie o jeden półton, gdy stosunek ich częstotliwości wynosi $\sqrt[12]{2} \approx 1.059463094359295$. Jest to skala *równomiernie temperowana*.

Jan Sebastian Bach *Das Wohltemperierte Klavier*

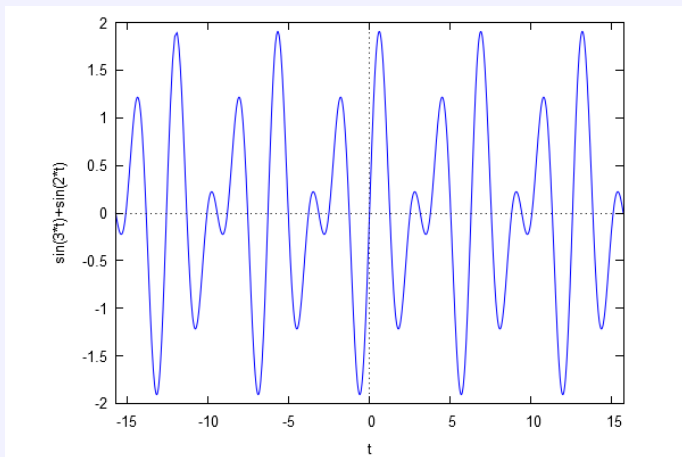
[http:](http://pl.wikipedia.org/wiki/Das_Wohltemperierte_Klavier)

[//pl.wikipedia.org/wiki/Das_Wohltemperierte_Klavier](http://pl.wikipedia.org/wiki/Das_Wohltemperierte_Klavier)



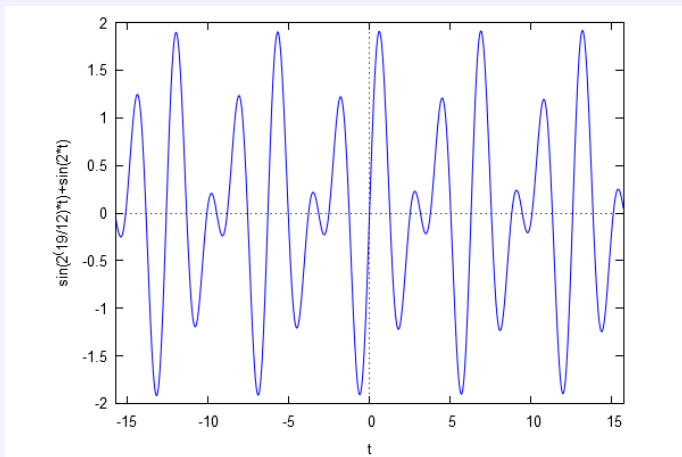
Kwinta naturalna

$$\sin(t) + \sin(3t/2), \quad -5\pi \leq t \leq 5\pi.$$



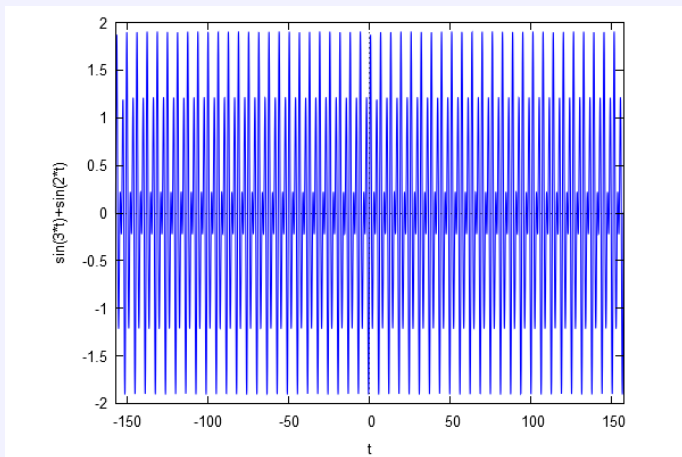
Kwinta temperowana

$$\sin(t) + \sin\left(2^{7/12}t\right), \quad -5\pi \leq t \leq 5\pi.$$



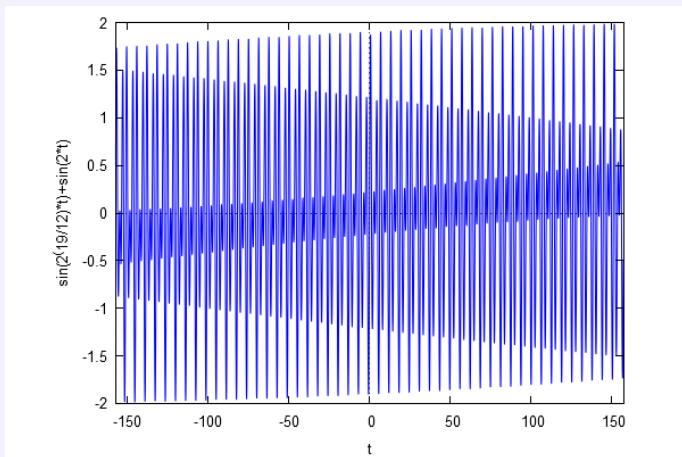
Kwinta naturalna

$$\sin(t) + \sin(3t/2), \quad -50\pi \leq t \leq 50\pi.$$



Kwinta temperowana

$$\sin(t) + \sin\left(2^{7/12}t\right), \quad -50\pi \leq t \leq 50\pi.$$



Tryby rozdzielczości

Skrót	Nazwa	Rozdzielczość
VGA	VGA	640 × 480
SVGA	Super VGA	800 × 600
XGA	eXtended Graphics Array	1024 × 768
SXGA	Super XGA	1280 × 1024
UXGA	Ultra XGA	1600 × 1200
HDTV	High-Definition TV	1920 × 1080
QXGA	Quard XGA	2048 × 1536



VGA, SVGA – obsługa pamięci

- Pamięć obrazu dzielona jest na płyty (bloki) po 64 KB każdy.
- W każdym bloku jednemu bitowi odpowiada jeden punkt.
 - VGA – 4 bloki
 - SVGA – minimum 8 bloków



RAMDAC

Random Access Memory Digital to Analog Converter.

Konwerter zawiera 4 funkcjonalne bloki:

- pamięć SRAM do przechowywania mapy kolorów,
- trzy przetworniki cyfrowo-analogowe (C/A), po jednym dla każdego koloru podstawowego modelu RGB
 - R – czerwonego,
 - G – zielonego,
 - B – niebieskiego.



Kolory EGA

- 0 – czarny (black)
- 1 – niebieski (blue)
- 2 – zielony (green)
- 3 – siny (cyan)
- 4 – czerwony (red)
- 5 – fioletowy (magenta)
- 6 – brązowy (brown)
- 7 – jasnoszary (light gray)
- 8 – szary (gray)
- 9 – jasnoniebieski (light blue)
- 10 – jasnozielony (light green)
- 11 – jasnosiny (light cyan)
- 12 – jasnoczerwony (light red)
- 13 – różowy (pink)
- 14 – żółty (yellow)
- 15 – biały (white)



Akceleratory grafiki

Umieszczane są w pobliżu pamięci obrazu na karcie graficznej. Akceleratory wykonują część procesu dekodowania obrazu video, zwalniają więc CPU z dużej liczby operacji.

Odzyskana w ten sposób moc obliczeniowa jest do dyspozycji systemu operacyjnego.

Odciążone magistrale systemowe umożliwiają szybszy dostęp do zasobów: pamięci RAM, dysków i peryferii.



Akceleratory grafiki 3D

Funkcje:

- obliczanie współrzędnych obiektów podlegających przemieszczeniom,
- przeskalowywanie obiektów, których rozmiary zmieniają się w trakcie ruchu,
- obliczenia barw obiektów oświetlanych,
- usuwanie (clipping) fragmentów brył niewidocznych dla obserwatora,
- projekcja obrazu wypadkowego na płaszczyznę 2D.



Wyjścia kart graficznej

Od lewej: HDMI, D-SUB, DVI.



Funkcje obliczeniowe kart graficznych

Karty graficzne, oprócz swojej roli w renderowaniu grafiki i obsłudze wyświetlacza, są wykorzystywane do przyspieszania obliczeń w zastosowaniach naukowych, w tym w symulacjach naukowych, analizach danych i badaniach. Dodatkowo, za sprawą swojej mocy obliczeniowej, są często używane do treningu modeli sztucznej inteligencji, ponieważ potrafią szybko przetwarzać duże ilości danych potrzebnych do uczenia maszynowego i głębokich sieci neuronowych.



Funkcje obliczeniowe kart graficznych

Karty graficzne, oprócz swojej roli w renderowaniu grafiki i obsłudze wyświetlacza, są wykorzystywane do przyspieszania obliczeń w zastosowaniach naukowych, w tym w symulacjach naukowych, analizach danych i badaniach. Dodatkowo, za sprawą swojej mocy obliczeniowej, są często używane do treningu modeli sztucznej inteligencji, ponieważ potrafią szybko przetwarzać duże ilości danych potrzebnych do uczenia maszynowego i głębokich sieci neuronowych.

Karty graficzne o dużej mocy obliczeniowej są wykorzystywane w procesie wydobywania, zwłaszcza w przypadku niektórych kryptowalut takich jak Ethereum, które korzystają z algorytmu [Proof of Work \(PoW\)](#).

Źródło: Wikipedia



Gigabyte GeForce RTX 4070

GAMING OC 12GB GDDR6X 192bit DLSS 3

Układ graficzny: GeForce RTX™ 4070

Taktowanie rdzenia: 2475 – 2565 MHz

Rdzenie CUDA: 5888

Pamięć: 12 GB

Rodzaj pamięci: GDDR6X

Taktowanie pamięci: 21000 MHz

Szyna pamięci: 192 bit

Złącze karty: PCI-Express 4.0

Złącza: DisplayPort x 3, HDMI 2.1 x 1

Maks. obsługiwana rozdzielczość: 7680 x 4320

Obsługa wielu monitorów: 4

Typ chłodzenia: aktywne

Liczba wentylatorów: 3

Rekomendowana moc zasilacza: 700 W

Wtyczka zasilania: 1 x 16 pin



CUDA

Compute Unified Device Architecture – opracowana przez firmę Nvidia (2007) uniwersalna architektura procesorów wielordzeniowych (głównie kart graficznych) umożliwiająca wykorzystanie ich mocy obliczeniowej do rozwiązywania ogólnych problemów numerycznych w sposób wydajniejszy niż w tradycyjnych, sekwencyjnych procesorach ogólnego zastosowania.

Integralną częścią architektury CUDA jest oparte na języku programowania C środowisko programistyczne wysokiego poziomu, w którego skład wchodzi m.in. specjalny kompilator (nvcc), debugger (cuda-gdb, który jest rozszerzoną wersją debugera gdb umożliwiającą śledzenie zarówno kodu wykonywanego na CPU, jak i na karcie graficznej), profiler oraz interfejs programowania aplikacji.

Źródło: Wikipedia



CUDA – linki

Strona producenta:

<https://developer.nvidia.com/cuda-zone>

Programowanie:

[https://bulldogjob.pl/readme/
pisanie-blyskawicznie-szybkiego-kodu-z-cuda](https://bulldogjob.pl/readme/pisanie-blyskawicznie-szybkiego-kodu-z-cuda)

[http://mariuszrurarz.cba.pl/wp-content/uploads/2016/
09/programowanie_procesorow_graficznych_
GPU-denkowski.pdf](http://mariuszrurarz.cba.pl/wp-content/uploads/2016/09/programowanie_procesorow_graficznych_GPU-denkowski.pdf)

