Technologie cyfrowe

Artur Kalinowski

Zakład Cząstek i Oddziaływań Fundamentalnych

Pasteura 5, pokój 4.15

Artur.Kalinowski@fuw.edu.pl

Semestr letni 2014/2015

Organizacja zajęć

- Wykład: czwartek 14:15 – 16:00, Pasteura 5, sala 2.03

- Ćwiczenia: Pasteura 5, środa 12:30 – 15:15, sala 1.28, grupa 2
  piątek 13:30 – 16:15, sala 1.30, grupa 1

Zaliczenie:

- Ćwiczenia: **zaliczenie praktyczne (dwa sprawdziany), dopuszcza do egzaminu**
- Wykład: **egzamin pisemny (test)**
  Termin: 18.06.2015, 9:00 – 12:00, sala: 1.02

Terminy i sale wszystkich sprawdzianów można znaleźć tutaj:

https://www.fuw.edu.pl/~ppw/ipz/?m=exams#I%20rok%20optyka%20okular.

Materiały

- Slajdy z wykładu:
  http://brain.fuw.edu.pl/edu/Slajdy_z_wykładów_2015

- Skrypt prof. Piotra Durki:
  http://brain.fuw.edu.pl/edu/TI:Technologia_Informacyjna/skrypt

- Książki wprowadzające do informatyki:
  Piotr Gawrysiak, Cyfrowa rewolucja, PWN 2008
  Zbigniew Polański, Wpędzeni do komputerowego raju, 2010

- Książki podawane w trakcje kolejnych wykładów.

- Wikipedia i inne źródła w internecie (używane z ostrożnością)

Lista obecności

O czym będzie ten wykład?

Cyfrowy świat

Informacja (dane)

- kodowanie
- zapis, przychowywanie
- przesyłanie
- bezpieczeństwo danych

przetwarzanie danych

Obliczenia

- algorytmy
- struktury danych
- programy komputerowe

Urządzenia ("sprzęt")

Informacja: od tego co jest do bitu

Rzeczywistość

Przekaz słowny: mowa

rzędu 10000 słów
(trudna w kopiowaniu, wymaga złożonego nośnika, bardzo podatna na przekłamania)

Przekaz pisemny: alfabet

Kodowanie dzwięków mowy przez skończony, niewielki, zbiór znaków:
Egipt, około 2700 p.n.e – 24 hieroglify
Czasy współczesnie – alfabet łacińki, 26 znaków
(kopiowanie: od żmudnego przepisywania, przez maszynę drukarską po xero, podatny na przekłamania przy ręcznym kopiowaniu, nośnik mało złożony technicznie)

Przekaz cyfrowy: liczby

litery alfabetu ponumerowane, dowolna (przeliczalna) liczba znaków możliwa do zapisu. Liczby zapisywane w formacie binarnym.

Przekaz cyfrowy: liczby

litery alfabetu ponumerowane, dowolna (przeliczalna) liczba znaków możliwa do zapisu. Liczby zapisywane w formacie binarnym.
(wymaga zaawansowanej technologii do zapisu i odczytu, prawie zerowy koszt i czas zapisu i odczytu. Daje możliwość kontroli poprawności zapisu i odczytu)

Zapis cyfrowy:

każda informacja jest kodowana przy użyciu liczb. Odczyt i zapis informacji wymagają w szczególności znajomości "kodowania" - czyli przepisu zamiany złożonej informacji (np. obraz, tekst) na liczby.
Proces zamiany postaci analogowej ("zwykłej") na cyfrowną nzywa się **digitalizacją** (*z ang. digit – cyfra*) lub ucyfrowieniem.

Liczby

Systemy liczbowe:

zbiory reguł jednolitego zapisu i nazewnictwa liczb

System dziesiętny (decymalny): oparty na potęgach liczby 10

cyfry: **0 1 2 3 4 5 6 7 8 9**

System dwójkowy (binarny): oparty na potęgach liczby 2

cyfry: **0 1**

System szesnastkowy (hexadecymalny): oparty na potęgach liczby 16

“cyfry”: **0 1 2 3 4 5 6 7 8 9 A B C D E F**

Liczby

$$Liczba:(a_n a_{n-1} \dots a_0)_b = \sum_{k=0}^{k=n} a_k \cdot b^k$$

cyfry

podstawa: 2, 10, 16, lub inna

Ta sama liczba zapisana w różnych systemach:

$$(a_n a_{n-1} \dots a_0)_{b1} = (c_n c_{n-1} \dots c_0)_{b2}$$

Liczby

Liczba miesięcy w roku kalendarzowym:

Dziesiętnie:

$$(12)_{10}=\sum_{k=0}^{k=1} a_k \cdot 10^k = 2\cdot 10^0 + 1\cdot 10^1$$

Binarnie:

$$(1100)_2=\sum_{k=0}^{k=3} a_k \cdot 2^k = 0\cdot 2^0 + 0\cdot 2^1 + 1\cdot 2^2 + 1\cdot 2^3$$

Szesnastkowo:

$$(C)_{16}=\sum_{k=0}^{k=0} a_k \cdot 16^k = C\cdot 16^0$$

Liczby

Liczba miesięcy w roku kalendarzowym:

Dziesiętnie:

$$(12)_{10} = \sum_{k=0}^{k=2} a_k \cdot 10^k = 2 \cdot 10^0 + 1 \cdot 10^1$$

Binarnie:

$$(110\boxed{0})_2 = \sum_{k=0}^{k=3} a_k \cdot 2^k = 0 \cdot 2^0 + 0 \cdot 2^1 + 1 \cdot 2^2 + 1 \cdot 2^3$$

Pojedyncza cyfra w zapisie binarnym – najmniejsza jednostka informacji: **bit.**

Liczba 1100 ma cztery bity.

Liczby

Maksymalna liczba jakią można zapisać na 4 bitach:

$$(1111)_2 = \sum_{k=0}^{k=3} a_k \cdot 2^k = 1\cdot 2^0 + 1\cdot 2^1 + 1\cdot 2^2 + 1\cdot 2^3 =$$

$$= 2^4 - 1 = 15$$

Maksymalna liczba jakią można zapisać na n bitach to **2^n-1,** czyli włącznie z zerem daje to **2^n** liczb.

8 bitów (bajt) – 256 liczb

16 bitów – 65 356

24 bity – 16 777 216

32 bity – 4 294 967 296

Liczby

Konwencje zapisu:

liczba bez przedrostka: zwykle oznacza zapis w systemie dziesiętnym: 12

liczba z przedrostkiem **0b** oznacza zapis w systemie dwójkowym: 0b1100

liczba z przedrostkiem **0x** oznacza zapis w systemie szesnastkowym: 0xC

Ważne wielokrotności w systemie dwójkowym:

2^3 bitów = 8 bitów – bajt, 1B = 8 bitów

$2^{10} = 1024$ – kilo, np. Kb – kilo bit

$2^{20} = 1024 \cdot 1024 = 1048576$ – mega, np. Mb – mega bit

$2^{30} = 1024 \cdot 1024 \cdot 1024 = 10737441824$ – giga, np. Gb – giga bit

Uwaga:

Kb (kib) oznacza kilo w systemie dwójkowym,

kb – w systemie dziesiętnym (nie zawsze przestrzegane)

Dlaczego system binarny?

Reprezentacja:

Liczby, w dowolnym formacie muszą być jakoś przechowywane "w środku" komputera. Najprościej można reprezentować 0 i 1: **0 – nie ma, 1 – jest**. Komputery są urządzeniami elektronicznymi, więc

jest – **napięcie na jakimś elemencie przekracza ustalony próg**

nie ma – **oznacza, że napięcie jest poniżej tego progu**

Operacje arytmetyczne również są bardzo proste w reprezentacji binarnej:

0+0 = 0, 1+1=10, 0+1 = 1

Sumy kontrolne

Sumy kontrolne (*ang. checksum*)**:** Do informacji którą chcemy zapisać dołączamy liczby uzyskane w wyniku operacji matematycznych wykonanych na zapisywanych danych. Przy odczycie suma kontrolna jest liczona jeszcze raz i sprawdzana z wartością zapisaną.

Numer PESEL: przykładowy numer PESEL **44051401458 (abcdefghijk)**

cyfry [1-6] – data urodzenia (RRMMDD)

cyfry [7-10] – numer serii z oznaczeniem płci

cyfra [10] – płeć

cyfra [11] – cyfra kontrolna

Wzór na obliczenie cyfry kontrolnej (jedenestej, czyli k):

k = 1a + 3b + 7c + 9d + 1e + 3f + 7g + 9h + 1i + 3j

Uwaga: Poprawność sumy kontrolnej nie oznacza zawsze poprawności numeru PESEL. Jedynie znacznie zmiejsza prawdopodobieństwo popełnienia błędu.

Korekcja błędów

Korekcja błędów (*ang. Error Checking and Correction, ECC*)
korzystając z sum kontrolnych możemy nie tylko wykrywać błędy, ale także próbować je korygować.

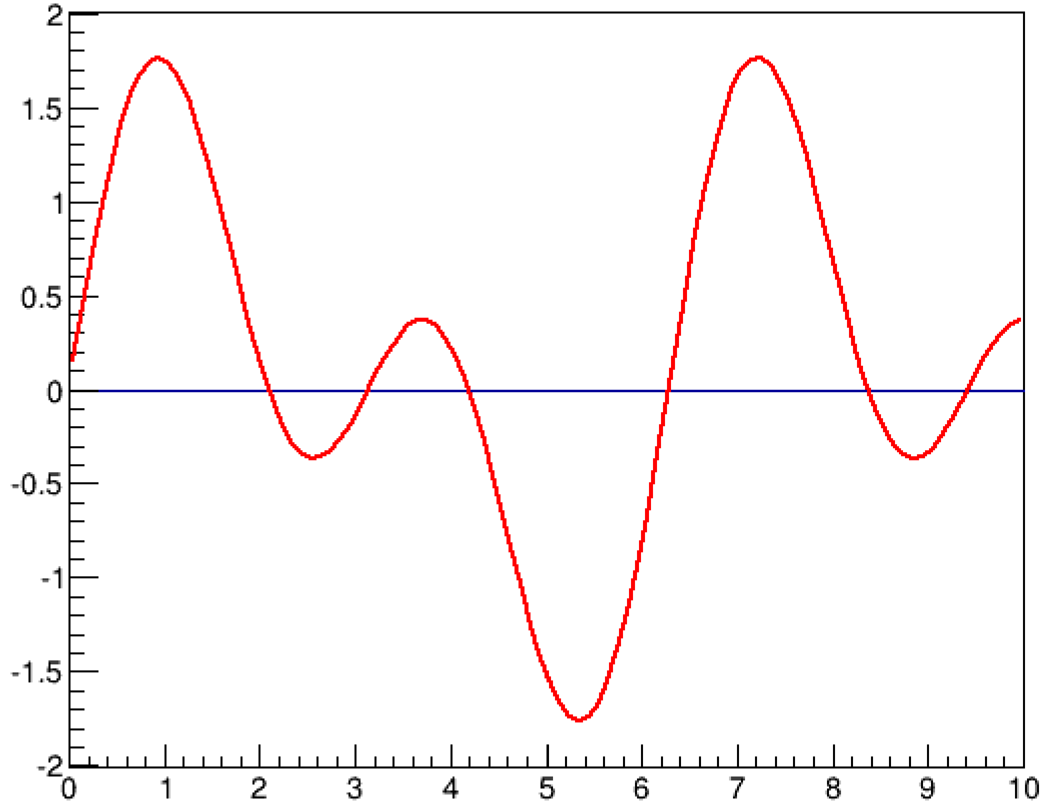
Na płytach audio CD algorytmy korekcji błedów pozwalają na wykrycie błędów w odczycie. Błędnie odczytane bloki danych mogą być opuszczone przy odtwarzaniu muzyki. Brakujące dane mogą być odtworzone przy pomocy interpolacji liniowej sąsiednich próbek (dla danych, np. programu komputerowego to nie działa!).

Metoda EEC używana w audio CD pozwala na wykrycie 99,9985% błędów

Sygnał analogowy

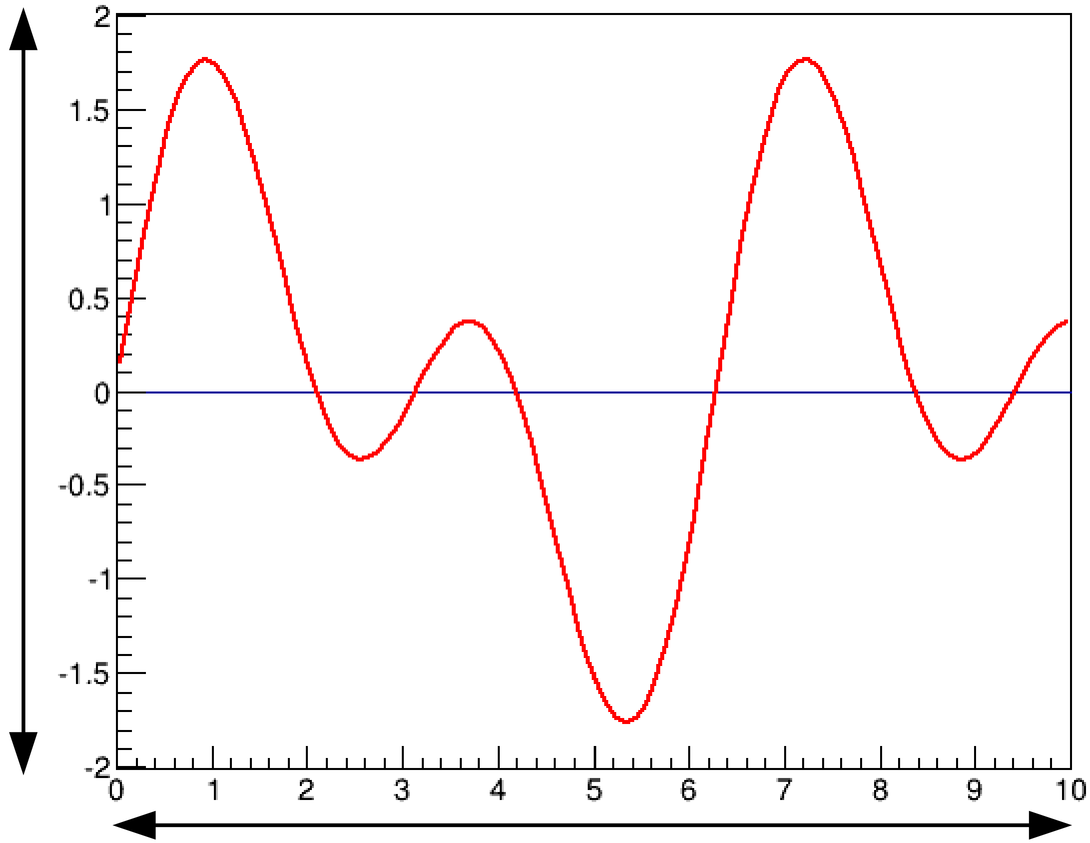
Sygnał – zmienny w czasie przebieg pewnej wartości, np. głośności dzwięku.

Analogowy – dowolna wartość między minimum i maksimum jest możliwa. Wielkość minimum i maksimum mogą być “nieskończone”. Zmienność w czasie jest ciągła i dowolna.



Sygnał analogowy

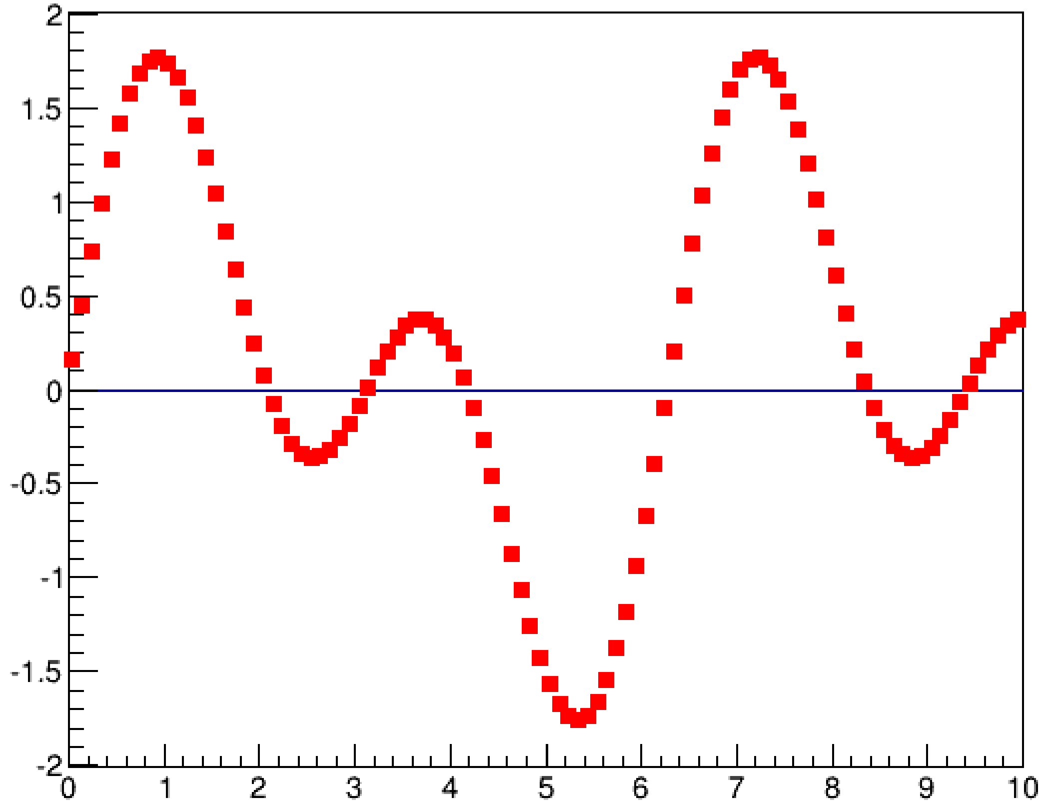
Ciągły rozkład obserwowanej wartości.



Dowolna zmienność w czasie.
Zmiany mogą być bardzo szybkie (duża częstość) lub bardzo wolne (mała częstość).

Sygnał cyfrowy

Sygnał – zmienny w czasie przebieg pewnej wartości, np. głośności dzwięku.
Cyfrowy – tylko skończona liczba wartości między minimum i maksimum jest dopuszczalna. Wielkość minimum i maksimum muszą być skończone. Liczba próbek w jednostce czasu musi być skończona.

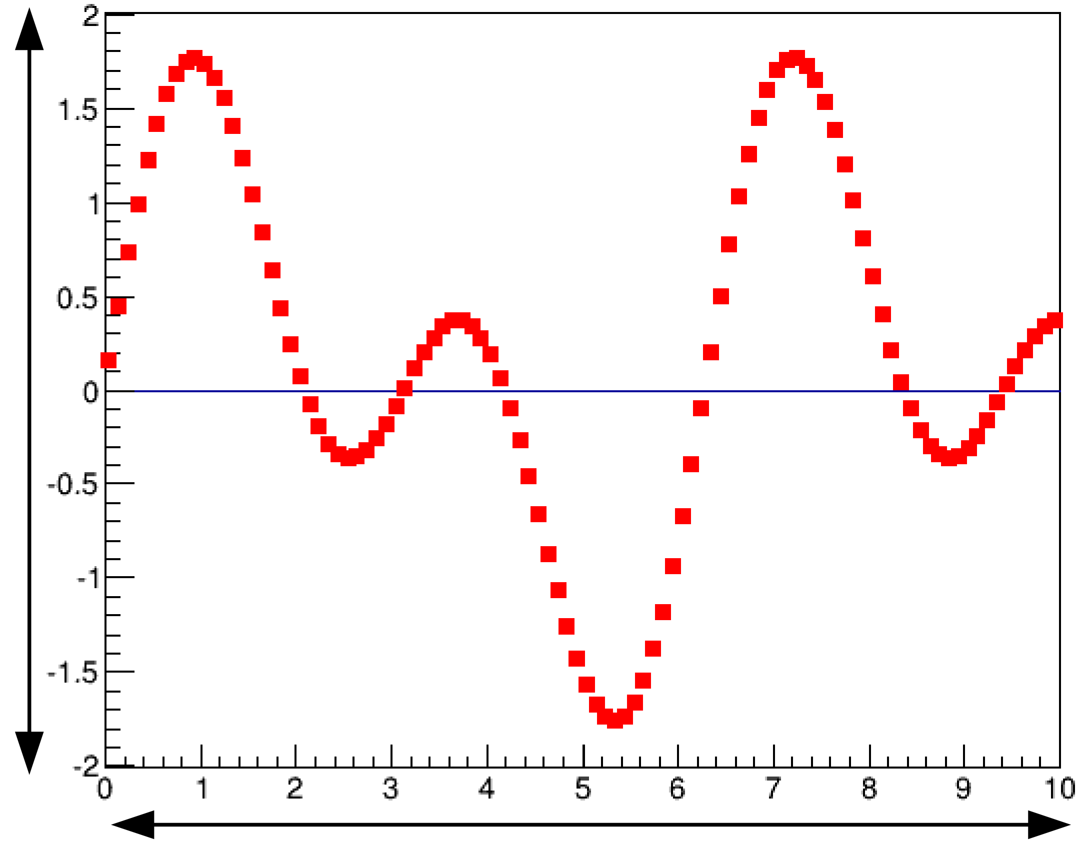


Digitalizacja

Każdej obserwowanej wartości przypisujemy numer przedziału do którego ona należy. Liczba przedziałów dana przez liczbę bitów przeznaczonych na zapis wartości.

Np. dla 8 bitów:

zakres [-2,+2] dzielimy na 256 przedziałów. Pomiarowi "0" przypisujemy wartość 129, bo należy on do 129 przedziału o granicach [0,0.0156[



Liczba punktów pomiarowych w czasie zależy od częstotliwości próbkowania. 44.1 kHz oznacza 44100 próbek na sekundę.

Interpolacja

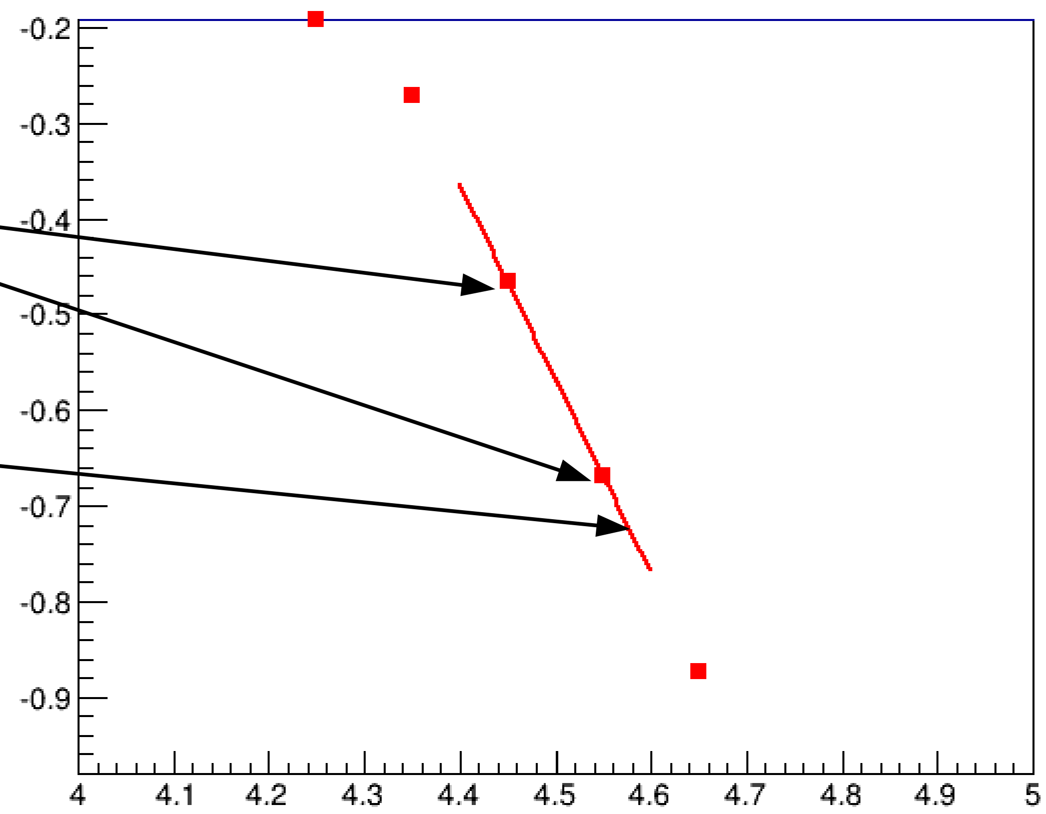
Wartości pomiędzy kolejnymi próbkami w czasie możemy wyznaczyć analitycznie (ze wzoru). Najprostrzy sposób to interpolacja liniowa.

Zapisane dane w dwóch kolejnych chwilach: (y_1,t_1), (y_2,t_2)

Prosta przechodząca przez dwa punkty:

$y = a \cdot t + b$,

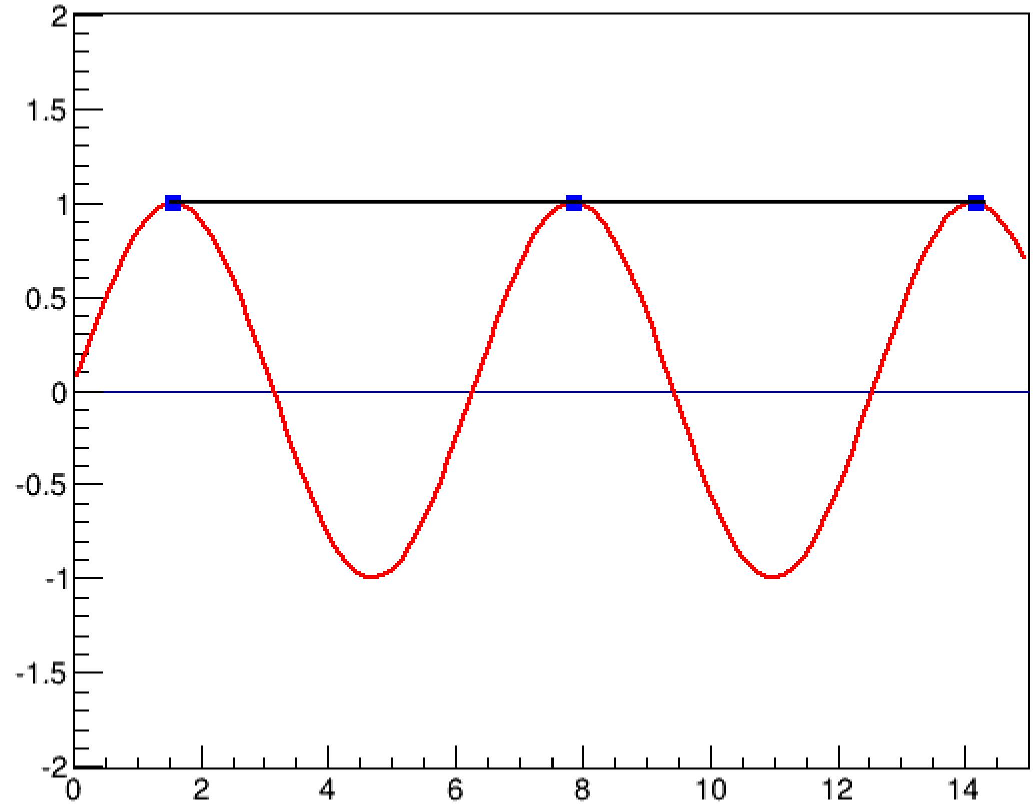
$y_{1/2} = a \cdot t_{1/2} + b$



Ze wzoru na prostą możemy policzyć y dla dowolnej wartości czasu t. Zwykle wynik będzie bliski rzeczywistemu tylko dla przedziału $[t_1,t_2]$

Twierdzenie Nyquista

Załóżmy prosty sygnał sinusoidalny: $f(t) = \sin(\omega t)$, $\omega = 2\pi/T$. Próbkujemy raz na okres, czyli z częstością $1/T$



W tej sytuacji interoplacja liniowa da nam prostą o wartości stałej w czasie! Straciliśmy całą informację o zależności od czasu.

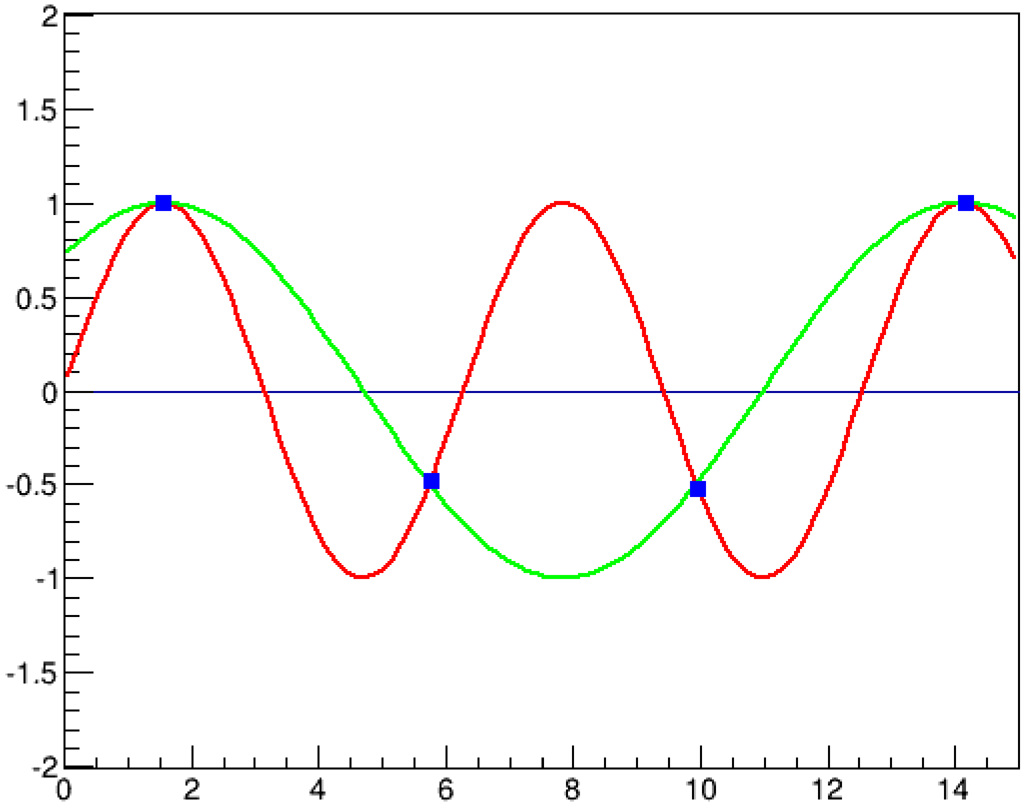
Efekt stroboskopowy

https://www.youtube.com/watch?v=4XkywLza-9E

Twierdzenie Nyquista

Załóżmy prosty sygnał sinusoidalny: $f(t) = \sin(\omega_0 t)$, $\omega_0 = 2\pi/T$. Próbkujemy raz na okres, czyli z częstością 1/T

Co się stanie jeśli będziemy próbkowali z częstością $\omega_p = 1.5\omega_0$, czyli 3 razy na 2 okresy?



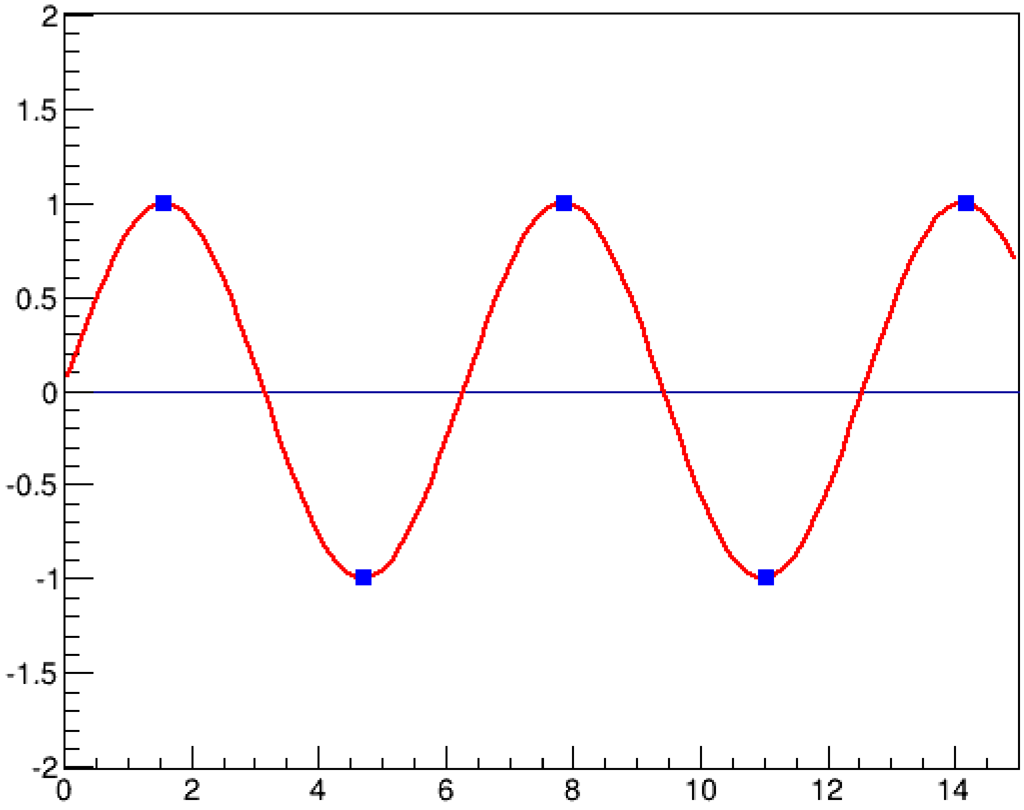
Sygnał cyfrowy będzie wyglądał identycznie jak byśmy obserwowali na wejściu sygnał o częstości $\omega = \omega_p - \omega_0 = (1.5-1)\omega_0 = 0.5\omega_0$

→

Aliasing: pojawienie się w sygnale składowych o błędnych częstotliwościach (aliasów). $\omega_a = |\omega_0 - \omega_p|$

Twierdzenie Nyquista

Załóżmy prosty sygnał sinusoidalny: $f(t) = \sin(\omega_0 t)$, $\omega_0 = 2\pi/T$. Próbkujemy raz na okres, czyli z częstością 1/T



Dopiero gdy częstość próbkowania będzie conajmniej $\omega_p = 2\omega_0$ będziemy mogli odtworzyć oryginalny sygnał

Tw. Nyquista: aby móc dokładnie odtworzyć ciągły sygnał o ograniczonym paśmie z jego próbek, należy zastosować częstość próbkowania równą conajmniej podwojonej najwyższej częstości sygnału.

Twierdzenie Nyquista

Tw. Nyquista: aby móc dokładnie odtworzyć ciągły sygnał o ograniczonym paśmie z jego próbek, należy zastosować częstość próbkowania równą conajmniej podwojonej najwyższej częstości sygnału.

Zakresy częstości:		Częstości próbkowania:	
Mowa	100 Hz – 3.5 kHz	telefonia	8 kHz
dzwięki odbierane przez człowieka:	20 Hz – 20 kHz	płyty CD	44.1 kHz
		radio FM	22.05 kHz

Zapis audio CD

Standardowa płyta CD: 333 000 sektorów

sektor: 2352 bajty danych audio

Cała płyta CD: 333 000 · 2352 = 783 216 000 000 bajtów

Zapis dźwięku na płycie CD:

- częstotliwość próbkowania: 44.1 kHz
- liczba bitów używana do kodowania natężenia dźwięku: 16

Sekunda dźwięku stereo z płyty CD: **2 · 16 · 44100** = 1411200 bitów

Minuta dźwięku: 1411200 · 60/8/1024/1024 = **10 MiB**

Płyta audio CD mieści: 783 216 000 000/ 10 MiB/min = **74' muzyki**

Na każde 24 bajty muzyki jest dodatkowo zapisywane 8 bajtów EEC.

Digitalizacja obrazu

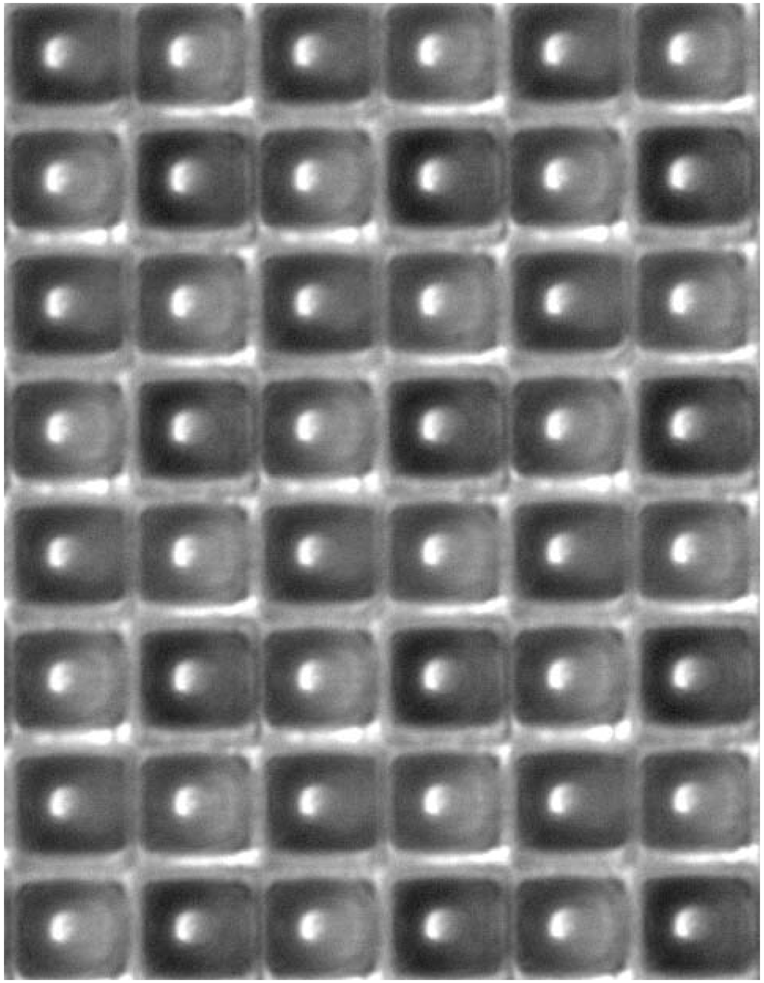
Fotografia cyfrowa: pomiar jasności pikseli (punktów) matrycy, na którą pada światło.

Rozmiar matrycy: rzędu milionów pikseli

Rozmiar pojedynczego piksela: **rzędu 1x1 μm**

Liczba bitów kodujących obraz: zależy od ADC (*ang. Analog to Digital Converter*), około **12**

Jak jest rozpoznawany kolor?



Photographs by Jack/The Landingfield and used with permission

Digitalizacja obrazu

Fotografia cyfrowa: pomiar jasności pikseli (punktów) matrycy, na którą pada światło.

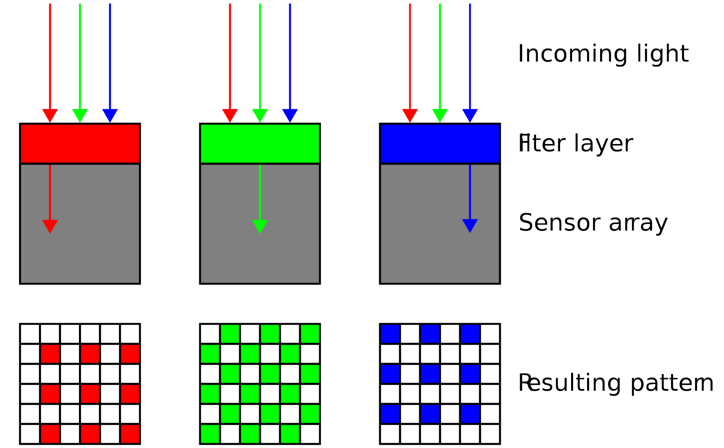
Rozmiar matrycy: rzędu milionów pikseli

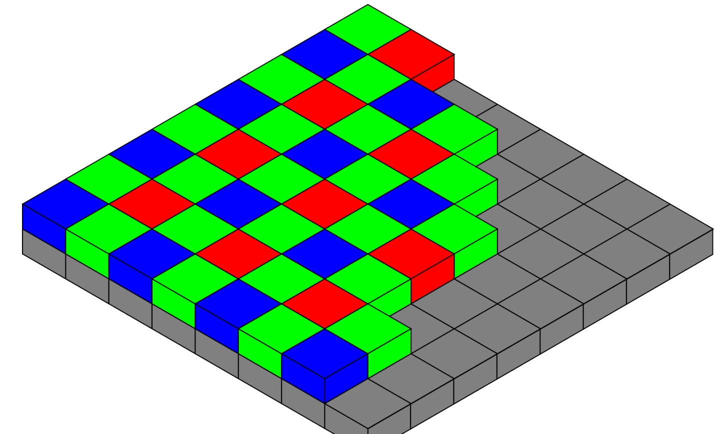
Rozmiar pojedynczego piksela: **rzędu 1x1 µm**

Liczba bitów kodujących obraz: zależy od ADC, około **12**

Jak jest rozpoznawany kolor?

Każdy piksel rozpoznaje jeden kolor (filtr kolorowy z przodu): R G B





http://en.wikipedia.org/wiki/Bayer_filter

Digitalizacja obrazu

Fotografia cyfrowa: pomiar jasności pikseli (punktów) matrycy, na którą pada światło.

Rozmiar matrycy: rzędu milionów pikseli

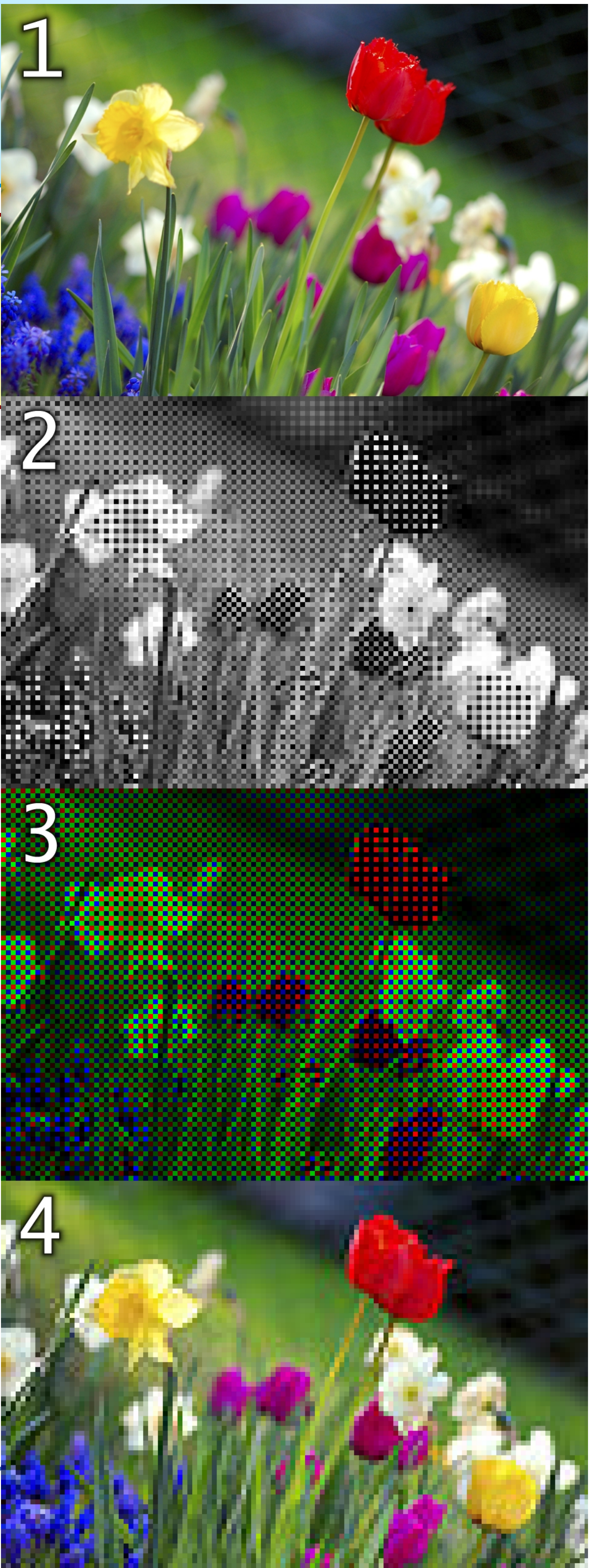
Rozmiar pojedynczego piksela: **rzędu 1x1 μm**

Liczba bitów kodujących obraz: zależy od ADC, około **12**

Jak jest rozpoznawany kolor?

Każdy piksel rozpoznaje jeden kolor (filtr kolorowy z przodu): R G B

Finalny kolor jest jest średnią z sąsiadów

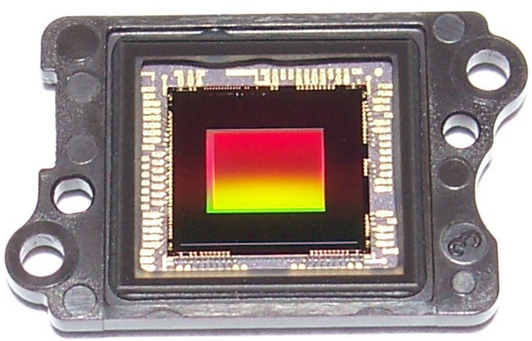


http://en.wikipedia.org/wiki/Bayer_filter

Digitalizacja obrazu

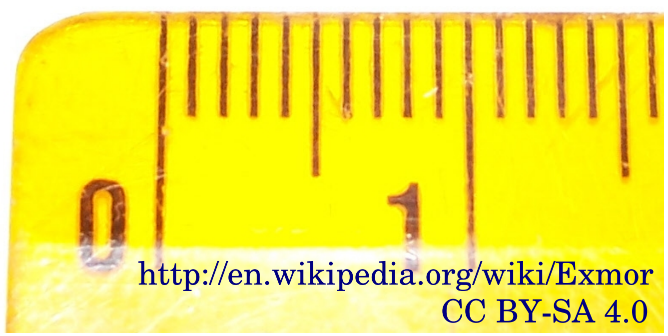
Fotografia cyfrowa: pomiar jasności pikseli (punktów) matrycy, na którą pada światło.

Rozmiar matrycy: rzędu milionów pikseli
Rozmiar pojedynczego piksela: **rzędu 1x1 µm**
Liczba bitów kodujących obraz: **12**



Weźmy matrycę 10 MP.
Ile danych trzeba zapisać dla jednego zdjęcia?

$$\mathbf{10 \cdot 10^6 \cdot 12 / 8 / 1024 \; / 1024 = 14.3 \; MiB!}$$



http://en.wikipedia.org/wiki/Exmor