

Technologie cyfrowe

Artur Kalinowski

Zakład Cząstek i Oddziaływań
Fundamentalnych

Pasteura 5, pokój 4.15

Artur.Kalinowski@fuw.edu.pl



Semestr letni 2014/2015

- 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: 8/10.04.2015, 27/29.05.2015
- Wykład: egzamin pisemny (test)
Termin: 19.06.2015, 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>.

Kodowanie: tablica zamiany znaków alfabetu, cyfr i znaków specjalnych na liczby.

Kodowanie ASCII (*ang. American Standard Code for Information Interchange*) używa 7 bitów do kodowania 128 znaków (w tym znaków specjalnych) używanych w języku angielskim. **Nie zawiera polskich znaków**

0	NUL	16	DLE	32	SPC	48	0	64	@	80	P	96	`	112	p
1	SOH	17	DC1	33	!	49	1	65	A	81	Q	97	a	113	q
2	STX	18	DC2	34	"	50	2	66	B	82	R	98	b	114	r
3	ETX	19	DC3	35	#	51	3	67	C	83	S	99	c	115	s
4	EOT	20	DC4	36	\$	52	4	68	D	84	T	100	d	116	t
5	ENQ	21	NAK	37	%	53	5	69	E	85	U	101	e	117	u
6	ACK	22	SYN	38	&	54	6	70	F	86	V	102	f	118	v
7	BEL	23	ETB	39	'	55	7	71	G	87	W	103	g	119	w
8	BS	24	CAN	40	(56	8	72	H	88	X	104	h	120	x
9	HT	25	EM	41)	57	9	73	I	89	Y	105	i	121	y
10	LF	26	SUB	42	*	58	:	74	J	90	Z	106	j	122	z
11	VT	27	ESC	43	+	59	;	75	K	91	[107	k	123	{
12	FF	28	FS	44	,	60	<	76	L	92	\	108	l	124	
13	CR	29	GS	45	-	61	=	77	M	93]	109	m	125	}
14	SO	30	RS	46	.	62	>	78	N	94	^	110	n	126	~
15	SI	31	US	47	/	63	?	79	O	95	_	111	o	127	DEL

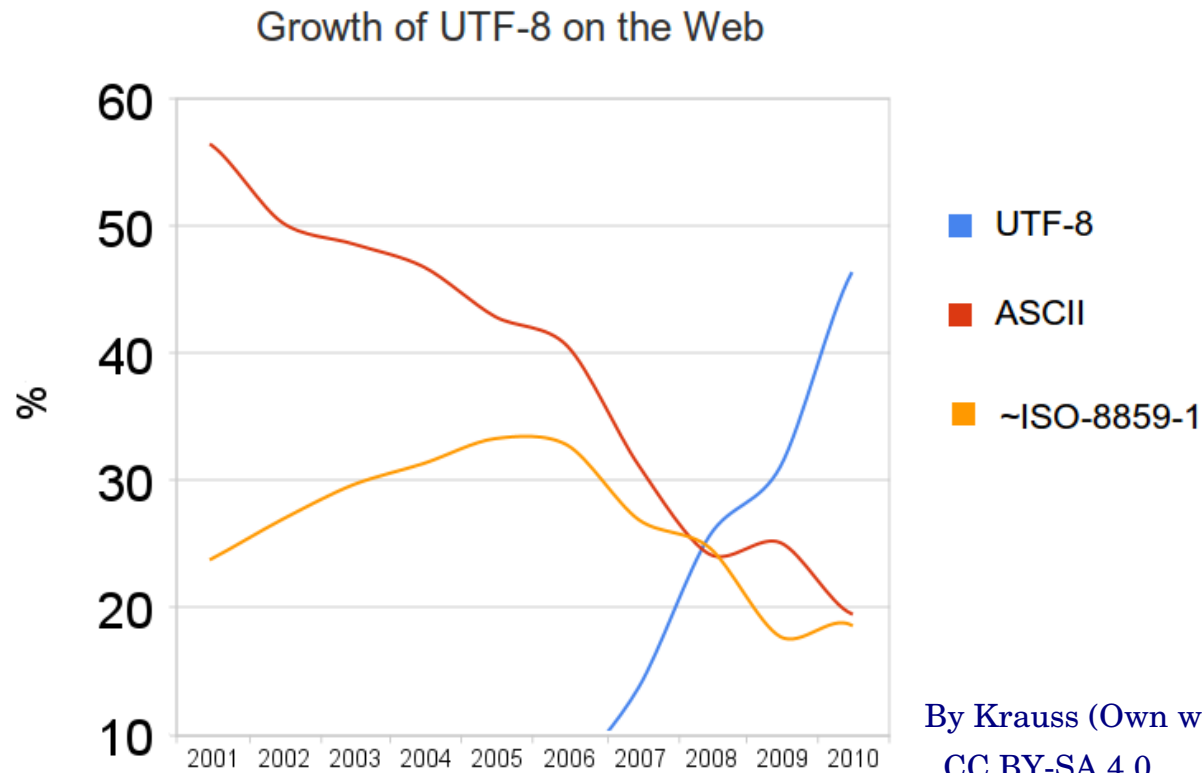
Do niedawna dominującym rozszerzeniem ASCII obejmującym polskie znaki był 8-bitowy standard ISO 8859-2 (Latin-2) kodujący 191 znaków pisma łacińskiego i języków środkowo europejskich.

Unicode: kodowanie (=przypisanie numerów do znaków) mające w założeniu obejmować wszystkie języki świata.

UTF-8: system kodowania Unicode używający dla oszczędności od 8 do 32 bitów. Często używane znaki są kodowane przy użyciu 8 bitów, inne przy pomocy 2 do 4 bajtów.

$$2^8 = 256$$

$$2^{32} = 4E9$$



Jaka ilość danych jest potrzebna by zapisać tekst w kodowaniu ISO 8859-2?

Strona tekstu to około 3000 znaków, jeden znak to 8 bitów

3000 x 1 bajt / 1024 \approx 3 kB

Gruba książka to 341 stron: $341 \times 3 \text{ kB} = 1 \text{ MB}$

Biblioteka UW: **2 miliony książek: $2\text{M} \times 1 \text{ MB} \approx 2 \text{ TB}$**

Dysk twardy o pojemności 2 TB: $\sim 400 \text{ PLN}$

Czy to oznacza, że zapis cyfrowy dyskwalifikuje zapis analogowy?

Kod ASCII (dec): 065 108 097 032 109 097 032 107 111 116 097 046

Kod ASCII (hex): 416c 6120 6d61 206b 6f74 612e 0a

$$(65)_{10} =$$

$$4 \times 16 + 1 = (41)_{16}$$

$$(108)_{10} =$$

$$6 \times 16 + 12 = (6C)_{16}$$

0	NUL	16	DLE	32	SPC	48	0	64	@	80	P	96	`	112	p
1	SOH	17	DC1	33	!	49	1	65	A	81	Q	97	a	113	q
2	STX	18	DC2	34	"	50	2	66	B	82	R	98	b	114	r
3	ETX	19	DC3	35	#	51	3	67	C	83	S	99	c	115	s
4	EOT	20	DC4	36	\$	52	4	68	D	84	T	100	d	116	t
5	ENQ	21	NAK	37	%	53	5	69	E	85	U	101	e	117	u
6	ACK	22	SYN	38	&	54	6	70	F	86	V	102	f	118	v
7	BEL	23	ETB	39	'	55	7	71	G	87	W	103	g	119	w
8	BS	24	CAN	40	(56	8	72	H	88	X	104	h	120	x
9	HT	25	EM	41)	57	9	73	I	89	Y	105	i	121	y
10	LF	26	SUB	42	*	58	:	74	J	90	Z	106	j	122	z
11	VT	27	ESC	43	+	59	;	75	K	91	[107	k	123	{
12	FF	28	FS	44	,	60	<	76	L	92	\	108	l	124	
13	CR	29	GS	45	-	61	=	77	M	93]	109	m	125	}
14	SO	30	RS	46	.	62	>	78	N	94	^	110	n	126	~
15	SI	31	US	47	/	63	?	79	O	95	_	111	o	127	DEL

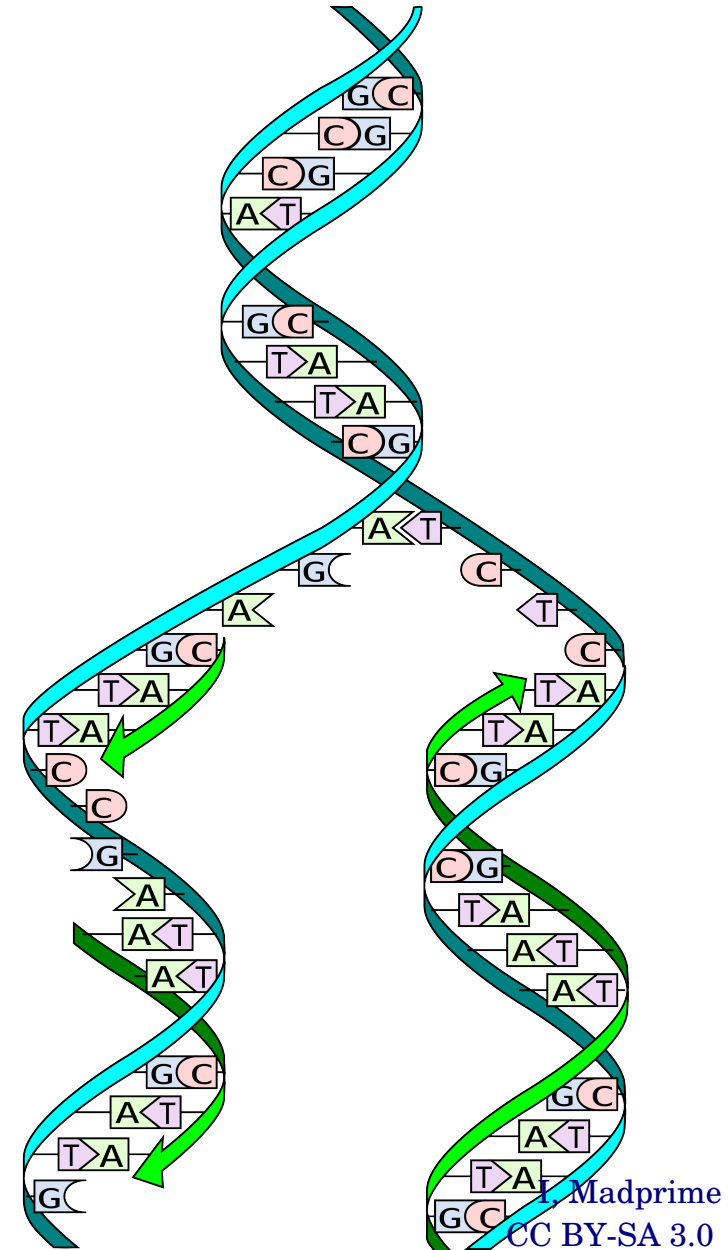
Jaka ilość danych jest potrzebna by zapisać pełen ludzki kod DNA?

Genomy (46 chromosomów) pojedynczego człowieka składają się z 6 miliardów par bazowych A – T, T – A, G – C i C – G

Cztery pary wymagają dwu bitów do kodowania: 00, 01, 10, 11.

$6G \times 4 \text{ bity} / 8 / 1024 / 1024 / 1024$
 $\approx 1.4 \text{ GB}$

Czy te 1.4 GB wystarcza by odtworzyć człowieka?



Madprime
CC BY-SA 3.0

Dyski optyczne



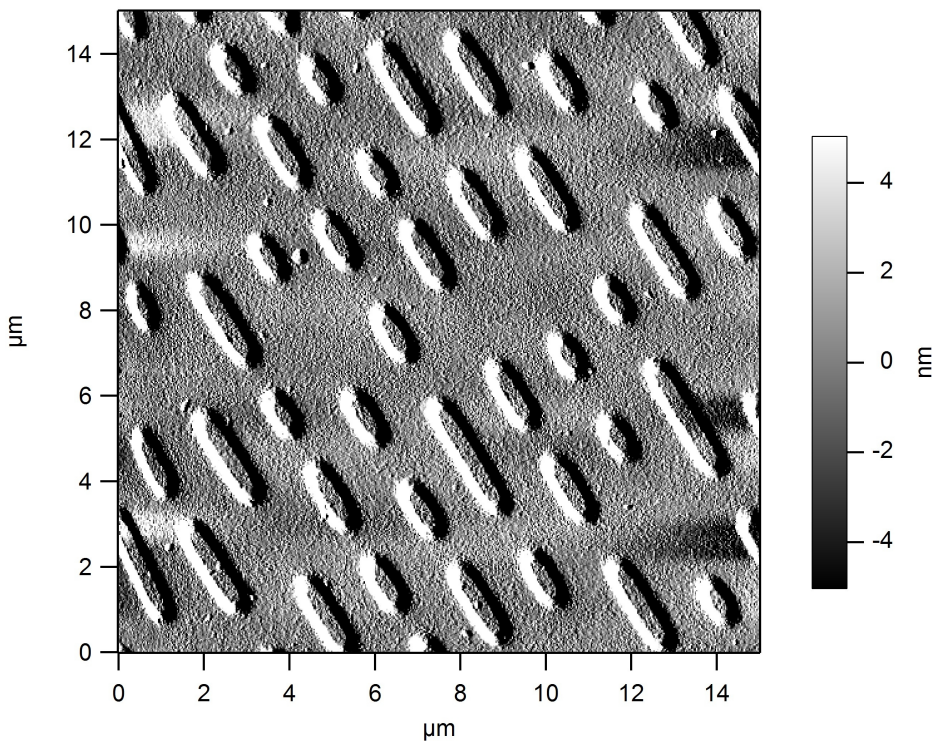
http://en.wikipedia.org/wiki/Optical_disc

CC BY-SA 3.0

Dysk optyczny: nośnik danych w którym bity są kodowane w postaci obszarów odbijających lub nie odbijających światła.

Jeden – **brak odbicia** Zero – **odbicie**

Powierzchnia odbijająca jest pokryta wgłębieniami (*ang. pits*) i polami (*ang. lands*)

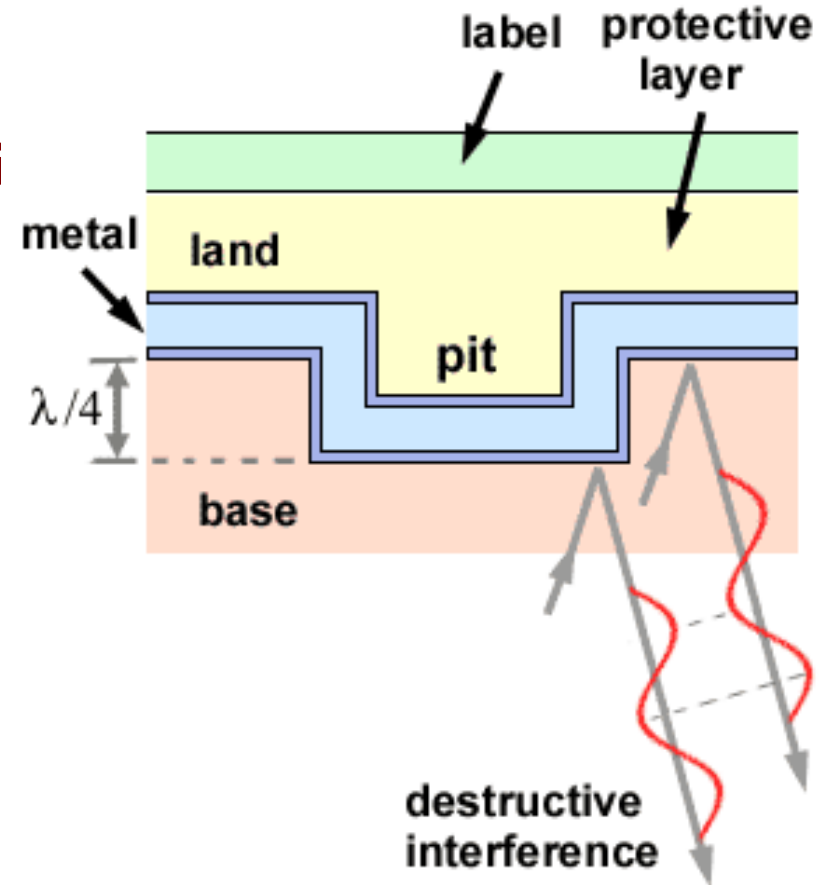


Spiral Data Track on Optical Discs
(CD, DVD and High Definition Discs)

Dysk optyczny: nośnik danych w którym bity są kodowane w postaci obszarów odbijających lub nie odbijających światła.

Głębokość wgłębień to $1/4$ długości fali lasera odczytującego → **na brzegu wgłębienia następuje interferencja destrukcyjna fali powodująca brak odbicia.**

Różne formaty płyt optycznych (CD, DVD, Blu-ray) różnią się długością fali lasera odczytującego. Im mniejsza długość, tym mniejsze mogą być wgłębienia i pola, także w długości i szerokości)

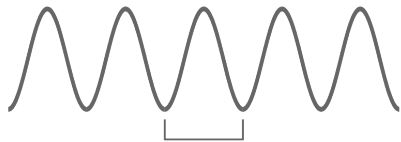
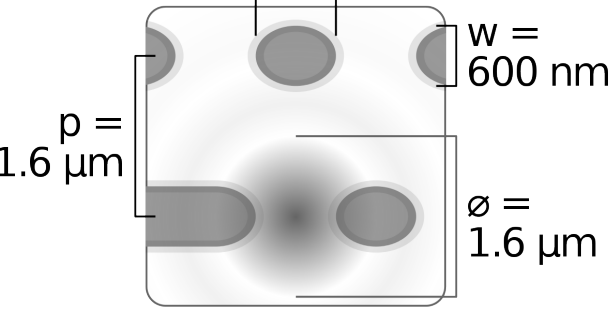


http://www.hk-phy.org/articles/cdrom/cdrom_e.html

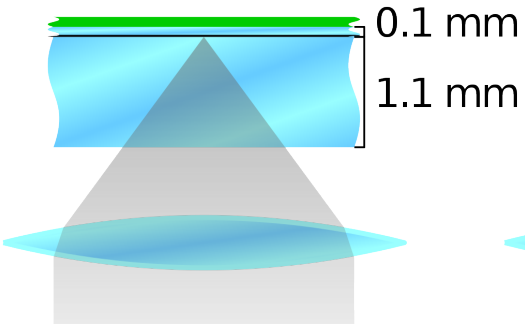
Dyski optyczne

CD

$l = 800 \text{ nm}$

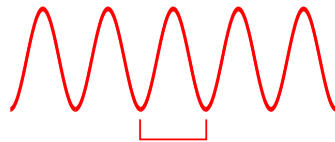
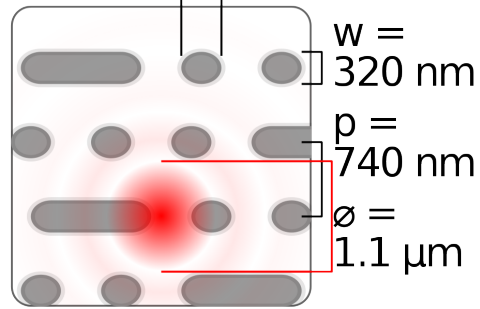


$\lambda = 780 \text{ nm}$

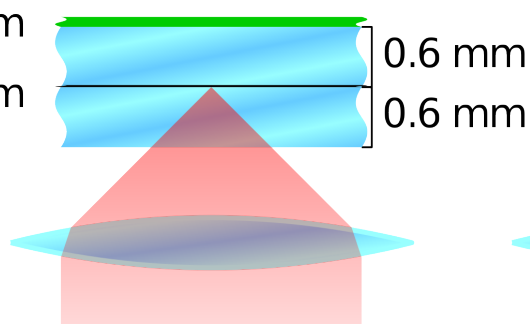


DVD

$l = 400 \text{ nm}$

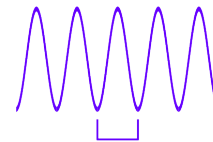
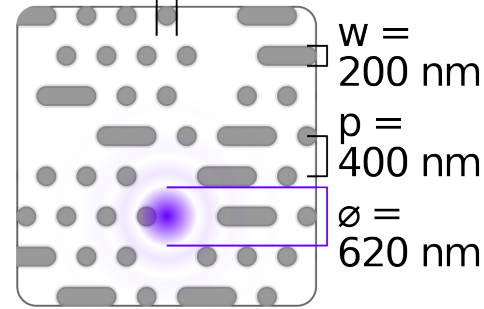


$\lambda = 650 \text{ nm}$

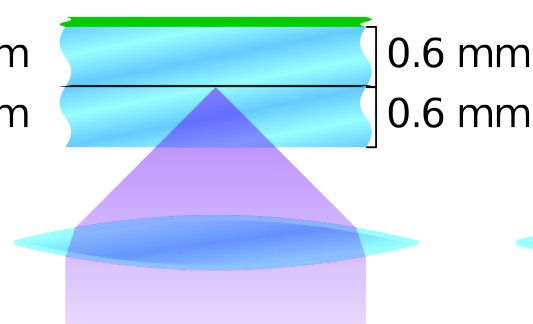


HD DVD

$l = 200 \text{ nm}$

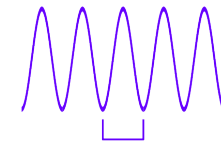
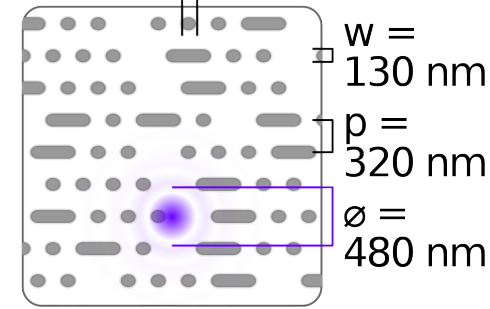


$\lambda = 405 \text{ nm}$

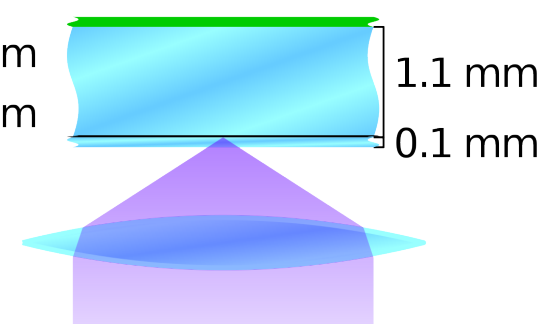


Blu-ray

$l = 150 \text{ nm}$



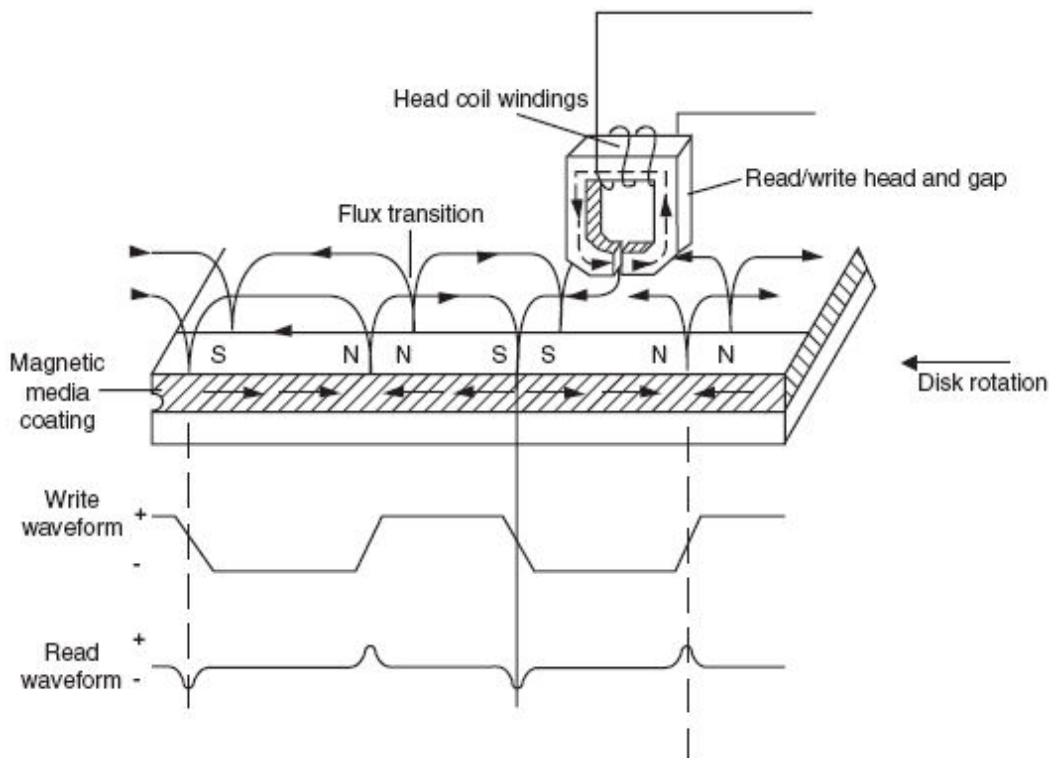
$\lambda = 405 \text{ nm}$



http://en.wikipedia.org/wiki/Optical_disc
CC BY-SA 3.0

Dysk magnetyczny: nośnik danych w którym do zapisu danych używa się obszarów o różnym kierunku namagnesowania.

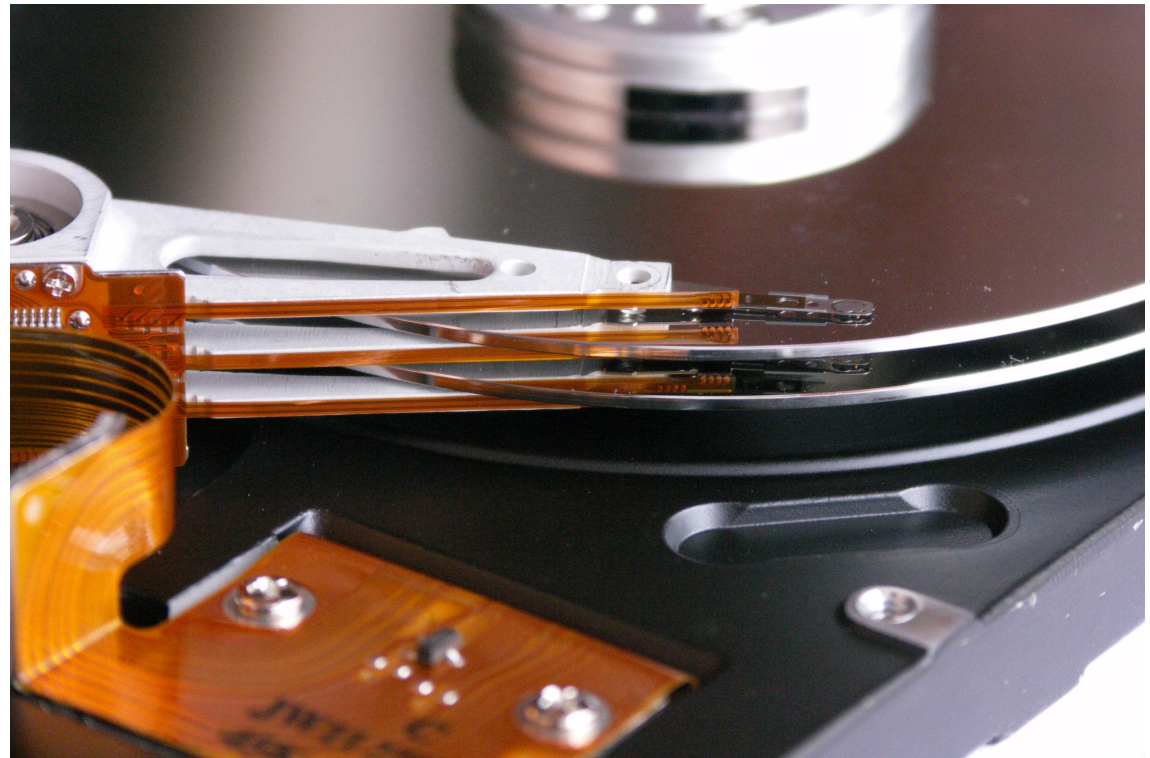
W dyskach twardych granice między obszarami różnego namagnesowania są odczytywane jako “0” lub “1”



By Evan-Amos (Own work)
CC BY-SA 3.0

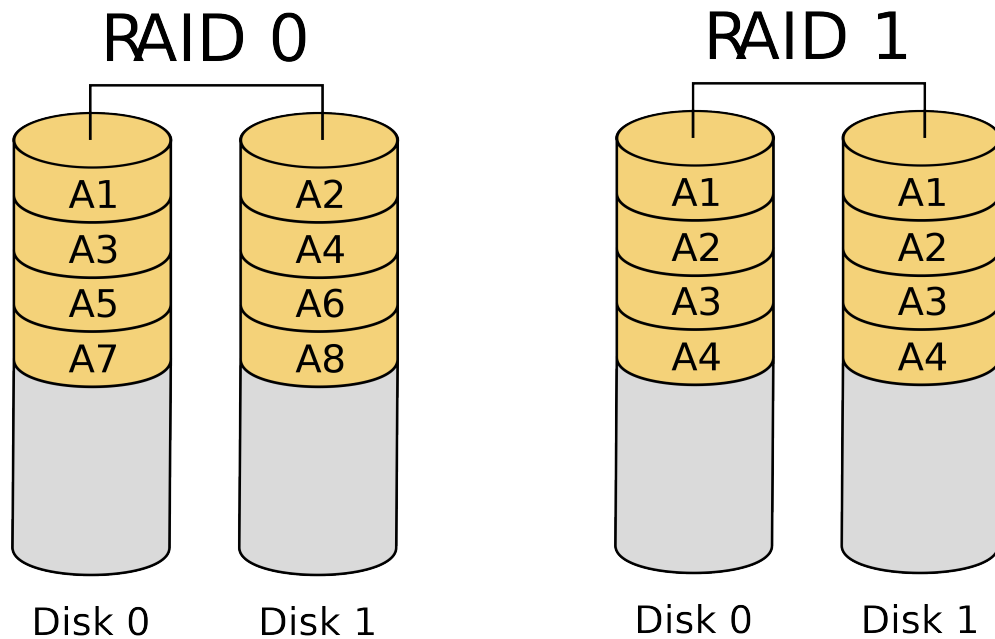
Jeden dysk twardy zawiera zwykle kilka talerzy.

Dyski obracają się z prędkością 7 do 10 tys. obrotów na min.



<http://gadgetadda.com>

RAID (ang. *Redundant Array of Independent Disks*): dane są zapisywane jednocześnie na kilku dyskach. Awaria jednego dysku nie oznacza straty danych. Duże ilości danych zawsze są przechowywane na macierzach dyskowych



Taśmy magnetyczne

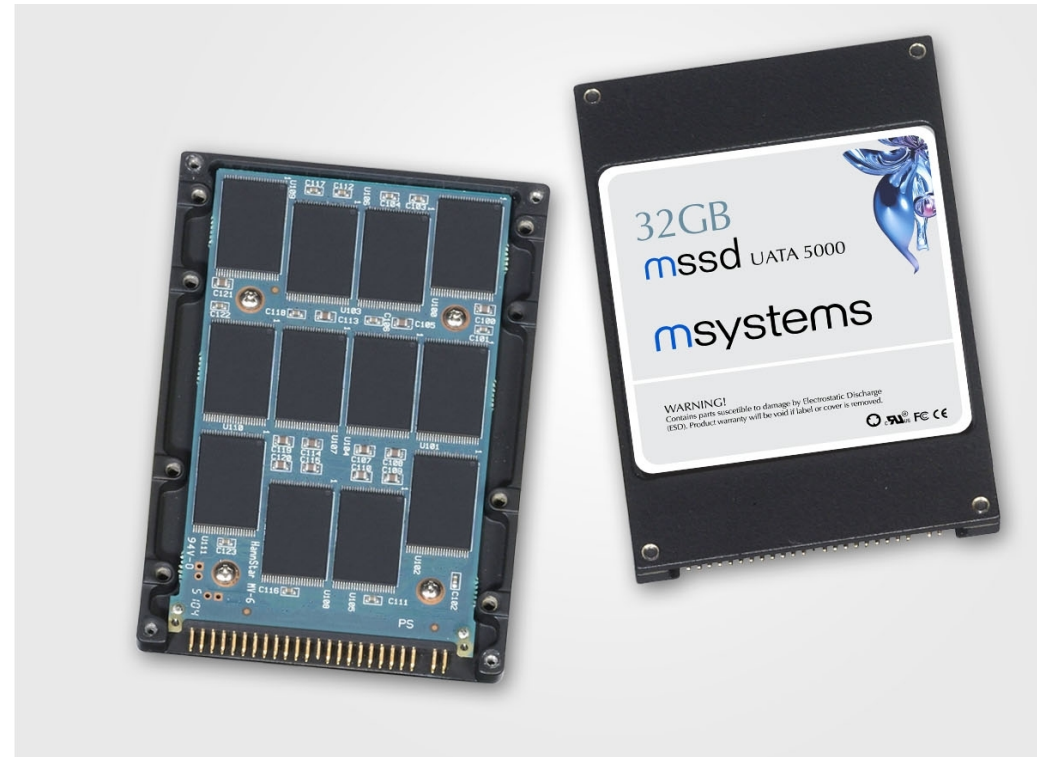


CERN

Solid State Drive: nośnik danych w którym do zapisu danych używa się stanu komórek układu scalonego.

W odróżnieniu od płyt CD czy dysków twardych SSD nie posiadają żadnych ruchomych części. Zapis i odczyt danych następują przez zmianę stanu komórki układu scalonego.

Dane w dyskach SSD są przechowywane w modułach pamięci flash, taki jak w dyskach USB (*ang. pendrive*)



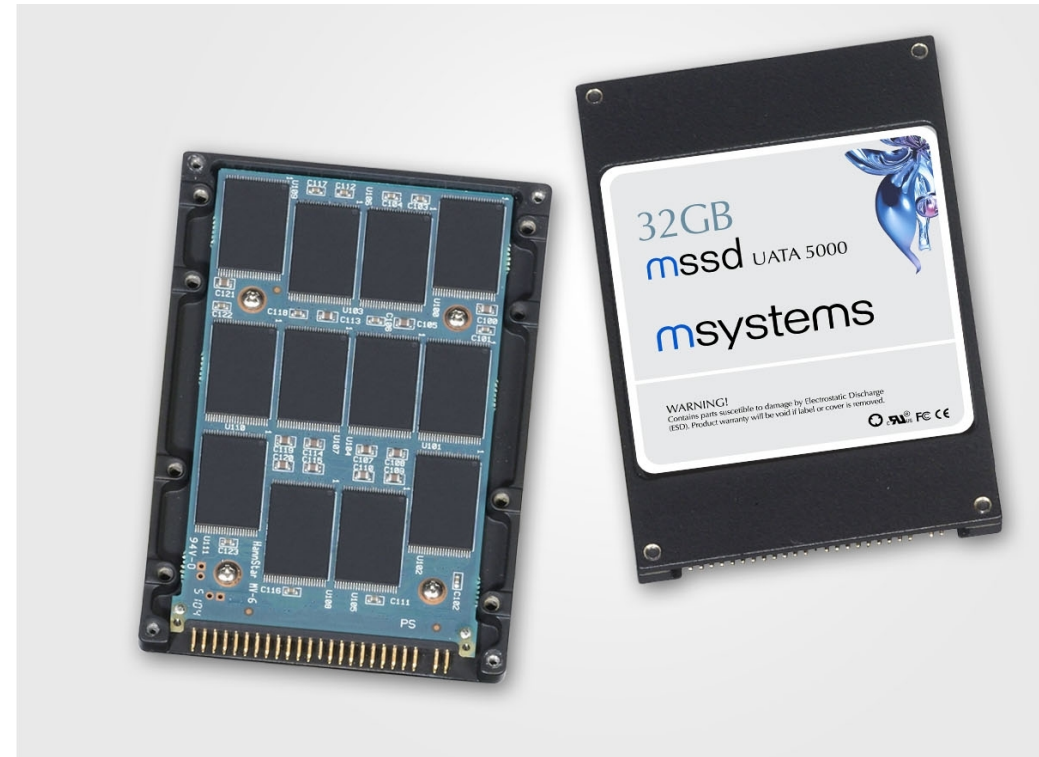
By Rochellesinger at en.wikipedia
CC BY-SA 2.5

Zalety SSD:

- szybszy dostęp do danych
- bezgłówna praca
- mniejsze zużycie prądu
- znacznie większa odporność na warunki zewnętrzne (upadki, kurz, temperatura)

Wady SSD:

- mniejsza pojemność
- krótsza żywotność (ograniczona liczba cykli zapis – odczyt), ale i tak wystarczy na ~10 lat



By Rochellesinger at en.wikipedia

CC BY-SA 2.5

Plik (*ang. file*): uporządkowany zbiór danych. Nazwa pliku nie jest częścią pliku. Rodzaje plików: tekstowe **i binarne** (czyli nie tekstowe).

odczyt	interpretacja
01000011010011110011111100111111	trzydzieści dwa bity
0.747303187847137451	liczba rzeczywista (32-bitowa)
67 79 63 63	cztery liczby całkowite (8-bitowe)
C O ? ?	litery (ASCII)
0100001101001111 0011111100111111	dwie „paczki” po szesnaście bitów
20291 16191	dwie liczby całkowite (16-bitowe)

Kompresja: zmiana sposobu zapisu danych w celu zmniejszenia liczby bitów koniecznych do zapisania informacji.

Algorytm: przepis na wykonanie czynności, zwykle jakieś operacji na danych.

Kompresja bezstratna: z postaci skompresowanej można odtworzyć całkowitą informację wejściową. Kompresja polega na zastosowaniu algorytmów używających redundantne (=powtarzające się) sekwencje danych.

Kompresja stratna: z postaci skompresowanej nie można uzyskać wyjściowego zestawu danych. Kompresja polega na usunięciu części informacji, a następnie zastosowaniu algorytmów używanych w kompresji bezstratnej.

Kompresja bezstratna: z postaci skompresowanej można odtworzyć całkowitą informację wejściową. Kompresja polega na zastosowaniu algorytmów usuwających redundantne (=powtarzające się) sekwencje danych.

Najprostszy wariant: kodowanie długości serii (*ang. Run Length Encoding, RLE*):

Sekwencja oryginalna: `wwwwwwiiiikkkkkkkkiiippppppppeeeeeddddiia`
bajtów: $37 \times 1 = 37$

Sekwencja skompresowana: `5w4i7k3i6p5e4d2i1a`
bajtów: $18 \times 1 = 18$

Stopień kompresji: $37/18 = 2$

ZIP: jeden z najbardziej popularnych formatów kompresji bezstratnej. Pliki w formacie ZIP mają rozszerzenie .zip. Kompresja jest uzyskana między innymi przez użycie algorytmów kodowania Huffmana.

kodowanie Huffmana: dla danego zbioru znaków i stowarzyszonych z nim prawdopodobieństw wystąpienia jest tworzona lista słów kodowych (ciągów bitowych), których długość jest odwrotnie proporcjonalna do prawdopodobieństwa. To oznacza, że im częściej dany symbol występuje (może wystąpić) w ciągu danych, tym mniej zajmie bitów.

Sekwencja oryginalna (tekst): [wwwwiiiikkkkkkkiiippppppeeeeddddiia](http://www.wiiikkkkkkkiiippppppeeeeddddiia)

Sekwencja oryginalna (ASCII): 01110111 01110111 01110111 01110111
01110111 01101001 01101001 01101001 01101001 01101011 01101011
01101011 01101011 01101011 01101011 01101011 01101001 01101001
01101001 01110000 01110000 01110000 01110000 01110000 01110000
01100101 01100101 01100101 01100101 01100101 01100100 01100100
01100100 01100100 01101001 01101001 01100001

Słowa kodowe wg. kodowania Huffmana:

a – 1100, d – 1101, e – 100, i – 01, k – 00, p – 111, w – 101

Sekwencja skompresowana: 101 101 101 101 101 01 01 01 01 00 00 00 00
00 00 00 01 01 01 111 111 111 111 111 111 100 100 100 100 100 1101 1101
1101 1101 01 01 1100

bajtów: 18



Plik (*ang. file*): uporządkowany zbiór danych. Nazwa pliku nie jest częścią pliku.

Rodzaj pliku jest zwykle zakodowany w “rozszerzeniu” – trzech literach nazwy występujących po kropce.

```
pliki tekstowe ASCII          *.txt, *.asc
obrazy                        *.jpg, *.jpeg, *.gif, *.bmp, *.pcx, *.png
video                         *.mpg, *.mpeg, *.avi
dźwięk                        *.mp3, *.wav, *.au, *.mid
pliki dla programów TEX i LATEX  *.tex
teksty programów (C, Pascal, Java...) *.c, *.pas, *.java...
pliki dla WWW                 *.html, *.htm
programy MS Windows          *.exe
skrypty MS Windows           *.bat, *.vbs, *.vsh
PostScript (książki, rysunki)   *.ps, *.eps
Portable Document Format       *.pdf
archiwa skompresowanych plików  *.zip, *.rar, *.arj, *.gz, *.tgz, *.Z, *.tar.Z
```

Piotr J. Durka <http://durka.info>

Plik (*ang. file*): uporządkowany zbiór danych. Nazwa pliku nie jest częścią pliku.

Rodzaj pliku jest zwykle zakodowany w “rozszerzeniu” – trzech literach nazwy występujących po kropce.

MS Word (edytor tekstów)	*.doc
MS Excel (arkusz kalkulacyjny)	*.xls, *.xlw
MS Access (baza danych)	*.mdb
DBase (baza danych)	*.db
Corel Draw! (program graficzny)	*.cdr
Corel PhotoPaint	*.cpt

Format danych: sposób kodowania informacji. Dla każdego typu danych (tekst, dźwięk, obraz, inne) istnieje wiele formatów danych, kodujących tę samą informację.

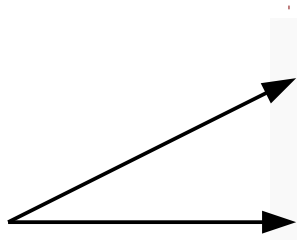
Niektóre formaty plików tekstowych:

- .txt – format podstawowy
- .xml – język przeznaczony do reprezentowania danych w strukturalizowany sposób
- .doc – format binarny kodujący formatowanie tekstu w programach MS Office
- .docx – format kodujący formatowanie tekstu w programach MS Office
Informacje o kodowaniu są zapisane w postaci pliku XML
- .odt – format kodujący używany przez program Open (Libre) Office

XML (*ang. eXtensible Markup Language*): uniwersalny język znaczników przeznaczony do reprezentowania różnych danych w strukturalizowany sposób.

znacznik

```
<?xml version="1.0" encoding="UTF-8"?>
<ksiazka-telefoniczna kategoria="bohaterowie książek">
  <!-- komentarz -->
  <osoba charakter="dobry">
    <imie>Ambroży</imie>
    <nazwisko>Kleks</nazwisko>
    <telefon>123-456-789</telefon>
  </osoba>
  <osoba charakter="zły">
    <imie>Alojzy</imie>
    <nazwisko>Bąbel</nazwisko>
    <telefon/>
  </osoba>
</ksiazka-telefoniczna>
```



DOCX: format zapisu plików tekstowych używany przez najnowsze wersje pakietu MS Office

```
[akalinow@hepzc3 TI]$ file Tekst.docx
Tekst.docx: Zip archive data, at least v2.0 to extract
[akalinow@hepzc3 TI]$ █
```


DOCX: format zapisu plików tekstowych używany przez najnowsze wersje pakietu MS Office

```
[akalinow@hepzc3 TI]$ file Tekst.docx
Tekst.docx: Zip archive data, at least v2.0 to extract
[akalinow@hepzc3 TI]$ unzip Tekst.docx
Archive:  Tekst.docx
  inflating: _rels/.rels
  inflating: word/settings.xml
  inflating: word/_rels/document.xml.rels
  inflating: word/fontTable.xml
  inflating: word/styles.xml
  inflating: word/document.xml
  inflating: docProps/app.xml
  inflating: docProps/core.xml
  inflating: [Content_Types].xml
[akalinow@hepzc3 TI]$ █
```

```
-<w:document>
- <w:body>
- <w:p>
- <w:pPr>
  <w:pStyle w:val="style0"/>
</w:pPr>
- <w:r>
  <w:rPr/>
  <w:t>Ala ma kota.</w:t>
</w:r>
</w:p>
- <w:sectPr>
  <w:type w:val="nextPage"/>
  <w:pgSz w:h="16838" w:w="11906"/>
  <w:pgMar w:bottom="1134" w:footer="0" w:gutter="0" w:header="0" w:left="1134" w:right="1134" w:top="1134"/>
  <w:pgNumType w:fmt="decimal"/>
  <w:formProt w:val="false"/>
  <w:textDirection w:val="lrTb"/>
</w:sectPr>
</w:body>
</w:document>
```

```

- <w:document>
- <w:body>
  - <w:p>
    - <w:pPr>
      <w:pStyle w:val="style0"/>
    </w:pPr>
  - <w:r>
    - <w:rPr>
      <w:b/>
      <w:bCs/>
    </w:rPr>
    <w:t xml:space="preserve">Ala </w:t>
  </w:r>
- <w:r>
  - <w:rPr>
    <w:b w:val="false"/>
    <w:bCs w:val="false"/>
  </w:rPr>
  <w:t>ma kota.</w:t>
</w:r>
</w:p>
- <w:sectPr>
  <w:type w:val="nextPage"/>
  <w:pgSz w:h="16838" w:w="11906"/>
  <w:pgMar w:bottom="1134" w:footer="0" w:gutter="0" w:header="0" w:left="1134" w:right="1134" w:top="1134"/>
  <w:pgNumType w:fmt="decimal"/>
  <w:formProt w:val="false"/>
  <w:textDirection w:val="lrTb"/>
  <w:docGrid w:charSpace="0" w:linePitch="240" w:type="default"/>
</w:sectPr>
</w:body>
</w:document>

```

Grafika rastrowa (bitmapa): reprezentacja obrazu jako dwuwymiarowej siatki punktów o różnych kolorach.

Kolor każdego piksela jest kodowany jako suma trzech podstawowych kolorów:

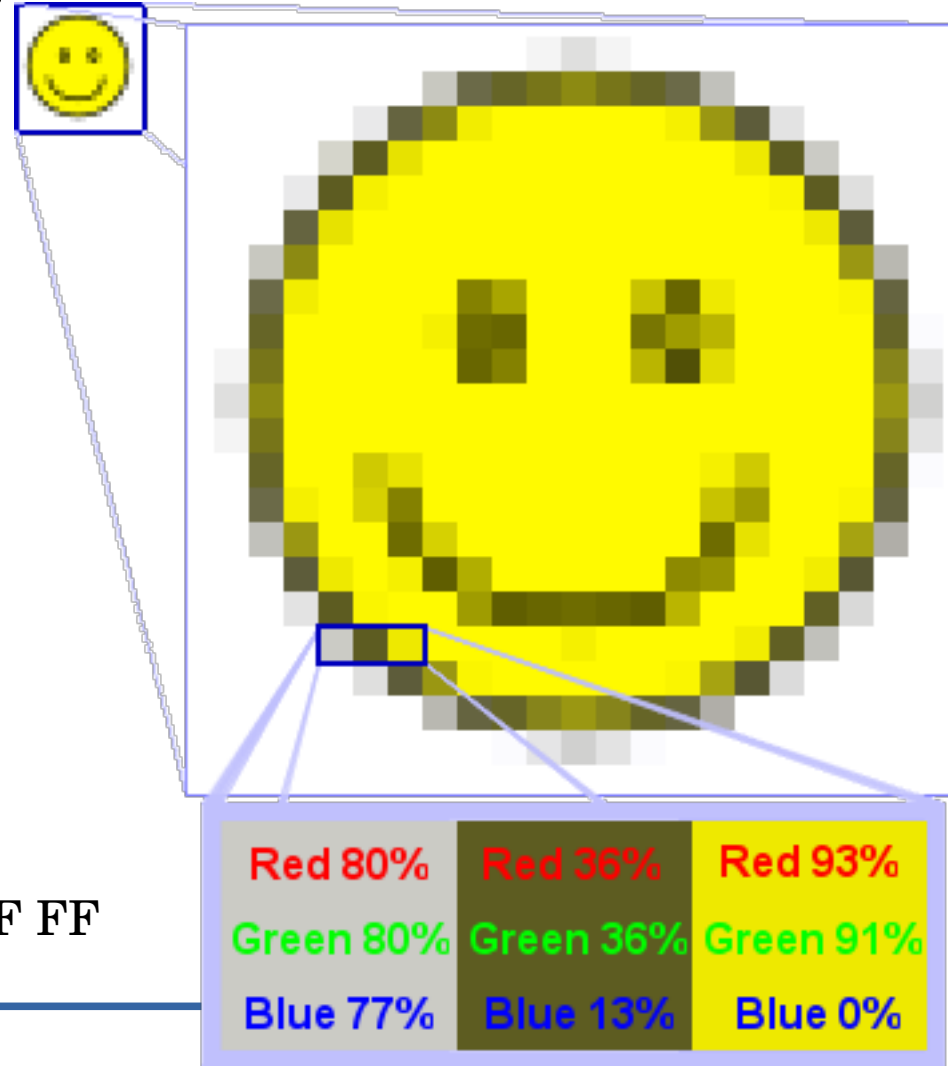
czzerwony, **niebieski**, **zielony** (RGB)

Zwykle przeznaczają się 1 bajt na poziom każdego z kolorów, co daje:

$2^8 = 256$ poziomów dla każdego z kolorów
(min – 0x0, max – 0xFF)

3x8 = 24 bity na kodowanie kolorów
24 bity koduje $2^{24} = 16777216$ kolorów

Jakie to kolory? 00 00 00 , 00 FF 00, FF FF FF



Ile miejsca zajmuje rysunek rastrowy o rozdzielczości 1280x1024?

**24 bity x 1280 x 1024 / 8 / 1024 / 1024 =
3.75 MiB**

Formaty graficzne z kompresją bezstratną:
PNG, TIFF,

Formaty graficzne z kompresją stratną:
JPEG

(kompresja JPEG najlepiej odzwierciedla tekstury,
najsłabiej krawędzie)

Przykład obok: jakość 100 – kompresja 2.6:1
jakość 50 – kompresja 15:1
jakość 1 – kompresja 144:1



By Toytoy at en.wikipedia

CC BY-SA 3.0
Wykład 2

Ile miejsca zajmuje rysunek rastrowy o rozdzielczości 1280x1024?

**24 bity x 1280 x 1024 / 8 / 1024 / 1024 =
3.75 MiB**

Formaty graficzne z kompresją bezstratną:
PNG, TIFF,

Formaty graficzne z kompresją stratną:
JPEG

(kompresja JPEG najlepiej odzwierciedla tekstury,
najsłabiej krawędzie)

Przykład obok: jakość 100 – kompresja 2.6:1
jakość 50 – kompresja 15:1
jakość 1 – kompresja 144:1



By Toytoy at en.wikipedia

CC BY-SA 3.0
Wykład 2

RAW: format zapisu zdjęć cyfrowych zachowujący pełną (surową, *ang. raw*) informację o danych uzyskanych przez matrycę światłoczułą, bez żadnych modyfikacji (i kompresji) które zwykle wykonuje oprogramowanie aparatu zapisując zdjęcie w formatach wyższego poziomu.



©Kamil Rumiński, <http://www.studio-progresiv.pl/format-raw-w-fotografii/>

Nie wszystkie aparaty cyfrowe pozwalają na zapis zdjęć w formacie RAW.