

情報科学

(1) 情報の表現

目次

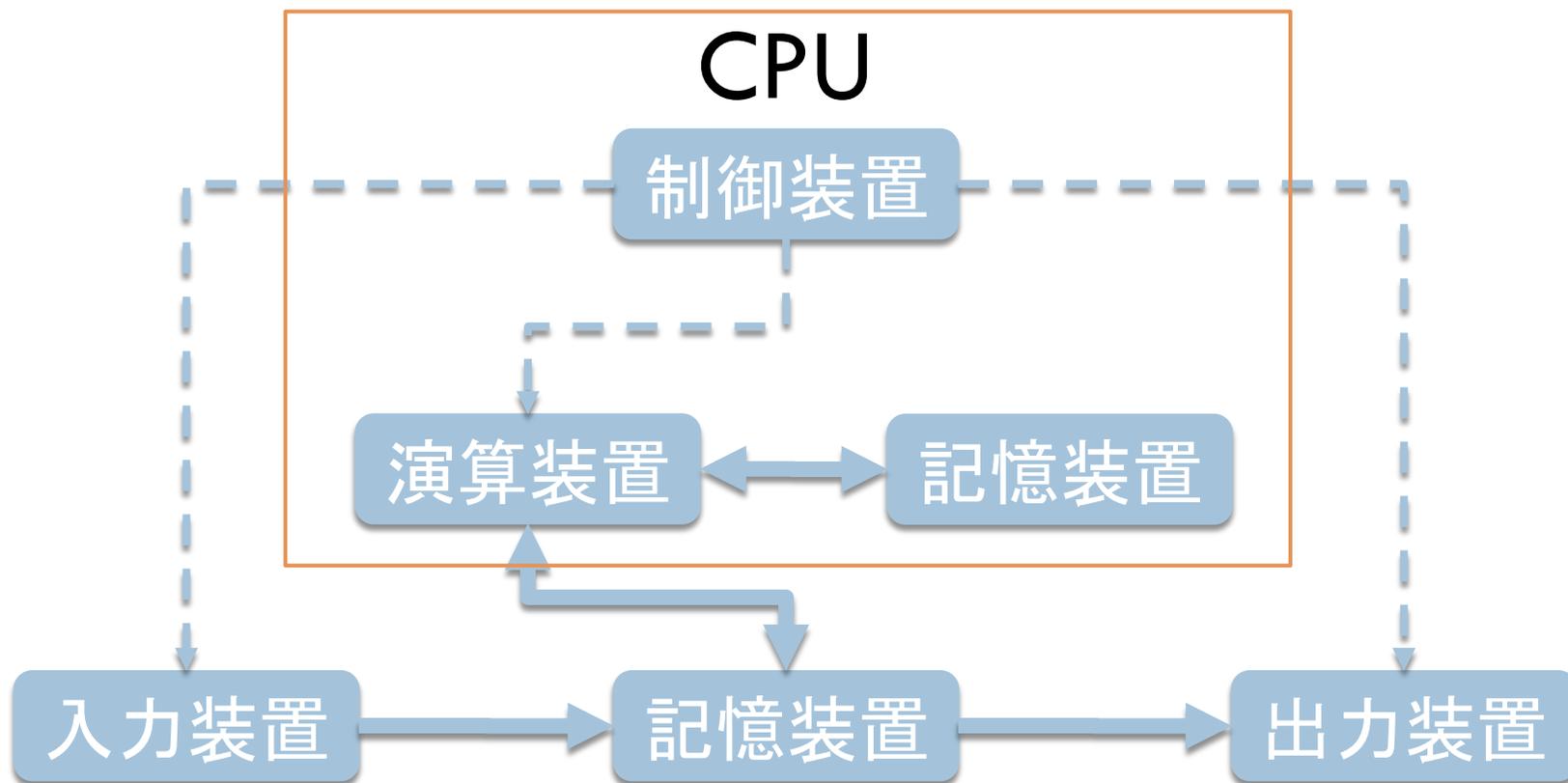


- コンピュータの5大機能
- コンピュータ内での情報表現
 - 数値の表現
 - 文字の表現
 - アナログ情報とデジタル情報

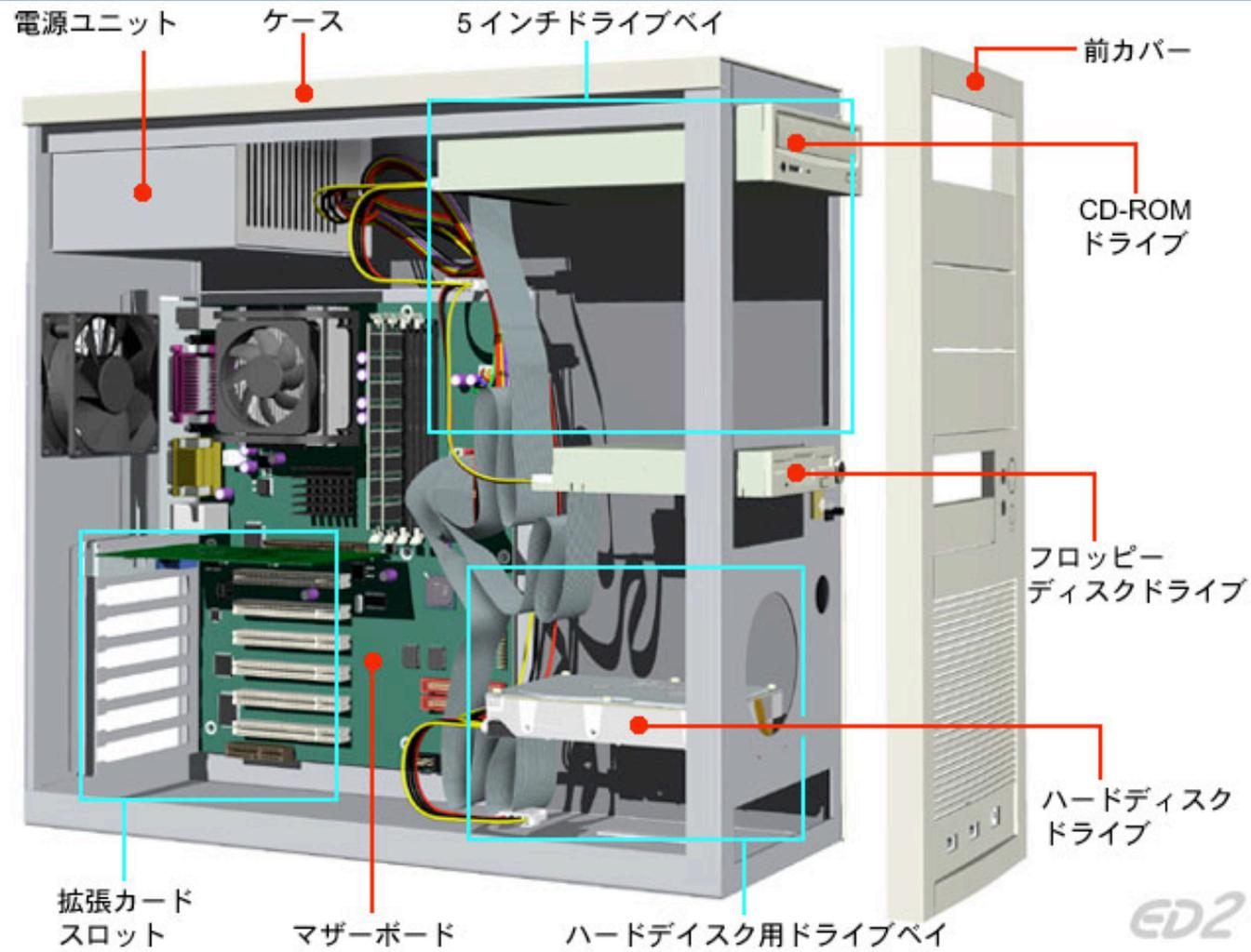
コンピュータの5大機能

- 入力: 情報を受け取る
- 出力: 情報を送り出す
- 記憶: 情報を蓄積する
- 演算: 情報を組み合わせて計算を行う
- 制御: 他の機能を制御する

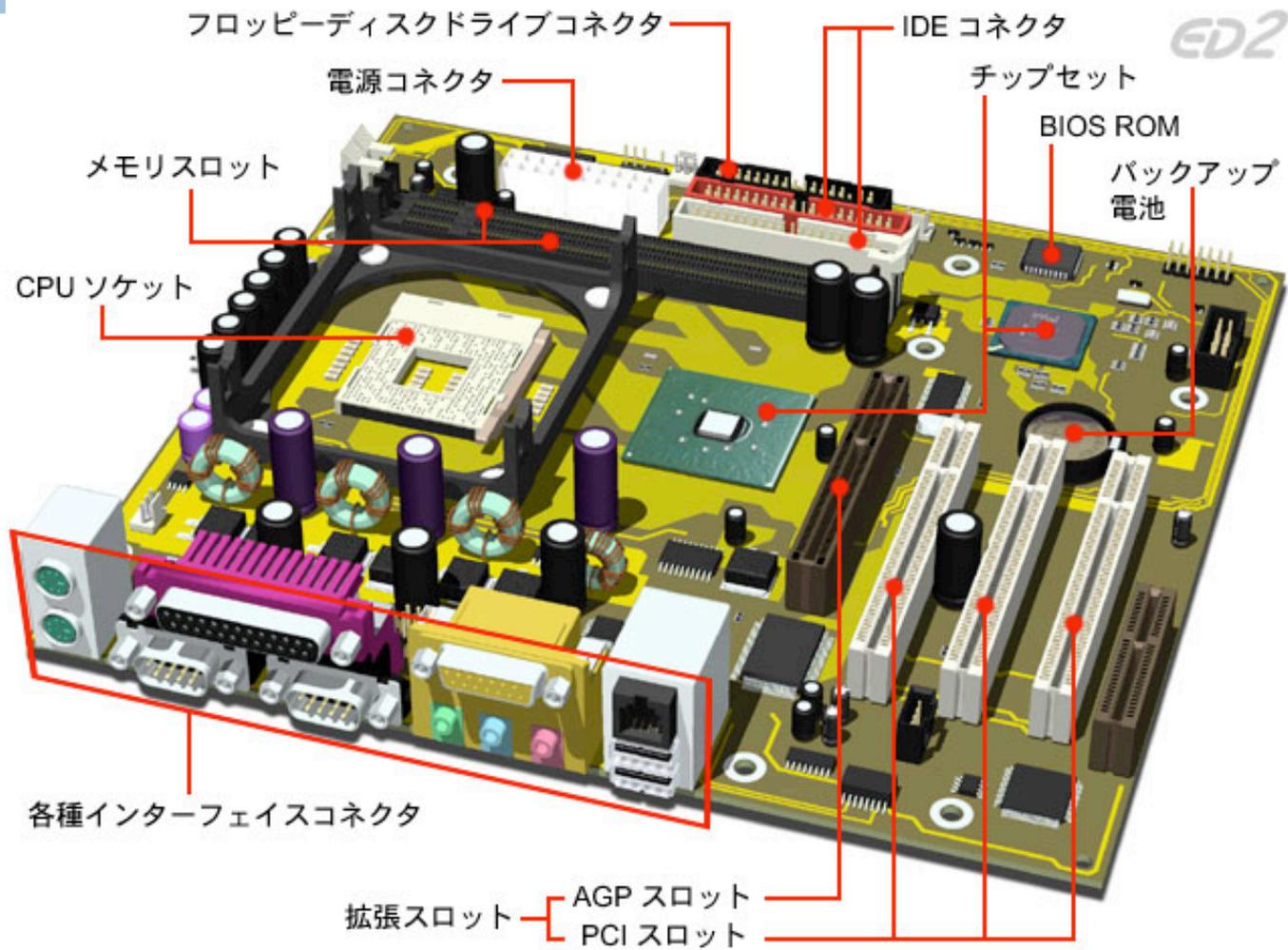
コンピュータの構成



コンピュータの内部構造



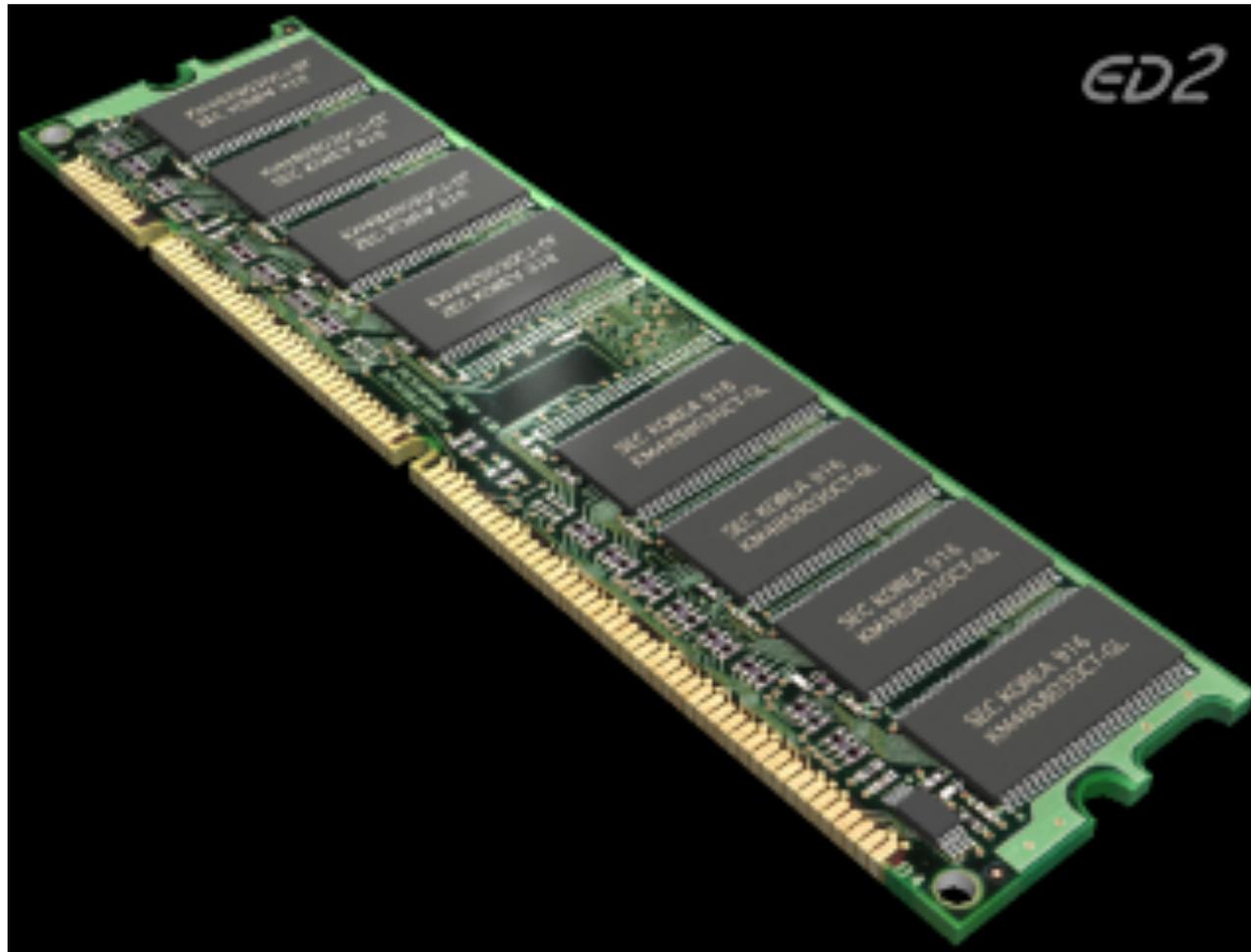
マザーボード



記憶装置

- 主記憶装置 = CPUが直接読み書きできる
- 補助記憶装置
 - 磁気ディスク(ハードディスク)
 - 光ディスク(CD, DVD, BD)

主記憶装置の構造



主記憶装置の構造

- スイッチが並んだようなもの（ONかOFF）
- 珠が1つしかないそろばんのようなもの

- 1GBのメモリ = 8,589,934,592個のスイッチ

スイッチ1つで表せる情報

- 授業があるかないか



授業あり



休講

- 勝ったか負けたか



勝った



負けた

スイッチ2つで表せる情報

□ 授業の成績

ON ON



優

ON OFF



良

OFF ON



可

OFF OFF



不可

情報量の単位 (bit, byte)

- 1ビット=1個のスイッチで表現できる量=2状態
- Nビット=N個のスイッチで表現できる量= 2^N 状態
- 8ビット=1バイト(byte) (1Bとも表現)
- $2^{10}B \doteq 1000B = 1KB$
- $2^{20}B \doteq 1000KB = 1MB$
- $2^{30}B \doteq 1000MB = 1GB$
- $2^{40}B \doteq 1000GB = 1TB$

なんで1バイトは8ビット？

- コンピュータにとって単位は 2^N が便利である
(人間にとっては 10^N が便利)
- はじめに普及したのが8ビットを1単位として扱うコンピュータだった
- 最近のコンピュータは32ビットか64ビットを1単位としているものが多い

コンピュータ内での情報の表現

- すべての情報をON/OFFの組み合わせで表現
 - 数値: 2進数
 - 文字: 文字コード表
 - 音声・画像・動画:
アナログ情報をデジタル情報に変換(量子化)

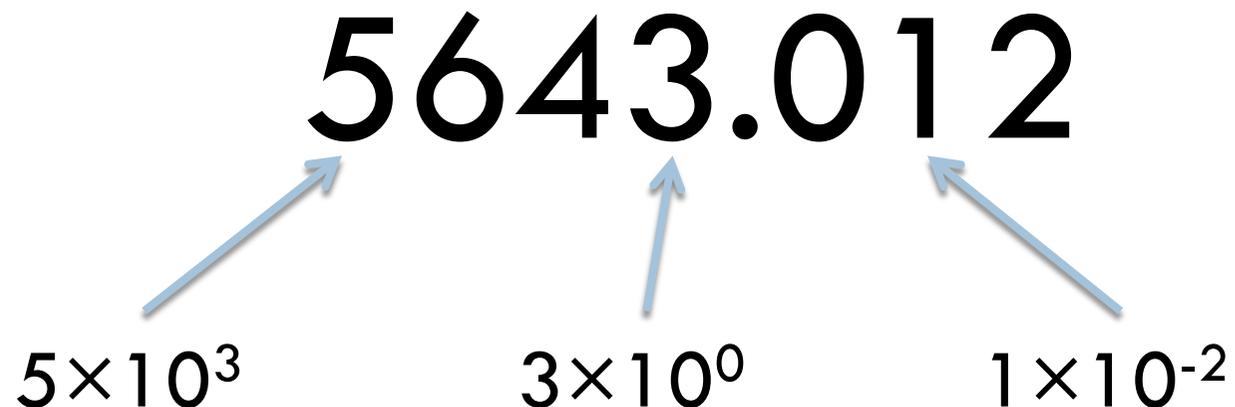
数値の表現

片手でいくつまで数えられる？

- 5だと思った人はスイッチの話を思いだそう
- 指がつりそうになったら無理しない
- 失礼なポーズもあるけど今日は気にしない

数値の表現(10進数)

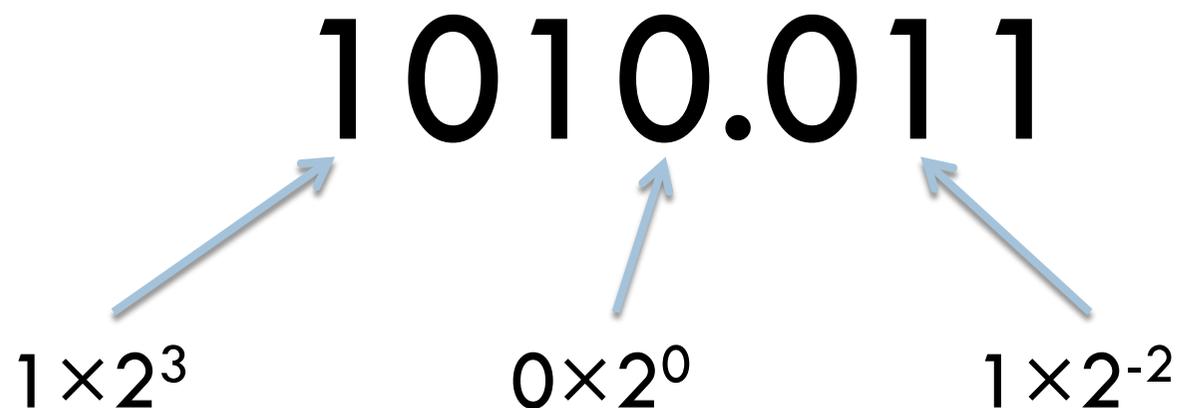
- 10進数: 各桁は0~9の10状態を表現



N桁目の重みは 10^N

数値の表現(2進数)

- 各桁は0か1の2状態を表現



N桁目の重みは 2^N

10進数から2進数への変換

- 整数部分: 2で割った余りを並べる
- 小数部分: 2を掛けて整数部分を並べる
- 例) 10進数 $6.375 = 2$ 進数 110.011

$$\begin{array}{l} 6 \div 2 = 3 \quad \text{あまり} \quad 0 \\ \leftarrow \\ 3 \div 2 = 1 \quad \text{あまり} \quad 1 \\ \leftarrow \\ 1 \div 2 = 0 \quad \text{あまり} \quad 1 \end{array}$$

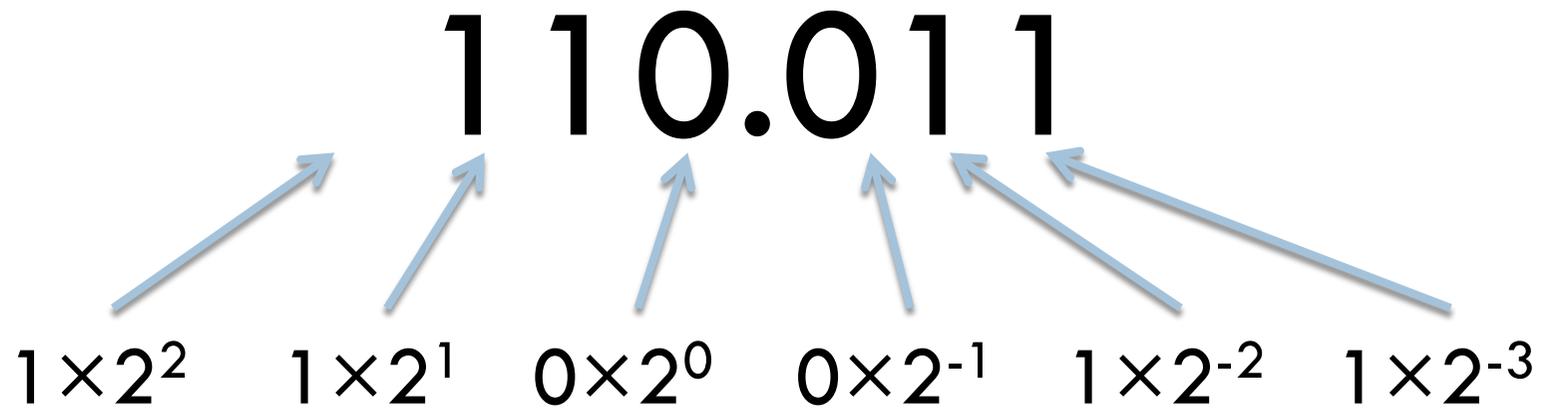
整数部分 = 110

$$\begin{array}{l} 0.375 \times 2 = 0.75 \\ \leftarrow \\ 0.75 \times 2 = 1.5 \\ \leftarrow \\ 0.5 \times 2 = 1 \end{array}$$

小数部分 = 0.11

2進数から10進数への変換

□ 例) 2進数 110.011 = 10進数 6.375


$$110.011$$
$$1 \times 2^2 \quad 1 \times 2^1 \quad 0 \times 2^0 \quad 0 \times 2^{-1} \quad 1 \times 2^{-2} \quad 1 \times 2^{-3}$$

$$4 + 2 + 0 + 0 + 0.25 + 0.125 = 6.375$$

16進数

- 各桁は0~9、A~Fの16状態
- N桁目の重みは 16^N
- 例
 - 2進数 0101 = 16進数 5 = 10進数 5
 - 2進数 1100 = 16進数 C = 10進数 12
 - 2進数 0101 1100 = 16進数 5C = 10進数 92
- 10進数 \leftrightarrow 2進数よりも変換するのが楽
- 8ビットがちょうど2文字で表せる

負の数の表現

- 方法1: 最上位で符号を表す(0=正、1=負)
 - 例) $5 = 00101$, $-5 = 11011$
- 方法2: 2の補数
 - 正の整数A に足した時、桁が1つ増える数のうち最も小さい数
 - 例) 0101 の「2の補数」は 1011

$$\begin{array}{r} 00101 \\ +11011 \\ \hline 100000 \end{array}$$

「2の補数」の求め方

- 各桁の0と1を反転させて1を加える
 - 例) 0101の「2の補数」を求める

0101 \longrightarrow 1010 \longrightarrow 1011
反転 1を加える

コンピュータにおける数値表現 (自然数)

- 8ビット(1バイト)単位の固定長
 - ▣ 8ビット 0~255
 - ▣ 16ビット 0~65,535
 - ▣ 32ビット 0~4,294,967,295
- 負の数も表したい場合(符号あり)
 - ▣ 8ビット -128~127
 - ▣ 16ビット -32768~32767

実数の表現（浮動小数点表現）

- 浮動小数点表現 $A = m \times b^e$
 - m: 仮数部、b: 基数部、e: 指数部
- とても大きな数や小さな数を少ないビット数で表現
- 同じ数を複数の表し方ができてしまう
 - 例) 123000の浮動小数点表現
 - 123×10^3
 - 12.3×10^4
 - 1.23×10^5
 - 0.123×10^6

IEEE 754 形式

- 表し方が同じになるように決めたルール
- 基数は2
- 単精度(32ビット)
符号1ビット、指数部8ビット、仮数部:23ビット
- 倍精度(64ビット)
符号1ビット、指数部11ビット、仮数部:52ビット

誤差

- 桁あふれ誤差: データの桁数超過による誤差
 - オーバーフロー: 最大値を超える
 - アンダーフロー: 最小値を下回る
- 丸め誤差: 数値を有効桁数に納めるために発生する誤差
 - 切り捨て, 切り上げ, 四捨五入によって発生
- 打ち切り誤差: 浮動小数点の演算を打ち切ることで発生
 - 例) 円周率

誤差(つづき)

- 桁落ち誤差: 演算途中の有効桁数の不足による誤差
 - 絶対値のほぼ等しい数の演算で発生
- 情報落ち: 大きい値と小さい値の演算で小さい値が消えること

〇〇年問題？



- 2000年問題

- 2036年問題

- 2038年問題

文字の表現

コンピュータにおける文字の表現

文字のデジタル化の例

JIS X 0201コード表に基づいて、1文字ずつ文字コードに置き換える。

This is a pen.

文字コード
(10進数表記)

84 104 105 83 32 105 83 32 97 32 112 101 110 46

文字列データ

84,104,105,83,32,105,83,32,97,32,112,101,110,46

2進数の文字列データ

1010100,1101000,1101001,1010011,0100000,1101001,1010011,
0100000,1100001,0100000,1110000,1100101,1101110,0101110

パソコンに記憶



文字コード表の例

JIS X 0201コード表

※赤の数字が10進数表記の文字コード
緑の数字が16進数表記の文字コード

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0	NUL 0 00	TC1 1 01	TC2 2 02	TC3 3 03	TC4 4 04	TC5 5 05	TC6 6 06	BEL 7 07	FE0 8 08	FE1 9 09	FE2 10 0A	FE3 11 0B	FE4 12 0C	FE5 13 0D	SO 14 0E	SI 15 0F
1	TC7 16 10	DC1 17 11	DC2 18 12	DC3 19 13	DC4 20 14	TC8 21 15	TC9 22 16	TC10 23 17	CAN 24 18	EM 25 19	SUB 26 1A	ESC 27 1B	IS4 28 1C	IS3 29 1D	IS2 30 1E	IS1 31 1F
2	空白 32 20	! 33 21	" 34 22	# 35 23	\$ 36 24	% 37 25	& 38 26	' 39 27	(40 28) 41 29	* 42 2A	+ 43 2B	, 44 2C	- 45 2D	. 46 2E	/ 47 2F
3	0 48 30	1 49 31	2 50 32	3 51 33	4 52 34	5 53 35	6 54 36	7 55 37	8 56 38	9 57 39	: 58 3A	; 59 3B	< 60 3C	= 61 3D	> 62 3E	? 63 3F
4	@ 64 40	A 65 41	B 66 42	C 67 43	D 68 44	E 69 45	F 70 46	G 71 47	H 72 48	I 73 49	J 74 4A	K 75 4B	L 76 4C	M 77 4D	N 78 4E	O 79 4F
5	P 80 50	Q 81 51	R 82 52	S 83 53	T 84 54	U 85 55	V 86 56	W 87 57	X 88 58	Y 89 59	Z 90 5A	[91 5B	¥ 92 5C] 93 5D	^ 94 5E	_ 95 5F
6	` 96 60	a 97 61	b 98 62	c 99 63	d 100 64	e 101 65	f 102 66	g 103 67	h 104 68	i 105 69	j 106 6A	k 107 6B	l 108 6C	m 109 6D	n 110 6E	o 111 6F
7	p 112 70	q 113 71	r 114 72	s 115 73	t 116 74	u 117 75	v 118 76	w 119 77	x 120 78	y 121 79	z 122 7A	{ 123 7B	 124 7C	} 125 7D	~ 126 7E	DEL 127 7F
8	128 80	129 81	130 82	131 83	132 84	133 85	134 86	135 87	136 88	137 89	138 8A	139 8B	140 8C	141 8D	142 8E	143 8F
9	144 90	145 91	146 92	147 93	148 94	149 95	150 96	151 97	152 98	153 99	154 9A	155 9B	156 9C	157 9D	158 9E	159 9F
A	160 A0	161 A1	162 A2	163 A3	164 A4	165 A5	166 A6	167 A7	168 A8	169 A9	170 AA	171 AB	172 AC	173 AD	174 AE	175 AF
B	176 B0	177 B1	178 B2	179 B3	180 B4	181 B5	182 B6	183 B7	184 B8	185 B9	186 BA	187 BB	188 BC	189 BD	190 BE	191 BF
C	192 C0	193 C1	194 C2	195 C3	196 C4	197 C5	198 C6	199 C7	200 C8	201 C9	202 CA	203 CB	204 CC	205 CD	206 CE	207 CF
D	208 D0	209 D1	210 D2	211 D3	212 D4	213 D5	214 D6	215 D7	216 D8	217 D9	218 DA	219 DB	220 DC	221 DD	222 DE	223 DF
E	224 E0	225 E1	226 E2	227 E3	228 E4	229 E5	230 E6	231 E7	232 E8	233 E9	234 EA	235 EB	236 EC	237 ED	238 EE	239 EF
F	240 F0	241 F1	242 F2	243 F3	244 F4	245 F5	246 F6	247 F7	248 F8	249 F9	250 FA	251 FB	252 FC	253 FD	254 FE	255 FF

上位4ビット
下位4ビット

制御コード

ASCIIコードに準拠した部分
(¥はASCIIコードでは \)

未使用

JISで拡張された半角カタカナ部分

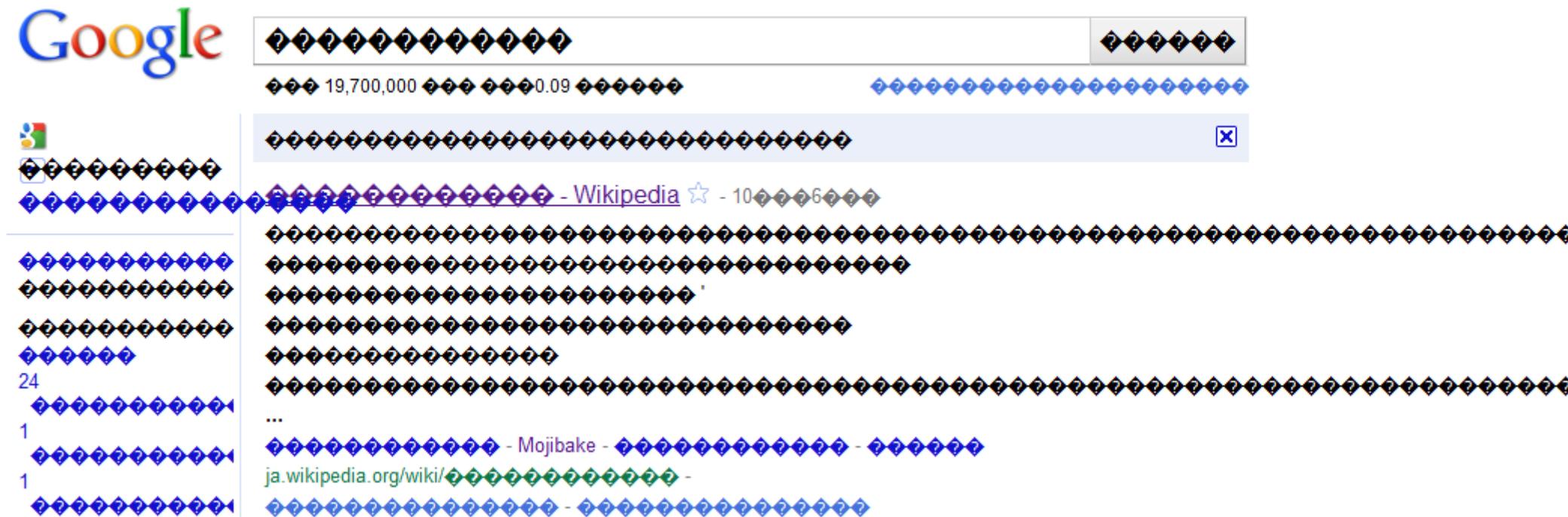
未使用

いろいろな文字コード

- ASCIIコード: 1文字を7ビットで表現
 - A~Z, a~z, 0~9
 - 記号
 - 制御文字 (改行など)
- 日本語
 - JIS, Shift-JIS, EUC-JP など
- Unicode: 世界中の文字を1つの表で表現
 - UTF-8, UTF-16 など

文字化け

- 使う文字コードを間違えているときに起きる問題
 - 「エンコード」で正しい文字コードを指定するとなおる

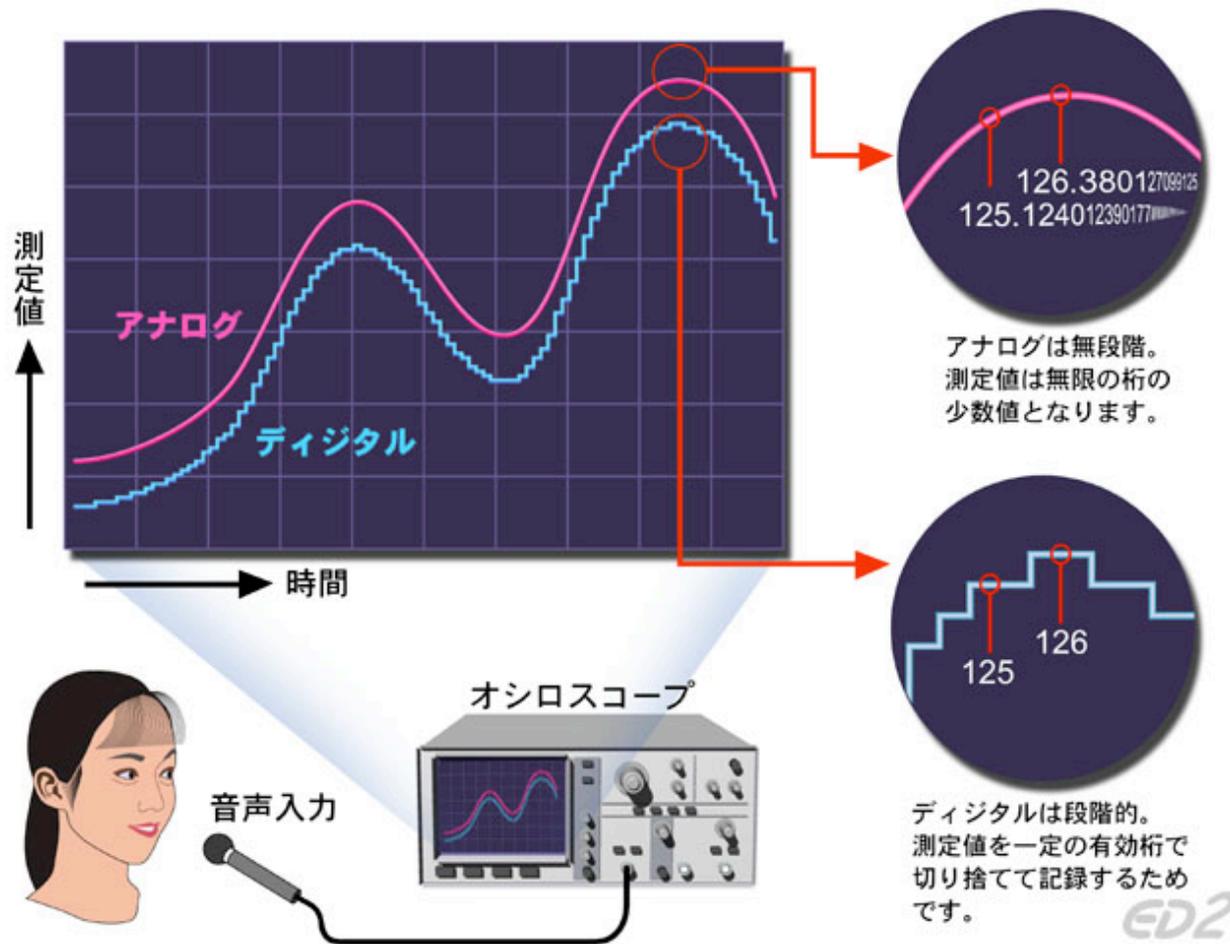


アナログ情報とデジタル情報

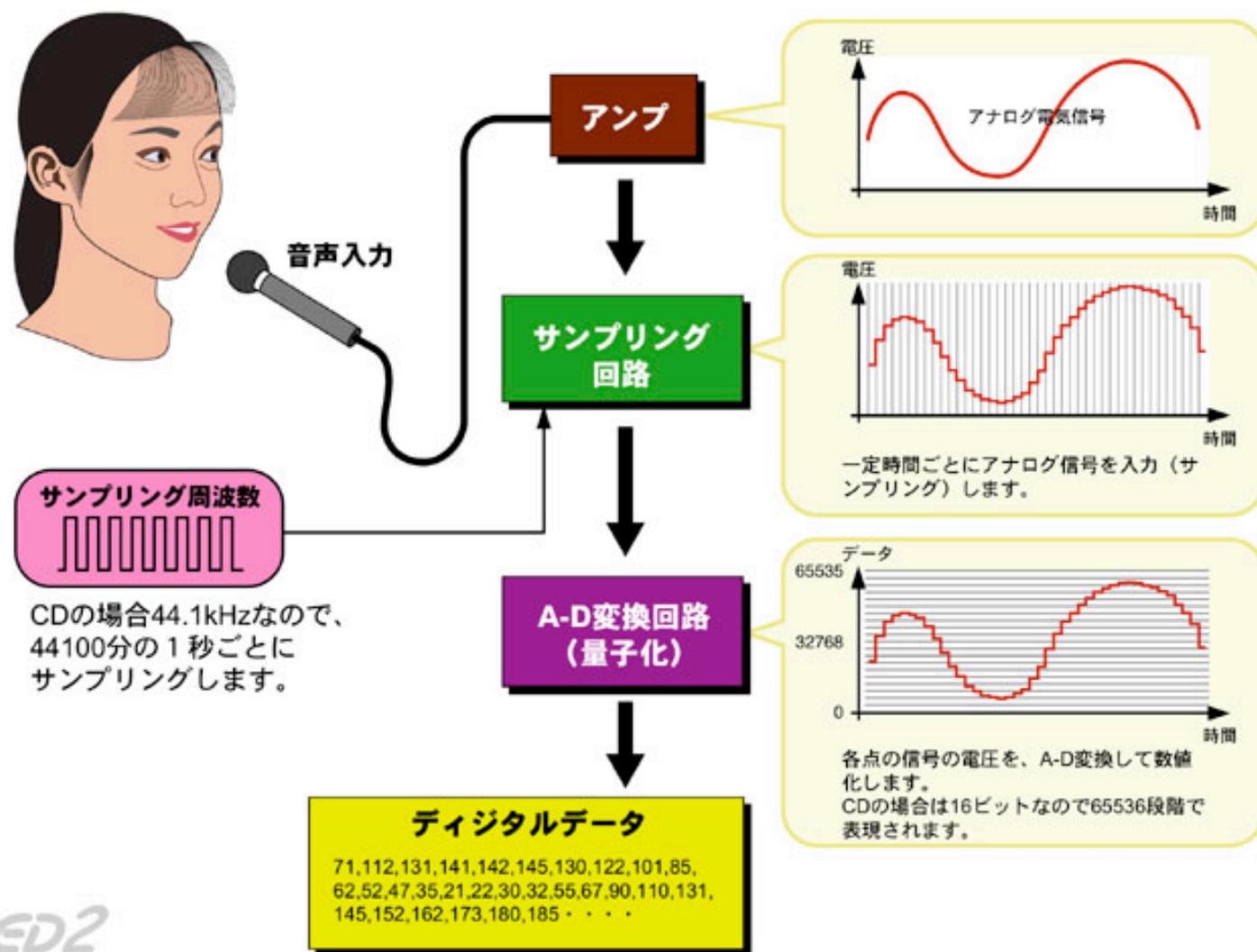
アナログ情報とデジタル情報

- アナログ：連続量を連続量で表現
 - 温度計：気温を体積で表現
 - スピードメーター：車の速度を角度で表現
- デジタル：連続量を離散的な値として表現
 - 気温や速度を数値で表現

情報のデジタル化

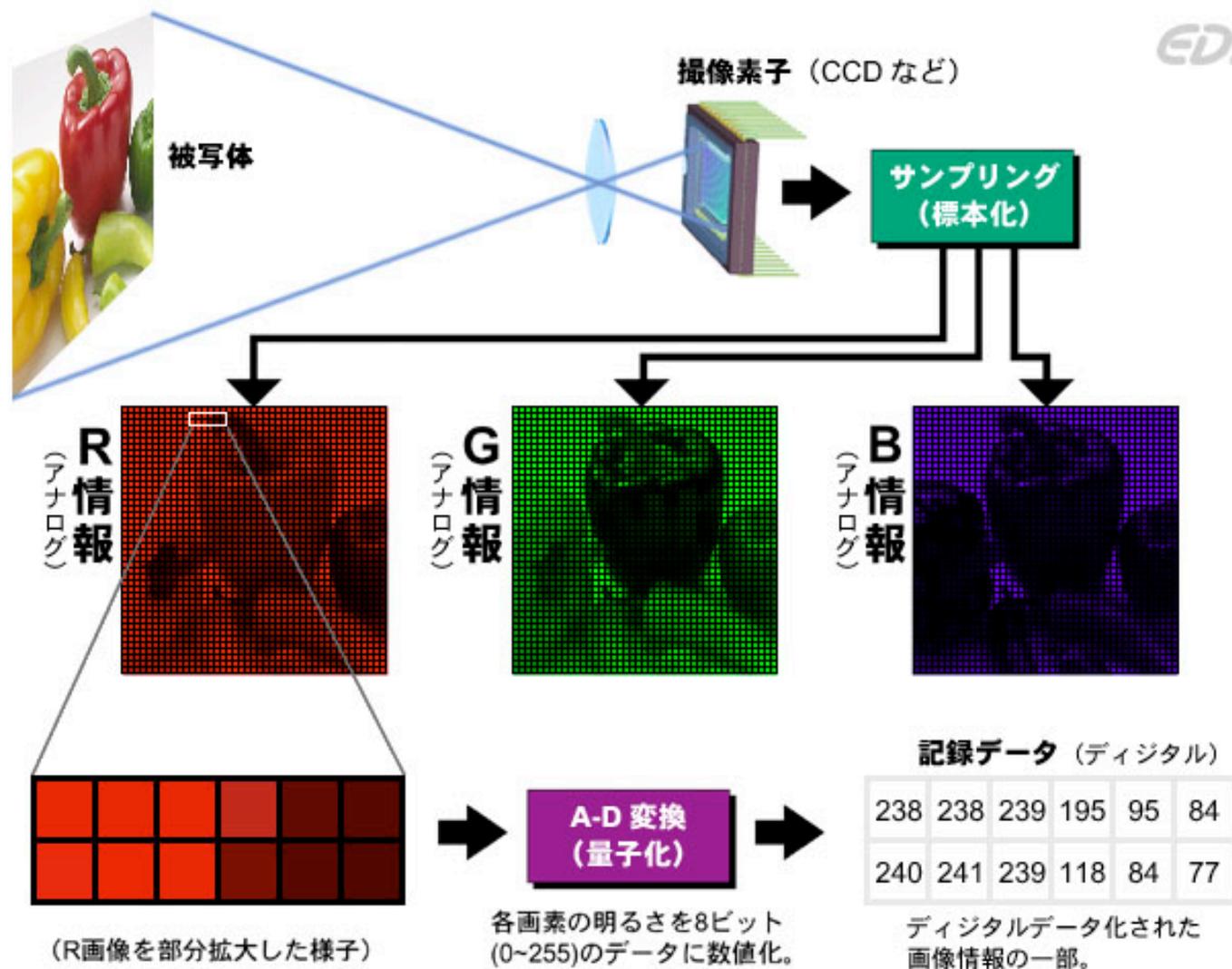


音声情報



画像情報

ED2



デジタル化に伴う誤差

- 元の情報との間に誤差が生じる(量子化誤差)
- 解像度・周波数・ビット数が大きいほど誤差は少なくなる
 - 地上波デジタル 1440 x 1080
 - BSデジタル 1920 x 1080

アナログ vs. デジタル

- デジタルの欠点
 - 量子化誤差がある
- デジタルのメリット
 - コンピュータで扱える
 - 減衰・雑音に強い
- 最近ではデジタルが主流
 - 音楽はレコードからCDに
 - テレビは地上波デジタル放送に