Data & Memory

Part 1

1

Binary Numbers

Bit of This and a Bit of That

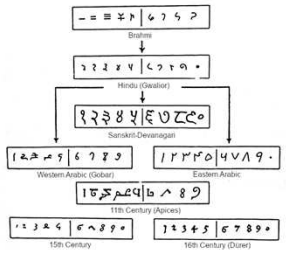
2

What is a Number?

- *Hindu-Arabic Number System*
  - positional grouping system
  - each position represents an increasing power of 10
  - used throughout the World
- Binary numbers
  - based on the same system
  - use powers of **2** rather than 10

3

Evolution of a Genius System



4

Base 10 Number

The number 1783 is ...

10^4	10^3	10^2	10^1	10^0
10000	1000	100	10	1
0	**1**	**7**	**8**	**3**

1000 + 700 + 80 + 3 = 1783

5

Binary Number Example

The number 0100 1010 is ...

2^7	2^6	2^5	2^4	2^3	2^2	2^1	2^0
128	64	32	16	8	4	2	1
0	**1**	**0**	**0**	**1**	**0**	**1**	**0**

64 + 8 + 2 = 74

6

Binary Number Example

The number 1101 1011 is ...

2^7	2^6	2^5	2^4	2^3	2^2	2^1	2^0
128	64	32	16	8	4	2	1
1	**1**	**0**	**1**	**1**	**0**	**1**	**1**

128 + 64 + 16 + 8 + 2 + 1 = 219

7

Hexadecimal Numbers

- Writing out long binary numbers is cumbersome and error prone
- As a result, computer scientists often write computer numbers in hexadecimal
- Hexadecimal is base-16
  - we only have 0 ... 9 to represent digits
  - So, hex uses **A ... F** to represent **10 ... 15**

8

Hexadecimal Numbers

Hex	Decimal	Binary
0	0	0000
1	1	0001
2	2	0010
3	3	0011
4	4	0100
5	5	0101
6	6	0110
7	7	0111

Hex	Decimal	Binary
8	8	1000
9	9	1001
A	10	1010
B	11	1011
C	12	1100
D	13	1101
E	14	1110
F	15	1111

9

Hex Example

The number 7AC is ...

16^4	16^3	16^2	16^1	16^0
65536	4096	256	16	1
0	**0**	**7**	**A**	**C**

(7×256) + (10×16) + (12×1) = 1964

10

Converting Binary to Hex = Easy

- Since $16^1 = 2^4$, a single hex character can represent a total of 4 bits
- Convert every 4-bits to a single hexadecimal digit

A				**7**			
1	0	1	0	0	1	1	1

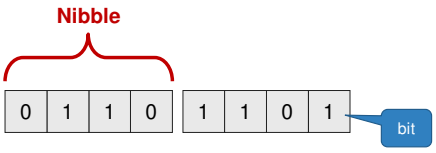
11

Bits and Bytes

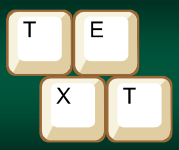
- Everything in a *modern* computer is stored using combination of ones and zeros
- *Bit* is one binary digit
  - either 1 or 0
  - shorthand for a bit is b
- *Byte* is a group of 8 bits
  - e.g. **1101 0100**
  - shorthand for a byte is B

12

The Byte



13



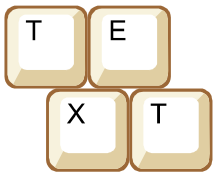
Text in Programming Languages

Press Any Key to Continue

14

How Text Is Stored

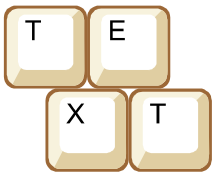
- Computer often store and transmit textual data
- Examples:
  - punctuation
  - numerals 0 – 9
  - letter
- Each of these symbols is called a *character*



15

Characters

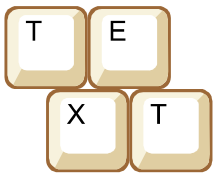
- Processors rarely know what a "character" is, and instead store each as an integer
- In this case, each character is given a unique value
- For instance
  - "A", could have the value of 1
  - "B" is 2
  - "C" is 3, etc…



16

Characters

- Characters and their matching values are a *character set*
- There have been many characters sets developed over time



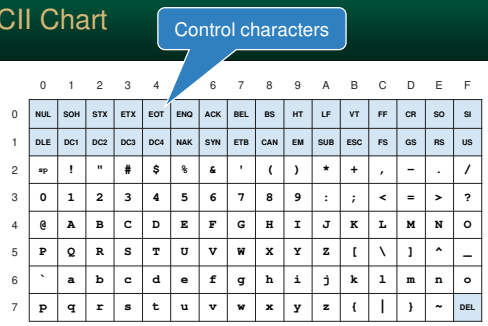
17

Character Sets

- ASCII
  - 7 bits – 128 characters
  - uses a full byte, one bit is not used
  - created in the 1967
- EBCDIC
  - Alternative system used by old IBM systems
  - Not used much anymore

18

ASCII Chart



19

ASCII Codes

- Each character has a unique value
- The following is how "OMG" is stored in ASCII

	Decimal	Hex	Binary
O	79	4F	0100 1111
M	77	4D	0100 1101
G	71	47	0100 0111

20

ASCII Codes

- ASCII is laid out very logically
- Alphabetic characters (uppercase and lowercase) are 32 "code points" apart

	Decimal	Hex	Binary
A	65	41	01000001
a	97	61	01100001

21

ASCII Codes

- $32^1 = 2^5$
- 1-bit difference between upper and lowercase letters
- Printers can easily convert between the two

	Decimal	Hex	Binary
A	65	41	01000001
a	97	61	01100001

22

ASCII: Number Characters

- ASCII code for 0 is 30h
- Notice that the actual value of a number character is stored in the lower nibble
- So, the characters 0 to 9 can be easily converted to their binary values

0	0011 0000
1	0011 0001
2	0011 0010
3	0011 0011
4	0011 0100
5	0011 0101
6	0011 0110
7	0011 0111
8	0011 1000
9	0011 1001

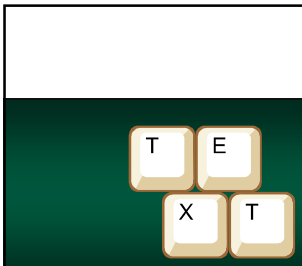
23

ASCII: Number Characters

- Character → Binary
  - clear the upper nibble
  - Bitwise And: 0000 1111
- Binary → Character
  - set the upper nibble to 0011
  - Bitwise Or: 0011 0000

0	0011 0000
1	0011 0001
2	0011 0010
3	0011 0011
4	0011 0100
5	0011 0101
6	0011 0110
7	0011 0111
8	0011 1000
9	0011 1001

24



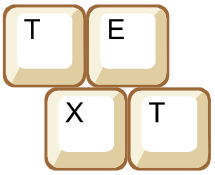
Evolution of ASCII

From ASCII to Unicode

25

Evolution of ASCII

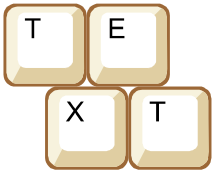
- The ASCII Character Set has gone through a few minor revisions this it was first created
- But, the 1967 standard is still used today



26

Evolution of ASCII

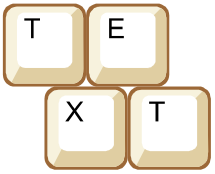
- Back in the 1960's – there were no computer monitors
- Nor did you enter programs using a keyboard
- Instead…
  - programs were inputted using punched cards or tape
  - all output was printed



27

Evolution of ASCII

- The initial 1963 version didn't include lowercase letters
- As shocking as it sounds, there was a time where lowercase letters were considered obsolete
- All-caps was called "modern"



28

Original 1963 ASCII Proposal

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0	NUL	SOM	EOA	EOM	EOT	WRU	RU	BEL	FE0	HT	LF	VT	FF	CR	SO	SI
1	DC0	DC1	DC2	DC3	DC4	ERR	SYN	LEM	S0	S1	S2	S3	S4	S5	S6	S7
2	sp	!	"	#	$	%	&	'	(	)	*	+	,	–	.	/
3	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
4	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
5	P	Q	R	S	T	U	V	W	X	Y	Z	[	\	]	↑	←
6																
7													ACK		ESC	DEL

29

1967 Revision Added Lowercase

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
Changed →	NUL	SOH	STX	ETX	EOT	ENQ	ACK	BEL	BS	HT	LF	VT	FF	CR	SO	SI
1	DLE	DC1	DC2	DC3	DC4	NAK	SYN	ETB	CAN	EM	SUB	ESC	FS	GS	RS	US
2	sp	!	"	#	$	%	&	'	(	)	*	+	,	–	.	/
3	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
4	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
5	P	Q	R	S	T	U	V	W	X	Y	Z	[	\	]	^	_
New → 6	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
7	p	q	r	s	t	u	v	w	x	y	z	{	|	}	~	DEL

30

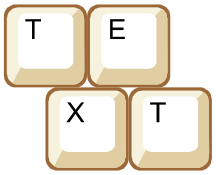
Control Characters

- Many of the control characters were designed for transferring data (SOH, STX, FS, GS, etc…)
- Other control characters we designed for printing

31

Evolution of ASCII

- Also note that very useful characters like "→" were ultimately removed from the 1963 design
- And a new character, Backspace (BS), was added
- Why?



32

Backspace Control Character

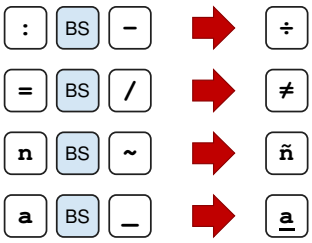
- Printers, at the time, were basically classic typewriters
- The backspace character (BS) was used to print 2+ symbols on top of each other
- This would essentially create a new character on the paper

33

IBM Ball-Printers

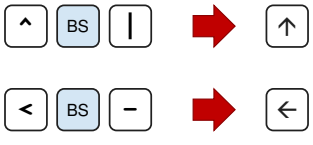
34

Power of the Backspace



35

Recreating the Removed Arrows



36

DEL Control Character

- You might have noticed an odd control character located at 7F
- This is the "Deleted" control character and was used with punched cards and tape

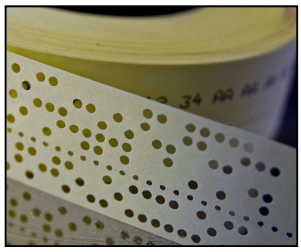
37

ASCII Chart

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0	NUL	SOH	STX	ETX	EOT	ENQ	ACK	BEL	BS	HT	LF	VT	FF	CR	SO	SI
1	DLE	DC1	DC2	DC3	DC4	NAK	SYN	ETB	CAN	EM	SUB	ESC	FS	GS	RS	US
2	sp	!	"	#	$	%	&	'	(	)	*	+	,	-	.	/
3	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
4	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
5	P	Q	R	S	T	U	V	W	X	Y	Z	[	\	]	^	_
6	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
7	p	q	r	s	t	u	v	w	x	y	z	{	|	}	~	DEL

38

Punched Tape



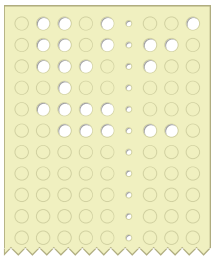
39

Bill Gates – Original Altair BASIC

Intel 8080 machine code

40

Example Punched Tape: `int x;`



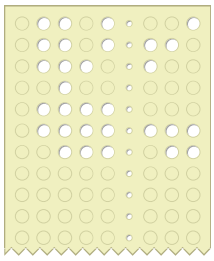
```
i
n
t

x
>
```

Student punched the wrong hole! They meant `;`

Do they have to punch an entire new tape?

41

Example Punched Tape: `int x;`



```
i
n
t

x
DEL
;
```

Just punch all remaining holes. It becomes DEL and is ignored by the card or tape reader.

42

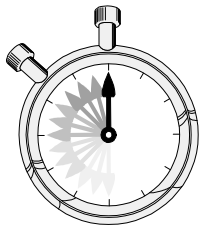
Times have changed

A New Character Set Enters the Scene

43

Times have changed…

- Computers have changed quite a bit since the 1960's
- As a result, most of these clever control characters are no longer needed
- Backspace, DEL, and numerous others are obsolete



44

Only Control Characters Still Used

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0	NUL	SOH	STX	ETX	EOT	ENQ	ACK	BEL	BS	HT	LF	VT	FF	CR	SO	SI
1	DLE	DC1	DC2	DC3	DC4	NAK	SYN	ETB	CAN	EM	SUB	ESC	FS	GS	RS	US
2	sp	!	"	#	$	%	&	'	(	)	*	+	,	-	.	/
3	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
4	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
5	P	Q	R	S	T	U	V	W	X	Y	Z	[	\	]	^	_
6	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
7	p	q	r	s	t	u	v	w	x	y	z	{	|	}	~	DEL

45

Unicode Character Set

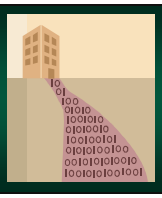
- ASCII is only good for the United States
  - Other languages need additional characters
  - Multiple competing character sets were created
- Unicode was created to support every spoken language
- Developed in Mountain View, California

46

Unicode Character Set

- Originally used 16 bits
  - that's over 65,000 characters!
  - includes every character used in the World
- Expanded to 21 bits
  - 2 million characters!
  - now supports every character ever created
  - … and emojis
- Unicode can be stored in different formats

47



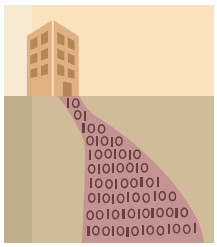
Computer Memory

Its… um…. I forgot….

48

Computer Memory

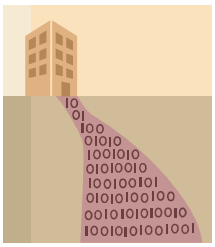
- Programs access and manipulate memory far more than you realize
- So, understanding it…
  - is vital to becoming a great assembly programmer
  - and understanding computer architecture



49

What is Memory?

- Memory is essentially an enormous array
- It is also, sometimes, referred to as *storage*
- It stores both running programs and their related data



50

Memory Addresses

- Memory is divided into a storage locations that can hold 1 byte (8 bits)
- Each an be accessed using an *address*
  - unique value that refers to that specific byte
  - used to locate the exact byte the processor wants

Memory	
0	01000100
1	01000011
2	01101111
3	01101111
4	01101011

51

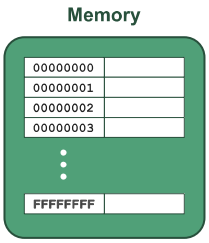
What is Memory?

- Each address is conceptually the same as an "index" in arrays
- … and you will write access memory as would an array

Memory	
0	01000100
1	01000011
2	01101111
3	01101111
4	01101011

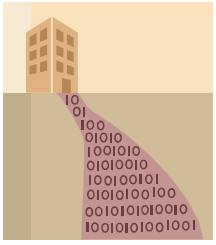
52

Memory is a Hardware Array



53

Memory Contains Data & Programs



- Data and programs are just binary numbers (stored in a series of bytes)
- …and are stored together
- Appreciating this is vital to understanding computer architecture

54

Memory Contains Data & Programs

55