



CS 208

M, 22 Jan 2024



I want to store  
`int x = 259;`  
 in memory.

0x103

An int is 4 bytes  
 0x00 00 01 03

"big  
 endian"

"little endian"

"Indian" - Gulliver's Travels

"On Holy Wars & a Plea  
For Peace"

# Codepoints

U+0041



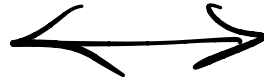
A

U+03B1



α

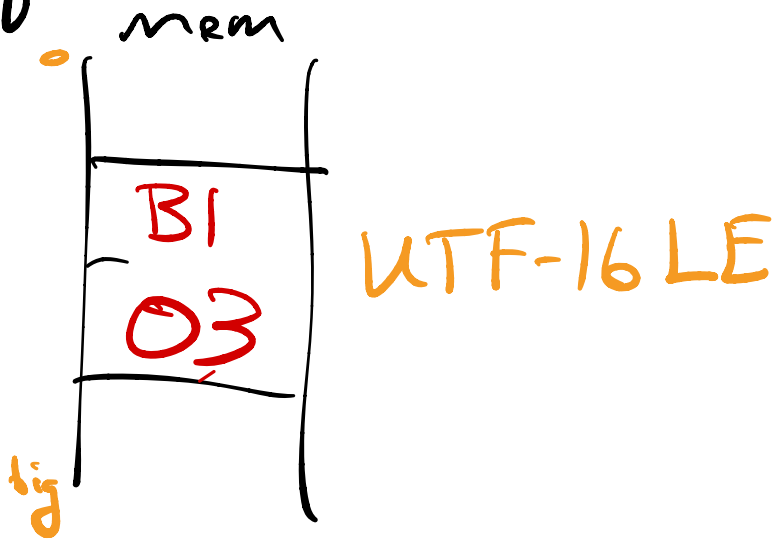
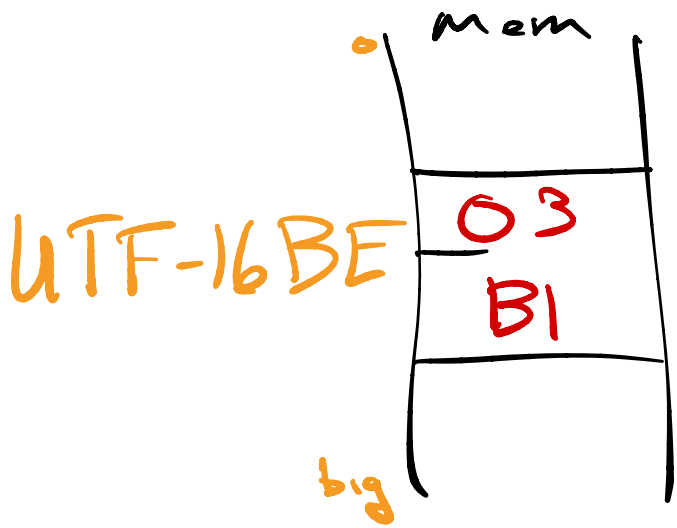
U+1F600



Wanna store characters in memory

UTF-32

Obvious techniques



## Questions:

- What about bigger codepoints?

U+01F600

- If my text is ASCII, doesn't this waste a lot of space?

1992 ~ Ken Thompson proposed  
a scheme to fix UTF-16  
problems

U+03B1 ~ 2nd row of UTF-8  
Chart

~~0000 0011 1011 0001~~  
xxxxxx xxxxxx

110 01110

0xCE

10110001

0xB1