High Performance Computing

What is Memory?

Martin Raum

An initial description

> *Computer memory is a part of the hardware to store bits of information, both data and instructions, and load them.*

Bits are the singleton of classical (as opposed to quantum) computer information theory. A bit represents either either "true" or "false" usually denoted by 1 and 0.

This does not make reference to where the memory is located and what it is attached to.

We will initially focus on host memory and on data being stored.

Information is nothing but sequences of bits

Information, for instance, integers, floating points numbers, or strings are encoded as sequences of bits. This is also referred to as their binary representation.

Programming in C gives you access to these encodings; for the purpose of speed you may have to make use of them directly.

Reading bit representations

Eight bits are usually grouped and referred to as bytes. This yields $2^8 = 256$ possible values.

Bytes can be conveniently written as two-digit hexdecimal numbers. In contrast to the decimal system its digits run between 0 and 15 as opposed to 0 and 9.

Decimal representation															
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

Hexadecimal representation															
0	1	2	3	4	5	6	7	8	9	a	b	c	d	e	f

A first mental model

To start with we will consider memory as a sequence of bytes. The position of a byte in this sequence is called its (memory) address. Memory addresses are stored in pointers.

We can gain access to parts of specified size by “memory allocation”. This can happen implicitly or explicitly.

Allocated memory must be “freed”, implicitly or explicitly.