Direct Mapping

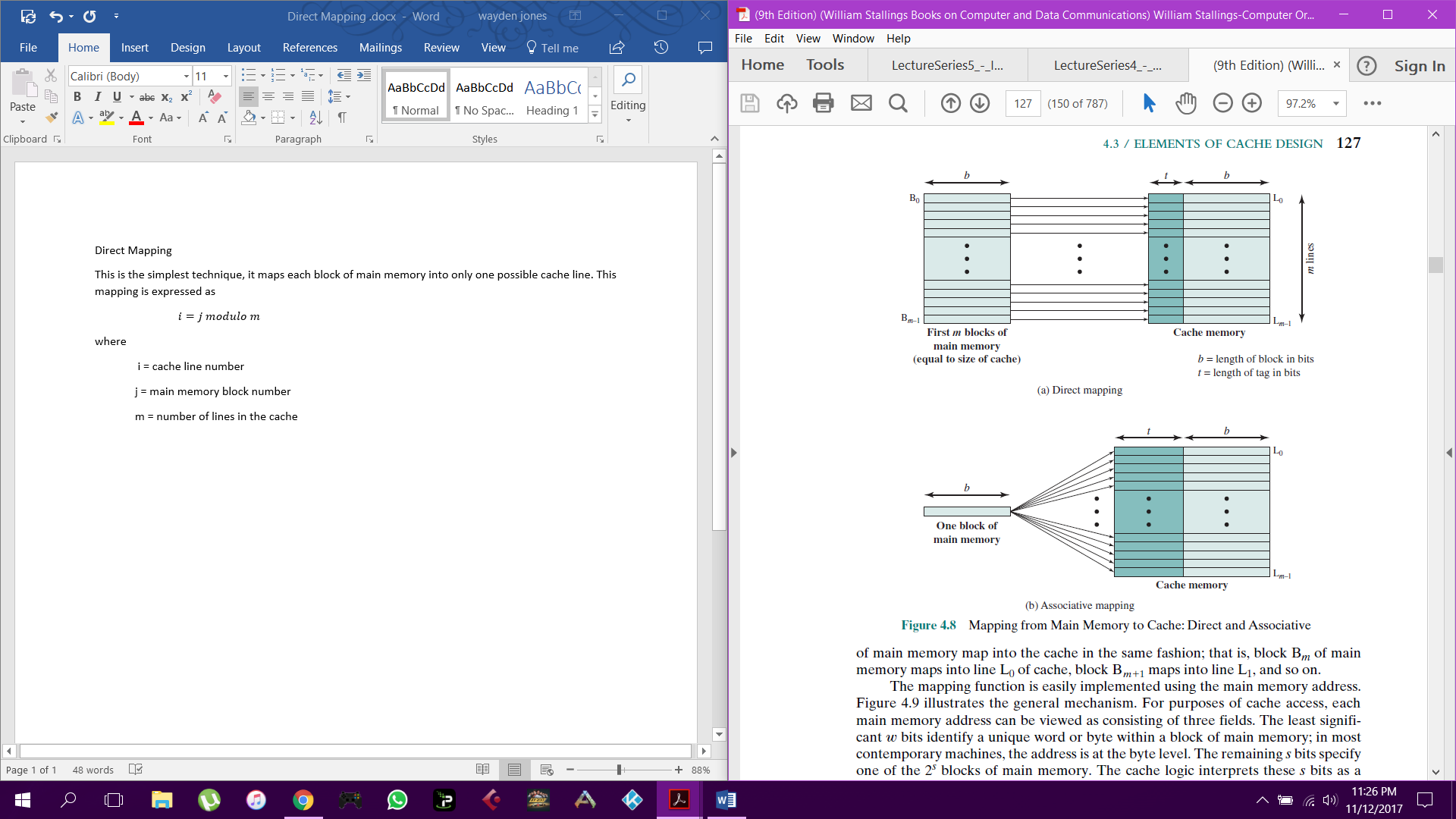
This is the simplest technique, it maps each block of main memory into only one possible cache line. This mapping is expressed as

where

i = cache line number

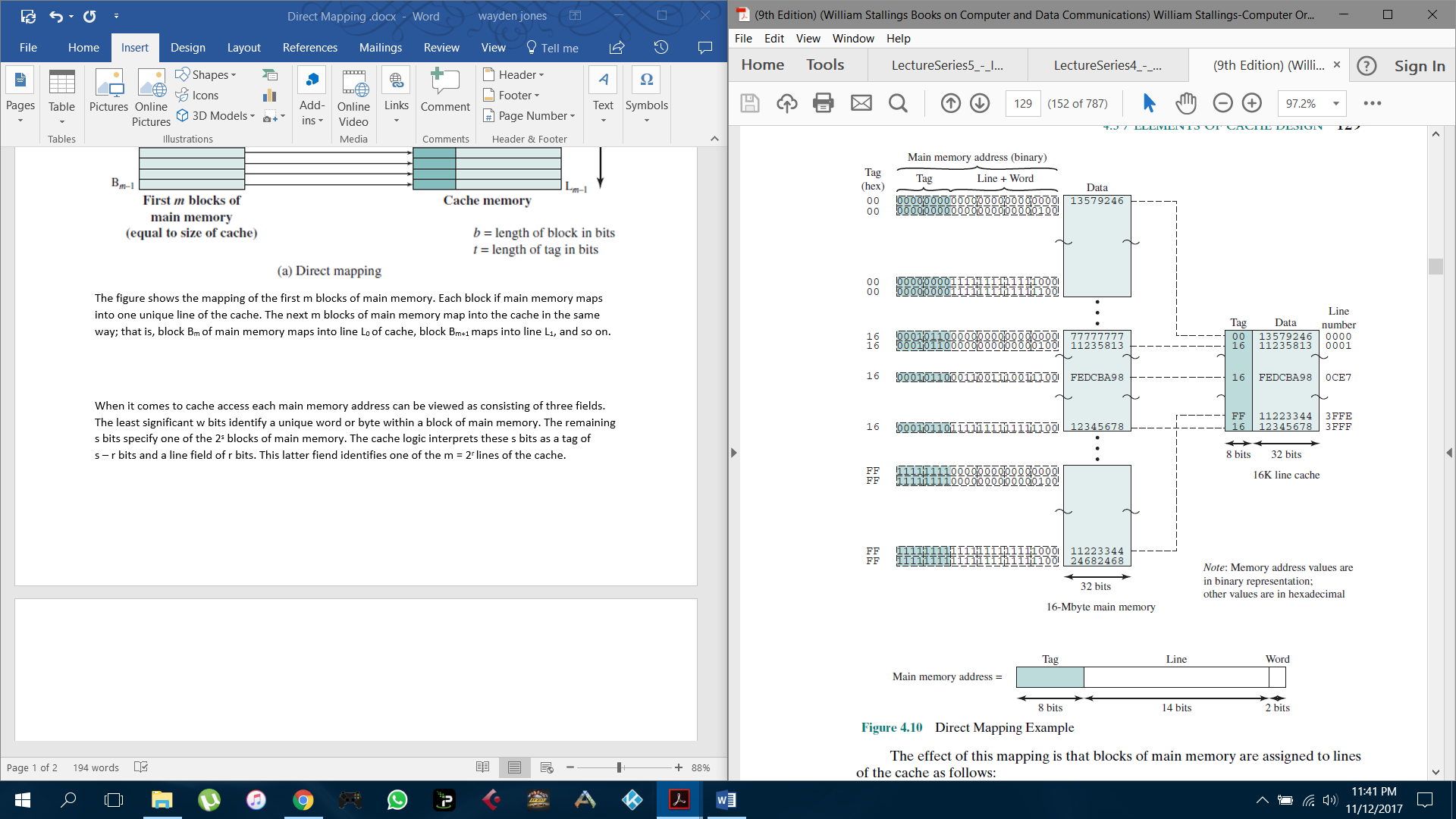
j = main memory block number

m = number of lines in the cache



The figure shows the mapping of the first m blocks of main memory. Each block if main memory maps into one unique line of the cache. The next m blocks of main memory map into the cache in the same way; that is, block Bm of main memory maps into line L0 ­of cache, block Bm+1 maps into line L1, and so on.

When it comes to cache access each main memory address can be viewed as consisting of three fields. The least significant w bits identify a unique word or byte within a block of main memory. The remaining s bits specify one of the 2s blocks of main memory. The cache logic interprets these s bits as a tag of s – r bits and a line field of r bits. This latter fiend identifies one of the m = 2r lines of the cache.



In the example, m = 16K = 214 and i = j modulo 214. The mapping becomes

Cache Line Starting Memory Address of Block

0 000000, 010000, …, FF0000

1 000004, 010004, …, FF0004

. .

. .

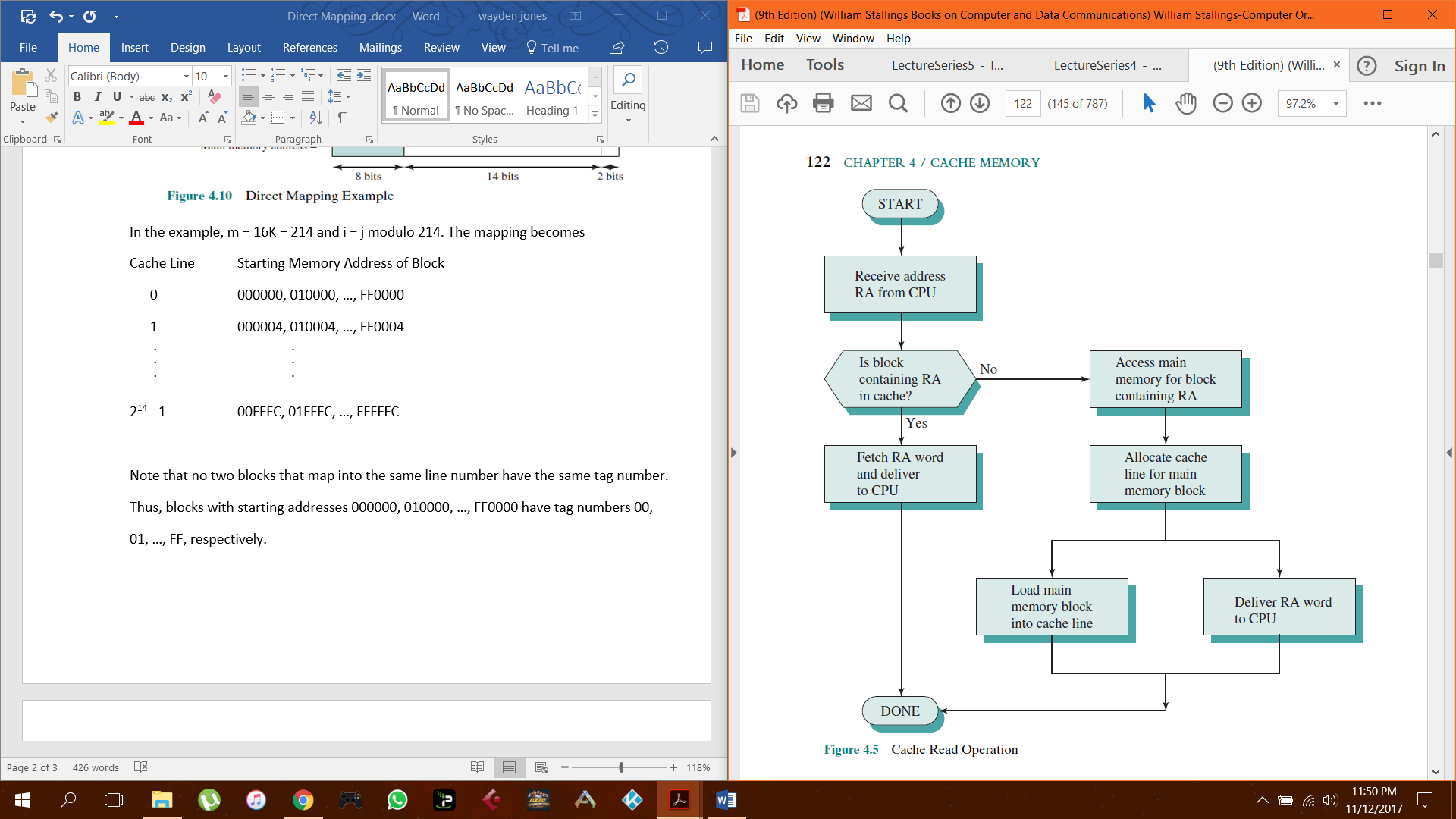
. .

214 - 1 00FFFC, 01FFFC, …, FFFFFC

Note that no two blocks that map into the same line number have the same tag number.

Thus, blocks with starting addresses 000000, 010000, …, FF0000 have tag numbers 00,

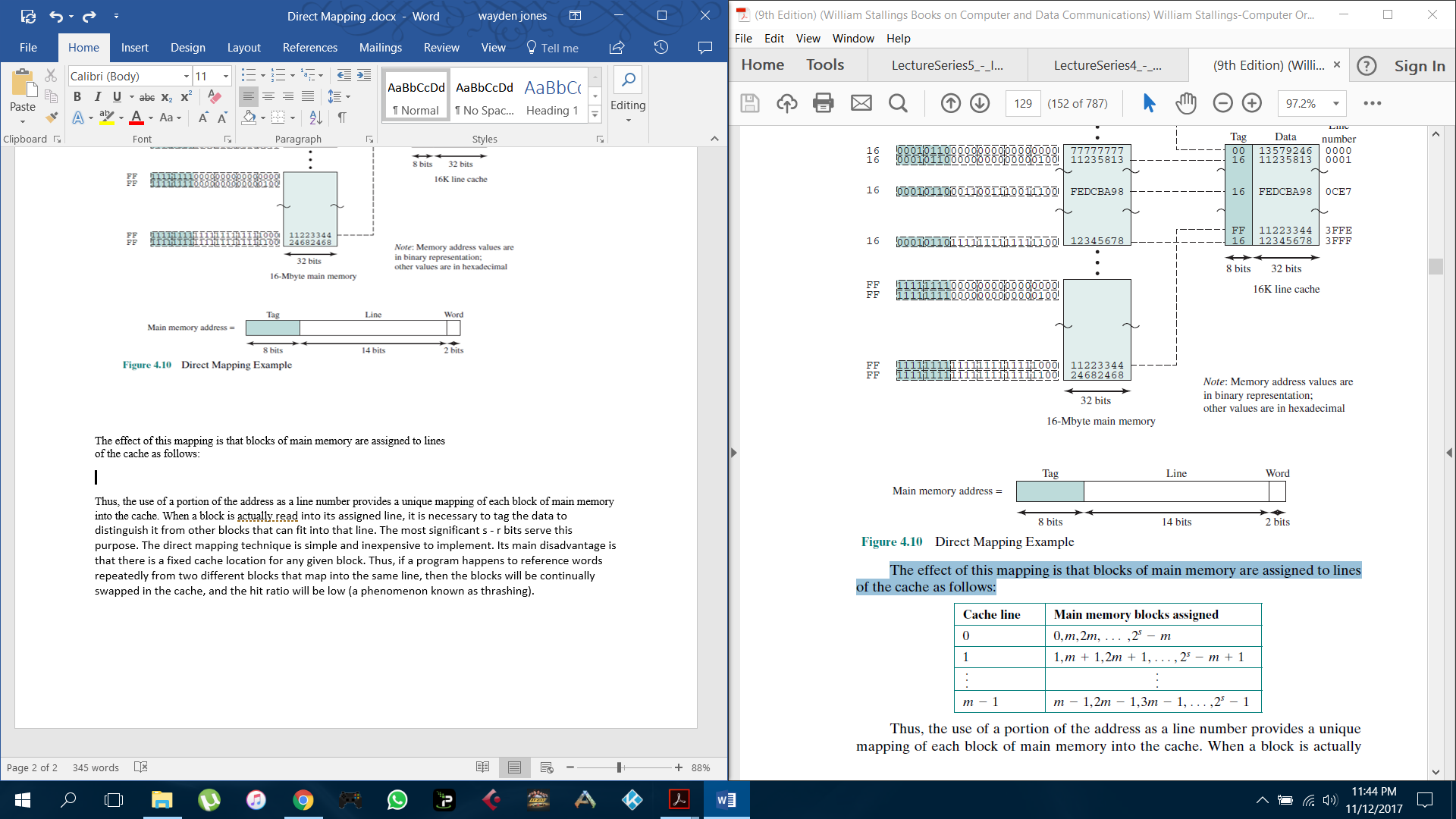
01, …, FF, respectively.



A read operation works as follows. The cache system is presented with a 24-bit address. The 14-bit line number is used as an index into the cache to access a particular line. If the 8-bit tag number matches the tag number currently stored in that line, then the 2-bit word number is used to select one of the 4 bytes in that line. Otherwise, the 22-bit tag-plus-line field is used to fetch a block from main memory. The actual address that is used for the fetch is the 22-bit tag-plus-line concatenated with two 0 bits, so that 4 bytes are fetched starting on a block boundary.

The effect of this mapping is that blocks of main memory are assigned to lines

of the cache as follows:



Thus, the use of a portion of the address as a line number provides a unique mapping of each block of main memory into the cache. When a block is actually read into its assigned line, it is necessary to tag the data to distinguish it from other blocks that can fit into that line. The most significant s - r bits serve this purpose. The direct mapping technique is simple and inexpensive to implement. Its main disadvantage is that there is a fixed cache location for any given block. Thus, if a program happens to reference words repeatedly from two different blocks that map into the same line, then the blocks will be continually swapped in the cache, and the hit ratio will be low (a phenomenon known as thrashing).