A parallel algorithm for matrix-vector/matrix-matrix multiplication refers to performing matrix operations on a computer, and achieving optimal computation speeds by utilizing parallel processing. Since a computer often has multiple processers, then the various tasks necessary to perform a matrix operation can be distributed among the multiple processors of the computer system. This can be done by partitioning the data into stripes (columns/rows), or into rectangular fragments (blocks).

Parallel processing can be difficult to program, because sending or receiving information from one processer to another takes up time, in addition to performing a calculation. Properly organizing a “network” of processers to work together to solve pieces of a problem and then assemble the pieces together can be difficult to implement, which is why partitioning data into simple “stripes” or “blocks” is the most commonly used approach.

Sources used.

<http://www.hpcc.unn.ru/mskurs/ENG/DOC/pp07.pdf>

<https://ac.els-cdn.com/0898122188902623/1-s2.0-0898122188902623-main.pdf?_tid=7a54f0ee-6a4f-496f-93a7-3dda15b7612a&acdnat=1550794300_e042d57e90d5e3476b2a1421f3a13e6d>