Library for matrices multiplication written as part of research works by scientists and experts in company *«Basic Projects»*. Main argument and motivation for Assembly language usage is extreme low level optimization for maximize performance speed. HPC or High Performance Computing and its practical usage for modeling complex systems is one of main activities of our group. Years of experience programming computational problems accumulated by our professionals since the days of the math co-processor Intel 8087 and the first personal computers IBM PC. Developed new theoretical methods and principles of computer modeling provide solution of tasks of computer simulation of «systems» and complex equipment, which were previously considered unsolvable. See also information in the end of the file.

The operation of matrix multiplication is realized using 256-bit AVX vector extensions and is optimized for multiprocessor server platforms. The current version of the library is intended for operating systems of Microsoft Windows and Intel's Sandy Bridge processors. The project is in continuous development. Planned support for new functional extensions (AVX512) and the operating systems of Linux, as well as the extension of the set of executed computational operations.

The library has been developed by two specialists

Parfus Alexander. Head of the IT Department. Candidate of technical Sciences. Developed computational algorithms and technology «extreme optimization» performance of a multilevel cache memory for efficient implementation of the procedure of matrix multiplication on modern multiprocessor systems with multilevel cache memory. Developed a model separation of functionality and interface between upper level and low-level computing library.

Manusov Ilya. Chief specialist of hardware and system programming. Performed a survey of critical performance factors. Programmed and debugged algorithms in Assembly language using vector extensions AVX256. Implemented benchmarks for computational procedures and standard operations read and write the cache memory of various levels and memory.

The library functions are executed as a single thread, however, is optimized for multithreaded processing on multiprocessor system. It is assumed that the management of parallel running threads is executed in the program top level.

The library is written in Assembly language for Intel x64. The result of translation are the DLL files that are designed to run on a 64-bit Microsoft Windows operating systems. As the translator uses Flat Assembler (FASM) version 1.71.17. The debugger FDBG for assembler code debug. The development environment of FASM Editor 2.0. For correct display of tabs and color coding comments it is recommended to use the development environment when working with the source code of the program.

The library uses a 256-bit vector extension of Intel AVX in the first generation and at this stage of development is optimized for Intel Sandy Bridge. Performance close to the peak corresponding to the implementation of the two vector operations (one addition and one multiplication) by each core for each cycle, reached for Sandy Bridge. As for the number of test examples, managed to surpass the performance of common mathematical library Intel MKL unit of interest. Soon will be published the results of comprehensive testing. On other processors supporting AVX256, the performance of the current version of the library may not be optimal. In development is a library that uses AVX512 vector extension to the Intel Xeon based on core the item has been added.

The library uses two basic algorithm, we select the top-level program depending on the dimensions of the multiplied matrices, with the aim of maximum performance in relation to a particular task. A more detailed description is in document

15-09-27_AVX_256_Specific_Specification_v200.DOC.

In the document

15-10-13_AVX_OS_DLL_Specification_v202.DOC

is description support libraries for the configuration of a computing platform, designed for the detection of functional extensions supported by the processor, the size of the cache and other settings.

Description of the basic library files of the configuration of a computing platform AVX_OS.

Translated main module:

AVX_OS.asm

Modules for functions debug:

Debug_DLL_DetectCache.asm

Debug_DLL_DetectVXLib.asm

Connected modules:

DetectCache.inc

DetectVXLib.inc

Description of the main computational libraries files AVX_256_Specific.

Translated main module:

Specific_AVX1_256_1.asm

<u>Modules for functions debug:</u>

Debug_DLL_InitLibrary.asm

Debug_DLL_PackA.asm

Debug_DLL_PackASmall.asm

Debug_DLL_PackB.asm

Debug_DLL_PackBSmall.asm

Debug_DLL_UnPackA.asm

Debug_DLL_UnPackASmall.asm

Debug_DLL_UnPackB.asm

Debug_DLL_UnPackBSmall.asm

Debug_MultiplyMatrixBig.asm

Debug_MultiplyMatrixSmall.asm

<u>Connected modules:</u>

<DIR> MMB

<DIR> MMS

Debug_Context.inc

InitLibrary.inc

PackASmall.inc

PackA_TRANSP.inc

PackB.inc

PackBSmall.inc

UnPackA.inc

UnPackASmall.inc

UnPackB.inc

UnPackBSmall.inc

Briefly «about us».

The core of the team are natives of the institutes of the Academy of Sciences of the former Soviet Union, highly qualified specialists, doctors and candidates of Sciences, whose area of activity – the problem solving simulation «large systems». They are representatives of the scientific school of Acad. Georgiy

Evgenyevich Pukhov, as well as other specialists.

For convenience let us consider a specific example of decision of tasks of modeling of gas transport systems (GTS). First of all we are talking about the decision of such basic tasks:

- identification of actual modes of operation of hydraulic structures and the parameters of its elements (the solution of a problem of calculus of variations);
- the solution of the optimal control of the operating modes of the GTS.

In both tasks address the actual transient behaviour operating modes of GTS. GTS can be seen at any level of complexity until level continental GTS.

When such tasks need to be prepared for the following dimensions of the system: the number of degrees of freedom (number of variables) may amount to «millions and tens of millions of unknowns» amount into account technological restrictions of the equipment of the GTS may amount to «hundreds of thousands and millions».

The «Curse of dimensionality» makes it impossible to solve such problems of the known mathematical methods.

> The term «Curse of dimension» (also known as «Curse of dimensionality difficulties» or Diculties), reflects a rapid growth in the number of required calculations that arise with the increase of object modeling. The considered problems belong to the exponential task complexity.

Until recently, the task of «Optimization of operating modes of GTS» was incapable of calculation. This is true in General for optimization problems (and reverse tasks) the so-called «Large systems».

At the heart of our development lays out the new principles and methods create accurate numerical models of complex GTS (and in General complex systems), provide high performance real time solution of tasks of computer simulation of «systems» and complex equipment, which were previously considered unsolvable.

Ensures the removal of the «Curse of dimensionality», the pressures on the problems of modeling and optimization, «large systems».

Methods and algorithms are developed «high level», in combination with the techniques of «extreme optimization» low-level computational algorithms that allow to solve the problems supercomputer level, using a computing platform for mass application.

Published source computational libraries are one example of such optimization.

Some algorithms for hardware accelerated computing is to be included in this project.

*   *   *

Contacts:

Head of development Alexander Parfus
E-mail: AVP@basic-projects.com
Chief specialist hardware Ilya Manusov
E-mail: im1969@yandex.ru