UDC 004.272.43.003.13

V.P. Ivaschenko, G.G. Shvachych, M.A. Tkach

PROSPECTS OF NETWORK INTERFACE INFINIPBAND IN MULTIPROCESSOR COMPUTER SYSTEMS FOR SOLVING TASKS OF CALCUALTIONS' AREA SPREADING

The article investigates the specific application InfiniBand network interface in a multiprocessor computer system for tasks aimed at increasing the computational domain. The main patterns regarding the computing time of a task depending on changes in the field of computing in a multiprocessor system. The carried out researches had the aim to determine the deceleration rate calculations. The rate is connected with an increase in the computational domain of multiprocessor system in comparison with a computer having the unlimited field of computing.

Keywords: multi-processor computer system, computing nodes, calculations ' area for multiprocessor systems, computing platforms.

Introduction

The need to use high-performance computing in the whole world belongs to the fundamentals of developing a strategic capacity and has an important scientific, technical and economic importance. Nowadays there are two main methods of improving productivity and performance of computing systems:

– the use of increasingly complicated element base;

– parallel execution of computational operations.

The first method needs a very significant investment. Experience of the Company CRAY which created supercomputer based on gallium arsenide showed that the development of fundamentally new element base for high performance computing systems is a daunting task even for such a well-known corporation. The second method dominates after the announcement of the U.S. government program: " Accelerated Strategic Computer Initiative » (ASCI). Taking the above in account we note that in recent years the process of creating high-performance systems developed mainly in the direction of combining many parallel processors

© Ivaschenko V.P., Shvachych G.G., Tkach M.A., 2014

for the solution of a large and complex task [1 – 3]. In this regard now we equate supercomputer and parallel (multi) computer system.

We note that the development of such systems is an actual problem. This is not only because of the fundamental limitations for the maximal possible speed of conventional serial computers but also because of constant existence of computational tasks for which capacity of existing computer equipment is insufficient. These tasks include, for example, the numerical modeling of hydrodynamics and metallurgical Thermo physics [4 – 5], the pattern recognition problem, optimization problems with a large number of parameters, climate modeling, genetic engineering calculations, design of integrated circuits, analysis of environmental pollution [6] as well as solving a wide range of multidimensional unsteady problems [7], etc.

However, there were not many researches devoted to the effectiveness of computing parallelization. This can be explained by the fact that this problem is extremely complex because the efficiency of parallelization of calculations depends on many factors. Along with this, we note that neglecting of these factors can negate the effect of increasing the number of processors used. Taking into account the marks, this work is aimed at disclosure the effectiveness of parallelization for a certain class of tasks solved by using multiprocessor computer systems.

Analysis of recent research and publications

Currently there appeared a unique opportunity to create inexpensive installations network technologies of the supercomputer's type: multiprocessor computer clusters. Until recently, there was doubt in the prospect of such a direction. However, with all the "pros" and "contras" the permanent residents in a list of Top500 : companies Cray, Sun, Hewlett-Packard and others had to make room passing forward a number of cluster solutions. On the other hand, now the market is rapidly developing, and manufacturers of networking solutions based on cLAN, Myrinet, ServerNet, SCI continue to improve their technologies, practically enabling to construct their own version of a supercomputer without any financial expenses. It is obvious today there are many different options for building cluster computing systems. However, one of the major differences in their design lies in the networking technology, the choice of which is determined by the class of tasks. For example, solving the metallurgy tasks via the mathematical modeling of high-speed heat treatment of the long length objects. One of the main problems can be formulated as follows: we have a differential grid of dimension M, the computation time of solving the task with using a single-processor system is determined by t. This parameter is critical. It is necessary to reduce significantly the computation time while preserving the value of M. Here we consider the problem of reducing the computation time by increasing the number of nodes in a cluster system. This approach is focused, for example, on the development of new technological processes in which the computation time is a critical value. In addition, similar problems often have to be solved in the fields of medicine, military equipment, etc.

Thus, the theme of designing cluster multiprocessor systems today is relevant, interesting and experiencing a stage of rapid development. It is clear that using high productive clusters (HPC) is an effective way to solve a wide range of topical problems. In our opinion, the new qualitative stage of development of multiprocessor cluster system lies in the use of new advanced network technologies. This is explained as follows. Network computing cluster system is fundamentally different from a network of workstations, although it demands to build cluster conventional network cards and hubs/switches that are used in the design of a network of workstations. However, in the case of a cluster computing system, there is one fundamental feature. Cluster network is primarily intended for computing processes, not for computers' communication. In this regard, the higher is the throughput speed of a computer network cluster, the faster will be solved users' parallel tasks performed at the cluster. Thus, the technical characteristics of computer networks is of the utmost importance for multi-cluster systems.

By now the problem of choice and analysis of network technology for modular multi-cluster systems was not well developed. In addition, practically there are no works devoted to the study of influence of network technologies on parallelization efficiency in modular multiprocessor cluster systems. In this regard studies, considered in this paper, are directed on prospects for InfiniBand network technology application for solving tasks with an expandable calculations' area.

Statement of the Problem Research

In this paper we consider the following problem. There is a differential grid having dimension M; the computation time for solving the task by using a single-processor system is determined by the value *t*. This parameter is not determinative. The principal is increasing the size of the net, and above the one which can be processed in a computer memory. This procedure is critical for a more detailed account or getting some new effects of the investigated processes. It is necessary to investigate the features of the calculations under this class of tasks based on the use of multiprocessor computer system having network interface InfiniBand.

Purpose and objectives of research

The purpose of this work is to study the specific application InfiniBand network interface in parallel computing for solving tasks related to the extension of the computational area.

First of all, it is necessary to:

1. Identify the basic regularity regarding the counting time for a task depending on changes in the field of multiprocessor computing system designed on the application of network interface InfiniBand. At the same time it is important to carry out the main analytical interconnections defining the dependence of the solution of the problem on the basic parameters of a multiprocessor system.

2. Explore the option of a hypothetical computer with unlimited memory and to hold the comparative analysis with a real multiprocessor system; compare features of formation computational area for this computer with analytical interconnection; show the results of the comparative analysis of the functioning of a real multiprocessor system and hypothetical computer with unlimited memory to determine the main factors affecting the efficiency of parallelization of a real computer system.

3. Run the modeling phase of the main performance characteristics of the problem being solved by the application of a multiprocessor system designed on the application network interface InfiniBand.

4. To carry out the research aimed at determining the rate of deceleration the calculations associated with an increase of the area of multiprocessor computing system, distributed over its nodes, in a comparison with a computer with an unlimited area of computing.

Derive analytical expressions for the rate of deceleration the calculations. Studies are directed on further development of the approach given below which was focused in [5 – 12]. The studies and developed to a multiprocessor system [13].

The main material of research

Thus, we consider the task of expanding the area of computing by increasing the number of nodes in a cluster system. We assume that the area is evenly distributed among computing nodes of a cluster system. For the convenience of research we also assume that the area in which the calculations are carried out has the shape of a circle. At the same time each node of the multiprocessor system corresponds to one sector of an isometric circle. Taking into account that each cluster node has available RAM R (Gbit), the total area of computing for a multiprocessor system is represented as the ratio:

$$S = N \cdot R, \qquad (1)$$

where N – number of nodes in a multiprocessor system.

In the conditions when the area of computing is maximally loaded and evenly distributed among the nodes of a multiprocessor system, one can define a formula to calculate the volume of the boundary data exchange (in Gbit). This formula is:

$$E_{\text{ex}} = m \cdot N \cdot \sqrt{\frac{S}{\pi}}, \qquad (2)$$

The value of m may be equal to unity for a unilateral regime of boundary data exchange or two for bilateral one. Note that at $N = 1$ it is completely obvious that the value of the amount of data exchange boundary (E_{ex}) vanishes.

Under such circumstances it is possible to determine – T_{ex} the time of data exchange among the boundary nodes of the cluster c. Note that the counting time iteration depends only on the power of the processor while the time of boundary data exchange is dictated by the size of the differential grid, the number of nodes of the cluster system and computer network capacity. Consequently, one can determine the value T_{ex} from ratio:

$$T_{ex} = \frac{E_{\text{ex}}}{V}. \qquad (3)$$

In the expression (3) V – is the cluster network capacity (Gbit / s). In general a communication network throughput in a multiprocessor system can be determined on the relation:

$$V = k \cdot d \cdot V_p, \qquad (4)$$

where V_p (Gbit /s) is the network capacity of a port, κ – is a number of computer network communication channels which operate simultaneously (the number of computer networks), d – half-duplex (d = 1) or duplex (d = 2) mode of a cluster computer network system. In this class of tasks all calculations are made on the basis of the differential grid. In addition, for the analysis of the effectiveness of a multiprocessor system the most important parameter is the calculation time per iteration (T_{it}).

$$T_{\mathrm{it}} = T_c^N + T_{ex} \quad . \tag{5}$$

Here T_c^N is the computing time for one iteration of a multiprocessor system (s):

Obviously, for the case where N = 1 we obtain that

$$T_{\mathrm{it}} = T_c^1 \, . \tag{6}$$

where T_c^1 – is the counting time per iteration for a single-processor computer system. For the case where N > 1 the total calculation time per iteration will be determined as (7) regarding expression (2).

$$T_{\mathrm{it}} = T_c^N + \frac{\mathrm{m} \cdot \mathrm{N} \cdot \sqrt{\frac{\mathrm{S}}{\pi}}}{\mathrm{V}} \, . \tag{7}$$

For the considering type of tasks we will specify the first term in (7). Wherein:

$$T_c^N = \frac{\mathrm{S}}{N \cdot V_c} \, , \tag{8}$$

V_c – counting rate of one iteration of the task for this type of processor and associated numerical methods which are determined experimentally. Using (1), the expression (8) is presented as following:

$$T_c^N = \frac{\mathrm{R}}{V_c} \, . \tag{9}$$

Analysis of (9) shows that T_c^N depends on the memory amount used by the CPU and on the speed of computation for one iteration for this node type multiprocessor system.Thus, we have all the prerequisites for the determination of total computation time per iteration for a multiprocessor system:

$$T_{\mathrm{it}} = \frac{\mathrm{R}}{\mathrm{V_c}} + \frac{\mathrm{m} \cdot \mathrm{N} \cdot \sqrt{\frac{\mathrm{N} \cdot \mathrm{R}}{\pi}}}{\mathrm{V}} \tag{10}$$

Analysis of (10) shows that at the increase of the computational area in N times the calculation time for the task grows as $N^{3/2}$ with a coefficient that depends on the amount of RAM node cluster network capacity and the character of communication among nodes , i.e.:

$$T_{\text{it}} = \mathrm{T}_c^N + N^{3/2} \cdot f \cdot (\mathrm{m}, R, V). \tag{11}$$

Analysis of (11) shows the perspective of the modern communication technologies, e.g., InfiniBand application and also the use of multi-core computing platforms.On the background of the conducted research let us consider the case of hypothetical computer with unlimited memory. Thus, taking into account the relation (6), we obtain:

$$T_c^1(\mathrm{S}) = \frac{\mathrm{S_i}}{\mathrm{V_c}} \tag{12}$$

In (12) the total area of the hypothetical computer calculations represented in the form:

$$\mathrm{S_i} = i \cdot R\,, \tag{13}$$

where i – coefficient that determines the change in the area of computing hypothetical computer. Analysis of the interconnections (12) and (13) shows that the increase of the total amount of computation in N times the calculation time increases linearly with some coefficient depending on the computational capabilities of the processor.

In accordance with the derived computational interconnections' the experiments for a computing platform equipped with a processor Intel E8400 3 GHz were conducted. Here features of the tasks of the solved class and the ones of the cluster system were accepted as the initial characteristics. It is shown in Table.1.

Table 1

Initial data for calculating system performance using a computing platform equipped with a processor Intel E8400 3 GHz

Vp	8 Gbit/s
T_c^1	100 s
Vc	14 10 ^{9} bit/s
R	24 Gbit
m	2
d	2
k	1

The obtained simulation results are listed in Table 2 .

Under this system simulation results showed the following general trend: in effect a significant impact of time for the boundary data exchange by the total time for solving the problem, on the background of expansion of area calculation time to solve the problem for real multi-processor system will increase significantly compared with the ideal computer.

Table 2

The results of calculating the basic characteristics of the system equipped with a processor Intel E8400 3 GHz

N	*Sn*	*Eex*	*Tex*	*Tc,n*	*Tit*	*Tid*
2	48,00	15,64	1,95	1,71	3,67	3,43
3	72,00	28,73	3,59	1,71	5,31	5,14
4	96,00	44,23	5,53	1,71	7,24	6,86
5	120,00	61,82	7,73	1,71	9,44	8,57
6	144,00	81,26	10,16	1,71	11,87	10,29
7	168,00	102,40	12,80	1,71	14,51	12,00
8	192,00	125,11	15,64	1,71	17,35	13,71
9	216,00	149,29	18,66	1,71	20,38	15,43
10	240,00	174,85	21,86	1,71	23,57	17,14
11	264,00	201,73	25,22	1,71	26,93	18,86
12	288,00	229,85	28,73	1,71	30,45	20,57
13	312,00	259,17	32,40	1,71	34,11	22,29
14	336,00	289,64	36,21	1,71	37,92	24,00
15	360,00	321,22	40,15	1,71	41,87	25,71
16	384,00	353,88	44,23	1,71	45,95	27,43
17	408,00	387,56	48,45	1,71	50,16	29,14
18	432,00	422,26	52,78	1,71	54,50	30,86
19	456,00	457,93	57,24	1,71	58,96	32,57
20	480,00	494,56	61,82	1,71	63,53	34,29

The obtained simulation results are displayed in the form of plots (Fig. 1).

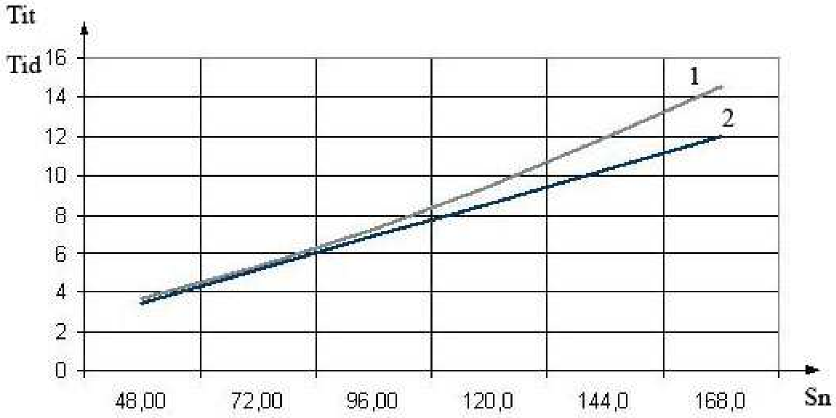

Figure 1 - Curves of the calculation time of one iteration depending on the size of the computational area for a multiprocessor system equipped with an Intel Pentium 4 3 GHz

As we can see in Fig. 1 counting time per iteration by increasing the computational area of a multiprocessor system varies according to the nonlinear dependence (curve 1, T_{it}). This dependence shows that an increase in the computational area in N times the calculation time of the problem grows as $N^{3/2}$ with a coefficient that depends on the amount of RAM for the node cluster network capacity and the nature of data exchange among the computing nodes. Counting time for one iteration of a hypothetical computer with unlimited memory, as was expected, increases linearly (Line 2, T_{id}). The angle of a line inclination is determined by the characteristics of the applied computing platform. Note that the experiments were carried out for a multiprocessor system shown in [13].

For the system described above simulation results showed the following general trend: significant impact time for the expensed area to solve the task of boundary data exchange by the total time for solving the task for real multiprocessor system will increase significantly compared with the ideal computer.

Further studies are directed at determining the deceleration calculations' rate (K) associated with an increase of the multiprocessor computing system area distributed over its nodes in comparison with one computer with an unlimited computing area. Obviously, such a deceleration rate will be determined by the ratio.

$$\mathrm{K}= \frac{\mathrm{T}_{\mathrm{c}}^{\mathrm{N}}}{\mathrm{T}_{\mathrm{c}}^{1}(\mathrm{S})}. \tag{14}$$

Equation (14) shows that this factor is calculated and, simultaneously, one has to take into account the increase of the computing area of a multiprocessor system distributed above the nodes. Using (5) we obtain:

$$\mathrm{K}=\frac{T_c^N+T_{ex}}{T_c^1(S)}. \tag{15}$$

In regarding (7 - 10), (12) and (13), the value of deceleration calculations' rate (K) can be represented in the form convenient for analysis:

$$\mathrm{K}= \frac{1}{\mathrm{N}}(1+\frac{T_{ex}}{T_c^N}). \tag{16}$$

Expression (16) can be represented in a form suitable for analysis:

$$K= \frac{1}{N}(1+K_1). \qquad (17)$$

In the expression (17) K_1 is defined as:

$$K_1= \frac{T_{ex}}{T_c^N}. \qquad (18)$$

This factor can be interpreted as the ratio of the active deceleration due to the fact that this value mainly affects the deceleration rate of calculations in general.

The features of the solved class of tasks and the ones of the cluster system were accepted as the initial characteristics and shown in Table.1. The simulation results are presented in the form of plots (Fig. 2).

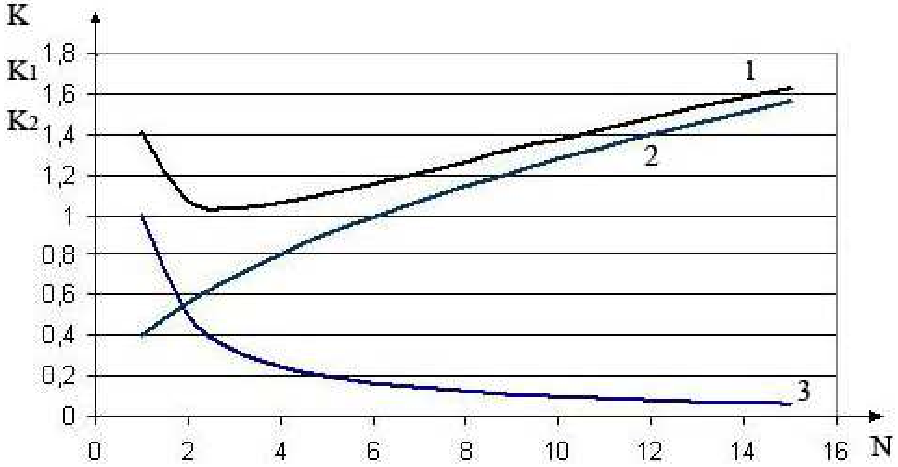

Figure 2. - Curves deceleration rate of the number of nodes in a multiprocessor system equipped with a processor Intel E8400 3 GHz

In Figure. 2 Line 1 shows the general trend of the deceleration of calculations rate. Line 2 shows the effect of time of the boundary data exchange on deceleration computing coefficient. At the same time line 3 shows the effect of the number of nodes in a multiprocessor system by the value of deceleration calculations' rate. These dependences have a significant effect on computing time of the boundary data exchange deceleration coefficient. Last underlines the need to perform a procedure matching the network interface and computational capabilities of the selected computing platform. Furthermore, it becomes apparent that at other equal conditions for a multiprocessor system we have a problem of optimal choice of the nodes' number to minimize the deceleration rate calculations.

Conclusions and prospects for further research

1. We revealed main patterns regarding the counting time for the task depending on changes in the computing area for a multiprocessor system. It is shown that at the increase of the total amount of computation in N times the calculation time increases in $N^{3/2}$.

2. We derived the main analytical interconnections determining the dependence of the task solution on the basic parameters of a multiprocessor system. Such interconnections showed that the calculation time for a task grows nonlinearly with a coefficient that depends on the amount of a node memory, network capacity and the nature of the cluster communication among computing nodes.

3. We considered an option of a hypothetical computer with unlimited memory and held its comparative analysis with a real multiprocessor system and derived analytical expressions defining features of the computational area for the formation of such a computer. To determine the main factors affecting the efficiency of parallelization of a real computer system there were analyzed two processes: functioning of a real multiprocessor system and functioning of a hypothetical one with unlimited memory. Studies showed the relevance of the reconciliation process components of the network interface and the processing capabilities of the selected computing platforms.

4. The research was focused on determining the calculations' deceleration rate associated with an increase in the area of multiprocessor computing system distributed over its nodes; the rate coefficient was compared with the one of a computer with an unlimited computing area. We derived analytical expressions for the calculations' deceleration rate. The decisive role of time for the boundary data exchange on the value of the calculations' deceleration rate was shown.

5. In their subsequent studies the authors intend to highlight features matching components of the network interface and the processing capabilities of the selected computing platforms and provide solution for the task of optimal choice of nodes in a multiprocessor system to minimize the calculations' deceleration rate.

LITERATURE

1. Лацис А.О. Как построить и использовать суперкомпьютер / А.О. Лацис. – М.: Бестселлер, 2003. – 240 с.
2. Гергель В.П. Основы параллельных вычислений для многопроцессорных вычислительных систем: учеб. пособие / В.П. Гергель, Р.Г. Стронгин. – Н.Новгород: Н.НГУ, 2003. – 184 с.
3. Beowulf Introduction & Overview [Електронний ресурс]. – Режим доступу: http://www.beowulf.org.
4. Коздоба Л. А. Вычислительная теплофизика / Л.А. Коздоба. – К.: Наук. думка, 1992. – 224 с.
5. Иващенко В.П. Параллельные вычисления и прикладные задачи металлургической теплофизики / В.П. Иващенко, Г.Г. Швачич, А.А. Шмукин // Системні технології. Регіональний збірник наукових праць. – Вип. 3(56). – Т. 1. – Дніпропетровськ, 2008.– С . 123 – 138.
6. Швачич Г.Г. К вопросу конструирования параллельных вычислений при моделировании задач идентификации параметров окружающей среды / Г.Г. Швачич // Математичне моделювання.– 2006.–№ 2 (14).–С. 23 – 34.
7. Швачич Г.Г. О параллельных компьютерных технологиях кластерного типа решения многомерных нестационарных задач / Г.Г. Швачич // Materibly IV mezinбrodnн videcko- praktickб konference [«Vedecky potencial sveta - 2007»]. – D. Technicke vedy. Matematika. Fyzika. Modernн informacni technologie. Vystavba a architektura. – Praha: Publishing House «Education and Science» s.r.o. – P. 35 – 43.
8. Швачич Г.Г. Математическое моделирование скоростных режимов термической обработки длинномерных изделий / Г.Г. Швачич, В.П. Колпак, М.А. Соболенко // Теория и практика металлургии. Общегосударственный научно-технический журнал.–2007.–№ 4– 5(59 – 60).–С. 61 – 67.
9. Сбитнєв Ю.І. Дослідження оцінки ефективності багатопроцесорної кластерної системи / Ю.І. Сбінтєв, Г.Г. Швачич, М.О. Ткач // VI Intrenational Conference “Strategy of Quality in Indastry and Education”, June, 1 – 8, 2010, Varna; Bulgaria. – Proceedings. – V. 2. – P. 288 – 296.
10. Швачич Г.Г. О проблеме исследования эффективности модульной кластерной системы / Г.Г. Швачич, Ю.І. Сбитнєв, М.О. Ткач // [Електронний ресурс]. – Режим доступа: http://cluster.linux-ekb.info/cuda1.php.
11. Іващенко В.П. Дослідження оцінок ефективності модульної багатопроцесорної кластерної системи / В.П. Іващенко, Г.Г. Швачич, Є.О. Башков // Наукові праці Донецького національного технічного університету. Серія “Інформатика, кібернетика та обчислювальна техніка”. – Вип. 13 (185). – Донецьк: ДонНТУ, 2011. – С. 33 – 43.
12. Башков Е.А. Исследование влияния сетевого интерфейса на эффективность модульной многопроцессорной системы / Е.А. Башков, В.П. Иващенко, Г.Г. Швачич // Наукові праці Донецького національного технічного університету. Серія “Інформатика, кібернетика та обчислювальна техніка”. – Вип. 14 (188). – Донецьк: ДонНТУ, 2011. – С. 89–99.
13. Башков І.О. Високопродуктивна багатопроцесорна система на базі персонального обчислювального кластера / Є.О. Башков, В.П. Іващенко, Г.Г. Швачич // Наукові праці Донецького національного технічного університету. Серія “Проблеми моделювання та автоматизації проектування”. – Вип. 9 (179). – Донецьк: ДонНТУ, 2011. – С.312 – 324.