

МАТЕМАТИЧНЕ МОДЕЛЮВАННЯ ТА ОБЧИСЛЮВАЛЬНІ МЕТОДИ

УДК 681.324

В. М. ДУБОВОЙ, Е. Д. НИКИТЕНКО

Винницький національний технічний університет, Вінниця

ОЦЕНКА ХАРАКТЕРИСТИК ПОСЛЕДОВАТЕЛЬНО-ПАРАЛЛЕЛЬНОГО ПРОЦЕССА СТАТИСТИЧЕСКОГО МОДЕЛИРОВАНИЯ

Анотація. У статті вирішується задача оцінки швидкості виконання перетворень при моделюванні статистичних характеристик систем на основі GPGPU комплексів послідовно-паралельної обробки інформації. Розглянуто узагальнену задачу координації послідовно-паралельного процесу статистичного моделювання для послідовних кроків перетворення вихідної інформації. Процеси координації проаналізовані з застосуванням діаграм Ганта. Правила виконання операцій моделювання ґрунтуються на операторному методі перетворень законів розподілу ймовірностей, який використовує інтегральні оператори. У роботі синтезована узагальнена модель координації паралельних процесів і перевірено її адекватність на прикладі її застосування до процесу обробки зображень. Розроблений підхід дозволяє сформулювати критерій для вибору оптимального розміру блоків даних при паралельній обробці і скоротити час моделювання.

Ключові слова: статистичне моделювання, модель координації, послідовно-паралельна обробка, діаграма Ганта.

Аннотация. В статье решается задача оценки скорости выполнения преобразований при моделировании статистических характеристик систем на основе GPGPU комплексов последовательно-параллельной обработки информации.

Рассмотрена обобщенная задача координации последовательно-параллельного процесса статистического моделирования для последовательных шагов преобразования исходной информации. Процессы координации проанализированы с применением диаграмм Ганта. Правила выполнения операций моделирования основываются на операторном методе преобразований законов распределения вероятностей, который использует интегральные операторы. В работе синтезирована обобщенная модель координации параллельных процессов и проверена ее адекватность на примере ее применение к процессу обработки изображений. Разработанный подход позволяет сформулировать критерий для выбора оптимального размера блоков данных при параллельной обработке и сократить время моделирования.

Ключевые слова: статистическое моделирование, модель координации, последовательно-параллельная обработка, диаграмма Ганта.

Abstract. The article is devoted to evaluation of speed of transformation to be performed while modeling the statistical characteristics of system based on GPGPU complex of sequential -parallel data processing. A generalized task of coordinating a sequential-parallel process of statistical modeling for sequential steps of transformation of the initial information. The processes of coordination analyzed using Gantt charts. Terms of performance of modeling operations based on the operator conversion method of distribution laws of probabilities, that uses integral operators. In the article were synthesized and tested for adequacy generalized model of coordination of parallel processes and considered its application to image processing. The developed approach allows you to create criteria for the selection of the optimum size of the data blocks in parallel processing and to reduce the simulation time.

Keywords: statistical modeling, coordination model, sequential -parallel processing, the Gantt chart.

Введение

Статистическое моделирование является одним из основных способов системного анализа и проектирования разнообразных систем массового обслуживания, информационно-измерительных систем и т.п. [1]. Такое моделирование связано с необходимостью выполнения большого объема преобразований, зависящего от размерности моделируемой системы и требуемой точности моделирования. Повышение скорости выполнения таких преобразований в системах реального времени является важной научно-прикладной проблемой.

Актуальность

Для статистического моделирования используются программно-аппаратные системы. Наибольшее распространение получили комплексы на основе GPSS [2], система моделирования на основе специализированных аппаратных вычислителей [3], параллельные системы моделирования на платформе GP GPU [4]. Управление их работой приводит к необходимости решения задачи координации [5]. В литературе [6-8] рассматриваются методы координации применительно к управлению технологическими процессами. Однако не рассмотрена задача распределения ресурсов во взаимосвязи с задачей синхронизации параллельных процессов. Аналогом ресурсов при обработке изображений является массив данных изображения. Отсутствие общего решения задачи координации процессов статистического моделирования не позволяет сформулировать критерий для оптимального управления процессом моделирования, что обуславливает актуальность задачи оценки характеристик последовательно-параллельного процесса статистического моделирования с учетом процессов координации обработки данных.

Постановка задачи

Рассмотрим обобщенную задачу координации последовательно-параллельного процесса статистического моделирования для G последовательных шагов преобразования исходной информации (рис. 1).

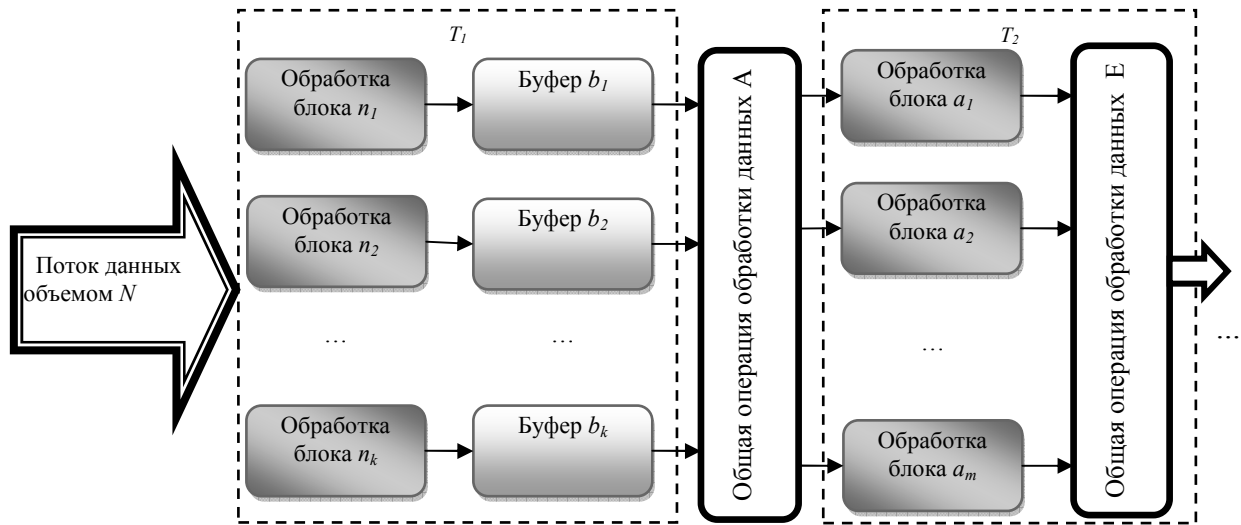


Рисунок 1 – Обобщенная структура последовательно-параллельной обработки данных

Определены правила выполнения операций моделирования. За основу определения этих операций принят операторный метод преобразования законов распределения вероятностей [3], использующий интегральные операторы вида

$$\beta_Y(y) = \Phi_{XY}(\beta_X(\bar{x}), A, W) = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \beta_X(\bar{x}) \varphi(x, y, A, W) d\bar{x},$$

где β - функция статистического распределения, φ - ядро преобразования, A – статистические моменты функции распределения, W – параметры преобразования.

Так преобразования для унарных нелинейных операций вида $y = N(x)$.

$$\varphi^{(1)}(x, y) = \delta[y - N(x)], \tag{1}$$

Для бинарных операций $y = N(x_1, x_2)$

$$\beta(y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \delta[y - N(x_1, x_2)] \beta(x_1) \beta(x_2) dx_1 dx_2. \tag{2}$$

Для интегро-дифференциальных преобразований, характеризуемых импульсной передаточной функцией $g(\tau)$:

$$\beta_Y(y) = \varphi^{(n)} \beta_X(x) = \int_{-\infty}^{+\infty} \beta_X(x_n - m_X^{(1)}(t)) \varphi^{(n)}(x_n, y, \omega) dx_n, \tag{3}$$

где $\varphi^{(n)}(x, y, \omega) = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \prod_{i=1}^{n-1} \beta_X(x_i - m_X^{(1)}(t)) \delta \left[y - (1-a)m_Y^{(1)} - a \sum_{i=1}^{n-1} x_{n-1}(t-i\tau) g(i\tau) \right] dx_1 \dots dx_{n-1};$

$$a = \sqrt{\frac{D_Y}{\Delta\tau D_X \sum_{i=1}^{n-1} g_0^2(i\Delta\tau)}}; \quad g_0(i\Delta\tau) = \frac{1}{\Delta\tau} \int_{i\Delta\tau}^{(i+1)\Delta\tau} g(\tau) d\tau; \quad \Delta\tau = \frac{G_{XX\max}}{D_X}; \quad n = \text{ent} \left[\frac{T_{\text{pr}}}{\Delta\tau} \right] + 1;$$

$$T_{pr} = \frac{\pi W_{max}}{\int_0^{\infty} W(\omega) d\omega}.$$

Значения функции распределения вероятностей поступают на вход схемы рис.1 параллельным потоком. Поток объемом N разделяется на кластеры количеством k . Каждый кластер информации $n_i, i = \overline{1..k}$, поступает на блок обработки информации, производительностью p_i . Каждому блоку соответствует буфер $b_i, i = \overline{1..k}$, максимальной вместимостью B_{max_i} . Информация после обработки поступает к общему блоку обработки данных A в определенной последовательности и обрабатывается с производительностью u_i . Обработанная информация поступает на параллельные блоки $a_j, j = \overline{1..m}$, и снова на блок обработки всей информации E .

Целью статьи является разработка подхода к оценке характеристик последовательно-параллельного процесса статистического моделирования с учетом процессов координации обработки данных. Для этого необходимо оценить временные характеристики системы моделирования при оптимальной синхронизации параллельных процессов.

Методы решения задачи

Для решения задачи представим операторные преобразования (1) – (3) в виде специализированных блоков, реализуемых на базе GP GPU.

Блок для нелинейного преобразования статистических данных, заданных в виде функций распределения, (рис. 2) состоит из коммутатора (К), блока развертки (БР), блока параллельного вычисления (БПВ) и запоминающего устройства (ЗУ). Функция распределения задается в виде n дискретных значений $\beta_X(x_i)$ и подается одновременно со значениями, которые вырабатывает блок развертки, на соответствующие входы коммутатора. Из коммутатора значения поступают в блок параллельного вычисления, где выполняется дифференцирование и умножение, полученный результат заносится в запоминающее устройство. При прохождении полного цикла моделирования на выходе запоминающего устройства воспроизводятся значения функции распределения после нелинейного преобразования в соответствии с формулой (1).

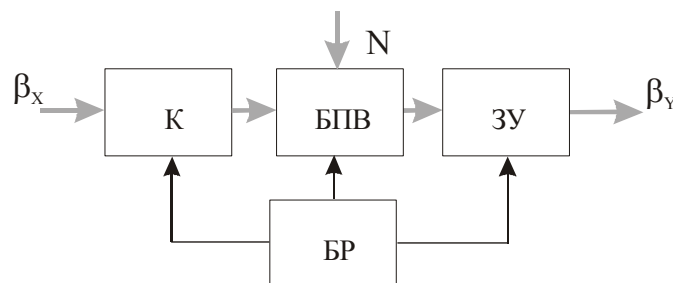


Рисунок 2 – Блок для нелинейного преобразования статистических данных, заданных в виде функций распределения

Структура блока суммирования статистических данных, заданных в виде функций распределения, представлена на рис.3, где обозначено: ГА – генератор аргумента, СС – схема совпадения, БЛП – блок линейного преобразования.

Блок работает следующим образом. Функции распределения задаются в виде n дискретных значений $\beta_{X1}(x_{i1}), \beta_{X2}(x_{i2})$ и подаются на соответствующие коммутаторы. На вход блока линейного преобразования поступает значение x_1 , вырабатываемое блоком развертки и формируется значение функции

$$a_1x_1 + a_0,$$

поступающее на управляющий вход схемы совпадения, где оно сравнивается с текущими значениями $(a_2y - x_2)$, которые формирует генератор аргумента. Таким образом, на коммутатор поступают значения, которые выбирают данные, индексы которых удовлетворяют соотношению

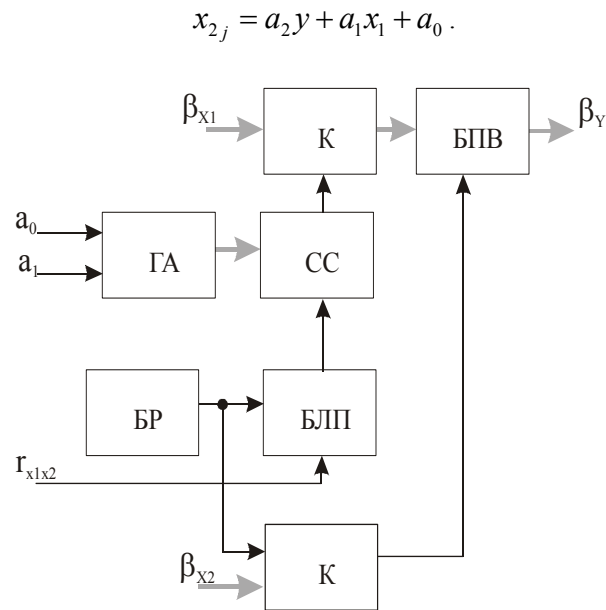


Рисунок 3 – Блок суммирования данных, заданных в виде функций распределения

Тогда на выходе коммутатора K формируются значения функции $\beta(a_2 y + a_1 x_1 + a_0)$, которые поступают на вход блока параллельного вычисления БПВ, на второй вход поступают значения функции $\alpha(x_{1j})$. В блоке параллельного вычисления выполняются операции умножения и интегрирования. При прохождении полного цикла преобразования на выходе БПВ формируются значения функции распределения суммы статистических данных β_Y в соответствии с формулой (2).

Блок для линейного интегро-дифференциального преобразования статистических данных, заданных в виде функций распределения, имеет структуру, представленную на рис.4

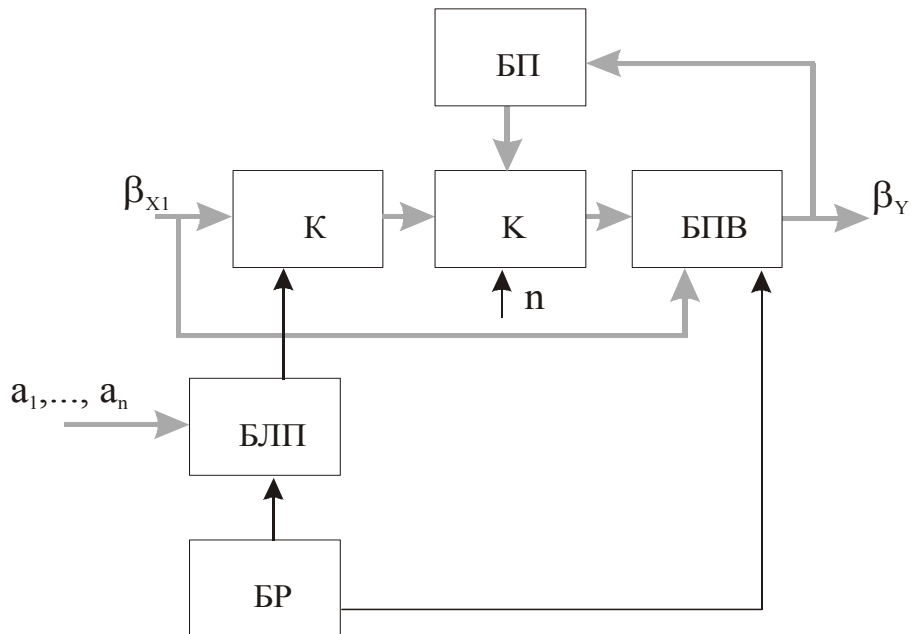


Рисунок 4 - Блок для линейного интегро-дифференциального преобразования данных, заданных в виде функций распределения

Функция распределения задается в виде n дискретных значений $\beta_X(x_i)$ и подается одновременно на входы коммутатора и блока параллельного вычисления. В блоке развертки формируются текущие

значения x и подаются на входы блока параллельного вычисления и блока линейного преобразования. В блок линейного преобразования также подаются значения постоянных коэффициентов $a_i, i = 0, \dots, n$, значения которых вычисляются в соответствии с формулой для ядра преобразования (3). На выходе блока линейного преобразования формируются значения

$$a_0 - a_n \sum_{i=1}^{n-1} x_i b_i - (a_0 - a_n) m_X$$

На управляющий вход второго коммутатора подается значение n , которое равно количеству циклов интегрирования. С выходов блока параллельного вычисления, где производится умножение и интегрирование, значения подаются в блок памяти. При прохождении $(n - 1)$ циклов интегрирования на выходе блока параллельного вычисления формируются функции распределения βY после линейного интегро-дифференциального преобразования в i -й времени. При прохождении полного цикла моделирования на выходе блока интегрирования формируются значения обобщающей функции βY после линейного интегро-дифференциального преобразования для дискретных значений t .

Таким образом, последовательное выполнение операций статистического моделирования сочетается с параллельным преобразованием данных на каждой операции. Между операциями происходит перемутация данных от каналов вывода предыдущей операции на каналы вывода последующей операции, которая выполняется последовательно. Такой конвейерный метод требует координации процессов поступления данных от одной операции к другой. Целью координации процессов является уменьшение общего времени моделирования T .

Модель координации процессов обработки данных покажем в виде диаграммы Ганта (рис. 5) [9].

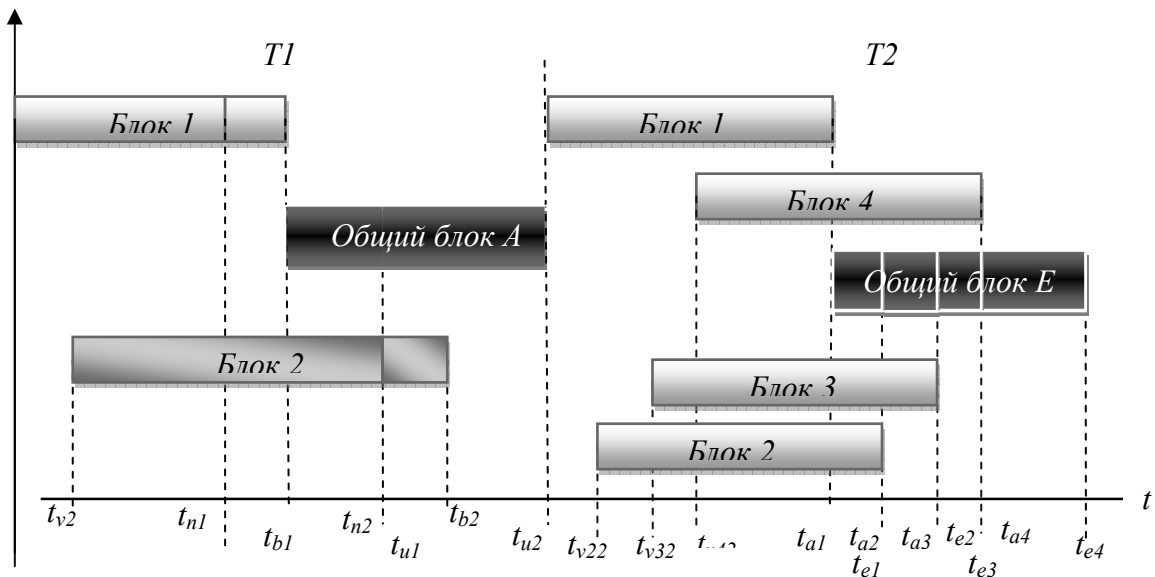


Рисунок 5 – Диаграмма Ганта для последовательно-параллельной обработки данных

Время работы системы состоит из суммы времени выполнения всех последовательных шагов:

$$T = T_1 + T_2 + \dots + T_G \tag{4}$$

Поскольку определен порядок поступления элементов на блок совместной обработки данных, то можно сказать, что первый блок обработает информацию раньше, а последний - позже. Тогда

$$T_1 = T_{1_k} + T_v, \tag{5}$$

где T_v – время простоя блока, которое зависит от порядка поступления соответствующего кластера в блок совместной обработки данных.

Очевидно

$$T_{1_i} = t_{n_i} + t_{b_i}, \quad (6)$$

где t_{n_i} - время работы блока обработки, которое зависит от производительности p_i этого блока и количества данных n_i

$$t_{n_i} = \frac{n_i}{p_{n_i}} \quad (7)$$

t_{b_i} - время буферизации данных в блоке памяти, представляет собой разницу между временем заполнения t_{nb_i} и временем отдачи данных t_{ab_i} .

Время заполнения буфера памяти и время вывода данных из буфера памяти:

$$t_{nb_i} = \frac{B_{\max_i}}{p_{n_i}}, \quad (8)$$

$$t_{ab_i} = \frac{B_{\max_i}}{u_{n_i} - p_{n_i}}, \quad (9)$$

где u_i - производительность обработки кластера данных n_i в блоке A .

Время нахождения данных в блоке памяти:

$$t_{b_i} = t_{nb_i} - t_{ab_i}. \quad (10)$$

Тогда T_{1_i} :

$$T_{1_i} = \frac{n_i}{p_{n_i}} + \left(\frac{B_{\max_{n_i}}}{p_{n_i}} - \frac{B_{\max_{n_i}}}{u_{n_i} - p_{n_i}} \right) = \frac{n_i(u_{n_i} - p_{n_i}) + B_{\max_{n_i}}(u_{n_i} + 2p_{n_i})}{p_{n_i}(u_{n_i} - p_{n_i})} \quad (11)$$

Время обработки всех данных в блоке A :

$$T_A = \frac{N \cdot \eta}{u_{n_i}}, \quad (12)$$

где N - общее количество данных, η - коэффициент сжатия при обработке данных.

Блок T_2 аналогичен T_1 , но с другим количеством данных и другим количеством кластеров, то есть

$$T_{2_j} = \frac{a_j}{p_{a_j}} + \left(\frac{B_{\max_{a_j}}}{p_{a_j}} - \frac{B_{\max_{a_j}}}{u_{a_j} - p_{a_j}} \right) = \frac{a_j(u_{a_j} - p_{a_j}) + B_{\max_{a_j}}(u_{a_j} + 2p_{a_j})}{p_{a_j}(u_{a_j} - p_{a_j})}, \quad (13)$$

где a_j - количество данных в кластере, j - порядковый номер кластера, $j = \overrightarrow{1..m}$.

Время обработки всех данных в блоке E :

$$T_E = \frac{A \cdot \eta_1}{u_{ai}} \quad (14)$$

где A - общее количество информации, η_1 - коэффициент сжатия.

Общее время T для двух аналогичных блоков будет состоять из суммы времен «медленных» блоков:

$$T = \left(\frac{n_k(u_{n_k} - p_{n_k}) + B_{\max_{n_k}}(u_{n_k} + 2p_{n_k})}{p_{n_k}(u_{n_k} - p_{n_k})} + T_l \right) + \frac{N \cdot \eta}{u_{ni}} + \left(\frac{a_m(u_{am} - p_{am}) + B_{\max_{am}}(u_{am} + 2p_{am})}{p_{am}(u_{am} - p_{am})} + T_l \right) + \frac{A \cdot \eta_1}{u_{ai}} \quad (15)$$

Для проверки адекватности модели (15) построим частные зависимости $t_{n_i} = f(n_i)$, $t_{n_i} = f(p_{n_i})$, $t_{b_i} = f(B_{\max_{ni}})$, $t_{b_i} = f(p_{n_i})$, $t_{b_i} = f(u_{n_i})$, получаемые из (15) фиксации остальных переменных. Параметры зависимостей пронумерованы и приведены к диапазону (0, 100).

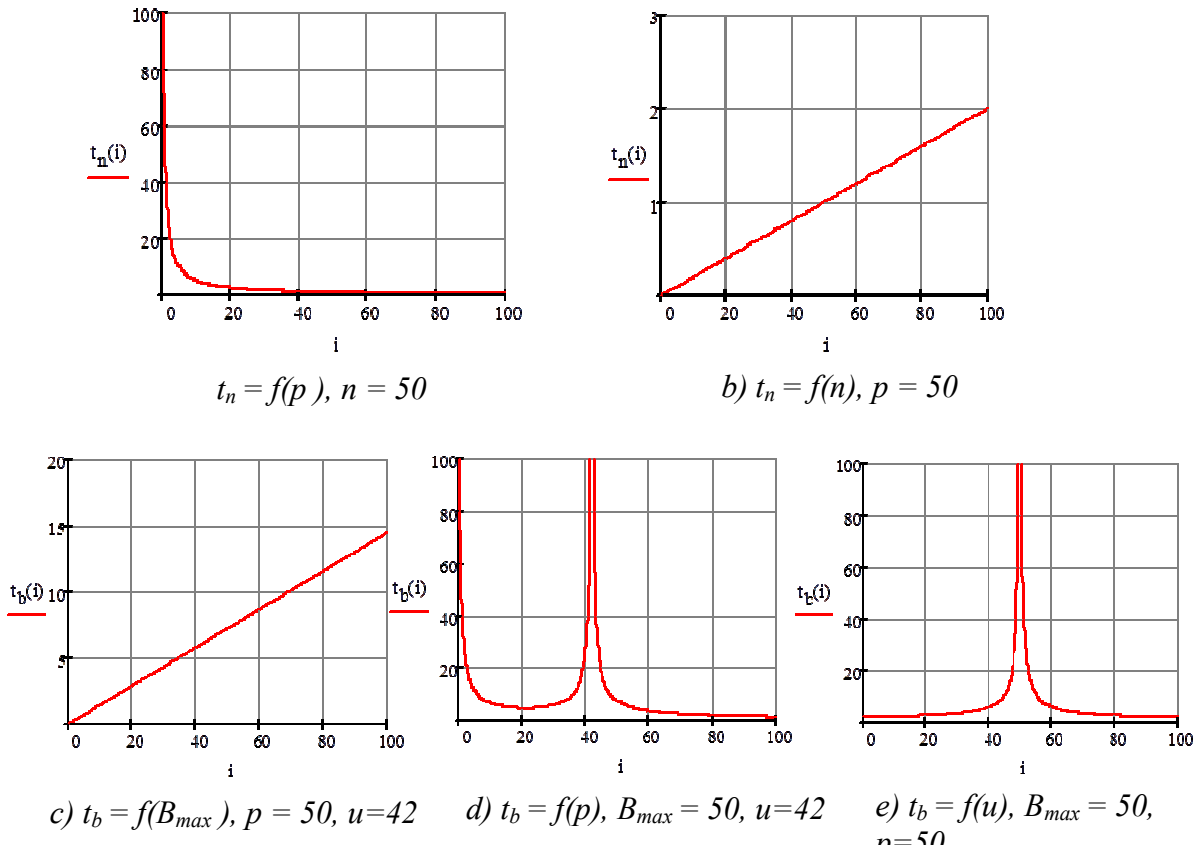


Рисунок 6 – Зависимости $t_{n_i} = f(n_i)$, $t_{n_i} = f(p_{n_i})$, $t_{b_i} = f(B_{\max_{ni}})$,
 $t_{b_i} = f(p_{n_i})$, $t_{b_i} = f(u_{n_i})$

Анализ зависимостей (рис. 6) показывает соответствие расчетной и имитационной модели. Так, рис. 6, а показывает убывание времени выполнения операции с ростом производительности, а рис. 6, б и рис. 6, с показывают рост времени выполнения отдельной операции от количества данных и размера буфера. Наиболее интересна и нетривиальна зависимость на рис. 6, d, которая показывает наличие минимума времени буферизации при изменении производительности обработки и размера блоков данных. Оптимизация размера блоков позволит сократить общее время моделирования.

Выводы

Разработанный подход к оценке характеристик последовательно-параллельного процесса статистического моделирования с учетом процессов координации обработки данных позволяет сформировать критерий для выбора оптимального размера блоков данных при параллельной обработке и сократить время моделирования.

Список литературы

1. Лифшиц А.Л. Э.А. Статистическое моделирование систем массового обслуживания / Лифшиц А.Л., Мальц Э.А. – М.: Советское радио. – 1978. – 248 с.
2. Общецелевая система моделирования GPSS World / Exponenta.ru (Электронный документ) Режим доступа: <http://www.exponenta.ru/soft/others/gpss/gpss.asp>
3. Глонь О.В. Моделювання систем керування в умовах невизначеності. Монографія./ Глонь О.В., Дубовой В. М. – Вінниця: УНІВЕРСУМ – Вінниця, 2003. – 174 с.
4. What is GPU accelerated computing / NVIDIA (Электронный документ) Режим доступа: <http://www.nvidia.com/object/what-is-gpu-computing.html>
5. Woodside C. M. Fast Allocation of Processes in Distributed and Parallel Systems / C. M. Woodside, and G. G. Monforton // IEEE Transactions on parallel and distributed systems . – 1993. – P. 164-174.
6. Katrenko A.V. Mechanisms of coordination in complex hierarchical systems. / A.V. Katrenko I.V. Savka - 2008 - URL: http://vlp.com.ua/files/16_1.pdf
7. Ladanyuk A.P. System task coordination continuous technological complexes type / A.P. Ladanyuk, D.A. Shumyhay, R.O. Boyko - URL:http://dspace.nuft.edu.ua/jspui/bitstream/123456789/4444/1/Sh_3.pdf.
8. Mesarovic, M. D. Theory of Hierarchical Multilevel Systems / M. D. Mesarovic, D. Macko, Y. Takahara,. — Published by Academic Press, Inc., 1972. – 344 p.
9. Прийняття рішень в управлінні розгалуженими технологічними процесами : [монографія] / В. М. Дубовой, Г. Ю. Дерман, І. В. Пилипенко, М. М. Байас. — Вінниця : ВНТУ, 2013. — 223 с.

Данные про авторов

Дубовой Владимир Михайлович - д.т.н., проф., зав. кафедрой компьютерных систем управления, Факультет компьютерных систем и автоматики, Винницкий национальный технический университет.

Никитенко Елена Дмитриевна - к.т.н., доцент кафедры компьютерных систем управления, Факультет компьютерных систем и автоматики, Винницкий национальный технический университет.