

ФОРМАЛИЗОВАННЫЙ СИНТЕЗ ВАРИАНТОВ ОРГАНИЗАЦИИ ПАРАЛЛЕЛЬНЫХ ВЫЧИСЛЕНИЙ ДЛЯ ВЕКТОРНО-МАТРИЧНЫХ ОПЕРАЦИЙ И ПРОБЛЕМА ВЫБОРА АЛГОРИТМА¹

Рассматривается проблема создания алгоритмов, изначально ориентированных на параллельную обработку. Предлагается решение этой проблемы для векторно-матричных операций, позволяющее синтезировать для каждой операции данного класса множество вариантов организации вычислений. По сути, найденное решение является моделью, порождающей требуемое для параллельной обработки разнообразие пространственно-временных алгоритмов. Это разнообразие представляется в виде пространства вариантов организации параллельных вычислений - алгоритмов, характеризующихся изначальной независимостью от архитектуры вычислительных систем. Поэтому множество синтезируемых вариантов и формальные правила их образования создают основу для согласования алгоритмических и архитектурных характеристик параллельных вычислительных систем, в ходе которого решается проблема выбора алгоритма.

FORMALIZED SYNTHESIS OF VARIANTS OF PARALLEL COMPUTATION ORGANIZATION FOR VECTOR-MATRIX OPERATIONS AND ALGORITHM CHOICE PROBLEM / O.V. Klimova (Institute of Engineering Science, 34 Komsomolskaya, Ekaterinburg 620219, Russia, E-mail: klimova@imach.uran.ru). Formalized synthesis of variants of parallel computation organization for vector-matrix operations is considered in this paper. The developed formal model of the parallel computation organization is the basis of the synthesis. The set of synthesized variants and formal instructions of their creation allow one to solve the algorithm choice problem.

1. Введение

Проблема эффективности параллельных вычислений является актуальной особенно в свете все нарастающих темпов внедрения параллельных принципов обработки данных. Путь повышения эффективности – согласование алгоритмических и архитектурных характеристик параллельных вычислительных систем, реализуемое в ходе совместных исследований алгоритмов и архитектур. Объектом исследований на этом пути

¹Настоящая работа выполнена при частичной поддержке РФФИ (грант № 11-01-12126) и программы Президиума РАН № 18 "Алгоритмы и математическое обеспечение для вычислительных систем сверхвысокой производительности" при поддержке УрО РАН (проект 12-П-1-1028).

является организация вычислений, характеризующаяся множеством различных вариантов. Целью исследований является обоснованный выбор варианта организации вычислений, обеспечивающий эффективность обработки в заданных условиях её реализации. Основной задачей таких исследований является совместное изучение алгоритмов и архитектур. Однако для обеспечения эффективности и возможности реализации таких исследований, прежде всего, необходимо описать объект исследований в среде его изучения, то есть в пространственно-временной среде. Это означает, что требуемое описание должно порождать разнообразие вариантов организации вычислений, должно быть формальным и архитектурно независимым. Наличие такого описания – формального инструмента позволит синтезировать множество параллельных алгоритмов – вариантов организации вычислений для выбранной операции и выполнять совместные исследования алгоритмов и архитектур. Таким формальным инструментом должна стать модель организации параллельных вычислений. Но как построить такую модель? Сложность этой задачи была отмечена многими исследователями [1-3]. Путь реструктуризации [1-2] известных последовательных алгоритмов и их программ, направленной на построение параллельных алгоритмов, не привел к формированию общего описания различных вариантов организации вычислений. Более того, этот путь ориентирован на адаптацию структур последовательных алгоритмов к конкретным условиям их параллельной обработки. Поэтому результаты, полученные на этом пути, не могут быть рассмотрены в качестве основы для реализации совместных исследований алгоритмов и архитектур. В то же время, во многих современных работах [4,5] подчеркивается важность и актуальность таких исследований, а также приводятся результаты, позволяющие реализовать эти исследования на основе найденных приемов описания разнообразия параллельных алгоритмов. Однако, известных результатов, относящихся к области разработки моделей организации параллельных вычислений, чрезвычайно недостаточно. Очевидно, что такие модели не могут быть универсальными, они должны разрабатываться для выделенного класса операций. Для создания модели необходимо установить общий закон, позволяющий выявлять внутренний параллелизм последовательных алгоритмов и представлять их в виде композиционных форм. Компонентами этих форм должны стать независимые вычислительные процессы. Именно такой путь предлагается использовать в работе [6] для построения параллельных алгоритмов изначально ориентированных на параллельную обработку. Основная сложность реализации этого пути заключается в установлении требуемого закона, приводящего к образованию композиционных форм. Обнаружение такого закона позволит установить связь между последовательной и параллельной обработкой данных. В свою очередь, именно эта связь позволит осуществить естественный переход от алгоритма к модели. Такие модели станут теми формальными инструментами, которые будут порождать алгоритмы (варианты организации вычислений), изначально ориентированные на параллельную обработку. С помощью таких моделей можно будет синтезировать пространство решений, задаваемое параметрами модели и отношениями их значений.

Итак, формальные инструменты с вышеуказанными возможностями необходимы для проведения совместных исследований алгоритмов и архитектур. Именно с помощью таких исследований можно лучше понять – что мы делаем, создавая параллельные архитектуры? Именно от исследований этого вопроса зависит эффективность использования параллельной обработки. Поэтому решения, получаемые в ходе таких исследований, должны стать основой для создания перспективной вычислительной техники. Базой же для проведения таких исследований должны стать формальные инструменты – модели организации параллельных вычислений. В работах [7-11] были представлены результаты, которые привели к созданию модели организации параллельных вычисле-

ний для операций цифровой обработки сигналов (дискретного преобразования Фурье (ДПФ), свертки, корреляции, различных ортогональных преобразований). Модель также применима для операций перемножения матриц и умножения матрицы на вектор, причем, размерность всех операций может быть любой.

В предлагаемом докладе представляются возможности модели, связанные с синтезом различных вариантов организации вычислений для широкого класса векторно-матричных операций, а также иллюстрируется решение проблемы выбора параллельного алгоритма с помощью созданного формального инструмента.

2. Формальная модель синтеза вариантов организации вычислений

Основными характеристиками построенной модели являются её архитектурная независимость и способность порождать разнообразие вариантов организации вычислений для множества конкретных операций ЦОС. Наличие этих характеристик у созданного формального инструмента позволяет сказать о его изначальной ориентации на описание вычислений, протекающих в пространственно-временной среде. Такая ориентация модели свидетельствует о её способности к решению вопросов согласования алгоритмических и архитектурных характеристик вычислительных систем путем выбора вариантов организации параллельных вычислений в заданных условиях их реализации.

Рассмотрим требования, которые были предъявлены к модели, способной сделать указанный выбор обоснованным. Итак, модель должна обеспечивать:

- преимущество описания параллельных и последовательных вычислений;
- масштабируемость вычислений;
- разнообразие вычислительных структур;
- децентрализацию и модульность вычислений;
- детерминизм описания вычислительных компонент и законов их координации;
- синтез вариантов организации параллельных вычислений.

Созданная модель удовлетворяет всем вышеуказанным требованиям. Модель была построена путем выполнения преобразований вычислительных сред, порождаемых операциями ЦОС. Основными из этих операций являются циклическая свертка длины N

$$C(t) = \sum_{q=0}^{N-1} x(t-q)y(q)$$

и N – точечное ДПФ входной последовательности $x(t)$

$$S(\alpha) = \sum_{t=0}^{N-1} x(t)W_N^{t\alpha},$$

где

$$W_N = e^{-2\pi i / N}.$$

Вычислительные среды представленных операций, как и многих иных операций ЦОС, характеризуются выполнением N^2 операций умножения и $N \cdot (N - 1)$ операций сложения, реализуемых в едином вычислительном процессе над отсчетами входной последовательности $x(t)$. Суть выполненных преобразований – формирование многопроцессных описаний организации вычислений рассматриваемых операций в пространственно-временной среде. Вместо одного вычислительного процесса, обрабатывающего входную последовательность $x(t)$, было получено N_1 независимых вычислительных

процессов, предполагающих одновременную обработку укороченных последовательностей данных $x_j(t_j)$, полученных из исходной последовательности $x(t)$, где

$$N = N_l \cdot h_l, \quad j = 0, \dots, N_l \quad \text{и} \quad t_l = 0, \dots, h_l.$$

Для выполнения перехода от однопроцессного описания вычислений к многопроцессному, по сути, означающего переход от алгоритма к модели, были введены и реализованы два следующих этапа эквивалентных преобразований:

- этап 1 – декомпозиция данных;
- этап 2 – декомпозиция операций ЦОС.

Особенностью первого этапа является его функциональный характер, означающий представление входной последовательности $x(t)$ с помощью найденной композиционной формы

$$x(t) = \sum_{j=0}^{N_l-1} x_j^*(t-j),$$

позволивший выполнить второй этап и получить различные композиционные формы для операций ЦОС. Таким образом, в основе создания требуемого формального инструмента лежит декомпозиция данных, позволившая наделить порождаемую ею модель организации параллельных вычислений свойствами, перечисленными в начале данного раздела. В основе декомпозиции данных лежат две операции, выполняемые во временной области над отсчетами входной последовательности. Это операции сдвига и масштабирования. Проиллюстрируем суть указанных операций. Пусть $t = t_l \cdot N_l + j$, тогда $t_l = (t - j) / N_l = ((t_l \cdot N_l + j) - j) / N_l$, отсчеты входной последовательности $x(t)$ в точках $t = t_l \cdot N_l + j$ сдвигаются на $j = 0, \dots, N_l$ позиций влево, а затем масштабируются по времени, где N_l коэффициент масштабирования. В результате этих действий формируются N_l наборов данных $x_j(t_l)$, порождающих N_l независимых процессов, предполагающих одновременную обработку. На рисунке 1 представлены варианты структур входной последовательности данных при $N=16$. Применение указанных операций сдвига и масштабирования к приведенным на рисунке 1 наборам данных позволит сформировать входные последовательности $x_j(t_l)$, соответствующие пространственно-временной обработке.

В процессе реализации второго этапа – декомпозиции операций ЦОС, обеспечивающего процесс перехода от алгоритма к модели, были получены их различные композиционные формы [7-9], порождающие множество вариантов организации вычислений. Каждый элемент этого множества определяется значением параметра N_l , $N_l = N / h_l$. Таким образом, меняя значения N_l , можно синтезировать различные варианты организации вычислений для различных векторно-матричных операций. Обобщая результаты, полученные в процессе такого синтеза, можно отметить единство их вычислительных структур.

Так схемы, представленные на рисунке 2 и иллюстрирующие алгоритм свертки и модель организации её параллельных вычислений, могут быть легко преобразованы в схемы, представляющие вычисления ДПФ. При этом, основным параметром, характеризующим вариант организации вычислений, является параметр N_l . Так при $N_l = 1$ созданная модель превратится в исходный последовательный алгоритм, в то время, как множество иных возможных значений этого параметра будет поддерживать возмож-

ность синтеза различных вариантов организации вычислений в пространственно-временной среде. При $N = 2^n$ число таких вариантов определяется как $(n-1)$.

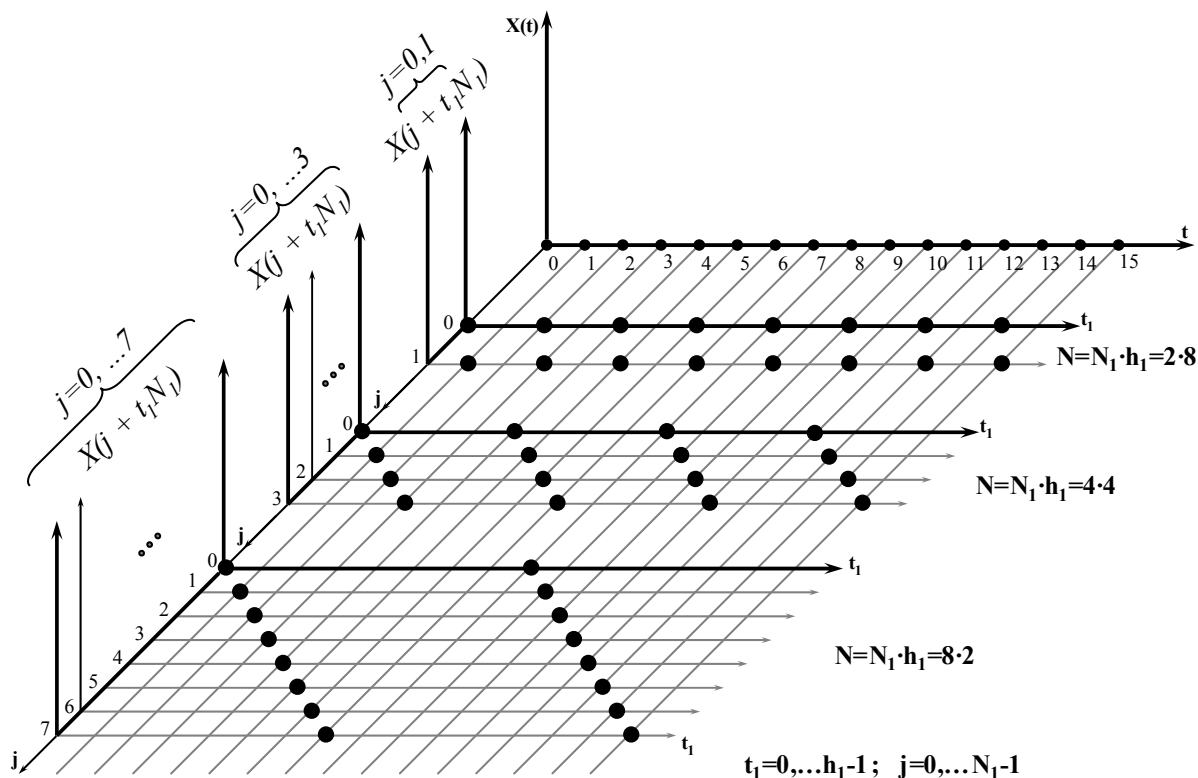


Рис. 1. Варианты структур входной последовательности данных

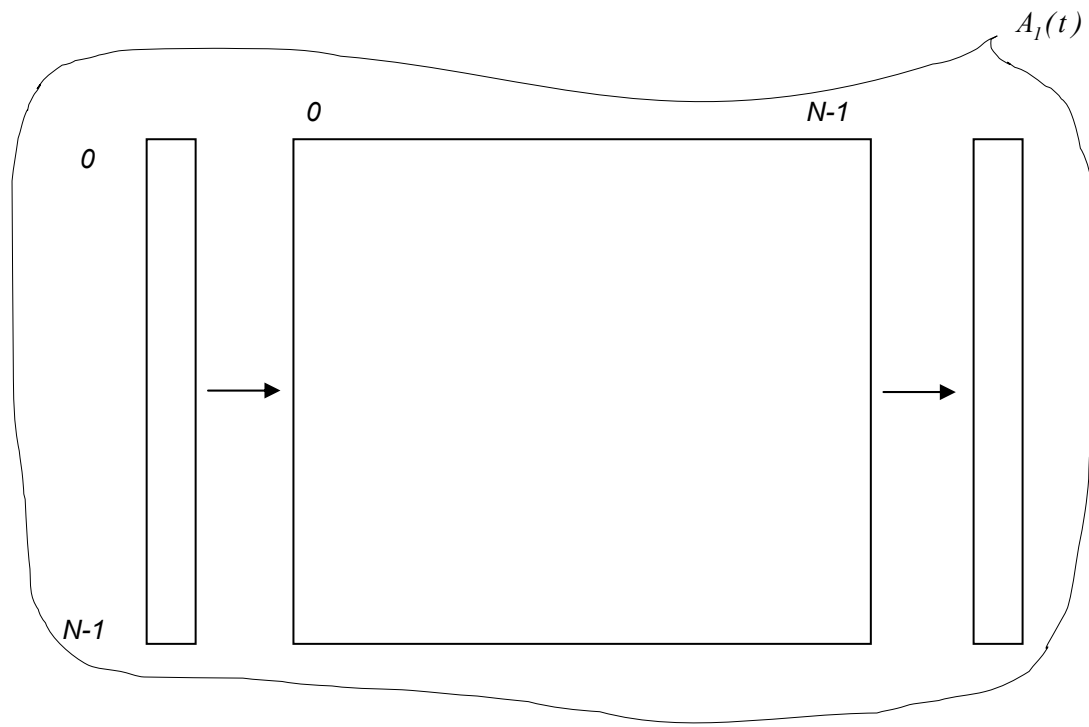
Таким образом, созданная модель порождает требуемое разнообразие вариантов организации параллельных вычислений, является архитектурно независимой и отвечает предъявляемым к ней вышеперечисленным требованиям. На самом деле параметрами модели также являются: вид операции; метод декомпозиции; метод вычислений; вид алгоритма; размерность операции. Однако, для того чтобы на общем уровне продемонстрировать реализацию обоснованного выбора на основе предложенной модели можно использовать лишь её ключевой параметр N_1 .

3. Модель и проблема выбора варианта параллельного алгоритма

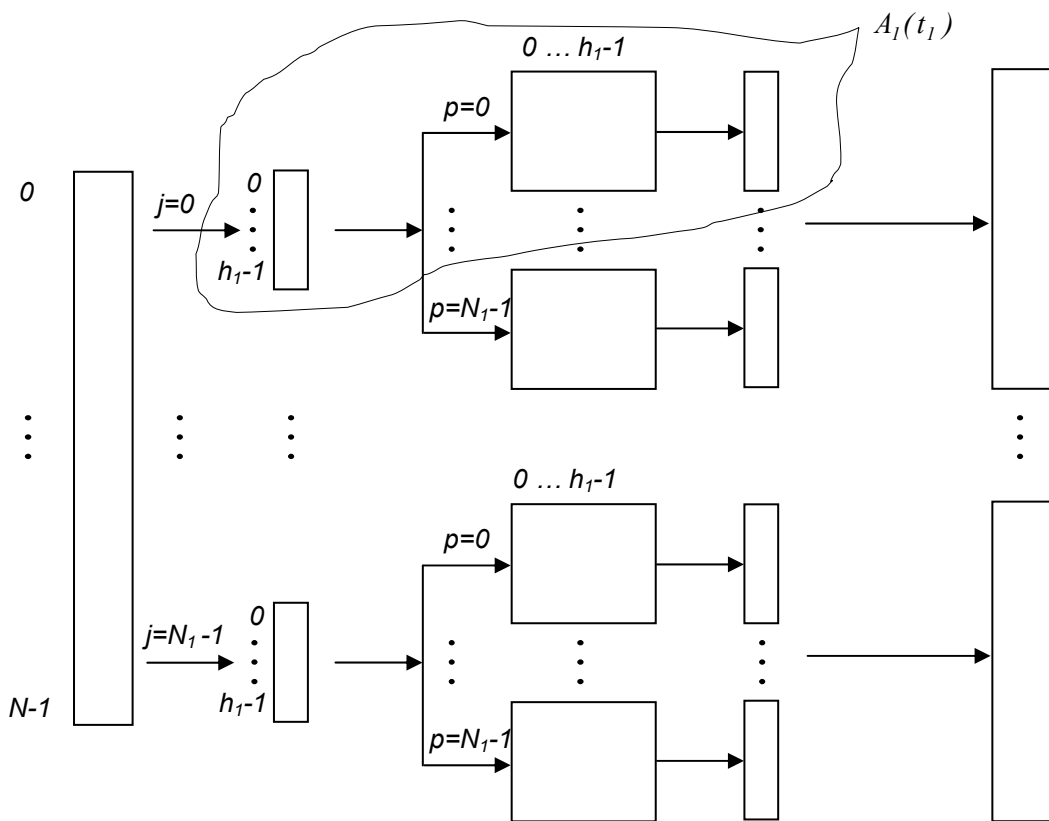
Формирование пространства решений с помощью разработанной модели обеспечивает реализацию синтеза вариантов организации вычислений, управление изменениями структуры вычислений, обоснованный выбор параллельных алгоритмов и проектирование их вариантов. Реализация этих шагов позволит согласовать алгоритмические и архитектурные параметры, а следовательно, позволит повысить эффективность параллельных вычислений. Продемонстрируем возможности модели, основанные на изменении значений параметра N_1 и направленные на реализацию обоснованного выбора параллельных алгоритмов.

Прежде всего, рассмотрим вариант алгоритма с максимально возможным параллелизмом, для этого варианта $N_1 = 2^{n-1}$ и $h_1=2$ при $N = 2^n$. Проведем оценку выигрыша

по быстрдействию от использования такого варианта для алгоритма ДПФ при условии аппаратной поддержки максимального параллелизма.



а



б

Рис. 2. Обобщенные схемы представления алгоритма (а) и модели (б)

Трудоёмкость алгоритмов оценим с помощью числа операций, необходимых для их реализации. Тогда трудоёмкость последовательного алгоритма можно оценить величиной $N^2 + N(N-1)$ операций умножения и сложения соответственно. Будем считать равноценной сложность выполнения операций сложения и умножения. Тогда трудоёмкость одной j -той ветки параллельного алгоритма можно оценить величиной $h_j^2 + h_j(h_j-1) + N$. Чем меньше h_j , тем меньше сложность ветки. Для $h_j=2$ эта трудоёмкость минимальная, её формула отличается от общей, так как в ней нет операций умножения, а сложений 2. Тогда выигрыш от использования параллельного алгоритма с максимальным параллелизмом можно оценить следующим образом: $(N^2 + N(N-1)) / (2+N) = N_j(2N-1) / (N_j+1)$. Такой выигрыш по времени возможен при одновременной реализации вычислений в $N_j=N/2$ вычислительных процессах.

Однако в реальных ситуациях существуют жесткие ограничения на одновременную обработку. Исходя из этого, доопределим модель, введя в неё параметр k , показывающий число одновременно обрабатываемых процессов в рассматриваемой вычислительной системе. При этом $k \leq N_j$, в общем случае обработка различных процессов будет параллельно-конвейерной, а величина N_j / k определит глубину конвейера. Общая трудоёмкость параллельно-конвейерных вычислений будет определяться величиной $(h_j^2 + h_j(h_j-1) + N) N_j / k$, а выигрыш от использования такой обработки данных можно представить следующим образом

$$(N^2 + N(N-1)) k / (h_j^2 + h_j(h_j-1) + N) N_j = k(2N-1) / (2h_j + N_j - 1).$$

Тогда, для различных значений параметра k необходимо выбирать алгоритмы, характеризующиеся такими соотношениями значений параметров (h_j, N_j) , для которых выражение $(2h_j + N_j - 1)$ будет принимать минимальные значения. Выбор варианта параллельного алгоритма, способного обеспечить наилучшую эффективность параллельной обработки, сводится к выбору наилучшего соотношения между параметрами (h_j, N_j) . Таким образом, пользуясь разработанным формальным инструментом можно выполнять обоснованный выбор вариантов организации вычислений, обеспечивая тем самым эффективность параллельной обработки данных.

4. Заключение

Создание формальной архитектурно-независимой модели организации вычислений для векторно-матричных операций позволило синтезировать и анализировать множество возможных вариантов организации вычислений в пространственно-временной среде. Одной из основных возможностей модели является выполнение обоснованного выбора параллельного алгоритма. Реализация этой возможности позволяет повысить эффективность параллельной обработки за счет согласования архитектурных и алгоритмических характеристик параллельной вычислительной системы. Показаны пути использования этой возможности. Для выполнения согласования алгоритмических и архитектурных характеристик в модель необходимо ввести параметры архитектуры, то есть доопределить модель. Далее необходимо установить функциональную зависимость трудоёмкости реализации вычислений, как от алгоритмических, так и от архитектурных параметров. Выбор алгоритма осуществляется на основе полученной функциональной зависимости с учетом заданных ограничений на аппаратные и временные затраты при реализации вычислений. Таким образом, выбор основан на проектировании вариантов, возможность реализации такого проектирования обеспечивается свойствами разработанной модели.

СПИСОК ЛИТЕРАТУРЫ

1. *Воеводин В.В.* Вычислительная математика и структура алгоритмов. М.: Изд-во МГУ, 2006.
2. *Воеводин В.В., Воеводин Вл.В.* Параллельные вычисления. СПб.: БХВ – Петербург, 2002.
3. *Кун С.* Матричные процессоры на СБИС. М.: Мир, 1991.
4. *G.G. Lee, Y.K. Chen, M. Mattavelli, and E.S. Jang,* Algorithm/Architecture Co-Exploration of Visual Computing: Overview and Future Perspectives // IEEE Trans. Circuits and Systems for Video Technology, vol. 19, no. 11, pp. 1576-1587, Nov. 2009.
5. *Gwo Giun (Chris) Lee, He-Yuan Lin, Chun-Fu Chen, and Tsung-Yuan Huang,* Quantifying Intrinsic Parallelism Using Linear Algebra for Algorithm/Architecture Coexploration // IEEE Trans. on Parallel and Distributed Systems, vol. 23, no. 5, pp. 944–957, May 2012.
6. *Edward A. Lee.* The Problem with Threads // IEEE Computer, v. 39, № 5, pp. 33-42, may 2006.
7. *Климова О.В.* Единый подход к построению быстрых алгоритмов и распараллеливанию вычислений дискретного преобразования Фурье // Изв. РАН. Теория и системы управления. - 1999. - №3. – С. 68-75.
8. *Климова О.В.* Параллельная архитектура процессора свертки произвольной длины с использованием числовых преобразований Рейдера // Изв. РАН. Техн. кибернетика. - 1994. - №2. – С. 183-191.
9. *Klimova O.* Decomposition on a Group and Parallel Convolution and Fast Fourier Transform Algorithms. Parallel Computing Technologies 4th International Conference, PaCT-97 // Proceedings, p.358-363 Springer-Verlag, Berlin, LNCS1277.
10. *Климова О.В.* Способы управления изменениями структуры параллельных алгоритмов цифровой обработки сигналов // Параллельные вычисления и задачи управления PACO'2008: труды IV международной конференции, г. Москва, 27–29 октября 2008 г. – Институт проблем управления им. В.А. Трапезникова РАН. С. 1033–1041.
11. *Климова О.В.* Эволюция способов организации вычислений для операций цифровой обработки сигналов: от алгоритма к модели // Вестник Томского Государственного Университета. Управление, вычислительная техника и информатика. 2011. № 1.С. 31-38.