

УДК 681.3

Э.И. Ватутин, канд. техн. наук, доцент, кафедра вычислительной техники, ЮЗГУ (e-mail: evatutin@rambler.ru)

В.С. Титов, д-р техн. наук, профессор, зав. кафедрой вычислительной техники, ЮЗГУ (e-mail: titov-kstu@rambler.ru)

СРАВНЕНИЕ МЕТОДОВ СИНТЕЗА РАЗБИЕНИЙ ГРАФ-СХЕМ ПАРАЛЛЕЛЬНЫХ АЛГОРИТМОВ С ИСПОЛЬЗОВАНИЕМ ДВУМЕРНЫХ ДИАГРАММ

Приведены результаты вычислительных экспериментов, целью которых являлось сравнение эвристических методов синтеза разбиений параллельных алгоритмов логического управления при различных значениях технологических ограничений и размера граф-схем алгоритмов. Эксперименты выполнены с использованием грид-системы BOINC в рамках проекта Gerasim@home. Полученные двумерные карты уточняют и дополняют сформулированные ранее положения об областях преимущества методов синтеза разбиений.

Ключевые слова: система логического управления, проектирование логических мультиконтроллеров, разбиения, параллельный алгоритм, дискретная комбинаторная оптимизация, BOINC, Gerasim@home

Одним из перспективных подходов к синтезу многомодульных однородных систем логического управления (СЛУ) является их реализация в базе логических мультиконтроллеров (ЛМК) [1, 2]. При этом возникает ряд задач дискретной оптимизации, одной из которых является задача отыскания субоптимального разбиения априорно известной граф-схемы параллельного алгоритма логического управления. Для ее решения известны [1–4] различные эвристические методы, существенно отличающиеся по ряду ключевых характеристик, к которым в первую очередь относятся трудоемкость реализации, асимптотическая временная и емкостная сложности, состав оптимизируемых частных показателей качества и интегральное качество получаемых решений.

Общепринятой стратегией сравнения эвристических методов является использование априорно заданной выборки $\Lambda = \{G_1^0, G_2^0, \dots, G_K^0\}$ из K тестовых примеров G_k^0 , $k = \overline{1, K}$ (граф-схем $G = \langle A, V \rangle$ алгоритмов в данном случае, где A – множество вершин и V – множество дуг) с целью выяснения средних значений оптимизируемых показателей качества (и, соответственно, доверительных интервалов) и построение кривых полученного распределения, показывающих, в скольких случаях качество оптимизации выбранного критерия x ухудшается на заданную величину $\Delta\gamma_x$ (в абсолютных или относительных единицах). Подобное сравнение методов приведено в работе [5] (см. рис. 1 в), временные затраты на выполнение подобного эксперимента составляют величину порядка нескольких минут в расчете на один эксперимент.

Спецификой рассматриваемой задачи является наличие существенно-го влияния на качество оптимизации частных критериев не только «внутренних» числовых параметров (например, коэффициентов весовых или штрафных функций) и алгоритмических особенностей (например, выполнение тех или иных вспомогательных преобразований) методов синтеза разбиений, но и значений технологических ограничений X_{\max} (ограничение на число принимаемых контроллером сигналов логических условий), Y_{\max} (ограничение на число выдаваемых контроллером сигналов микроопераций) и W_{\max} (ограничение на емкость памяти контроллера), диктуемых элементной базой, и размера N граф-схемы алгоритма управления, определяемого числом вершин в ее составе. Учитывая возможность беспрепятственного обхода ограничения Y_{\max} путем дублирования мультиконтроллера в СЛУ [1–4], фактически задача исследования поведения эвристических методов синтеза разбиений сводится к исследованию различных областей пространства параметров $\{X_{\max}, W_{\max}, N\}$. В работе [6] было выполнено частичное его исследование путем изменения одного из параметров в заданном диапазоне при неизменных значениях остальных (см. рис. 1 г, д), в результате чего были получены одномерные графики зависимостей γ_x (среднее значение критерия x) и ρ_x (вероятность получения решения с наименьшим значением минимизируемого критерия x) от изменяемого параметра $y \in \{X_{\max}, W_{\max}, N\}$. Временные затраты для получения подобных графиков зависимостей составляют величину порядка нескольких часов процессорного времени в расчете на один эксперимент. Для полученных графиков была выявлена зонная структура: в различных областях значений ограничений (слабые, сильные, очень сильные) методы демонстрировали различное качество решений, не позволяя выявить однозначно лучшего.

В данной работе приведены новые результаты сравнения методов синтеза разбиений с использованием двухпараметрических диаграмм (карт), показывающих зависимость значений γ_x и ρ_x от пары изменяемых параметров y_1 и $y_2 \in \{X_{\max}, W_{\max}, N\}$, $y_1 \neq y_2$ (рис. 1 б).

Построение подобных карт является достаточно ресурсоемким процессом (в особенности с ростом N), однако поставленная задача является слабосвязанной и, следовательно, может быть эффективно распараллелена.

1. Отдельные элементы пространства параметров (точки на карте), являющиеся входными параметрами для псевдослучайного генератора граф-схем алгоритмов и методов синтеза разбиений, могут быть обработаны параллельно.
2. Отдельные разбиения $Sep_{F_i}(A_j^0)$ в составе выборки граф-схем алгоритмов Λ , соответствующей каждой из точек на карте, могут быть построены параллельно (F_i – i -й метод синтеза разбиений).

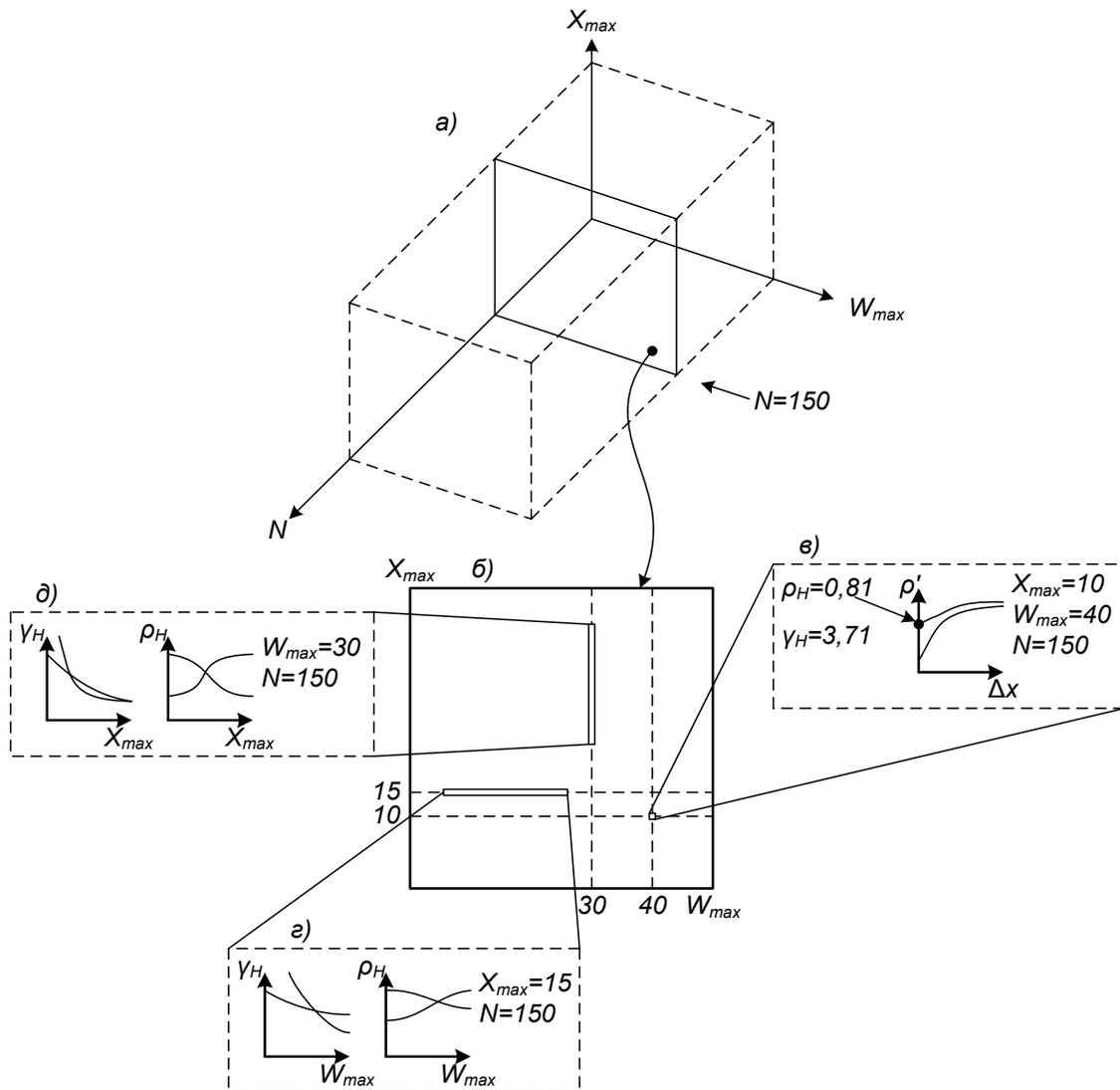


Рис. 1. Пространство параметров (а), карта в разрезе параметров (X_{max} , W_{max}) (б) и ранние вычислительные эксперименты (в, г, д)

Последовательная часть расчетов (постобработка) включает в себя расчет значений γ_x и ρ_x для полученных в ходе вычислительных экспериментов выборок значений критериев качества.

В качестве инструментария для организации параллельных вычислений использованы технические возможности платформы добровольных распределенных вычислений BOINC (Berkeley Open Infrastructure for Network Computing) [7], активно используемой в ряде широко известных проектов распределенных вычислений (SETI@Home, Einstein@home, LHC@Home и др.) [8]. Инфраструктура BOINC представляет собой распределенную грид-систему из более чем 6,5 млн. существенно неоднородных, географически удаленных вычислительных узлов, поддерживаемых более чем 2 млн. добровольцев, характеризующуюся суммарной интегральной производительностью порядка 5–6 PFLOPS [9].

Учитывая особенности и рекомендации к организации процесса вычислений на данной платформе, вычисления были распараллелены на уровне координат исследуемого пространства параметров (первая стратегия среди представленных выше), каждая из выборок граф-схем алгоритмов обрабатывалась последовательно в рамках одного расчетного задания, время выполнения которого составляло от нескольких секунд (область малых значений N) до 5–6 часов (область больших значений N). Использование потенциала многоядерных процессоров и многопроцессорных машин при этом реализовано с использованием возможностей клиентской программы BOINC Manager путем параллельного выполнения нескольких заданий с использованием SIMD-принципа (в данном случае, один расчетный модуль, запускаемый с разными входными параметрами) с близкой к 100% загрузкой всех доступных процессорных ядер, что весьма удобно в данном случае, т.к. позволяет не производить более трудоемкое и менее эффективное в задаче исследования пространства параметров ручное распараллеливание расчетного кода методов синтеза разбиений.

Серверная часть кода, реализованная в рамках проекта распределенных вычислений Gerasim@home [10], выполняет функции управления выдачей расчетных заданий (исходный файл с параметрами расчета в формате XML размером несколько сотен байт) на удаленные клиентские машины по их запросу из программы BOINC Manager, а также приема полученных результатов (результатирующий бинарный файл с выборкой значений показателей качества размером около 200 КБ без учета сжатия). С целью защиты от потенциальных ошибок различного характера каждое расчетное задание было обработано в трех экземплярах на разных машинах (т.н. кворум) с последующей побайтной сверкой полученных результатов (т.н. валидация).

Затраты вычислительного времени на проведенные эксперименты с использованием грид-инфраструктуры составили в общей сложности около 170 лет машинного времени (51,6 эксафлоп или 818 ГГц-лет), достигнутый выигрыш во времени обработки по сравнению с реализацией расчетов на одной машине составил 155 раз (оценки выполнены круглосуточно работающего компьютера с процессором Intel Core 2 Duo E6300 1,86 ГГц). В расчетах приняли участие более 1000 добровольцев (1100 компьютеров) из 61 страны мира, обеспечивая среднюю производительность проекта по данным [9] на уровне 1 TFLOPS. Объем полученных в ходе экспериментов «сырых» данных составил 235 ГБ.

Постобработка результатов заключалась в расчете средневыборочных значений критериев γ_x по каждой из выборок в отдельности и вероятностей ρ_x путем сопоставления выборок, полученных разными методами, для выбранной точки пространства параметров. Время постобработки составило величину порядка 5–23 часов для каждого вычислительного эксперимента, основным ограничением при этом является скорость чтения большого объема данных с жесткого диска.

В результате постобработки были получены карты (двумерные массивы), представляющие собой набор срезов пространства параметров, общим объемом 96 МБ. Полученные карты были использованы в дальнейшем для детального сравнения методов синтеза разбиений, результаты которого представлены ниже.

С целью исследования пространства параметров были выполнены ряд вычислительных экспериментов, основными среди которых являются:

1. $3 \leq W_{\max} \leq 200$, $1 \leq N \leq 600$ при $X_{\max} = +\infty$ и $Y_{\max} = +\infty$;
2. $4 \leq X_{\max} \leq 150$, $1 \leq N \leq 500$ при $W_{\max} = +\infty$ и $Y_{\max} = +\infty$.

Во всех выполненных экспериментах максимальное число логических условий $X(G^0) = 200$ и максимальное число микроопераций $Y(G^0) = 200$, объем псевдослучайной выборки граф-схем алгоритмов $K = 5000$.

На рис. 2–5 приведены экспериментально полученные зависимости, показывающие абсолютные значения оптимизируемых показателей качества. Полученные зависимости подтверждают сделанные ранее [5, 6] выводы о том, что по мере усиления ограничения значения показателей качества увеличиваются. Приведенные зависимости показывают, что усиление ограничения X_{\max} в меньшей степени сказывается на ухудшении абсолютных показателей качества по сравнению с усилением ограничения W_{\max} . При этом ограничение X_{\max} практически не влияет на увеличение степени дублирования микроопераций.

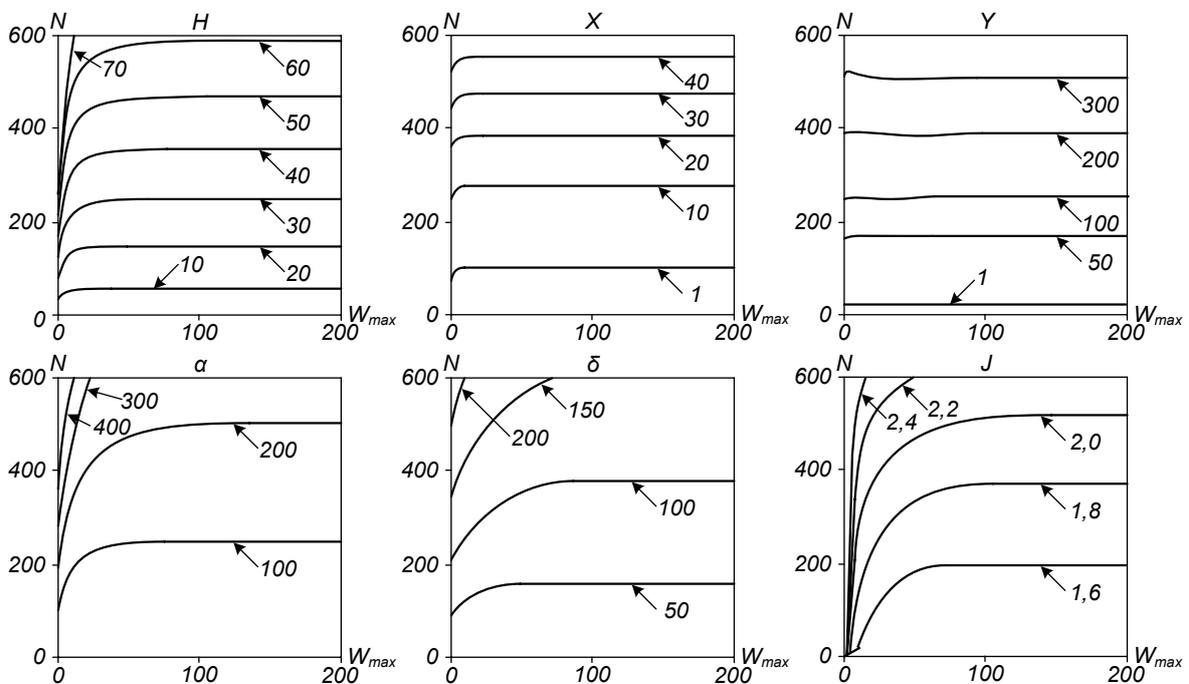


Рис. 2. Линии уровня усредненных значений γ_x показателей качества для метода С.И. Баранова (эксперимент 1)

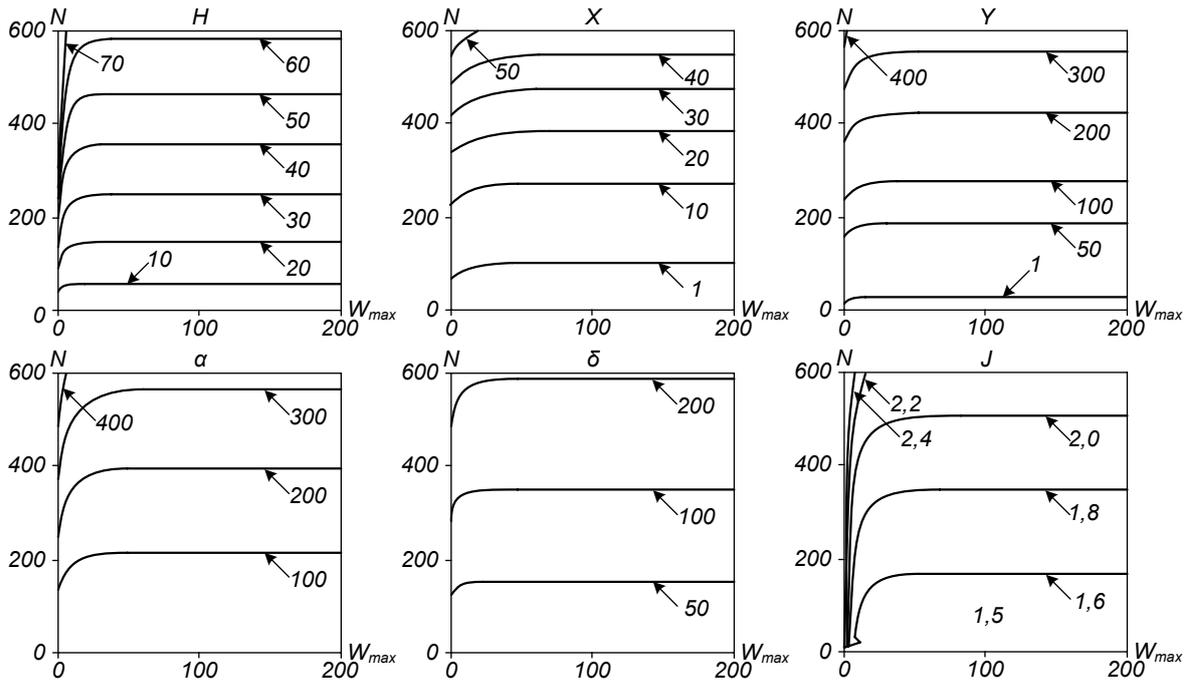


Рис. 3. Линии уровня усредненных значений γ_x показателей качества для параллельно-последовательного метода (эксперимент 1)

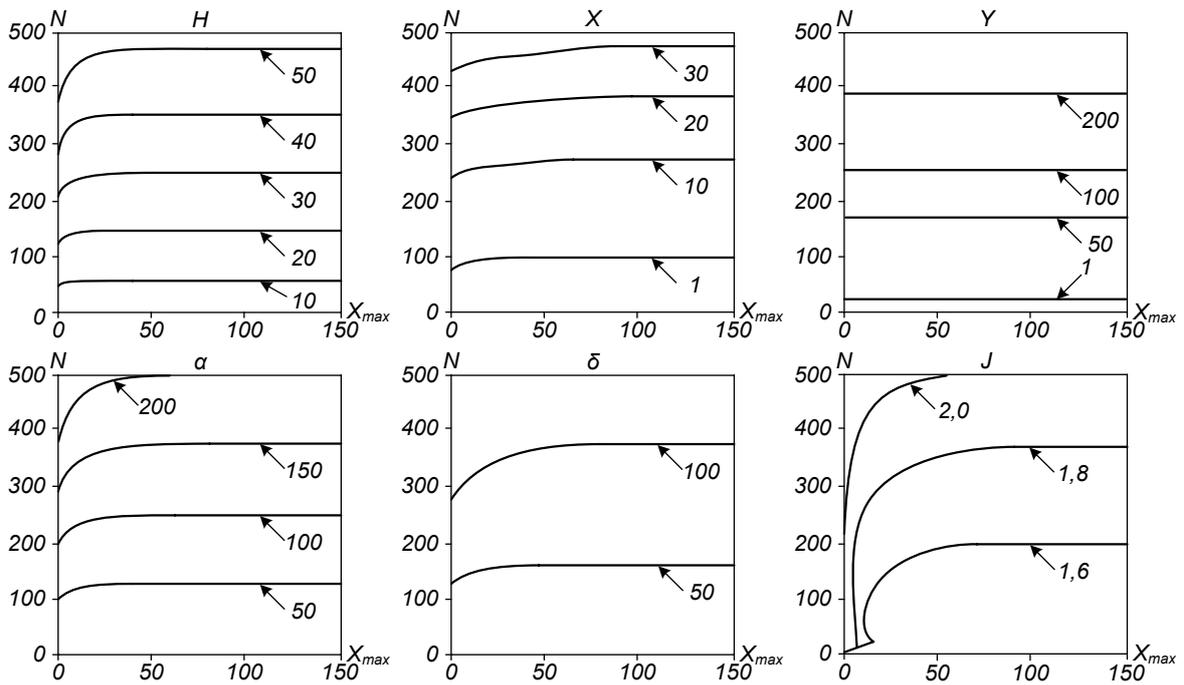


Рис. 4. Линии уровня усредненных значений γ_x показателей качества для метода С.И. Баранова (эксперимент 2)

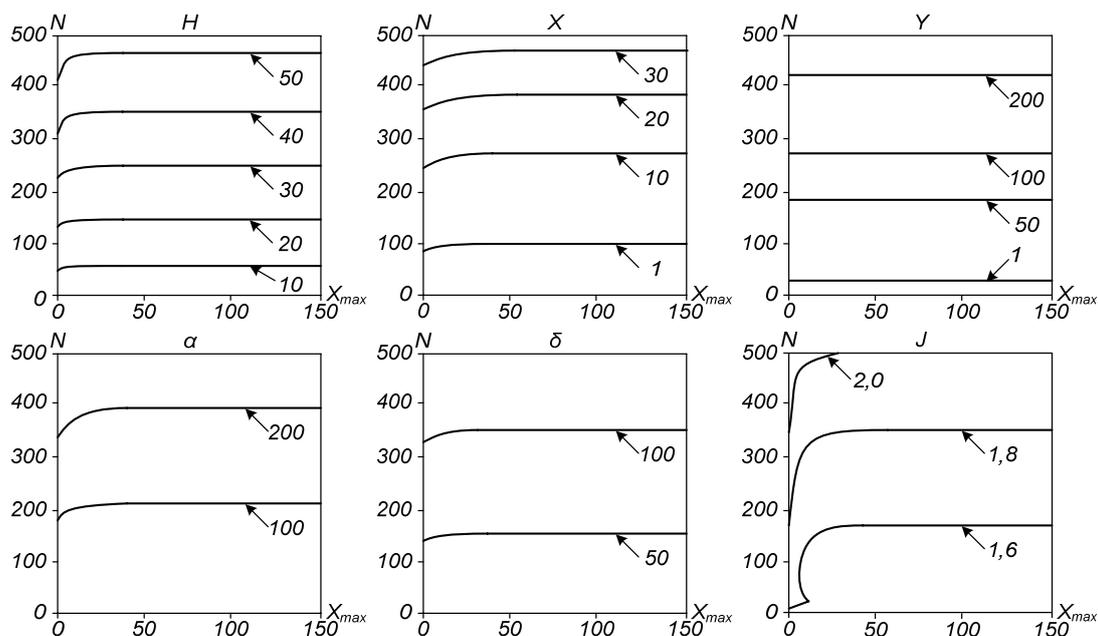


Рис. 5. Линии уровня усредненных значений γ_x показателей качества для параллельно-последовательного метода (эксперимент 2)

На рис. 7–10 приведены линии уровня для вероятностей ρ_x получения решения с минимальным значением выбранного показателя качества.

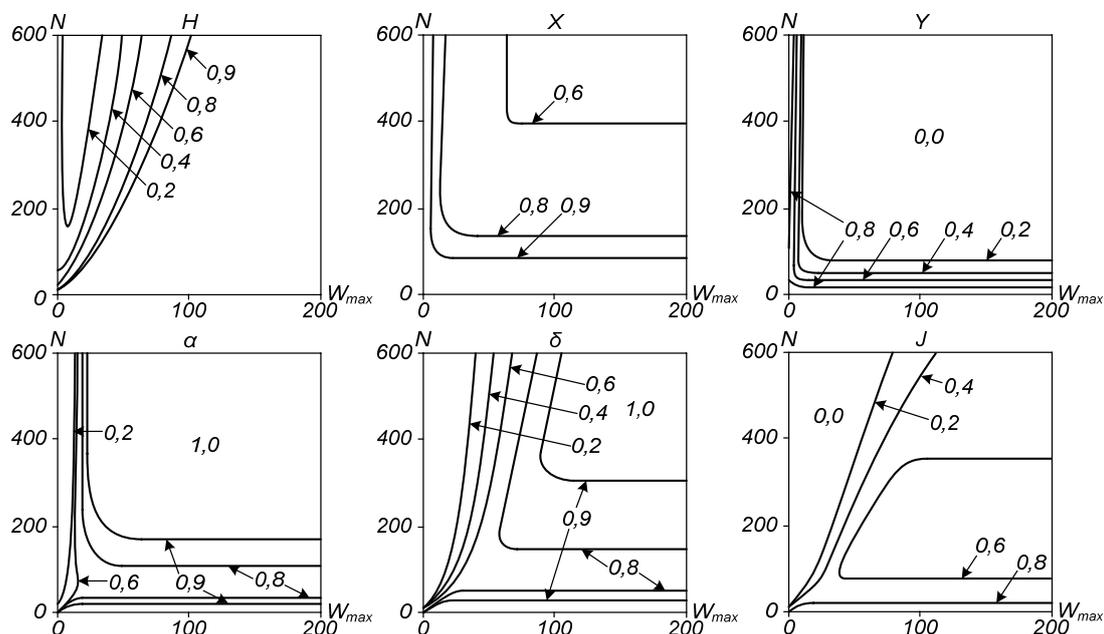


Рис. 6. Линии уровня значений вероятности ρ_x получения показателя качества с минимальным значением для метода С.И. Баранова (эксперимент 1)

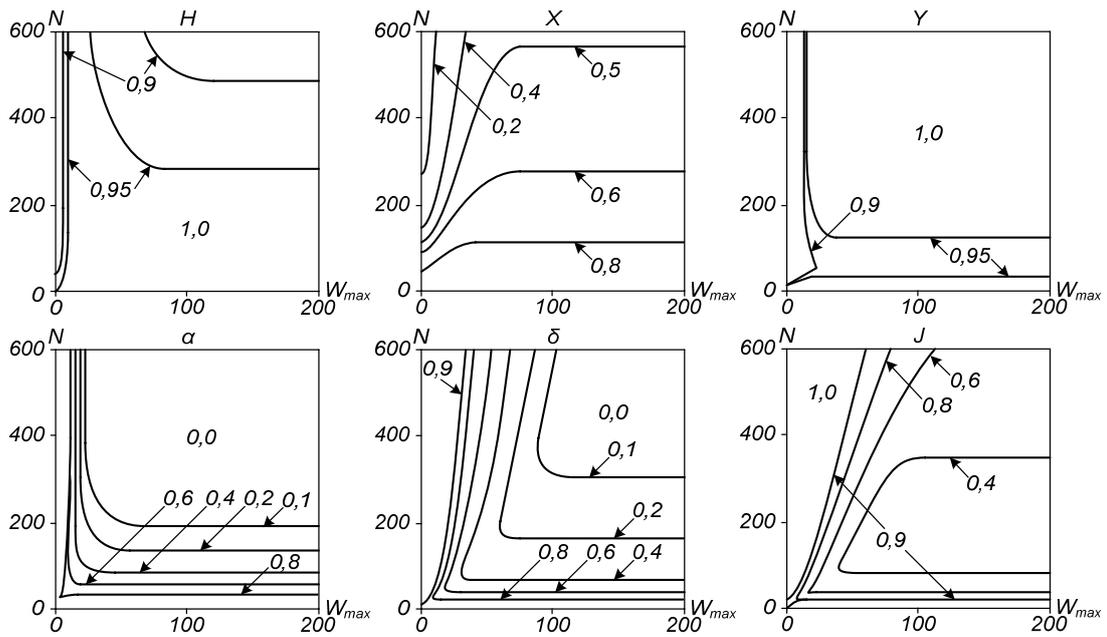


Рис. 7. Линии уровня значений вероятности ρ_x получения показателя качества с минимальным значением для параллельно-последовательного метода (эксперимент 1)

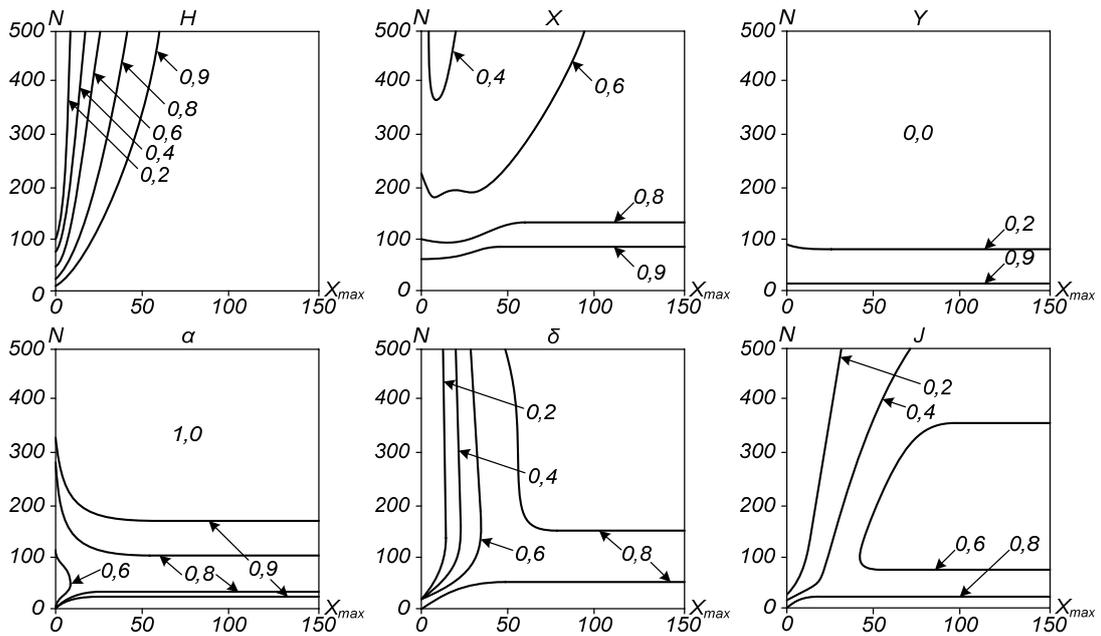


Рис. 8. Линии уровня значений вероятности ρ_x получения показателя качества с минимальным значением для метода С.И. Баранова (эксперимент 2)

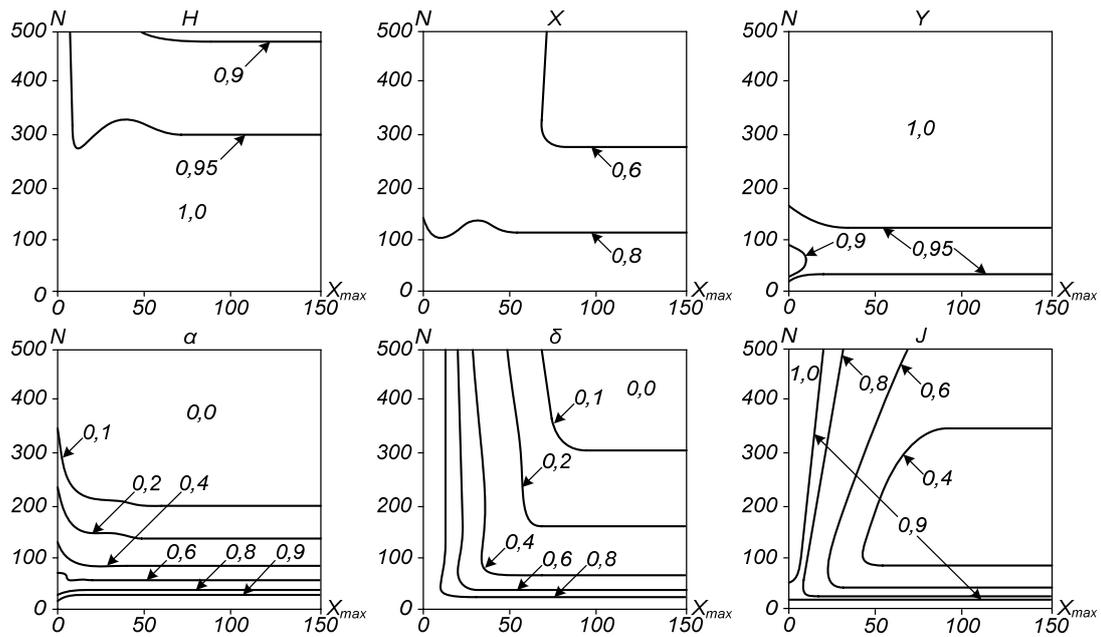


Рис. 9. Линии уровня значений вероятности ρ_x получения показателя качества с минимальным значением для параллельно-последовательного метода (эксперимент 2)

Приведенные карты позволяют оценить вероятность получения разбиения с минимальным значением показателя качества в зависимости от области на карте.

Полученные данные в целом подтверждают сделанные ранее [5, 6] выводы о наличии ряда зон в пространстве параметров, причем предпочтительной зоной для метода С.И. Баранова является зона слабых или отсутствующих ограничений, а для параллельно-последовательного метода – зона сильных ограничений. Новыми являются следующие исключения из общего правила:

1. Преимущество параллельно-последовательного метода по оптимизации распределения микроопераций и интегральному критерию качества нивелируется в области очень сильных ограничений по емкости памяти контроллера ($W_{\max} \leq 10$).
2. Преимущество параллельно-последовательного метода по оптимизации сложности сети межблочных связей отсутствует при любых значениях ограничения на число принимаемых сигналов логических условий X_{\max} при размере граф-схемы алгоритма управления $N \geq 40$.

Полученные двумерные зависимости являются гораздо более подробными и удобными для анализа по сравнению с полученными ранее одномерными [6]. Они позволяют подтвердить сделанный ранее вывод о том, что однозначного предпочтения нельзя отдать ни одному из методов синтеза разбиений. Параллельно-последовательный метод обеспечивает высокое качество решений по интегральному показателю за счет оптимизации числа блоков

и распределения микроопераций, однако уступает при распределении логических условий в области сильных ограничений, в минимизации сложности сети межблочных связей и интенсивности межблочных взаимодействий в области слабых ограничений с ростом размера граф-схемы алгоритма. Метод С.И. Баранова наоборот демонстрирует ухудшение интегрального качества решений по мере усиления ограничения, выигрывая при этом по распределению логических условий, сложности сети межблочных связей и интенсивности межблочных взаимодействий в области слабых ограничений.

Авторы статьи выражают благодарность всем добровольцам, принявшим участие в расчетах и обсуждении деталей адаптации расчетного модуля под BOINC грид, а также лично Валяеву Сергею Юрьевичу (SerVal), разработчику серверного кода проекта Gerasim@home под операционную систему Microsoft Windows, за предоставление технической возможности использования проекта Gerasim@Home для организации распределенных вычислений и ряд ценных советов и замечаний.

Список литературы

1. Организация и синтез микропрограммных мультимикроконтроллеров / Зотов И.В., Колосков В.А., Титов В.С. и др. Курск: изд-во Курск, 1999. 368 с.
2. Архитектура параллельных логических мультимикроконтроллеров / Емельянов С.Г., Зотов И.В., Титов В.С. М: Высшая школа, 2009. 233 с.
3. Комбинаторно-логические задачи синтеза разбиений параллельных алгоритмов логического управления при проектировании логических мультимикроконтроллеров / Э.И. Ватутин, И.В. Зотов, В.С. Титов и др. Курск: изд-во КурскГТУ, 2010. 200 с.
4. Ватутин Э.И. Проектирование логических мультимикроконтроллеров. Синтез разбиений параллельных граф-схем алгоритмов. Saarbrücken: Lambert Academic Publishing, 2011 г. 292 с.
5. Ватутин Э.И., Волобуев С.В., Зотов И.В. Комплексная сравнительная оценка методов выбора разбиений при проектировании логических мультимикроконтроллеров // Идентификация систем и задачи управления (SICPRO'08). М.: ИПУ РАН, 2008. С. 1917–1940.
6. Ватутин Э.И., Волобуев С.В., Зотов И.В. Комплексный сравнительный анализ качества разбиений при синтезе логических мультимикроконтроллеров в условиях присутствия технологических ограничений // Параллельные вычисления и задачи управления (PACO'08). М.: ИПУ РАН, 2008. С. 643–685.
7. <http://boinc.berkeley.edu>
8. <http://ru.wikipedia.org/wiki/BOINC>
9. <http://boincstats.boincstats.com>
10. <http://gerasim.boinc.ru>