

УДК 04.92+532.5

doi 10.26089/NumMet.v17r106

**ПРИМЕНЕНИЕ АРХИТЕКТУРЫ MULTIGPU+CPU
ДЛЯ ЗАДАЧ ПРЯМОГО ЧИСЛЕННОГО МОДЕЛИРОВАНИЯ
ЛАМИНАРНО-ТУРБУЛЕНТНОГО ПЕРЕХОДА**

Н. М. Евстигнеев¹, О. И. Рябков²

Обобщаются данные о применении различных параллельных вычислительных архитектур при численном моделировании задач ламинарно-турбулентного перехода (ЛТП). Обычно анализ ЛТП основан на рассмотрении статистических параметров: корреляций пульсаций скорости, энергетических спектров и др. Анализ ЛТП как нелинейной динамической системы в дополнение к уже указанному анализу основан на анализе собственных значений якобиана, вида аттракторов систем в фазовом пространстве и собственных значений матрицы монодромии. В результате строятся бифуркационные сценарии и диаграммы. Это дает возможность проследить механизм усложнения для рассматриваемых задач при ЛТП при изменении выбранных параметров (чисел Рейнольдса, Маха, Фруда и др.). Рассмотрение процесса ЛТП с точки зрения нелинейных динамических систем накладывает требования точности и быстродействия используемых алгоритмов решения задач. Начиная с 2008 г. в наших работах используются GPU- и multiGPU-архитектуры совместно с CPU. За это время было рассмотрено восемь постановок задач ЛТП. Для численного моделирования применялись различные методы высокого порядка. В настоящей статье для каждого класса методов рассматриваются характерные вычислительные операции, приводятся использованные библиотеки и выполняется сравнение эффективности разработанных алгоритмов и примененных библиотек с CPU-версиями кода, а также между собой. Показано, что в среднем на один GPU по сравнению с CPU ускорение варьируется от 5 до 35 раз. В связи со сложностью алгоритмов при MPI CPU- и multiGPU-подходах ускорение редко бывает линейным и оно пропорционально степенной функции с показателем 0.78–0.81. Для multiGPU-анализа алгоритмы тестировались на пяти GPU. Обсуждаются результаты при гибридном применении CPU+multiGPU для одной из задач.

Ключевые слова: multiGPU, гибридная архитектура GPU и CPU, прямое численное моделирование, ламинарно-турбулентный переход, динамические системы, численные методы высокого порядка.

1. Введение. Настоящая статья посвящена обобщающему анализу результатов ускорений параллельных вычислений для задач ламинарно-турбулентного перехода (ЛТП). Во многих статьях [6–22] не уделяется данному вопросу достаточного внимания. В нашей работе рассматриваются различные методы численного моделирования и особенности, возникающие при этом с точки зрения технологии реализации и программирования параллельных вычислений. Основное внимание уделяется multiGPU-подходу.

В общем случае анализ ЛТП основан на построении аттракторов численных аппроксимаций начально-краевых задач (НКЗ) и их сечений сечениями Пуанкаре. Пусть M — фазовое пространство, а φ — функция действия фазового потока изучаемой динамической системы на фазовое пространство, порождающая траекторию $\gamma \in M \times \mathbb{R}$. Сечение Пуанкаре S — это сечение пространства M в некоторой точке $m \in M$, такое, что S трансверсально траектории γ . Таким образом, изучается множество точек пересечения $A = \gamma \cap S$. Технически для реализации данного метода расчетные данные записываются из некоторой точки физического или функционального пространства за весь период расчета. После этого по имеющимся данным строится сечение Пуанкаре и множество точек пересечения наносится на плоскость. Более детально см. [6].

Настоящая статья состоит из семи разделов. В первом разделе в общих терминах ставятся задачи, рассмотренные в ряде других работ. Далее излагаются архитектуры, на которых производятся исследования. В последующих четырех разделах обсуждаются элементы алгоритмов и полные численные решения различными методами на multiGPU-архитектурах. В последнем разделе делается общий вывод по применимости multiGPU и гибридных алгоритмов в задачах ЛТП.

¹ Институт системного анализа РАН, просп. 60-летия Октября, 9, 117312, Москва; ст. науч. сотр., e-mail: evstigneevnm@yandex.ru

² Институт системного анализа РАН, просп. 60-летия Октября, 9, 117312, Москва; науч. сотр., e-mail: roi-techsup@yandex.ru

2. Исходные начально-краевые задачи. Рассмотрим задачи ЛТП, которые решались с применением (multi)GPU и гибридных архитектур. В общем случае начально-краевые задачи можно разделить на задачи для несжимаемых и сжимаемых течений. Принципиальное различие заключается в процедуре расчета при численном моделировании. В первом случае требуется либо проекция в пространство соленоидальных вектор-функций, либо разложение решения на потенциальное и соленоидальное. Во втором случае требуется решение задачи распада произвольного разрыва на элементах конечномерного приближения [1]. Для экономии места сформулируем типичную НКЗ для двух случаев.

1. *Задана область $\Omega \subset \mathbb{R}^k$, $k = 2, 3$, а также краевые и начальные условия. Для каждого значения бифуркационного параметра (например, числа Рейнольдса R) из заданного отрезка требуется найти $\mathbf{u} : \Omega \times (0, T] \rightarrow \mathbb{R}^k$, $p : \Omega \times (0, T] \rightarrow \mathbb{R}$, такие, что на Ω выполняется следующая система уравнений Навье–Стокса для вязкой несжимаемой жидкости:*

$$\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla) \mathbf{u} = -\nabla p + \frac{1}{R} \Delta \mathbf{u} + \mathbf{f}, \quad (1)$$

$$\nabla \cdot \mathbf{u} = 0. \quad (2)$$

2. *Задача для течения вязкого газа записывается следующим образом.*

Задана область $\Omega \subset \mathbb{R}^3$, а также краевые и начальные условия. Определим скалярные функции f как $f : \Omega \times (0, T] \rightarrow \mathbb{R}$ и вектор-функции \mathbf{f} как $\mathbf{f} : \Omega \times (0, T] \rightarrow \mathbb{R}^3$. Тогда для каждого значения бифуркационного параметра из заданного отрезка найти такие \mathbf{u} , ρ и E , что выполняется система уравнений

$$\begin{aligned} \frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x_j} [\rho u_j] &= 0; & \frac{\partial}{\partial t} (\rho u_i) + \frac{\partial}{\partial x_j} [\rho u_i u_j + p \delta_{ij} - \tau_{ji}] &= g_i, \quad i = 1, 2, 3; \\ \frac{\partial}{\partial t} (\rho E) + \frac{\partial}{\partial x_j} [\rho u_j E + u_j p - u_i \tau_{ij}] &= 0; & E &= \frac{1}{2} \rho \mathbf{u}^2 + \rho e; \\ p &= (\gamma - 1) \left(E - \frac{1}{2} \rho \mathbf{u}^2 \right); & \tau_{ij} &= 2\nu S_{ij}; \\ S_{ij} &= \frac{1}{2} \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right) + \frac{1}{3} \delta_{ij} \frac{\partial u_k}{\partial x_k}. \end{aligned} \quad (3)$$

Здесь E — скалярная функция полной энергии газа; e — скалярная функция внутренней энергии газа; γ — показатель адиабаты газа (1.4 для воздуха); p — скалярная функция давления; \mathbf{u} — вектор-функция скорости газа; ρ — скалярная функция плотности газа; \mathbf{g} — внешняя сила; ν — динамическая вязкость газа. Подразумевается суммирование по одинаковым индексам.

Размерность конечномерного аппроксимирующего пространства выбирается из оценки размерности аттрактора для каждой задачи при верхнем значении бифуркационного параметра, которое соответствует наибольшей оценке размерности [6, 22]. Ниже в каждом подразделе статьи приведен максимальный масштаб дискретизации. В большинстве случаев расчеты проводились на более грубых аппроксимациях, когда предварительные результаты совпадали. Это также дало возможность провести сопоставление эффективности вычислительных методов при разных размерах задач.

2.1. Задача течения вязкой несжимаемой жидкости с периодической возмущающей силой (задача А.Н. Колмогорова) на \mathbb{T}^k , $k = 2, 3$. Данные задачи рассматривались в работах [20, 21], а также на некоторых конференциях и в докладах.

Особенности НКЗ: область $\Omega = \mathbb{T}^k(\alpha)$, $0 < \alpha \leq 1$, α — параметр удлинения тора:

$$\Omega = \begin{cases} [0, 2\pi/\alpha] \times [0, 2\pi] \times [0, 2\pi], & R = \frac{\sqrt{A}}{\nu} \left(\frac{L_y L_z}{4\pi^2} \right)^{3/4}, & \mathbf{f} = (\sin(ny) \sin(mz); 0; 0)^T & \text{для } k = 3; \\ [0, 2\pi/\alpha] \times [0, 2\pi], & R = \frac{\sqrt{A}}{\nu} \left(\frac{L_y}{2\pi} \right)^{3/2}, & \mathbf{f} = (\sin(ny); 0)^T & \text{для } k = 2. \end{cases}$$

Здесь m, n — параметры силового воздействия; A — амплитуда силового воздействия; L_j — размеры физического пространства; ν — кинематическая вязкость жидкости.

В связи с периодичностью задачи для решения уравнения (1) применяются два метода — метод Галеркина и спектральный метод. Для двухмерной задачи решение ищется в терминах ω - Ψ (“завихренность–функция тока”, более детально см. в [20]) и приводится к конечномерной системе обыкновенных дифференциальных уравнений (ОДУ) методом Бубнова–Галеркина. В качестве конечномерного пространства выбирается линейная оболочка, натянутая на пространство специальных сплайн-функций (S-сплайны [4]). В

результате получается система ОДУ, для решения которой применяется n -стадийный метод Рунге–Кутты. Для трехмерной задачи применяется метод Петрова–Галеркина и спектральный метод с применением дискретного преобразования Фурье (ДПФ) [21]. Для метода Петрова–Галеркина используются пробные функции из пространства S -сплайн-функций, а поверочные функции — полиномы Якоби.

Во всех системах тензор валентности три (свертка за счет нелинейности) заменяется на операцию произведения в физическом пространстве для ускорения расчета. Данная операция выполняется с применением либо квадратур наивысшей алгебраической степени точности (для методов Галеркина), либо с применением быстрого преобразования Фурье.

Размерность пространств:

- для спектрального метода: $512 \times 128 = 65\,536$ для 2D-задачи, и $512 \times 128 \times 128 = 8\,388\,608$ для 3D-задачи (с учетом действительности функций в спектральном методе);
- для метода Галеркина: $240 \times 60 = 14\,400$ для 2D-задачи и $40 \times 40 \times 81 = 129\,600$ для 3D-задачи.

Основными операциями для данных задач являются матрично-векторные операции и дискретное преобразование Фурье.

2.2. Задача течения вязкой жидкости в канале с расширением. Здесь рассматриваются две задачи — 3D-задача течения с уступа и 2D-задача течения проводящей жидкости в канале с расширением. Область $\Omega = \Omega_1 \setminus (\Omega_2 \cap \Omega_3)$, $\Omega_1 = L \times H (\times W)^*$, $\Omega_2 = l \times h (\times W)^*$, $\Omega_3 = (L - l) \times h$, где $L > l$, $H > h$. Выражение в скобках (*) присутствует только для 3D-задачи (т.е. глубина), а также ставится несимметричное расширение, т.е. $\Omega_3 = \emptyset$. На границах Ω ставятся граничные условия — стенки (условия непротекания и прилипания), периодические, входные и выходные. Более детально постановки задач описаны в работах [8, 16]. В задаче течения проводящей жидкости в канале НКЗ ставится аналогично, дополняясь системой уравнений Максвелла и соответствующими граничными условиями.

Для задачи течения с уступа применяется метод расщепления с проектором Гельмгольца–Хаджа. Это приводит к необходимости решения уравнения Пуассона на каждом шаге метода интегрирования системы ОДУ (например, для метода Рунге–Кутты k -го порядка требуется решать уравнение Пуассона k раз для одного полного шага по времени).

В связи со сложностью геометрии для данных задач использовались гибридные конечно-объемные и конечно-элементные методы. Для аппроксимации нелинейных частей уравнений применялись схемы WENO (Weighted Essentially Non-Oscillatory) 9 и 11 порядков (еще не опубликованы), а для аппроксимации линейных членов — метод конечных элементов с применением полиномов Лагранжа шестой степени. Размерность конечномерного пространства для 3D-задачи течения с уступа составляет $512 \times 256 \times 128 = 16\,777\,216$ для периодического по направлению z граничного условия и $512 \times 256 \times 256 = 33\,554\,432$ для граничного условия стенки с одной стороны. Для 2D-задачи течения проводящей жидкости — $1024 \times 256 = 262\,144$.

Основные операции — расчет конвективных членов уравнения WENO-схемой, решение уравнения Пуассона и применение матрицы к вектору для метода конечных элементов (МКЭ).

2.3. Задача конвекции Рэлея–Бенара. Классическая задача свободной конвекции рассматривалась в работах [9, 11] для системы приближения Обербека–Буссинеска. Рассмотрены две постановки задачи: $\Omega_1 = h \times \mathbb{T}^2$ в цилиндрической геометрии и $\Omega_2 = L \times H \times W$ — кубическая область, выбранная потому, что в книге [3] имеется экспериментальное исследование начальных бифуркаций. Уравнения (1), (2) дополняются уравнением переноса температуры $T_t + (\mathbf{u}, \nabla)T = \Delta T$ и силовым воздействием $\mathbf{f} = \text{Ra Pr} (T - T_0)(0; 0; -1)^T$. Бифуркационными параметрами являются число Прандтля $\text{Pr} = \nu/\chi$ и число Рэлея $\text{Ra} = \beta g h^3 \Delta T / \nu \chi$, построенные по временному масштабу, связанному с масштабом диффузии. Здесь χ — температуропроводность жидкости; β — коэффициент температурного расширения; h — высота слоя; ΔT — перепад температуры на слое; g — ускорение свободного падения. Для Ω_1 применялся спектральный метод на тригонометрических полиномах с введением фиктивного условия на верхних границах [9]. Свертка заменяется на переход в физическое пространство с применением ДПФ. Для Ω_2 с граничными условиями на стенке применялся комбинированный конечно-объемный + конечно-элементный метод. Размерность конечномерного аппроксимирующего пространства для Ω_1 составляет $256 \times 256 \times 256 = 16\,777\,216$ элементов, а для Ω_2 — $292 \times 292 \times 292 = 24\,897\,088$.

Основные операции — ДПФ, расчет конвективных членов уравнения WENO-схемой, решение уравнения Пуассона и применение матрицы к вектору для МКЭ.

2.4. Конвективная сверхзвуковая неустойчивость течения газа. Задача ставится как обтекание твердого тела потоком вязкого газа при $M_\infty = 2$ для системы уравнений (3). Особенностью задачи является образование сверхзвуковых и дозвуковых зон как в самой области расчета, так и на границе. Это приводит к необходимости рассмотрения граничных условий неотражающего типа. Бифуркационные

параметры не выбираются, а рассматривается поведение фазового пространства в зависимости от точки в физической области расчета, т.е. параметром выступает координата конфигурационного пространства. Для аппроксимации вязких членов применяется разрывной метод Галеркина на полиномах Лежандра, для аппроксимации невязких членов при решении задачи Римана — метод WENO 9 и 11 порядков для реконструкции потоков [17]. Размерность задачи $800 \times 265 \times 256 = 52\,428\,800$.

Основные операции — решение задачи распада произвольного разрыва, расчет потоков схемами WENO, применение матрицы к вектору для разрывного метода Галеркина.

2.5. Неустойчивость Кельвина–Гельмгольца течения вязкого газа. Задача ставится в классическом виде слоя смешения двух разноплотностных течений, двигающихся с различной скоростью. Критерием устойчивости течения является число Ричардсона, определяющееся как $Ri = \frac{\Delta\rho gh}{\rho_A \Delta U^2}$, где g — ускорение свободного падения, h — высота слоя смешения, ΔU — разность скоростей сдвигового слоя, ρ_A — средняя плотность жидкости (газа), $\Delta\rho$ — разность плотностей сдвиговых слоев.

На основе линейного анализа устойчивости отмечается, что турбулентное перемешивание наступает при $Ri < 1/4$. Область Ω — прямоугольная с размерами $X = 2.8$, $Y = 0.3$, $Z = 0.7$, сила гравитации направлена по направлению оси $-Y$. Граничные условия: условия симметрии по оси Y при $y = 0$ и $y = 0.3$, условия периодичности по оси Z , условия свободного истечения из области по оси X при $x = 2.8$. При $x = 0.0$ задаются начальное возмущение потока с амплитудой $\delta \leq 1 \times 10^{-5}$. В данной задаче также требуется применение неотражающих граничных условий [19]. Сила гравитации подбирается так, чтобы обеспечить неустойчивость по критерию Ричардсона. Давление задается так, чтобы максимальное число Маха составляло 0.8. Расчет числа Ричардсона выполняется по указанному выражению. Здесь $h = 0.15$, $\Delta\rho = 0.4$, $\rho_A = 0.8$. Для учета вязкости вводится число Рейнольдса, определяемое как $R = \Delta U h \rho_A / \nu$. Получается двухпараметрическая задача, где бифуркационными параметрами являются Ri и R .

Как и в задаче выше, для аппроксимации вязких членов применяется разрывной метод Галеркина на полиномах Лежандра, для аппроксимации невязких членов при решении задачи Римана — метод WENO 9 и 11 порядков для реконструкции потоков [17]. Размерность конечномерного пространства — $800 \times 350 \times 350 = 98\,000\,000$.

Аналогично предыдущей задаче, основные операции — решение задачи распада произвольного разрыва, расчет потоков схемами WENO, применение матрицы к вектору для разрывного метода Галеркина.

3. Мини multi-GPU кластер. За долгое время проведения работ сменилось много поколений графических сопроцессоров и центральных процессоров. Исторически все работы выполнены на GPU компании NVIDIA и центральных процессорах компании INTEL. Расчеты в работах с 2007 до 2009 гг. проводились с использованием двух GTX 8800, с 2009 по 2010 гг. — Tesla C1060, с 2010 по 2013 гг. — четырех k20, с 2013 г. по настоящее время — пяти GK110-B1 на базе GeForce. Центральные процессы — Intel Xeon начиная с L7445, 4 ядра X4 и до E5 Ivy Bridge, 16 ядер X2. Для современного сравнения все методы были повторно переметрированы на конфигурацию собранного multi-GPU кластера: Intel Xeon E5 Ivy Bridge 16 ядер, пять GPU NVIDIA GTX TITAN BLACK (GK110-B1), конфигурация собрана с общей памятью 64Gb. Такая конфигурация является относительно дешевой и обладает достаточной вычислительной мощностью для проведения прямого численного моделирования задач ЛТП на начальных стадиях. Все метрики выполнены на данной машине, если не оговорено обратное. Отдельно приводятся данные по производительности в количествах операций с плавающей точкой в секунду (FLOPS). Однако по опыту расчетов авторы считают оценку FLOPS недостаточной, поэтому приводится сопоставление различных методов по времени выполнения. Все вычисления выполнены в арифметике с плавающей точкой с двойной точностью.

4. Дискретное преобразование Фурье. Для расчета ДПФ используются две реализации алгоритма быстрого преобразования Фурье (БПФ): CUFFT (CUDA v.6.5) и FFTW (3.3.4). Для ускорения расчета CUFFT использовался в режиме двух GPU, а FFTW — в однопоточном и многопоточном (16 потоков) режиме. На рис. 1 показано сопоставление времени расчета одного полного преобразования $\mathbb{R}^3 \rightarrow \mathbb{C}^3$ и $\mathbb{C}^3 \rightarrow \mathbb{R}^3$ на реальных данных соответствующих задач при применении одной и двух GPU при размерах области вида $2^p \times 2^q \times 2^n$, $\{p, q, n\} \in \mathbb{N}$. Если сопоставить результаты более раннего теста при версиях CUFFT (CUDA 1.1/2.0) и FFTW(3.2) [5], то видно, что библиотека CUFFT улучшена по сравнению с предыдущими версиями. Максимальное ускорение составило 17 раз по сравнению с однопоточным FFTW и в 3.6 раза по сравнению с 16-поточным FFTW. При этом для задач малой размерности (2D задачи, тестовые 3D задачи) использование CUFFT не является оправданным. Работа при двух GPU для CUFFT оправдана начиная с размерности задачи 128^3 и выше. При этом для рассмотренных выше задач все результаты были получены на одном GPU ввиду уменьшения конечномерного пространства при сравнении результатов. Если линейные размеры области не являются степенью двойки, то эффективность CUFFT

падает, а ускорение по сравнению с 16-поточной FFTW составляет 2.4 раза.

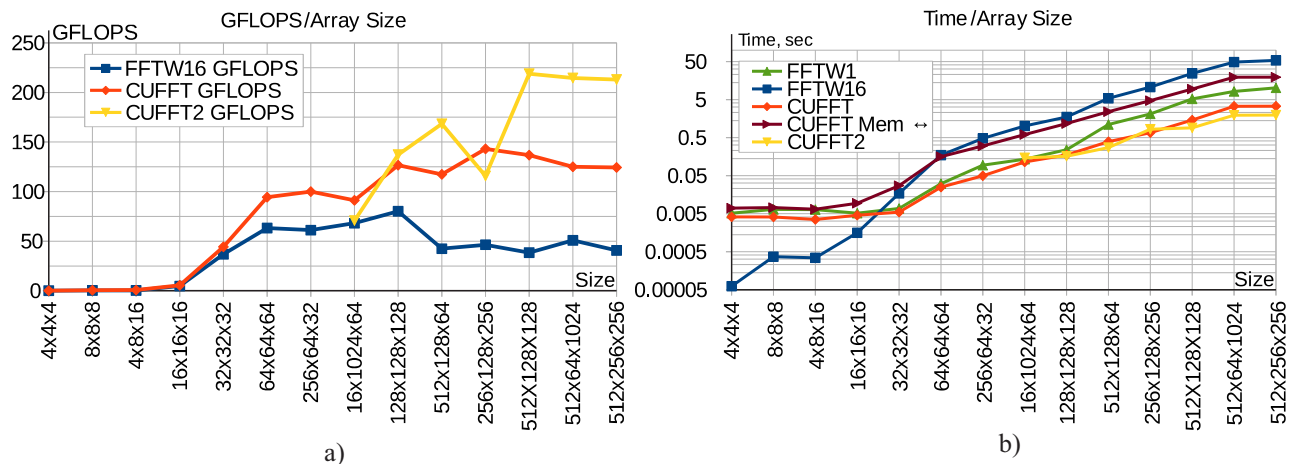


Рис. 1. Сопоставление производительности библиотек FFTW и CUFFT на 3D задаче А.Н. Колмогорова: а) производительность относительно размера задачи, б) время выполнения относительно размера задачи

В целом, для спектральных (псевдоспектральных) методов применение CUFFT и CUDA для остальной части кода приводит к ускорению в пять-шесть раз по сравнению с 16-поточной OpenMP версией кода.

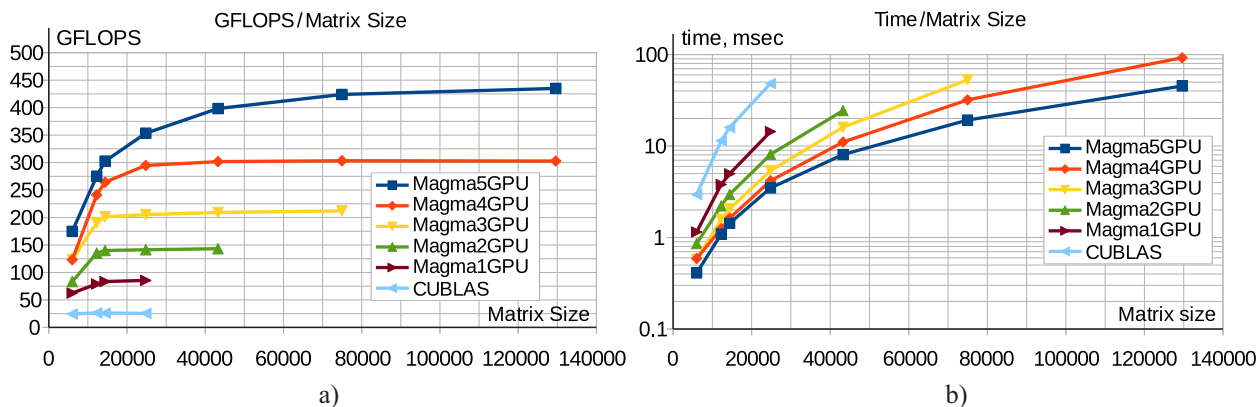


Рис. 2. Производительность $y = \alpha Ax + \beta y$ для MAGMA в режиме Multi-GPU и CUBLAS: а) производительность относительно размера задачи, б) время выполнения одного шага относительно размера задачи

5. Применение матрицы к вектору. Данная операция применяется, в основном, в методах Галеркина и МКЭ, при этом матрицы могут быть как плотные, так и разреженные (МКЭ). Для плотных матриц применяются библиотеки CUBLAS и MAGMA. Схожие по структуре, они имеют принципиальное различие, заключающееся в том, что MAGMA может быть скомпилирована под конкретную архитектуру с автоматическим учетом количества GPU, установленных на одном узле кластера. Поэтому указанная выше конфигурация миникластера является достаточно успешной. На рис. 2 показано сравнение производительности и чистого времени для операций `magmblas_dsymv_mgpu` и `cublasDsymv`.

Таким образом, достигается производительность 85 GFLOPS на одной GPU и 435 GFLOPS на пяти GPU. Из полученных результатов следует, что применение MAGMA оправдывается при расчете задач ЛТП и позволяет значительно сократить трудозатраты при написании кода.

6. Схема WENO и обмен данных на multiGPU. Вопрос о коммуникациях для multiGPU не поднимался при использовании библиотек. При написании собственных частей кода данный вопрос возникает. Во всех реализациях используется MPI-обмен данными, в котором выделяется буфер в оперативной памяти, относящейся к центральному процессору. Альтернативой такому подходу является DirectGPU, поддерживаемый CUDA начиная с версии 5.5. Авторы остановились на применении первого подхода, поскольку он может быть напрямую применен при переносе кода с одного узла на несколько. Минусами

такого подхода является загрузка центрального процессора при передаче данных и большее время, затрачиваемое на перенос данных. На практике в вышеуказанных задачах затраты на передачу данных составляли 0.3–0.5% от всех затрат (согласно замерам, полученным с использованием программы `nvprof`). При применении неявных методов затраты на пересылку возрастают и составляют 1.5–3%. Снижения затрат можно добиться асинхронным вызовом пересылки данных.

Отдельно кратко следует упомянуть достаточно сложную процедуру вычисления схемы WENO. Поскольку шаблон схемы является достаточно громоздким (17×3 точек шаблона для WENO-9 и 21×3 точек шаблона для WENO-11), вычисление весов полиномов Лагранжа выполнялось в четырех (шести для WENO-11) отдельных ядрах. Сетка красилась в “цвета”, и каждое ядро работало на своем цвете, таким образом минимизируется `coalescing`. После сборки весов результаты объединялись в промежуточных ядрах, после чего выполнялась окончательная сборка TVB (Total Variation Bounded) схемы с использованием весов шаблонов. Несмотря на сложность, такой метод оказался в 4.5 раза быстрее, чем наивный метод.

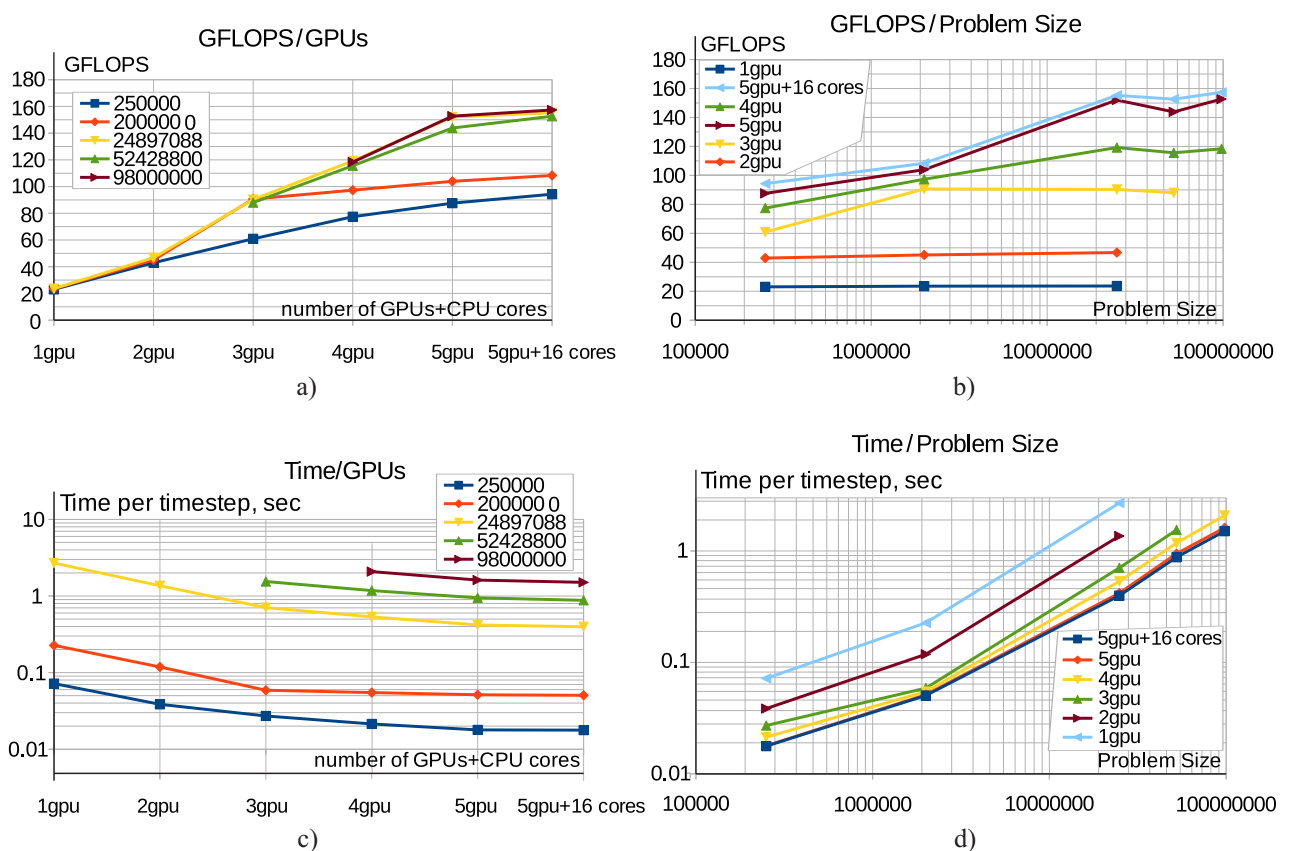


Рис. 3. Производительность всего кода для моделирования течения вязкого газа:

- производительность всего кода относительно количества GPU (для различных размеров задач),
- производительность всего кода относительно размера задачи, с) время одного шага по времени относительно количества GPU (для различных размеров задач),
- время одного шага по времени относительно размера задачи

Реализация `multiGPU+CPU` была выполнена на задаче о неустойчивости Кельвина–Гельмгольца. Суть в том, что рассмотрение характеристических граничных условий вместе с буферным слоем приводит к большому использованию регистров для GPU, что значительно замедляет выполнение соответствующих ядер. Поэтому расчет характеристических граничных условий выполняется на CPU на основе MPI. Результаты, представленные на рис. 3 для данной конфигурации, получены на машине с двумя CPU, 16 ядер на каждый. Один центральный процессор отвечал за коммуникацию между GPU, другой — за расчет граничных условий. В конфигурации с одним CPU прироста производительности не было обнаружено.

Характерный локальный минимум на рис. 3b как раз связан с выходными граничными условиями, на которых возникает трансзвуковой режим течения. При применении CPU+`multiGPU` данный минимум уменьшается.

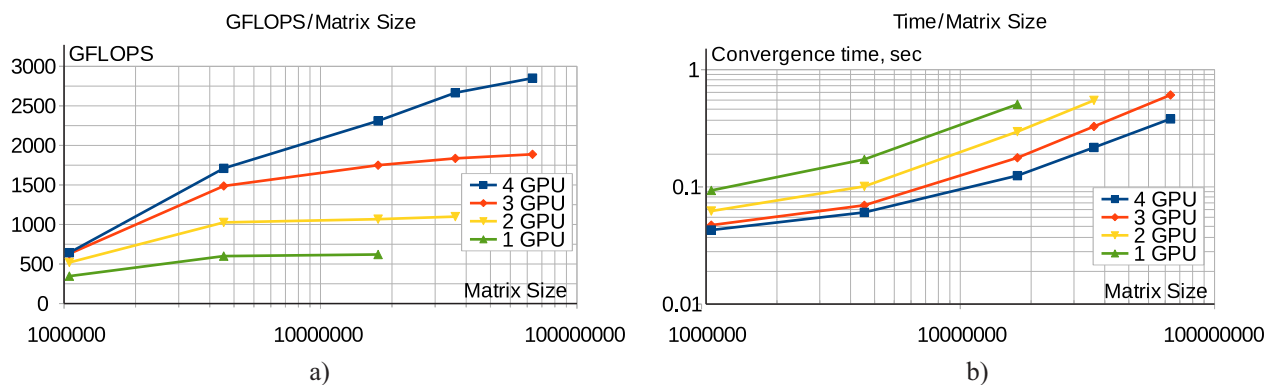


Рис. 4. Производительность при решении уравнения Пуассона геометрическим многосеточным методом: а) производительность относительно размера задачи, б) время одного шага по времени относительно размера задачи

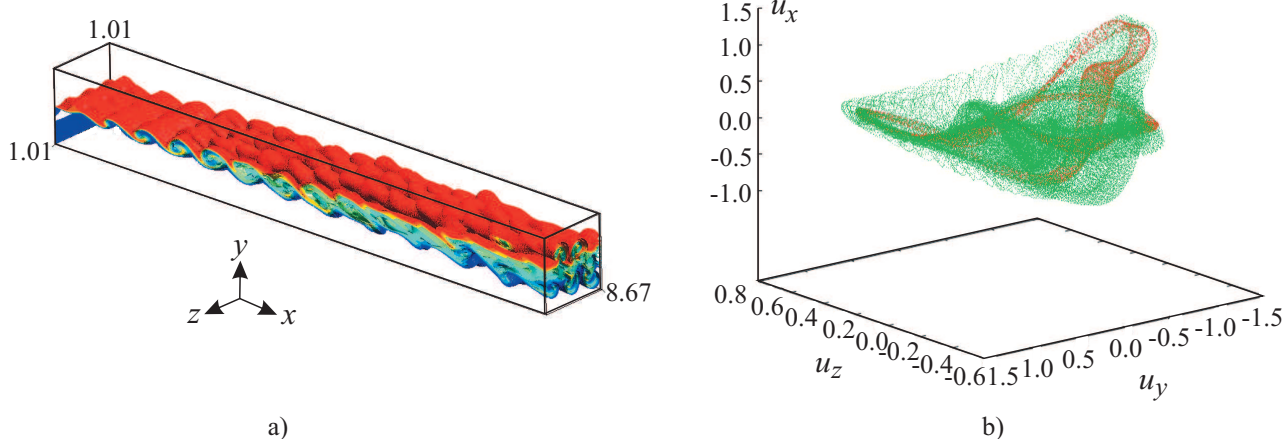


Рис. 5. Один из режимов при ЛТП в задаче Кельвина–Гельмгольца: а) изоповерхности плотности, б) проекция четырехмерного инвариантного тора

7. Уравнение Пуассона и решение линейных систем. Для решения уравнения Пуассона применяется многосеточный геометрический метод (МГМ) [10]. Он же применяется при необходимости решения систем линейных алгебраических уравнений. Такая необходимость возникает при использовании неявных методов, цель которых — выйти на квазистационарное решение перед применением высокоточных явных методов (ограничение по устойчивости при прямом численном моделировании задач ЛТП менее строгое, чем ограничение по точности). В отличие от работы [10] на верхнем уровне МГМ применяются различные методы крыловского типа. Из практики использования наиболее удачным оказался метод BiCGstab (BiConjugate Gradient Stabilized).

При условии использования пролонгаторов и рестрикторов высокого порядка (с 4 до 8) [10] удается добиться асимптотики $\mathcal{O}(N_x N_y N_z)$ для времени сходимости при решении уравнения Пуассона и $\mathcal{O}((N_x N_y N_z)^{1.15-1.2})$ при применении неявных методов и решении неявной части для разрывных методов Галеркина (где необходимо). На рис. 4 показана производительность при решении уравнения Пуассона для задачи течения с уступом. Оптимальным для данной задачи оказался W-цикл, 5 сеток с изменением размера в 2 раза, пролонгаторы/рестрикторы шестого порядка. При этом на последних ветках пролонгаторов применяются BiCGstab-методы в тестовом режиме, для оценки скорости сходимости и следующая пролонгирующая ветвь подправляется с учетом этих результатов. Критерий сходимости по норме C на верхнем уровне: $\varepsilon \leq 5.0 \times 10^{-15}$.

Как видно из рис. 4, на данной задаче удалось добиться практически идеального масштабирования, характерного для многосеточных геометрических методов. Кроме того, производительность одного GPU составила 620.33 GFLOPS, что составляет 32.5% от максимальной теоретической производительности на операциях с плавающей точкой двойной точности.

8. Общий анализ результатов. За длительное время написания кодов для прямого численного

моделирования ЛТП авторы пришли к выводу, что относительно небольшие машины, построенные на архитектуре multiGPU+CPU, являются оптимальными при требованиях к сеткам не больше 100 миллионов элементов. Более низкая производительность по сравнению с НРС-кластерами может быть компенсирована монопольным доступом к машине. В дальнейшем, после отладки всех узлов на такой машине, задачи с большим количеством элементов могут быть легко перенесены на промышленный кластер. Наиболее затратными для данной архитектуры до сих пор остаются эллиптические задачи, неявные методы и ограничение на количество регистров на ядро.

В целом, использование нами численных методов на данной архитектуре позволило в 5–7 раз ускорить расчеты по сравнению с более ранними конфигурациями, упомянутыми выше. Ускорение по сравнению с CPU-версиями кодов составляет 5–6 раз, при использовании библиотек до 35, а в отдельных задачах и до 50 раз. Несмотря на все еще слабую поддержку двойной точности в GPU, полученной производительности уже достаточно, чтобы достигать очень детального разрешения течения. Например, режим течения и проекция фазового портрета в \mathbb{R}^3 (и его многомерные сечения) для задачи Кельвина–Гельмгольца показаны на рис. 5. Видно, что точности достаточно, чтобы найти течение, в котором был обнаружен устойчивый четырехмерный инвариантный тор, хотя в данной задаче наблюдается масштабный эффект мультиустойчивости.

Работа поддержана РФФИ (код проекта 14–07–00123).

Статья рекомендована к публикации Программным комитетом Международной научной конференции “Параллельные вычислительные технологии” (ПаВТ–2016; <http://agora.guru.ru/pavt2016>).

СПИСОК ЛИТЕРАТУРЫ

1. *Toro E.F.* Riemann solvers and numerical methods for fluid dynamics. Berlin: Springer-Verlag, 1999.
2. *Sarris I.E., Kassinos S.C., Carati D.* Large-eddy simulations of the turbulent Hartmann flow close to the transitional regime // *Phys. Fluids*. 2007. **19**. doi: 10.1063/1.2757710
3. *Фрик П.Г.* Турбулентность: подходы и модели. Москва–Ижевск: Институт компьютерных исследований, 2003.
4. *Силаев Д.А., Коротаев Д.О.* Решение краевых задач с помощью S-сплайна // *Компьютерные исследования и моделирование*. 2009. **1**, № 2. 161–171.
5. CUFFT 1.1 / 2.0 vs FFTW 3.1.2 (x86_64) vs FFTW 3.2 (Cell) comparison. URL: http://www.sharcnet.ca/~merz/CUDA_benchFFT/.
6. *Evstigneev N.M., Magnitskii N.A.* FSM scenarios of laminar-turbulent transition in incompressible fluids // *Nonlinearity, Bifurcation and Chaos. Theory and Applications*. Rijeka: InTech, 2012. 250–280.
7. *Евстигнеев Н.М., Магницкий Н.А., Сидоров С.В.* Новый подход к объяснению природы турбулентности вязкой несжимаемой жидкости // *Труды ИСА РАН*. 2008. **33**. 49–65.
8. *Евстигнеев Н.М., Магницкий Н.А., Сидоров С.В.* О природе турбулентности в задаче движения жидкости за уступом // *Дифференциальные уравнения*. 2009. **45**, № 1. 69–73.
9. *Евстигнеев Н.М., Магницкий Н.А., Сидоров С.В.* О природе турбулентности в конвекции Рэлея–Бенара // *Дифференциальные уравнения*. 2009. **45**, № 6. 909–912.
10. *Евстигнеев Н.М.* Интегрирование уравнения Пуассона с использованием графического процессора технологии CUDA // *Вычислительные методы и программирование*. 2009. **10**. 268–274.
11. *Евстигнеев Н.М., Магницкий Н.А.* О возможных сценариях перехода к турбулентности в конвекции Рэлея–Бенара // *Доклады Академии наук*. 2010. **433**, № 3. 1–5.
12. *Евстигнеев Н.М., Магницкий Н.А.* Нелинейная динамика в начально-краевой задаче течения жидкости в уступа для гидродинамического приближения уравнений Больцмана // *Дифференциальные уравнения*. 2010. **46**, № 12. 1794–1798.
13. *Evstigneev N.M., Magnitskii N.A., Sidorov S.V.* Nonlinear dynamics of laminar-turbulent transition in three dimensional Rayleigh–Benard convection // *Communications in Nonlinear Science and Numerical Simulation*. 2010. **5**, N 10. 2851–2859.
14. *Евстигнеев Н.М.* Применение графического процессора для ускорения численного сеточного метода Больцмана с энтропийной стабилизацией // *Труды ИСА РАН*. 2010. **53**, № 14. 111–123.
15. *Evstigneev N.M., Magnitskii N.A.* Nonlinear dynamics of laminar-turbulent transition in back facing step problem for Boltzmann equations in hydrodynamic limit // *AIP Conf. Proc.* 2010. **1281**. 896–900.
16. *Евстигнеев Н.М., Магницкий Н.А., Рябков О.И.* Численное исследование перехода к турбулентности в задаче о двумерном течении вязкой сжимаемой проводящей жидкости в канале с симметричным расширением // *Труды ИСА РАН*. 2012. **62**, № 1. 55–62.
17. *Евстигнеев Н.М., Магницкий Н.А.* Особенности фазового пространства уравнений динамики газа для транзвуковой начально-краевой задачи // *Труды ИСА РАН*. 2012. **62**, № 4. 85–102.
18. *Евстигнеев Н.М., Магницкий Н.А.* Нелинейная динамика начальной стадии ламинарно-турбулентного перехода в задаче развития неустойчивости Кельвина–Гельмгольца // *Труды ИСА РАН*. 2013. **63**, № 3. 45–52.

19. Евстигнеев Н.М., Магницкий Н.А. О развитии неустойчивости Кельвина–Гельмгольца на начальной стадии ламинарно-турбулентного перехода в вязком газе // Труды ИСА РАН. 2014. **64**, № 3. 41–52.
20. Евстигнеев Н.М., Магницкий Н.А., Силаев Д.А. Качественный анализ динамики в задаче А.Н. Колмогорова течения вязкой несжимаемой жидкости // Дифференциальные уравнения. 2015. **51**, № 10. 1302–1314.
21. Евстигнеев Н.М., Магницкий Н.А. Численное исследование нелинейной динамики в обобщенной задаче течения А.Н. Колмогорова // Системный анализ и информационные технологии (САИТ'2015). Труды шестой международной конференции (Светлогорск, 10–20 июня, 2015г.) Светлогорск: Издательский центр БФУ им. И. Канта, 2015. 49–55.
22. Евстигнеев Н.М. Об одном способе осреднения уравнений динамики сжимаемой и несжимаемой жидкости // Вестник МГОУ. Серия ФМ. 2010. № 2. 47–52.

Поступила в редакцию
05.02.2016

Application of MultiGPU+CPU Architecture for the Direct Numerical Simulation of Laminar-Turbulent Transition

N. M. Evstigneev¹ and O. I. Ryabkov²

¹ *Institute for Systems Analysis, Russian Academy of Sciences; prospekt 60-letiya Oktyabrya, 9, Moscow, 117312, Russia; Ph.D., Senior Scientist, e-mail: evstigneevnm@yandex.ru*

² *Institute for Systems Analysis, Russian Academy of Sciences; prospekt 60-letiya Oktyabrya, 9, Moscow, 117312, Russia; Ph.D., Scientist, e-mail: roi-techsup@yandex.ru*

Received February 5, 2016

Abstract: The use of various parallel computational architectures for the direct numerical simulation (DNS) of laminar-turbulent transition problems (LTTPs) is summarized. Usually, DNS results are analyzed on the basis of a set of statistical parameters: pulsation velocities correlations, energy spectra, etc. When a dynamical system analysis approach is applied to DNS results, it is necessary to evaluate additional parameters: eigenvalues of Jacobi matrices, phase space attractors, and eigenvalues of monodromy matrices, etc. This allows one to present results as bifurcation scenarios and diagrams and bring up more details concerning LTTP scenario as functions of bifurcation parameters (e.g., Reynolds, Mach, and Froude numbers). This problem is computationally expensive and algorithms are complex. This brings up more demands on the hardware efficiency and software algorithm optimization. We are using GPU and multiGPU together with CPU architectures since 2008 for this kind of DNS. We considered eight different LTTPs since then. Various high-order methods were applied. In this paper we show typical computational operations for each class of problem. We illustrate the application of libraries and algorithms, perform efficiency benchmarking across GPUs and with CPU versions. It is shown that in general one GPU is 5 to 35 times faster than CPU. The acceleration is worse than linear and is proportional to a power function with an exponent between 0.78 and 0.81. We use five GPUs for multiGPU and show CPU+multiGPU efficiency for one of the problems under consideration.

Keywords: multiGPU, hybrid GPU and CPU architecture, direct numerical simulation, laminar-turbulent transition, dynamical systems, high order numerical methods.

References

1. E. F. Toro, *Riemann Solvers and Numerical Methods for Fluid Dynamics* (Springer, Berlin, 1999).
2. I. E. Sarris, S. C. Kassinos, and D. Carati, “Large-Eddy Simulations of the Turbulent Hartmann Flow Close to the Transitional Regime,” *Phys. Fluids* **19** (2007). doi: 10.1063/1.2757710
3. P. G. Frik, *Turbulence: Approaches and Models* (Inst. Komp’yut. Issled., Moscow; Izhevsk, 2003) [in Russian].
4. D. A. Silaev and D. O. Korotaev, “Solving of Boundary Tasks by Using S-Spline,” *Komp. Issled. Model.* **1** (2), 161–171.
5. CUFFT 1.1 / 2.0 vs FFTW 3.1.2 (x86_64) vs FFTW 3.2 (Cell) comparison.
http://www.sharcnet.ca/~merz/CUDA_benchFFT/. Cited February 15, 2016.

6. N. M. Evstigneev and N. A. Magnitskii, "FSM Scenarios of Laminar-Turbulent Transition in Incompressible Fluids," in *Nonlinearity, Bifurcation and Chaos. Theory and Applications* (InTech, Rijeka, 2012), pp. 250–280.
7. N. M. Evstigneev, N. A. Magnitskii, and S. V. Sidorov, "A New Approach to the Explanation of the Nature of Turbulence in a Viscous Incompressible Fluid," *Tr. ISA RAN* **33** (12), 49–65 (2008).
8. N. M. Evstigneev, N. A. Magnitskii, and S. V. Sidorov, "On the Nature of Turbulence in a Problem on the Motion of a Fluid behind a Ledge," *Differ. Uravn.* **45**, (1), 69–73 (2009) [*Differ. Equ.* **45** (1), 68–72 (2009)].
9. N. M. Evstigneev, N. A. Magnitskii, and S. V. Sidorov, "On the Nature of Turbulence in Rayleigh–Benard Convection," *Differ. Uravn.* **45**, (6), 890–893 (2009) [*Differ. Equ.* **45** (6), 909–912 (2009)].
10. N. M. Evstigneev, "Numerical Integration of Poisson's Equation Using a Graphics Processing Unit with CUDA-Technology," *Vychisl. Metody Programm.* **10**, 268–274 (2009).
11. N. M. Evstigneev and N. A. Magnitskii, "On Possible Scenarios of the Transition to Turbulence in Rayleigh–Benard Convection," *Dokl. Akad. Nauk* **433** (3), 318–322 (2010) [*Dokl. Math.* **82** (1), 659–662 (2010)].
12. N. M. Evstigneev and N. A. Magnitskii, "Nonlinear Dynamics in the Initial-Boundary Value Problem on the Fluid Flow from a Ledge for the Hydrodynamic Approximation to the Boltzmann Equations," *Differ. Uravn.* **46**, (12), 1794–1798 (2010) [*Differ. Equ.* **46** (12), 1794–1798 (2010)].
13. N. M. Evstigneev, N. A. Magnitskii, and S. V. Sidorov, "Nonlinear Dynamics of Laminar-Turbulent Transition in Three Dimensional Rayleigh–Benard Convection," *Commun. Nonlinear. Sci. Numer. Simul.* **15** (10), 2851–2859 (2010).
14. N. M. Evstigneev, "Application of GPU for the Acceleration of the Lattice Boltzmann Method with Entropy Stabilization," *Tr. ISA RAN* **53** (14), 111–123 (2010).
15. N. M. Evstigneev and N. A. Magnitskii, "Nonlinear Dynamics of Laminar-Turbulent Transition in Back Facing Step Problem for Boltzmann Equations in Hydrodynamic Limit," *AIP Conf. Proc.* **1281**, 896–900 (2010).
16. N. M. Evstigneev, N. A. Magnitskii, and O. I. Ryabkov, "Numerical Simulation of the Transition to 2D Flow of a Viscous Compressible Conducting Fluid in a Symmetrically Expanding Channel," *Tr. ISA RAN* **62** (1), 55–62 (2012).
17. N. M. Evstigneev and N. A. Magnitskii, "On Phase Space Peculiarities of Gas Dynamics Equations for a Supersonic Initial-Boundary Value Problem," *Tr. ISA RAN* **62** (4), 85–102 (2012).
18. N. M. Evstigneev and N. A. Magnitskii, "Nonlinear Dynamics of the Initial Stage for the Laminar-Turbulent Transition in the Kelvin–Helmholtz Instability Problem," *Tr. ISA RAN* **63** (3), 45–52 (2013).
19. N. M. Evstigneev and N. A. Magnitskii, "Kelvin–Helmholtz Instability Evolution at the Initial Stage of the Laminar-Turbulent Transition in a Viscous Gas," *Tr. ISA RAN* **64** (3), 41–52 (2014).
20. N. M. Evstigneev, N. A. Magnitskii, and D. A. Silaev, "Qualitative Analysis of Dynamics in Kolmogorov's Problem on a Flow of a Viscous Incompressible Fluid," *Differ. Uravn.* **51**, (10), 1302–1314 (2015) [*Differ. Equ.* **51** (10), 1292–1305 (2015)].
21. N. M. Evstigneev and N. A. Magnitskii, "Numerical Simulation of Nonlinear Dynamics in the Generalized Kolmogorov's Flow Problem," in *Proc. 6th Int. Conf. on System Analysis and Informational Technologies, Svetlogorsk, Russia, June 10–20, 2015* (Baltic Federal Univ., Svetlogorsk, 2015), pp. 49–55.
22. N. M. Evstigneev, "About One Averaging Method of the Equations of a Compressible and Incompressible Fluid," *Vestn. Mosk. Gos. Obl. Univ., Ser. Fiz. Mat., No. 2*, 47–52 (2010).