

Е. В. БОЛГОВА, С. В. ИВАНОВ, Е. А. ГРИНИНА,
П. М. А. СЛОТ, А. В. БУХАНОВСКИЙ

ПАРАЛЛЕЛЬНЫЕ АЛГОРИТМЫ МОДЕЛИРОВАНИЯ ДИНАМИЧЕСКИХ ПРОЦЕССОВ НА КОМПЛЕКСНЫХ СЕТЯХ

Предложен способ формального описания динамики комплексных сетей. Рассмотрены схемы распараллеливания алгоритмов при моделировании динамических процессов в комплексных сетях.

Ключевые слова: *комплексная сеть, динамический процесс, случайный граф, схема распараллеливания.*

Комплексная сеть представляет собой граф с динамически эволюционирующей во времени структурой, при этом число состояний графа предполагается достаточно большим, чтобы интерпретировать закономерности процессов на нем средствами статистической механики [1]. Примерами комплексных сетей служат социальные сети (знакомств,

соавторства ученых [2]), информационные (цитирования в научных статьях [3], интернет-ссылок [4]), технологические (Интернет как сеть компьютеров, транспортные и электрические сети) и биологические (сети нейронов мозга, взаимодействующих протеинов, генетические сети). Развитие методов прямого моделирования процессов на комплексных сетях во многом обусловлено возможностями их реализации на параллельных вычислительных архитектурах, как для ускорения вычислительных операций, так и для работы со структурами данных большого объема.

В настоящей работе, продолжающей исследования авторов [5], предлагается формальный математический аппарат, позволяющий описывать динамические системы на основе моделей комплексных сетей в различных предметных областях, и рассматриваются аспекты параллельной реализации таких сетей.

Микромасштабная вероятностная модель динамики комплексной сети. Комплексная сеть представляется взвешенным неориентированным графом G , который определяется как совокупность (V, E) конечного множества вершин V , $\dim(V) = N$, и множества ребер E , состоящего из неупорядоченных пар (u, v) , где $u, v \in V$ и $u \neq v$. Каждая вершина характеризуется своей степенью, т.е. числом инцидентных ей ребер. Микромасштабная модель описывает динамику комплексной сети на основе знаний о состоянии каждой из вершин (и соответствующих им ребер) сети.

Комплексные сети от других графовых структур отличает возможность их вероятностного описания. Она не ограничивается частотным определением вероятности, пригодным для сетей с очень большим количеством узлов, но формально позволяет ввести вероятностное пространство $(\Omega, B_\Omega, P_\Omega)$, включающее в себя следующие элементы.

1. Ω — пространство элементарных событий. Пусть V_i — множество всех вершин веса i (возможно, бесконечное), а $E_{i,k}$ — множество всевозможных графов-звеньев, инцидентных паре вершин, одна из которых имеет вес i , а другая — k : $E_{i,k} = \{e = \{u, v\} : u \in V_i, v \in V_k\}$, тогда $\Omega = \{v \in V_i, i = 1, \dots, N_1; e \in E_{i,k}, i, k = 1, \dots, N_1\}$.

2. B_Ω — сигма-алгебра подмножеств Ω . Любой граф, содержащий вершины весов $i = 1, \dots, N_1$ может быть составлен из элементов множества Ω и соответственно рассмотрен как подмножество множества Ω .

3. P_Ω — сигма-аддитивная мера на множестве Ω (вероятностная мера), отражающая вероятностные закономерности формирования топологии комплексной сети. $P_\Omega(G) = p_G$ — вероятность того, что из всех возможных графов (элементов B_Ω) комплексная сеть представляется графом G .

Использование формального представления вероятностного пространства позволяет описывать динамические процессы на сетях даже в тех случаях, когда число узлов графа ограничено. В целом эволюция комплексной сети может быть описана посредством оператора Γ сети, действующего на множестве B_Ω :

$$\begin{aligned} (V, E, f)_{t+1} &= \Gamma(V, E, f)_t, \\ (V, E, f)_{t=0} &= (V_0, E_0, f_0). \end{aligned} \tag{1}$$

Оператор Γ может быть представлен как совокупность нескольких различных операторов

$$\Gamma = \bigotimes_{m=1}^M \Gamma_m, \tag{2}$$

каждый из которых соответствует определенному динамическому компоненту эволюции сети, в общем случае это добавление новых вершин (Γ_1), удаление из сети вершин (Γ_2), добавление новых связей (Γ_3), разрушение существующих связей, удаление ребер (Γ_4). В общем случае эти операторы некоммутативны. Например, при эволюции комплексной сети по типу случайного графа

$$\Gamma = \Gamma_3 \otimes \Gamma_1 \otimes \Gamma_2 \otimes \Gamma_4. \quad (3)$$

Будем предполагать, что из сети на каждом шаге (в каждый момент времени) могут быть удалены только те вершины, которые не имеют инцидентных им ребер. При этом всякая вершина, не имеющая инцидентных ребер, может быть удалена из сети с вероятностью q_V , а любое ребро может быть удалено из сети с вероятностью q_E . Далее, пусть p_V — вероятность появления одной новой вершины в сети, а p_E — вероятность появления ребра, инцидентного двум произвольным вершинам. Компоненты эволюционного оператора (3) можно представить, используя описание комплексной сети как вероятностного пространства.

Пусть B_e — множество ребер, удаляемых из графа G в момент времени t (G_t), $B_e \subset B_\Omega$. Тогда для оператора Γ_4 , учитывая $P(G_t \cap B_e) = q_E^{\#B_e}$, получаем

$$\Gamma_4 = \langle G_t / (G_t \cap B_e), B_e \subset E_t | P(G_t \cap B_e) \rangle = \langle G_t / (G_t \cap B_e), B_e \subset E_t | q_E^{\#B_e} \rangle. \quad (4)$$

Здесь индекс # означает мощность множества.

Обозначим через B_v множество вершин, удаляемых из графа G_t , $B_v \subset B_\Omega$. Тогда для оператора Γ_2 , учитывая $P(G_t \cap B_v) = q_V^{\#B_v}$, получаем

$$\Gamma_2 = \langle G_t / (G_t \cap B_v), B_v \subset V_t | P(G_t \cap B_v) \rangle = \langle G_t / (G_t \cap B_v), B_v \subset V_t | q_V^{\#B_v} \rangle, \quad (5)$$

пусть теперь B_v — множество новых вершин, добавляемых в граф G_t , $B_v \subset B_\Omega$, тогда оператор Γ_1 , с учетом $P(B_v) = p_V^{\#B_v}$, имеет вид

$$\Gamma_1 = \langle G_t \cup B_v, B_v \subset V_t | P(B_v) \rangle = \langle G_t \cup B_v, B_v \subset V_t | p_V^{\#B_v} \rangle, \quad (6)$$

B_e — множество новых ребер, добавляемых в граф G_t , $B_e \subset B_\Omega$, оператор Γ_3 , с учетом $P(B_e) = p_E^{\#B_e}$, имеет вид

$$\Gamma_3 = \langle G_t \cup B_e, B_e \subset E_t | P(B_e) \rangle = \langle G_t \cup B_e, B_e \subset E_t | p_E^{\#B_e} \rangle. \quad (7)$$

Дополнительно в (3) могут участвовать дополнительные операторы Θ , отвечающие за эволюцию отдельных узлов сети (переход из состояния в состояние); могут использоваться операторы Γ_2 и Γ_4 .

В качестве примера на рис. 1 приведены три формы развития процессов в контактной сети носителей ВИЧ $l \in N_1$ (○ — „здоровый“, ⊗ — „инфицированный“, ● — „больной/умерший“, ⊙ — новый узел, которого не было) [6]. Каждый узел может быть „инфицирован“; по истечении инкубационного периода он выбывает из сети; при этом соответствующие связи удаляются из сети. На рис. 1, а представлен процесс в „замороженной“ сети со статическими связями и постоянным количеством узлов. На рис. 1, б количество узлов сети постоянно, но связи между ними могут изменяться в течение времени вне зависимости от

действия оператора Θ . На рис. 1, в сеть имеет переменное количество узлов и изменяемые связи между ними, т.е. задействованы все составляющие (3).

Модель (3)—(7) описывает эволюцию сети как случайного графа. Однако в ряде случаев такое приближение не является удовлетворительным. Это требует перейти от описания событий, определяемых (4)—(7) в форме модели Бернулли, к более сложным вероятностным схемам, например, с использованием условных распределений добавления (удаления) связи у вершины, которая уже имеет k связей (preferential attachment [7]).

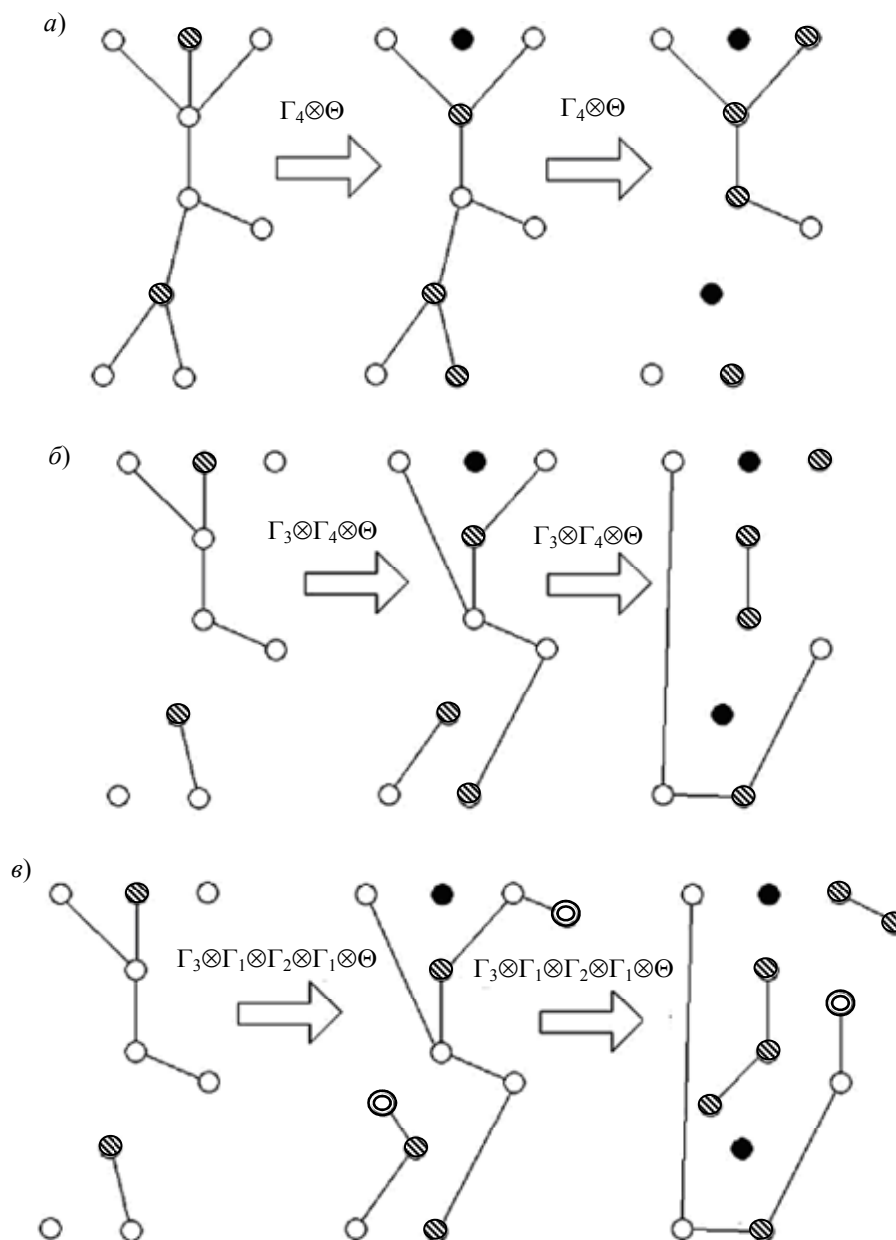


Рис. 1

Макромасштабная вероятностная модель динамики комплексной сети. Макроскопическая модель описывает динамику процесса в терминах вероятностных характеристик сети (общее количество вершин, общее количество ребер, количество вершин каждого конкретного веса), которые, в свою очередь, могут быть получены усреднением значений величин, характеризующих микроскопические свойства сети. Эти величины являются интегральными количественными характеристиками воздействия эволюционного оператора (2) на сеть.

Применяя к (1) оператор определения числа вершин заданного веса $\rho_l = \rho_l(t)$ — количество вершин веса l , $l \in N_1$, можно получить систему дифференциальных уравнений

$$\frac{d\rho_l(t)}{dt} = F(\rho_1(t), \rho_2(t), \dots, \rho_{N_1}(t), \lambda_1, \lambda_2, \dots, \lambda_p), \quad (8)$$

где λ_i — макропараметры, характеризующие процессы в конкретной сети, определяемые, например, через вероятности в соотношениях (4)—(7). Следует отметить, что сама по себе такая модель является детерминированной и не отражает стохастического характера процессов в комплексной сети.

Для приведенного выше примера эволюции сети (3)—(7) в форме случайного графа система (8) вырождается в единственное дифференциальное уравнение, описывающее зависимость $\rho(t)$, которая определяется балансом между количеством удаленных из сети вершин и количеством новых в каждый момент времени t . Всякая вершина сети, имеющая степень k , в результате действия оператора $\Gamma_2 \otimes \Gamma_4$ удаляется из сети с вероятностью $q_E^k q_V$, тогда как под действием оператора $\Gamma_3 \otimes \Gamma_1$ с вероятностью $p_E^k p_V$ появляется новая, имеющая степень k :

$$\frac{d\rho(t)}{dt} = \rho(t) \sum_{k=0}^K (p_E^k p_V - q_E^k q_V) p_k(t), \quad (9)$$

где K — максимальная степень вершины в сети, а $p_k = p_k(t)$ — закон распределения степеней.

В общем случае вид уравнений (8) зависит от характеристик распределения связей в структуре сети. Например, для эпидемиологической сети ВИЧ система (8) приобретает вид:

$$\left. \begin{aligned} \frac{ds_k(t)}{dt} &= -s_k(t) \sum_{l=1}^k f(l, \lambda) L(k, l, p), \\ \frac{d\rho_k(t)}{dt} &= -\mu \rho_k(t) + s_k(t) \sum_{l=1}^k f(l, \lambda) L(k, l, p), \\ \frac{dr(t)}{dt} &= \sum_k \mu \rho_k(t). \end{aligned} \right\} \quad (10)$$

Здесь s обозначает число индивидов, r — число изолированных от контактов (заболевших или умерших), ρ_k — количество инфицированных узлов, у которых есть ровно k связей. При этом $L(k, l, p) = C_k^l p^l (1-p)^{k-l}$ задает биномиальное распределение контактов, где k — общее число узлов в группе, а l — доля узлов, подверженных заражению. Функция $f(l, \lambda) = 1 - (1-\lambda)^l$ определяет вероятность изоляции узла (переход в группу r). Здесь λ — вероятность изоляции, а μ — вероятность заражения, интерпретируемые как макропараметры сети.

Модели в форме (8)—(10) являются более простыми с вычислительной точки зрения, однако менее гибкими, чем модель (1)—(7). Однако они позволяют анализировать основные характеристики процессов на сетях в терминах, присущих классическим динамическим системам. Например, на рис. 2 приведены результаты моделирования (кривая) развития эпидемии ВИЧ в США посредством модели (10), точки — результат измерений. На рис. 2, а, б представлено количество заболевших (r) и фазовый портрет эпидемии в среде гомосексуального населения (dr), а на рис. 2, в, г — в среде гетеросексуального населения. Из рисунка

видно, что фазовые портреты для различных популяций имеют принципиально разную форму, это является, в первую очередь, следствием различия структуры контактной сети.

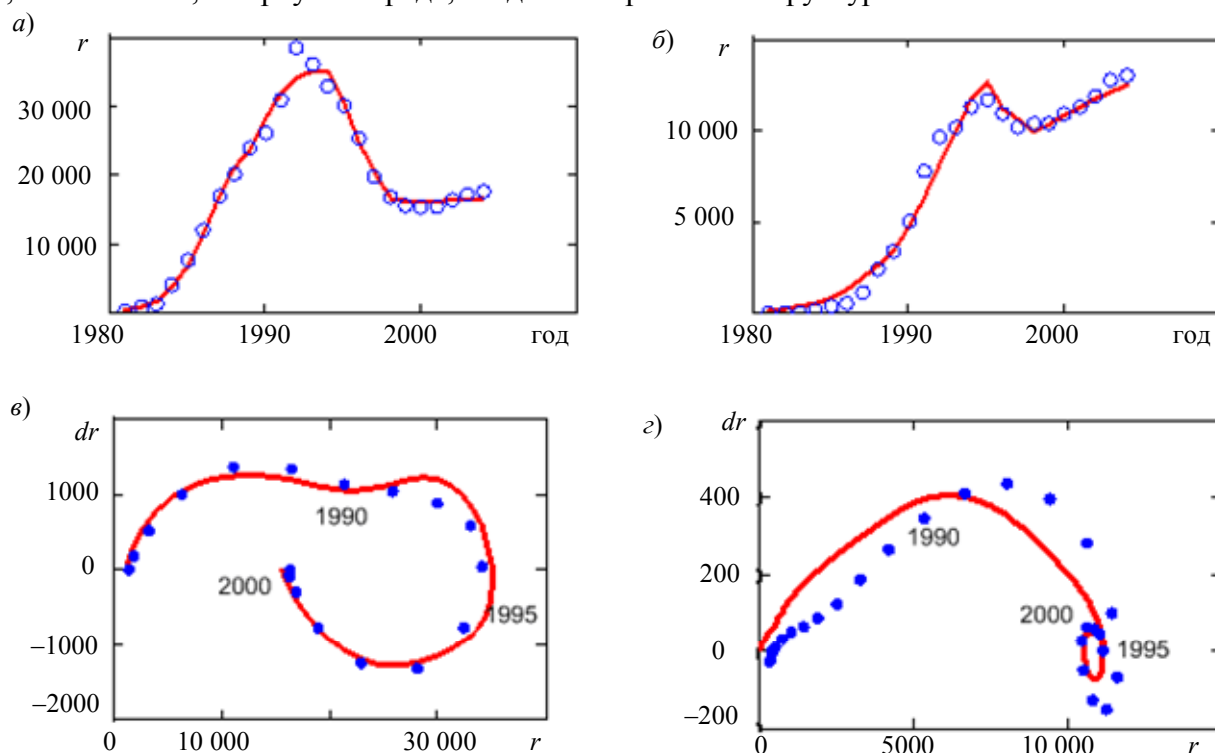


Рис. 2

Параллельные алгоритмы моделирования процессов на комплексных сетях. В последовательном алгоритме моделирования динамики комплексной сети по модели (1)—(7) условно можно выделить следующие этапы: формирование сети в начальный момент времени, эволюция сети, сбор и анализ необходимых данных (расчет макропараметров). Параллельные алгоритмы формирования сети заданной структуры подробно рассмотрены в работе [5]. Для моделирования эволюционной динамики сети можно рассмотреть следующие схемы распараллеливания для p вычислителей.

Островная схема. На каждом из вычислителей независимо формируется сеть размером $N_i = N/p$, после чего к ней применяется оператор (2). Результаты расчета макропараметров усредняются по всем p вычислителям. Такая схема отличается простотой реализации, однако имеет ограниченную эффективность, поскольку скорость сходимости статистических оценок (как следствие — параллельное ускорение) в данном случае будет пропорциональна \sqrt{p} . Однако ее применение может быть оправдано в том случае, когда макропараметры расчета варьируются от вычислителя к вычислителю (т.е. моделируется неэргодический процесс).

Блочная схема. Для применения оператора эволюции (2) необходимо сеть разделить на блоки (количество блоков $2p$) с одинаковым числом узлов сети $N_i = N/p$ в каждом, при этом любой из полученных блоков будет связан со всеми остальными. Применение оператора (2) осуществляется за одну итерацию для пары блоков на одном вычислителе, после чего происходит пересылка блоков между процессорами, этот процесс повторяется до тех пор, пока блоки не провзаимодействуют во всех возможных сочетаниях в соответствии с заранее построенной „турнирной“ таблицей. Реализация блочной схемы требует существенных накладных расходов на передачу данных между вычислителями на каждом шаге моделирования.

Для сопоставления производительности параллельных алгоритмов в рамках островной и блочной схем рассмотрим кривую изоэффективности, определяемую через равенство параллельных ускорений для обоих алгоритмов:

$$p(nm(1+\varphi) - 2N\bar{d}n\varphi) = N\bar{d} + 24N + nm. \quad (11)$$

Здесь N — количество вершин в сети, \bar{d} — средняя степень вершины, n — число шагов эволюционного алгоритма, m — количество макропараметров сети, подлежащих исследованию, $\varphi = V_0\lambda$, где λ — безразмерный коэффициент, определяющий отношение времени, затрачиваемого на пересылку единицы одного объема данных V_0 между вычислителями, ко времени, затрачиваемому на одну арифметическую операцию. При этом оптимальное количество вычислителей, приводящих к наибольшему ускорению, вычисляется как

$$p_{\max}^{\text{б.с}} = \sqrt{\frac{\delta + n\tau}{2n\varphi}}, \quad p_{\max}^{\text{о.с}} = \sqrt{\frac{24N + N\bar{d}(1 + \delta + n\tau) + nm}{nm(1 + \varphi)}}, \quad (12)$$

δ — безразмерный параметр, характеризующий отношение времени, затрачиваемого на создание ребра, к времени на одну вычислительную операцию, τ — безразмерный коэффициент, характеризующий отношение числа перестраиваемых вершин сети к общему количеству узлов N в сети. Из сопоставления (11) и (12) видно, что во всех случаях до достижения пика ускорения производительность островной схемы будет выше, чем у блочной. Однако этот приоритет является условным, поскольку при использовании островной модели не учитываются связи между отдельными подсетями, которые моделируются независимо друг от друга на различных вычислителях, число таких связей может быть $N\bar{d}/p$. Поэтому выбор схемы распараллеливания должен целиком определяться целями исследования и учитывать не только формальное ускорение, но и соответствие целей расчета предметной постановке задачи (например, моделирования связной сети не менее определенного размера).

Заключение. Динамически изменяющиеся комплексные сети являются эффективным инструментом описания сложных физических и социальных явлений. Для описания динамических моделей на базе комплексных сетей в работе предложен формальный математический аппарат, учитывающий основные аспекты сетевых моделей (стохастический характер сети, иерархию уровней моделирования и т.п.). Для проведения эффективных исследований сетевых моделей проанализированы основные подходы к распараллеливанию алгоритмов построения и эволюции комплексных сетей. В качестве примера для демонстрации удобства описания моделей на базе предложенного формализма и подтверждения эффективности параллельных алгоритмов была использована сетевая модель распространения инфекции ВИЧ.

Работа выполнена в рамках реализации Постановлений № 218 и 220 Правительства Российской Федерации при частичной поддержке ФЦП „Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2007—2012 гг.“.

СПИСОК ЛИТЕРАТУРЫ

1. Newman M. E. J. The Structure and Function of Complex Networks // SIAM Rev. 2003. Vol. 45, N 2. P. 167—256.
2. Redner S. How popular is your paper? An empirical study of the citation distribution // Europ. Phys. J. B. 1998. Vol. 4. P. 131—134.
3. Price D. J. de S. A general theory of bibliometric and other cumulative advantage processes // J. of the American Soc. for Information Sci. 1976. Vol. 27. P. 292—306.
4. Broder A., Kumar R., Maghoul F. et al. Graph structure in the web // Computer Networks. 2000. Vol. 33. P. 309—320.
5. Иванов С. В., Колыхматов И. И., Бухановский А. В. Параллельные алгоритмы моделирования комплексных сетей // Изв. вузов. Приборостроение. 2008. Т. 51, №10. С. 5—12.
6. Sloot P. M. A., Ivanov S. V. et al. Stochastic simulation of HIV population dynamics through complex network modeling // Intern. J. of Computer Mathematics. 2008. P. 1175—1187.

7. Barabási A.-L., Albert R. Emergence of scaling in random networks // Science. 1999. Vol. 286 (5439). P. 509—512.

Сведения об авторах

- Екатерина Владимировна Болгова** — НИИ Научно-технических компьютерных технологий Санкт-Петербургского государственного университета информационных технологий, механики и оптики; младший научный сотрудник; E-mail: katerina.bolgova@gmail.com
- Сергей Владимирович Иванов** — канд. техн. наук; НИИ Научно-технических компьютерных технологий Санкт-Петербургского государственного университета информационных технологий, механики и оптики; старший научный сотрудник; E-mail: Sergey.v.ivanov@rambler.ru
- Екатерина Александровна Гринина** — канд. физ.-мат. наук; Санкт-Петербургский государственный университет информационных технологий, механики и оптики; E-mail: grinina@tprs.ru
- Петрус Мария Арнольдус Слоот** — доктор философии в области компьютерных наук, профессор; Университет города Амстердам; E-mail: p.m.a.sloot@uva.nl
- Александр Валерьевич Бухановский** — д-р техн. наук, профессор; НИИ Научно-технических компьютерных технологий Санкт-Петербургского государственного университета информационных технологий, механики и оптики; директор; E-mail: avb_mail@mail.ru

Рекомендована НИИ НКТ

Поступила в редакцию
15.05.11 г.