



НАНОСТРУКТУРЫ

математическая физика и моделирование



Nanostructures. Mathematical physics & modelling

НАНОСТРУКТУРЫ

математическая физика и моделирование

Nanostructures.
Mathematical Physics & Modelling

2016, volume 14(2)

Наноструктуры. Математическая физика и моделирование

Редколлегия:

**И.В. Волович, В.Л. Голо, В.В. Гусаров, П.Н. Дьячков, Р.Г. Ефремов,
М.В. Карасев (зам. главного редактора), Ю.Е. Лозовик, М.А. Мазо,
В.П. Маслов (главный редактор), А.В. Махиборода (ответственный
секретарь), А.Ю. Морозов, С.А. Никитов, Р.А. Сурис, А.Н. Тихонов,
В.А. Тулин, В.Е. Фортов, А.С. Холево, А.Р. Хохлов, А.В. Чаплик,
Л.А. Чернозатонский, К.В. Шайтан**

Электронная версия журнала размещается на сайте
<http://nano-journal.ru>

Адрес редакции:
123458, Москва, ул. Таллинская, д. 34, каб. 429
+7 (495) 916-88-76
nanostructures@hse.ru

Москва

© 2016, Европейский центр по качеству

Содержание

И.А. Амелюшкин, А.Л. Стасенко

Взаимодействие нанокapелъ аэрозольного потока с твердым телом 5

А.В. Махиборода, А.В. Ильичёв, А.А. Подобин, А.В. Царёв

Проблемы реализации массового динамического параллелизма. II 25

В.А. Морозов

**Математическое моделирование динамики заселенности состояний
трехуровневой наночастицы при спектроскопических переходах 69**

А.А. Тужилин

Продолжаем чистить Protein Data Bank 85

Информация и правила для авторов 101

Contents

I.A. Amelyushkin ¹ , A.L. Stasenko	
Interaction of aerosol flow nanodroplets with a solid body	5
A.V. Makhiboroda, A.V. Ilichev, A.A. Podobin, A.V. Tsarev	
Problems of implementation of massive dynamic parallelism. II	25
V.A. Morozov	
Mathematical modeling of states population dynamics for a three-level nanoparticle in spectroscopic transitions	69
A.A. Tuzhilin	
Protein Data Bank: Next Cleaning	85
The information and rules for authors	101

ВЗАИМОДЕЙСТВИЕ НАНОКАПЕЛЬ АЭРОЗОЛЬНОГО ПОТОКА С ТВЕРДЫМ ТЕЛОМ*

И.А. Амелюшкин¹, А.Л. Стасенко²

¹ *Центральный аэрогидродинамический институт (ЦАГИ);*

² *Московский физико-технический институт (МФТИ)*

Amelyushkin_Ivan@mail.ru, Stasenko@serpantin.ru

Поступила 16.05.2016

Развита физико-математическая модель и соответствующий численный код для исследования взаимодействия капель воды с твердым телом и предсказания начала его обледенения в аэрозольном влажном потоке. Методом молекулярной динамики получены зависимости макрохарактеристик малых капель от их размеров и дано сравнение с другими теоретическими исследованиями. Приведены оценки области параметров аэрозольного потока и физико-химических свойств материала обтекаемого тела, обеспечивающих отсутствие его обледенения.

УДК 533.6.071.08.632.57

1. Введение

Процесс обледенения летательных аппаратов исследуется на протяжении почти всей истории авиации. Экспериментальное и численное исследования роста льда на телах, помещенных в поток воздуха, содержащего капли, проводились более 80 лет (напр. [1, 2]). Физические процессы, которые происходят до того, как частицы воды достигают поверхности элементов конструкции летательного аппарата, были исследованы в [3, 4]. Кристаллизация и примерзание к поверхности летательного аппарата переохлажденной капли после удара являются фундаментальными проблемами механики многофазных течений и физики конденсированного состояния.

* Работа выполнена при поддержке Российского Научного Фонда, грант 16-19-10472

При любой температуре газ состоит не только из молекул (мономеров), но также из молекулярных кластеров – наночастиц со специфическими свойствами и законами поведения. В частности, их присутствие не делает необходимым достижения критического зародыша a_c^* бесконечно большого радиуса при насыщении пара ($p_v = p_{sv}(T)$), значение которого следует из известной формулы Томсона $a_c^* = 2\sigma_l / \rho_l R_v T \ln(p_v / p_{sv})$, где σ_l , ρ_l – коэффициент поверхностного натяжения и плотность капли соответственно, R_v – удельная газовая постоянная пара. Индексы: l – жидкость, v – пар, sv – насыщенный пар.

В настоящей работе под каплей понимается молекулярный кластер, состоящий более чем из $N = 100$ молекул. Эволюцию таких частиц исследовать методами квантовой химии чрезвычайно сложно. Следует отметить, что определение даже агрегатного состояния этой капли становится проблематичным: часть ее молекул ведет себя как молекулы газа, другие – как молекулы жидкости, в то время как некоторые имеют кинетическую энергию, мгновенное значение которой соответствует температуре, близкой к абсолютному нулю. После удара о холодную поверхность летательного аппарата частицы воды могут сформировать тонкий слой толщиной в несколько диаметров молекул, который далее будет способствовать интенсивному росту льда уже независимо от физико-химических свойств материала обтекаемого тела и его рельефа.

Традиционно используемые теоретические и численные подходы могут оказаться необоснованными, поскольку известные законы механики сплошной среды и кинетической теории для малых частиц не выполняются. Использование методов квантовой химии [5], которые основаны на приближенном численном решении уравнения Шредингера, могут привести к неоднозначности решений и чрезвычайно трудоемки для случая больших молекулярных кластеров ($N \geq 20$). При уменьшении характерных размера или времени течений жидкости и газа экспериментальное исследование становится также затруднительным [6]. Методы же частиц в гидродинамике (MPS – moving particle simulations [7] и SPH – smoothed particle hydrodynamics [8]) недостаточно адекватно описывают теплофизику и фазовые переходы жидкости, взаимодействующей с обтекаемым телом и окружающим потоком газа.

В настоящей работе взаимодействие частиц воды с твердым телом (в частности, с летательным аппаратом), обтекаемым влажным аэрозольным потоком, исследуется методом молекулярной динамики. Благодаря современной вычислительной технике, данный подход позволяет исследовать физические процессы в объемах материала размером до кубического микрометра [9], что соответствует миллиардам частиц.

В отличие от аналогичных исследований (например, [10]), в настоящей работе учтен ряд существенных физических явлений, в частности, наличие воздуха и более точная модель поведения молекул твердого тела. Развита модель взаимодействия частицы воды с поверхностью летательного аппарата, позволяющая оценить условия полета, при которых частицы воды с самого начала не будут примерзать к поверхности летательного аппарата.

Что касается броуновской динамики, в предлагаемой статье рассматривается капля воды в «последний момент» перед столкновением с поверхностью твердого тела; поэтому влияние стохастических движений по мере ее приближения к поверхности окружающего воздуха не учитывается.

2. Процесс формирования малой частицы при стремлении к термодинамическому равновесию

Поскольку размер и структура наночастицы, состоящей из заданного числа молекул, заранее неизвестны, прежде всего, исследуется процесс ее стремления к внутрен-

нему равновесию и к равновесию с несущей средой при фиксированных значениях температуры, плотности и состава несущей газовой смеси.

Метод молекулярной динамики, основанный на классической ньютоновской механике, применялся ранее без учета присутствия воздуха в [10] и в [11] для малых молекулярных соединений. В настоящей работе использован известный сферически симметричный потенциал парного межмолекулярного взаимодействия Леннарда-Джонса (ЛД) при исследовании свойств жидкости, осредненных по времени и по объему. Как известно, зависимость потенциальной энергии парного взаимодействия $U(r_{ij})$ от расстояния между центрами молекул r_{ij} может быть представлена в виде суммы энергии ориентационного, индукционного и дипольного взаимодействия, а также энергии отталкивания:

$$U(r_{ij}) = U_{\text{or}}(r_{ij}) + U_{\text{ind}}(r_{ij}) + U_{\text{disp}}(r_{ij}) + U_{\text{rep}}(r_{ij});$$

$$U_{\text{or}}(r_{ij}) = -\frac{p_e^4}{24\pi^2 \varepsilon_0^2 k_B T} \frac{1}{r_{ij}^6}; \quad U_{\text{ind}}(r_{ij}) = -\frac{\alpha p_e^2}{8\pi^2 \varepsilon_0^2} \frac{1}{r_{ij}^6}; \quad U_{\text{disp}}(r_{ij}) = -\frac{3}{4} \alpha^2 E_I \frac{1}{r_{ij}^6}; \quad U_{\text{rep}}(r_{ij}) \sim \frac{1}{r_{ij}^{12}}.$$

Здесь p_e – дипольный момент молекулы, α – ее поляризуемость, E_I – энергия ионизации, ε_0 и k_B – электрическая постоянная и постоянная Больцмана соответственно. Первое, второе и третье слагаемые уравнения представляют собой силы притяжения, которые действуют между i -й и j -й молекулами. Энергия отталкивания соответствует известному квантово-механическому принципу запрета Паули, согласно которому два одинаковые фермиона (частицы с полуцелым спином) не могут одновременно находиться в одном и том же квантовом состоянии. Выражение для силы F_i , действующей на i -ю молекулу воды, имеет следующий вид:

$$F_i(r_i(t)) = \sum_{j=1}^{N_{\text{H}_2\text{O}}} \left\{ -\nabla U_{\text{H}_2\text{O}}^{ij} \right\} + \sum_{k=1}^{N_{\text{Air}}} \left\{ -\nabla U_{\text{H}_2\text{O-Air}}^{ik} \right\} + \sum_{m=1}^{N_{\text{Wall}}} \left\{ -\nabla U_{\text{H}_2\text{O-Wall}}^{im} \right\}.$$

Частицы воды, состоящие из $50 \div 15000$ молекул, были сформированы и уравновешены при различных температурах, соответствующих температурам полета летательного аппарата в аэрозольном потоке.

В работе учтены потенциалы взаимодействия не только молекул воды друг с другом, но и с молекулами азота, составляющими основную часть воздуха (кислород не рассматривался), а также с атомами твердого тела.

Расчетное пространство вокруг частицы – кубическая область; условия симметрии для молекул поставлены для всех границ с газом. Отметим, что принятый в настоящей работе характерный размер кубической расчетной области на один-два порядка больше характерного размера рассматриваемой наночастицы. На «дне» кубической области молекулы воды и воздуха взаимодействуют с атомами твердого тела в зависимости от их физических свойств и температуры стенки. В предположении потенциального характера взаимодействия между атомами или молекулами уравнения движения численно интегрировались согласно алгоритмам, близким к описанным в [11]. Приведем предложенный в [12] и использованный нами численный метод интегрирования уравнений движения молекул, который обеспечивает хорошую консервативность:

$$r_i(t + \Delta t) = r_i(t) + v_i(t)\Delta t + \left(\frac{2}{3} a_i(t) - \frac{1}{6} a_i(t - \Delta t) \right) \Delta t^2,$$

$$\mathbf{a}_i(t + \Delta t) = \frac{-\sum_j \nabla U_{ij} (\mathbf{r}_i(t + \Delta t) - \mathbf{r}_j(t + \Delta t))}{m_i},$$

$$\mathbf{v}_i(t + \Delta t) = \mathbf{v}_i(t) + \left(\frac{1}{3} \mathbf{a}_i(t + \Delta t) + \frac{5}{6} \mathbf{a}_i(t) - \frac{1}{6} \mathbf{a}_i(t - \Delta t) \right) \Delta t.$$

Здесь $\mathbf{r}_i$, $\mathbf{v}_i$ и $\mathbf{a}_i$ – радиус-вектор, скорость и ускорение i -й молекулы, соответственно, $\mathbf{r}_j$ – радиус-вектор j -й молекулы.

Во всех расчетах задавались и контролировались значения основных характеристик молекулярной системы (температуры, плотности, давления, распределения молекул по скоростям) в системе отсчета, связанной с центром масс частицы. При вычислении скорости центра масс частицы предполагалось, что молекула принадлежит частице, если в малой окрестности молекулы плотность частицы больше 10 кг/м^3 , что определяет границу частицы.

В численных расчетах оказалось, что «выращиваемая» наночастица изменяет свою форму квазипериодично: в результате возникают тепловые волны, которые обусловлены неоднородным распределением энергии молекул при их тепловом движении.

Ниже определены макрохарактеристики частиц воды в терминах параметров межмолекулярного взаимодействия. Параметры потенциала Леннарда-Джонса взяты из констант уравнения состояния Ван-дер-Ваальса: $\sigma_{\text{H}_2\text{O}} = 2.6 \text{ \AA}$, $\varepsilon_{\text{H}_2\text{O}} = 809 k_B \cong 10^{-20} \text{ Дж}$.

2.1. Распределение молекул по скоростям

В начальный момент времени молекулам были приданы скорости согласно известному закону распределения Максвелла–Больцмана в системе отсчета, связанной с центром масс капли. В ходе численного моделирования капли воды наиболее быстрые молекулы покидали каплю и, как следствие, изменялся закон распределения молекул по скоростям. В конце концов наступало динамическое равновесие. Амплитуды колебаний уменьшались с увеличением начального числа молекул в капле.

На рис. 1а показано распределение молекул воды по скоростям в каплях различных размеров (штрих-пунктирные линии) и известная функция Максвелла, соответствующая заданной температуре, а также изменение распределения молекул по скоростям с течением времени.

2.2. Радиальное распределение температуры

Локальное значение температуры рассчитывалось как средняя кинетическая энергия молекул в системе центра масс частицы, движущейся в аэрозольном потоке:

$$T(r_k) = \frac{2}{3} \frac{1}{k_B} \sum_{i=1}^{\Delta N_k(r_k)} \frac{m_{\text{H}_2\text{O}} |\mathbf{v}_i|^2}{2}.$$

На рис. 1б показаны результаты расчета радиального распределения температуры и его изменения со временем в каплях различных размеров. Видно, что в связи с малым количеством молекул в капле и существенно неравномерным распределением их по скоростям локальное значение температуры существенно колеблется около среднего значения 0°C . Следует также отметить, что распределение температуры внутри капли существенно неоднородно.

Численный эксперимент показал, что кинетическая энергия молекул (как и температура) внутри капли меньше, чем на межфазной границе. Этот эффект, как и большинство других молекулярных эффектов, уменьшается с увеличением размера капли.

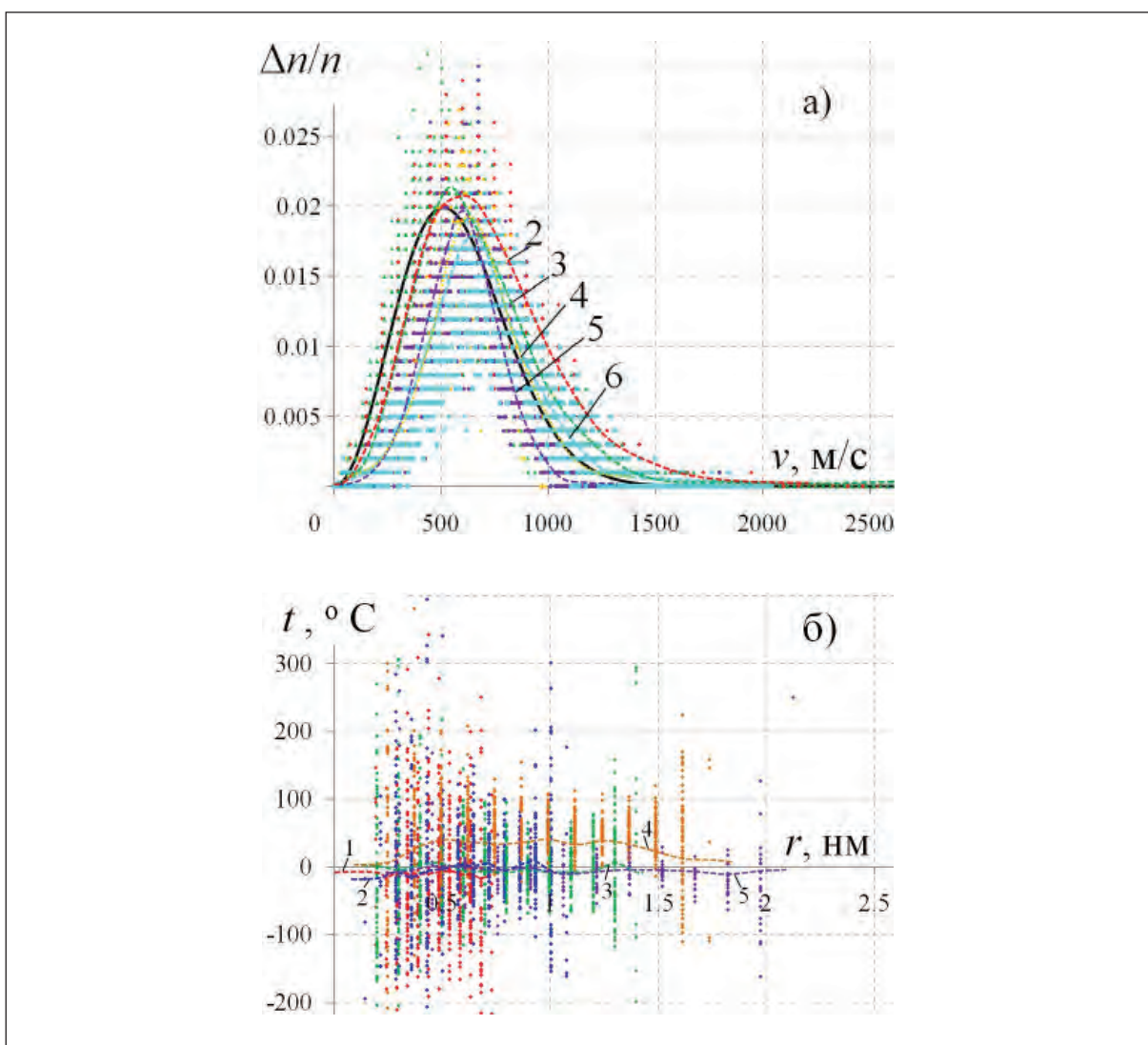


Рис. 1. а) Распределение молекул по скоростям внутри каплей различных размеров и его сравнение с функцией Максвелла (сплошная линия); б) радиальное распределение температуры в капле воды; кривые 1, 2, 3, 4, 5 и 6 соответствуют каплям, состоящим из 55, 160, 350, 650, 1100 и 2500 молекул соответственно.

2.3. Распределение локальной величины плотности

Тепловое движение молекул приводит к колебаниям плотности. Для расчета радиального распределения плотности $\rho(r)$ (r – расстояние от центра масс частицы) частицы разобьем ее на сферические слои. Разделив число молекул, попавших в каждый слой, на объем этого слоя ΔV_d и умножив на массу одной молекулы $m_{\text{H}_2\text{O}}$, получим выражение для радиального распределения массовой плотности в капле воды:

$$\rho(r_k) = m_{\text{H}_2\text{O}} \Delta N(r_k) / \Delta V(r_k);$$

$$\Delta V(r_k) = (4/3)\pi(r_k + \Delta r/2)^3 - (4/3)\pi(r_k - \Delta r/2)^3.$$

На рис. 2а показана зависимость радиального распределения локальной массовой плотности в каплях различных размеров при температуре окружающего влажного воздуха приблизительно 10 °C (относительная влажность воздуха приблизительно 5 %).

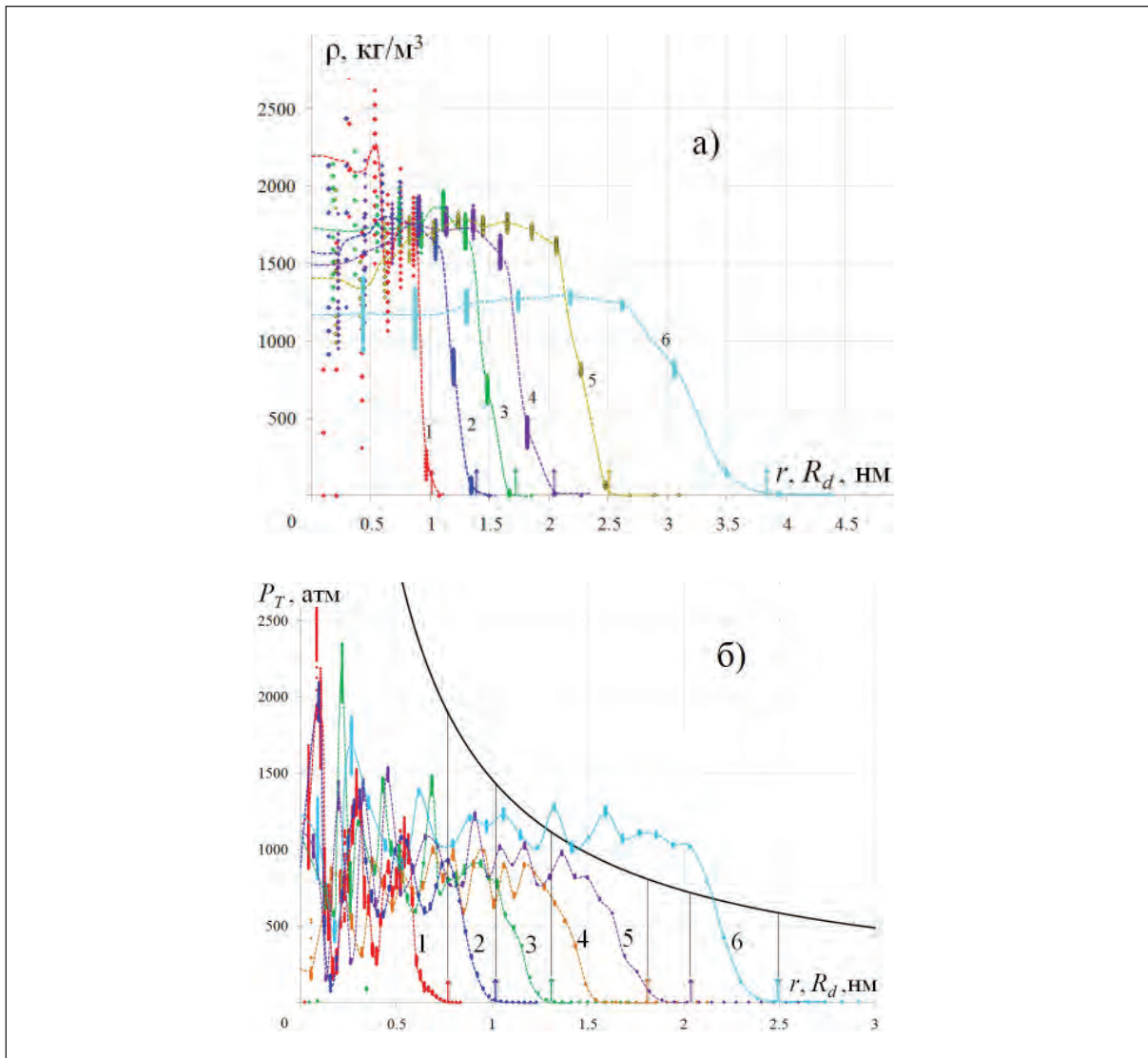


Рис. 2. а) Радиальное распределение массовой плотности частиц воды в зависимости от их «размеров» R_d , указанных вертикальными стрелками; б) радиальные распределения давления внутри малых капель воды в зависимости от их размеров и давление Лапласа (сплошная кривая)

Первая кривая соответствует частице, состоящей из $N = 160$ молекул; последняя (шестая) – из 4660 молекул. Видно, что плотность частицы увеличивается с уменьшением ее размера. С ростом радиуса капли R_d ее плотность стремится к табличному значению $\rho_\infty = 1000 \text{ кг/м}^3$. Для сравнения приведем оценку характерной плотности «частицы» воды, состоящей из одной молекулы: $\rho(\sigma_{\text{H}_2\text{O}}/2) = m_{\text{H}_2\text{O}} / \left(\frac{4}{3} \pi (\sigma_{\text{H}_2\text{O}}/2)^3 \right) \approx 3250 \text{ кг/м}^3$. Это значение более чем в три раза превышает значение плотности жидкой воды при нормальных условиях.

Следует отметить, что локальное значение массовой плотности частицы больше вблизи межфазной границы и сильно зависит от радиуса кривизны частицы. Рис. 2а показывает плавное уменьшение среднего значения плотности с удалением от межфазной границы к центру частицы. Чем больше «толщина» межфазной границы, тем меньше пик плотности на межфазной границе. Кроме того, тепловые колебания радиального распределения плотности исчезают с увеличением размера частицы в связи с меньшим вкладом теплового движения каждой молекулы воды в

результатирующее значение распределения плотности капли. Напомним, что на основе оптических измерений солнечного света, рассеянного облаками, было обнаружено [14], что плотность капли в атмосферных облаках может быть в 2 – 2.5 раза больше, чем плотность воды при нормальных условиях.

Согласно классическим представлениям [15], межфазная граница имеет переходный слой жидкость-пар, плотность которого постепенно уменьшается вблизи границы. На самой «границе» плотность скачком изменяется на конечную величину, а с удалением от этой «поверхности» плотность пара монотонно падает до термодинамически равновесного значения $\rho_{sv}(T)$, соответствующего заданной температуре. Для оценки характерной толщины δ_σ этого переходного слоя имеется её связь с коэффициентом поверхностного натяжения σ_l и радиусом $\sqrt{\langle r_m^2 \rangle}$ действия когезионных сил:

$$\delta_\sigma = \left(\frac{a_{vw} \langle r_m^2 \rangle}{\sigma_l} \right) \cdot (\rho_l - \rho_{sv}(T))^2 \cong 5 \text{ нм.}$$

Здесь $a_{vw} = 5.5 \cdot 10^3 \text{ Н} \cdot \text{м}^4 / \text{кг}$ – параметр уравнения состояния Ван-дер-Ваальса. Радиус принят равным характерному размеру молекулы воды – параметру длины Леннарда-Джонса $r_m \cong \sigma_{\text{H}_2\text{O}} = 2.6 \text{ \AA}$. Использована интерполяция табличных данных для насыщающих паров воды; коэффициент поверхностного натяжения $\sigma_l = 0.073 \text{ Н/м}$. Таким образом, с классической точки зрения вся наночастица представляет собой «переходный слой».

2.4. Распределение давления

Давление в жидкости складывается из двух составляющих:

$$P(r_k) = P_T(r_k) + P_\sigma(r_k).$$

Первая из них обусловлена кинетической энергией движения молекул. Построим концентрические воображаемые феры радиусов r_k , центры которых совпадают с центром капли. Эта часть давления рассчитана как поток импульса молекул через каждую сферу. Пусть $N(t)$ – количество молекул воды, которые проходят через k -ю сферу в момент времени t . Их радиальный импульс равен

$$\Delta p_r^k(t) = \sum_{i=1}^{N(t+\Delta t)-N(t)} |m_{\text{H}_2\text{O}} \cdot \mathbf{V}_r^{i,k} \cdot \mathbf{i}_r| = m_{\text{H}_2\text{O}} \sum_{i=1}^{N(t+\Delta t)-N(t)} |\mathbf{V}_r^{i,k}|,$$

где $|\mathbf{V}_r^{i,k}| = |\mathbf{V}_r^{i,k}(r_k, \theta_i, \varphi_i)| = v_x^{i,k} \sin \theta_i \cos \varphi_i + v_y^{i,k} \sin \theta_i \sin \varphi_i + v_z^{i,k} \cos \theta_i$ – радиальная составляющая скорости i -й частицы, проходящей через k -ю сферу, $\mathbf{i}_r$ – единичный вектор, направленный из центра капли (орт радиального направления), φ_i и θ_i – азимутальная и полярная координаты (в системе координат, связанной с центром масс капли) i -й молекулы в момент ее прохождения через k -ю сферу.

Таким образом, радиальное распределение давления может быть рассчитано как сила, которая действует на воображаемую поверхность с площадью $4\pi r_k^2$ или суммарный радиальный импульс в единицу времени, приходящийся на единицу площади:

$$P_T(r_k) = \frac{\Delta p_r^k(t)}{4\pi r_k^2 \Delta t} = \frac{m_{\text{H}_2\text{O}}}{4\pi r_k^2 \Delta t} \sum_{i=1}^{N(t+\Delta t)-N(t)} |\mathbf{V}_r^{i,k}(t)|.$$

В настоящей работе значение $\varepsilon_{\text{H}_2\text{O}}/\sigma_{\text{H}_2\text{O}} \cong 6828$ атм. использовано в качестве масштаба давления.

Вторая составляющая в выражении для давления обусловлена полем сил межмолекулярного взаимодействия:

$$P_{\sigma}(r_k) = \frac{1}{4\pi} \int_0^{4\pi} P_{\sigma}(r_k) d\Omega = \frac{1}{4\pi} \int_0^{2\pi} \int_0^{\pi} \frac{1}{4\pi r_k^2} \left(\sum_i \nabla U(\mathbf{r}_i - \mathbf{r}_k) \cdot \frac{\mathbf{r}_k}{|\mathbf{r}_k|} \right) \sin \theta d\theta d\varphi.$$

На рис. 26 показаны результаты численных расчетов радиального распределения давления в каплях воды в зависимости от их размеров (кривые 1, 2 и 3 соответствуют каплям, которые состоят из 650, 1080 и 2450 молекул соответственно). Стрелками показаны «радиусы» этих капель. Сплошная кривая соответствует известной формуле Лапласа $\Delta P = 2\sigma_l / R_d$.

2.5. Распределение напряженности поля (удельной силы) межмолекулярного взаимодействия

Эта величина может быть определена как сумма градиентов энергии межмолекулярного взаимодействия с обратным знаком, деленная на массу одной молекулы:

$$E_{\sigma}(\mathbf{r}) = \frac{-\sum_i \nabla U(\mathbf{r}_i - \mathbf{r})}{m_{\text{H}_2\text{O}}}. \text{ В случае малой капли радиальная составляющая напряженно-}$$

сти поля межмолекулярного взаимодействия максимальна в центре капли. С увеличением ее размера максимум радиальной составляющей напряженности достигается на межфазной границе, а силовые линии направлены к ее центру. Таким образом, межфазную границу можно определить не только по значению плотности, но и по напряженности поля сил межмолекулярного взаимодействия.

2.6. Распределение внутренней энергии

Эта энергия системы складывается из кинетической энергии, обусловленной тепловым движением молекул, и потенциальной энергии межмолекулярного взаимодействия. Тепловая энергия определяется как средняя кинетическая энергия движения молекул. Для вычисления радиального распределения энергии межмолекулярного взаимодействия $\Delta W(r_k)$ в частице разобьем ее на K концентрических слоев с радиусами r_k . Для числа молекул $\Delta N(r_k)$, которые заключены в k -м сферическом слое (радиуса r_k и толщиной Δr , которая взята на порядок меньшей радиуса частицы), рассчитывается сумма энергии взаимодействия со всеми другими молекулами:

$$\Delta W(r_k) = \frac{3}{2} \Delta N_k(r_k) k_B T(r_k) + \sum_{j=1}^{\Delta N_k(r_k)} \sum_i U_{ij}(\mathbf{r}_i - \mathbf{r}_j).$$

Особый интерес представляет значение средней потенциальной энергии межмолекулярного взаимодействия E_1 , приходящееся на одну молекулу капли в зависимости от ее положения в пространстве относительно межфазной границы

$$E_1(r_k) = \frac{\sum_{j=1}^{\Delta N_k(r_k)} \sum_i U_{ij}(\mathbf{r}_i - \mathbf{r}_j)}{\Delta N_k(r_k)}.$$

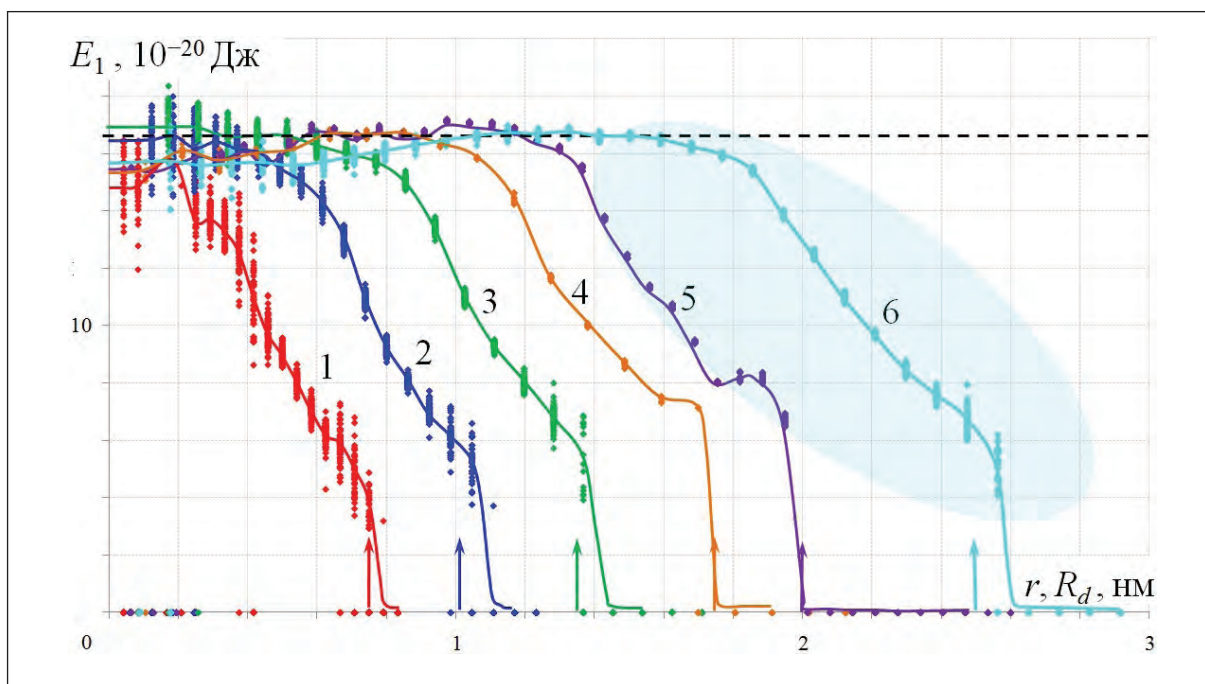


Рис. 3. Радиальное распределение средней энергии межмолекулярного взаимодействия в расчете на одну молекулу; стрелки указывают условные радиусы R_d частиц; кривые 1, 2, 3, 4, 5 и 6 соответствуют каплям, состоящим из 55, 160, 350, 650, 1100 и 2500 молекул соответственно.

Эта энергия максимальна внутри капли вдали от ее границы и уменьшается по мере удаления от центра капли (рис. 3). Для капель с размерами много большими характерной толщины межфазной границы капли значение E_1 – величина постоянная внутри капли вплоть до поверхностного слоя межфазной границы и не зависит от размера капли.

2.7. Распределение плотности энергии межмолекулярного взаимодействия

Для практических расчетов представляет интерес плотность энергии межмолекулярного взаимодействия, которая определяется как $\rho_w = \Delta W(r_k)/\Delta V(r_k)$. Здесь $\Delta V(r_k)$ – объем тонкого сферического слоя, который содержит $\Delta N(r_k)$ молекул, $\Delta W(r_k)$ – суммарная энергия взаимодействия $\Delta N(r_k)$ молекул друг с другом и с окружающими молекулами. Распределение плотности энергии межмолекулярного взаимодействия в каплях различных размеров показано на рис. 4.

2.8. Коэффициент поверхностного натяжения

Исследованию зависимости поверхностного натяжения от размера капли посвящено множество статей и монографий (например, [16–19]). С уменьшением характерных размеров объекта все большую роль начинают играть молекулярные и квантовые эффекты, и значения макропараметров начинают существенно отличаться от соответствующих значений в справочниках. В качестве оценки, «коэффициент поверхностного натяжения одной молекулы воды» $\sigma_l (N=1)$ может быть вычислен как отношение энергии распада (диссоциации) молекулы $E_{\text{Dis}} = 12.6$ эВ к площади поверхности сферы с характерным радиусом молекулы воды $\sqrt[6]{2}\sigma_{\text{H}_2\text{O}}$. $\sigma_l (N=1) = E_{\text{Dis}} / (4\pi\sigma_{\text{H}_2\text{O}}^2 \sqrt[3]{2}) \approx 1.9$ Н/м.

Следует отметить, что термин «поверхностное натяжение» связан не только с поверхностью капли. Силы межмолекулярного взаимодействия стремятся вернуть каплю

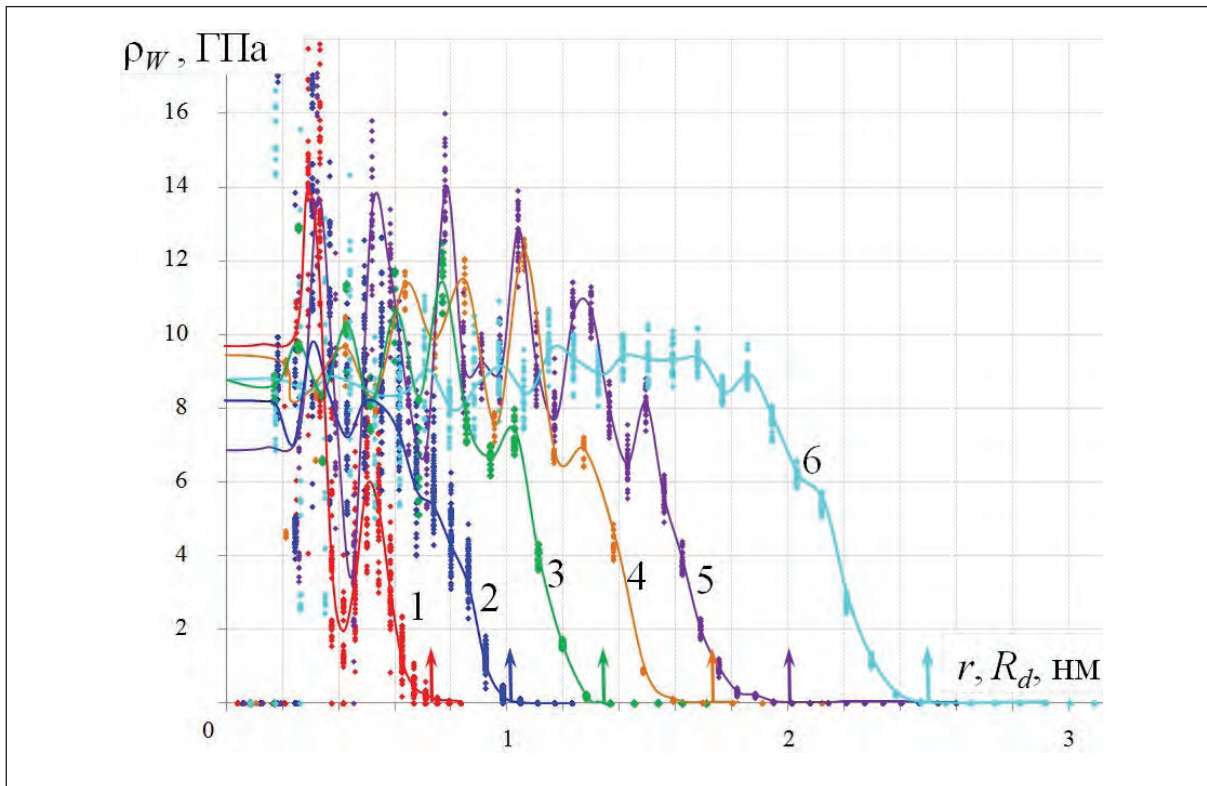


Рис. 4. Радиальное распределение плотности энергии межмолекулярного взаимодействия; стрелки и цифры показывают то же, что и на предыдущем рисунке.

к ее первоначальной форме. В [18] утверждается, что отказ от модели поверхностного натяжения (и замена способа оценки прочности жидких и твердых частиц терминами прочности и текучести материала) оказывается более приемлемым для современных моделей эволюции диспергируемых сред.

Коэффициент поверхностного натяжения нанокapли может быть найден как интегральная работа сил (действующих в радиальном к центру капли направлении) межмолекулярного взаимодействия в расчете на единицу площади:

$$\sigma_l(R_d, T) = \int_0^{\infty} \frac{1}{4\pi r^2} \left(-\frac{\partial W_{\sigma}(r, T)}{\partial r} \right) dr \cong \frac{\rho(R_d, T) \cdot R_d(T) \cdot (\sigma_l^{\infty}(T))^2}{\rho_l^{\infty} a_{vW} < r_m^2 > (\rho_l^{\infty} - \rho_{sv}(T))^2}.$$

Здесь индекс ∞ означает бесконечное число молекул $N \gg 1$. Таким образом, характер зависимости $\sigma_l(R_d)$ будет иметь вид, качественно напоминающий при малых размерах капли известные теоретические поправки [5, 16, 17].

На рис. 5 показано сравнение полученной зависимости (кривая 1) коэффициента поверхностного натяжения σ капли от ее радиуса R с другими известными физико-математическими моделями: 2 – $\sigma_l(r) = \sigma_l^{\infty} / (1 + 2\delta / R_d)$ – формула Толмена [16], 3 – результаты квантово-химических расчетов [5].

3. Взаимодействие частицы воды с поверхностью твердого тела

Для исследования поведения частицы в процессе ее столкновения с твердым телом (в зависимости от их свойств, а также параметров окружающей среды) учтем следующее. Амплитуды колебаний атомов твердого тела в аэрозольном потоке можно рассчи-

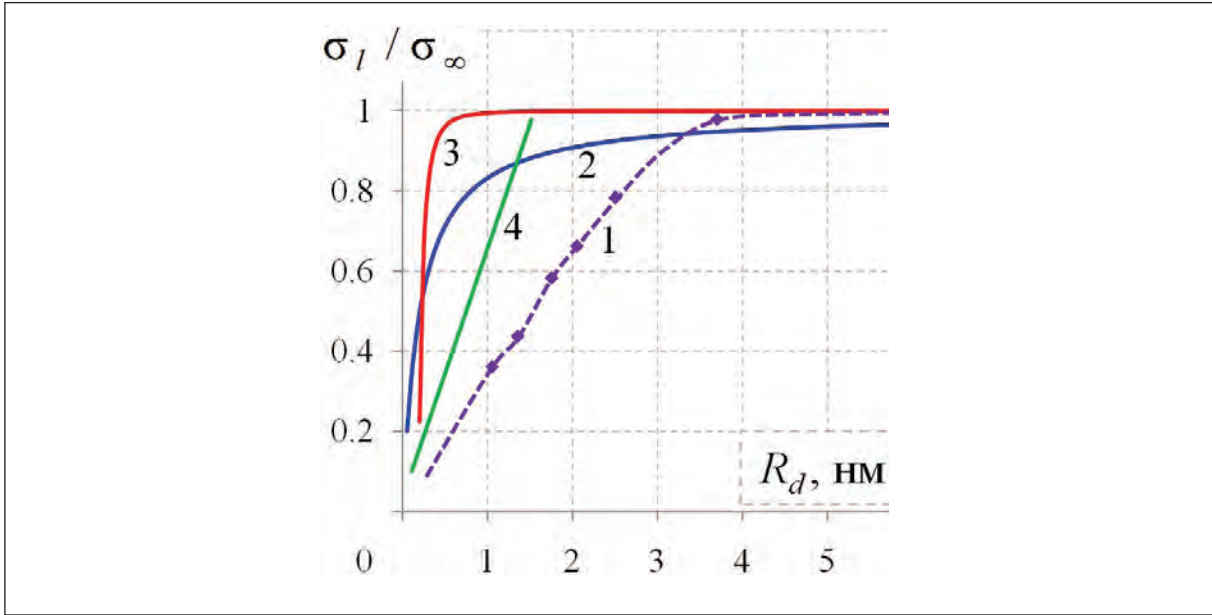


Рис. 5. Зависимость коэффициента поверхностного натяжения капли от ее размера

тать из энергетических соображений. Энергия фононов i -й акустической ветви, обладающей волновым вектором $\mathbf{k}$, равна

$$E_i(\mathbf{k}) = \hbar \omega_i(\mathbf{k}) \left[(\exp[\hbar \omega_i(\mathbf{k}) / k_B T] - 1)^{-1} + \frac{1}{2} \right],$$

где $\omega_i(\mathbf{k})$ – частота моды колебаний, $\hbar = h/2\pi$ – постоянная Планка. С другой стороны, можно представить эту энергию как сумму средних кинетической и потенциальной энергий колебаний атома массы m_A с амплитудой $\Delta a(\mathbf{k})$. Для гармонического осциллятора средняя кинетическая энергия равна потенциальной, так что

$$E_i(\mathbf{k}) = 2 \frac{m_A \omega_i^2(\mathbf{k}) \cdot \Delta a_i^2(\mathbf{k})}{2} = \hbar \omega_i(\mathbf{k}) \left[(\exp[\hbar \omega_i(\mathbf{k}) / k_B T] - 1)^{-1} + \frac{1}{2} \right].$$

Усредняя амплитуду колебаний атомов $\Delta a_i(\mathbf{k})$ в кристаллической решетке по зоне Бриллюэна и всем ветвям колебаний, получим:

$$\Delta a_i^2(T) = \frac{9\hbar^2}{4m_A k_B \theta_D} I(T/\theta_D).$$

Здесь $I(T/\theta_D) = 1 + 4 \left(\frac{T}{\theta_D} \right)^2 \int_0^{\theta_D/T} \frac{x dx}{e^x - 1}$, θ_D – температура Дебая материала твер-

дого тела.

Свойства кристаллической решетки и результаты численного расчета даны в таблице 1. Здесь n – число атомов, приходящихся на одну ячейку кристалла, a – постоянная кристаллической решетки.

Таблица 1. Свойства кристаллических решеток некоторых материалов при температуре -10°C

Материал	ρ , кг/м ³	μ , г/моль	n	a , Е	E , ГПа	θ_D , К	Δa , Е	ω , 10^{12}c^{-1}
Al	2700	27	4	4.05	68	394	0.17	16.7
H ₂ O (лед)	917	18	2	2.85	5	192	0.60	8.3

Для твердого тела с кубической элементарной ячейкой кристаллической решетки длина одной ячейки равна $a = \sqrt[3]{n\mu/\rho N_A}$; μ – молярная масса вещества. Для льда с гексагональной ячейкой имеем $a = \sqrt[3]{2n\mu_{\text{ice}}/3\sqrt{3}\rho_{\text{ice}} N_A}$.

Сравним два подхода к оценке характеристик поведения частиц твердого тела. С одной стороны, из квантовой теории следует, что

$$k_w \Delta a^2 / 2 = k_w (9\hbar^2 / 8k_B \theta_D) I(T / \theta_D) = m_A < c_w^2 > / 2 = 3k_B T / 2 = U,$$

где $< c_w^2 >$ – среднеквадратичная скорость атомов твердого тела.

С другой стороны, из теории упругости можно получить следующую оценку коэффициента «жесткости» межмолекулярного взаимодействия:

$$k_w = ES/a = E \gamma (a / \sqrt[3]{n})^2 / a = m_A J^2(\theta_D T) / I(T / \theta_D).$$

$$\text{Здесь } J(\theta_D T) = 2k_B \sqrt{\theta_D T} / 3\hbar.$$

Расчеты показали, что для рассмотренных материалов коэффициент $\gamma = \frac{k_w a}{E(a / \sqrt[3]{n})^2} = m_A \frac{J^2(\theta_D T)}{E(\mu / (n\rho N_A))^{1/3} I(T / \theta_D)}$, определяющий различие между описа-

нием колебаний атомов с помощью квантовой теории и теории упругости, приблизительно равен единице ($\gamma \approx 1$) в широком диапазоне температур, которые соответствуют условиям полета летательного аппарата. Из соотношений

$$< (\omega \Delta a)^2 > \approx < \omega^2 > < \Delta a^2 >, < c_w^2 > \approx k_B T / m_A$$

получим оценку для частоты колебаний атомов твердого тела

$$< \omega^2 > = J^2(\theta_D T) / I(T / \theta_D).$$

Аналогично уравнению Ван-дер-Ваальса, константы уравнения состояния твердого тела (Ми – Грюнайзена) можно связать с характеристиками взаимодействия элементов его структуры [20, 21]. В широком диапазоне температур (до 10^4 К) его внутреннюю энергию и давление можно представить в виде двух слагаемых

$$U(\rho_w, T) = U_G + U_T; \quad P(\rho_w, T) = P_G + P_T.$$

Здесь первые слагаемые (с индексом G) описывают упругие свойства «холодного» тела при гидростатическом сжатии, а вторые (с индексом T) – эффект гармонических колебаний атомов решетки, определяемых температурой.

Учтем следующие соотношения:

$$U_G(\rho_w) = \int_{\rho_w^0}^{\rho_w} \frac{P_G(\rho_w)}{\rho_w^2} d\rho_w; \quad P_G(\rho_w) = \rho_w^2 \frac{dU_G}{d\rho_w};$$

$$U_T(T) = c_w^0 T; \quad P_T(T) = \rho_w \Gamma(\rho_w) U_T;$$

Здесь нижний предел интеграла ρ_w^0 – плотность тела при фиксированных значениях параметров (например, при $P = T = 0$); $\Gamma(\rho_w)$ – коэффициент Грюнайзена для гармонических колебаний атомов, который можно представить в виде $\Gamma(\rho_w) = \Gamma^0 - \Gamma' \frac{\rho_w}{\rho_w^0}$.

Эта линейная аппроксимация справедлива для многих металлов в широком диапазоне плотностей. В первом приближении тепловая энергия решетки равна $U = \frac{3}{2} R_w T$, где

$R_w = \frac{3R_0}{\mu_w}$, μ_w – молярная масса вещества твердого тела, R_0 – универсальная газовая постоянная. (При низких температурах значение U может быть уточнено, например, с учетом теории Дебая).

Для описания гидростатических слагаемых часто используется потенциал Борна–Майера, так что

$$U_G(\rho_w) = \frac{3A}{b\rho_w^0} \exp[b(1-R)] - \frac{3K}{\rho_w^0 R^{m-1}(m-1)};$$

$$P_G(\rho_w) = \frac{A}{R^2} \exp[b(1-R)] - \frac{K}{R^m}, \quad \text{где } R = \left(\frac{\rho_w}{\rho_w^0} \right)^{-\frac{1}{3}}, \quad m > 0.$$

В качестве примера приведем в таблице 2 полуэмпирические значения констант для алюминия ($m = 4$).

В этих терминах тепловая энергия и частота колебаний атомов кристаллической решетки равны:

$$U \cong \frac{m_A \langle c_w^2 \rangle}{2} = \frac{m_A \langle (\omega \Delta a_w)^2 \rangle}{2} \sim \frac{m_A \langle \omega^2 \rangle \langle \Delta a_w^2 \rangle}{2},$$

$$\langle \omega^2 \rangle = \frac{2U}{m_A \langle \Delta a_w^2 \rangle} = \frac{8k_B \theta_D}{9\hbar^2} \frac{U(T)}{I(T/\theta_D)}.$$

Таблица 2. Параметры уравнения состояния твердого тела (алюминия)

ρ_w^0 , кг/м ³	A, ГПа	b	K, ГПа	Γ^0	Γ'	c_w^0 , Дж/(кг·К)
2700	24.43	10.332	26.18	2.77	-0.71	896

Интеграл $I(T/\theta_D)$ определен выше. Согласно молекулярно-кинетической теории,

$$\langle \omega^2 \rangle = \frac{3k_B T}{m_A \langle \Delta a_w^2 \rangle} = \frac{4k_B \theta_D}{9\hbar^2} \frac{3k_B T}{I(T/\theta_D)} = \frac{4k_B^2 \theta_D T}{3\hbar^2 I(T/\theta_D)}.$$

В настоящей работе метод молекулярной динамики использовался только для верхнего тонкого слоя обтекаемого твердого тела (рис. 6). Поскольку моделирование динамики всех атомов кристаллической решетки обтекаемого тела требует чрезвычайно больших вычислительных затрат, для расчета взаимодействия молекул аэрозоля с более «глубокими» атомами обтекаемого тела получены новые потенциалы взаимодействия. Предположив, что взаимодействие молекулы воды с атомом поверхности описывается потенциалом Леннарда-Джонса с параметрами $\varepsilon_w = Ak \cdot \varepsilon_{H_2O}$ и $\sigma_w \approx \sigma_{H_2O}$, разобьем твердое тело на множество плоских слоев толщины dy , параллельных поверхности твердого тела (рис. 6). Здесь число Ak – параметр подобия, который определяет отношение сил адгезии к силам когезии. Разобьем каждый из этих слоев на кольца радиуса r шириной dr . Тогда количество молекул в слое высотой dy и площадью $2\pi r dr$ будет равно $dN = 2\pi r dr dy \rho_w / m_{0w}$ (ρ_w – плотность материала обтекаемого тела), а потенциал взаимодействия между молекулой воды и тонким кольцом сечением $dr dy$ внутри твердого тела имеет следующий вид:

$$\begin{aligned} dU &= dN 4\varepsilon_w \left(\left(\frac{\sigma_w}{\sqrt{y^2 + r^2}} \right)^{12} - \left(\frac{\sigma_w}{\sqrt{y^2 + r^2}} \right)^6 \right) \cos \alpha = \\ &= 2\pi r dr dy \frac{\rho_w}{m_A} 4\varepsilon_w \left(\left(\frac{\sigma_w}{\sqrt{y^2 + r^2}} \right)^{12} - \left(\frac{\sigma_w}{\sqrt{y^2 + r^2}} \right)^6 \right) \frac{y}{\sqrt{y^2 + r^2}}; \\ U(h) &= 8\pi \frac{\rho_w}{m_A} \varepsilon_w \int_{y=-\infty}^h y dy \int_{r=0}^{\infty} \left(\left(\frac{\sigma_w}{\sqrt{y^2 + r^2}} \right)^{12} - \left(\left(\frac{\sigma_w}{\sqrt{y^2 + r^2}} \right)^6 \right) \right) \frac{r dr}{\sqrt{y^2 + r^2}} = \\ &= \frac{8}{15} \frac{\rho_w}{m_A} \pi \varepsilon_w \sigma_w^3 \left(\frac{5}{33} \left(\frac{\sigma_w}{h} \right)^9 - \left(\frac{\sigma_w}{h} \right)^3 \right). \end{aligned}$$

Этот новый потенциал использован в настоящей работе при вычислениях взаимодействия частиц воды с поверхностью твердого тела.

На рис. 7 первая кривая представляет собой обычный потенциал Леннарда-Джонса парного взаимодействия между молекулами воды $U(r)$. Вторая – потенциал взаимодействия молекулы воды с полубесконечным водоемом в зависимости от ее высоты h до его поверхности. Третья и четвертая кривые соответствуют энергии взаимодействия молекулы воды с гидрофильной и гидрофобной поверхностями соответственно. Последняя кривая 5 соответствует гипотетической супергидрофобной поверхности, атомы которой отталкивают молекулы воды на любом расстоянии. В этом случае сила отталкивания обратно пропорциональна третьей степени расстояния молекулы воды от поверхности твердого тела.

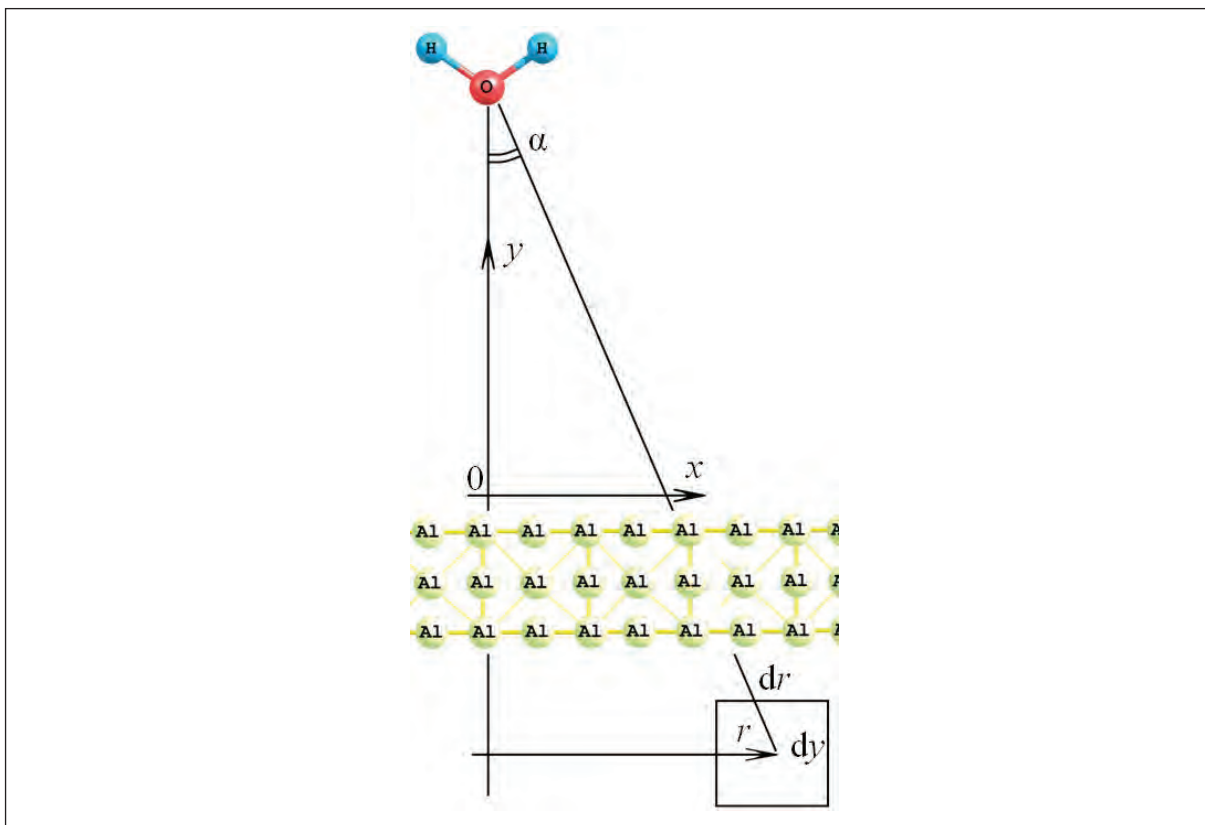


Рис. 6. Схема интегрирования энергии взаимодействия молекулы воды с атомами обтекаемого тела (алюминия).

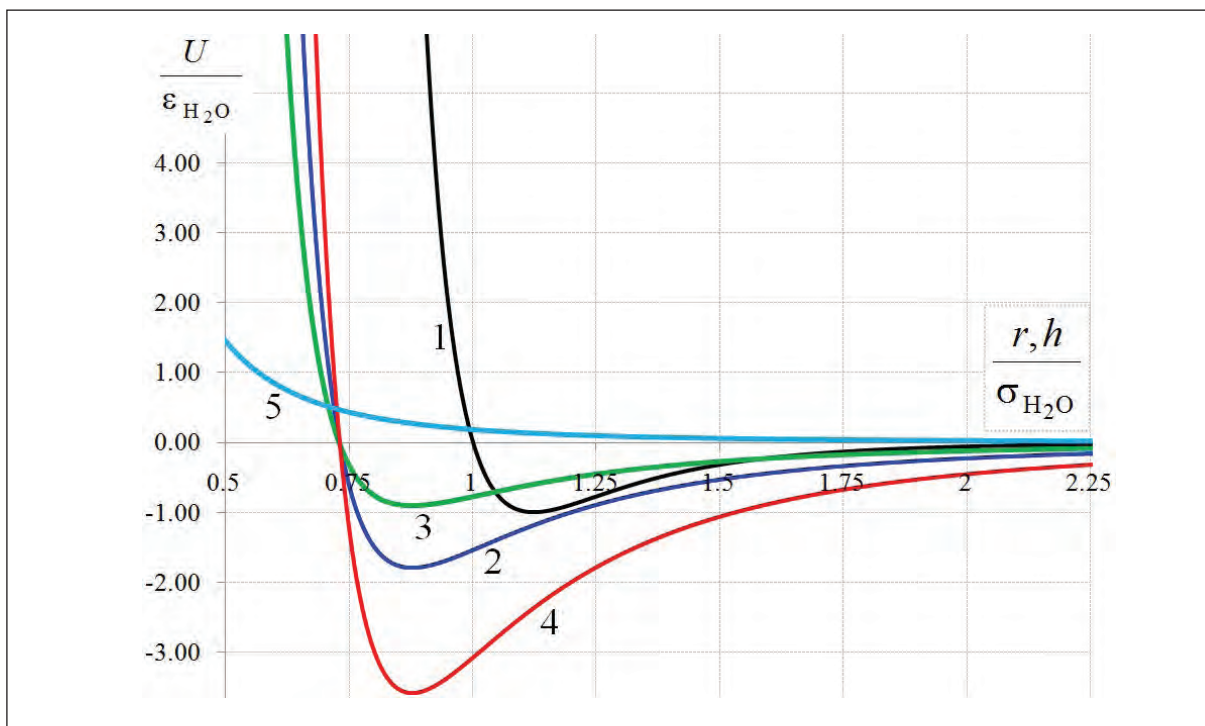


Рис. 7. Сравнение потенциалов межмолекулярного взаимодействия: 1 – потенциал Леннарда-Джонса парного взаимодействия молекул воды; 2 – потенциал взаимодействия молекулы воды с жидким телом из воды в зависимости от расстояния до его поверхности при $Ak = 1$; 3 – потенциал взаимодействия молекулы воды с гидрофобным телом в зависимости от расстояния до его поверхности при $Ak = 0.5$; 4 – то же, но с гидрофильным телом при $Ak = 2$; 5 – то же с супергидрофобной поверхностью при $Ak < 0$.

4. Область гидрофобности в пространстве физико-механических параметров твердого тела и условий обтекания

При исследовании поведения частицы в процессе столкновения определялись различные физические величины: область соприкосновения, углы смачивания, распределение температуры, давления, плотности энергии и др. Рис. 8 представляет собой примеры взаимодействия частиц с гидрофильной (вверху) и гидрофобной поверхностями: можно видеть, что при одной и той же температуре наночастица прилипает к гидрофильной поверхности и отскакивает от гидрофобной. На этом рисунке представлены формы и положения наночастиц в последующие моменты времени. Несущая газовая смесь (молекулы воздуха и водяного пара) показаны двойными (N_2 и O_2) и тройными (H_2O) кружками. Видно, в частности, что наночастицы мало похожи на привычную шаровую каплю воды, ограниченную сферической поверхностью, которой приписывается коэффициент поверхностного натяжения σ : она находится в состоянии постоянного обмена молекулами с несущим газом, так что число молекул N , входящих в состав наноклапты, хаотически изменяется вокруг некоторого среднего значения.

Проведенные исследования позволяют найти область (в многопараметрическом пространстве параметров аэрозольного потока – скорость, температура и др., а также физических свойств материала поверхности обтекаемого тела), в которой частицы воды не будут примерзать к элементам конструкции летательного аппарата. На рис. 9 показан пример расчета такой области (сечение многомерной области плоскостью V , A_k) для воды радиуса 1.3 нм ($N \approx 350$ молекул) не будут примерзать к обтекаемому телу.

Результаты этой работы могут быть использованы также для оценки коэффициентов восстановления компонент скорости микрочастиц (окислов металлов, кристаллов льда), сталкивающихся с твердым телом [22, 23].

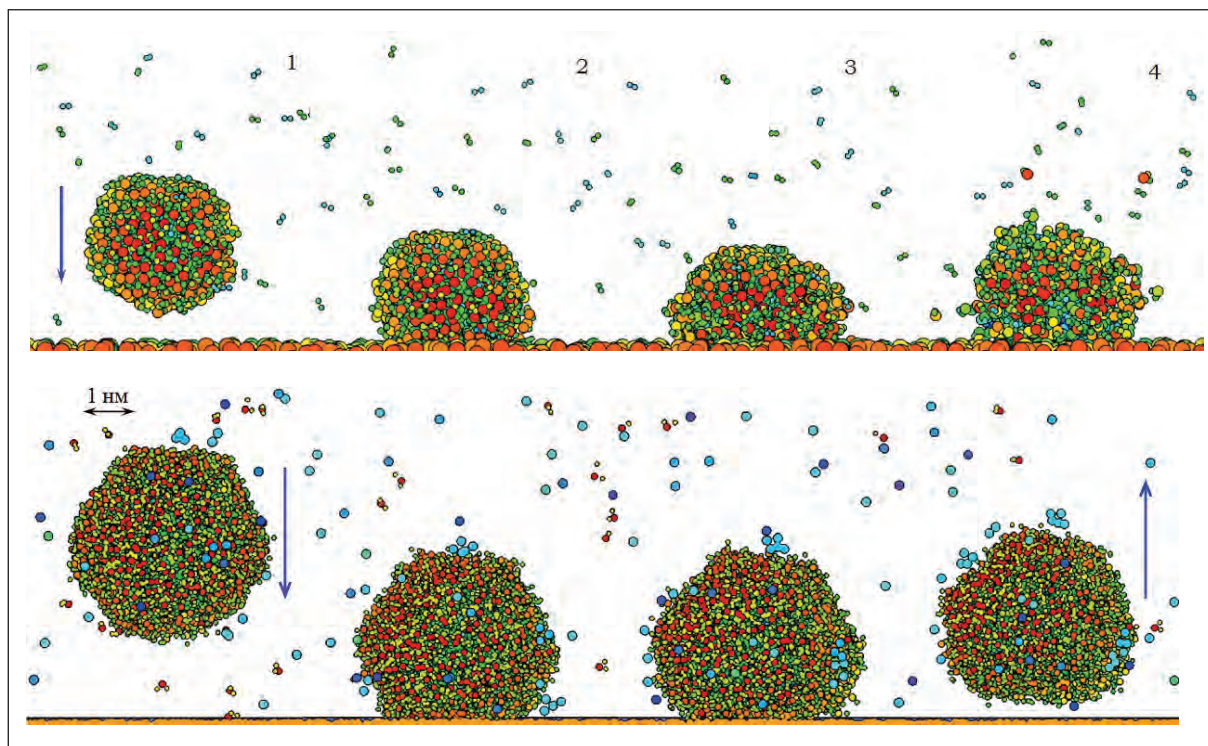


Рис. 8. Столкновение капли воды: а) с гидрофильной поверхности; б) с гидрофобной поверхностью; температура капли, окружающего газа и обтекаемого тела $-10\text{ }^{\circ}\text{C}$. Нормальная скорость соударения 100 м/с. Несущая газовая смесь (воздух и водяной пар) неподвижна (дно пограничного слоя)

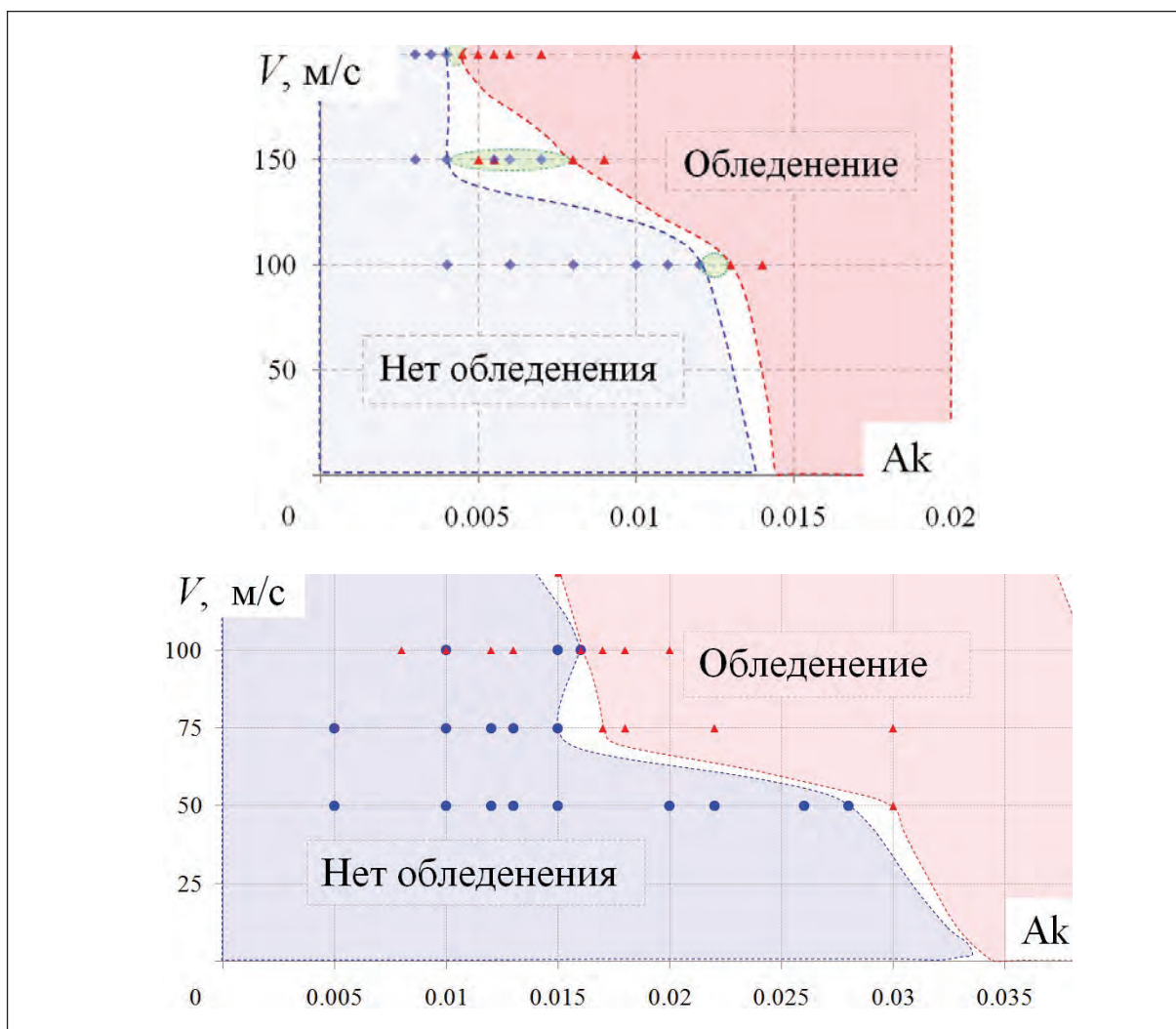


Рис. 9. Оценка области отсутствия обледенения в пространстве скорости наноаэрозольного потока V и степени гидрофобности Ak поверхности обтекаемого тела при температуре $t = -10^\circ\text{C}$. Белый промежуток соответствует неустойчивой гидрофобности.

Заключение

Развита физико-математическая модель и соответствующий численный код для исследования макро-характеристик наночастицы в терминах законов межмолекулярного взаимодействия. Результаты работы позволяют оценить область в многопараметрическом пространстве физико-механических свойств обтекаемого тела и условий полета, в которой нанокapли не будут примерзать к поверхности летательного аппарата.

Литература

1. Böttger R. Investigation on the possibility to use disc and cone for aerodynamic drag measurement in a process of ice accretion // IZWB FB, Berlin, 1935, 462, 12 pp.(in German).
2. Cebeci T., Kafyeke F. Aircraft icing // Ann. Rev. Fluid Mech., 2003, 35, 11–21.
3. Стасенко А.Л., Толстых А.И., Широбоков Д.А. К моделированию оледенения самолёта: Динамика капель и поверхность смачивания // Математическое моделирование, 2001, 13, 6, 81–86.
4. Кашевараев А.В., Стасенко А.Л. Вынужденная кристаллизация капель перед телом, движущимся в переохлажденном облаке // Математическое моделирование, 2010, 22, 2, 139–147.
5. Артюхин А.С., Егоров Б.В., Забабурин Е.А., Кощеев А.В., Маркачев Ю.Е., Плеханов Е.А., Уфимцев И.С., Хлопков А.Ю., Чугреев А.Л. Кинетика формирования ультралегкой фракции нейтральных и заряженных кластеров в газодинамических потоках летательного аппарата // ХФ, 2004, 24, 28–46.
6. Zeng Yi, Fan Li-Wu, Xiao Yu-Qi, Yu Zi-Tao, Cen Ke-Fa. Experimental investigation of melting of nanoparticle-enhanced phase change materials (NePCMs) in a bottom-heated vertical cylindrical cavity // International Journal of Heat and Mass Transfer, 2013, 66, 111–117.
7. Heng Xie, Seiichi Koshizuka, Yoshiaki Oka. Modelling of a single drop impact onto liquid film using particle method // Int. J. Numer. Math. Fluids, 2004, 45, 1009–1023.
8. Tao Jiang, Jie Ouang, Binxin Yang, Jinlian Ren. The SPH method for simulating a viscoelastic drop impact and spreading on an inclined plate // Comput. Mech., 2010, 45, 573–583.
9. Кривцов А.М. Деформирование и разрушение твердых тел с микроструктурой. М.: Физматлит, 2007, 340 стр.
10. Matsumoto M. Nanoscale bubbles and droplets. – Investigation with molecular simulations. Disc from the 8th Pacific Symposium on Flow Visualization and Image Processing (PSFVIP8). М.: Lomonosov Moscow State University, 2011, ISBN 978-5-8279-0093-1.
11. Allen M., Tildesley D. Computer Simulation of Liquids / Clarendon Press, London, 1987, 385 pp.
12. Hockney R.W., Eastwood J.W. Computer Simulation Using Particles, 1981, McGraw-Hill, New York, 540 pp.
13. Evans D.J., Murad S. Singularity-free algorithm for molecular dynamics simulation of rigid polyatomics // Mol. Phys., 1977, 34, 327 – 331.
14. Невзоров А. Н. О внутреннем механизме кристаллизации метастабильной жидкой воды и об его эффектах, влияющих на внутриоблачные процессы // Известия РАН, Физика атмосферы и океана, 2006, 42, (6), 830–838.
15. Радченко И.В. Молекулярная физика. М.: Наука, 1965, 480 стр.
16. Tolman, R.C. Effect of droplet size on surfaces tension // J. Chem. Phys., 1949, 17, 333–338.
17. Русанов А.И. Фазовые равновесия и поверхностные явления. Л.: Химия, 1967, 388 стр.
18. Иголкин С.И. Натяжение пленки и давление Лапласа под искривленной поверхностью жидкости // Прикл. физика, 2007, 5, 21–28.
19. Хайдаров Г.Г., Хайдаров А.Г., Мацек А.Ч. Физическая природа поверхностного натяжения жидкости // Вестник СПбГУ, сер. 4 (Физика, Химия), 1, 2001, 3–8.
20. Mie G. Zur kinetischen Theorie der einatomigen Körper // Ann. Phys., 1903, 11, 657–672.
21. Grüneisen E. Theorie des festen Zustandes einatomiger Elemente // Ann. Phys., 1912, 39, 257–306.
22. Кудин О.К., Нестеров Ю.Н., Токарев О.Д., Флакман Я.Ш. Экспериментальное исследование натекания высокотемпературной струи запыленного газа на преграду // Ученые записки ЦАГИ, 2013, XLIV (6), 105–115.
23. Кашевараев А.В., Стасенко А.Л. взаимодействие частиц различной формы с несущим континуальным потоком // Ученые записки ЦАГИ, 2014, XLV(5), 3–17.

INTERACTION OF AEROSOL FLOW NANODROPLETS WITH A SOLID BODY

I.A. Amelyushkin¹, A.L. Stasenko²

¹ *Central Aerohydrodynamic Institute;*

² *Moscow Institute of Physics and Technology*

Amelyushkin_Ivan@mail.ru, Stasenko@serpantin.ru

Received 16.05.2016

Physico-mathematical model and the corresponding numerical code are derived to investigate a water droplet interaction with a solid body and to predict its icing start-up in the humid aerosol flow. Dependences of macro-characteristics of the small droplets on their dimensions are calculated using the molecular dynamics simulation method, and their comparisons with other theoretical data are presented. Numerical estimations of the domain in the space of the aerosol flow parameters and physicochemical properties of the flown-around solid body's material, which exclude the body icing are presented.

ПРОБЛЕМЫ РЕАЛИЗАЦИИ МАССОВОГО ДИНАМИЧЕСКОГО ПАРАЛЛЕЛИЗМА. II

А.В. Махиборода, А.В. Ильичёв, А.А. Подобин, А.В. Царёв

*Департамент прикладной математики МИЭМ,
Национальный исследовательский университет «Высшая школа экономики»*

makhiboroda@yandex.ru

Поступила 17.05.2016

В первой части статьи было отмечено, что освоение массового динамического параллелизма становится основным инструментом обеспечения устойчивого роста производительности вычислительных средств. Показано также, что на пути реализации массового параллелизма имеются два основных препятствия. Это специфические затраты на поддержку параллелизма, которые растут опережающими темпами с ростом числа процессорных элементов, и низкая производительность обменной среды, поддерживающей межпроцессорные пересылки данных.

Во второй части проводится оценка прогнозов развития обменной среды в части совершенствования её алгоритмов и повышения пропускной способности. Делаются экспертные заключения о возможностях расширения диапазона эффективного параллелизма до значений в пределах сотен процессоров. При этом показатель в несколько тысяч процессоров на кристалле уже достигнут, а данные о текущем состоянии технологии позволяют прогнозировать рост числа процессоров до десятков тысяч и более на ближайшие несколько лет. Поиск путей дальнейшего расширения диапазона эффективной реализации параллелизма является актуальным и направлен на освоение принципиально новых форм организации вычислений. Излагаются основные идеи архитектуры самоопределяемых данных, и даётся краткий очерк построения математического аппарата дискретной динамики. Понятийный аппарат дискретной динамики позволяет сформулировать и решить задачу программирования процессов в архитектуре самоопределяемых данных.

УДК 681.3.06

Введение

В предыдущем разделе мы рассмотрели конкретный пример, который позволяет определить некую общую тенденцию и дать качественную оценку динамики роста производительности параллельных систем при наращивании параллелизма в широком

диапазоне значений. При организации работы параллельных структур кроме основного объёма вычислительных работ, распределённых на множестве обрабатывающих процессоров, неизбежно возникают дополнительные издержки в виде затрат на обслуживание параллелизма. Издержки обслуживания параллелизма в основном представлены пересылками данных между процессорными элементами и процедурами доступа к общей памяти и всегда растут с ростом параллелизма. Следовательно, динамика роста производительности параллельных систем формируется как результат наложения двух взаимоисключающих тенденций – сокращения времени выполнения задачи вследствие параллелизма и роста издержек на обслуживание параллелизма. Длительный монотонный рост производительности при наращивании числа каналов обработки в данной ситуации невозможен. В общем случае динамика изменения производительности параллельной системы будет проходить стадию роста, затем входить в насыщение и далее падать. Реальные задачи проектирования и эксплуатации параллельных систем заключаются в том, чтобы вывести точку насыщения графика роста производительности в область более высоких значений и более полного использования параллелизма. К поставленной цели можно двигаться по двум направлениям – увеличивать производительность обменной среды и снижать объёмы обменных операций. Увеличение продуктивности межпроцессорных обменов задача инженерная, это выбор физической среды, повышение полосы пропускания, совершенствование алгоритмов и протоколов обмена. Сокращение объёмов обменных операций это более сложная задача, которая носит системный характер и требует анализа и корректировки архитектурных принципов и форм организации вычислительных процессов.

Цель настоящей статьи состоит в том, чтобы обозначить границы роста эффективности параллелизма вследствие совершенствования обменной среды, а также обосновать необходимость и наметить перспективы разработки принципиально новой архитектурной идеи, поддерживающей массовый динамический параллелизм.

1. Основные характеристики и направления развития коммутационных сетей

Начиная с определённого момента в процессе роста числа транзисторов на кристалле, разработчики чипов стали осваивать новое направление развития – создание многоядерных изделий, в которых число процессорных элементов исчисляется сотнями и тысячами. На период с 2020 года и далее прогнозируется достижение показателя степени интеграции в триллион транзисторов на кристалле, что позволит довести число ядер до миллиона [1]. В проектировании и производстве процессорных элементов накоплен огромный опыт, обеспечивающий большое разнообразие типов архитектур, ориентированных на разные виды применений. В тоже время средства обеспечения межпроцессорных обменов не получили достаточного развития и на данный момент являются самым консервативным и проблемным сегментом внутренней структуры кристалла. Очевидно, что при организации согласованной работы больших ансамблей процессорных элементов ведущая роль отводится средствам поддержки межпроцессорных обменов и коммутации. Наиболее полно тематика построения высокоскоростных коммутационных сетей межпроцессорных обменов представлена в области производства и эксплуатации суперкомпьютеров.

Рассмотрим основные характеристики и направления развития средств коммутации и межпроцессорных обменов, которые в дальнейшем будем называть сети интерконнекта. Для нас важно оценить две принципиально отличающиеся технологии построения обменной среды – с коммутацией каналов и коммутацией пакетов. Традиционной считается технология коммутации каналов. При коммутации каналов имеется некий пул процессорных элементов и над ним строится обменная среда, которая должна обеспечить возможность соединения каждого элемента с каждым для осуществления обмена данными. При значительных объёмах пула обеспечить полнодоступное соединение всех участников обмена со всеми технически невозможно. Поэтому строятся разнообразные системы переключателей с частичным доступом, в которых пара обменивающихся абонентов монополизует коммутационные ресурсы на время сеанса обмена данными. При этом часть путей обмена на время сеанса становится недоступной для других участников пула, а часть может одновременно обслуживать другие запросы на обмен. Технология коммутации каналов предполагает, что в составе системы имеются средства управления коммутацией и есть протокол, определяющий алгоритмику обменов. Участники пула должны посылать в систему управления запросы на обмен, система управления должна прокладывать путь обмена и после завершения сеанса восстанавливать исходное состояние коммутационной среды. Технические характеристики коммутационных систем определяют объём обслуживаемого пула абонентов, полосу пропускания и скорости пересылки данных, временные затраты на выполнение протокольных событий доступа к сети, вероятность отказа в доступе и среднее время ожидания в очереди на обслуживание. Важное обстоятельство заключается в том, что динамика функционирования обменной среды носит стохастический характер. Это может создавать определённые проблемы при построении параллельных вычислительных систем. В ряде приложений необходимо учитывать, что синхронные алгоритмы погружаются в аппаратную среду со стохастической динамикой, и это может привести к нежелательным последствиям.

Сети с технологией коммутации пакетов состоят из множества узлов, каждый из которых имеет фиксированное число постоянных некоммутируемых связей со смежными узлами. Данные оформляются как сетевые пакеты, которые могут продвигаться по сети от узла к узлу различными маршрутами. В узлах сети располагается оборудование, управляющее продвижением пакетов. В узлах сети могут располагаться, в том числе и процессорные элементы. Отличительный признак сетей с коммутацией пакетов состоит в том, что обменная среда может совмещаться со средой обработки. Такие сети легко масштабируются и хорошо приспособлены для построения больших многопроцессорных систем. Топология пакетной сети обеспечивает связность всех элементов процессорного пула, поскольку для каждой пары обменивающихся процессорных элементов всегда есть, связывающий их путь.

Важнейшей характеристикой коммутационной сети является топология её построения. Имеется огромный опыт разработки разнообразных топологий коммутационных сетей и обширная литература по этой теме, достаточно ознакомиться, например, с монографией [2]. Мы прокомментируем только некоторые из них, которые наиболее популярны в настоящее время у разработчиков суперкомпьютеров и мультипроцессорных систем. Сравнительные характеристики и анализ различных систем коммутации, применяемых в суперкомпьютерах можно найти в [3, 4]. На рис. 1 приведена структура коммутационной сети, известная под названием «толстое дерево».

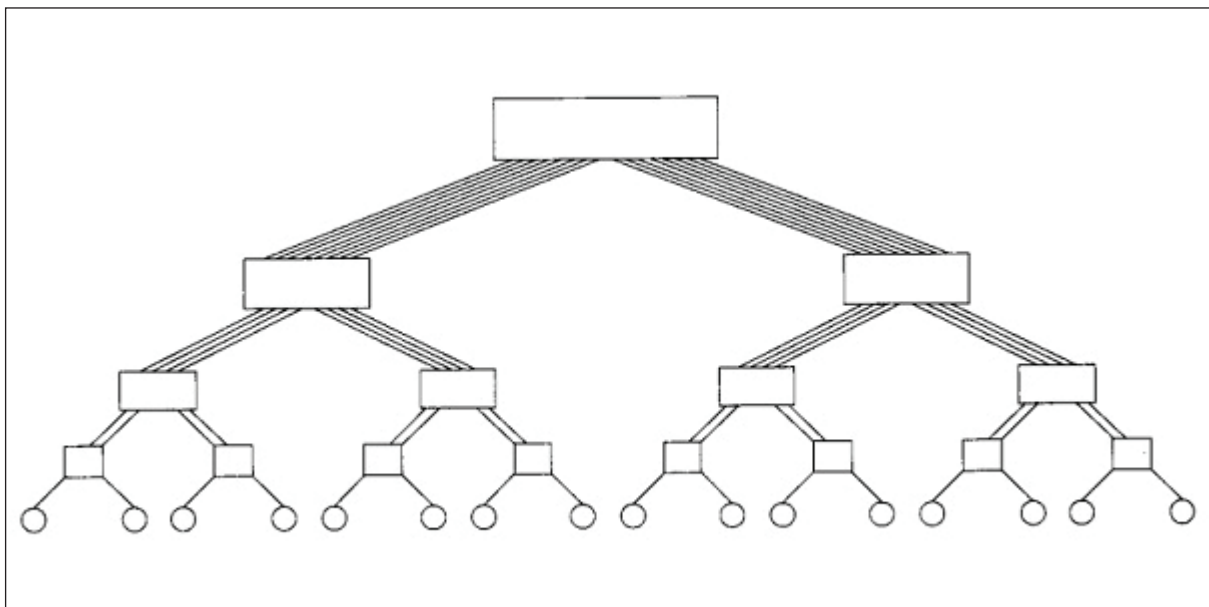


Рис. 1. Схема объединения узлов коммутаторами в топологии «толстое дерево».

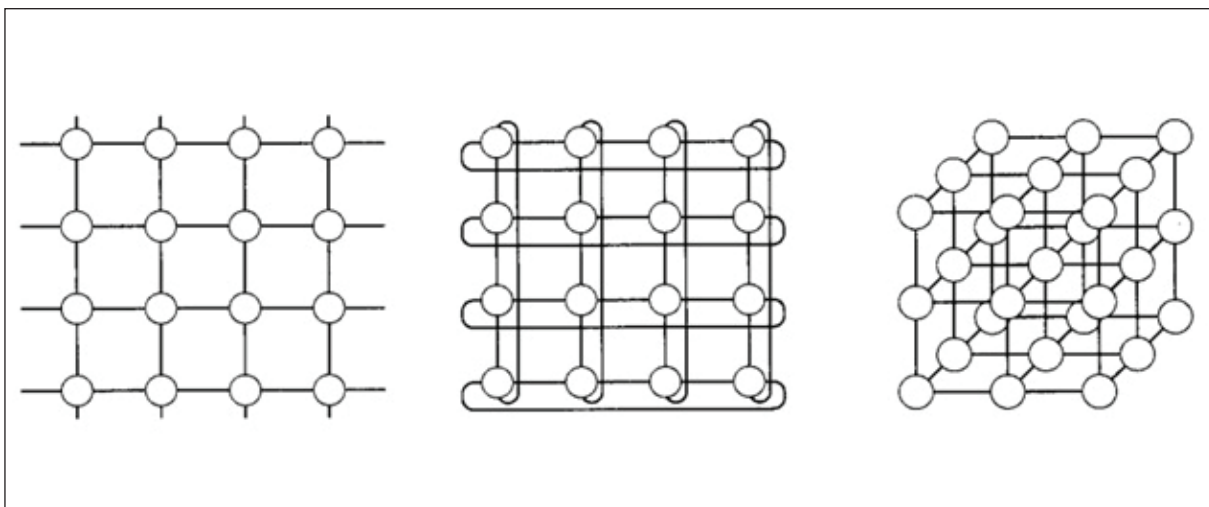


Рис. 2. Схема объединения узлов непосредственно друг с другом в топологии решётки и тора.

Это типичный пример построения сети по технологии коммутации каналов. Сеть «толстое дерево» имеет ограничения на рост объёма процессорного пула. «Толстое дерево» эффективно работает при объединении сотен процессорных элементов, при переходе тысячного рубежа издержки начинают расти взрывообразно. Более перспективной для построения мультимедийных систем оказываются топологии решётки, тора и многомерных торов, изображённые на рис. 2.

Активно разрабатываются гибридные структуры, сочетающие разные топологии в иерархических построениях, это структуры «стрекоза» и «бабочка» [5]. Проектирование коммутационных сетей решает сложные и противоречивые задачи – рост объёмов процессорного пула вызывает рост размерности и сложности сети, а это приводит к увели-

чению числа переключателей, удлинению маршрутов движения данных и к усложнению алгоритмов маршрутизации. Складывается парадоксальная ситуация, при которой сталкиваются две взаимоисключающие тенденции. С ростом размерности сети растёт число абонентов и путей передачи данных, что означает увеличение пропускной способности сети в целом. В тоже время увеличивается сложность маршрутов для каждого конкретного сеанса обмена. Рост числа переключателей и буферных зон на пути движения данных и усложнение алгоритмов маршрутизации существенно понижают скорость обменов, что не способствует росту суммарной продуктивности сети. Ситуацию пытаются переломить применением дорогостоящих высокочастотных элементов и оптических кабелей. При этом приходится искать компромиссы, связанные с ростом энергопотребления и высокой стоимостью аппаратуры [6, 7, 8].

По имеющимся на данный момент сведениям во внутренней среде кристалла для построения многоядерных структур используется каскадная многоуровневая коммутация на базе общей шины. В перспективе принятие решений по структуре внутренней обменной среды будет обусловлено выбором архитектуры кристалла и новыми возможностями технологии многослойной 3D компоновки. Можно предполагать, что предпочтительными окажутся топологии решётки и тора. Для понимания перспектив становления многоядерных структур на кристалле необходимо сделать оценки эффективности использования параллелизма в широком диапазоне значений числа процессорных элементов. С этой целью был сделан пробный расчёт ожидаемых значений коэффициента ускорения от параллелизма для конкретного алгоритма быстрого преобразования Фурье (БПФ) с проживанием по времени на $2^{16} = 65536$ точек. Разметка шкалы N (числа процессорных элементов) была сделана по некоторым правдоподобным предположениям наращивания N по степеням двойки с определённым вариантом многоуровневой компоновки на иерархии общих шин. Так, например, для первой точки шкалы компоновка представляет кластер из 16 процессоров на общей шине, для второй точки компоновка достигается объединением 16 кластеров на шине второго уровня и т. д. Исходные данные расчёта приведены в таблице 1.

Таблица 1

Количество ядер	Время выполнения	Коэффициент ускорения
1	10485760	1
16	917824	11,42
16x16 (256)	446032	23,51
16x32 (512)	363088	28,88
32x32 (1024)	672640	15,59
16x16x8 (2048)	672770	15,59
16x16x16 (4096)	601024	17,45
16x16x32 (8192)	600480	17,46
16x32x32 (16384)	605750	17,2
32x32x32 (32768)	702640	14,92

Время выполнения алгоритма рассчитывалось так же как и в первой части настоящей статьи, где суммировалось время выполнения счета, время выполнения процедур доступа к сети и время передачи данных. Для каждой точки расчет был индивидуальным, поскольку затраты времени на обмен данными определялись структурой шинной компоновки (см. первый столбец таблицы). Относительно характеристик шины были сделаны идеальные предположения о том, что шина всегда доступна и задержек доступа не бывает, шина широкополосная и ведёт обмен машинными словами, тактовая частота шины равна тактовой частоте процессора, а затраты на протокол доступа минимальны и равны одному такту. Полученный график роста коэффициента ускорения приведен на рис. 3.

Пик производительности достигается при числе процессоров равном 512 и позволяет получить ускорение в 28.88 раз. Далее начинается неуклонное падение производительности. Провал графика при значениях N 1024 и 2048 объясняется структурой компоновки и условиями балансирования коротких и длинных передач. Но по мере продвижения в область больших значений N доминантой становится время пересылки данных, а все сопутствующие компоненты и эффекты становятся относительно малыми. Возрастающий параллелизм производит возрастающие объёмы межпроцессорных обменов. При этом самые высокоскоростные и широкополосные сети интерконнекта не в состоянии изменить ход графика ускорения существенным образом. Напоминаем, что приведенный расчет представляет частный случай оценки условий распараллеливания конкретного алгоритма, но он позволяет выявить общую тенденцию. А суть этой общей тенденции такова, что системы с массовым параллелизмом на уровне десятков и сотен

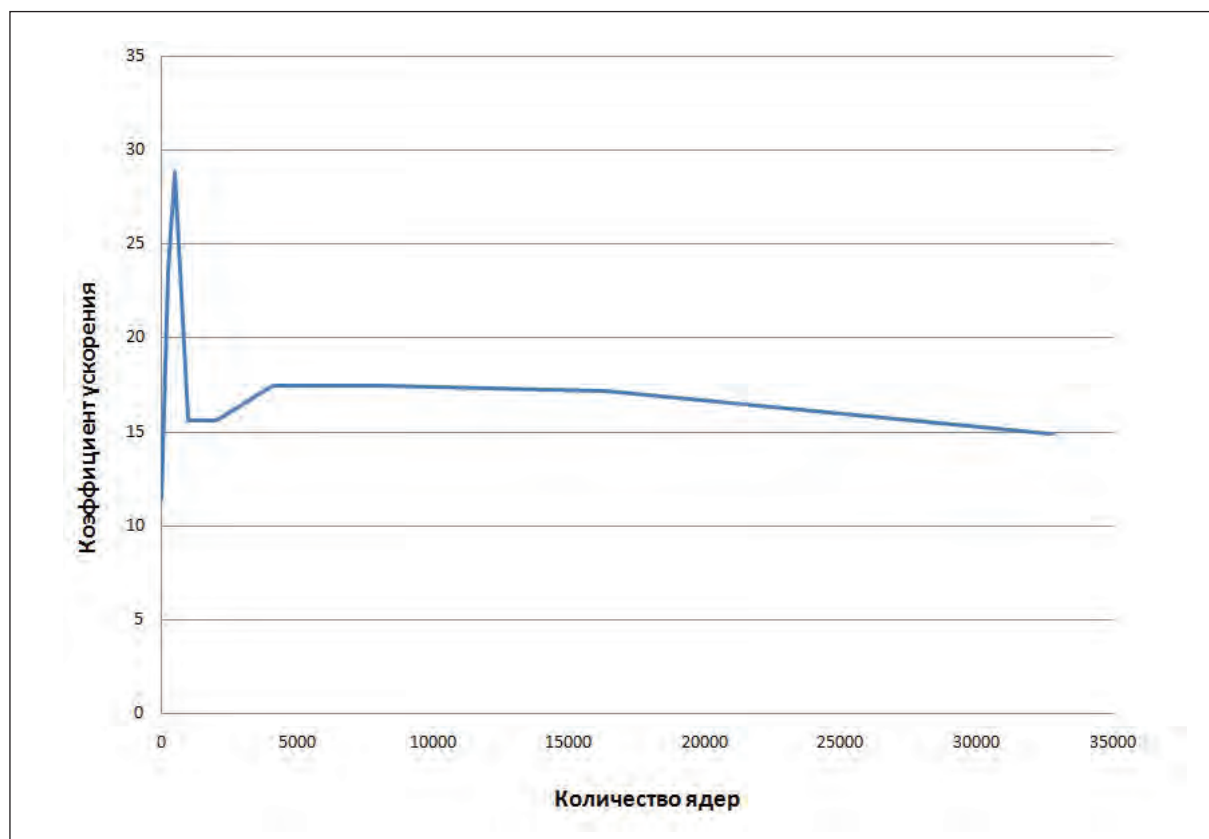


Рис. 3. График роста коэффициента ускорения для выполнения алгоритма БПФ

тысяч процессорных ядер будут очень чувствительны к типам решаемых задач и не смогут исполнить роль универсальных вычислительных средств широкого назначения.

2. Оценки эффективности организации вычислений

Расположение точки насыщения роста производительности определяется соотношением временных затрат на полезную работу и на издержки разных видов. Естественным образом возникает вопрос – почему в рассмотренных нами в первой части статьи примерах динамика роста производительности параллельных систем ограничивается столь низкими значениями коэффициентов ускорения и степени параллелизма. Это 3-х кратное ускорение при 6 процессорных элементах в примере с моделированием редукционно-поточковой архитектуры и 6-х кратное ускорение при 12 узлах обработки во втором примере [9].

Если определяющим фактором являются соотношения временных затрат на полезную работу и на издержки необходимо провести оценку существующих форм организации вычислений с целью определения источников издержек и возможности их радикального сокращения. Необходимо учитывать, что при распараллеливании процессов издержки, свойственные одному процессорному элементу размножаются кратно показателю параллелизма. Кроме того, порождаются дополнительные издержки обслуживания параллелизма, которые прибавляются к общему грузу издержек и таким образом суммарные издержки растут опережающим темпом. В современных условиях, когда произошло введение критерия эффективности кристалла по показателю числа операций на единицу затрат энергии учет издержек и определение их соотношения к полезной работе кристалла становится критическим показателем, от которого зависит возможность реализации массового параллелизма.

Рассмотрим основные положения классической архитектуры с точки зрения оценки издержек и эффективности организации процессов. Напомним, что все промышленно выпускаемые в настоящее время вычислительные средства построены по классической архитектуре. Неклассические архитектуры существуют как проекты или модели и не переходят в стадию промышленной реализации. Сущность архитектурной идеи наиболее полно воплощается в знаковой системе, средствами которой осуществляется программирование машины и управление внутренними процессами.

В архитектуре классической машины фон Неймана управление вычислительными процессами осуществляет простейшая знаковая система, называемая линейным императивным языком. Элементами императивного языка являются команды, которые в совокупности образуют функционально полный набор базисных функций. Функциональная полнота базисного набора команд позволяет декларировать тезис алгоритмической универсальности машины - средствами системы команд можно запрограммировать любой алгоритм. Программа представляет собой линейную последовательность команд.

Основой аппаратной реализации знаковой системы является машинное слово – это битовая строка определённой длины. Длина или разрядность машинного слова это одна из основных характеристик машины. В языке машины существует две версии интерпретации машинных слов – слова команды и слова операнды. Команда реализована на аппаратном уровне как битовая строка, разбитая на несколько битовых секций или полей, имеющих своё содержательное назначение. Далее для экспертной оценки эффективности организации процессов мы будем использовать некую гипотетическую ма-

шину с фиксированным трёхадресным форматом команды и с памятью, в которой осуществляется адресный доступ к машинным словам. Это не соответствует реальной организации современных машин. Так, например доступ в память осуществляется по байтам, что необходимо для поддержки переменных форматов команд и данных. Команды и данные размещаются в разных блоках памяти и читаются независимо по разным каналам доступа. И таких отличий от предлагаемой к рассмотрению гипотетической машины огромное число, но все эти отличия не меняют суть архитектурной идеи, а их учёт не только усложнит, но и сделает невозможной экспертную оценку эффективности организации вычислений.

Первое битовое поле кодирует операцию и называется КОП. КОП интерпретируется как число в двоичной системе и далее дешифрируется аппаратурой по таблице соответствия. Следующие три поля это так называемые адресные поля, которые обозначаются как A_1 , A_2 и A_3 . Это числа в двоичной системе счисления, по которым осуществляется доступ в память. Память состоит из массива регистров, имеющих индивидуальные номера или адреса. В регистрах памяти размещаются машинные слова. По предъявлению адреса осуществляется доступ к заданному регистру для чтения либо записи. Содержательно команда представляет элементарную процедуру вычислительного процесса, поддерживающую бинарную операцию с двумя входными операндами и одним результатом. Так, например, A_1 и A_2 могут обозначать адреса, по которым в памяти размещаются два входных операнда, а A_3 обозначает адрес для записи результата. Процессор связан с памятью каналом доступа, память служит местом размещения машинных слов. Память не различает слова как команды или операнды и это позволяет записывать в неё и программы и данные. Программист условно разделяет пространство памяти на сегменты для раздельного размещения программ и данных. Вычислительный процесс представляет собой поток обменов процессора с памятью, в котором циркулируют машинные слова, имеющие статус команд и операндов. Различение машинных слов на команды и операнды происходит при чтении. При этом команды читаются в устройство управления, а операнды в арифметико-логический блок.

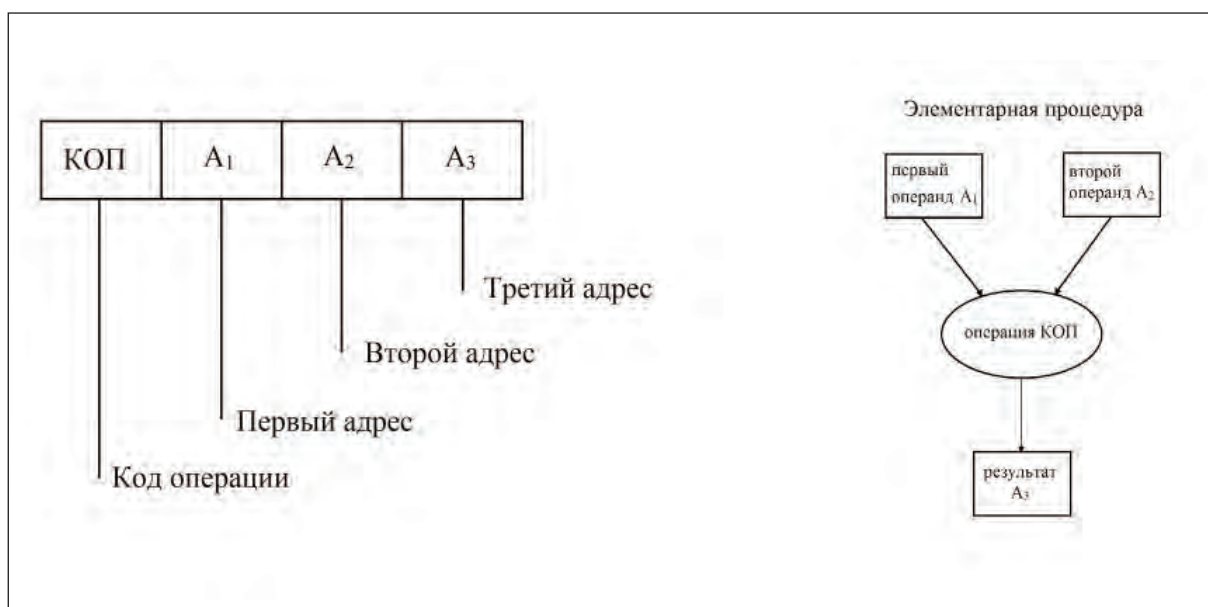


Рис. 4. Структура команды

Изложенное до сих пор позволяет зафиксировать одно очень важное следствие – знаковая система классической машины не является языком описания процессов вычисления и обработки данных. Язык машины является средством описания процессов преобразования состояний памяти. Команды, из которых состоит программа, привязаны не к данным, а к адресам памяти. Для того что бы процесс преобразования состояний памяти превратился в процесс обработки данных необходимо принять меры, обеспечивающие размещение нужных данных в нужных адресах памяти строго к тому моменту, когда соответствующая команда будет активирована и осуществит обращения к заданным адресам. Эти меры осуществляются как на аппаратном, так и на программном уровнях и являются основными источниками издержек и накладных потерь, сопровождающих вычислительный процесс. Для оценки эффективности организации процессов в гипотетической машине нам необходимо рассмотреть временную диаграмму и установить соотношение временных затрат на полезную работу и на сопутствующие обеспечивающие события, которые являются накладными потерями ресурса времени. В итоге нам необходимо оценить коэффициент полезного действия (КПД) машины как процентное соотношение разных работ на временной шкале.

Основным элементом временной диаграммы машины является такт выполнения команды, который состоит из следующих пяти микротактов: чтения команды, чтения первого операнда, чтения второго операнда, выполнения операции и записи результата. Сделаем допустимое в данном случае огрубление и будем считать, что все микротакты имеют равное время выполнения. Полезным квантом временной диаграммы считается только один микротакт – выполнение команды. Все остальные являются обеспечивающими и составляют накладные потери или издержки. Тогда получается, что на уровне выполнения одной команды языка машины во временной диаграмме полезный квант занимает 20%, а на издержки приходится 80%.

При построении простейшей процедуры некий стереотипный фрагмент программы повторяется многократно. Например, обрабатывается таблица, имеющая некоторое число атрибутов в строке и несколько сотен строк. Процедура состоит из программы обработки строки и цикла, повторяющего эту программу многократно по числу строк. Мы знаем, что адресные поля команды привязывают её к конкретным регистрам памяти. Это позволяет написать программу процедуры обработки строки таблицы. При повторном вызове процедуры программу обработки строки таблицы необходимо привязать уже к другим адресам, в которых размещается другая строка таблицы. При этом существует запрет на изменения в теле программы в ходе её выполнения. (В противном случае будет утерян принцип стереотипности и повторяемости программы). Проблема перемещаемости процедуры над данными решается через механизм косвенной адресации. При косвенной адресации в адресных полях команды не указывается непосредственный адрес размещения операнда, а указывается некий фиксированный в памяти регистр, в котором находится реальный адрес операнда. Далее по этому адресу осуществляется обращение в память за нужным операндом. При косвенной адресации для доступа к операнду необходимо осуществить два цикла обращения в память. Сначала обратиться в фиксированный регистр, извлечь, содержащийся в нём адрес и далее по этому адресу извлечь операнд. Известно, что время исполнения команды в основном определяется временем цикла обращения в память. Следовательно, полученные нами 20% КПД необходимо снизить вдвое до 10%. И это ещё не всё. Для реализации перемещаемости процедуры над данными необходимо перед началом каждого следующего

прохода в цикле заменить адреса в фиксированных регистрах. Это осуществляется за счёт приращения адресов на некий фиксированный шаг. Таким образом, возникает целый пласт вычислительных работ, который называется адресная арифметика. Объёмы преобразований адресной арифметики равны объёмам основной обработки, поскольку необходимо модифицировать адреса всех операндов. Следовательно, КПД урезается ещё вдвое до 5%.

Из структуры команды, приведенной на рис. 4 следует, что возможности прямой адресации памяти из адресных полей команды весьма ограниченные. Если длина машинного слова равна 32 разрядам, каждое поле команды содержит по 8 разрядов. Это значит, что непосредственно из адресных полей команды можно адресовать не более чем 256 слов. По этой причине трёхадресный формат команды оказался проблемным и наибольшее распространение получил двухадресный формат, который позволяет при 32-х разрядном слове иметь два адресных поля по 12 разрядов и непосредственно адресовать 4096 слов. А это означает, что для реализации элементарной процедуры обработки в общем случае понадобится более одной команды, что только ухудшает временной баланс издержек. С учётом некоторой статистики применения коротких и длинных форматов команд мы вынуждены снизить КПД до 3%.

В современном компьютере адресное пространство оперативной памяти исчисляется гигабайтами. Дальнейшее расширение объёмов оперативной памяти стало возможным при осуществлении механизма многоступенчатой адресации. При ступенчатой адресации, например при двухступенчатой, пространство памяти расчленяется на страницы, регистры которых доступны из адресного поля команды. Страницам, в свою очередь, присваиваются свои адреса или номера страниц. Для реального доступа в память необходимо совместить адресное поле команды и номер страницы. При этом номер страницы представляет старшую группу разрядов, а адресное поле команды младшую группу разрядов реального адреса обращения в память. Номер страницы хранится в специальном индексном регистре в составе процессора. Это так называемый блок регистров общего назначения, сохраняющих вычислительную обстановку процессора и в частности его позиционирование в общем поле памяти. Назначение номера страницы и его занесение в индексный регистр осуществляет загрузчик, специальная системная программа низкого уровня, осуществляющая управление ресурсами памяти и обслуживающая загрузку приложения в память. Мы не станем здесь излагать детали этого достаточно громоздкого механизма, а только вынуждены будем констатировать, что осуществление перемещаемости приложения в памяти большого объёма и механизм многоступенчатой адресации урезают оценку КПД ещё наполовину до 1,5%.

При оценке эффективности организации вычислений, приведенной выше, мы ограничились рассмотрением уровня простейшей процедуры, погружённой в аппаратную среду. Это заведомо очень упрощённая ситуация. На самом деле необходимо понимать, что реальные условия работы компьютера осуществляются в системной среде, поддерживающей мультизадачные режимы. Выполнение системных функций и обработка прерываний ложатся на тот же единственный процессор и продолжают нагружать временную диаграмму. Так что в реальности эффективность функционирования современного классического компьютера по нашей экспертной оценке не превышает одного процента. В качестве метафоры можно утверждать, что на одно полезное действие приходится не менее сотни обеспечивающих, которые являются издержками принятых форм организации процессов.

Традиция проектирования аппаратных средств вычислительной техники складывалась в условиях подавляющей высокой стоимости аппаратуры. Поэтому издержки укладывались в последовательные ряды и выполнялись на одном оборудовании последовательно. По мере развития технологии микроэлектронного производства принцип экономии аппаратных затрат перестал быть доминирующим. Современные инженерные решения создают другую ситуацию, отличную от той на которой мы провели вышеприведенную экспертную оценку эффективности функционирования компьютера. Так, например, в современных процессорах программы и данные размещаются в разных блоках памяти, а выборки команд и данных осуществляются по разным каналам и совмещаются во времени, что заметно улучшает временную диаграмму. Текущие операнды подкачиваются порциями в сверхоперативную память и обеспечивают бесперебойную работу арифметического блока. И таких решений множество. Суть их сводится к тому, что издержки выносятся из последовательных рядов временной диаграммы и размещаются в дополнительной аппаратуре, которая работает параллельно с основной. Но главное заключается в том, что груз издержек остаётся прежним. Издержки не устраняются, а поглощаются дополнительной аппаратурой, а аппаратура это и есть основной ресурс кристалла, выраженный в числе транзисторов. Технология обеспечивает размещение миллиарда транзисторов на кристалле, и есть две стратегии его заполнения. Можно сохранить примитивную структуру процессора с последовательным размещением издержек во времени и построить кристалл, содержащий 100 ядер. А можно усложнить процессор разгрузить его временную диаграмму и разместить издержки на дополнительной аппаратуре. Но тогда на кристалле будет размещаться не 100 низкоскоростных ядер, а 20 высокоскоростных. Суммарная пропускная способность кристалла останется прежней, поскольку определяющим является факт существования сотни обеспечивающих действий на одно полезное. Это константа, характеризующая архитектурную идею. Именно введение критерия эффективности по соотношению числа полезных действий на единицу затрат энергии позволяет осуществить объективную оценку эффективности инженерных решений. Вынос обеспечивающих действий из временной диаграммы и размещение их на дополнительной аппаратуре создаёт иллюзию повышения эффективности при условии, что стоимость транзисторов ничтожна. Но энергию эти действия потребляют при любом размещении и при разных инженерных решениях никуда не деваются. Выделенные нами при экспертной оценке накладные потери являются следствием архитектурной идеи и порождаемых ею форм организации процессов. Инженерные ухищрения не устраняют издержки, а манипулируют их размещением, и только вновь введенный критерий энергоэффективности кристалла позволит запустить тенденцию к критической оценке и пересмотру архитектурной концепции.

Начинать работы по созданию высокопараллельных структур с таким грузом издержек и такой низкой эффективностью организации процессов, по меньшей мере проблематично, а по результатам, рассмотренных нами в первой части статьи примеров, бесперспективно. При распараллеливании издержки тиражируются кратно степени параллелизма. Но при этом неизбежно возникают новые специфические издержки на обслуживание параллелизма. Мероприятия по организации параллельной работы множества процессорных элементов несут в себе тот же порок низкоэффективных форм организации процессов. Т.е. при обслуживании параллелизма вновь генерируются десятки и сотни накладных расходных событий на единицы полезных. Поэтому при решении целого ряда актуальных задач с высоким потенциалом параллелизма насыщение роста

производительности наступает практически сразу при 10 – 15 процессорах и значениях 3-х, 4-х кратного ускорения. Параллелизм блокируется низкоэффективными формами организации вычислительных процессов. Факт существования эффекта насыщения не является фатальным. Точку насыщения можно перемещать и в перспективе обеспечивать реализацию высоких значений параллелизма. Фатальной является крайне низкая эффективность организации процессов в классической архитектуре.

Существуют ли альтернативные решения, позволяющие строить вычислительные системы с более высокими показателями эффективности организации вычислительных процессов? Имеются ли приемлемые решения в области разработки неклассических архитектур? Наиболее значительным событием в этом направлении было создание архитектуры потока данных, известной как архитектура Data Flow. Первые публикации по архитектуре потока данных [10] были восприняты с большим энтузиазмом. Машина потока данных действительно представляет собой принципиально новый архитектурный проект, поскольку язык машины потока данных существенно отличается от линейного императивного языка машины фон Неймана. Перечислим эти отличия. Язык классической машины описывает программу как линейную последовательность команд. Команда представляет фиксированную элементарную процедуру вычислительного процесса. Адресные поля команды ссылаются не на данные, а на адреса регистров памяти. Команды исполняются в порядке их следования в записи программы. Здесь следует вновь напомнить, что внутренний язык классической машины не является языком описания вычислительных процессов, он описывает процессы изменения состояний памяти. Подгон процедур изменения состояний памяти к вычислительному процессу есть основное содержание работы программиста, но это содержание не фиксируется средствами языка машины. Это есть источник некорректности технологии программирования классических машин. А главный порок классической архитектуры это антагонизм последовательной записи процесса с его параллельной и асинхронной природой.

Основные положения архитектуры Data Flow мы уже излагали в первой части настоящей статьи [9]. Однако требование удобства чтения текста обязывают нас изложить эти положения повторно. Язык машины потока данных существенно отличается от языка классической машины по всем перечисленным позициям. Основной лексической единицей языка машины потока данных является пакет. Структура пакета приведена на рис. 5.

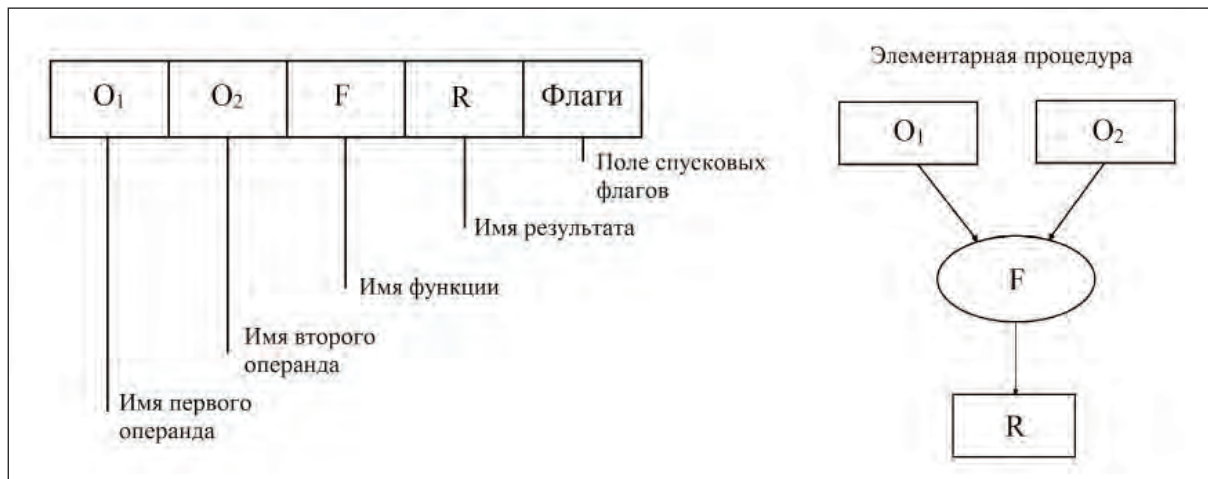


Рис. 5. Структура Data Flow пакета

Data Flow пакет также как и команда описывает элементарную процедуру вычислительного процесса, но в полях пакета записываются имена данных и это делает его языком описания вычислений. Элементарная процедура, описанная пакетом, не фиксирована, как в системе команд классической машины. Принимается гибкая система обозначений, при которой функция пакета может быть машинной командой или программой процедуры. Данные могут быть элементарными операндами или структурами данных. Такая гибкость очень важна при программировании параллелизма. Пакеты размещаются в памяти пакетов в произвольном порядке. В совокупности они образуют запись графа вычислительного процесса. Связность графа поддерживается взаимными ссылками имён аргументов и результатов. Память данных в потоковой машине является многофункциональной. Память данных осуществляет доступ к данным по именам, ведёт учёт их наличия и взаимодействует с памятью пакетов путём отправки фишек с именами поступивших данных. Фишки, принятые в памяти пакетов иницируют операции над спусковыми флагами. В ситуации, когда спусковые флаги пакета сигнализируют о наличии поименованных аргументов, пакет переводится в активное состояние и выставляет маркер готовности. В каждый текущий момент времени в памяти пакетов образуется множество готовых пакетов, которые отправляются на обработку, а в результате их обработки порождаются новые данные, которые спускают на обработку новые порции пакетов. Логика функционирования знаковой системы (языка) машины потока данных не требует специальных мер программирования параллелизма. Знаковая система самостоятельно извлекает динамический параллелизм из текущего состояния данных.

Перечисленные качественные характеристики архитектурной идеи, лежащей в основе проекта Data flow, воспринимаются как более предпочтительные для построения систем с массовым параллелизмом. Однако до настоящего времени неизвестно ни одного варианта их промышленной реализации - проект Data flow не выходит из стадии исследования и разработки. Наш опыт модельных испытаний системы с элементами Data flow, описанный в первой части настоящей статьи даёт тот же результат быстрого прекращения роста производительности на уровне 3-х кратного ускорения при 6 процессорах. Для того, что бы нащупать путь выхода из этого тупика необходимо понять, что объединяет оба проекта, несмотря на обозначенные нами существенные их различия. Общим для классической архитектуры и для проекта Data flow является принцип опережающей разметки трассы процесса. В обоих случаях знаковая система, лежащая в основе архитектуры, является средством создания символьной копии процесса. В классической машине символьная копия процесса существует в виде последовательности команд, а в архитектуре Data flow в виде потокового графа. В классической архитектуре команды считываются синхронно по одной в порядке следования в записи, а в Data flow активируются асинхронно группами в соответствии с логикой спуска по готовности данных. Но и в том и в другом случае активированные команды и пакеты не являются реальными агентами осуществления процессов обработки и вычислений, это только ссылки на данные и на функции. В след за активацией фрагментов трассы процесса разворачиваются потоки обеспечивающих событий, необходимых для поиска, доступа в память, чтения и транспорта данных к местам локализации функций обработки, а за тем по транспорту и размещению в памяти результатов обработки. Основным источником многочисленных накладных потерь являются работы по обслуживанию трассы процесса. Причём, запись трассы в виде потокового графа в Data flow в от-

личие от линейной последовательности в классической машине, порождает более сложные и объёмные работы по её обслуживанию. Проблема многократного использования стереотипных фрагментов трассы при линейной последовательной записи требует выделения двух точек записи – начала и конца процедуры. Для многократного повторного использования фрагмента потокового графа необходимо обозначить множество входных и выходных полюсов, привязанных к соответствующим вершинам графа. Далее необходимо организовать очереди фишек на входных полюсах, а сами фишки разметить по принадлежности к поколениям данных, возникающим в потоке обработки. Эта проблема получила название задачи раскраски фишек. А затем решать проблему синхронизации потоков фишек на всех входных полюсах процедуры. И всё это ещё не процесс вычислений – это только работы по обслуживанию записи трассы процесса. И это далеко не единственная проблема проекта, но она позволяет понять, почему проект Data flow до сих пор остаётся проектом.

Поиск аналогии в другой области. Похожая ситуация имела место на начальном этапе разработки проекта глобальной сети связи в конце 60-х годов [11]. Географически распределённые сети связи континентального и межконтинентального масштаба управлялись централизованно через центры коммутации. Сети функционировали на базе технологии коммутации каналов. Для осуществления сеанса связи через центры коммутации прокладывался маршрут, который монополизировал дорогостоящие магистральные ресурсы на время сеанса для пары абонентов. Для быстрого обслуживания плотных потоков требований на соединение приходилось наращивать число параллельно работающих каналов. Когда проектировщики задумались о построении глобальной единой автоматизированной системы связи (ЕАСС) и стали подсчитывать необходимые объёмы капитальных вложений оказалось, что проект по стоимости неподъёмный для самых богатых стран и их совместных усилий. Тогда инженеры связи сделали экспертные оценки эффективности использования магистралей связи. Оказалось, что эффективность их загрузки чудовищно низкая и лежит в пределах нескольких процентов. Существовала проблема выбора - или строить и вводить в эксплуатацию гигантские объёмы новых магистралей или решить системную проблему и поискать более продуктивные технологии связи с лучшими показателями эффективности. Результат нам известен - была создана технология коммутации пакетов, современная глобальная паутина это сеть пакетной коммутации [12].

В пакетной сети отсутствует централизованное управление и коммутация каналов. Сети равномерно покрывают территории как связные совокупности узлов, где каждый узел связан со смежными узлами, каналами открытого общего доступа. Информация, циркулирующая в каналах, имеет цифровое представление. Сообщения фрагментируются на относительно небольшие порции, которые оформляются в сетевые пакеты. Сетевой пакет это формализованная запись, в которой есть блок фиксации содержательной информации и ряд сопровождающих функциональных полей. Содержательная информация может быть графической, текстовой или акустической, но сеть к этому безразлична. Интерпретация содержательной информации осуществляется терминальным оборудованием, принадлежащим абонентам. Сеть становится унифицированной. Пакеты порождаются терминальным оборудованием пользователей, большими массами вбрасываются в сеть и начинают самостоятельно продвигаться по сети от узла к узлу. Ресурсы сети никогда не монополизированы отдельными сеансами связи и всегда открыты для доступа. Алгоритмы обработки пакетов в узлах построены таким образом,

что пакет может самостоятельно выбирать маршрут и прокладывать свой путь к заданному адресату. Вся необходимая управляющая информация содержится в функциональных полях пакета, сопровождающих поле содержательной информации. В узлах имеются полные каталоги маршрутов и кроме того узлы периодически обмениваются между собой служебными сообщениями, в которых обозначают свою текущую загрузку. Таким образом, пакет на каждом шаге продвижения от узла к узлу уточняет свой маршрут и имеет возможность обхода перегруженных участков сети. В результате сетевая технология пакетной коммутации позволила увеличить продуктивность сетей и эффективность использования магистральных ресурсов на порядки.

Подводя итог обсуждения проблемы эффективности методов организации вычислений можно отметить, что необходимость существенного обновления архитектурной идеи вполне назрела и контуры новой архитектуры определились. Далее мы сформулируем перечень основополагающих принципов новой архитектурной концепции.

Отказ от принципа раздельного существования потоков команд и потоков данных. Должен функционировать один поток - поток данных, сопровождаемых управляющими функциональными полями по аналогии с организацией сетевых пакетов в сетях пакетной коммутации. Основные лексические элементы знаковой системы, организующей вычислительный процесс должны представлять собой самоопределяемые операнды. Самоопределяемый операнд состоит из поля, несущего содержательную информацию, и набора сопровождающих функциональных полей, несущих всю управляющую информацию, детерминирующую поведение и виды активности самоопределяемого операнда.

Отказ от принципа опережающей разметки трассы процесса как следствие отказа от потока команд, которые и были средством формирования символьной копии процесса. По сути это отказ от принципа сосредоточенного управления процессом. Символьная копия процесса отсутствует, а вся управляющая информация распределена по непосредственным участникам процесса – самоопределяемым операндам. Трасса процесса теперь не актуализована в записи программы, а проявляется косвенно в динамике вычислений как результат наблюдения за процессом. Это и есть принцип распределённого управления.

Отказ от технологии коммутации каналов означает, что обмен данными не должен существовать как отдельная операция, связывающая параллельные фрагменты процесса. При сосредоточенном управлении процессами параллелизм порождает необходимость обмена данными как отдельный компонент процесса, требующий отдельной физической обменной среды. При этом обменные операции, порождают накладные затраты, растущие с ростом параллелизма. В новой архитектуре принимается принцип сетевой организации вычислений, при которой движение данных совмещается с процессом обработки. Практически это означает устранение значительных объёмов паразитных пересылок данных и максимальный параллелизм необходимых обменов данными на низовом уровне, не требующем специальных коммутационных сред.

Перечисленные принципы соответствуют тенденциям, которые фактически сложились в практике построения многопроцессорных систем. Наиболее полное воплощение сетевых принципов организации процессов просматривается в структурах с топологией тора [3, 4, 5]. В тороидальных топологиях применяется принцип пакетной маршрутизации данных и главное заключается в том, что в узлах сети размещаются процессорные элементы, а в целом узлы сети совмещают функции обработки и коммутации данных.

Осталось сделать лишь завершающее усилие – расширить функции распределённого управления маршрутизацией пакетов до функций управления вычислениями и далее распространить сетевые принципы управления на все уровни организации вычислений.

3. Архитектура самоопределяемых данных и принципы распределённого управления

Идея построения архитектуры самоопределяемых данных имеет определённую предысторию, некоторые её фрагменты изложены в ряде публикаций и патентов прошедших лет [13, 14, 15]. Однако в прошлые десятилетия идея не воспринималась как конкурентоспособная. Технология позволяла удовлетворять требования закона Мура и наращивать производительность вычислительных средств исключительно за счёт роста тактовой частоты и усложнения процессорного элемента. Кроме того смена принципов организации вычислений нарушала требование преемственности и переносимости, наработанных ранее программных продуктов и резкая смена архитектуры была экономически не целесообразна. Но в настоящее время, когда реализация массового динамического параллелизма становится основным направлением развития вычислительных средств, старый проект может стать востребованным и занять определённую нишу в существующем разнообразии областей применения.

Рассмотрим базовые принципы функционирования архитектуры самоопределяемых данных. Основным элементом знаковой системы является самоопределяемый операнд. Это упорядоченная запись, состоящая из содержательного поля, в котором размещается операнд и нескольких сопровождающих функциональных битовых полей. Формат самоопределяемого операнда приведен на рис. 6.

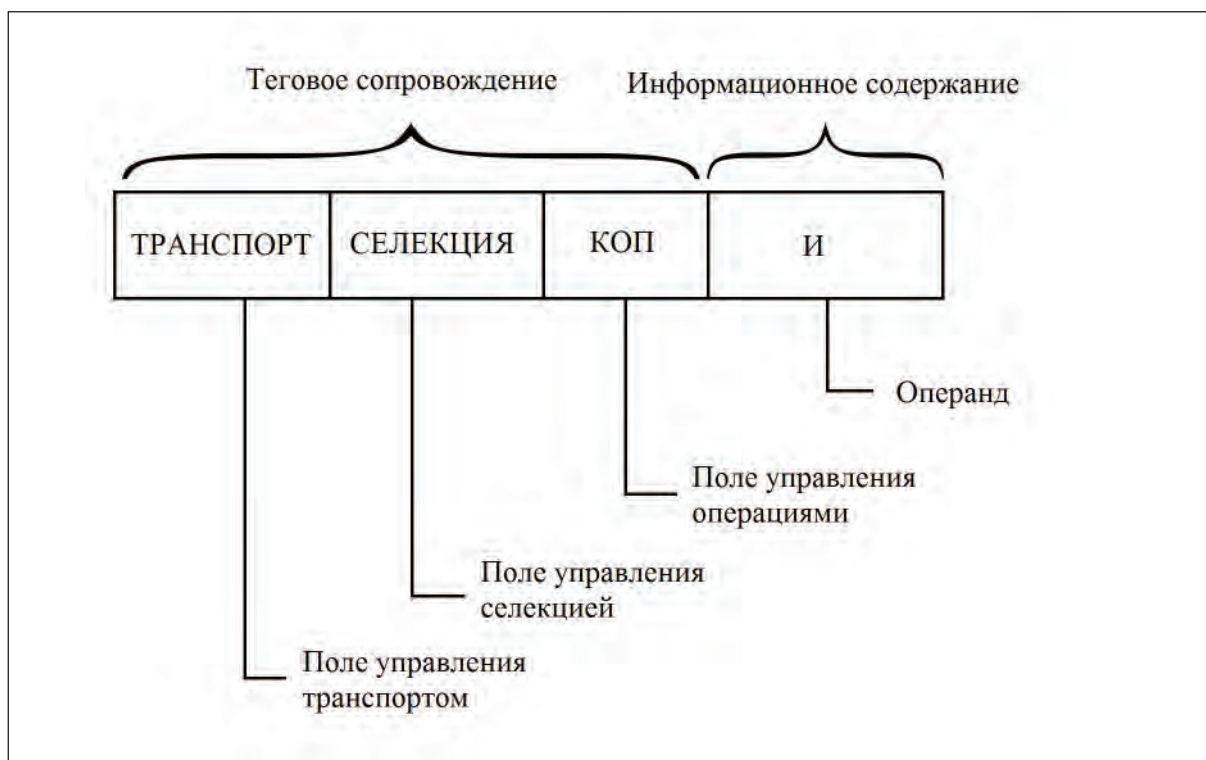


Рис. 6. Формат самоопределяемого операнда

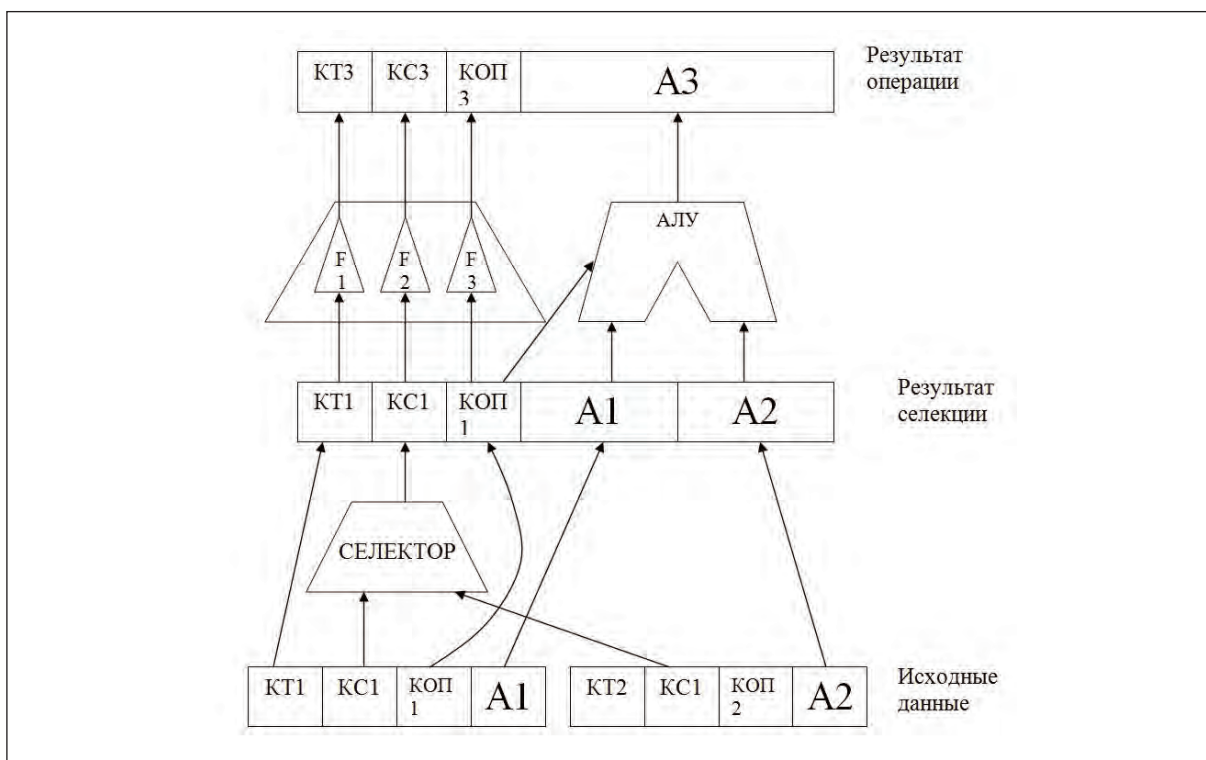


Рис. 7. Схема взаимодействия самоопределяемых операндов

Набор функциональных полей называется теговым сопровождением и включает в себя поле управления операциями, поле управления селекцией пар операндов и поле управления транспортом. Кроме перечисленных функциональных полей в теговом сопровождении может быть предусмотрено резервное поле, предназначенное для оперативного расширения управляющих функций в случае необходимости. Это означает, что набор управляющих полей открыт и может корректироваться по текущей обстановке. В приведенном формате рассматривается минимальный набор управляющих полей. Смысл управления состоит в том, что самоопределяемый операнд должен осуществлять селекцию пары для выполнения бинарной операции, определять тип операции обработки и принимать решения о перемещении по транспортным путям вычислительного устройства.

Рассмотрим логику взаимодействия самоопределяемых операндов проиллюстрированную диаграммой на рис. 7.

В нижней части схемы изображены два операнда, предназначенные для выполнения заданной бинарной операции. Операция селекции может осуществляться аппаратно или программно и смысл селекции состоит в том, что селектор отбирает пары операндов по кодовым состояниям полей селекции. На выходе селектора формируется блок, называемый парой. Пара состоит из двух информационных полей A_1 и A_2 , предназначенных для выполнения бинарной операции. К паре пристыковывается теговое сопровождение одного из операндов, второй тег после селекции отбрасывается. Отселектированная пара направляется в канал обработки, который состоит из двух секций – арифметического блока и преобразователя теговых кодов. Арифметический блок это стандартное арифметико-логическое устройство (АЛУ), которое управляется кодом операции из тегового сопровождения пары. Секция преобразования тегов содержит для

каждого тегового поля свой унарный функциональный преобразователь. В результате прохождения через канал обработки образуется новый операнд A_3 с новым теговым сопровождением.

Вычислительное устройство состоит из множества каналов обработки и путей перемещения, в которые загружается множество исходных операндов. Поведение исходных операндов определяется кодами их теговых сопровождений. В результате реализации предписанных тегими событий порождается второе поколение операндов с новыми кодами тегов, определяющими следующий шаг обработки.

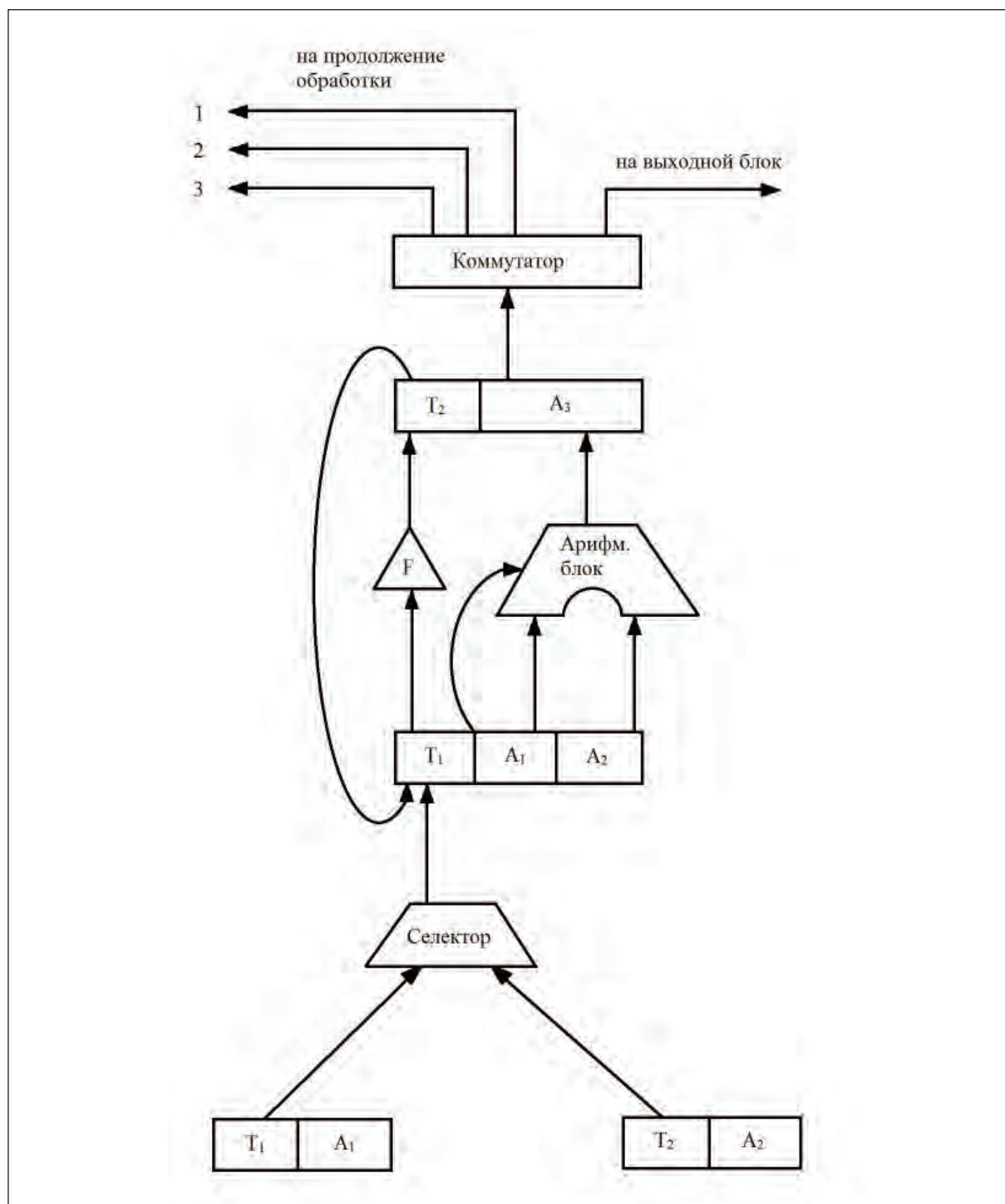


Рис. 8. Схема взаимодействия упрощённых самоопределяемых операндов

Для более детального рассмотрения динамики развития процесса обработки введём упрощённый вариант построения самоопределяемых операндов. Будем считать, что селекция пар операндов осуществляется по совпадению кодов селекции. Это даёт возможность совместить в одном поле управление селекцией и кодирование операции, поскольку для селекции важен только факт совпадения кодовых значений пары селективируемых операндов. Также с целью упрощения ситуации можно опустить поле управления транспортом. В результате мы станем рассматривать упрощённый самоопределяемый операнд с одним теговым кодом. Схема взаимодействия таких упрощённых операндов изображена на рис. 8.

Селектор вылавливает в потоке самоопределяемых данных операнды с совпадающими значениями теговых кодов и формирует из них пару. Пара операндов поступает в канал обработки, где над информационными полями выполняются арифметические операции, а теговые коды проходят через дискретный функциональный преобразователь. На выходе канала обработки формируется новый операнд с новым значением тега. На выходе канала обработки изображён коммутатор, в котором определяется выбор направления передачи операнда. Дешифрирующая схема коммутатора анализирует определённые разряды тега и принимает решение о перемещении операнда, либо в выходной блок для формирования выходных данных, либо в одно из трёх направлений продолжения обработки. Приведенная схема иллюстрирует, как именно самоопределяемый операнд прокладывает путь своих перемещений в аппаратной среде.

Далее новые поколения операндов вновь вылавливаются селекторами и передаются в каналы обработки, где теговые коды вновь проходят через функциональный преобразователь. Путь превращений теговых кодов выделен на схеме дугообразной стрелкой, замыкающей вход и выход функционального преобразователя. Динамика развития вычислительного процесса осуществляется следующим образом. В секции преобразования теговых кодов всех каналов обработки загружается определённая функция F . В вычислительное устройство вводится набор исходных операндов, с заданными значениями тегов. В силу принятых принципов построения аппаратуры из исходного набора операндов селекторами отбирается множество пар, которые загружаются в каналы обработки для выполнения бинарных арифметических операций. Таким образом осуществляется первый шаг обработки, который полностью определяется состояниями теговых кодов, присвоенных исходному набору операндов. В результате выполнения первого шага на выходе каналов обработки формируется второе поколение операндов с новыми значениями тегов. Далее поведение вновь образованных операндов полностью определяется значениями новых теговых кодов и таким образом осуществляется следующий шаг. Динамическим ядром процесса обработки является рекуррентный генератор кодовых последовательностей, который просматривается на схеме рис. 8. Рекуррентный генератор работает в режиме самообращения и на каждом шаге преобразования передаёт выходной теговой код на вход преобразователя тегов F в качестве нового аргумента.

В классической машине динамика вычислений формируется в результате последовательного считывания программы, которая представляет собой статическую запись опережающей разметки трассы процесса. В самоопределяемых данных процесс развивается динамически шаг за шагом и задаётся в аналитической форме в виде функции преобразования теговых кодов F . При этом важно отметить, что в данном случае аналитическое представление динамики процесса не является функцией от времени, а является функцией от предыдущего состояния. В архитектуре самоопределяемых дан-

ных роль программы как детерминанта процесса выполняет функция преобразования тегов **F** и начальные состояния теговых кодов в исходном наборе операндов.

Можно ли в описанной логике взаимодействия самоопределяемых операндов сконструировать осмысленный вычислительный процесс? Рассмотрим конкретный пример на базе введенного ранее упрощённого представления самоопределяемых операндов с одним теговым полем, совмещающим кодирование селекции пар и арифметических операций. Для простоты и наглядности примера будем считать, что тег имеет минимальную разрядность три бита, а функция дискретного кодового преобразователя **F** есть простая операция сдвига битового набора вправо на один разряд. Трёхразрядный битовый код имеет всего восемь возможных состояний и заданное отображение **F** можно представить в виде таблицы. Отображение **F** также можно представить в виде графа, в котором кодовые состояния это вершины графа, а рёбра определяются как пары $x; F(x)$. Назовёт такой граф графом кодовых переходов. Таблица отображения **F** и граф кодовых переходов приведены на рис. 9. На графе кодовых переходов для удобства чтения вершины обозначены как целые десятичные числа. Необходимо также назначить кодовым значениям тега арифметические операции и зафиксировать соответствие в виде таблицы, по которой арифметический блок будет интерпретировать теговые коды. Таблица кодирования арифметических операций также приведена на рис. 9. Поскольку кодовых состояний 8, а арифметических операций только 4, целесообразно привязать лишние коды к имеющимся операциям повторно. В данном случае соответствие не должно быть обязательно взаимно однозначным. Код «0» присваивается операции завершения процедуры и интерпретируется аппаратурой как останов.

И так мы сделали все необходимые определения и можем проследить как будет развиваться вычислительный процесс. Графическое изображение процесса в виде графовой диаграммы приведено на рис. 10. Прямоугольниками на диаграмме обозначены операнды, состоящие из двух полей – тега и информационного поля. Пирамидками обозначены процессы, включающие операции селекции пар, арифметические операции и операции преобразования тегов.

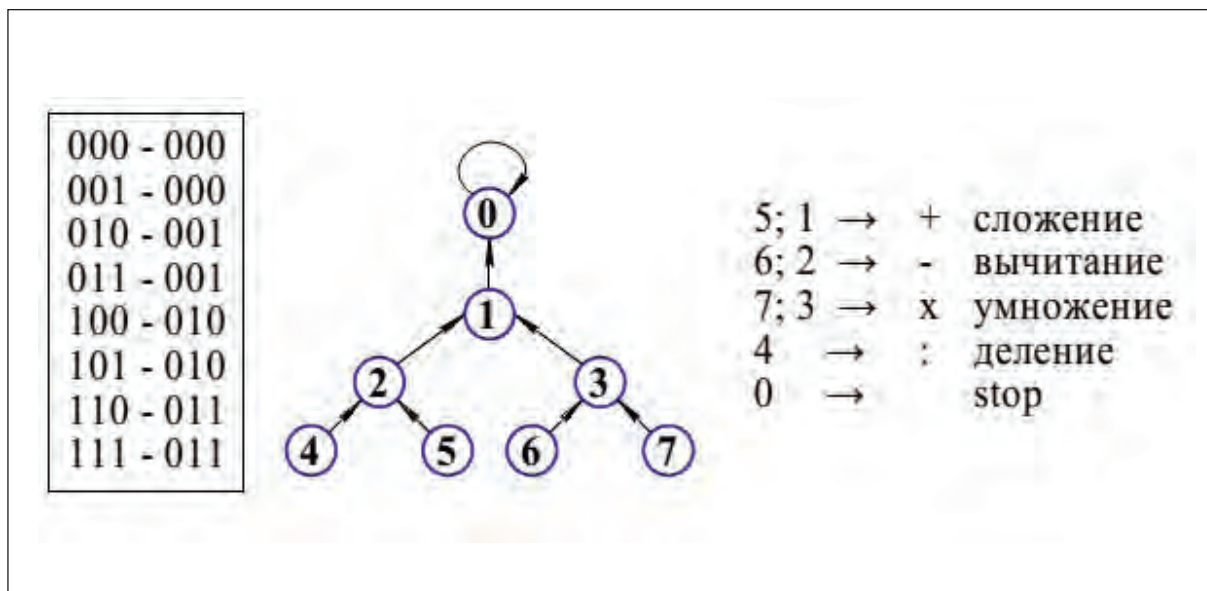


Рис. 9. Таблица отображения **F** и граф кодовых переходов

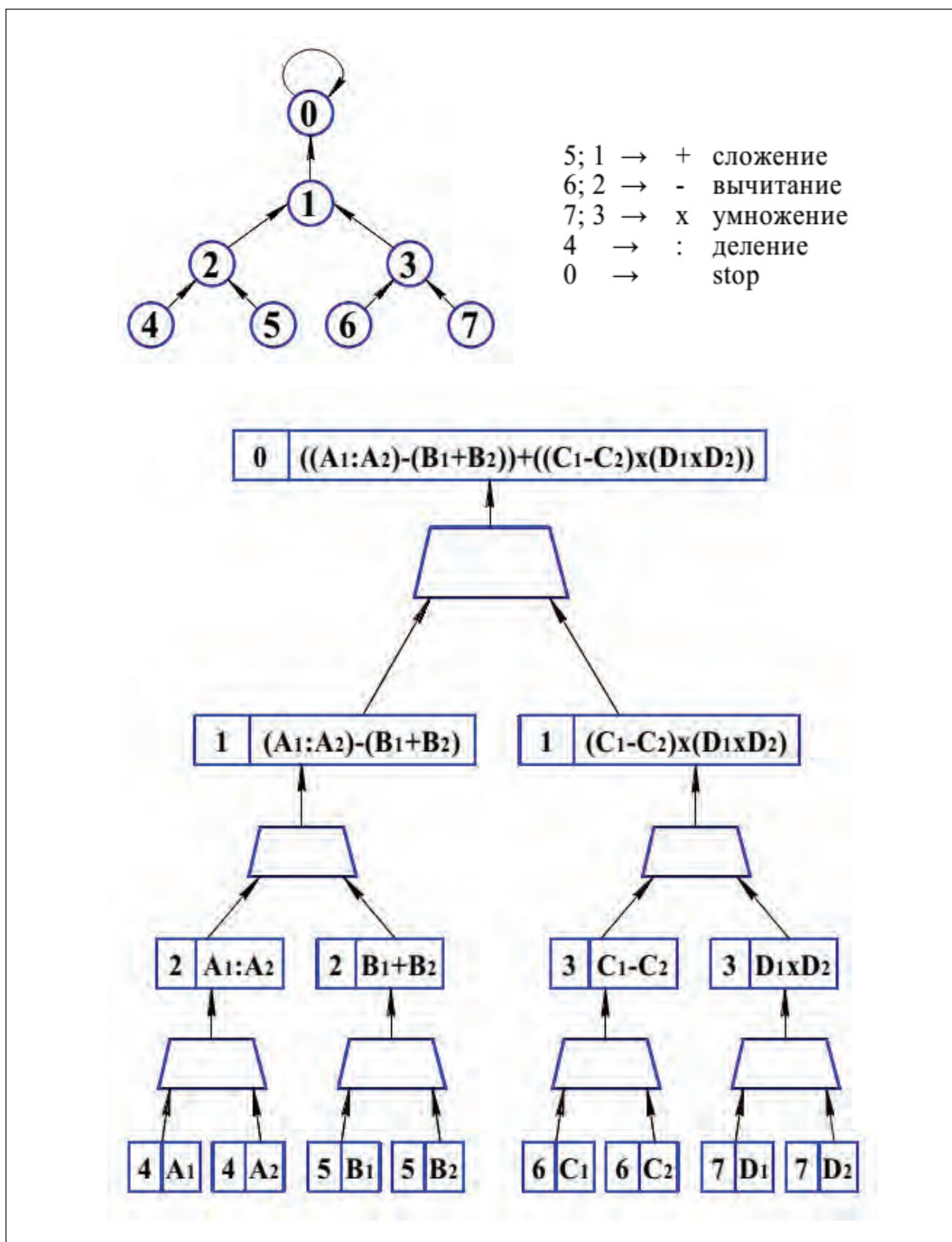


Рис. 10. Диаграмма вычислительного процесса

Перед началом работы во все каналы обработки вычислительного устройства загружаются преобразователи тегов **F** и таблицы интерпретации теговых кодов. Далее вводится исходный набор операндов. Исходные операнды заполняют нижний слой диаграммы на рис. 10. Теговые коды исходных операндов позволяют отселектировать пары

и направить их в каналы обработки, что отображено во втором слое диаграммы. Третий слой диаграммы заполнен результатами выполнения первого шага обработки, представляющими второе поколение операндов. Формирование содержания полей операндов второго поколения можно проследить по графу кодовых переходов и по таблице интерпретации тегов, которые для наглядности приведены на рис 10. Преобразования тегов прослеживаются на графе кодовых переходов, а выполняемые арифметические операции определяются таблицей кодирования операций. В информационные поля операндов записывается операция вычисления нового операнда. Во втором поколении операндов в полном соответствии с графом кодовых переходов образуются пары с одинаковыми тегами, которые на следующем шаге обработки вновь отлавливаются селекторами и передаются в каналы обработки. И таким образом события разворачиваются до появления тега со значением «0», что интерпретируется аппаратурой как операция останова и вывода результата в выходной блок.

На данной иллюстрации при изображении операндов следующих поколений, отличных от исходных и порождаемых в процессе обработки, в информационное поле вместо символического обозначения операндов вписаны процедуры их вычисления. Это необходимо для иллюстрации поэтапной сборки заключительного арифметического выражения, описывающего всю наблюдаемую вычислительную процедуру.

Рассмотренный пример показывает, что при заданной логике взаимодействия операндов и некоторых произвольно выбранных начальных условиях в системе развивается детерминированный процесс, соответствующий вычислению определённого арифметического выражения. Параллелизм извлекается аппаратурой из текущего состояния данных и не требует специальных мер синхронизации, привязки данных к каналам обработки и других мер программирования и обслуживания параллелизма. При условии, что во все каналы обработки загружена одна функция преобразования тегов и одна таблица интерпретации тегов, любой операнд может обрабатываться в любом канале. Где операнд оказался в текущий момент, там он и обрабатывается. При этом привязка операнда к определённой позиции на графе процесса обработки содержится в его теговом сопровождении и неотделима от него. Таким образом, архитектура самоопределяемых данных ликвидирует накладные потери в виде многочисленных паразитных пересылок данных из мест хранения к местам обработки и обратно.

Для наглядности и простоты примера мы выбрали малое значение разрядности тега, равное трём. Но даже в этом простейшем случае вычисляемое арифметическое выражение оказалось относительно протяжённым и не тривиальным. Достаточно увеличить теговый код на один разряд и число вершин в графе кодовых переходов удвоится, соответственно удвоится и сложность вычислительной процедуры. При 8-разрядном теге граф кодовых переходов с топологией бинарного дерева будет состоять из 256 вершин и 8 этажей, что позволит поддерживать достаточно сложные и многоэлементные вычислительные процедуры. При этом затраты на программирование и поддержку программы (вернее того, что является аналогом программы) останутся незначительными. В классической архитектуре сложность программирования и сложность алгоритма связаны очень просто. Для простой процедуры надо написать десять строк программы, а для сложной тысячу строк. А в самоопределяемых данных для усложнения программируемой процедуры на порядок достаточно только увеличить разрядность тегового кода примерно на три – четыре разряда.

В предыдущем разделе при оценке эффективности разных форм организации вычислений рассматривались механизмы, поддерживающие многократное использование стереотипных фрагментов программы или так называемых стандартных процедур. Было показано, что как в классической архитектуре, так и в архитектуре Data Flow обеспечение перемещаемости процедуры и привязка тела программы процедуры к разным данным является источником огромных издержек и в программном и в аппаратном обеспечении. Посмотрим как эта проблема решается в архитектуре самоопределяемых данных. Рассмотрим тот же пример с обработкой табличных данных, в котором строка таблицы содержит десяток атрибутов, а таблица в целом состоит из нескольких сотен строк. Создаётся процедура обработки строки, для неё формируется функция преобразования тегов **F** и набор теговых кодов, присваиваемых исходным операндам из строки таблицы. Для тиражирования процедуры и привязки её к разным строкам таблицы достаточно ввести в теговое сопровождение операндов дополнительное поле индексной метки строк таблицы. (В первичных определениях констатировалось, что набор теговых полей открыт для пополнения и введение новых полей не нарушает основы архитектуры и не требует перестраивать конструкцию машины). Тогда все операнды, принадлежащие одной строке таблицы, будут иметь одно значение в индексном поле, а разные строки будут иметь разные индексные метки. Поле индексной метки будет участвовать в операции селекции, но не будет участвовать в преобразовании теговых кодов. Далее можно загружать таблицу в вычислительную среду в произвольном темпе и в произвольном порядке. Операнды, принадлежащие одной строке всегда отсеlectionируются в один процесс обработки строки, а атрибуты разных строк никогда не пересекутся на всех стадиях выполнения процедуры. Любая процедура в архитектуре самоопределяемых данных может быть тиражирована с заданной кратностью путём расширения тегового сопровождения. Дополнительные разряды тега это необходимая плата за обеспечение требуемого свойства. Можно предполагать, что в данном случае издержки окажутся ощутимо меньше чем в классической архитектуре.

В рассмотренном примере реализации процедуры вычисления арифметического выражения функция преобразования тегов, которая является динамическим ядром процесса, порождает граф кодовых переходов с топологией бинарного дерева. Существует целый класс функций, заданных на конечных дискретных наборах и порождающих бинарные деревья разной размерности и с разным распределением кодов по вершинам графа. Понятно, что процесс вычисления арифметического выражения на базе двухместных операций имеет топологию бинарного дерева и может быть размещён на подходящем графе кодовых переходов, порождаемом функцией преобразования тегов. Означает ли это, что в архитектуре самоопределяемых данных можно реализовать только процессы с топологией бинарного дерева?

Рассмотрим возможности расширения разнообразия вычислительных процедур в самоопределяемых данных. Поскольку в машинной обработке приняты двухместные базовые операции, базовые процессы имеют топологию бинарного дерева. Но можно усложнить топологию реального процесса путём склеивания бинарных деревьев в определённых вершинах. Это потребует механизма многократного вхождения некоторых операндов в граф вычислений либо обмена операндами между разными поддеревьями. Создать набор непересекающихся бинарных деревьев можно при условии, что теговый код имеет достаточно большую разрядность и порождает большое дерево. Тогда при одной порождающей функции в графе кодовых переходов можно выделить набор

непересекающихся поддеревьев. Локализация поддеревьев определяется выбором листовых кодов. Для привязки групп исходных операндов к своим фрагментам процесса необходимо присвоить им в качестве тегов листовые коды выбранных поддеревьев. Многократное вхождение определённых операндов в граф вычислений можно обеспечить созданием в исходном наборе операндов ряда копий операнда с разными тегами, поскольку именно тег позиционирует место операнда на графе.

Широкие возможности многократного вхождения операндов в граф процесса обеспечивает некоторое усложнение операции селекции путём маскирования части разрядов теговых полей, в частности полей индексных меток. В этом случае возможен множественный отклик при селекции и организация векторной обработки.

Наиболее важным инструментом расширения динамики процессов в системе самоопределяемых данных является механизм операций с шаблонами. Это аналог команд перехода в классической машине. Команда перехода несёт в своём составе адрес новой точки в теле программы и нарушает естественный порядок следования команд путём замены состояния счётчика команд на адрес принудительного перехода. В самоопределяемых данных позиционирование операнда на графе процесса определяется теговым кодом. Смысл операции с шаблоном заключается в том, чтобы принудительно сменить теговый код операнда и перебазировать его в другой фрагмент графа процесса. Шаблон имеет тот же формат, что и самоопределяемый операнд и так же путём селекции находит своего партнёра, но в информационном поле шаблона вместо операнда записан новый тег для партнёра по селекции. При выполнении операции с шаблоном операнд в информационном поле не меняется, а получает новый тег из информационного поля шаблона. Операции с шаблоном позволяют организовать разнообразные условные и безусловные переходы, итеративные циклы и рекурсивные конструкции.

Рассмотренный пример показывает, что организация процессов в архитектуре самоопределяемых данных строится на базе двух компонент – динамического ядра и интерпретатора. Динамическое ядро порождает каркас процесса в виде графа кодовых переходов, а интерпретатор читает порождаемые коды теговых сопровождений операндов и придаёт им содержательное значение, например в виде арифметических операций. Архитектурная идея, не требуя изменений своей сути, допускает разнообразие интерпретаций. Информационные поля операндов могут содержать не только числа, а действия могут выполнять не только арифметические операции. Это могут быть символьные строки и текстовые объекты, это могут быть фрагменты графики или структур данных. Рассмотренный нами пример можно воспринимать как вычисление заданного значения, а можно считать его процессом сборки символической записи из базовых лексических единиц. Возможно значительное разнообразие применений архитектурной идеи самоопределяемых данных для построения реальных информационных систем.

Рассмотренный нами пример строился таким образом, чтобы продемонстрировать, что в принятой логике взаимодействия самоопределяемых операндов заданные начальные условия разворачиваются в определённый и осмысленный вычислительный процесс. Но это подтверждает принципиальную состоятельность выдвинутой архитектурной идеи только частично. Необходимо так же показать возможность решения обратной задачи – по заданному процессу построить начальные условия, которые смогут развернуться в этот процесс. Решение этой задачи означает возможность построения технологии программирования процессов в системе самоопределяемых данных. Из рассмотренного примера следует, что технология программирования самоопределяе-

мых данных существенно отличается от привычной технологии программирования классических машин. Суть этого отличия состоит в том, что эвристическое программирование в виде прямой записи символьной копии процесса без привлечения математических инструментов более невозможно. Для построения технологии программирования самоопределяемых данных требуется специальный математический аппарат описания процессов и целый ряд прикладных математических инструментов, разработанных на базе этого аппарата.

4. Основные понятия дискретной динамики

4.1. Работы В.И. Арнольда, основные принципы

Мы будем опираться на ряд фундаментальных работ В.И. Арнольда, которые тематически объединяются одним названием – исследование геометрии функциональных пространств, заданных на конечных множествах. Работы изложены в целом ряде публикаций [16, 17, 18].

Смысл этих работ заключается в том, что исследуется базовое понятие математики – структура, которая задаётся функциональным оператором на конечном множестве элементов. Основная масса исследований структур базируется на аналитической записи оператора и представляется как система манипуляции символическими конструкциями. Арнольд предложил представлять оператор и структуру в виде ориентированного графа. Графовое представление возможно лишь в тех случаях, когда множество, образующее структуру конечно. Граф структуры строится по правилу – элементы структуры являются вершинам графа, а рёбра определяются как пары $x; F(x)$. Возможно только однократное вхождение элемента в граф. Арнольд назвал этот граф **монадой**. Этим архаичным названием подчёркивалось фундаментальное значение введенного понятия как древнего первоосновного. Традиция исследования монад восходит к Ньютону, Лейбницу и возможно имеет и более ранних предшественников [18]. Арнольд отмечает, что первоначально построение монад носило эмпирический характер и оказалось интригующим и увлекательным занятием. В последствии ученики и аспиранты применили компьютеры для построения монад и дело двинулось более высокими темпами.

Основные свойства монад достаточно очевидны и следуют из первичных определений. В монаде всегда есть хотя бы один цикл, это следует из конечности исходного множества. Поскольку оператор образующий структуру всюду определён, в некоторый момент для результата применения оператора не хватит элементов, и результат замкнётся на ранее использованных. Если граф монады содержит несколько циклов, множество распадается на непересекающиеся классы элементов, тяготеющих к своему циклу. Это происходит, потому что между циклами рёбер не может быть, в противном случае нарушается функциональность отображения F . На рис. 11 приводится перечень возможных структур графов монад.

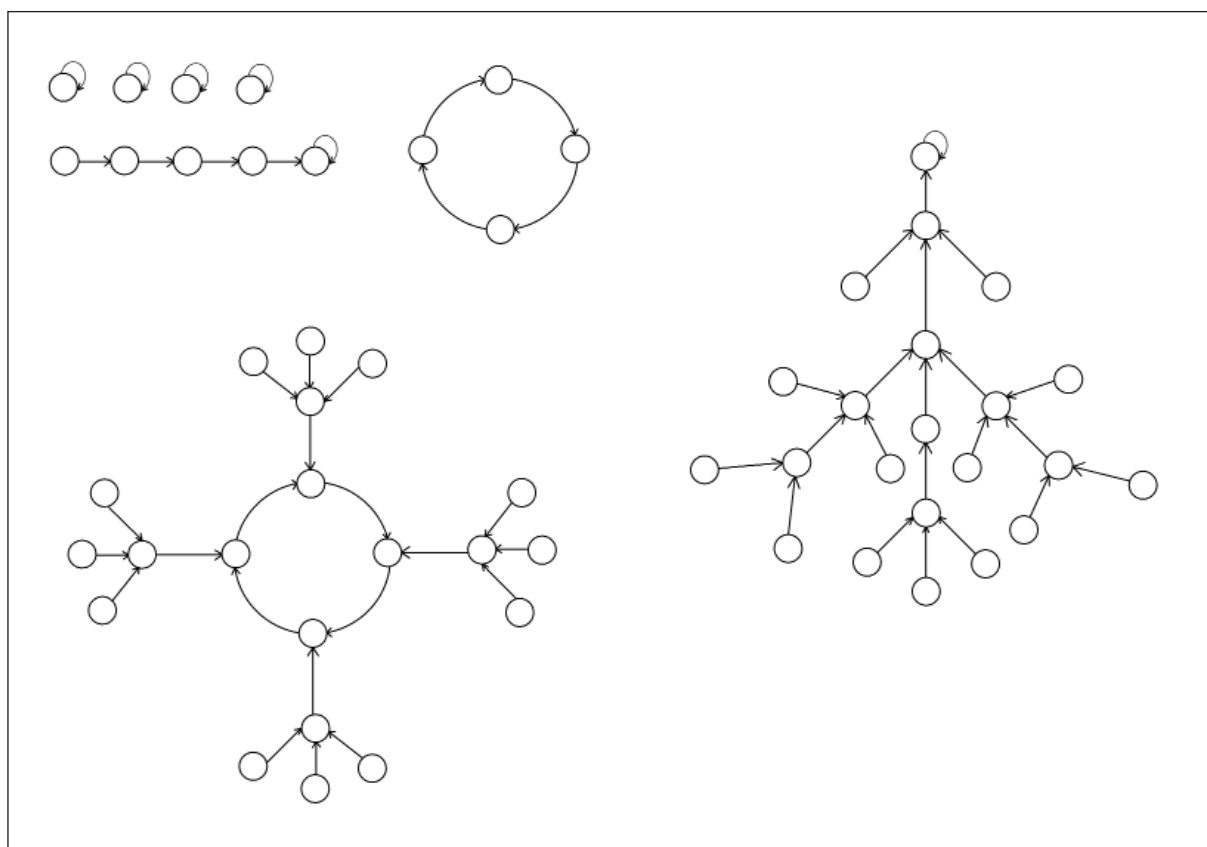


Рис. 11. Примеры графов монад

Множество элементов, тяготеющих к своему циклу, Арнольд назвал аттрактором. Исследования конечных структур, представленных монадами использовались Арнольдом для построения шкалы оценки сложности математических объектов, для исследований в теории чисел и ряде других направлений.

Представление структуры в виде монады позволяет увидеть тонкие закономерности и свойства, которые ускользают или маскируются при аналитической записи. Так, например, рассмотрим структуру, образованную дискретным функциональным преобразованием, заданным на множестве состояний компьютерного регистра. Оператор может представлять собой арифметическое соотношение или булеву функцию, составленную из набора побитовых булевых операций. Построение монад для данной структуры обнаруживает неожиданный факт - структура графа монады зависит от числа элементов образующего множества. При изменении разрядности регистров и неизменном функциональном операторе структура графа претерпевает значительные изменения. Пример иллюстрируется на рис. 12. Оператор F представляет собой следующую последовательность элементарных операций: над исходным операндом выполняется операция циклического сдвига на один разряд, далее полученный результат складывается с исходным операндом по модулю 2. На рис. 12 приводятся графики заданного отображения F при разных значениях разрядности регистров и соответствующие им графы монады.

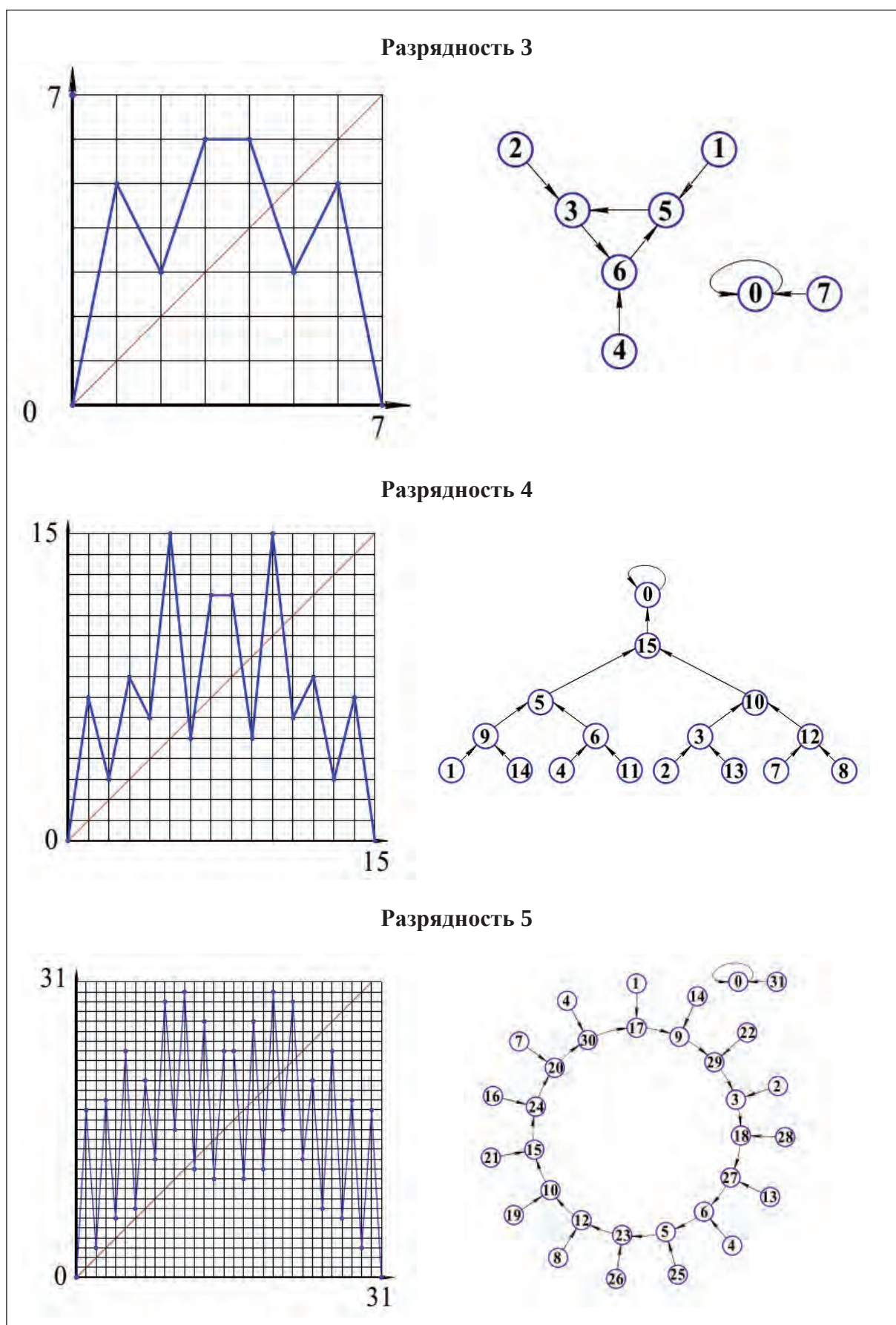


Рис. 1. Схема объединения узлов коммутаторами в топологии «толстое дерево».

Данные получены эмпирически и демонстрируют целый ряд нетривиальных закономерностей. Интересующий нас односвязный граф с топологией бинарного дерева проявляется при разрядности 4, затем наблюдается при разрядности 8 и далее повторяется с определённой периодичностью. На приведенных иллюстрациях мы вынуждены ограничиться простейшими картинками при малой разрядности регистров, но этого достаточно для понимания существенных изменений топологии графа монады при изменении числа исходных элементов. Этот эффект проявляется не для всех функций, образующих структуру. Отразить эти факты в аналитической записи структуры либо невозможно, либо очень непросто.

Мы примем за основу метод графового представления структур на конечных множествах с целью разработки дискретной динамики, как инструмента описания процессов в среде самоопределяемых данных.

4.2. Структуры на регистровых состояниях, граф кодовых переходов

Мы не будем рассматривать абстрактные множества, а ограничимся изучением вполне конкретных ситуаций во внутренней среде компьютера. Исходными множествами в нашем случае будут конечные наборы регистровых состояний, которые можно представлять, как битовые векторы определённой разрядности, либо записывать их как целые положительные натуральные числа. Операторы, образующие структуры на множествах регистровых состояний представляют собой дискретные преобразователи, которые можно реализовать аппаратно как логические схемы либо программно как наборы компьютерных команд. Оператор задаётся как функция и реализует функциональное отображение. Метод исследования структуры основан на представлении целочисленной функции в виде графа. Здесь требуется разъяснение – какая связь существует между функцией и графом.

Рассмотрим график целочисленной функции, заданной на конечном отрезке. Это будет совокупность точек на целочисленной решётке. Линия, соединяющая точки носит условный характер и может быть опущена. Совокупность точек на квадратной целочисленной решётке можно рассматривать как матрицу смежности, задающую ориентированный граф, рёбра которого есть совокупность пар вида $x; F(x)$. Так, что представление функции в виде ориентированного графа в данной ситуации совершенно естественно, хотя и непривычно. Поскольку функция определена на регистровых состояниях, станем называть этот граф графом кодовых переходов и обозначать G_F , граф, порождаемый функцией F .

Каждая точка графика целочисленной функции имеет две проекции - на ось x и на ось y . Если график функциональный, то на все точки оси x всегда имеется одна и только одна проекция графика. На точки оси y может проецироваться любое число точек графика, в том числе и ни одной. Из этого следует, что на графе кодовых переходов каждая вершина всегда имеет одно и только одно исходящее ребро. Входных рёбер может быть сколько угодно от 0 до N , где N число состояний регистра. Следовательно, все возможные графы кодовых переходов образуют специфический класс, графов, ограниченный определёнными правилами структурообразования.

На рис. 13 можно проследить, как свойства графика функции F проецируются на свойства графа кодовых переходов

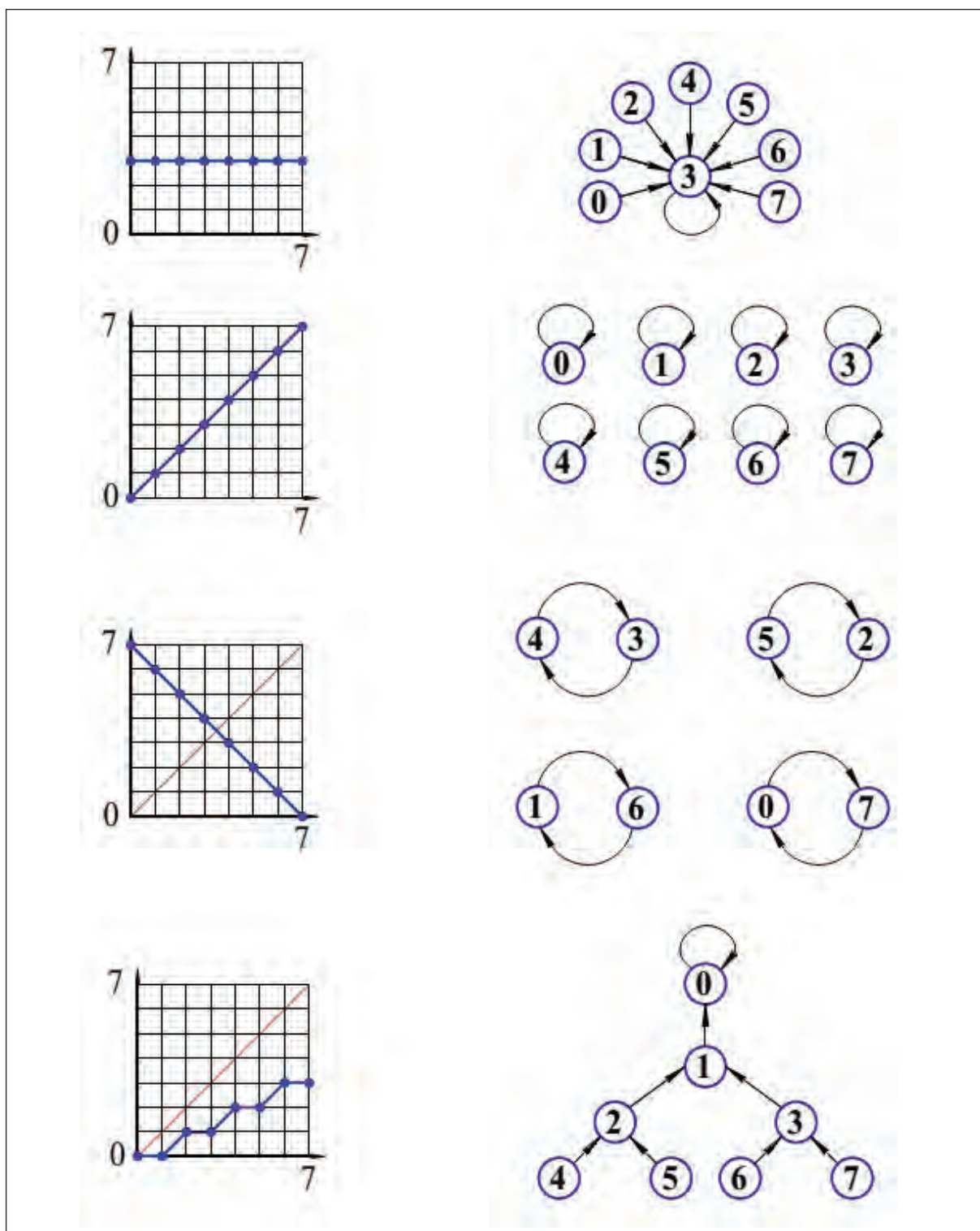


Рис. 13. Взаимосвязи графика функции и порождаемого графа G_F

показано, что динамика поведения самоопределяемых данных порождается процессом смены значений теговых кодов, которые на каждом шаге обработки проходят через функциональный преобразователь. Когда значения функции используются в качестве аргументов, выполняется операция суперпозиции, и она применяется многократно, что изображено на рис. 14 а. А поскольку в этой длинной цепи на всех шагах преобразований функции F одна и та же, мы имеем операцию автосуперпозиции. Сокращенно операцию автосуперпозиции можно изобразить в виде замкнутой схемы, приведенной на рис. 14 б.

По сути это рекуррентный генератор, порождающий кодовые последовательности. Если мы хотим понять динамику развития кодовых последовательностей, порождаемых рекуррентным генератором мы должны обратиться к графу кодовых переходов. Порождаемые рекуррентным генератором кодовые последовательности есть маршруты на графе кодовых переходов и их динамика определяется структурой графа.

А теперь можно сделать определения основных понятий дискретной динамики:

Дискретный аттрактор это динамическая система, имеющая в своём составе множество регистровых состояний R , функциональный оператор F , отображающий R в R и оператор автосуперпозиции F .

Фазовый портрет дискретного аттрактора это граф кодовых переходов G_F порождаемый функцией F .

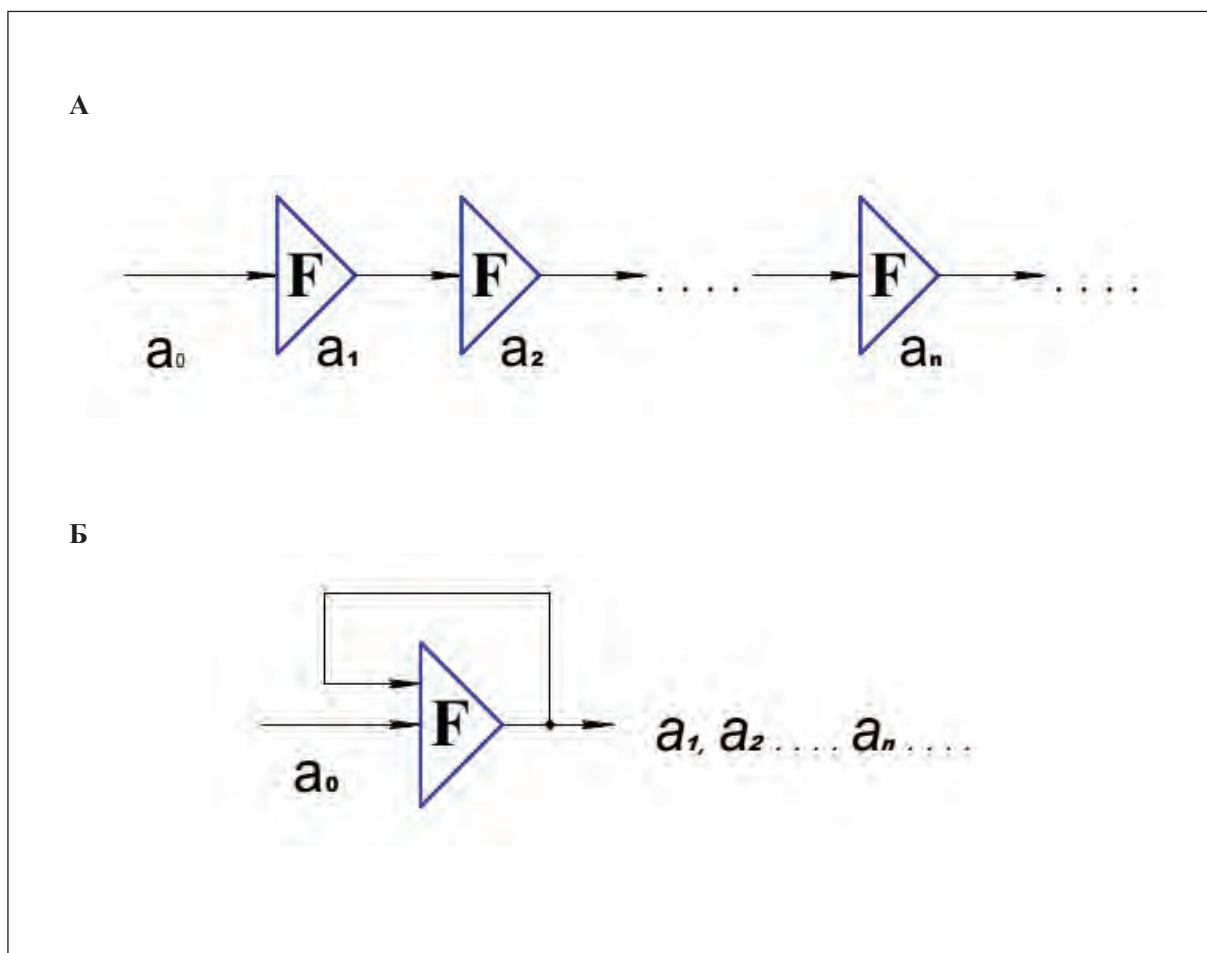


Рис. 14. Операция автосуперпозиции

Состояние совокупности самоопределяемых операндов в текущий момент фиксируется состояниями их теговых кодов. Теговые коды размещаются на вершинах графа кодовых переходов, т. е. на фазовом портрете дискретного аттрактора. Смена теговых кодов может осуществляться только переходом в смежные вершины на фазовом портрете.

Дискретный аттрактор задаёт функциональное пространство, в котором существуют самоопределяемые данные. Граф кодовых переходов формирует геометрию этого пространства как совокупность путей преобразования теговых кодов и именно таким образом исполняет роль фазового портрета аттрактора.

Силовой агрегат или движок, который осуществляет перемещения на фазовом портрете это процедура автосуперпозиции, реализованная как рекуррентный генератор. Рекуррентный генератор это процедура самообращения, подающая выходной результат вновь на вход преобразователя. Именно самообращающаяся процедура исполняет роль динамического ядра процесса.

Фазовый портрет это математическая абстракция, Граф кодовых переходов не реализуется физически, например, аппаратно как схемный лабиринт, по которому перемещаются самоопределяемые операнды или программно как сетевая структура данных. Фазовый портрет задаётся косвенным образом как результат взаимосвязанных трансформаций теговых кодов. Обширные фазовые портреты могут поддерживаться очень скромными аппаратными затратами. Если теговый код содержит 16 двоичных разрядов, а программа реализации оператора F состоит из менее чем десяти машинных команд, аттрактор поддерживает фазовый портрет в виде графа кодовых переходов, состоящего из сотен тысяч вершин. Уровень компрессии средств фиксации аттрактора в данном случае может составлять примерно 4 - 5 порядков. По этой причине появляется возможность не хранить детерминанты процесса в одном устройстве, осуществляющем сосредоточенное управление вычислениями, а разместить средства фиксации дискретного аттрактора непосредственно в каждом операнде и осуществить принцип распределённого управления.

Если все теговые сопровождения операндов обрабатываются одним преобразователем F , все они принадлежат одному фазовому портрету, а их теговые коды позиционируют всех и каждого на определённых вершинах графа кодовых переходов. В результате множество никак не связанных единым управлением операндов функционирует как строго согласованный ансамбль,двигающийся по заданным траекториям и реализующий заданный процесс. Уместно назвать этот эффект **функциональной когерентностью**.

Функциональная когерентность обеспечивается высокой степенью компрессии средств фиксации дискретного аттрактора. А компрессия возможна вследствие глубокого вырождения комбинаторики структурообразующих факторов, что означает потерю возможности задавать любые мыслимые и не мыслимые структуры фазовых портретов. В условиях компрессивного представления средств фиксации аттракторов возможные фазовые портреты образуют узкий класс структур, подчиняющихся жёстким ограничениям. Означает ли это потерю универсальности программирования любых приложений – вопрос дискуссионный. Во всяком случае, следует учитывать, что не любая последовательность команд классической машины может быть осмысленной программой.

Все кодовые траектории в дискретном аттракторе сходящиеся и завершаются попаданием в цикл на графе G_F . Если цикл состоит из одной вершины, процесс останавливается. Это точка покоя дискретного аттрактора. Если цикл содержит множество вершин, аттрактор бесконечно повторяет заданную последовательность, что можно интерпретировать как постоянное циклическое выполнение определённого действия, например сканирования и контроля текущих параметров системы управления. Дискретный аттрактор обладает свойством устойчивости. Если внешнее воздействие или повреждение выбросит его из точки покоя, аттрактор сам, в соответствии с его устройством начнёт двигаться, а любые траектории движения аттрактора есть маршруты на фазовом портрете, которые всегда сходятся к точке покоя. В точке покоя механика аттрактора не выключается. Рекуррентный генератор кодовых последовательностей устроен таким образом, что в результате автосуперпозиции на выходе бесконечно порождается один и тот же код. Для внешнего наблюдателя это выглядит как останов. На самом деле дискретный аттрактор никогда не останавливается. Останов дискретного аттрактора носит условный характер, это стабильное состояние динамического равновесия, которое можно интерпретировать как непрерывный контроль равновесного состояния.

Дискретный аттрактор как динамическая система обладает определённой спецификой, которая в первую очередь заключается в том, что при формировании динамики не используется категория времени. Динамика дискретного аттрактора основана на отношениях предшествования, и каждое следующее состояние порождается как функция от предыдущего. В простейшем случае, когда параллельный процесс развивается в рамках одного аттрактора по единому фазовому портрету, все события увязаны отношениями предшествования, определяемыми графом кодовых переходов. Но при усложнении ситуации, предполагающей взаимодействие процессов, размещённых на разных аттракторах, придётся решать проблему определения одновременности событий в функциональных пространствах, в которых размещены и двигаются самоопределяемые данные. С этой целью придётся вводить понятия меры и измерения расстояния на фазовых портретах. При этом понадобится осуществлять переход от представления процесса как функции от предыдущего состояния к его представлению как функции от номера шага или длины расстояния на фазовом портрете. Это одна из важнейших задач развития дискретной динамики.

4.4. Конкатенация

Актуальной задачей дискретной динамики является конструирование фазовых портретов с заданными свойствами или, что то же, построение графов кодовых переходов с заданной топологией. Первые шаги становления дискретной динамики сделаны чисто эмпирически. С помощью простой инструментальной программы задавались рекуррентные генераторы с определённой разрядностью регистров и определённой функцией преобразования кодов и для каждого генератора строился граф кодовых переходов G_F . Таким образом, была сформирована базовая библиотека рекуррентных генераторов, представляющая базовые конструкции. Базовые конструкции известны и изображены на рис. 13. В вольной терминологии мы можем их перечислить. Это наборы самовозвратных полюсов, цепи, деревья, кольца, и розетки, представляющие собой деревья и цепи, пристыкованные к вершинам кольца. Самоовозвратными полюсами мы называем петли или минимальные циклы, состоящие из одной вершины. При этом граф может быть связан или иметь несколько компонент связности со своими циклами.

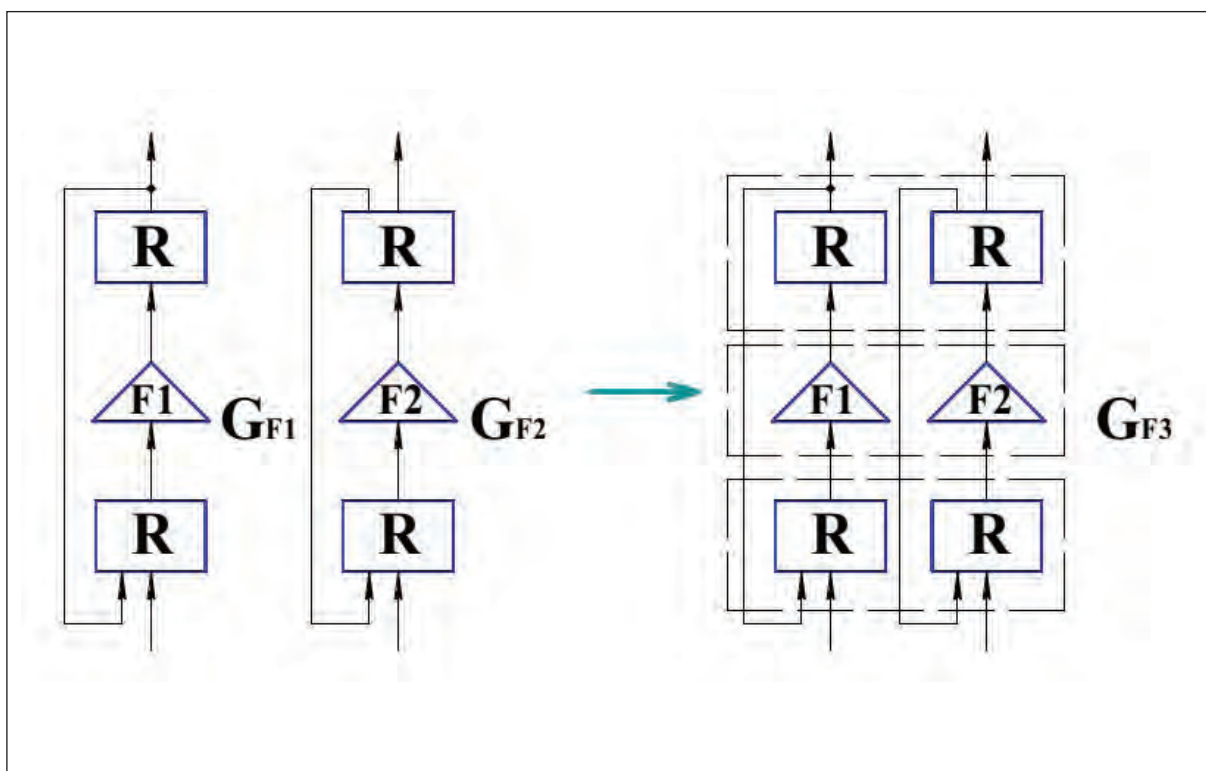


Рис. 16. Конкатенация рекуррентных генераторов

Для создания инструментальных средств дискретной динамики необходимо найти процедуру, позволяющую строить сложные конструкции фазовых портретов из простых базовых. Такая процедура существует и называется конкатенация. В наших условиях конкатенация означает совмещение двух и более регистровых секций в определённом порядке. В результате конкатенации образуется новый регистр, в котором правая секция выполняет роль младших разрядов, а левая секция роль старших разрядов объединённого регистра. Если регистры входят в состав рекуррентных генераторов и над ними построены функциональные преобразователи, в результате конкатенации образуется новый рекуррентный генератор, в котором и регистры и преобразователи образуются путём совмещения участников операции конкатенации. Иллюстрация операции конкатенации приведена на рис. 15.

Операция конкатенации объединяет два рекуррентных генератора, заданных функциями $F1$ и $F2$. Каждый из участников операции порождает свой граф кодовых переходов G_{F1} и G_{F2} . В результате конкатенации и совмещения двух исходных рекуррентных генераторов образуется новый результирующий, который порождает новый граф кодовых переходов G_{F3} . Структура нового графа может существенным образом отличаться от структур участников конкатенации, а примитивная инженерная операция совмещения регистровых секций с математической точки зрения реализует нетривиальную операцию над графами. Далее приведены некоторые примеры операций над графами, возникающими при конкатенации рекуррентных генераторов.

Так, например, при конкатенации двух бинарных деревьев получается тернарное дерево. При этом максимальная глубина дерева сохраняется, а кратность ветвления умножается. Результаты конкатенации двух деревьев изображены на рис. 16.

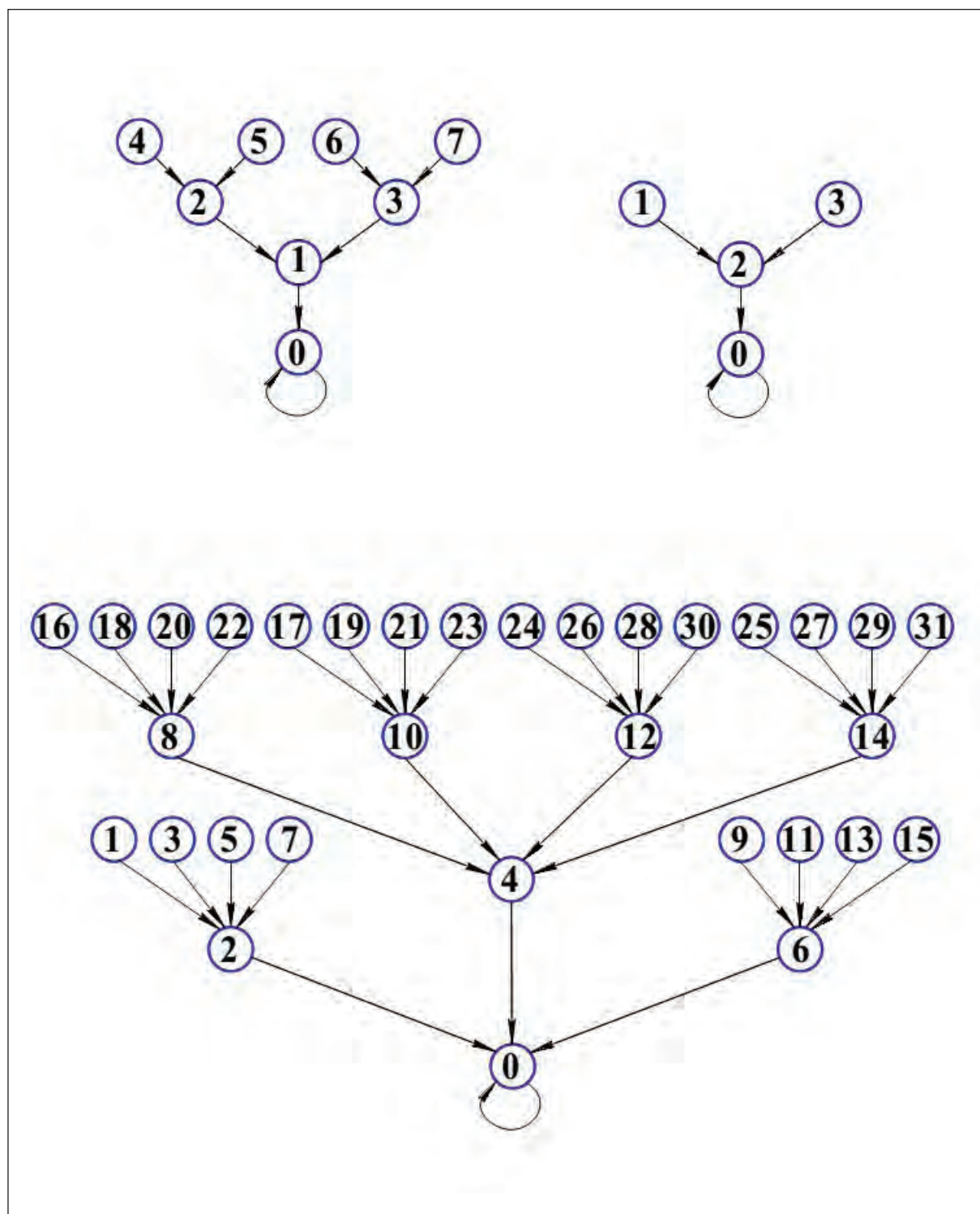


Рис. 16. Конкатенация двух деревьев

При конкатенации дерева и набора полюсов получается набор деревьев, т.е. связный граф превращается в многосвязный и число компонент связности равно числу полюсов в одном из участников конкатенации. Результаты приведены на рис. 17.

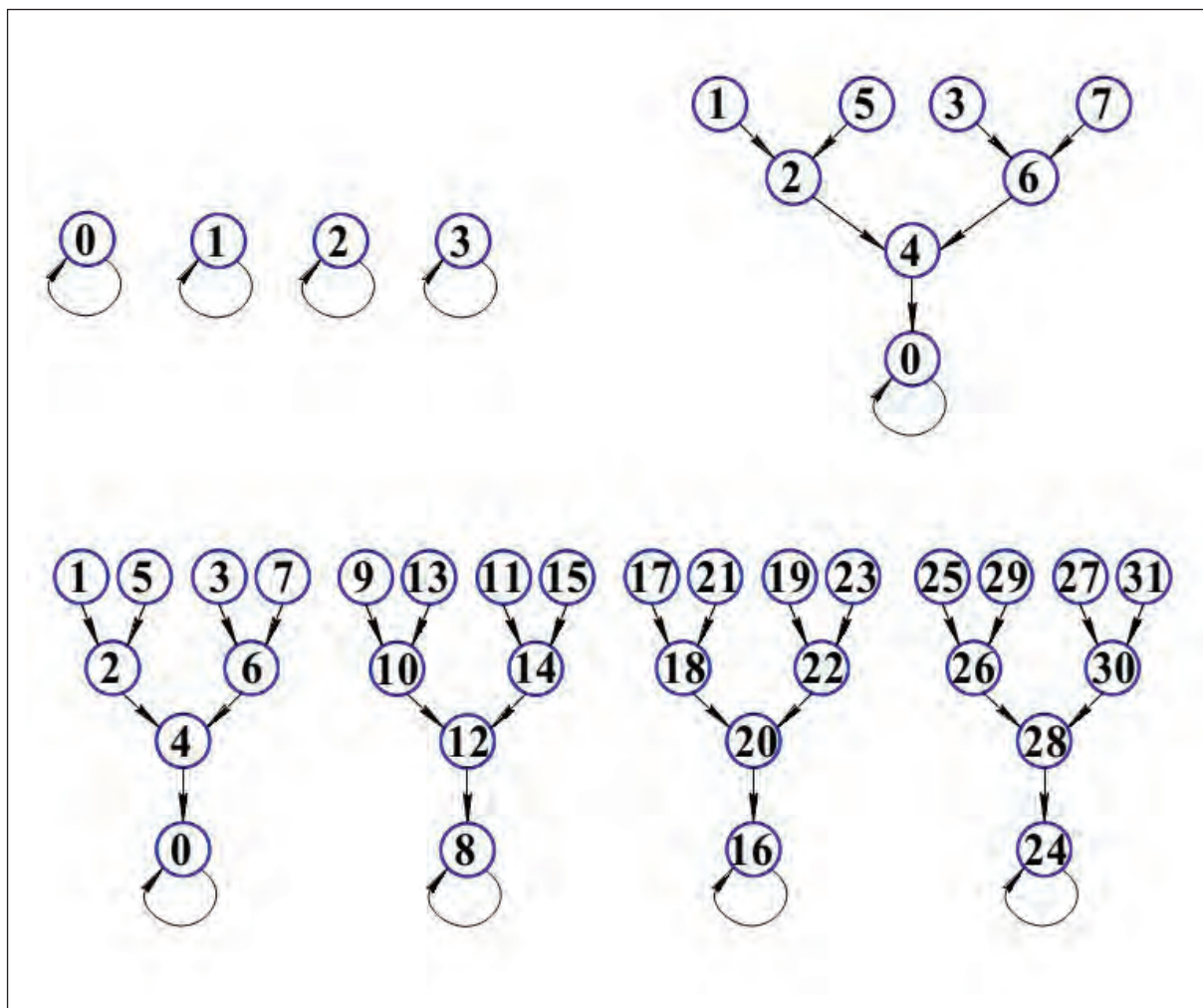


Рис. 1. Схема объединения узлов коммутаторами в топологии «толстое дерево».

При конкатенации кольца и дерева, происходит образование розетки, представляющей собой набор деревьев пристыкованных к вершинам кольца.

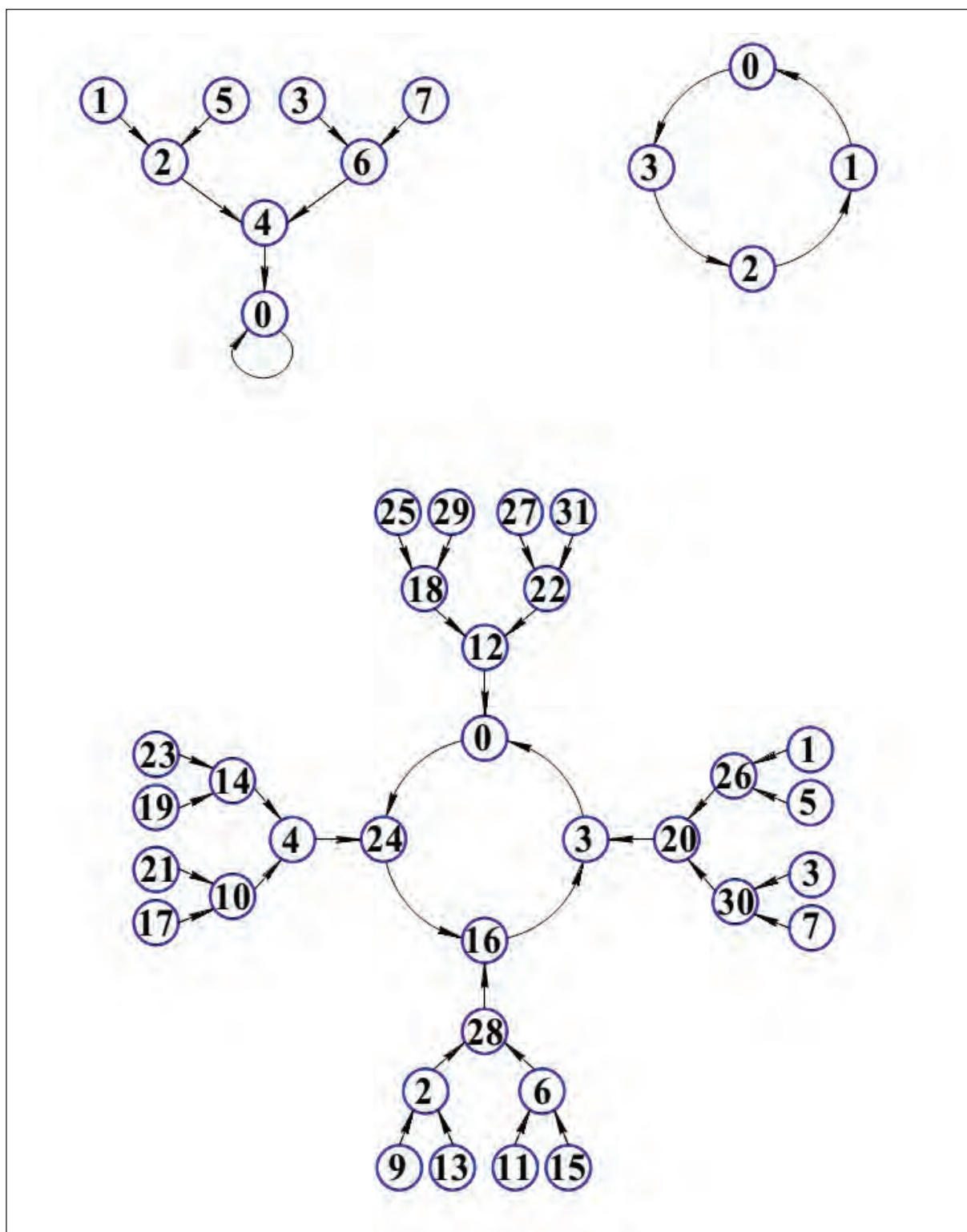


Рис. 18. Конкатенация дерева и кольца

Эмпирически удалось сформировать базу данных по конкатенации разных вариантов графов кодовых переходов и выделить набор практических приёмов конструирования структур, необходимых для решения ряда практических задач. Дальнейшая разработка этого направления должна привести к обоснованию свойств операций конкатенации и построению алгебры операций над графами кодовых переходов.

4.5. Репеллер и реверсивный аттрактор

До сих пор мы рассматривали варианты использования графа кодовых переходов как динамического ядра дискретного аттрактора, в котором продвижение по фазовому портрету происходит в направлении от листьев к корню дерева и циклу, которым завершается компонента связности. Для практических приложений дискретной динамики важно также обеспечить возможность продвижения по фазовому портрету в обратном направлении – от корня дерева к листьям. Эта задача не может решаться средствами функционального оператора, поскольку обратное отображение является многозначным и должно поддерживать ветвящийся процесс. Для поддержки ветвящегося процесса можно использовать гирлянду функциональных преобразователей. Гирлянда изображена на рис. 19.

Для примера выбрана гирлянда из трёх функциональных преобразователей, их число может меняться. Принцип действия гирлянды следует из её схемного изображения. При подаче на вход некоторого начального значения кода A_0 гирлянда откликается тремя выходными значениями A_1 , A_2 , A_3 . Далее каждое из них может подаваться на вход гирлянды и порождать на выходах следующие три значения. В данном случае рекуррентный генератор, построенный на базе гирлянды, порождает ветвящийся процесс, который носит расходящийся характер. Таким образом, вводится новый базовый элемент дискретной динамики - **дискретный репеллер**. **Дискретный репеллер** это расходящийся процесс, порождаемый гирляндой функциональных преобразователей и операцией автосуперпозиции гирлянды. Фазовый портрет репеллера это расходящийся граф кодовых переходов гирлянды. Репеллер по определению не имеет остановки и порождает бесконечный расходящийся процесс. Вследствие конечности набора кодов, на котором определены функциональные преобразователи гирлянды рано или поздно запас кодов исчерпывается и на выходе гирлянды появляется код, который ранее уже имел вхождение в вершины графа кодовых переходов. Это означает, что фрагмент дерева, начинающийся с данной вершины, будет воспроизведен репеллером вновь.

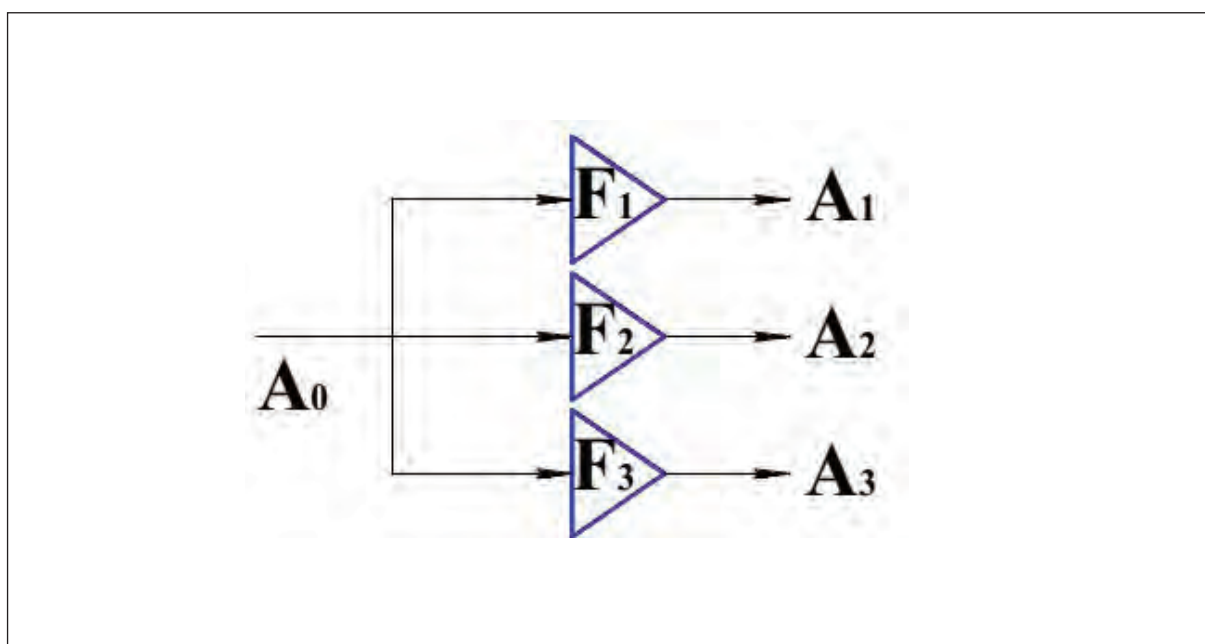


Рис. 19. Гирлянда образующая дискретный репеллер

В целом фазовый портрет репеллера будет состоять из бесконечного повторения конечного набора фрагментов, графа кодовых переходов, заданного гирляндой. Исследование динамики развития процессов в репеллерах это следующий и очень содержательный раздел дискретной динамики.

Опираясь на понятия дискретный аттрактор и дискретный репеллер можно сформулировать важное для приложений дискретной динамики понятие **реверсивный аттрактор**. Допустим, что задан дискретный аттрактор с определенным фазовым портретом. Для построения реверсивного аттрактора необходимо синтезировать репеллер, фазовый портрет которого совпадает с фазовым портретом аттрактора с точностью до инверсии направления рёбер графа кодовых переходов. Реверсивный аттрактор это сопряжённая пара аттрактора и репеллера, обеспечивающая продвижение процесса по заданному фазовому портрету в прямом и обратном направлениях. Задача эта решается следующим образом. Исходный аттрактор задаётся рекуррентным генератором на базе определённой целочисленной функции. График функции представляется как совокупность точек на целочисленной квадратной решётке. График изображён на рис. 20 а.

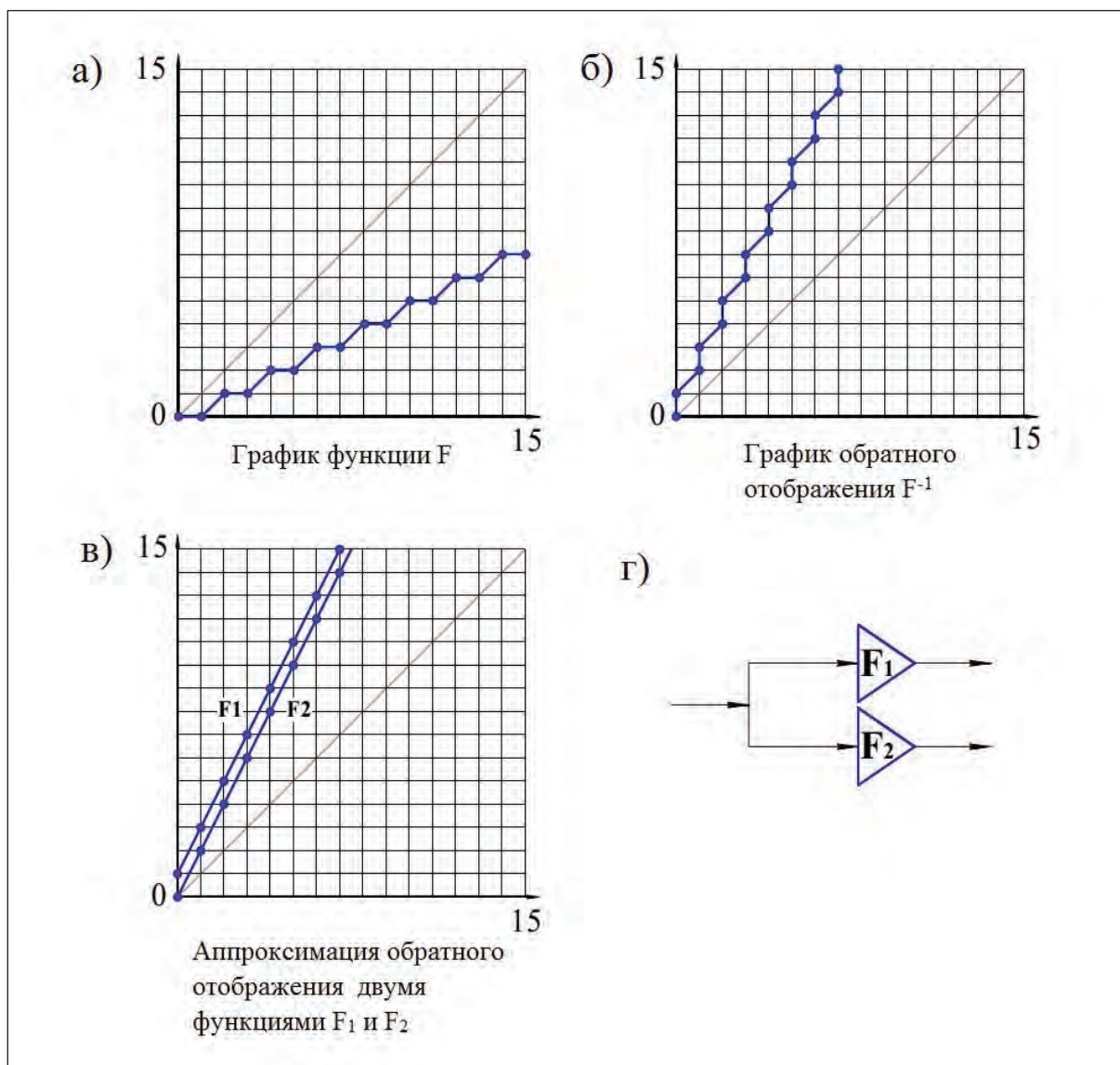


Рис. 1. Схема объединения узлов коммутаторами в топологии «толстое дерево».

По данному графику необходимо построить обратное отображение. Для этого надо выполнить операцию замены координат, что достигается поворотом исходного графика вокруг главной диагонали. График обратного отображения изображён на рис. 20 б. Обратное отображение не является функциональным и далее надо решить задачу синтеза набора сопряжённых функций, покрывающих, полученный график обратного отображения. Перечисленные действия изображены на рис. 20.

Эта задача не имеет общего решения. В каждом конкретном случае путём подбора можно найти множество вариантов покрытия графика обратного отображения функциональными графиками и искать среди них приемлемые по критерию удобства реализации сопряжённых функций. Из полученного таким образом набора сопряжённых функций составляется гирлянда для реализации репеллера, дополнительного к заданному аттрактору.

4.6. Маскирование и сильное ветвление

Рассмотренные до сих пор средства формирования графов кодовых переходов предполагают, что каждое регистровое состояние из множества, образующего структуру имеет однократное вхождение в граф кодовых переходов. В практике применения дискретных аттракторов для формирования вычислительных работ необходимо обеспечить многократную повторяемость кодов операций на графе вычислительного процесса. Это достигается путём наложения маски на образующие регистры и чтения из под маски только части разрядов, необходимых для кодирования операций. Рассмотрим конкретный пример. Необходимо построить граф кодовых переходов со структурой сильно ветвящегося дерева с кратностью ветвления равной 8 и заданной глубиной 4 уровня. При этом ставится задача обеспечить возможность полной нумерации предшественников каждой вершины. Выберем определённую вершину и рассмотрим набор из восьми вершин, являющихся предшественниками данной. Коды вершин предшественников должны быть устроены таким образом, чтобы при наложении на них определённой маски можно было извлечь три разряда, задающих полную нумерацию предшественников от 0 до 7, или в двоичном выражении от комбинации 000 до 111. По принятому условию это свойство должно выполняться для всех вершин кроме листовых, которые не имеют предшественников.

Искомый рекуррентный генератор, поддерживающий заданное сильно ветвящееся дерево будет сформирован в результате конкатенации трёх генераторов, поддерживающих четырёхэтажные бинарные деревья и построенные на базе уже известной нам функции сдвига вправо на один разряд. По известным нам свойствам операции конкатенации бинарных деревьев глубина сохраняется равной 4, а кратности ветвления перемножаются и дают требуемое значение 8. В итоге получается полное четырёхэтажное дерево с кратностью ветвления равной 8. А теперь наложим на образующие регистры маску, которая позволяет извлечь из каждой секции участвующей в конкатенации один младший крайний правый разряд. Трёхразрядный код, извлечённый в результате маскирования, содержит полную нумерацию вершин предшественников. Остальные разряды полного кода позиционируют вершины на дереве в целом. На рис. 21 приведен фрагмент сильно ветвящегося дерева с реальными значениями кодов вершин и показано извлечение полной нумерации вершин предшественников.

вым портретом в виде сильно ветвящегося дерева, в котором содержатся бинарные поддеревья вычисления всех возможных арифметических выражений, не превышающих определённый уровень сложности, ограниченный глубиной дерева вычисления. Работа аттрактора запускается загрузкой в вычислитель набора исходных операндов. Теговые коды исходных операндов позиционируют начальное состояние процесса на фазовом портрете. Дальнейшие события разворачиваются детерминировано и определяются конструкцией аттрактора и структурой фазового портрета. Правильно загруженный набор исходных операндов извлекает из фазового портрета своё бинарное поддерево. Следовательно, для программирования заданного вычисления необходимо отыскать, соответствующие ему листовые вершины и присвоить их коды исходным операндам в виде теговых сопровождений.

Для отыскания листовых вершин заданного вычисления необходимо наложить заданное бинарное дерево на общее сильно ветвящееся. Предполагается, что нумерация вершин предшественников увязана в таблицу кодирования арифметических операций. Размещение начинается с корневой вершины. Выбирается заключительная операция заданного вычисления и ставится маркер на вершину предшественницу корня, номер которой соответствует заключительной операции. Два операнда, участвующие в заключительной операции порождаются выполнением двух предшествующих арифметических операций, что обнаруживается при раскрытии скобок в направлении от главной заключительной операции. На следующем шаге необходимо поставить маркерные метки на двух вершинах, предшествующих отмаркированной на первом шаге. При этом следует отбирать вершины с номером, соответствующим кодам предшествующих арифметических операций. Если бы предшествующие операции всегда были разными, можно было бы ограничиться ветвлением общего дерева с кратностью равной четырём. Но в общем случае обе предшествующие операции могут быть одинаковыми. По этой причине в номерах предшествующих вершин каждая операция должна быть представлена дважды. Вот почему при четырёх операциях понадобилось дерево с кратностью ветвления равной восьми. Описанная процедура повторяется до исчерпания операций в исходной записи арифметического выражения. Последние маркерные метки отмечают листовые вершины бинарного дерева искомого вычислительного процесса. Результат программирования в данном случае это назначение теговых кодов исходным операндам. При загрузке в вычислитель исходные операнды запускают вычислительный процесс и извлекают требуемое бинарное поддерево путём прохождения своих маршрутов на фазовом портрете.

Разработанные ранее инструментальные средства дискретной динамики позволяют сконструировать компилятор, осуществляющий программирование процессов вычисления арифметических выражений в архитектуре самоопределяемых данных. Для этого есть инструменты формирования сильно ветвящихся деревьев, есть инструменты маскирования кодов для организации полной нумерации вершин предшественников, есть методы построения реверсивных аттракторов, обеспечивающих движение по фазовому портрету в прямом и обратном направлениях. Описанный компилятор можно сделать полностью автоматическим.

Поддержка полного сильно ветвящегося дерева необходима для работы компилятора, который должен содержать пути для вычисления всех возможных арифметических выражений. После выбора заданного выражения и завершения компиляции поддержка полных теговых кодов при исполнении процесса становится избыточной. Возможна оптимизация, которая означает сокращение разрядности тегов. Сокращение раз-

рядности теговых кодов должно происходить при условии, что общая конструкция графа кодовых переходов не нарушается, выбранное бинарное поддереву сохраняется, а лишние ветви полного сильно ветвящегося дерева отсекаются. Такая задача может быть поставлена и есть предположения по её решению. Задача не имеет общего решения, это означает, что в каждом конкретном случае эффективность сокращения разрядности теговых сопровождений будет различной.

Заключение

Достигнутый уровень технологии микроэлектронного производства обусловил основные направления развития вычислительных средств на ближайшее десятилетие. В соответствии с законом Мура отрасль будет рентабельна при условии соблюдения определённых темпов роста производительности вычислительных средств и выполнения определённых технологических ограничений, в частности соблюдением заданных показателей энергоэффективности. В целом это означает, что главным инструментом роста производительности становится освоение массового параллелизма вычислений. В данной работе проведено исследование динамики роста производительности высокопараллельных структур и выявлен целый ряд парадоксальных явлений. На примере конкретной, но типичной задачи показано, что максимум производительности достигается в диапазоне нескольких сотен процессорных элементов, а в области высоких значений параллелизма (десятков и сотен тысяч) производительность катастрофически падает. В диапазоне пика производительности эффективность использования потенциала параллелизма не превышает 10 %. Определяющим фактором при этом является архитектурная идея. Принятые формы организации вычислений базируются на разделении зоны хранения и зоны обработки данных, которые строятся по различным технологиям. Анализ баланса временных затрат показывает, что львиная доля приходится на пересылки данных из зон хранения в зоны обработки и обратно. Это определяет показатель эффективности процессорного элемента, который не превышает 1%. Выражаясь метафорически на одно полезное действие, собственно обработки данных, приходится не менее 100 действий, обеспечивающих доступ в память и пересылки данных. Попытка увеличения производительности вычислений за счёт параллелизма и использования множества процессорных элементов требует разрезки информационной структуры алгоритма и размещения её в разных процессорных элементах. А это приводит к тому, что часть пересылок обретает статус межпроцессорных обменов и требует особой коммутационной среды. По мере увеличения параллелизма информационная структура алгоритма дробится на всё более мелкие фрагменты, и всё более и более крупные объёмы пересылок становятся межпроцессорными. Для больших значений параллелизма баланс временных затрат обретает гипертрофированный характер. Доля затрат на вычисления становится исчезающе малой, а доля обменных операций становится быстро растущей доминантой и никакие усложнения и ускорения коммутационной среды не могут переломить эту динамику. Увеличение степени параллелизма при больших значениях приводит к устойчивому падению производительности.

Далее есть два пути - сохранение классической архитектуры и поиск задач большой размерности с малой связностью данных и отказ от классической архитектуры и разработка новых форм организации вычислений и новых технологий программирования. Каждое из этих двух направлений займёт свою нишу в широком спектре применений. В данной работе излагается пример разработки принципиально новой архитектурной

идеи, основанным положением которой является совмещение зон хранения и обработки данных. Архитектура самоопределяемых данных поддерживает массовый мелкозернистый параллелизм на уровне обработки отдельных операндов и при этом не порождает крупные массивы межпроцессорных обменов. Принципы, заложенные в архитектуре самоопределяемых данных, создают предпосылки для создания новых технологий программирования, основанных на применении специальных математических инструментов. В статье даётся краткий очерк основ дискретной динамики и приводится пример использования основных понятий дискретной динамики для моделирования и программирования вычислений в системе самоопределяемых данных.

Литература

1. Сборник трудов 7-й Международной конференции по физике и технологии гетероструктурной СВЧ-электроники 25 мая 2016 г. // Москва МИФИ, 2016, 169 стр.
2. *Duato J., Yalamanchili S., Ni L. M.* Interconnection networks: an engineering approach. // Morgan Kaufmann, 2003, 600 pp.
3. *Стегайлов В. В., Норман Г. Э.* Проблемы развития суперкомпьютерной отрасли в России: взгляд пользователя высокопроизводительных систем // Программные системы: теория и приложения, 2014, 1 (1), 111-152.
4. *Норман Г. Э. и др.* Зачем и какие суперкомпьютеры эксафлопсного класса нужны в естественных науках // Программные системы: теория и приложения, 2015, 6 (4), 243-311.
5. *Kim J., Dally W. J., Abts D.* Flattened butterfly: a cost-efficient topology for high-radix networks // ACM SIGARCH Computer Architecture News, 2007, 35 (2), 126-137.
6. *Yadav S., Mishra R., Gupta U.* Performance evaluation of different versions of 2D Torus network // Computer Engineering and Applications (ICACEA), 2015 International Conference on Advances in IEEE, 2015, 178-182.
7. Resources for Embedded and IoT Designers // URL: <http://www.intel.com/design/network/products/optical/cables/index.htm>.
8. *Dally W. J., Poulton J. W.* Digital systems engineering // Cambridge University Press, New York, NY, 1998, 663 pp.
9. *Махиборода А.В., Ильичёв А.В., Подобин А.А.* Проблемы реализации массового динамического параллелизма I // Наноструктуры математическая физика и моделирование, 2016, 14 (1), 47-67.
10. *Dennis J. B.* First version of a data flow procedure language // Programming Symposium, Springer Berlin Heidelberg, 1974, 362-376.
11. *Дурнев В. Г., Стандрик В. Д.* Основы построения систем передачи ЕАСС. // М.: Радио и связь, 1985, 208 стр.
12. *Мизин И. А., Богатырев В. А., Кулешов А. П.* Сети коммутации пакетов. //М.: Радио и связь, 1986. 408 стр.
13. *Махиборода А.В. и др.* Система потоковой обработки информации с интерпретацией функциональных языков // Авторское свидетельство № 1697084 опубликовано 29. 06. 1989 г. Положительное решение 25.05 1990 г.
14. *Мизин И. А., Махиборода А. В.* Концепция самоопределяемых данных и архитектура распределённых систем // Информационные технологии и вычислительные системы, Москва, Наука, 1995, 1, 21-31.
15. *Махиборода А. В.* Проблемы создания вычислительных средств с непроцедурными внутренними языками // Управляющие системы и машины, 1993, 3, 51-62.
16. *Арнольд В. И.* Топология и статистика формул арифметики // Успехи математических наук, 2003, 58, 4 (352), 3-28.
17. *Арнольд В. И.* Топология алгебры: комбинаторика операции возведения в квадрат // Функциональный анализ и его приложения, 2003, 37 (3), 20-35.
18. *Арнольд В. И.* Сложность конечных последовательностей нулей и единиц и геометрия конечных функциональных пространств //Публичная лекция, 2006, 13 стр.

PROBLEMS OF MASSIVE DYNAMIC PARALLELISM IMPLEMENTATION. II

A.V. Makhiboroda, A.V. Ilichev, A.A. Podobin, A.V. Tsarev

*Department of Applied Mathematics MIEM,
National Research University HSE*

makhiboroda@yandex.ru

Received 17.05.2016

In the first part of this paper, it is noticed that the development of massive dynamic parallel computations becomes the main tool for ensuring the stable growth of the processing power of computing facilities. It is also shown that the implementation of massive parallel computations meets two major obstacles. These are specific costs of the parallelism support which grow faster than the number of processors and a poor performance of the interconnection networks supporting the interprocessor communications.

In the second part of this paper, the forecasts of the interconnection network development are evaluated from the standpoint of improving its algorithms and increasing its bandwidth. The expert opinions about the possibilities of extending the effective parallelism range to the value of hundreds of processors are given. At the same time, systems with several thousands of processors on a chip have already been constructed, and the current state of the art in these technologies allows predicting an increase in the number of processors to more than tens of thousands in the next few years. Searching the ways to extend the effective implementation range of parallelism is a rather urgent problem directed to develop principally new forms of algorithms. The basic ideas of self-defined data architecture are discussed and a brief outline of the construction of discrete dynamics mathematical formalism is given. The conceptual apparatus of discrete dynamics allows formulating and solving the problem of programming the processes in self-defined data architecture.

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ДИНАМИКИ ЗАСЕЛЕННОСТИ СОСТОЯНИЙ ТРЕХУРОВНЕВОЙ НАНОЧАСТИЦЫ ПРИ СПЕКТРОСКОПИЧЕСКИХ ПЕРЕХОДАХ

В.А. Морозов

Институт органической химии им. Н.Д. Зелинского Российской академии наук

morozov@mail.ioc.ac.ru

Поступила 18.05.2016

Проведено математическое моделирование динамики заселенности состояний изолированной трёхуровневой наночастицы при преобразовании ею длительного импульса монохроматического облучения, а также при спонтанном излучении из состояния с наибольшим значением энергии. Использовались два подхода к проведению такого моделирования, один из которых основан на использовании численных решений оптических уравнений Блоха для частицы, а другой – решений уравнения Шредингера для амплитуд вероятности заселенности состояний составной системы из частицы, квантованного поля облучения и вторичного излучения. Полученные результаты моделирования динамики для случая спонтанного излучения согласуются при применении как первого, так и второго из этих подходов, но существенно различаются для случая преобразования частицей длительного импульса облучения. Приведено объяснение происхождения таких различий.

УДК 535.14

Введение

Трехуровневую модель отдельной наночастицы (изолированного иона, хромофорной группы молекулы, полупроводниковой квантовой точки) часто используют при моделировании динамики заселенности ее состояний. К первым по времени из работ, проводивших такое моделирование при описании результатов детектирования вторичного излучения от одного иона бария или ртути в магнитной ловушке, можно отнести работы [1–3]. Моделирование динамики заселенности трех выделенных состояний иона, одно из которых метастабильно, в этих работах было проведено на основе ис-

пользования решений оптических уравнений Блоха (ОУБ). Эти решения, дополненные квантовыми представлениями о поле излучения, использовались для объяснения происхождения наблюдавшихся нерегулярных спадов и подъемов (от максимального значения до полного исчезновения) значений интенсивности резонансной флуоресценции. Для более последовательного и полного описания этого эффекта во многих последующих работах сами ОУБ модифицировались так, чтобы отражать проявления этих скачков интенсивности соответствующей динамикой заселенности состояний иона и объяснять, например, отсутствие отсчета при детектировании вторичного излучения (“нулевое измерение”) и эффект антигруппировки фотонов вторичного излучения. Использование таких модифицированных уравнений Блоха при рассмотрении динамики заселенности состояний частицы в настоящее время относят к общему методу описания преобразования излучения наночастицей, который получил название “метод квантовых скачков” (список работ см., например, в [4–6]).

Достигнутый к настоящему времени прогресс в технике фотодетектирования излучения позволяет получать данные временной развёртки вторичного излучения одиночных молекул, изолированных молекулярных комплексов и других “квантовых точек” (см., например, в [7]). При моделировании динамики заселенности состояний таких наночастиц также часто используют трехуровневую модель, динамику заселенности состояний которой описывают на основе того или иного варианта ОУБ, не переходя к тем их модификациям, о которых упомянуто выше. Например, при описании мониторинга фотохимической окислительно-восстановительной реакции отдельной молекулы флавинадениндинуклеотид в холестероле [8]; закономерностей мерцания флуоресценции сложной молекулы с аббревиатурой DiI [9]; кинетики резонансной люминесценции одиночной полупроводниковой квантовой точки [10]; наведенного тепловым резервуаром эффекта антигруппировки, сжатых состояний и усиления света во вторичном излучении квантовой точки, содержащей две пары электрон-дырка [11]; управления оптической бистабильностью молекулы, содержащей две квантовых точки [12]. В этих работах отмечается, например, что использование результатов моделирования динамики фотохимических реакций в терминах заселения состояний одной молекулы при обсуждении динамики реакции намного плодотворнее обычных рассмотрений химической кинетики, применяемых для ансамблей молекул, а моделирование динамики заселенности состояний полупроводниковых квантовых точек дает возможность оценки параметров релаксационных процессов, знание которых может быть использовано при создании нового типа наноэлектронных устройств.

Следует, однако, иметь в виду, что формализм обычных ОУБ содержит существенные допущения и приближения (см., например, в [13]) относительно общего представления и деталей описания взаимодействия света с частицей по сравнению с формализмом квантовой теории излучения (КТИ), использующим строгое – относительно физического смысла – представление о процессе заселения состояний частицы и соответствующее строгое его описание, основанное на решении уравнения Шредингера для составной системы из частицы и квантованного поля излучения. В отличие от формализма ОУБ, который рассматривает молекулу как замкнутую квантовую систему и не учитывает изменения состояния облучения и появление вторичного излучения, формализм КТИ рассматривает молекулу как открытую квантовую систему, описывает изменение состояния поля облучения и появление вторичного излучения, учет поглощения которого макроскопическим окружением (“детектирование” или “регистрация” фото-

нов вторичного излучения) существенно влияет на динамику заселения состояний частицы (по [14] – “динамику спектроскопических переходов”). При таких различиях формализмов ОУБ и КТИ представляет интерес иметь сопоставление результатов применения этих двух подходов для моделирования динамики заселенности состояний частицы.

Сопоставление результатов ОУБ- и КТИ-моделирования динамики заселенности состояний трехуровневой изолированной квантовой частицы, одно из невырожденных возбужденных состояний которой не распадается (“темное” состояние), проведено в [15, 16] для случая монохроматического резонансного облучения частицы в основном состоянии. Установлено существенное различие результатов при использовании таких подходов. Выявлены характерные черты установленных различий и предложено объяснение их происхождения как следствия отмеченных выше различий в представлениях о преобразовании света частицей, принятых в формализмах применявшихся подходов.

В настоящей работе приведены результаты аналогичного сопоставления ОУБ- и КТИ-моделирования динамики заселенности состояний возбуждаемой светом 3-х уровневой частицы с учетом радиационного распада двух невырожденных ее возбужденных состояний как при преобразовании монохроматического резонансного облучения, так и при спонтанном излучении частицей из состояния с наибольшей энергией, в котором она оказывается после действия короткого импульса света.

Моделирование на основе оптических уравнений Блоха

Будем полагать, что значения энергии E_n стационарных состояний модели частицы $n = 1, 2, 3$ таковы, что $E_3 > E_2 \gg E_1$ или $E_3 \gg E_2 > E_1$. Пример одного из таких относительных расположений уровней E_n приведен на рис. 1 в произвольных единицах с указанием учитываемых переходов частицы между соответствующими состояниями.

При использовании формализма ОУБ динамика вероятности заселённости m -того состояния изолированной частицы определяется функцией $\rho_{mm}(t)$, являющейся решением уравнения Лиувилля для $\dot{\rho}_{mn}(t)$ с добавлением слагаемых, учитывающих релаксационный распад этих состояний (см., например, в главе 5 монографии [5]).

Для определенности полагаем, что частица, расположенная в центре декартовой системы координат (X, Y, Z) , облучается импульсом монохроматического света, распространяющимся вдоль оси Z и поляризованном вдоль оси X , напряжённость электрического поля которого в месте расположения частицы $\mathbf{e}(t) = 2e_0 \Theta(t) \Theta(T - t) \cdot \sin \omega_L t$, $\Theta(t) = 1$ при $t \geq 0$ и $\Theta(t) = 0$ при $t < 0$, T – длительность импульса облучения, Оператор энергии взаимодействия между частицей и полем излучения имеет вид $V = -(\mathbf{e}\mathbf{d})$, где $\mathbf{d}$ – оператор дипольного момента частицы. Ограничимся рассмотрением облучения светом с частотой ω_L , равной частоте резонансного перехода $\omega_{31} = \hbar^{-1}(E_3 - E_1)$ и, имея в виду использование справедливого при этом резонансного приближения для системы ОУБ, матричный элемент оператора V по состояниям частицы 1 и 3 будем использовать в виде $V_{13} = i\hbar\Omega_R \cdot \exp(i\omega_L t)$, где $\Omega_R = \hbar^{-1}(\mathbf{e}_0 \mathbf{d}_{13})$ – частота Раби, $\mathbf{d}_{13}$ – матричный элемент оператора $\mathbf{d}$ по состояниям частицы 1 и 3.

При принятом резонансном облучении в системе ОУБ для рассматриваемой частицы можно пренебречь слагаемыми, пропорциональными матричным элементам V_{12} и V_{23} . Используя такое приближение и не учитывая “чисто фазовую релаксацию”, для

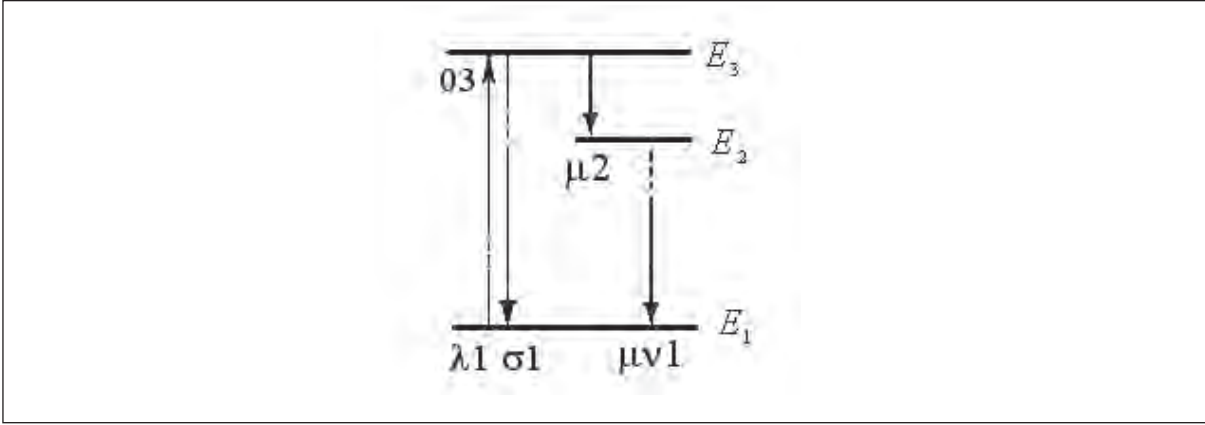


Рис. 1. Схема расположения уровней энергии трёхуровневой модели частицы с указанием состояний составной системы из частицы, поля излучения и учитываемых переходов между этими состояниями.

изучения динамики заселенности состояний принятой модели частицы ограничимся использованием решений следующей системы ОУБ:

$$\begin{aligned}\dot{\rho}_{11} &= 2\gamma_{3\rightarrow 1}\rho_{33} + 2\gamma_2\rho_{22} + 2\rho_{31}\cdot\Omega_R, \\ \dot{\rho}_{22} &= 2\gamma_{3\rightarrow 2}\rho_{33} - 2\gamma_2\rho_{22}, \\ \dot{\rho}_{33} &= -2\gamma_3\rho_{33} - 2\rho_{31}\cdot\Omega_R, \\ \dot{\rho}_{31} &= -\gamma_3\cdot\rho_{31} + (\rho_{33} - \rho_{11})\cdot\Omega_R = \dot{\rho}_{13},\end{aligned}$$

где $\gamma_{3\rightarrow 1}$, $(\gamma_{3\rightarrow 2})$ – константа скорости радиационного распада состояния 3 с переходом частицы в состояние 1, (в состояние 2); γ_2 – константа скорости радиационного распада состояния 2 с переходом частицы в состояние 1; $\gamma_{3\rightarrow 1} + \gamma_{3\rightarrow 2} \equiv \gamma_3$. Будем полагать, что $\gamma_3 \ll \omega_{31}$ и $\gamma_2 \ll \omega_{21} \equiv \hbar^{-1}(E_2 - E_1)$. Заметим, что интенсивность линии резонансного РР рассматриваемой частицы пропорциональна значению константы $\gamma_{3\rightarrow 1}$, а соответствующего КР – константы $\gamma_{3\rightarrow 2}$. (см., например, [17]).

Форма записи уравнений для $\dot{\rho}_{mn}(t)$ выбрана так, чтобы значения введенных констант затухания состояний частицы совпадали с соответствующими значениями радиационных скоростей переходов частицы при описании динамики состояний составной системы из частицы и поля излучения в том смысле, что для спонтанного излучения частицы из состояния 3 было бы справедливо равенство $\rho_{33}(t) = \exp(-2\gamma_3 t) = |b_{03}(t)|^2$, где $b_{03}(t)$ – амплитуда состояния 03 составной системы (см. следующий раздел).

1. Динамика заселенности состояний при спонтанном излучении.

Для случая спонтанного излучения частицы, оказавшейся в состоянии 3 после взаимодействия с импульсом облучения, длительность которого много меньше $\tau \equiv \gamma_3^{-1}$ (“времени жизни” состояния 3), в приведенной системе ОУБ полагаем $\Omega_R = 0$. Решение системы получившихся при этом уравнений для $\gamma_3 \neq \gamma_2$ имеет вид:

$$\begin{aligned}\rho_{33}(t) &= \exp(-2\gamma_3 t), \\ \rho_{22}(t) &= \gamma_{3\rightarrow 2} [\exp(-2\gamma_3 t) - \exp(-2\gamma_2 t)] / (\gamma_2 - \gamma_3), \\ \rho_{11}(t) &= 1 + [\gamma_{3\rightarrow 2} \exp(-2\gamma_2 t) - (\gamma_2 - \gamma_{3\rightarrow 1}) \exp(-2\gamma_3 t)] / (\gamma_2 - \gamma_3),\end{aligned}$$

а при $\gamma_3 = \gamma_2 \equiv \gamma$ имеет вид:

$$\begin{aligned}\rho_{33}(t) &= \exp(-2\gamma t), \\ \rho_{22}(t) &= 2\gamma t \exp(-2\gamma t), \\ \rho_{11}(t) &= 1 - (1 + 2\gamma t) \exp(-2\gamma t).\end{aligned}$$

2. Динамика заселенности состояний при преобразовании монохроматического облучения.

Для случая облучения частицы в состоянии 1 импульсом света, длительность которого много больше τ , в пределе $T \rightarrow \infty$ приведенная система ОУБ имеет стационарное решение:

$$\rho_{11} = \gamma_2(\Omega_R^2 + \gamma_3^2) \cdot A, \quad \rho_{22} = \gamma_{3 \rightarrow 2} \Omega_R^2 \cdot A, \quad \rho_{33} = \gamma_2 \Omega_R^2 \cdot A, \quad \text{где}$$

$$A^{-1} = \Omega_R^2(\gamma_{3 \rightarrow 2} + 2\gamma_2) + \gamma_2 \gamma_3^2.$$

Ограничимся рассмотрением облучения такой интенсивности, которому соответствует значение $\Omega_R = 0.25 \gamma_3$. Когда это не приводит к недоразумению, конкретные значения констант γ_2 , $\gamma_{3 \rightarrow 1}$ и $\gamma_{3 \rightarrow 2}$ будут приводиться в единицах γ_3 .

Примеры установившихся значений заселенности состояний частицы при трех конкретных значениях γ_2 (0.05, 0.01, 0.001) для трех различных частиц, одна из которых характеризуется значениями $\gamma_{3 \rightarrow 1} = 0.99$, $\gamma_{3 \rightarrow 2} = 0.01$; другая – $\gamma_{3 \rightarrow 1} = \gamma_{3 \rightarrow 2} = 0.5$; третья – $\gamma_{3 \rightarrow 1} = 0.01$, $\gamma_{3 \rightarrow 2} = 0.99$ приведены в таблице.

Как видно из таблицы, перераспределение первоначальной заселенности основного состояния частицы со временем по всем трем состояниям зависит от значений констант $\gamma_{3 \rightarrow 1}$, $\gamma_{3 \rightarrow 2}$, γ_2 . Заметим, что это перераспределение одинаково для одинакового значения отношения $\gamma_2 / \gamma_{3 \rightarrow 2}$. Например, когда это отношение равно 0.01 распределение заселенности по состояниям частицы одинаково как для частицы с $\gamma_{3 \rightarrow 1} = 0.99$, $\gamma_{3 \rightarrow 2} = 0.01$, так и для частицы с $\gamma_{3 \rightarrow 1} = \gamma_{3 \rightarrow 2} = 0.5$ – см. третью и четвертую строку в таблице установившихся значений заселенности состояний. В общем же случае, можно отметить, что установившуюся заселенность можно характеризовать следующим образом: заселённость состояния 3 тем больше, чем больше $\gamma_{3 \rightarrow 1}$ и γ_2 , но при этом она значительно меньше заселенности состояний 1 и 2, а перераспределение заселенности между состояниями 1 и 2 тем больше, чем больше $\gamma_{3 \rightarrow 2}$ и чем меньше γ_2 . При $\gamma_2 = 0$ установившимся состоянием частицы можно считать полностью заселенное “темное” состояние 2 с “отключением” взаимодействия частицы с не изменившимся полем об-

Таблица установившихся значений заселённости состояний частицы

$\gamma_{3 \rightarrow 1}$	$\gamma_{3 \rightarrow 2}$	γ_2	ρ_{11}	ρ_{22}	ρ_{33}
0.99	0.01	0.05	0.9341	0.0110	0.0549
		0.01	0.8947	0.0526	0.0526
		0.001	0.6071	0.3571	0.0357
0.5	0.5	0.05	0.6071	0.3571	0.0357
		0.01	0.2500	0.7353	0.0147
		0.001	0.0328	0.9653	0.0019
0.01	0.99	0.05	0.4497	0.5238	0.0265
		0.01	0.1453	0.8462	0.0085
		0.001	0.0169	0.9821	0.0010

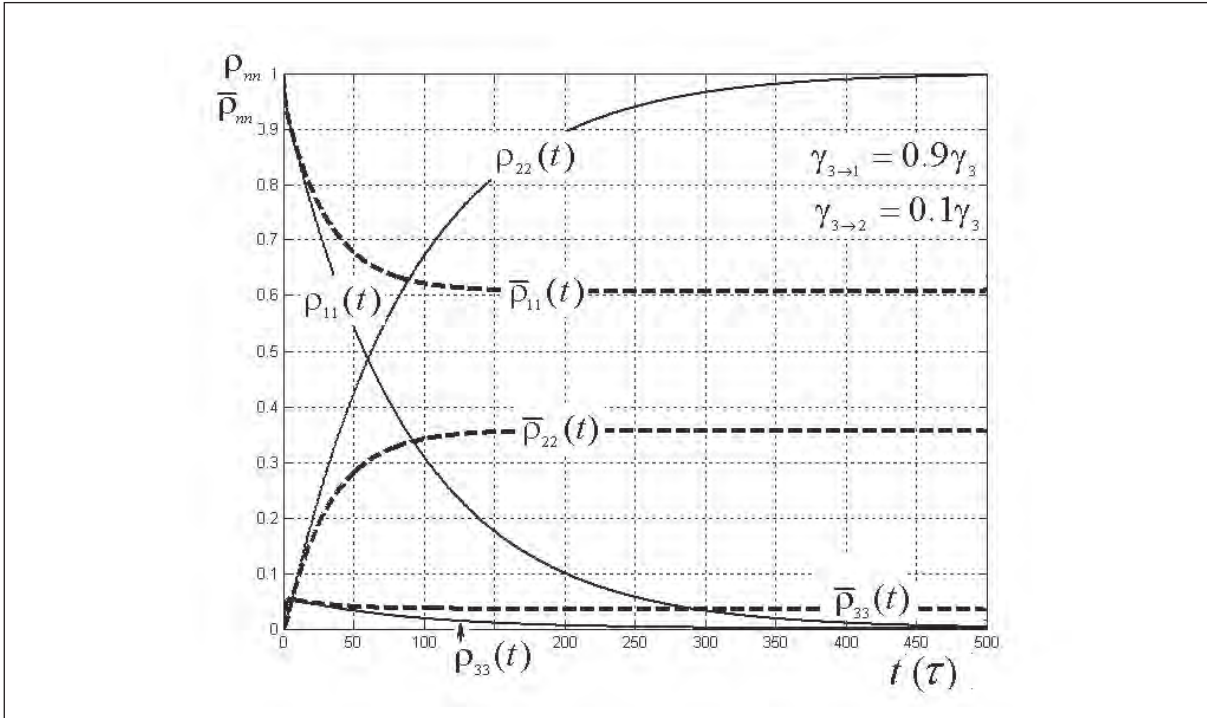


Рис. 2. Определённая на основе решения уравнений Блоха динамика заселённости состояний частицы при преобладании во вторичном излучении релеевского рассеяния над комбинационным рассеянием.

лучения, как в [15,16]. Установившиеся значения заселённости состояний частицы с $\gamma_2 > 0$ характеризуются перераспределением заселённости начального состояния 1 по состояниям 1 и 3, которое тем больше, чем больше отношение $\gamma_{3 \rightarrow 1} / \gamma_{3 \rightarrow 2}$. В зависимости от величины этого отношения при одном и том же значении γ_2 заселённость состояния 2 может быть больше заселённости состояния 1 и состояния 3, а может быть меньше. Например, $\rho_{22}(\infty) \geq \rho_{11}(\infty)$ при $\gamma_{3 \rightarrow 2} \geq 17\gamma_2$, но $\rho_{22}(\infty) \leq \rho_{11}(\infty)$ при $\gamma_{3 \rightarrow 2} \leq 17\gamma_2$.

Характерные примеры динамики заселённости состояний частицы, отображающиеся видом функций $\rho_{mn}(t)$, полученных в результате численных решений приведенной системы ОУБ для $\Omega_R = 0.25\gamma_3$ при начальном условии $\rho_{11}(0) = 1$ приведены на рис. 2 и рис. 3. Тонкие сплошные кривые отображают динамику заселённости состояний частицы с “темным” состоянием 2 и помечены как $\rho_{nn}(t)$, а утолщенные штриховые – с $\gamma_2 = 0.01$ и помечены как $\bar{\rho}_{nn}(t)$. На рис. 2 приведены функции $\rho_{nn}(t)$ и $\bar{\rho}_{nn}(t)$ для частицы, характеризующейся значениями $\gamma_{3 \rightarrow 1} = 0.9$ и $\gamma_{3 \rightarrow 2} = 0.1$ (преобладание во вторичном излучении РР над КР), а на рис. 3 – для частицы с $\gamma_{3 \rightarrow 1} = 0.1$ и $\gamma_{3 \rightarrow 2} = 0.9$. (преобладание КР над РР).

Сравнивая рис. 2 с рис. 3, можно заключить, что расселение исходного состояния 1, сопровождающееся заселением состояний 2 и 3, происходит тем быстрее, чем больше значение константы $\gamma_{3 \rightarrow 2}$. Важную роль в описании формирования расселения состояния 1 играет учет слагаемых $2\gamma_{3 \rightarrow 1}\rho_{33}$ и $2\gamma_2\rho_{22}$ в правой части уравнения для $\dot{\rho}_{11}$. По [18] следует полагать, что наличие этих слагаемых отражает учет “перепоглощения” частицей рассеиваемого света.

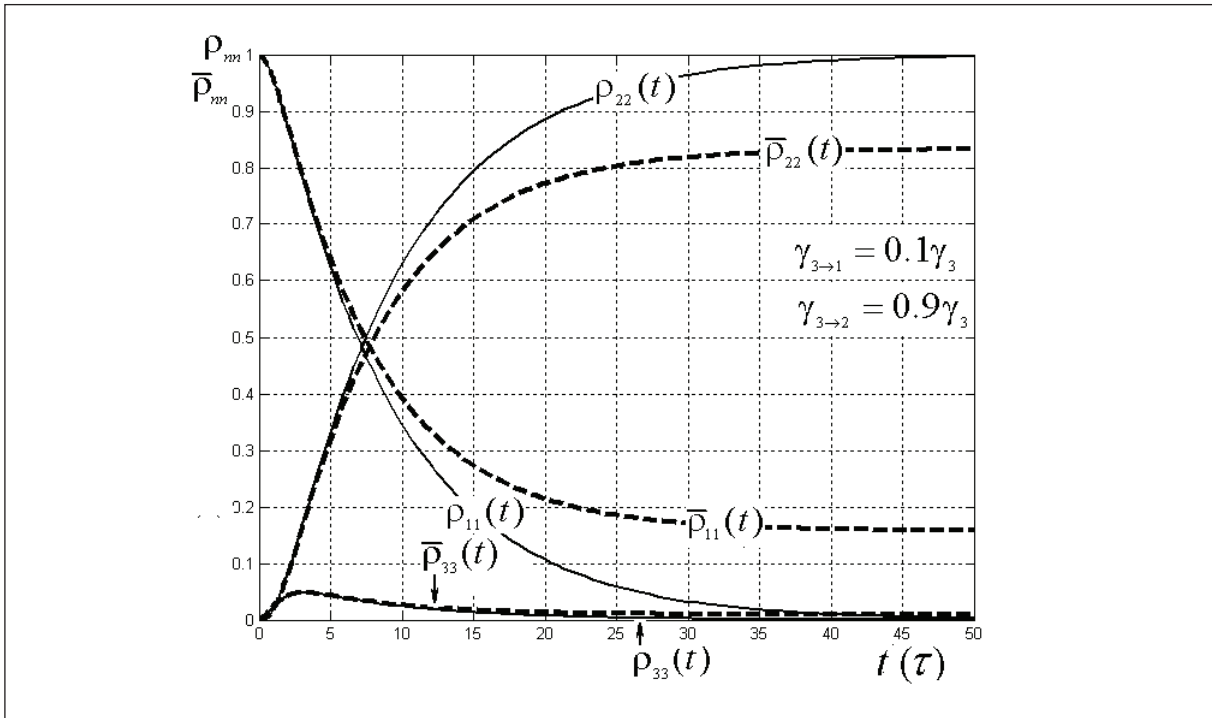


Рис. 3. Определённая на основе решения уравнений Блоха динамика заселённости состояний частицы при преобладании во вторичном излучении комбинационного рассеяния над релеевским рассеянием.

Моделирование на основе квантовой теории в излучения составной системе

1. Динамика заселенности состояний при спонтанном излучении

При рассмотрении спонтанного излучения полагаем, что в начальный момент времени частица находится в состоянии 3, а поле излучения не содержит фотонов. Со временем возникает состояние $\sigma 1$, означающее, что в поле излучения появился фотон сорта σ , а частица оказалась в состоянии 1, или состояние $\mu 2$, означающее, что в поле излучения появился фотон сорта μ , а частица оказалась в состоянии 2, или состояние $\mu \nu 1$, означающее, что в поле излучения появился фотон сорта μ и фотон ν , а частица оказалась в состоянии 1.

Для объёма L^3 пространства, в котором находится частица и рассматриваемое поле излучения, и для реального значения матричных элементов оператора дипольного момента частицы при введенных обозначениях для матричного элемента оператора V по состояниям $\sigma 1$ и 03 имеем выражение: $V_{\sigma 1}^{03} = iL^{-3/2} \sqrt{2\pi \cdot \hbar \omega_{31}} \cdot (d_{\sigma})_{13}$, где d_{σ} – проекция $\mathbf{d}$ на направление поляризации фотона σ , $(d_{\sigma})_{13}$ – матричный элемент оператора d_{σ} по состояниям 1 и 3.

Выражение для амплитуды вероятности $b_n(t)$ заселенности состояния n составной системы определяется решением уравнения Шредингера в представлении взаимодействия. При использовании Фурье-представления

$$b_n(t) = i(2\pi)^{-1} \int_{-\infty}^{+\infty} G_n(E) \cdot \exp[i(E_n - E)t/\hbar] dE$$

согласно уравнению Шредингера имеем

$$(E - E_n) \cdot G_n(E) = \sum_m V_{nm} G_m(E) + \delta_{n0},$$

где δ_{n0} – символ Кронекера, в котором цифра 0 означает, что частица находится в состоянии 3, а поле излучения не содержит фотонов при отсчитывании времени от того момента, когда возбудивший частицу короткий световой импульс перестал взаимодействовать с ней.

Решая систему уравнений для $G_n(E)$, находим

$$\begin{aligned} G_{03}(E) &= (E - E_{03} + i\hbar\gamma_3)^{-1}, \\ G_{\sigma 1}(E) &= V_{\sigma 1}^{03} G_{03}(E) \zeta(E - E_{\sigma 1}), \\ G_{\mu 2}(E) &= V_{\mu 2}^{03} G_{03}(E) (E - E_{\mu 2} + i\hbar\gamma_2)^{-1}, \\ G_{\mu\nu 1}(E) &= V_{\mu\nu 1}^{\mu 2} G_{\mu 2}(E) \zeta(E - E_{\mu 2}). \end{aligned}$$

Использованы обычно вводимые (см. [19]) обозначения:

$$\begin{aligned} \gamma_{3 \rightarrow 1} &= i\hbar^{-1} \sum_{\sigma} |V_{\sigma 1}^{03}|^2 \zeta(E - E_{\sigma 1}), \text{ аналогичное выражение для } \gamma_2, \text{ где } \zeta(E) = P/E + i\pi\delta(E), \\ P/E &- \text{ главное значение функции } 1/E, \delta(E) - \text{ функция Дирака, и выражение} \\ \gamma_{3 \rightarrow 2} &= i\hbar^{-1} \sum_{\mu} |V_{03}^{\mu 2}|^2 (E - E_{\mu 2} + i\hbar\gamma_2)^{-1}; \text{ как и раньше } \gamma_{3 \rightarrow 1} + \gamma_{3 \rightarrow 2} \equiv \gamma_3. \end{aligned}$$

Вывражениями $\gamma_{n \rightarrow m}$ подразумеваются реальные части приведенных выше этих величин, взятые при значениях энергии составной системы в состоянии m , а мнимые их части не учитываются [19], так что, например, $\gamma_{3 \rightarrow 1} = \pi\hbar^{-2} \sum_{\sigma} |V_{\sigma 1}^{03}|^2 \delta(E_{03} - E_{\sigma 1}) = 2\omega_{31}^3 d_{31}^2 / 3\hbar c^3 -$

константа скорости радиационного распада состояния частицы 3 с переходом в состояние 1. Как принято в предыдущем разделе, полагаем, что значения таких констант совпадают со значениями соответствующих констант в принятой системе ОУБ.

Используя эти решения, находим амплитуды $b_n(t)$. Например, для $b_{\mu 2}(t)$ получаем:

$$b_{\mu 2}(t) = \hbar^{-2} V_{\mu 2}^{03} \{ \exp[-i(\omega_{32} - \omega_{\mu} - i\gamma_3)t] - \exp(-\gamma_2 t) \} [\omega_{32} - \omega_{\mu} - i(\gamma_3 - \gamma_2)]^{-1}.$$

Полученные выражения для других амплитуд $b_n(t)$ выглядят громоздко, и приводятся не будут.

Следуя общим представлениям теории преобразования света веществом и квантовых измерений состояний открытых систем [20], будем полагать, что взаимодействие преобразующей свет частицы с естественным макроскопическим окружением приводит к “декогерентным” состояниям поля излучения при соответствующих состояниях частицы и непрерывно осуществляется реальная – как это отмечено и в [21] – регистрация (“измерение”) состояний частицы при детектировании поля излучения, т.е. при уничтожении соответствующего излученного фотона. Такая регистрация состояния части-

цы может иметь селективный характер, позволяя определять вероятности альтернативных парциальных ее эволюций, если поглощается только некоторый определенный вид фотонов, излучаемых частицей при ее переходе в рассматриваемое состояние. С учетом этого, заселенность состояния молекулы будем определять суммой вероятностей состояний составной системы по возможным состояниям фотонов, соответствующих рассматриваемой эволюции состояний частицы (как и в [15–17]). Такое описание динамики заселенности состояний частицы в [20] названо “селективным”, в отличие от основанного на решении системы ОУБ “неселективного” описания, при котором заселенность рассматриваемого состояния частицы напрямую не связана с характеристиками фотонов, излученных при переходе частицы в это состояние.

Руководствуясь принятыми определениями, суммируем квадраты модулей полученных выражений для амплитуд составной системы по состояниям соответствующих фотонов и сравнивая полученные выражения с приведенными выше решениями $\rho_{nn}(t)$ системы ОУБ, получаем равенства:

$$\begin{aligned} |b_{03}(t)|^2 &\equiv P_3(t) = \rho_{33}(t), \\ \sum_{\mu} |b_{\mu 2}(t)|^2 &\equiv P_2(t) = \rho_{22}(t), \\ \sum_{\sigma} |b_{\sigma 1}(t)|^2 &\equiv P_{(\sigma)1}(t) = \gamma_{3 \rightarrow 1} [1 - \exp(-2\gamma_3 t)] / \gamma_3, \end{aligned}$$

$$\sum_{\mu\nu} |b_{\mu\nu 1}(t)|^2 \equiv P_{(\mu\nu)1} = \gamma_2 \gamma_{3 \rightarrow 2} \{ \gamma_2^{-1} \gamma_3^{-1} + [\gamma_2^{-1} \exp(-2\gamma_2 t) - \gamma_3^{-1} \exp(-2\gamma_3 t)] (\gamma_2 - \gamma_3)^{-1} \},$$

так что $P_{(\sigma)1}(t) + P_{(\mu\nu)1}(t) = \rho_{11}(t)$.

Эти равенства можно интерпретировать как демонстрацию согласованности динамики заселенности состояний при спонтанном излучении рассматриваемой молекулы, предсказываемой результатом решения системы ОУБ, с динамикой соответствующих состояний, предсказываемой КТИ с учетом непрерывного реального детектирования всех видов фотонов, излученных молекулой при переходах в эти состояния. При этом результат подхода КТИ поясняет происхождение отдельных слагаемых в полученной решением системы ОУБ функции $\rho_{11}(t)$, как результата вкладов от однофотонных и двухфотонных переходов, и предсказывает окончательное значение заселенности состояния 1 в зависимости от детектирования тех или иных – σ или μ и ν – фотонов спонтанного излучения ($\gamma_{3 \rightarrow 1} / \gamma_3$ или $\gamma_{3 \rightarrow 2} / \gamma_3$).

2. Динамика при преобразовании монохроматического излучения

Ограничимся рассмотрением преобразования частицей, находящейся первоначально в основном состоянии 1, одного из N фотонов монохроматического облучения. Рассматриваемые при этом состояния составной системы обозначены индексами $\lambda 1$, 03 , $\mu 2$, $\sigma 1$, $\mu\nu 1$, которые приведены на рис. 1 возле горизонтальных линий, отображающих уровни энергии частицы. Состояние $\lambda 1$ означает, что частица находится в со-

стоянии 1, а поле облучения содержит N фотонов сорта λ ; 03 – частица в состоянии 3, а поле содержит $N-1$ фотонов λ ; $\sigma 1$ – частица в состоянии 1, а поле содержит $N-1$ фотонов λ и фотон резонансного релеевского рассеяния (РР) сорта σ ; $\mu 2$ – частица в состоянии 2, а поле содержит $N-1$ фотонов λ и фотон резонансного комбинационного рассеяния (КР) сорта μ ; $\mu \nu 1$ – частица в состоянии 1, а поле содержит $N-1$ фотонов λ , фотон сорта μ и фотон сорта ν , относящийся к энергетически релаксированной флуоресценции.

Для матричного элемента оператора V по состояниям $\lambda 1$ и 03 имеем выражение: $V_{\lambda 1}^{03} = iL^{-3/2} \sqrt{2\pi \cdot N \cdot \hbar \omega_{31}} \cdot (d_{\lambda})_{13} \equiv i\hbar \bar{\Omega}_R$, где d_{λ} – проекция $\mathbf{d}$ на направление поляризации фотона λ . Будем полагать, что $\bar{\Omega}_R = \Omega_R$.

Для нахождения амплитуд $b_n(t)$ проводим вычисления, аналогичные вычислениям при рассмотрении спонтанного излучения, полагая начальным состояние $0 = \lambda 1$. Получено, что не зависящие от значения γ_2 выражения $b_{\lambda 1}(t)$, $b_{03}(t)$, $b_{\sigma 1}(t)$ (и выражение $b_{\mu 2}(t)$ для частицы с $\gamma_2 = 0$) совпадают с соответствующими выражениями, полученными в [22,23] методом Гайтлера–Ма [19], и здесь они приводиться не будут. Используя эти выражения, определяем динамику расселения состояния 1, соответствующую поглощению фотона облучения частицей. Она определяется следующим выражением:

$$P_1(t) = |b_{\lambda 1}(t)|^2 = \exp(-\gamma_3 t) [\cosh^2 Rt + (1-\alpha)^{-1} \sinh^2 Rt],$$

Динамика заселенности состояния 3 при поглощении фотона облучения определяется следующим выражением:

$$P_3(t) = |b_{03}(t)|^2 = \alpha(1-\alpha)^{-1} \exp(-\gamma_3 t) \sinh^2 Rt],$$

Динамика заселенности состояния 1 при детектировании фотонов РР определяется следующим выражением:

$$P_{(\sigma)1}(t) = \sum_{\sigma} |b_{\sigma 1}(\omega_{\sigma}, t)|^2 = \gamma_{3-1} \gamma_3^{-1} \{1 - \exp(-\gamma_3 t) [\cosh 2Rt + 2R\gamma_3^{-1} \sinh 2Rt - \alpha](1-\alpha)^{-1}\},$$

где $\alpha \equiv 4\Omega_R^2 / \gamma_3^2 < 1$, $R = 0.5\gamma_3(1-\alpha)^{1/2}$.

Для частицы, у которой $\gamma_2 > 0$, получаем:

$$b_{\mu 2}(\omega_{\mu}, t) = \frac{V_{\mu 2}^{03} V_{03}^{\lambda 1}}{\hbar^2} \left[\frac{\exp[-(b+ix)t]}{(b-\gamma_2+ix)(a-b)} - \frac{\exp[-(a+ix)t]}{(a-\gamma_2+ix)(a-b)} - \frac{\exp(-\gamma_2 t)}{(a-\gamma_2+ix)(b-\gamma_2+ix)} \right],$$

где $x = \omega_{32} - \omega_{\mu}$; $a = 0.5\gamma_3(1-\sqrt{1-\alpha})$, $b = 0.5\gamma_3(1+\sqrt{1-\alpha})$.

Динамика заселённости состояния 2 такой частицы при детектировании фотонов КР определяется выражением

$$P_2(t) = \sum_{\mu} |b_{\mu 2}(\omega_{\mu}, t)|^2 = \gamma_{3 \rightarrow 2} F_{32}(a, b, \gamma_2, t),$$

где

$$F_{32}(a, b, \gamma_2, t) = \frac{\alpha \gamma_3^2}{4} \cdot \left[\frac{\exp(-2\gamma_2 t)}{(a - \gamma_2)(b - \gamma_2)(\gamma_3 - 2\gamma_2)} - \frac{1}{\gamma_3^2(1 - \alpha)} \left(\frac{\exp(-2at)}{(a - \gamma_2)} + \frac{\exp(-2bt)}{(b - \gamma_2)} \right) + \frac{4\exp(-\gamma_3 t)}{\gamma_3^2(1 - \alpha)(\gamma_3 - 2\gamma_2)} \right].$$

Для $b_{\mu\nu 1}(t)$ получено такое выражение:

$$b_{\mu\nu 1}(\omega_{\mu}, \omega_{\nu}, t) = \frac{V_{\mu\nu 1}^{\mu 2} V_{\mu 2}^{03} V_{03}^{\lambda 1}}{\hbar^3} \left[\frac{\exp[-i(x - i\gamma_2)t]}{[y + i(\gamma_2 - a)][y + i(\gamma_2 - b)](x - i\gamma_2)} - \frac{1}{(y + x - ia)(y + x - ib)(x - i\gamma_2)} + \frac{\exp[-i(y + x)t]}{i(a - b)} \left\{ \frac{\exp(-bt)}{(y + x - ib)[y + i(\gamma_2 - b)]} - \frac{\exp(-at)}{(y + x - ia)[y + i(\gamma_2 - a)]} \right\} \right],$$

где $y = \omega_{21} - \omega_{\nu}$.

Соответствующее выражение, определяющее динамику заселенности состояния 1 при детектировании фотонов μ и ν , имеет вид:

$$P_{(\mu\nu)1}(t) = \sum_{\mu\nu} |b_{\mu\nu 1}(\omega_{\mu}, \omega_{\nu}, t)|^2 = \gamma_{3 \rightarrow 2} \gamma_2 \Omega_R^2 F_{3 \rightarrow 2 \rightarrow 1}(a, b, t)$$

$$F_{3 \rightarrow 2 \rightarrow 1}(a, b, t) = \left[\frac{1}{\gamma_3 \gamma_2 \Omega_R^2} + \frac{\exp(-2\gamma_2 t)}{\gamma_2(a - \gamma_2)(b - \gamma_2)(2\gamma_2 - \gamma_3)} + \frac{8\exp(-\gamma_3 t)}{\gamma_3^3(1 - \alpha)(2\gamma_2 - \gamma_3)} + \frac{1}{\gamma_3^2(1 - \alpha)} \left(\frac{\exp(-2at)}{a(a - \gamma_2)} + \frac{\exp(-2bt)}{b(b - \gamma_2)} \right) \right]$$

В частности, при $t = \infty$ $P_{(\sigma)1}(\infty) = \gamma_{3 \rightarrow 1} / \gamma_3$ независимо от значения константы γ_2 ; $P_2(\infty) = \gamma_{3 \rightarrow 2} / \gamma_3$ для частицы с $\gamma_2 = 0$ и $P_{(\mu\nu)1}(\infty) = \gamma_{3 \rightarrow 2} / \gamma_3$ для частицы с $\gamma_2 > 0$, как это имеет место при спонтанном излучении.

Имеют место равенства:

$$\frac{d}{dt} \sum_{\sigma} |b_{\sigma 1}(\omega_{\sigma}, t)|^2 = 2\gamma_{3 \rightarrow 1} P_3(t), \quad \frac{d}{dt} \sum_{\mu\nu} |b_{\mu\nu 1}(\omega_{\mu}, \omega_{\nu}, t)|^2 = 2\gamma_2 P_2(t),$$

справедливые так же и для соответствующих выражений, полученных при рассмотрении динамики состояний частицы при спонтанном излучении.

Заметим, что приведенное выше выражение для динамики заселенности состояния 1 частицы при детектировании фотонов РР может быть записано в виде

$$P_{(\sigma)1}(t) = \sum_{\sigma} |b_{\sigma 1}(\omega_{\sigma}, t)|^2 = \gamma_{3 \rightarrow 1} F_{31}(a, b, t),$$

где $F_{31}(a, b, t) = F_{32}(a, b, \gamma_2 = 0, t)$.

Рассмотренная динамика заселенности состояний системы отражает преобразование частицей одного фотона поля облучения. В [23] получены выражения для установившихся значений заселенности состояний рассмотренной там трёхуровневой модели частицы при учёте возможности такого преобразования многих фотонов поля облучения, (названного “каскадным” релеевским рассеянием), которое обозначено следующей совокупностью символов:

$$\begin{aligned} N\lambda, 1 \rightarrow (N-1)\lambda, 3 \rightarrow (N-1)\lambda\sigma_1, 1 \rightarrow (N-2)\lambda\sigma_1, 3 \\ \rightarrow (N-2)\lambda\sigma_1\sigma_2, 1 \rightarrow (N-3)\lambda\sigma_1\sigma_2, 3 \dots \rightarrow \sigma_1\sigma_2 \dots \sigma_N, 1' \end{aligned}$$

где, например, $(N-k)\lambda\sigma_1\sigma_2 \dots \sigma_k, 1$ обозначает состояние системы, при котором поле облучения содержит $N-k$ фотонов λ , поле вторичного излучения содержит k фотонов релеевского рассеяния $\sigma_1, \sigma_2 \dots \sigma_k$, а молекула находится в состоянии 1. Для того, чтобы учесть неразличимость фотонов σ , волновые функции этих состояний были симметризованы относительно перестановки сортов фотонов ($\sigma_1 \leftrightarrow \sigma_2$ и т.д.). В [23] показано, что при использовании таких симметризованных волновых функций значение константы $\gamma_{3 \rightarrow 1}$ не зависит от числа преобразованных фотонов облучения, и выражения для $P_{(\sigma)1}(\infty)$ и $P_{(\mu)2}(\infty)$ имеют такой же вид, как и приведенные выше выражения, полученные для случая преобразования одного фотона поля облучения. На основании этого можно полагать, что полученные в настоящей работе выражения для заселенности состояний частицы справедливы и для случая каскадного релеевского рассеяния при сохранении принятых ограничений на значение интенсивности поля облучения. В [23] отмечено, что каскадное релеевское рассеяние двухуровневой молекулой впервые описано в [24]. Заметим, что такое преобразование света с учётом неразличимости рассеянных фотонов приведено в [25].

Рис. 4 и рис. 5 иллюстрируют вид полученных выражений $P_k(t)$ для $k = 1, 3, (\sigma)1, (\mu)2, (\mu\nu)1$ при $\Omega_R = 0.25\gamma_3$ ($\alpha = 0.25$). Так же, как и на рис. 2 и рис. 3, тонкие сплошные кривые соответствуют частице с $\gamma_2 = 0$ и помечены $P_k(t)$, а утолщенные штриховые и штрих-пунктирные кривые – частице с $\gamma_2 = 0.01$ и помечены как $\bar{P}_k(t)$. На этих рис. по оси абсцисс отложено время в единицах γ_3^{-1} , а по оси ординат – значения P_k и $\bar{P}_k$ для частицы с $\gamma_{3 \rightarrow 1} = 0.9$, $\gamma_{3 \rightarrow 2} = 0.1$ на рис. 4, а для частицы с $\gamma_{3 \rightarrow 1} = 0.1$, $\gamma_{3 \rightarrow 2} = 0.9$ на рис. 5.

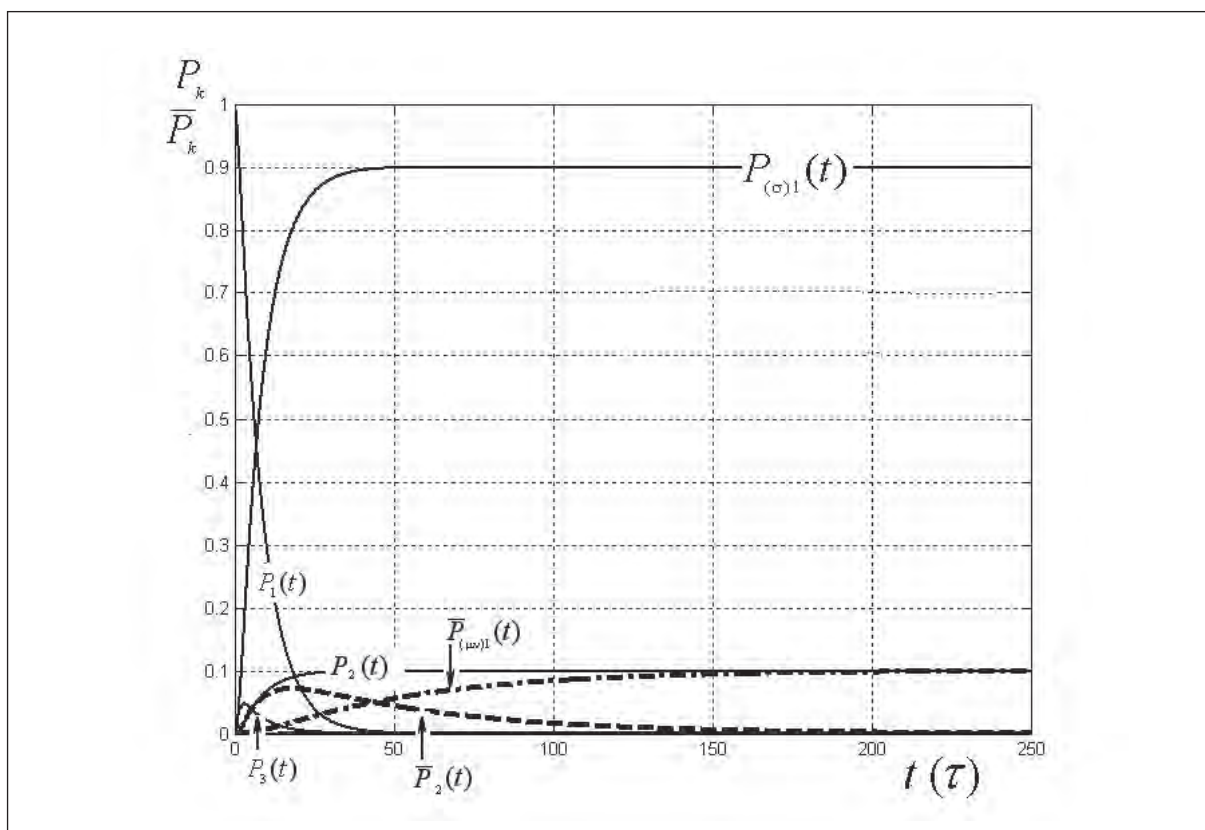


Рис. 4. Динамика заселённости состояний частицы по квантовой теории излучения при преобладании во вторичном излучении релеевского рассеяния над комбинационным рассеянием.

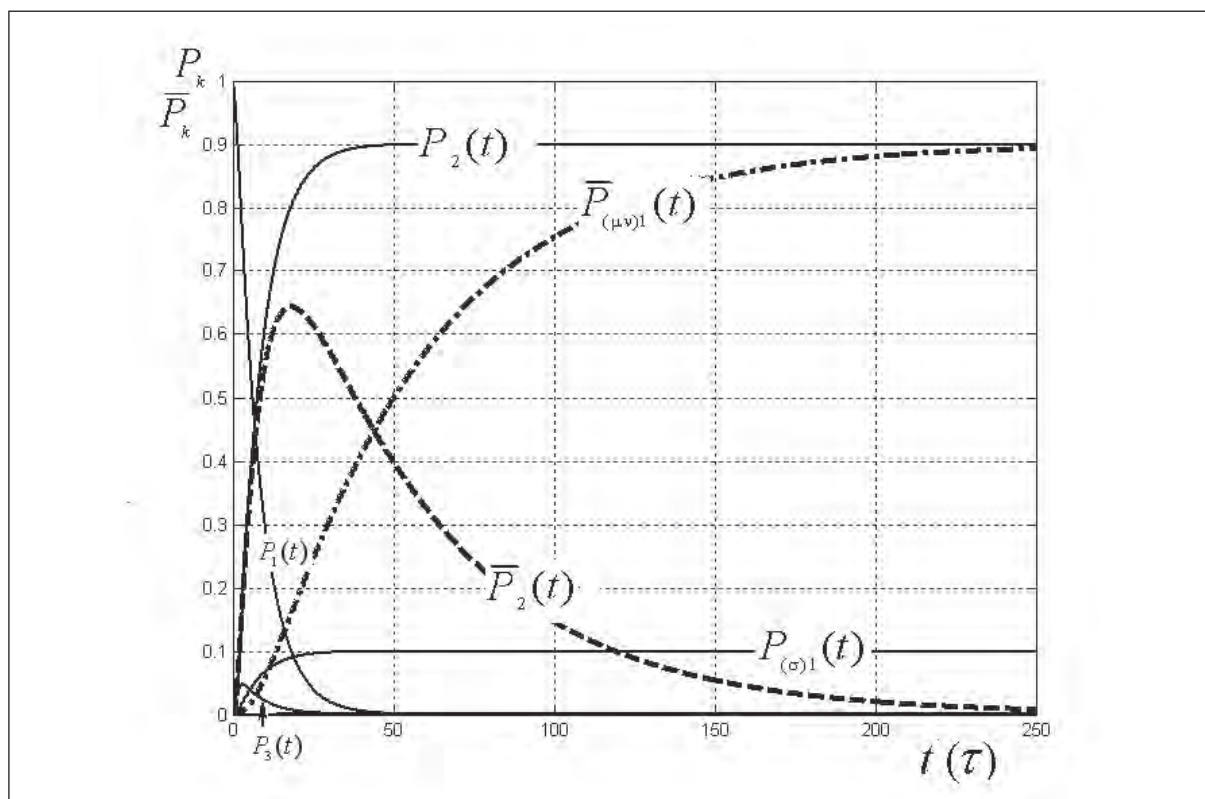


Рис. 5. Динамика заселённости состояний частицы по квантовой теории излучения при преобладании во вторичном излучении комбинационного рассеяния над релеевским рассеянием.

Сравнение результатов применения двух подходов

Как отмечено выше, полученные равенства между характеристиками динамики заселенности состояний частицы при спонтанном излучении, установленные при использовании ОУБ и при использовании КТИ, можно интерпретировать как демонстрацию согласованности результатов применения этих двух подходов в том смысле, что при переходе от “селективного” описания рассматриваемой динамики к “неселективному” получается одинаковый результат, но такая согласованность не имеет места при сравнении результатов моделирования динамики заселенности состояний частицы в случае преобразования ею монохроматического излучения.

Рисунки 2 ÷ 5 иллюстрируют характерные существенные различия динамики заселенности состояния n при моделировании на основе ОУБ – функции $\rho_{nn}(t)$, и на основе КТИ – функции $P_n(t)$. Для частицы с $\gamma_2 = 0$ эти различия обсуждены в [15,16]. Для частицы с $\gamma_2 > 0$ (в приведенном на рисунках случае с $\gamma_2 = 0.01 \gamma_3$) моделирование по ОУБ предсказывает, что начальное состояние частицы рассеяется тем быстрее, чем больше отношение $\gamma_{3 \rightarrow 2} / \gamma_{3 \rightarrow 1}$, а по КТИ – скорость этого рассеяния определяется значением константы γ_3 . Это различие обусловлено тем, что формализм ОУБ подразумевает такое преобразование облучения, при котором частица – как замкнутая квантовая система – возвращается в начальное состояние и, как это трактуется в [18], происходит “перепоглощение” рассеянного света, а согласно КТИ составная система не возвращается в начальное состояние при рассеянии света, поскольку $\sigma \neq \lambda$, и “перепоглощение” не происходит. При моделировании по ОУБ заселенность начального состояния перераспределяется со временем по всем состояниям частицы и при $t \gg \gamma_3^{-1}$ – см. таблицу – все состояния частицы заселены в той или иной мере (в случае частицы с $\gamma_2 = 0$ асимптотически полностью заселяется “темное” состояние 2), в то время как по КТИ при $t \gg \gamma_3^{-1}$ частица находится в состоянии 1 с вероятностью $P_{(\sigma)1}(\infty) = \gamma_{31} / \gamma_3$ при регистрации фотонов РР, или с вероятностью $P_{(\mu\nu)1}(\infty) = \gamma_{32} / \gamma_3$ при регистрации фотонов КР и флуоресценции. Разница между стационарными значениями заселенности соответствующих состояний частицы, полученными при использовании двух применявшихся подходов тем больше, чем больше отношение $\gamma_{3 \rightarrow 1} / \gamma_{3 \rightarrow 2}$, т.е. чем больше интенсивность РР частицы по отношению к интенсивности КР.

Заключительные замечания

Если руководствоваться отмеченным во Введении положением, что формализм моделирования динамики заселенности состояний составной системы методами квантовой теории излучения является более последовательным и строгим по сравнению с формализмом, основанном на решении оптических уравнений Блоха для частицы как замкнутой квантовой системы, то следует заключить, что установленные различия результатов этих двух подходов к моделированию динамики заселенности состояний трехуровневой частицы (проявляющиеся тем значительнее, чем больше интенсивность резонансного РР частицы по сравнению с соответствующим КР) свидетельствуют о существенном ограничении пригодности использования оптических уравнений Блоха для такого моделирования и необходимости проведения его на основе квантовой теории излучения.

Представляет интерес получить подтверждение такого заключения данными экспериментального определения динамики заселенности состояний трехуровневой частицы. Такое определение можно осуществить, например, по результатам измерения скорости и эффективности какой-либо фотореакции с переносом энергии возбуждения метастабильного состояния молекулы-реагента, рассматривающейся как трехуровневая частица. При выборе конкретных молекул органических соединений, подходящих для проведения такого исследования (“химической спектроскопии”), полезно иметь в виду установленные численные характеристики динамики заселённости состояний трехуровневой наночастицы.

Литература

1. Javanainen J. // Phys. Rev. A., 1986, **33**, P. 2121.
2. Arecchi F.T., Schenzie A., DeVoe R.G., Jungmann K., Brewer R.G. // Phys. Rev. A., 1986, **33**, P. 2124.
3. Schenzie A., DeVoe R.G., Brewer R.G. // Phys. Rev. A., 1986, **33**, P. 2127.
4. Plenio M. B., Knight P.L. // Rev. Mod. Phys., 1998, **70**, P. 101.
5. Скалли М.О., Зубайри М.С. // Квантовая оптика. М.: ФИЗМАТЛИТ 2003. 504 стр.
6. Zheng Y., Brown F.L.H. // J. Chem. Phys., 2003, **119**, P. 11814.
7. Осадько И.С. // Флуктуирующая флуоресценция наночастиц. М.: ФИЗМАТЛИТ 2011- 320 стр.
8. Lu H.P., Luying Xun, Xie X.S. // Science, 1998, **282**, P. 1877.
9. Yip. Wai-Tak., Hu. Dehong., Yu. Ji., D.A. Vanden Bout., P.F Barbara. // J. Phys. Chem. A., 1998, **102**, P. 7564.
10. Леонов М.Ю., Турков В.К., Рухленко И.Д., Фёдоров А.В. // Оптика и спектроскопия, 2012, **113**, С. 288; С. 295.
11. Huang Hui, Li Gao-Xiang, Gu Wen-ju, Ficek Z. // Phys. Rev. A., 2014, **90**, P. 023815
12. Jafarzadeh H., Sahrai M., Jamshidi-Chalen K. // Eur. Phys. J. D., 2014, **68**, P. 115.
13. Осадько И.С. // Селективная спектроскопия одиночных молекул. М.: ФИЗМАТЛИТ 2000. 320 стр.
14. Макомбер Дж.Д. // Динамика спектроскопических переходов. М.: “Мир” 1979. 345 стр.
15. Морозов В.А., Чувылкин Н.Д., Смоленский Е.А. // Докл. РАН., 2015, **461**, С. 300.
16. Морозов В.А., Чувылкин Н.Д., Смоленский Е.А. // Хим. Физика, 2015, **34**, С. 3.
17. Морозов В.А., Шорыгин П.П. // Опт. и спектр., 1987, **63**, С. 1235.
18. Mollow B.R. // Phys. Rev. A. 1975, **12**, P. 1919.
19. Гайтлер В. // Квантовая теория излучения. Изд. Ин. литературы. Москва 1956, С. 491 стр.
20. Менский М.Б. // Успехи физ. наук, 2003, **173**, С.1199.
21. Морозов В.А., Шорыгин П.П. // Журн. физ. химии, 1990, **64**, С. 289.
22. Mies F.H., Ben Aryeh Y. // J. Chem. Phys., 1981, **74**, P. 53.
23. Mies F.H. // J. Quant. Spectrosc. Radiat. Transfer., 1983, **29**, P. 237.
24. Smithers M.E., Freedhoff H. S. // J. Phys. B: Atom. Molec. Phys., 1975, **8**, P. 911.
25. Морозов В.А. // Опт. и спектр., 1969, **26**, С. 116.

MATHEMATICAL MODELING OF STATES POPULATION DYNAMICS FOR A THREE-LEVEL NANOPARTICLE IN SPECTROSCOPIC TRANSITIONS

V.A. Morozov

*N.D. Zelinsky Institute of Organic Chemistry
Russian Academy of Sciences, Moscow*

morozov@mail.ioc.ac.ru

Received 18.05.2016

The mathematical modeling of the states population dynamics for an isolated three-level nanoparticle when it transforms a long pulse of monochromatic radiation, as well as in the case of spontaneous emission from the state with the highest energy, was performed. Two approaches were used, one of which is based on the use of numerical solutions of the optical Bloch equations for a particle, and the other employs the solutions of the Schrödinger equation for the probability amplitudes of the states population of the system composed of a particle, the quantum field of radiation, and the secondary radiation. In the case of spontaneous emission the results of dynamics modeling are mutually consistent when both the first and the second of these approaches are applied, but the results differ significantly in the case where a long pulse of monochromatic radiation is transformed by a particle. An explanation of the origin of this difference is given.

ПРОДОЛЖАЕМ ЧИСТИТЬ PROTEIN DATA BANK

А.А. Тужилин

*Кафедра дифференциальной геометрии и приложений механико-математического
факультета МГУ имени М.В.Ломоносова*

tuz@mech.math.msu.su

Поступила 14.05.2016

В настоящей работе мы продолжаем исследование геометрии расстояний между атомами белков, начатой в [1] и [2]. В качестве базы данных координат мы используем Protein Data Bank. Мы приведем ряд новых соображений, позволяющих понять, что файл с координатами атомов белка содержит ошибки. В результате мы построим новую базу, которая в следующих работах будет использована для изучения геометрии графа ковалентных связей и его связи с другими экстремальными графами, например, с минимальными остовными деревьями. Эта работа продолжит изучение геометрии ломаных, построенных на последовательных альфа-углеродах, см. [3], [4].

УДК 514.8, 51-76, 57.087

1 Введение

Может, мы обидели кого-то зря,
Календарь закроет этот лист.
К новым приключениям спешим, друзья,
Эй, прибавь-ка ходу, машинист!

Из детской песенки “Голубой вагон”.

В предыдущих работах автора и его коллег [1], [2] был получен ряд результатов, демонстрирующих необходимость очень внимательного отношения к файлам, выложенным в знаменитой базе данных Protein Data Bank (в дальнейшем, для краткости, PDB). Оказалось, что многие файлы содержат как путаницу в обозначениях, так и грубые ошибки в приводимых координатах атомов белков. Примером первого типа недочетов может служить глицин, “атомный состав” которого в файлах из PDB, полученных с помощью ядерно-магнитного резонанса (ЯМР), имеет 31 представление:

{55705, {N, CA, C, O, H, HA2, HA3}}, {1576, {N, CA, C, O, H1, HA2, HA3}},
 {870, {N, CA, C, O, H1, H2, H3, HA2, HA3}}, {399, {N, CA, C, O, OXT, H, HA2, HA3}},
 {390, {N, CA, C, O}}, {384, {N, CA, C, O, H}}, {47, {N, CA, C, O, H, HA2}},
 {37, {CA}}, {37, {N, CA, C, O, HA2, HA3}}, {26, {N, CA, C, O, H2, HA2, HA3}},
 {15, {N, CA, C, O, H1, H2, HA2, HA3}}, {15, {N, CA, C, O, HA2, HA3, H1, H2, H3}},
 {13, {N, CA, C, O, H, HA}}, {13, {N, CA, C, O, HA2, HN, HA1}}, {9, {N, CA, C}},
 {8, {N, CA, C, O, H1, H2, H3}}, {7, {C, CA, H, HA2, HA3, N, O}},
 {5, {N, CA, C, O, OXT, H}}, {5, {N, CA, C, H, HA2, HA3}},
 {4, {N, CA, C, O, H, HA3, HA2}}, {2, {N, CA, C, O, H1, H2, H3, HA3}},
 {2, {N, CA, C, O, HA2, HA3, H1, H2}}, {2, {N, CA, C, O, OXT}}, {2, {CA, C}},
 {1, {N, CA, C, O, OXT, H, HA2, HA3, HXT}}, {1, {N, CA, C, O, H, HA2, HA3, H2, H3}},
 {1, {N, CA, C, O, H2, H, H3, HA2, HA3}}, {1, {N, CA, C, O, H2, H, HA2, HA3}}, {1, {N}},
 {1, {N, CA, C, OXT, H, HA2, HA3}}, {1, {N, CA, C, O, HA2}}, {1, {N, CA, C, O, H, HA3}}.

Яркими примерами ошибок второго типа являются файлы 2PDE.pdb и 2I2J.pdb, см. рис. 1 и рис. 2, на которых приведены изображения некоторых аминокислот, построенных по координатам из этих файлов.

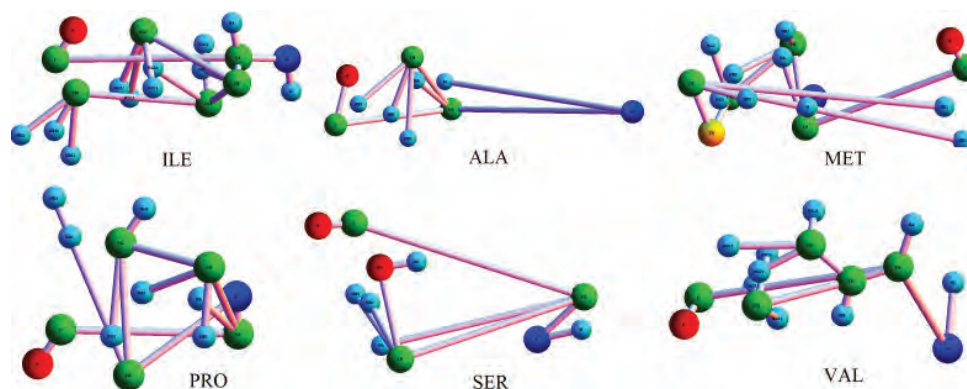


Рис. 1: Некоторые “аминокислоты” из файла 2PDE.pdb (1992 год).

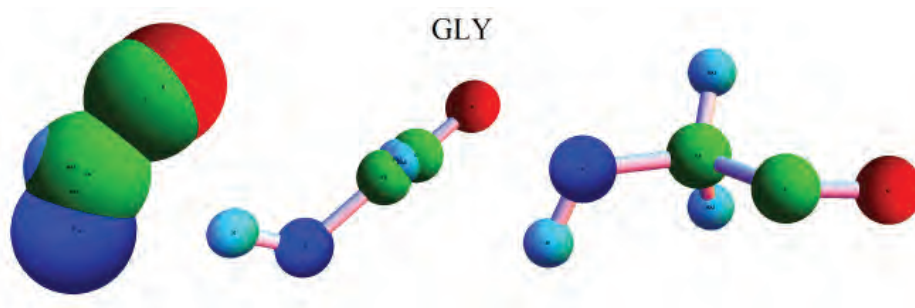


Рис. 2: Слева: два “глицина” из файла 2I2J.pdb (2006 год). Справа — “стандартный” глицин.

Так как первоначальной целью нашего исследования было изучение статистики геометрических свойств конформаций белков, нам нужна была более-менее надежная база данных белковых трехмерных структур. Обнаруженные проблемы с использованием PDB вынудили нас выкинуть примерно 2/3 файлов.

Любопытна реакция, вызванная нашей работой. Одни специалисты, увидев рис. 1, 2 и целый ряд других похожих иллюстраций, мгновенно отреагировали хорошо известной фразой “этого не может быть потому, что этого не может быть никогда” и посоветовали нам пойти искать свои ошибки. На это мы привели следующий фрагмент из файла 2I2J.pdb

АТОМ	14	N	GLY	A	2	-13.525	-1.614	-1.070	1.00	5.70	N
АТОМ	15	CA	GLY	A	2	-13.174	-1.687	-0.986	1.00	5.18	C
АТОМ	16	C	GLY	A	2	-12.623	-1.567	-0.783	1.00	4.25	C
АТОМ	17	O	GLY	A	2	-12.514	-1.508	-0.694	1.00	4.43	O
АТОМ	18	H	GLY	A	2	-13.535	-1.585	-1.035	1.00	5.83	H
АТОМ	19	HA2	GLY	A	2	-13.317	-1.721	-1.013	1.00	5.48	H
АТОМ	20	HA3	GLY	A	2	-13.165	-1.791	-1.040	1.00	5.56	H

взяли координаты $(-13.525, -1.614, -1.070)$ азота N, координаты $(-13.535, -1.585, -1.035)$ водорода H и на калькуляторе публично вычислили расстояние между этими атомами: оно оказалось приблизительно равным 0.0465403 ангстрема (эти азот и водород относятся к левой аминокислоте на рис. 2, где они почти совместились).

Кроме того, мы привели пример как полученного нами изображения аминокислот, так и результата визуализации этих аминокислот с помощью программы с сайта PDB, см. рис. 3.

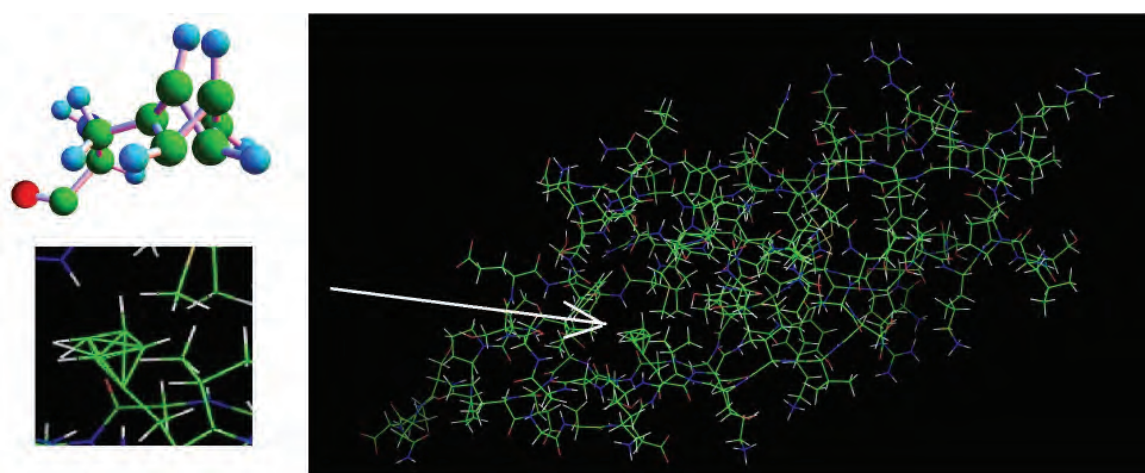


Рис. 3: Файл 2NPL.pdb, 2006 год.

Последнее свидетельствует о том, что обнаруженные нами “недочеты” могли быть выявлены стандартными средствами с PDB.

Другие специалисты сказали, что такие ошибки были лишь на ранней стадии развития, а затем мощные методы верификации все ошибки исправили. Особенно безупречной, по мнению специалистов, является программа CYANA. В ответ на это заявление мы мгновенно привели следующий пример, рис. 4.

Этот же файл может служить иллюстрацией к обсуждению различных моделей (для ЯМР): некоторые специалисты считают, что первая модель получается усреднением, поэтому тут возникновение ошибок вполне понятно. Обратите внимание, что на рис. 4 изображена аминокислота из **4-ой модели**. На рис. 5 приведен пример, в котором все 16 моделей содержат некорректно вычисленную аминокислоту.

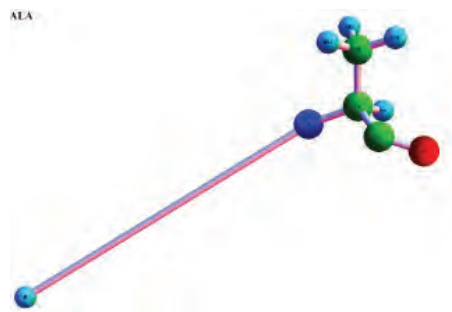


Рис. 4: Файл 2L5R.pdb, 4-ая модель, 19-ая аминокислота, 2010 год, список программ: TOPSP1 2.0, SPARKY 3.114, CYANA2.0.

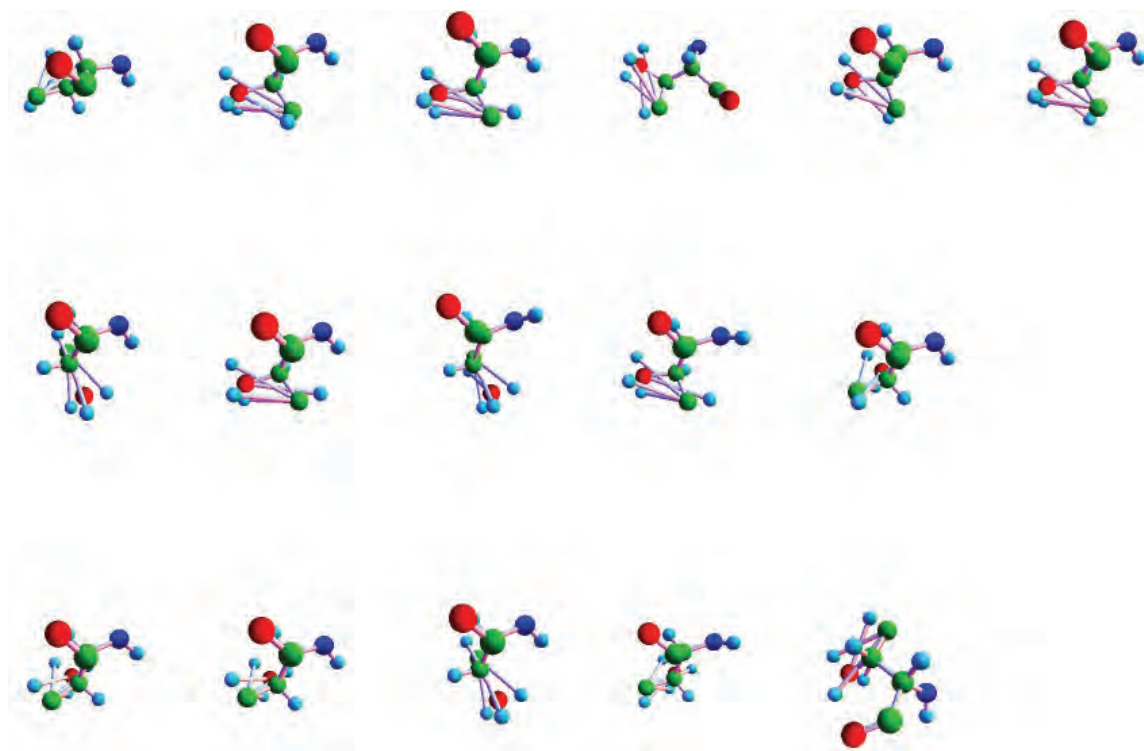


Рис. 5: Файл 1KWD.pdb (2002 год), в котором 10-ая аминокислота имеет некорректный вид во всех 16-ти моделях.

Справедливости ради заметим, что описанная только что сложность с первой моделью иногда действительно имеет место, рис. 6.

Обратим внимание еще на одну забавную особенность: в некоторых случаях визуализация на PDB “скрывает ошибки”, см. рис. 7.

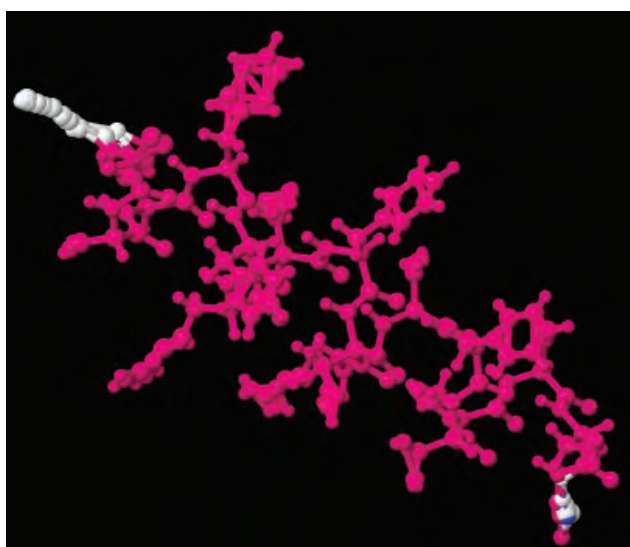


Рис. 6: Файл 2I2J.pdb, 2006 год, “плохая” только первая модель (остальные 33 модели — “хорошие”).

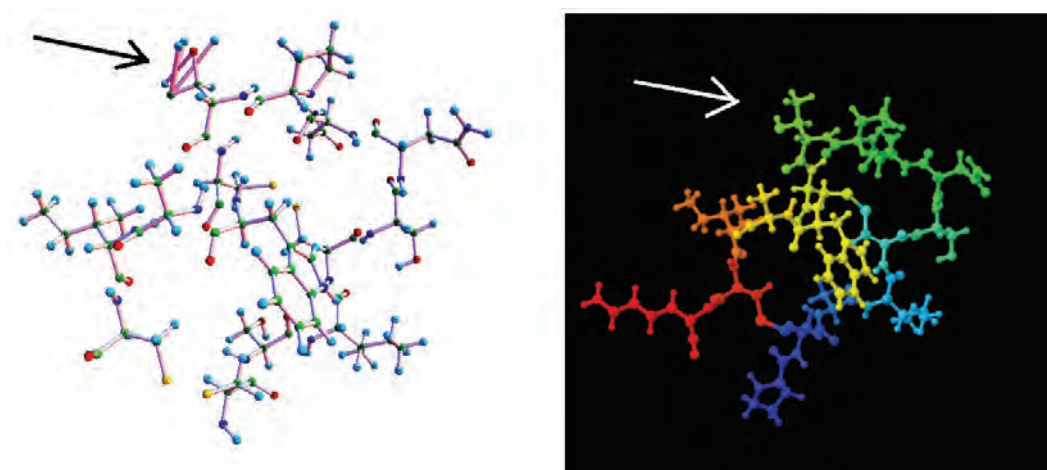


Рис. 7: Файл 1KWD.pdb, 2002 год, 10 аминокислота, неправильное обозначение водородов; на стандартном визуализаторе из PDB ошибка не видна.

Предыдущие обсуждения касались ЯМР. Многие специалисты согласились на том, что рентгено-структурный анализ (РСА) дает более точные результаты, причем тем точнее, чем меньше параметр RESOLUTION (разрешение). Чтобы понять, какое разрешение считать маленьким, мы построили график количеств файлов (ордината), имеющих те или иные значения RESOLUTION (абсцисса), см. рис. 8.

Из графика на рис. 8 видно, что максимальное число файлов соответствует разрешению

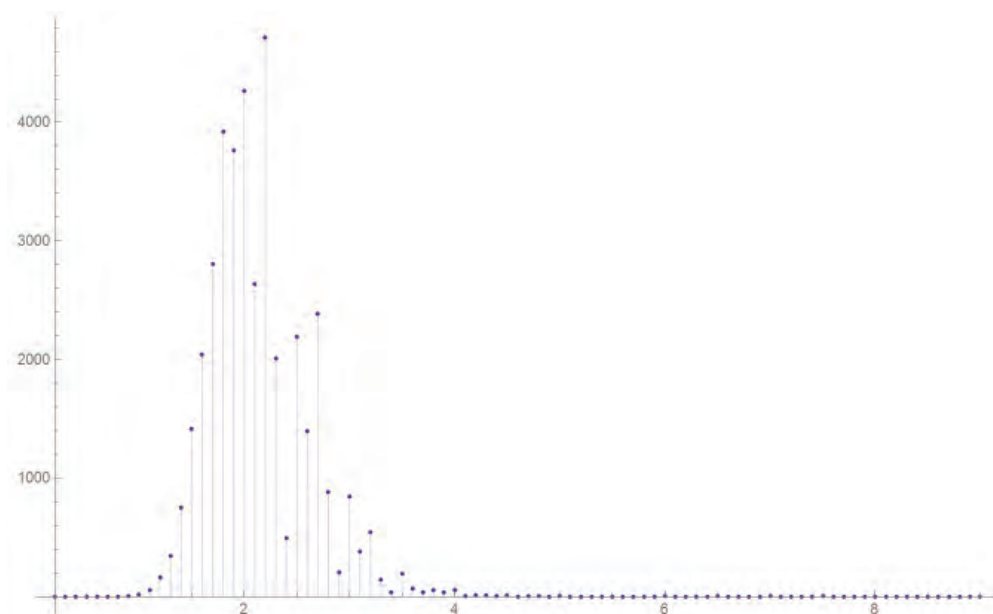


Рис. 8: График количеств файлов (ордината), имеющих те или иные значения RESOLUTION (абсцисса).

2.2, поэтому, учитывая разброс значений, разрешение 1.5 можно было бы отнести к маленьким. Приведенные рассуждения делают следующий пример контрпримером.

ARG

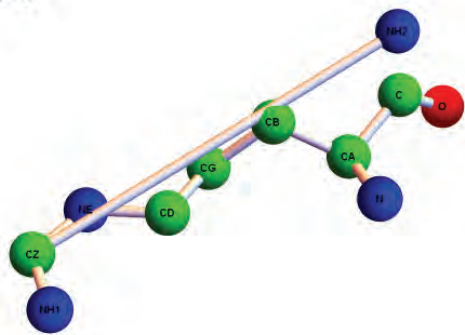


Рис. 9: Файл 1W5Q, 2004 год, RESOLUTION = 1.4.

Интересна также история ряда попыток публикации наших результатов. Мы начали с arXiv, где у нас размещены многочисленные наши математические работы, и где ни разу у нас не было проблем опубликовать очередной математический результат. Однако, к нашему искреннему удивлению, на сей раз статья принята не была: нам посоветовали обратиться в рецензируемый журнал. Когда же мы послали статью в один из таких журналов, рецензент сообщил нам, что информация о наличии ошибок в файлах из PBD давно и хорошо известна, и даже привел ссылки [5], [6], [7]. В первой из них, самой ранней (1996 год), в разделе “Correspondence”, содержится очень краткое, без единой иллюстрации, описание проделанной работы: говорится о проведенном вычислении средних длин ковалентных связей, средних углов между смежными ковалентными связями, отклонений от средних и т.п.; утверждается, что было найдено 76 классов разных проблем; приводится ссылка <http://www.sander.embl-heidelberg.de/johnny/>, видимо, на более подробное изложение; *в настоящее время эта ссылка не работает*.

Как видно из приведенного выше обсуждения, работа [5] осталась незамеченной рядом ведущих специалистов. Тем не менее, публикация такого типа позволяет “сведущему” рецензенту задать вопрос: а что нового есть в Вашей работе? И, естественно, нет никакой возможности объяснить, что вербальное сообщение о 76 проблемах в разделе “Correspondence” имеет меньше шансов на распространение, чем впечатление от рис. 1 и 2.

Итак, в предыдущих работах мы

- (1) вычислили наиболее часто встречающиеся “атомные составы” внутренних (неконцевых) аминокислот и отобрали только те файлы, в которых все внутренние аминокислоты такие;
- (2) выбросили все файлы с различными включениями (содержащие HETATM), описанными в разделе ATOM-TER;
- (3) выбросили все файлы со слишком большими отклонениями от средних длин ковалентных связей;
- (4) выбросили все файлы со слишком большими отклонениями от средних углов между смежными ковалентными связями.

Были также проанализированы частоты появления цис-конфигураций, степень плоскости пептидной группы, распределение расстояний между последовательными альфа-углеродами, подвижность аминокислот фиксированного типа.

Как уже было отмечено, мы получили “чистую” базу данных, состоящую для ЯМР примерно из 2500 файлов, а для РСА — из 30000 файлов. Одной из геометрических идей, которую мы хотели проверить на сей раз, является принадлежность ребер минимального остовного дерева, построенного на координатах атомов, графу ковалентных связей. Начав экспериментировать, мы обнаружили некоторые сбои: оказалось, что в ряде случаев ковалентно несвязанные ато-

мы находятся слишком близко друг к другу. Мы провели анализ не только длин ковалентных связей, как в работах [1] и [2], но и расстояний между всеми остальными парами атомов и, в результате, обнаружили еще ряд файлов, которые, скорее всего, следует отнести к содержащим ошибки. Именно об этом исследовании мы расскажем в настоящей статье. В следующей работе мы применим вычищенную с помощью описываемых здесь методов базу к исследованию экстремальной геометрии белков.

Автор выражает глубокую благодарность В.Л.Голо, фактически являющемуся для меня локомотивом написания биологических статей, а также моим коллегам и родственникам Г.А.Армееву, Е.А.Вилкул, А.О.Иванову, Ф.Ю.Попеленскому, Ж.Р.Тужилиной, К.В.Шайтану за многочисленные полезные обсуждения.

Работа выполнена при поддержке РНФ, соглашение № 14-50-00029.

2 Распределение интервалов длин разных типов ковалентных связей

В данной статье мы изучаем файлы, полученные исключительно с помощью ЯМР.

Эксперименты с минимальными остовными деревьями привели нас к необходимости выяснить, в каких пределах может меняться длина той или иной ковалентной связи, и какими бывают длины остальных расстояний (между несвязанными ковалентно атомами). Мы разбили все ковалентные связи на 100 видов, в соответствии с номенклатурой названий атомов в белках, и для каждого вида вычислили интервал между минимальным и максимальным значением длины этой связи по всем белкам. Вот начальный отрезок того, что мы получили:

```
{ "NZ" -> "HZ2", {0.9686681578332172, 1.084642337362875} }
{ "NZ" -> "HZ3", {0.9689220814905612, 1.0944066885760526} }
{ "CZ" -> "OH", {1.3299056357501453, 1.4159099547640697} }
{ "OH" -> "NN", {0.9379370981041326, 1.0296382860014484} }
```

Для наглядности, мы решили изобразить эти интервалы отрезками, пометив для каждого из них соответствующей внутренней точкой среднюю длину данной ковалентной связи (сделать последнее предложил К.В.Шайтан). Результаты представлены на рис. 10.

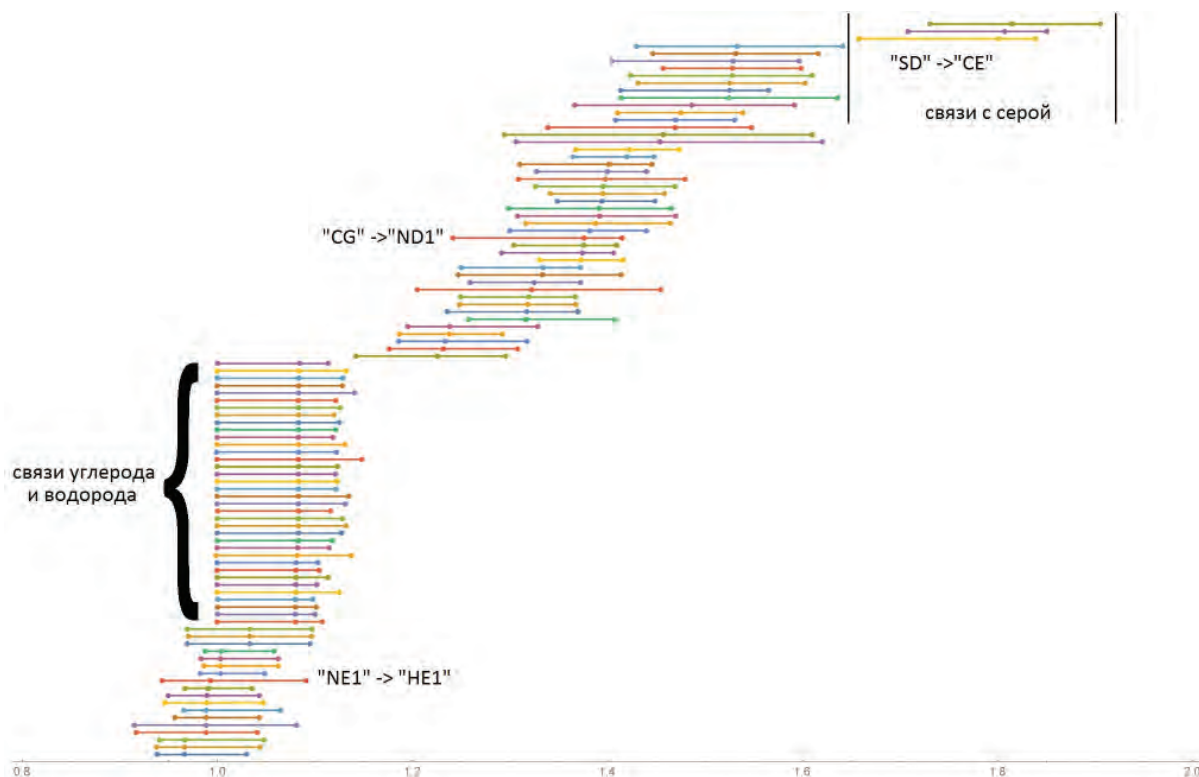


Рис. 10: Распределение интервалов длин ковалентных связей.

Отметим удивительную особенность полученной диаграммы: блок, соответствующий всем связям между углеродом и водородом, выровнен по левому краю. Г.А.Армеев высказал гипотезу, что этот эффект мог быть вызван жестким фиксированием минимальной длины этих связей в алгоритме нахождения конформации (для ряда файлов, в которых встретились самые короткие углеродно-водородные связи). Мы решили посмотреть, в каких файлах достигается эта левая граница.

Заметим, что во всем “углеродно-водородном блоке” левая часть интервала существенно больше правой. Этот эффект опять же может быть легко получен, если в одном из файлов имеются очень короткие углеродно-водородные связи. Мы решили отловить файлы, “виновные” в нарушении симметрии, и написали процедуру, которая выдает список всех файлов, содержащих связи, удлиняющие одну из частей интервала. Вот некоторые результаты для углеродно-водородного блока:

```
связь "CE3" -> "HE3":
"2MEX", {"CE3" -> "HE3", 0.999462}}
"2JRL", {"CE3" -> "HE3", 0.999725}}
"2DJM", {"CE3" -> "HE3", 1.033645}}
.....

связь "CH2" -> "HH2":
"2MEX", {"CH2" -> "HH2", 0.999708}}
"2JRL", {"CH2" -> "HH2", 1.000081}}
"1WRT", {"CH2" -> "HH2", 1.055652}}
.....

связь "CD1" -> "HD1":
"2MEX", {"CD1" -> "HD1", 0.999255}, {"CD1" -> "HD1", 0.999589}, ...}
"2JRL", {"CD1" -> "HD1", 0.999373}, {"CD1" -> "HD1", 0.999549}, ...}
"2OYW", {"CD1" -> "HD1", 0.999702}, {"CD1" -> "HD1", 1.00011}}
"2OYV", {"CD1" -> "HD1", 0.999925}, {"CD1" -> "HD1", 1.00053}}

связь "CD2" -> "HD2":
"2MEX", {"CD2" -> "HD2", 0.999551}, {"CD2" -> "HD2", 1.00003}, ...}
"2OYW", {"CD2" -> "HD2", 0.999709}, {"CD2" -> "HD2", 0.999807}}
"2JRL", {"CD2" -> "HD2", 0.999774}, {"CD2" -> "HD2", 0.999929}, ...}
"2OYV", {"CD2" -> "HD2", 0.999862}, {"CD2" -> "HD2", 1.00047}}
"1WM8", {"CD2" -> "HD2", 1.00014}}
"2CBH", {"CD2" -> "HD2", 1.02588}}
.....

связь "CG" -> "HG3":
"2MEX", {"CG" -> "HG3", 0.999049}, {"CG" -> "HG3", 0.999448}, ...}
"1WM8", {"CG" -> "HG3", 0.999391}, {"CG" -> "HG3", 0.999562}, ...}
"2JRL", {"CG" -> "HG3", 0.999454}, {"CG" -> "HG3", 0.999506}, ...}
"1ZUV", {"CG" -> "HG3", 0.999535}, {"CG" -> "HG3", 0.999798}, ...}
"2OYV", {"CG" -> "HG3", 0.999562}, {"CG" -> "HG3", 0.999576}, ...}
"2OYW", {"CG" -> "HG3", 0.999656}, {"CG" -> "HG3", 0.999974}, ...}
```

Отметим, что для некоторых связей списки оказались очень большими, не смотря на явное смещение среднего значения. Вот, например, начальный отрезок для связи "CA" -> "HA":

```
"1G10", {"CA" -> "HA", 0.998062}, {"CA" -> "HA", 0.998759}}
"1G11", {"CA" -> "HA", 0.998575}}
"2MEX", {"CA" -> "HA", 0.998962}}
"1X40", {"CA" -> "HA", 0.998964}}
"1WIZ", {"CA" -> "HA", 0.998968}}
```

```
"1CCV", {"CA" -> "HA", 0.998969}}
"1M3O", {"CA" -> "HA", 0.998984}}
"1X6D", {"CA" -> "HA", 0.998996}}
"1V5J", {"CA" -> "HA", 0.999057}}
"1V5T", {"CA" -> "HA", 0.999061}}
.....
```

Мы решили изобразить распределение количеств связей (ордината) с теми или иными длинами (абсцисса) для "CA" -> "HA", рис. 11.

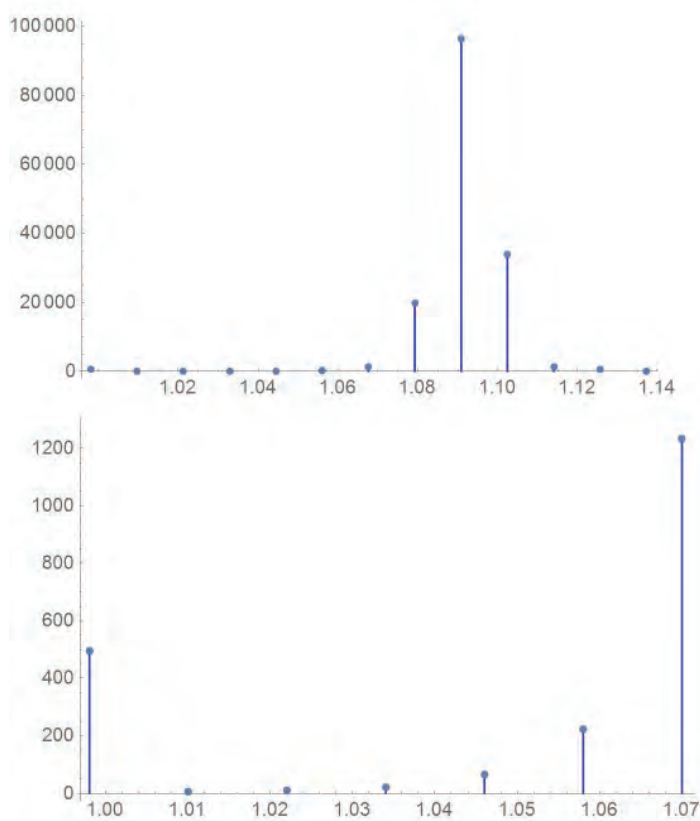


Рис. 11: Распределение количеств связей (ордината) с теми или иными длинами (абсцисса) для "CA"→ "HA".

Из диаграммы 11 видно, что в окрестности “длинного конца” интервала изменения длины связи "CA" -> "HA" сосредоточено очень большое число разных файлов, поэтому данный “перекос”, видимо, стоит считать адекватным.

Итак, список претендентов на удаление был получен следующим образом: в него мы поместили все файлы, в которых имеются связи длины не больше 1.01, и около левого конца интервала изменения длины связи сосредоточено “немного” файлов. В результате мы сформировали список из 7 элементов:

```
{"1WM8", "1WT8", "2JRL", "2MEX", "2OYV", "2OYW", "1ZUV"}.
```

Как легко видеть, на диаграмме 10 имеется еще много других типов связей, в которых возникает перекос, например, для связей "NE1" -> "HE1", "CG" -> "ND1", "SD" -> "CE".

Анализ, похожий на тот, который мы провели для "CA" -> "HA", показал, что наиболее подозрительной на наличие ошибок является связь "SD" -> "CE". Ниже мы приводим распределение количеств связей "SD" -> "CE" по длинам этих связей, рис. 12 и 13.

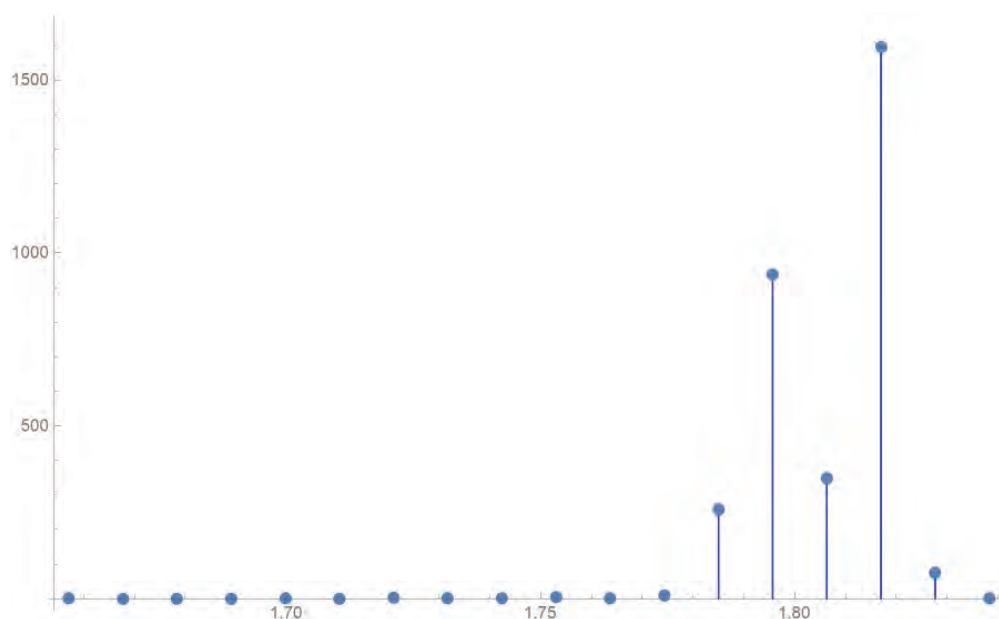


Рис. 12: Распределение количеств связей (ордината) с теми или иными длинами (абсцисса) для "SD" → "HA".

Из диаграммы 13 видно, что имеется ровно одна кратчайшая связь, относительно изолированная по своей величине. Непосредственное вычисление показывает, что эта связь принадлежит файлу 2LV1.pdb. Впрочем, относительно подозрительными являются также связи, длины которых расположены в пределах между 1.7 и 1.77 ангстрем. Не имея достаточных оснований считать последние файлы ошибочными, мы решили оставить их, а в список для исключения поместить только 2LV1.pdb.

Итак, этот этап “чистки” закончился исключением следующего списка файлов:

`{"1WM8", "1WT8", "2JRL", "2MEX", "2OYV", "2OYW", "1ZUV", "2LV1"}`.

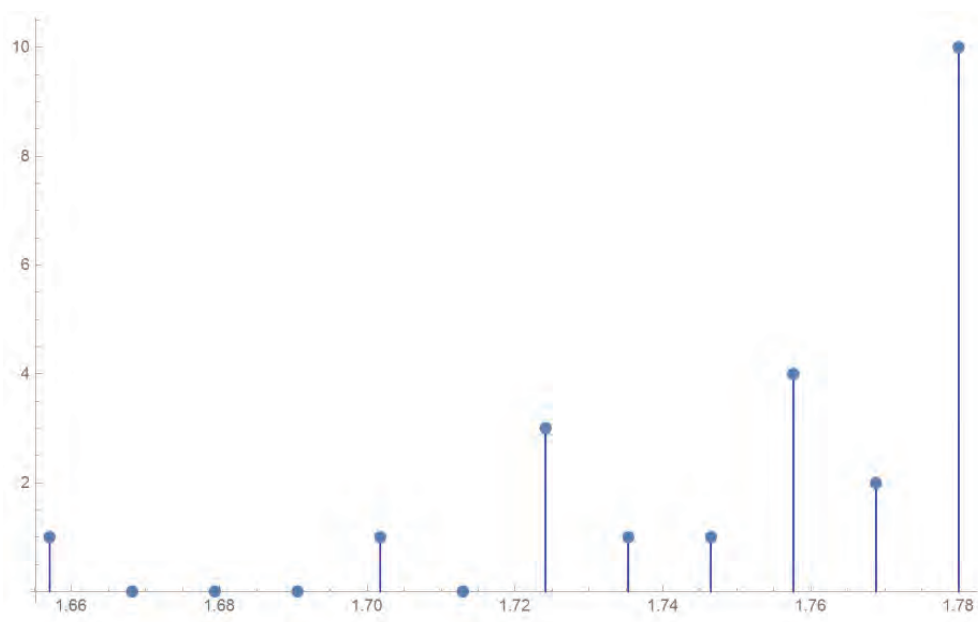


Рис. 13: Распределение количеств связей (ордината) с теми или иными длинами (абсцисса) для "SD" → "CE".

Мы пересчитали диаграмму распределения интервалов изменения длин связей, см. рис. 14.

Хорошо видно, что в результате мы избавились от выровненных левых концов блока связей углерод-водород, а также сделали распределение длин связи "SD" -> "CE" более симметричным.

3 Геометрия пар атомов в одной аминокислоте, не связанных ковалентно

В качестве следующего шага мы решили изучить расстояния между атомами, не связанными ковалентно. Мы начали с пар атомов, принадлежащих одной аминокислоте. На рис. 15 и 16 приведены диаграммы распределений количеств пар по расстояниям между ними.

Из диаграммы 16 видно, что имеется ровно одно очень маленькое расстояние, которое, как оказалось, равно 0.657274 ангстрем, принадлежит файлу 1Q2I.pdb и реализуется в 11-ой аминокислоте (лейцине) между атомами водорода "H" и "HD11".

Исключив из нашей базы файл 1Q2I.pdb, мы добились того, что расстояния между атомами в аминокислоте, не связанными ковалентно, стали не меньше 1.1 ангстрема.

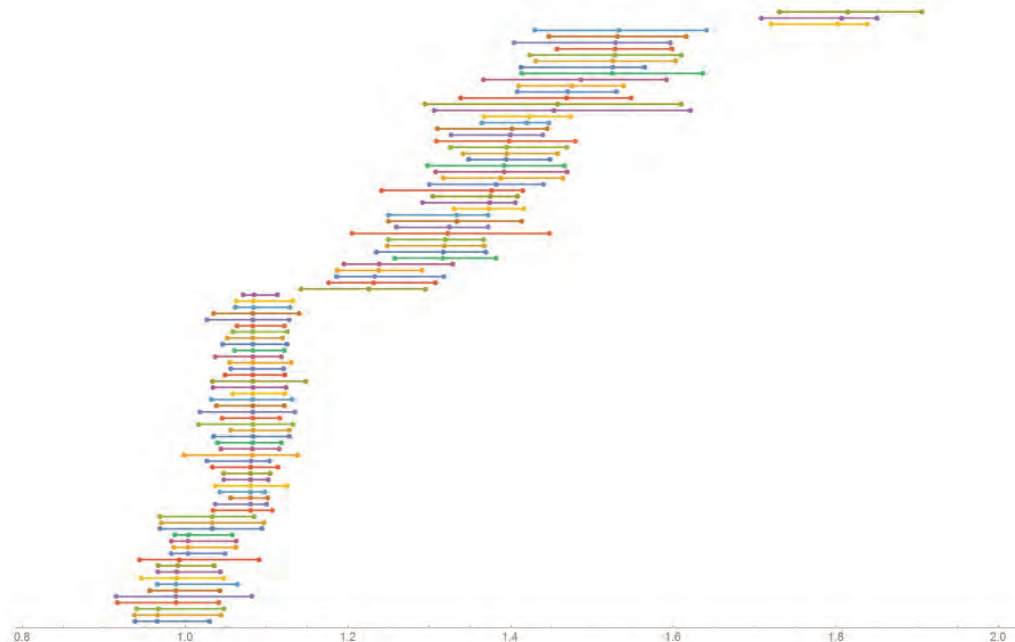


Рис. 14: Распределение интервалов длин ковалентных связей после первой “чистки”.

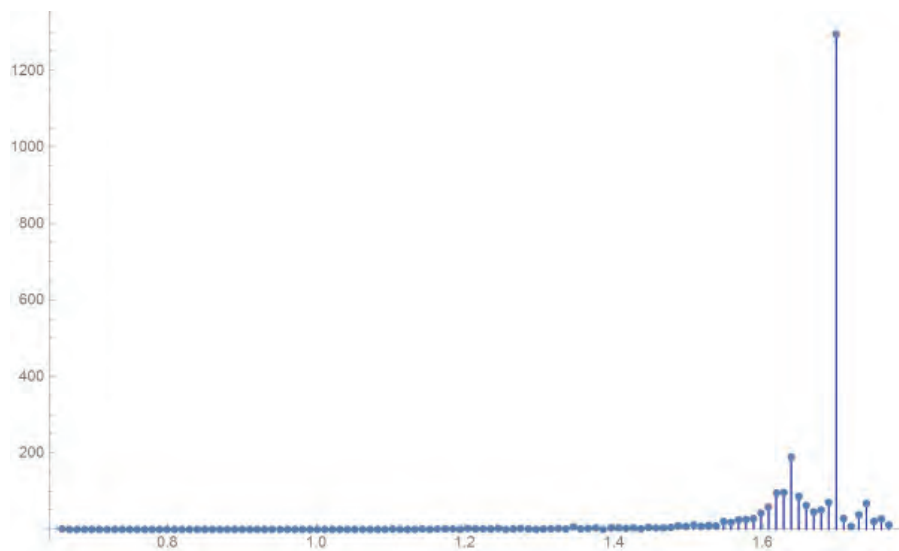


Рис. 15: Распределение количеств пар атомов из одной аминокислоты, не связанных ковалентно, по расстояниям между этими атомами.

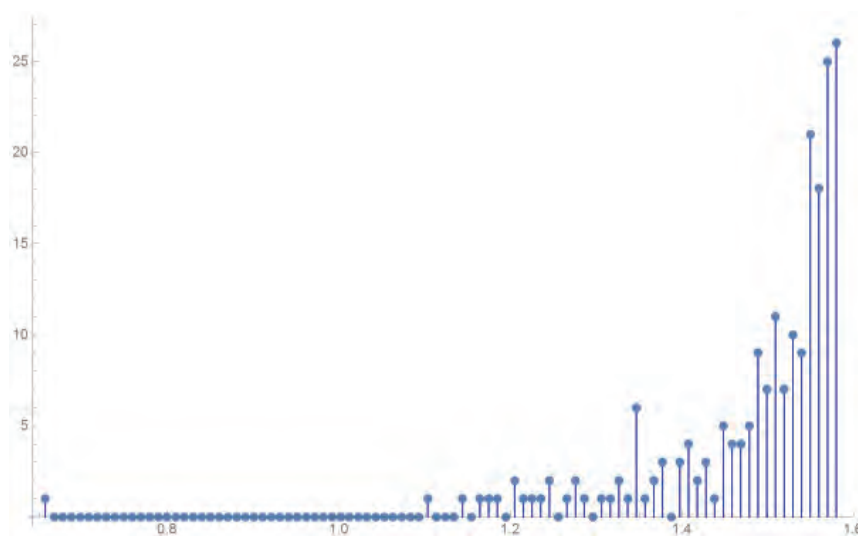


Рис. 16: Распределение количеств пар атомов из одной аминокислоты, не связанных ковалентно, по расстояниям между этими атомами.

4 Геометрия пар атомов в разных аминокислотах, не связанных ковалентно

Заключительный шаг “чистки” состоял в изучении расстояний между атомами, принадлежащими разным аминокислотам и несвязанным ковалентно (единственными ковалентными связями между такими атомами являются пептидные связи "C" -> "N").

Заметим, что таких пар очень много, поэтому прямой перебор всех пар приводит к большим временным затратам. Учитывая нашу дальнейшую цель изучения геометрии минимальных остовных деревьев, нам нужны расстояния, не превосходящие максимальную длину ковалентных связей, которая, как видно из диаграммы 14, не превосходит 2 ангстрем. Следующий трюк позволяет эффективно выделить все пары с расстояниями не больше 2 ангстрем.

- (1) Вычислим геометрические центры аминокислот.
- (2) Посчитаем “радиусы” аминокислот, определив их как наибольшие расстояния от центров до атомов аминокислоты. Оказалось, что максимальный радиус по всем аминокислотам всех рассматриваемых белков равен примерно 5.575 ангстрем.
- (3) Вычислим расстояния между всеми парами центров.
- (4) Выберем только те пары центров, которые расположены друг от друга не далее чем на $2 * 5.575 + 2$ ангстрема. Тем самым, мы добьемся того, что каждая пара атомов, расположенных друг относительно друга не далее чем на 2 ангстрема, лежит в некоторой паре аминокислот, чьи центры мы выбрали.
- (5) Образует список всех пар атомов по всем парам выбранных аминокислот.
- (6) Выкинем оттуда все пептидные связи.
- (7) Выберем из оставшихся пар все пары с расстояниями, не превосходящими 2 ангстрем.

Конечно, этот алгоритм можно улучшить, но для наших целей его оказалось вполне достаточно.

Приведем результаты, демонстрирующие распределения количеств связей по их длинам для выбранных выше пар, рис. 17 и 18.

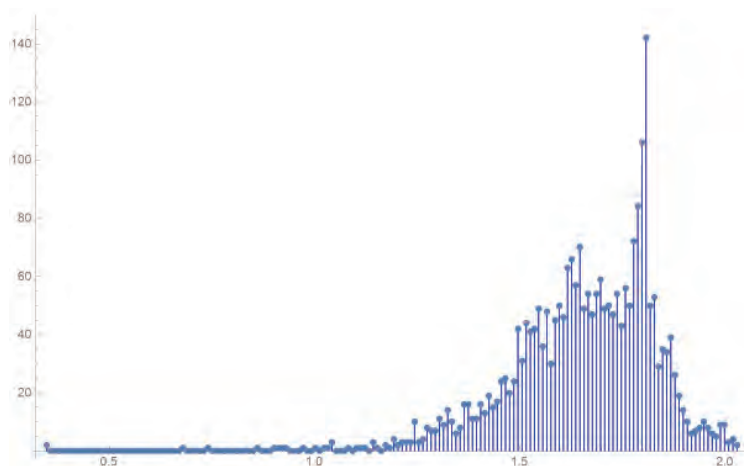


Рис. 17: Распределение количеств пар атомов из разных аминокислот, не связанных ковалентно, по расстояниям между этими атомами.

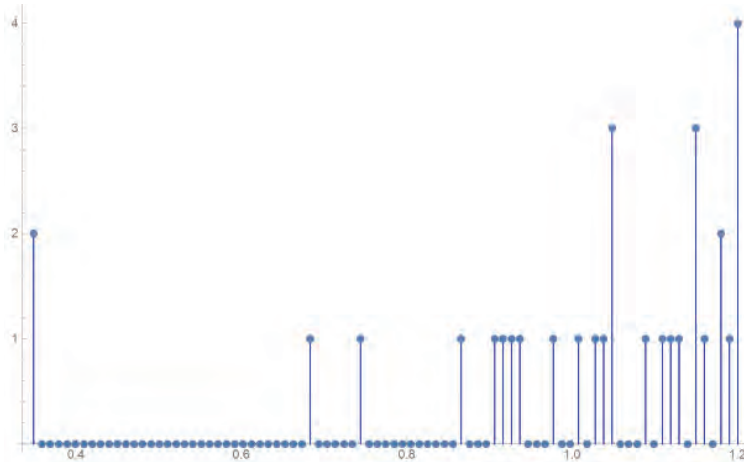


Рис. 18: Распределение количеств пар атомов из разных аминокислот, не связанных ковалентно, по расстояниям между этими атомами.

Из рис. 18 видно, что существуют ровно две совсем короткие связи. Их длины равны 0.356596 и 0.348223 ангстрем, и встречаются они в файлах 1APC.pdb и 2G0K.pdb соответственно. Все остальные расстояния не меньше 0.7 ангстрем.

Итак, окончательный список файлов, который мы решили выкинуть из нашей базы, такой:

1WM8, 1WT8, 2JRL, 2MEX, 2OYV, 2OYW, 1ZUV, 2LV1; 1Q2I; 1APC, 2G0K.

Так как после итоговой “чистки” диаграмма 14 практически не изменилась, мы не будем приводить ее здесь. Вместо этого мы приведем более компактную диаграмму изменений длин типов ковалентных связей, отождествив друг с другом разные водороды (H, HA, HB2, HB3, и т.д.), разные углероды, разные азоты и т.д. Результат показан на рис. 19.

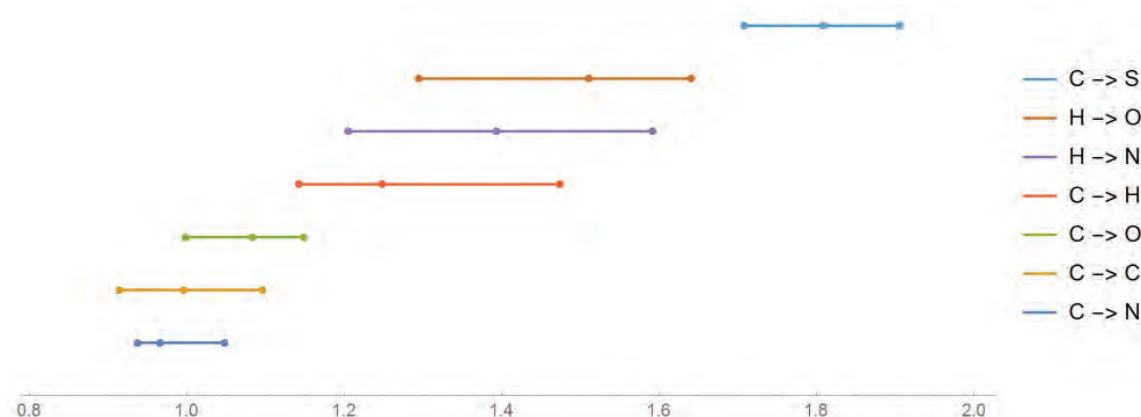


Рис. 19: Интервалы значений разных типов ковалентных связей.

В следующей статье мы расскажем о том, как очищенная нами база была применена для изучения связи экстремальных графов, построенных на координатах атомов белка, и графа ковалентных связей.

ВЫВОД. Для получения более надежной базы пространственных координат белков оказалось недостаточным изучение только средних значений и отклонений от них длин ковалентных связей и углов между смежными ковалентными связями. Предложена и реализована более тонкая методика “чистки”, что привело к более надежной базе, подготовленной для дальнейшего статистического исследования геометрии белков.

Список литературы

- [1] Иванов А.О., Мищенко А.С., Тужилин А.А. Геометрия аминокислот и полипептидов // Наноструктуры. Математическая физика и моделирование, 2014, **10**(1), 49–76; <http://nano-journal.ru>
- [2] Ivanov A.O., Mishchenko A.S., Tuzhilin A.A. Critical analysis of amino acids and polypeptides geometry // In Continuous and Distributed Systems: Theory and Applications, Springer, 2015, **2**, 29–74; <http://springer.com>
- [3] Ivanov A.O., Mishchenko A.S., Tuzhilin A.A. Geometry of space curves and applications to polymers conformations investigation // Educational Internet-Journal “Computer Graphics & Geometry”, 2007, **9**(2), 43–63.
- [4] Иванов А.О., Мищенко А.С., Тужилин А.А. Геометрия ломаных и полипептидов // Наноструктуры. Математическая физика и моделирование, 2014, **10**(1), 39–48; <http://nano-journal.ru>

- [5] *Hootd R.W.W., Vriend G., Sander Ch., Abola E.E.* Errors in protein structures // *Nature*, 1996, **381**, 272.
- [6] *Nabuurs S.B., Spronk Ch.A.E.M., Vuister G.W., Vriend G.* Traditional Biomolecular Structure Determination by NMR Spectroscopy Allows for Major Errors // *PLoS Comput Biol.*, 2006, **2**(2), 71–79.
- [7] *Joosten R.P., Joosten K., Cohen S.X., Vriend G., Perrakis A.* Automatic rebuilding and optimization of crystallographic structures in the Protein Data Bank // *Bioinformatics*, 2011 **27**(24), 3392–3398.

PROTEIN DATA BANK: NEXT CLEANING

A.A. Tuzhilin

*Faculty of Mechanics and Mathematics,
Lomonosov Moscow State University*

tuz@mech.math.msu.su

Received 14.05.2016

In the present paper we continue to investigate the distance geometry of proteins atoms. For the atoms coordinates database, we take Protein Data Bank (PDB) files. We present a few new ideas to test PDB-files for errors. As a result, we construct a new database which will be used in our next work for understanding the relation between the covalent bonds graph and some other extreme graphs like minimum spanning tree. This work continues our previous investigation of the geometry of polygonal lines constructed on coordinates of consecutive alpha-carbons.

Информация и правила для авторов

Общие положения

Журнал «Наноструктуры. Математическая физика и моделирование» (сокращенно: НМФМ) публикуется с 2009 года и является периодическим научным изданием. Электронная версия журнала размещается на сайте <http://www.nano-journal.ru>. Основная цель издания: представление новых теоретических и вычислительных методов моделирования наноструктур и мягкой материи, общих подходов в исследовании мезосистем, а также ключевых экспериментальных результатов в данной области и связанных с этим проблем математической физики.

Журнал НМФМ имеет междисциплинарный характер и в силу этого несет определенную образовательную направленность, а не только узко научную. Работы, представляемые в журнал, должны содержать вводные сведения, которые обеспечат понимание постановок задач и восприятие результатов не только прямыми специалистами. Определения понятий, объяснение обозначений и терминов, оценки характерных параметров, теоретические предпосылки и идеи, используемые методы, и т.п., должны быть кратко объяснены в тексте статьи, имея в виду читателей, специализирующихся в иных направлениях. Должны быть описаны базовые математические модели и уравнения. Во Введении и в последующих разделах очерчивается стратегия и основные трудности, это увязывается с используемыми моделями. Структура статьи ориентируется на прояснение общей логики и методики исследования, содержит резюмирующие выводы. В тексте должны быть рассмотрены характерные примеры (хотя бы, методические), ясно иллюстрирующие предлагаемые алгоритмы.

Журнал публикует научные обзоры, исследовательские статьи и краткие научные сообщения, а также избранные аналитические и информационно-образовательные материалы, тексты докладов и циклов лекций, прочитанных в университетах, научных центрах, на школах-семинарах, конференциях, нигде ранее не публиковавшиеся и не принятые к публикации в других изданиях. Язык публикации в журнале НМФМ, как правило, русский. Работы, представляемые в журнал, не могут иметь научно-популярный или компилятивный характер. Все статьи рецензируются и могут быть отклонены редколлегией журнала. В случае принятия работы к печати ее авторы передают издателью журнала НМФМ право на разовую безвозмездную публикацию текста и его размещение в электронной версии на сайте журнала. Перевод опубликованных в журнале статей на другие языки может осуществляться только с разрешения и при участии авторов.

Порядок представления статей

- В редакцию изначально представляются:
 - файл статьи, файлы с иллюстрациями;
 - сопроводительное письмо, можно в электронной форме, содержащее сведения об объеме статьи и обо всех авторах (фамилии, имена, отчества, полные названия мест работы, почтовый адрес с индексом, номер контактного телефона с кодом города, электронный адрес автора, ответственного за переписку с редакцией); предпочтительно, чтобы это письмо было выполнено на бланке учреждения, в котором работает кто-то из авторов, было заверено печатью и содержало утверждение о возможности открытого опубликования статьи;
 - файл с переводом на английский язык названия статьи, фамилий и инициалов авторов, аннотации, ключевых слов.
- Авторские файлы могут быть присланы на электронный адрес: papers@nano-journal.ru; (резервный адрес в случаях затруднений с пересылкой: nano@miem.edu.ru) или переданы в редакцию на любом электронном носителе. Авторы получают из редакции подтверждение о получении их материалов.
- Телефон (факс) редакции: +7 (495) 916-8876. Адрес редакции: Москва 109028, Б. Трехсвятительский пер., 3/12, Московский институт электроники и математики (МИЭМ), комн. 449.

Общие требования к представляемым файлам

- Допускается использование текстовых редакторов WORD и LATEX. К рабочим файлам должна быть приложена их pdf-копия. В названии файлов используется латинский алфавит, пробелы заменяются знаком `_`. Шапка статьи содержит название, инициалы и фамилии авторов, место работы, электронный адрес, краткую аннотацию, ключевые слова. В аннотации не следует использовать формулы и ссылки на текст работы или список литературы; в конце она должна содержать индекс УДК (к английской версии аннотации можно добавить индексы зарубежных рубрикаторов).
- Объем кратких сообщений 4-8 страниц, исследовательских статей, как правило, до 20 страниц, а обзоров – более 20 страниц. Верхняя граница согласуется с редколлегией. При подсчете объема нужно ориентироваться на страницы формата А4, шрифт 12, знаков в строке 80, интервалов между строками 1.
- Авторы не должны злоупотреблять сокращениями, составленными из заглавных начальных букв терминов. Предпочтительней каждый раз использовать полное наименование объекта. Возможно использование только устоявшихся аббревиатур.

Требования к файлам Word

- Рекомендуются шрифт – Times New Roman.
- Строки в пределах абзаца не должны разделяться символом возврата каретки (Enter).
- Нельзя использовать автоматическое создание сносков, автоматический перенос или автоматический запрет переносов, создание списков, автоматический отступ и т.п.
- Ссылки на список литературы даются цифрами в квадратных скобках: [1], [5,6,7], [1-9].
- Все без исключения формулы и обозначения размерности, даже состоящие из одной латинской буквы, и в тексте и вынесенные в отдельную строку, всегда набираются в формульном редакторе и никогда – в обычном текстовом редакторе.

- При создании таблицы рекомендуется использовать возможности Word или MS Excel. Таблицы, набранные вручную (с помощью большого числа пробелов), не принимаются.

Требования к иллюстрациям

- Иллюстрации представляются в отдельных файлах, черно-белыми. Они должны иметь разрешение не менее 600 dpi.
- Форматы файлов – TIFF, EPS, PSD, JPEG.

Требования к списку литературы

- Ф.И.О. авторов или редакторов выделяются курсивом.
- Для статей приводится название. Названия отделяются от выходных данных знаком //. Расположение выходных данных указано на образце ниже. Номер тома выделяется жирным шрифтом, номер выпуска дается в скобках. Указываются номера первой и последней страниц статьи, либо уникальный номер статьи и ее объем. Для книг желательно указывать их объем. Если известна ссылка на электронный архив или сайт, то ее желательно указать.

Фамилия И.О. Название статьи // Назв. журн., 2000, **1** (1), 1-6.

Family F.M. and Family F. Title of the paper // Name of the Journal, 2006, **73**, 165313, 9 pp.

Фамилия И.О., Фамилия И.О. Название книги // Наука, С.-П., 1999, 176 стр.

Family F.M. Title of the paper // In book: *Family F.M. (et al. eds)*, Title of the collection, Publisher, Boston, 2005, 9-24.

Family F.M. (ed.), Title of the collection // Publisher, N.Y., 2005, 324 pp.

Фамилия И.О. Название доклада // Доклад на конференции «Название конференции (место и дата проведения)»; ссылка на электронный ресурс.

Наноструктуры. Математическая физика и моделирование

Журнал зарегистрирован

в Министерстве РФ по делам печати,
телерадиовещания и средств массовых коммуникаций.

Свидетельство о регистрации

ПИ № ФС77-34934 от 29 декабря 2008 г.

Учредители

Московский государственный институт электроники и математики (МИЭМ),

ФГУ ГНИИ ИТТ «Информика»,

Европейский центр по качеству

Председатель Совета Учредителей

Тихонов А.Н.

Издатель

Европейский центр по качеству

ПОДПИСКА НА ЖУРНАЛ НМФМ

На первое полугодие 2016 г. подписаться на журнал можно в любом отделении связи по каталогу
Агентства Роспечать «Журналы России»,

рубрика «Физико-математические науки», подписной индекс 70017.

Редакция предлагает подписчикам возможность безвозмездно получить подборку прошлых
выпусков журнала. Пришлите на электронный адрес nano@miem.edu.ru (или на почтовый адрес:
123458, Москва, ул. Таллинская, д. 34, каб. 429, редакция журнала НМФМ)

копию подписной квитанции, а также адрес для отсылки выпусков.