ISSN 2224-8412

21(1) 2020

НАНОСТРУКТУРЫ

математическая физика и моделирование

Nanostructures. Mathematical physics & modelling

2020, том 21, № 1

НАНОСТРУКТУРЫ

математическая физика и моделирование

Nanostructures.
Mathematical Physics & Modelling

2020, volume 21(1)

Наноструктуры.
Математическая физика и моделирование

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

Содержание

Актуальные публикации прошлых лет

Contents

СТРУКТУРЫ БЛИЗОСТИ И ПРОСТРАНСТВА СХОДИМОСТИ

Е. М. Беркович

Главный редактор журнала «Семь искусств»
redaktor@7iskusstv.com

Поступила 08.11.2020

В работе рассматриваются пространства сходимости, в которых за основу берутся сходящиеся последовательности, а другие топологические свойства множеств и функций выводятся из этого понятия. Предел последовательности при этом может быть неединственным. Построенные таким образом пространства обладают рядом специфических свойств, в них, в частности, известную теорему Вейерштрасса можно сформулировать как необходимое и достаточное условие существования решения экстремальной задачи. Обсуждается применение пространств сходимости в исследовании задач оптимального управления.
Ключевые слова: сходимость, счетная компактность, замкнутость, непрерывность, полунепрерывность снизу, экстремум, оптимальное управление, теорема Вейерштрасса.

УДК 517.972.4, 515.122.8, 517.977.57.
DOI: 10.31145/2224-8412-2020-21-1-5-18

В разнообразных задачах моделирования физических объектов и систем исследователь нередко сталкивается с необходимостью оценить, насколько изучаемые объекты и системы близки или похожи друг на друга. С античных времен близость точек на прямой или плоскости оценивалась расстоянием между ними. Обобщение понятия расстояния на произвольные множества позволило создать крупный раздел математики – теорию метрических (см., например, «Энциклопедический словарь расстояний» [1]), а затем и еще более общих – топологических пространств (см., например, популярное введение в общую топологию [2]). В этих пространствах исходными являются понятия метрики или топологии, т. е. категории близости, а из них уже строятся другие связанные с ними конструкции: сходящиеся последовательности, компактные, открытые и замкнутые множества, непрерывные функции и так далее. Однако не всегда такая логика построения теории адекватна рассматриваемой задаче. Есть немало прикладных задач, в которых естественнее начинать со сходящихся последовательностей, а связанные с категорией близости свойства множеств и функций выводить из них, не привлекая сюда никаких метрик или топологий.

Впервые такой подход применил французский математик Морис Фреше в работе 1906 года [3], а затем эта идея была развита Казимиром Куратовским и описана им в монографии [4]. Фреше обозначал введенные им пространства буквой $\mathscr{L}$. Куратовский, уточнив это понятие, добавил звездочку: $\mathscr{L}^*$. Нами в 1972 и 1975 годах предпринята попытка обобщить подход Куратовского, допустив неединственность предела у сходящихся последовательностей [5] и [6]. Соответствующие пространства мы обозначаем через $\mathscr{L}^+$. Общая теория пространств сходимости строится на базе понятия фильтра, введенного Анри Картаном в 1937 году [7]. Основополагающей в этом направлении следует считать работу Густава Шоке, опубликованную в томе «Анналов университета Гренобля», датированном 1947-1948 годами [8]. Обзор современного состояния этой ветви теории можно найти в работах Шимона Далецки [9] и коллектива авторов [10].

В настоящей работе мы развиваем подход работ [5] и [6] и приводим некоторые новые результаты о необходимых и достаточных условиях существования решений экстремальных задач.

Считаю приятным долгом поблагодарить доктора физико-математических наук, профессора В.М. Тихомирова, доктора технических наук, профессора Б.Т. Поляка, доктора физико-математических наук, профессора А.Б. Успенского, прочитавших первоначальный вариант статьи и сделавших ценные замечания.

Сходимость

Прежде чем мы приступим к изложению формальной части статьи, постараемся ответить на естественный вопрос: зачем вводить новую математическую структуру, когда давно имеются изученные с разных сторон метрические и топологические пространства, в которых вполне уютно себя чувствуют сходящиеся последовательности и все другие связанные с ними свойства множеств и функций? На этот вопрос есть, по меньшей мере, два ответа. Во-первых, с чисто математической точки зрения развитие новой структуры – пространства сходимости – сулит много изящных и неожиданных результатов. В таких пространствах все функции могут оказаться непрерывными, а все последовательности – сходящимися. Здесь легко дать необходимое и достаточное условие существования экстремума и многое другое. Второй ответ связан с практическими приложениями, где новая структура смотрится более адекватной, чем старые.

Приведем простой пример: конструирование некоторой сложной системы, характеризующейся большим набором разнообразных технических параметров. Сравнивать близость различных конструкторских решений, опираясь только на эти параметры, малоперспективно. Трудно задать адекватную метрику на пространстве всех возможных решений, даже если каждый параметр решения задается конкретным числом. Поэтому сложно понять, в каком смысле некоторая последовательность конструкторских решений «сходится» к «оптимальному». Но можно пойти в другом направлении и определить последовательность решений сходящейся к какому-то «терминальному» решению, если последовательность значений критерия качества решений сходится, тогда задача приобретает смысл. Правда, предел некоторой последовательности решений, понимаемый в указанном смысле, перестает быть единственным. Ведь одно и то же значение критерия качества можно достичь на разных технических решениях. Так мы приходим к идее «пространств сходимости», в которых предел последовательностей может быть неединственным.

Мы будем обозначать введенные нами пространства $\mathscr{L}^+$, изначально допуская неединственность предела сходящейся последовательности. Приведем точное определение.

Определение. Пусть X – произвольное множество. Скажем, что оно является *пространством сходимости*, если среди всех последовательностей элементов множества X выделен класс так называемых *сходящихся последовательностей*, каждой из которых поставлено в соответствие непустое подмножество X, называемое *множеством ее пределов* (говорят также, что последовательность *сходится* к каждому элементу этого множества), причем выполнены следующие условия:

1) Если последовательность $\{x_n\}$, начиная с некоторого номера n_0, стационарна: $x_n = x_0$ при $n > n_0$[1], то она сходится, причем x_0 принадлежит множеству ее пределов.

2) Любая подпоследовательность сходящейся последовательности тоже сходится, причем множество пределов последовательности содержится в множестве пределов подпоследовательности.

3) Если последовательность не сходится к некоторому элементу множества X, то найдется такая ее подпоследовательность, никакая подпоследовательность которой не сходится к этому элементу.

[1] Такие последовательности будем называть квазистационарными.

Множество пределов последовательности $\{x_n\}$ будем обозначать выражением $LIMx_n$. Если x_0 – один из пределов последовательности $\{x_n\}$, то будем, как обычно, обозначать это так: $x_n \rightarrow x_0$.

В пространствах $\mathscr{L}^*$ Куратовского справедливы те же условия, но постулируется единственность предела каждой сходящейся последовательности [4 стр. 197]. Наше определение снимает это ограничение.

Сформулированные требования могут показаться непривычными, если сравнивать со сходящимися последовательностями в метрических или $\mathscr{L}^*$-пространствах. Там предел сходящейся последовательности единствен. И любая подпоследовательность сходящейся последовательности тоже сходится к тому же пределу, т. е. множества пределов последовательности и любой ее подпоследовательности совпадают – они состоят из одного и того же элемента.

Если последовательности составлены из элементов конечного множества X, то в метрических и $\mathscr{L}^*$-пространствах сходящимися могут быть только квазистационарные последовательности. В самом деле, предположим, что есть еще одна сходящаяся последовательность, в которой два элемента множества X встречаются бесконечное число раз. Если эта последовательность сходится к некоторому пределу, то к нему должна сходиться любая подпоследовательность. У рассматриваемой последовательности есть две стационарные подпоследовательности, каждая составленная из своего элемента множества X. К ним подпоследовательности и сходятся. Следовательно, у рассматриваемой последовательности минимум два предела. А этого быть не может. Следовательно, утверждение доказано.

В $\mathscr{L}^+$-пространствах положение другое. В них может быть случай, когда все последовательности, составленные из элементов конечного множества X, сходятся. Для этого достаточно назвать пределом любой последовательности произвольный элемент множества X. В этом случае несходящихся последовательностей просто нет, поэтому третье условие определения структуры сходимости выполняется автоматически, а первые два условия проверяются элементарно. Можно привести и другие примеры.

Рассмотренная структура в общем случае называется структурой тривиальной сходимости, или *тривиальной* $\mathscr{L}^+$*-структурой*.

Определение. *Структура тривиальной сходимости* на множестве X предполагает, что любая последовательность $\{x_n\}\subset X$ сходится, причем множество ее пределов совпадает со всем пространством: $LIMx_n = X$.

Очевидно, что в тривиальной структуре самое большое количество сходящихся последовательностей, которые можно построить из элементов заданного множества: ведь несходящихся последовательностей в ней вообще нет. Если ввести сравнение структур сходимости, как это сделано в нижеследующем определении, то тривиальная $\mathscr{L}^+$-структура – *самая слабая* из всех возможных структур сходимости на множестве X.

Определение. Пусть на множестве X заданы две $\mathscr{L}^+$-структуры τ и σ. Будем говорить, что структура сходимости τ *сильнее* структуры сходимости σ (или σ *слабее* τ), если каждый предел произвольной последовательности в смысле структуры τ (сокращенно τ-*предел*) является одновременно и σ-пределом. Другими словами, справедливо включение: $\tau\text{-}LIMx_n \subset \sigma\text{-}LIMx_n$.

Образно можно сказать, что сильносходящихся последовательностей *меньше*, чем слабосходящихся.

На шкале всех структур сходимости, определенных на заданном множестве X, тривиальная структура занимает один полюс – она самая слабая. А другой полюс представляет *самая сильная* структура, так называемая *структура дискретной сходимости*. В ней сходящимися являются только квазистационарные последовательности, причем сходятся они к тому элементу, из которого, начиная с некоторого номера, состоят. Первые два условия структуры сходимости очевидны. Покажем, что выполнено третье условие. Пусть последовательность $\{x_n\}$ не сходится к некоторому элементу $x_0 \in X$. Удалим из этой последовательности все элементы x_0, если они в ней присутствуют. Тогда должно остаться бесконечное число членов, иначе бы последовательность $\{x_n\}$ считалась бы квазистационарной и, по определению, сходилась бы к x_0. Никакая подпоследовательность оставшейся подпоследовательности не может сходиться к x_0, так как в ней нет ни одной квазистационарной подпоследовательности, состоящей из этого элемента. Таким образом, дискретная структура сходимости действительно определяет $\mathscr{L}^+$-пространство. Если x_0 – предел некоторой последовательности $\{x_n\}$ в смысле дискретной сходимости, то этот же элемент будет пределом и в смысле любой более слабой сходимости.

Отметим, что даже $\mathscr{L}^*$-пространство может не быть топологическим [4 стр. 199]. Тем более и $\mathscr{L}^+$-пространства не сводятся к топологическим.

А теперь рассмотрим, как в $\mathscr{L}^+$-пространствах вводится понятие непрерывности функций.

Непрерывность

Пусть заданы два $\mathscr{L}^+$-пространства X и Y и определена функция $f{:}X \to Y$. Обозначим через τ и σ структуры сходимости на множествах X и Y соответственно.

Определение. Функция f называется *непрерывной* относительно структур τ и σ (или τ,σ-*непрерывной*), если для любой сходящейся последовательности $\{x_n\} \subset X$ справедливо включение $f(LIMx_n) \subset LIMf(x_n)$.

Другими словами, из того, что элемент x_0 является одним из пределов последовательности $\{x_n\} \subset X$, следует, что $f(x_0)$ является одним из пределов последовательности $\{f(x_n)\} \subset Y$.

Непрерывность функции зависит от выбора структур сходимости на областях ее определения и значений. Если на множестве X задана структура дискретной сходимости, то любая функция, заданная на этом множестве, будет непрерывной независимо от структуры сходимости на множестве Y. В самом деле, для дискретной сходимости τ утверждение, что x_0 является одним из пределов последовательности $\{x_n\}\subset X$, означает, что $\{x_n\}$ – квазистационарная последовательность, состоящая, начиная с некоторого номера, из элемента x_0. Но тогда и последовательность $\{f(x_n)\}$ оказывается квазистационарной, и $f(x_0)$ является одним из ее пределов.

Справедливо и симметричное утверждение. Если структура сходимости σ на множестве Y значений функции тривиальная, то любая функция $f{:}X\to Y$ становится непрерывной независимо от структуры сходимости на множестве X. В самом деле, в силу тривиальности структуры сходимости на множестве Y последовательность $\{f(x_n)\}\subset Y$ становится сходящейся, как и любая другая последовательность из Y, причем пределом ее является любой элемент множества Y, в том числе и $f(x_0)$, где x_0 – один из пределов произвольной последовательности $\{x_n\}\subset X$. Что и требовалось доказать.

Если на множестве X заданы две структуры сходимости τ_1 и τ_2, причем τ_1 сильнее, чем τ_2, то из слабой непрерывности функции $f{:}X\to Y$, т.е. из τ_2-непрерывности, следует ее сильная непрерывность, т.е. τ_1-непрерывность. В самом деле, пусть произвольная последовательность $\{x_n\}\subset X$ сходится к элементу x_0 в смысле структуры τ_1. В силу определения сильной и слабой сходимостей элемент x_0 является пределом последовательности $\{x_n\}$ и в силу более слабой структуры сходимости τ_2. А в силу слабой непрерывности функции f последовательность $\{f(x_n)\}$ сходится к элементу $f(x_0)$, поэтому функция f непрерывна и в отношении более сильной структуры τ_1.

Симметричные рассуждения справедливы и для пространства Y. Пусть на нем заданы две структуры сходимости σ_1 и σ_2, причем σ_1 слабее чем σ_2. Тогда из сильной непрерывности функции f следует ее слабая непрерывность. Действительно, пусть x_0 – предел последовательности $\{x_n\}$, тогда из σ_2-непрерывности функции f следует, что $f(x_0)$ есть σ_2-предел последовательности $\{f(x_n)\}$. Тогда и подавно $f(x_0)$ есть предел последовательности $\{f(x_n)\}$ в смысле более слабой сходимости σ_1, т.е. функция f – σ_1-непрерывна.

Сказанное можно выразить такими словами: с усилением структуры сходимости на множестве X и ослаблением структуры сходимости на множестве Y непрерывных функций $f{:}X\to Y$ становится больше (условия непрерывности ослабляются).

Если зафиксировать структуру сходимости σ на множестве Y, то можно поставить задачу: найти самую слабую структуру сходимости τ на множестве X, при которой функция $f{:}X\to Y$ остается непрерывной. Такой структурой является следующая. Последовательность $\{x_n\}\subset X$ сходится к элементу x_0 в смысле структуры τ тогда и только тогда, когда последовательность $\{f(x_n)\}$ сходится к $f(x_0)$ в смысле структуры σ. Иногда говорят о «*сходимости* *по функционалу*» и обозначают эту структуру так: $\tau=f^{-1}(\sigma)$. Легко проверить, что так определенная структура действительно определяет $\mathscr{L}^+$-структуру сходимости. В самом деле, рассмотрим произвольную квазистационарную последовательность $\{x_n\}\subset X$, состоящую, начиная с некоторого номера, из элемента x_0. Тогда последовательность $\{f(x_n)\}$ тоже будет квазистационарной и состоять из элемента $f(x_0)$, начиная с того же номера. Тогда последовательность $\{f(x_n)\}$ должна сходиться к этому элементу в смысле структуры σ. Это означает, что $\{x_n\}$ сходится к x_0 в смысле структуры τ. Т.е. выполнено первое условие в определении структуры сходимости. Второе условие тоже легко проверяется. Если последовательность $\{x_n\}\subset X$ сходится к элементу x_0 в смысле структуры τ, то последовательность $\{f(x_n)\}$ сходится к $f(x_0)$ в смысле структуры σ. Любой подпоследовательности последовательности $\{x_n\}$ соответствует подпоследовательность последовательности $\{f(x_n)\}$, которая обязана сходиться к $f(x_0)$ в смысле структуры σ. Это означает, что выбранная подпоследовательность сходится к x_0 в смысле структуры τ. Т. е. второе условие структуры сходимости выполнено. Рассмотрим третье условие. Пусть последовательность $\{x_n\}\subset X$ не сходится к элементу x_0. Это означает, что последовательность $\{f(x_n)\}$ не сходится к $f(x_0)$. Тогда существует ее подпоследовательность, никакая подпоследовательность которой не сходится к $f(x_0)$. Соответствующая подпоследовательность последовательности $\{x_n\}$ и будет обладать нужным для третьего условия свойством.

Точно так же, как в метрических пространствах, вводится понятие полунепрерывных снизу функций.

Определение. Пусть в $\mathscr{L}^+$-пространстве X задана вещественная функция $f:X\to\mathbb{R}$. Она называется полунепрерывной снизу в точке $x_0\in X$, если для любой последовательности $\{x_n\}\subset X$, сходящейся к x_0, справедливо неравенство $f(x_0)\leq\underline{\lim}f(x_n)$.

Очевидно, что непрерывная в точке x_0 функция будет и полунепрерывной в этой точке, а обратное верно не всегда.

Замкнутость и счетная компактность

В топологических пространствах понятия открытых и замкнутых множеств являются первичными, и из них строится понятие сходимости последовательностей. Здесь мы идем встречным путем – опираясь на структуру сходимости строим открытые и замкнутые множества.

Определение. Пусть X – $\mathscr{L}^+$-пространство со структурой сходимости τ. Множество $U\subset X$ называется *замкнутым* относительно структуры τ (или τ-замкнутым), если для любой последовательности $\{x_n\}\subset U$ справедливо включение $\mathrm{LIM}x_n\subset U$.

Если последовательность не сходится, то множество ее пределов пусто, и включение автоматически выполняется. А если последовательность сходится, то любой ее предел должен принадлежать множеству U, в котором лежит вся последовательность.

Открытым множеством называют дополнение замкнутого множества.

Следующее определение задает еще одно свойство подмножеств пространств сходимости.

Определение. Множество $U \subset X$ называется *счетно-компактным* относительно структуры τ (или τ-счетно-компактным), если для любой последовательности $\{x_n\} \subset U$ существует ее сходящаяся подпоследовательность, хотя бы один предел которой принадлежит множеству U.

Можно сказать и так: пересечение множества пределов подпоследовательности с множеством U непусто.

Теорема. Любое замкнутое подмножество $U \subset X$ счетно-компактного пространства X счетно-компактно.

Для доказательства рассмотрим произвольную последовательность $\{x_n\} \subset U$. Так как пространство X счетно-компактно, то существует ее сходящаяся подпоследовательность. В силу замкнутости множества U все пределы этой подпоследовательности принадлежат U, т. е. множество U счетно-компактно.

В метрических и $\mathscr{L}^*$-пространствах с единственным пределом справедливо и обратное утверждение: счетно-компактное подмножество произвольного пространства замкнуто [4 стр. 203]. В $\mathscr{L}^+$-пространствах это не так.

Рассмотрим множество из двух элементов $X=\{a,b\}$, в котором структура сходимости тривиальна: любая последовательность сходится ко всему пространству X. Рассмотрим множество U, состоящее из одного элемента a. Очевидно, что это множество счетно-компактно: любая подпоследовательность стационарной последовательности тоже стационарна, причем один из ее пределов лежит в том же множестве – это элемент a. Но множество U не замкнуто: пределы стационарной последовательности, состоящей из элемента a, не лежат все в этом множестве: элемент b есть предел, но не принадлежит U.

Этот простой пример можно обобщить. Легко видеть, что в пространствах с тривиальной сходимостью любое непустое множество счетно-компактно, но замкнутым является только все пространство.

А на другом полюсе структур сходимости – в дискретной сходимости – замкнуто любое подмножество пространства, но счетно-компактны только конечные подмножества. В самом деле, если в множестве бесконечное число элементов, то можно построить последовательность, состоящую из различных элементов, а в ней не может быть квазистационарных подпоследовательностей, которые только и могут сходиться.

Теперь рассмотрим, как влияют усиление или ослабление структуры сходимости на свойства замкнутости и счетной компактности множеств. Так как из сильной сходимости следует слабая сходимость последовательностей, то из сильной счетной компактности множества следует его слабая счетная компактность. Другими словами, с усилением сходимости счетно-компактных множеств становится меньше.

Иначе реагирует на усиление сходимости замкнутость множеств. Если множество слабо замкнуто, то все слабые пределы любой его последовательности принадлежат этому множеству, а так как каждый сильный предел последовательности является одновременно ее слабым пределом, то и все сильные пределы тоже лежат в том же множестве. Значит, если множество слабо замкнуто, то оно и сильно замкнуто. Иначе говоря, с усилением сходимости замкнутых множеств становится больше.

Оптимизация

Рассмотрим $\mathscr{L}^{+}$-пространство X, на котором задано множество $U \subset X$ и вещественная функция (иногда говорят функционал) $f{:}X \to \mathbb{R}$. Мы хотим найти минимум функции f на множестве U, что сводится к двум задачам: поиску *оптимального значения функционала*, т.е. числа $f^{*}=\inf f(x)$, $x \in U$, и поиску *оптимального решения*, т.е. элемента $x^{*} \in U$, для которого $f(x^{*})=f^{*}$. Здесь через $\inf f(x)$, $x \in U$ обозначена точная нижняя грань f на множестве U.

Достаточное условие существования оптимального решения дает известная

Теорема (Вейерштрасса). Пусть множество U счетно-компактно, а функция f полунепрерывна снизу. Тогда оптимальное решение задачи оптимизации существует.

Доказательство, по сути, такое же, как для метрических пространств. По определению точной нижней грани существует так называемая минимизирующая последовательность $\{x_n\} \subset U$, для которой $f(x_n) \to f^{*}$. В силу счетной компактности множества U найдется ее подпоследовательность $\{y_n\}$, сходящаяся к элементу $x^{*} \in U$. Из полунепрерывности снизу функции f следует неравенство $f(x^{*}) \le \underline{\lim} f(y_n)$. Так как последовательность $f(x_n)$ сходится к числу f^{*}, то к этому же числу сходится любая ее подпоследовательность, в частности, $\{f(y_n)\}$. Поэтому $\underline{\lim} f(y_n)=\lim f(y_n)=f^{*}$. А так как $f(x^{*})$ не может быть меньше точной нижней грани f^{*}, то $f(x^{*})= f^{*}$. Другими словами, x^{*} есть искомое оптимальное решение рассматриваемой задачи оптимизации. Теорема доказана.

Наверно, можно и не повторять, что непрерывная функция автоматически является полунепрерывной снизу, поэтому в теореме Вейерштрасса можно было бы условие полунепрерывности снизу заменить на непрерывность.

Условие счетной компактности множества U в теореме Вейерштрасса можно немного ослабить и говорить не о свойствах всего множества, а только о свойствах минимизирующих последовательностей. Справедлива такая модификация теоремы Вейерштрасса.

Теорема. Пусть функция f полунепрерывна снизу и каждая минимизирующая последовательность имеет подпоследовательность, сходящуюся к элементу множества U. Тогда существует оптимальное решение исходной задачи оптимизации.

Оказывается, условия этой теоремы являются не только достаточными для существования оптимального решения, но и в некотором роде необходимыми.

Теорема (необходимое и достаточное условие). Для существования оптимального решения задачи оптимизации необходимо и достаточно, чтобы на множестве X существовала такая структура сходимости, относительно которой функция f полунепрерывна снизу и каждая минимизирующая последовательность имеет подпоследовательность, сходящуюся к элементу множества U.

Достаточность условий вытекает из приведенной выше модифицированной теоремы Вейерштрасса. Докажем, что сформулированные условия необходимы. Пусть $x^*\in U$ – оптимальное решение задачи оптимизации. Обозначим через σ структуру обычной сходимости числовых последовательностей в пространстве вещественных чисел. Тогда искомой структурой сходимости в пространстве X будет «сходимость по функционалу», т.е. $\tau=f^{-1}(\sigma)$. Мы уже убедились, что это действительно $\mathscr{L}^+$-структура. Функция f становится в этой структуре непрерывной. А каждая минимизирующая последовательность сходится к элементу x^*. Так же будет сходиться и каждая подпоследовательность минимизирующей последовательности. Теорема доказана.

Стоит напомнить, что с усилением структуры сходимости на множестве X требования к непрерывности функции ослабевают, а требование существования у каждой последовательности сходящейся подпоследовательности, наоборот, усиливается. Когда решается вопрос о существовании оптимального решения конкретной задачи оптимизации, следует выбирать структуру сходимости, учитывая ее влияние на условия теоремы Вейерштрасса.

Оптимальное управление

Рассмотрим простой модельный пример задачи оптимального управления. Представим себе тележку единичной массы, которая может двигаться по прямым рельсам. Положение тележки в каждый момент времени описывается координатой $s_1(t)$. На тележку действует сила u(t), которая может быть положительной или отрицательной в зависимости от направления ее действия. Величина приложенной силы ограничена по модулю некоторым числом, которое мы выберем за единицу измерения. В начальный момент t=0 тележка покоится в начале координат. Фиксирован некоторый достаточно большой максимальный промежуток времени $[0,T_{max}]$, на котором определены рассматриваемые ниже функции. Требуется как можно скорее перевести тележку в точку с координатой а так, чтобы тележка в ней остановилась.

Попробуем формализовать задачу, дать ее математическую формулировку. Управлением здесь является приложенная к тележке сила, т.е. функция $u(t)$, $0 \le t \le T$, ограниченная по величине: $|u(t)| \le 1$, $0 \le t \le T$ (здесь и ниже $T \in [0, T_{max}]$). Обозначим множество всех достаточно гладких (например, непрерывных) функций через E. Каждому управлению $u \in E$ соответствует траектория тележки, которую можно найти из Второго закона Ньютона F=ma, где сила F – это управление $u(t)$, $0 \le t \le T$, а ускорение *a* – это вторая производная от координаты тележки. Если через s_1 обозначить координату тележки, а через s_2 – ее скорость, задача определения траектории сводится к простой системе дифференциальных уравнений: $\dot{s}_1(t)=s_2(t)$, $\dot{s}_2(t)=u(t)$, $0 \le t \le T$. Состояние тележки в каждый момент времени t задается двумерным вектором $s(t)=(s_1(t), s_2(t))$. Начальное положение тележки задается равенствами $s_1(0)=0$; $s_2(0)=0$.

Обозначим через D пространство всех таких двумерных функций на отрезке $0 \le t \le T$. Из теории дифференциальных уравнений следует, что каждому управлению $u \in E$ соответствует траектория $s \in D$.

Конечное положение определяется такими условиями: $s_1(T)=a$; $s_2(T)=0$. Требуется выбрать такую функцию $u(t)$, $0 \le t \le T$, которая обеспечивает минимальное значение T среди всех возможных. Это так называемая *задача оптимального быстродействия*.

В этом простом примере мы имеем все составные части общей постановки *задачи оптимального управления*. Во-первых, определено множество E всех возможных управлений (множество всех непрерывных функций) и множество D всех возможных траекторий (множество двумерных вектор-функций). Между управлениями и траекториями установлено отношение $S: E \to D$, ставящее каждому управлению $u \in E$ траекторию $s=S(u) \in D$. В нашем случае такое отношение задается системой дифференциальных уравнений с начальными условиями. На управление наложены явные ограничения, т. е. выделено подмножество U множества E. В нашем случае это подмножество задается неравенством $|u(t)| \le 1$, $0 \le t \le T$. Кроме того, есть ограничения на траектории, т. е. в множестве D задано подмножество $H \subset D$, в котором должны лежать допустимые траектории. В рассматриваемом примере это ограничение на траектории $s(t)=(s_1(t), s_2(t))$ выглядит так: $s_1(T)=a$; $s_2(T)=0$. И наконец, задан функционал $f(u)=T$, т.е. время достижения тележкой заданного конечного состояния. В общем случае функционал может явно зависеть и от траектории, и от управления: $f(u)=G(S(u), u)$, где $G(s, u)$ – вещественная функция, определенная на произведении пространств E и D.

Говоря короче, задача оптимального управления в общем случае выглядит так:

минимизировать функционал $f(u)=G(S(u), u)$ при ограничениях:
$u \in U \subset E$
$S(u) \in H \subset D$

Мы не собираемся здесь решать или детально исследовать задачу оптимального управления ни в общем виде, ни для модельного примера. Приведем только теорему Вейерштрасса, адаптированную к этому случаю.

Теорема (Вейерштрасса). Пусть на множествах E и D заданы структуры сходимости τ и σ соответственно. Пусть множество $U \subset E$ τ-счетно-компактно, множество $H \subset D$ σ-замкнуто, отображение $S:E \rightarrow D$ (τ,σ)-непрерывно, функция $G(s,u)$ (τ,σ)-полунепрерывна снизу в каждой точке $s \in H$ и $u \in U$. Тогда в рассматриваемой общей задаче оптимального управления существует оптимальное решение.

В самом деле, нетрудно убедиться, что выполнены условия общей теоремы Вейерштрасса из предыдущего раздела. Функция $f(u)=G(S(u), u)$ является τ-полунепрерывной снизу в каждой точке $u \in U$. Множество $U_1=S^{-1}(H)$ τ-замкнуто, следовательно, пересечение $U_0=U_1 \cap U$ τ-счетно-компактно. Следовательно, оптимальное решение задачи минимизации функции $f(u)$ на множестве U_0 существует, что и требовалось доказать.

Отметим, что в задачах оптимального управления особенно наглядны преимущество пространств сходимости перед традиционными структурами близости типа метрики, нормы, скалярного произведения и т. п. В самом деле, оценка близости двух управлений, т. е. функций, заданных на отрезке $0 \leq t \leq T$, по этим традиционным структурам, мало говорит о близости соответствующих траекторий. А в ряде задач естественнее было бы сравнивать управления именно с точки зрения свойств их траекторий.

Например, та же тележка может попасть в конечное состояние $s_1(T)=b$; $s_2(T)=0$ под действием совершенно непохожих сил: постоянной, скачкообразной (сначала толкать в одну сторону, потом в другую) и т. д. По сути задачи нужно сравнивать именно траектории, а не управления, т. е. судить об управлениях именно по вызываемым ими движениям. Тогда взятая за основу структура сходимости «по траекториям», аналогичная рассмотренной выше сходимости «по функционалу», может привести не только к более простому выводу известных результатов, но и к получению результатов новых. Примером может служить исследование зависимости решений экстремальных задач от параметра и др. [5], [6]

Литература

1. *Деза, Елена и Деза, Мишель Мари*. Энциклопедический словарь расстояний // Наука, М., 2008, 445 стр.
2. *Келли, Джон Л.* Общая топология. Перевод с английского А.В. Архангельского // Наука, М., 1968, 384 стр.
3. *Fréchet, Maurice.* Sur quelques points du calcul fonctionnel // Rendiconti del Circolo Matematico di Palermo. 1906, **22**, 1–72.
4. *Куратовский, Казимир*. Топология. Том 1. Перевод М.Я. Антоновского. С предисловием П.С. Александрова // Мир, М., 1966, 595 стр.
5. *Беркович, Евгений*. О существовании оптимальных решений для одного класса двухэтапных стохастических экстремальных задач // В сборнике: Приближенные методы решения задач оптимального управления и некоторых некорректных обратных задач, МГУ, М., 1972, 17-41.
6. *Беркович, Евгений*. О существовании оптимальных решений одной многоэтапной стохастической экстремальной задачи // Вестник Московского университета. Серия «Математика, механика», 1975, (4), 19-25.
7. *Cartan, Hanri*. Théorie des Filtres // Comptes Rendus de l'Académie des Sciences. Paris, 1937, **205**, 595-598.
8. *Choquet, Gustav*. Convergences // Annales de l'université de Grenoble, Sect. Sci. Math. Phys. (N.S.). 1947-1948 г., **23**, 57–112.
9. *Dolecki, Szymon*. An initiation into convergence theory // In book: Frédéric Mynard and Elliott Pearl, editors. Beyond Topology. Contemporary Mathematics Series A.M.S., 2009, **486**, 115-162.
10. *Bentley, H.L., Herrlich, H. и Lowen-Colenbunders, E*. Convergence // Journal of Pure and Applied Algebra. 1990, **68** (1–2), 27-45.

PROXIMITY STRUCTURES AND CONVERGENCE SPACES

E. M. Berkovich

redaktor@7iskusstv.com

Received 08.11.2020

The paper considers convergence spaces, in which converging sequences are taken as a basis, and other topological properties of sets and functions are derived from this concept. In this case, the limit of the sequence may not be unique. The spaces constructed in this way have a number of specific properties, in which, in particular, the well-known Weierstrass theorem can be formulated as a necessary and sufficient condition for the existence of a solution to an extremal problem. The use of convergence spaces in the study of optimal control problems is discussed.

Keywords: convergence, countable compactness, closedness, continuity, lower semicontinuity, extremum, optimal control, Weierstrass theorem.

МЕТОД САМОСОГЛАСОВАННЫХ УРАВНЕНИЙ ДЛЯ РЕШЕНИЯ ЗАДАЧ РАССЕЯНИЯ ЭЛЕКТРОМАГНИТНЫХ ВОЛН НА ФОТОННЫХ КРИСТАЛЛАХ

А.Ю. Ветлужский

Институт физического материаловедения СО РАН
E-mail: vay@ipms.bscnet.ru

Поступила 12.10.2020

Описывается численный метод решения задач дифракции электромагнитных волн на двумерных фотонных кристаллах. Особенностью метода является использование аналитических выражений, описывающих рассеяние на одиночном цилиндрическом элементе. Входящие в выражения неизвестные коэффициенты, учитывающие взаимное влияние рассеивателей, определяются путем построения и решения систем самосогласованных уравнений. В отличие от большинства численных методов данный подход при его использовании позволяет получить информацию о характеристиках поля только в определенных точках фотонного кристалла. Отсутствие необходимости расчета поля во всей области пространства, занимаемой рассматриваемой многоэлементной системой, обуславливает высокую эффективность данного метода. В работе выполнен сравнительный анализ спектров пропускания фотонных кристаллов, полученных рассматриваемым методом, с экспериментальными данными и численными результатами, полученными с использованием других подходов.
Ключевые слова: численные методы, дифракция, рассеяние, фотонные кристаллы, спектры пропускания.

УДК 51.73
DOI: 10.31145/2224-8412-2020-21-1-19-28

Введение

Одним из интересных объектов исследований в оптике и радиофизике последних десятилетий, с которым связывают разнообразные перспективы практического применения, являются фотонные кристаллы (ФК) – среды с периодически меняющейся в одном, в двух или в трех направлениях в пространстве диэлектрической проницаемостью с характерным масштабом периодичности, сопоставимым с длиной волны электромагнитного излучения.

Современная концепция ФК была сформулирована в работе [1]. Ключевое понятие теории фотонных кристаллов – запрещенная зона, означающее полосу частот, в пределах которой подавляется распространение электромагнитных волн через ФК. Физическая природа такого подавления заключается в брэгговском рассеянии излучения на периодических неоднородностях среды. Таким образом, спектр пропускания любого ФК представляет собой чередование запрещенных и разрешенных зон, при этом в диапазоне последних излучение практически свободно проходит через ФК.

Области практического применения ФК существенным образом зависят от характера пространственной периодичности кристалла. Одномерные ФК могут, например, рассматриваться как основа эффективных резонаторов радио- и оптического диапазонов [2], двумерные кристаллы – как базовый элемент различного рода волноводных и преобразующих световые потоки устройств [3], трехмерные, способные к формированию полных запрещенных зон кристаллы могут обеспечить полный контроль спонтанного испускания фотонов, что позволит, в принципе, создавать беспороговые лазеры [4]. Однако, несмотря на уникальные свойства трехмерных структур и проводимые в последние годы их всесторонние теоретические исследования, их непосредственная практическая реализация для использования в оптическом диапазоне все еще представляет значительные трудности. Поэтому двумерные ФК, создание которых современными технологическими методами не составляет сложностей, а интересные физические свойства представляются на сей день даже более разнообразными, чем у их трехмерных аналогов, вызывают особый интерес в силу возможности непосредственного практического применения.

Теоретические методы, использовавшиеся для изучения свойств ФК на всех этапах истории их исследований, весьма разнообразны. Поскольку задача возбуждения ФК электромагнитным полем – типичная дифракционная задача, изначально использовались хорошо развитые к моменту появления работы [1] применительно к задачам рассеяния на дифракционных решетках различной размерности аналитические и численно-аналитические методы (например, методы матриц передачи и матриц рассеяния, исходящие из физической постановки проблемы, метод полуобращения матричных операторов и модифицированный метод вычетов, оперирующие математической формулировкой краевой дифракционной задачи и т.д. [5]). По мере развития вычислительной техники все большую популярность приобретали строгие численные методы. К числу последних относятся и широко используемые в последние годы прямые методы численного решения уравнений Максвелла – метод конечных элементов [6] и метод конечных разностей во временной области [7].

На наш взгляд, общим недостатком упомянутых численных методов при их использовании для решения дифракционных задач рассеяния волн на системах тел является получение подчас избыточного количества информации, и, как следствие, высокие требования к вычислительным ресурсам и значительное время, требуемое для расчетов. Действительно, т.к. данные методы относятся к сеточным методам решения дифференциальных уравнений, применяя их к задаче возбуждения ФК произвольной геометрии, получаем полное представление об амплитудно-фазовом распределении поля во всей рассматриваемой области пространства. Однако, многие электродинамические задачи, связанные с ФК, ограничиваются необходимостью получения амплитудных спектров пропускания таких структур в заданной полосе частот либо определения интенсивности поля в некоторых ключевых точках ФК без анализа его характеристик в других областях. Рассмотрение простого в алгоритмизации, не требовательного к вычислительным ресурсам численного метода, реализующего такой подход, является целью настоящей работы.

1. Метод самосогласованных уравнений

Данный метод был впервые предложен в [8] для анализа дифракции волн на параллельных цилиндрах. Позднее использовался в [9] для получения аналитических выражений, описывающих процессы рассеяния волн на двух цилиндрических и сферических телах. Принципиальным недостатком метода являлась невозможность его практического применения для систем более чем из двух тел из-за громоздкости получающихся выражений. Для решения дифракционных задач, касающихся систем большого числа рассеивателей, он начал применяться лишь с середины 90-х годов прошлого века по мере развития вычислительной техники [10].

Рассмотрим его на примере задачи возбуждения системы N параллельных диэлектрических бесконечно протяженных цилиндров кругового сечения нитью синфазного (электрического или магнитного) тока, ориентированной вдоль элементов структуры. Запишем решение двумерного неоднородного уравнения Гельмгольца в цилиндрической системе координат для поля, рассеянного произвольным -ым цилиндром (j = 1,2…N), в виде разложения по азимутальным гармоникам:

$$u_{\text{рас}}(\boldsymbol{r},\boldsymbol{r}_j) = \sum_{m=-\infty}^{\infty} P_{mj} H_m^{(1)}(k|\boldsymbol{r}-\boldsymbol{r}_j|) e^{in\varphi_{r-r_j}}, \tag{1}$$

где k – волновое число, $\mathrm{H}_{\mathrm{m}}^{(1)}$ – функция Ханкеля первого рода -ого порядка, $\varphi_{\mathrm{r-rj}}$ – угол, образованный вектором $\boldsymbol{r}$-$\boldsymbol{r}_j$ и осью x системы координат (рис. 1).

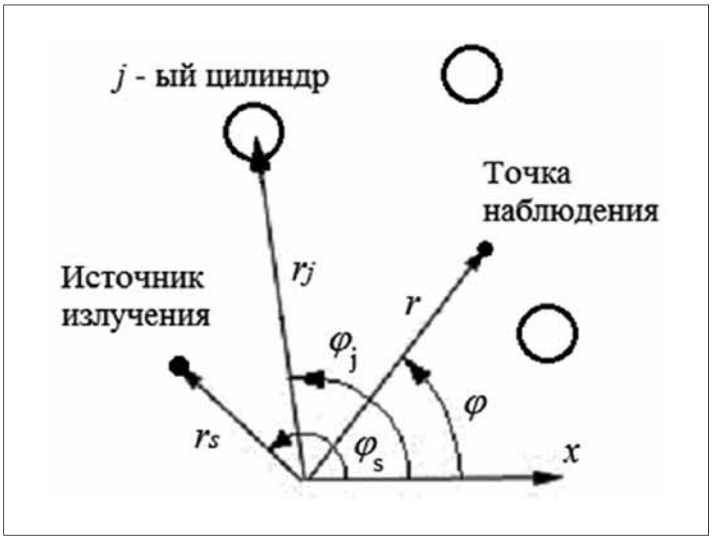


Рис. 1. Концептуальная схема рассматриваемой задачи.

Полное падающее на некоторый -ый цилиндр поле ($i = 1,2\ldots N$, $i \neq j$) представим в виде суммы поля источника и полей, рассеянных остальными элементами системы:

$$u^i_{\text{пад}}(\boldsymbol{r}) = u_0(\boldsymbol{r}) + \sum_{j=1, j\neq i}^{N} u_{\text{рас}}(\boldsymbol{r}, \boldsymbol{r}_j). \tag{2}$$

Это же поле можно описать в виде разложения, аналогичного (1):

$$u^i_{\text{пад}}(\boldsymbol{r}) = \sum_{m=-\infty}^{\infty} B_{mj}\, J_m\left(k\left|\boldsymbol{r} - \boldsymbol{r}_j\right|\right) e^{in\varphi_{r-r_j}}. \tag{3}$$

Радиальная зависимость поля здесь выражена через функцию Бесселя -ого порядка, поскольку она не имеет особенностей при $\boldsymbol{r} \to \boldsymbol{r}_j$.

Для определения неизвестных коэффициентов P_{mj} и B_{mj} выразим рассеянное поле $u_{\text{рас}}(\boldsymbol{r}, \boldsymbol{r}_j)$ для каждого -ого цилиндра через волновые функции i-ого элемента ($i \neq j$). Для этого используем теорему сложения для цилиндрических функций [11] (рис. 2):

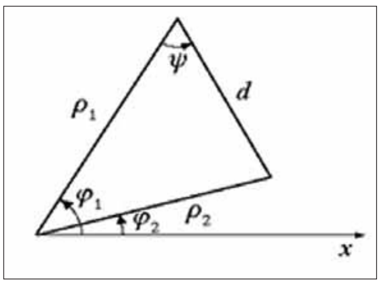


Рис. 2. Геометрическая иллюстрация к теореме сложения для цилиндрических функций.

$$Z_m(\alpha d)e^{im\psi} = \sum_{k=-\infty}^{\infty} Z_{m+k}(\alpha\rho_1)\,J_k(\alpha\rho_2)e^{ik(\varphi_1-\varphi_2)}, \tag{4}$$

где $\rho_1 > \rho_2$, Z – произвольная цилиндрическая функция.

Применяя (4) к (1) получаем:

$$u_{\text{рас}}(\boldsymbol{r},\boldsymbol{r}_j) = \sum_{m=-\infty}^{\infty} C_{mj,i}\,J_m(k|\boldsymbol{r}-\boldsymbol{r}_i|)e^{in\varphi_{r-r_i}}, \tag{5}$$

где

$$C_{mj,i} = \sum_{l=-\infty}^{\infty} P_{lj}H_{l-m}^{(1)}\big(k|\boldsymbol{r}-\boldsymbol{r}_j|\big)e^{i(l-n)\varphi_{r-r_j}}. \tag{6}$$

Для дальнейших преобразований поле источника представим в аналогичной формулировке – в виде разложения по волновым функциям -ого цилиндра:

$$u_0(\boldsymbol{r}) = \sum_{m=-\infty}^{\infty} D_{mi}\,J_m(k|\boldsymbol{r}-\boldsymbol{r}_i|)e^{in\varphi_{r-r_i}}, \tag{7}$$

где коэффициенты D_{mi} связаны с известной комплексной амплитудой поля источника A_0 и местоположением последнего, определяемым вектором $\boldsymbol{r}_s$:

$$D_{mj} = A_0H_m^{(1)}(k|\boldsymbol{r}_i-\boldsymbol{r}_s|)e^{im\varphi_{r_i-r_s}}.$$

Подставляя выражения (3), (5) и (7) в (2), получим:

$$B_{mi} = D_{mi} + \sum_{j=1, j\neq i}^{N} C_{mj,i}. \tag{8}$$

В последнем выражении неизвестными являются два коэффициента: B_{mi} и P_{lj}. Для их определения необходимо установить между ними дополнительное соотношение, что достигается решением простейшей задачи дифракции волн на одиночном цилиндре. Удовлетворяя граничным условиям на его поверхности, получаем хорошо известное выражение (ограничимся его формулировкой для *TM* волн, т.е. для параллельной ориентации вектора напряженности электрического поля $\boldsymbol{E}$ относительно цилиндров):

$$\frac{P_{mi}}{B_{mi}} = \frac{nJ_m(ka)J_m'(nka) - J_m'(ka)J_m(nka)}{J_m(nka)H_m^{(1)\prime}(ka) - nH_m^{(1)}(ka)J_m'(nka)}. \tag{9}$$

Здесь n – показатель преломления материала цилиндра.

Применяя последовательно описанную процедуру ко всем цилиндрам и учитывая (6) и (9), получаем систему N самосогласованных линейных неоднородных уравнений (8), в которой каждый неизвестный коэффициент P_{mi}, описывающий возбуждение соответствующего цилиндра, определяется через коэффициенты, характеризующие состояние других цилиндров системы. Разрешая систему стандартными методами, окончательно находим поле в произвольной точке пространства с использованием следующего выражения:

$$u(\boldsymbol{r}) = u_0(\boldsymbol{r}) + \sum_{i=1}^{N} \sum_{m=-\infty}^{\infty} P_{mi} H_m^{(1)}(k|\boldsymbol{r} - \boldsymbol{r}_i|) e^{in\varphi_{r-r_i}} .$$

Достоинствами описанного метода является простота алгоритмизации и программной реализации; получение сколь угодно высокой точности расчетов при учете соответствующего числа пространственных гармоник; непосредственная применимость для произвольного взаимного положения источника излучения, рассеивателей и точки наблюдения в отличие от большинства численных методов; возможность получения информации не только о результирующем поле в системе, но и о полях, рассеиваемых отдельными элементами структуры. Основной недостаток – возможность расчета электромагнитных полей только в системах объектов, задачи дифракции на которых имеют аналитическое решение.

2. Верификация метода

Рассмотрим две задачи возбуждения двумерных ФК, элементы которых образуют квадратные решетки, локально плоскими волнами. В первом случае полагаем ФК состоящим из идеально проводящих цилиндров ($n \to \infty$), что допустимо в пренебрежении тепловыми потерями в металле в СВЧ диапазоне, коэффициент заполнения структуры $f = {\pi a^2}/{d^2} =$ = $7.85 \cdot 10^{-3}$, период $d = 1$ см, радиус цилиндров $a = 0.05$ см. Количество элементов в структуре – 121 (11 на 11). Во втором – считаем диэлектрическую проницаемость элементов $\varepsilon = 9$ (оксид алюминия Al_2O_3), коэффициент заполнения структуры $f = 12.56 \cdot 10^{-2}$, $d = 1$ см, $a = 0.2$ см. Количество элементов – 361 (19 на 19). На рис. 3 (а) и (б) представлены спектры пропускания таких ФК, рассчитанные с использованием метода самосогласованных уравнений.

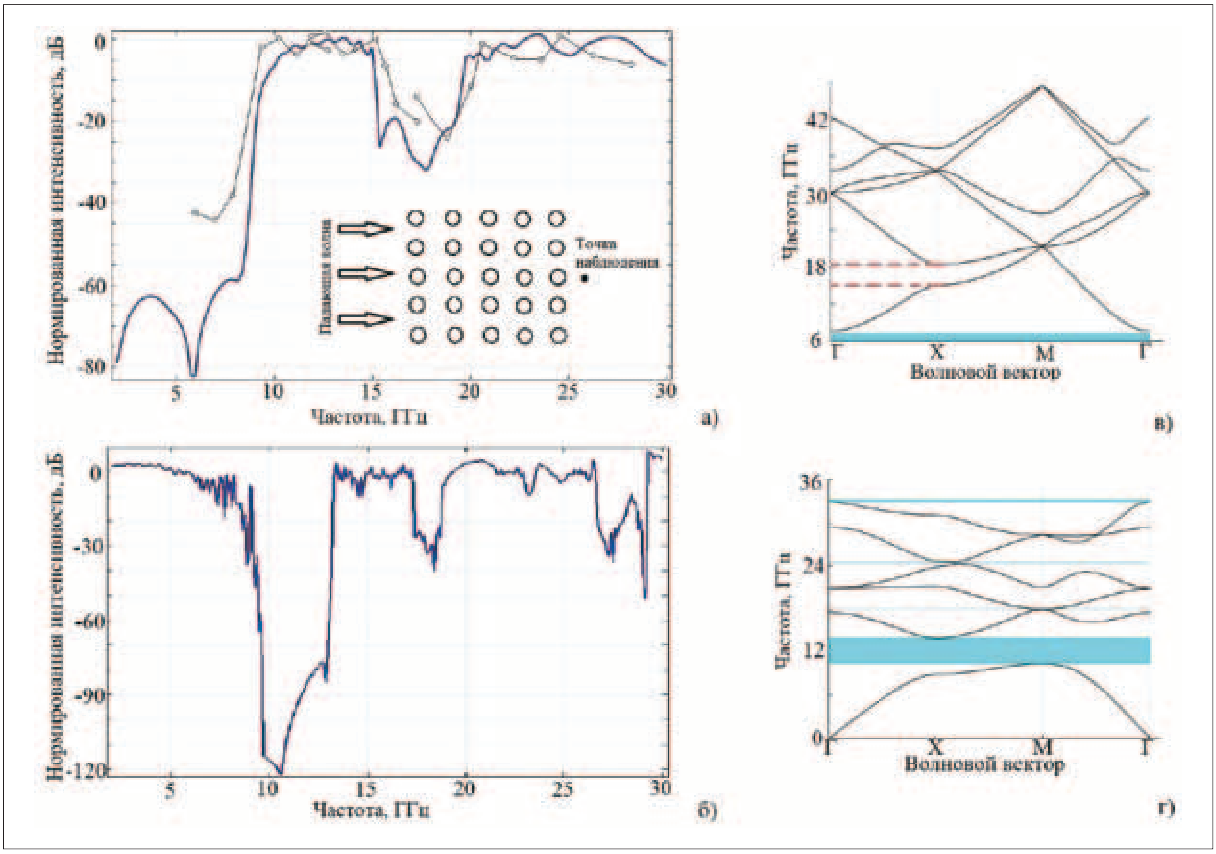


Рис. 3. (а, б) – плавные кривые – рассчитанные спектры пропускания металлического и диэлектрического ФК соответственно, ломаная кривая на рис. (а) отображает результаты экспериментов; (в, г) – дисперсионные диаграммы, соответствующие металлической и диэлектрической структурам, где Г, X, M – точки высокой симметрии, ограничивающие неприводимую зону Бриллюэна в пространстве волновых векторов, полные запрещенные зоны указаны выделенными цветом прямоугольными областями, пунктирными линиями на рис. (в) обозначена неполная запрещенная зона в направлении ГX.

Спектры определялись в направлении распространения ГX в пространстве волновых векторов или, используя понятие двумерных индексов Миллера, в направлении (10). Ломаная линия на рис. (а) описывает полученные автором экспериментальные данные по прохождению волн через металлический ФК в СВЧ диапазоне. Рис. 3 (в) и (г) иллюстрируют рассчитанные методом разложения по плоским волнам [12] дисперсионные диаграммы обоих рассматриваемых ФК.

Хорошее согласие между теоретически и экспериментально полученными данными, а также наблюдаемое соответствие частотных диапазонов формирования как полных, так и неполных запрещенных зон ФК, определенных разными методами, позволяют сделать вывод об адекватности получаемых методом самосогласованных уравнений результатов реальным электродинамическим характеристикам двумерных ФК.

Заключение

В работе рассмотрен один из методов решения задач рассеяния волн на многоэлементных двумерных системах, позволяющий определять спектральные характеристики таких объектов при произвольном положении источника и точки наблюдения поля, пространственное распределение интенсивности поля в системе, амплитудно-фазовые характеристики излучения, формируемого как всей совокупностью рассеивателей, так и отдельными элементами структуры и т.д. Корректность получаемых описываемым методом результатов подтверждена сопоставлением полученных с его помощью спектров пропускания двумерных ФК с экспериментальными и полученными с использованием другого теоретического подхода данными. Ограничением описываемого метода является необходимость существования аналитических решений задач рассеяния волн на отдельных элементах рассматриваемой системы.

Литература

1. *Yablonovitch E*. Inhibited spontaneous emission in solid-state physics and electronics // Phys. Rev. Lett., 1987, 58 (20), 2059-2062.
2. *Vetrov S. Y., Pankin P. S., Timofeev I. V*. The optical Tamm states at the interface between a photonic crystal and a nanocomposite containing core–shell particles // J. Opt., 2016, 18 (6), 65106, 10 pp.
3. *Шабанов В.Ф., Ветров С.Я., Шабанов А.В*. Оптика реальных фотонных кристаллов // Изд-во СО РАН, Новосибирск, 2005, 230 стр.
4. *Noda S., Fujita M., Asano T.* Spontaneous-emission control by photonic crystals and nanocavities // Nature Photonic, 2007, 1 (8), 449-458.
5. *Lourtioz J.-M., Benisty H., Berger V.* et al. Photonic Crystals: Towards Nanoscale Photonic Devices // Springer, 2008, 513 pp.
6. *Jin J.M., Riley D.J*. Finite Element Analysis of Antennas and Arrays // John Wiley & Sons, Hoboken, NJ, 2008, 440 pp.
7. *Nagra A.S., York R.A*. FDTD analysis of wave propagation in nonlinear absorbing and gain media // IEEE Transactions on Antennas and Propagation, 1998, 46, 334-340.
8. *Tversky V.* Multiple scattering of radiation by an arbitrary configuration of parallel cylinders // J. Acoust. Sos. Am., 1951, 24 (1), 42-46.
9. *Иванов Е.А*. К решению задачи о дифракции плоской волны на двух круговых цилиндрах в случае коротких волн // Радиотехника и электроника, 1966, 11 (5), 931-942.
10. *Ветлужский А.Ю., Ломухин Ю.Л., Михайлова О.Г*. Эффект прозрачности объемных решеток // Радиотехника и электроника, 1998, 43 (7), 797-799.
11. *Корн Г., Корн Т.* Справочник по математике для научных работников и инженеров // Наука, М., 1970, 720 стр.
12. *Лозовик Ю.Л., Эйдерман С.Л.* Зонная структура сверхпроводящих фотонных кристаллов // Физика твердого тела, 2008, 50 (11), 1944-1947.

METHOD OF SELF-CONSISTENT EQUATIONS FOR SOLVING THE PROBLEMS OF ELECTROMAGNETIC WAVES SCATTERING BY PHOTONIC CRYSTALS

A.Yu. Vetluzhsky

Institute of Physical Materials Science SB RAS
E-mail: vay@ipms.bscnet.ru

Received 12.10.2020

One of the numerical methods for solving problems of scattering of electromagnetic waves by two-dimensional photonic crystals is considered. A feature of the method is the use of analytical expressions describing diffraction by a single element of the system. The unknown coefficients included in these expressions, taking into account the mutual influence of the scatterers, are determined by constructing and solving systems of self-consistent equations. In contrast to most numerical methods this approach allows one to obtain information on the amplitude-phase or spectral characteristics of the field only at definite points of the structure. The absence of the need to determine the field parameters in the entire area of space occupied by the considered multi-element system determines the high efficiency of this method. The paper compares the results of calculating the transmission spectra of two-dimensional photonic crystals by the considered method with experimental data and numerical results obtained using other approaches.

Keywords: umerical methods, diffraction, scattering, photonic crystals, transmission spectra.

INTRABAND OPTICAL ABSORPTION INDUCED BY RASHBA SPIN-ORBIT COUPLING IN TWO-DIMENSIONAL ELECTRON GAS

G.B. Ibragimov, R.Z. Ibaeva

Institute of Physics, Academy of Sciences of Azerbaijan,
131 H.Javid ave., 1143, Baku, Azerbaijan

Received 21.07. 2020

Taking into account the Rashba spin – orbit coupling, the magnetoabsorption of infrared electromagnetic radiation by free carriers with scattering by optical, piezoelectric, and acoustic lattice vibrations in a quasi-two-dimensional system is studied.
Keywords: Rashba spin-orbit coupling, intraband magnetoabsorption, quasi-two-dimensional electron system

УДК 537.6, 538.9
DOI: 10.31145/2224-8412-2020-21-1-29-36

In recent years, considerable interest has been observed in the study of quantum states and transport Taking into account the spin – orbit interaction due to their applications in nanostructure physics. Interest in spin-dependent phenomena has recently increased significantly in connection with the rapidly developing direction of spintronics [1]. Particular attention is paid to semiconductor heterostructures, since the level of technology for their growth makes them the basis of future spintronics devices. A powerful method for studying spin properties is optical phenomena studies in a magnetic field. The quantum states of electrons and holes in semiconductor structures, where the spin – orbit interaction is associated with the absence of symmetry center that limits the structure potential, were studied in a number of theoretical and experimental works [2–5]. Taking into account the spin – orbit interaction in the system leads to a mixing of the electron states related to different magnetic subbands and, as a consequence, to a nontrivial structure of the energy spectrum and spin polarization [6]. Spin – orbit interaction also leads to the possibility of resonance transitions of conduction electrons in a magnetic field between Landau levels at frequencies representing linear combinations of cyclotron and Zeeman frequencies [7]. In recent years, much attention has been paid to the study of the optical properties of low-dimensional nanostructures. In particular, intraband transitions in quantum wells were considered in [8–10], in quantum wires [11, 12]. Intraband magnetic absorption of electromagnetic radiation by quantum nanostructures was studied in [13-15].

In this connection, it is of interest to study the effect of spin – orbit interaction on the intraband magnetoabsorption of electromagnetic radiation in quantum wells. In the present work, the intraband magnetic absorption of electromagnetic radiation of linear polarization by a 2D electron gas with Rashba spin-orbit interaction is theoretically investigated.

The Hamiltonian describing the quantum-mechanical motion of an electron in a two-dimensional system in a constant uniform perpendicular magnetic field (H || oz) taking into account the Rashba spin-orbit interaction and the Zeeman splitting has the form [3]:

$$H = \frac{P^2}{2m^*} + \frac{a}{\hbar}\left(\sigma_x P_y - \sigma_y P_x\right) + \frac{1}{2} g\mu_B B\sigma_z \tag{1}$$

P is the momentum operator, σ is the Pauli matrix, μ *is the* Bohr magneton, α is the Rashba spin-orbit interaction constant, g is the Lande factor, and ħ is the Planck constant. For the vector potential of the magnetic field, the Landau gauge A = (0, H • x, 0) was chosen.

In paper [3], it was analitically solved the problem of the quantum states of an electron described by the Hamiltonian (1). So, the electronic spectrum is a discrete levels mixed in pairs

$$E_n^{\pm} = \hbar\omega_c n \pm \frac{1}{2}\sqrt{\left(\hbar\omega_c + 2g\mu_B H\right)^2 + \frac{8\alpha^2}{l_H^2} n} \tag{2}$$

$n = 1,2,3,\ldots.\ E_0^{+} = \left(\hbar\omega_c / 2 + g\mu_B B\right)$

where $\omega_c = \dfrac{eH}{m^* c}$ is the cyclotron frequency.

The wave functions in this case had the form

$$\Psi_n^+(x,y) = \frac{e^{ik_y y}}{\sqrt{2\pi A_n}} \begin{pmatrix} D_n \Phi_n\left(\dfrac{x+x_0}{l_H}\right) \\ \Phi_n\left(\dfrac{x+x_0}{l_H}\right) \end{pmatrix} \tag{3}$$

for the branch E_n^+,

$$\Psi_n^-(x,y) = \frac{e^{ik_y y}}{\sqrt{2\pi A_n}} \begin{pmatrix} \Phi_{n-1}\left(\dfrac{x+x_0}{l_H}\right) \\ -D_n \Phi_n\left(\dfrac{x+x_0}{l_H}\right) \end{pmatrix} \tag{4}$$

for the branch E_n^-.

The full wave function

$\Psi(x,y,z) = \Psi(x,y)\xi_0(z),$

where

$\xi_0(z) = \sqrt{\dfrac{2}{d}} \sin\left(\dfrac{l\pi z}{d}\right)$

In expressions (3), (4) Φ(z) is the oscillation function, $l_H = \sqrt{\hbar c / eH}$ isthemagneticlength, $x_0 = k_y l_0^2$, $D_n = \dfrac{\sqrt{2n}\alpha / l_H}{E_0 + \sqrt{E_0^2 + 2n\alpha^2 / l_H^2}}$, $A_n = 1 + D_n^2$

Calculation of coefficient of light absorption by free carriers in quasi-two dimensional systems was carried out in the second order of perturbation theory. The rate of transition from the state kn to the state k′n ′ is determined in this case by the following formula:

$$W_i = \frac{2\pi}{\hbar} \sum_{fq} \left[\left| \langle f | M_+ | i \rangle \right|^2 \delta\left(E_f - E_i - \hbar\Omega - \hbar\omega_q\right) + \left| \langle f | M_- | i \rangle \right|^2 \delta\left(E_f - E_i - \hbar\Omega + \hbar\omega_q\right) \right] \tag{5}$$

where

$$\left\langle f\left|M_{\pm}\right|i\right\rangle=\sum_{\alpha}\left(\frac{\left\langle f\left|H_{R}\right|\alpha\right\rangle\left\langle\alpha\left|V_{s}\right|i\right\rangle}{E_{i}-E_{\alpha}\mp\hbar\omega_{q}}+\frac{\left\langle f\left|V_{s}\right|\alpha\right\rangle\left\langle\alpha\left|H_{R}\right|i\right\rangle}{E_{i}-E_{\alpha}-\hbar\Omega}\right)$$

here $\hbar\Omega, \hbar\omega_q$ are photon and phonon energy respectively.

We choose the direction of polarization of the photons along the x axis. Then the electron-photon interaction operator is written in the form:

$$H_{\text{int}}=-\frac{e}{m^{*}c}\left(P-\frac{eA}{c}\right)A_{0}+\frac{ea}{c\hbar}\sigma_{y}A_{0}$$

or

$$H_{\text{int}}=-\frac{i|e|\hbar}{m^{*}c}A_{0}\frac{\partial}{\partial x}-\frac{|e|}{c}\frac{\alpha}{\hbar}A_{0}\sigma_{y} \tag{6}$$

where A_0 is the amplitude of the electromagnetic wave, with the volume concentration of photons. Using expression (6) for H_{int} and wave functions (3), (4), we find

$$\left\langle\Psi_{n}^{+}(x,y)\left|H_{\text{int}}\right|\Psi_{n+1}^{-}(x,y)\right\rangle=\frac{1}{\sqrt{A_{n}A_{n+1}}}\frac{e}{c}A_{0}\left[\left\langle\Phi_{n}\left|\frac{p_{x}}{m}\right|\Phi_{n+1}\right\rangle+D_{n}D_{n+1}\left\langle\Phi_{n-1}\left|\frac{p_{x}}{m}\right|\Phi_{n}\right\rangle-i\frac{\alpha}{\hbar}D_{n+1}\right] \tag{7}$$

Using the following relations

$$\left\langle\Phi_{n}\left|\frac{p_{x}}{m*}\right|\Phi_{n+1}\right\rangle=i\sqrt{(n+1)/2}\,l_{H}\omega_{0} \quad \text{and} \quad \left\langle\Phi_{n-1}\left|\frac{p_{x}}{m*}\right|\Phi_{n}\right\rangle=i\sqrt{n/2}\,l_{H}\omega_{0}$$

for expression (7) we get:

$$\left\langle\Psi_{n}^{+}(x,y)\left|H_{\text{int}}\right|\Psi_{n+1}^{-}(x,y)\right\rangle=\frac{-il_{H}\omega_{0}}{\sqrt{2A_{n}A_{n+1}}}\left[\sqrt{n+1}+D_{n+1}\left(D_{n}\sqrt{n}+\sqrt{2}\,\alpha m*l_{H}/\hbar^{2}\right)\right] \tag{8}$$

The matrix element of the electron – phonon interaction has the following form:

$$\left|\left\langle k_{y}'n'l'\left|V_{s}\right|k_{y}nl\right\rangle\right|^{2}=C_{j}'^{2}\delta_{k_{y}',k_{y}\pm q_{y}}F_{nn'}^{\pm}\left(q_{x}q_{y}\right)\Lambda_{ll'}\left(q_{z}\right)$$

V_s is the energy operator of interaction between electron and phonon, C_j is the function, characterizing the interaction between electrons and phonons,

$$F_{nn'}^{\pm}(q_n) = \left| < \Psi_v(x,y) \right| e^{i(q_x x + q_y y)} \left| \Psi_{v'}(x,y) \right|^2$$

$$F_{nn}^{+}(q_{II}) = B_n^{n'}(\xi)\left[\sqrt{\frac{n}{n'}} D_n D_{n'} L_{n'-1}^{n-n'}(\xi) + L_{n'}^{n-n'}(\xi)\right]$$

$$F_{nn}^{-}(q_{II}) = B_n^{n'}(\xi)\left[\sqrt{\frac{n}{n'}} L_{n'-1}^{n-n'}(\xi) + D_n D_{n'} L_{n'}^{n-n'}(\xi)\right]^2$$

where

$$B_n^{n'} = \left(n'! \big/ n!\right)\xi^{n-n'} e^{-\xi} \delta_{k'_y k_q + q_y} \qquad \xi = q_{II}^2 l_H^2 / 2$$

$$\Lambda_{ll'}(q_z) = \left| \frac{2}{d}\int_0^d dz \exp(i q_z z) \sin\left(\frac{l' \pi z}{d}\right) \sin\left(\frac{l \pi z}{d}\right) \right|^2$$

$$\int_0^{\infty} \Lambda_{ll'}(q_z) dq_z = \frac{2\pi}{d}\left(1 + \frac{1}{2}\delta_{ll'}\right)$$

$$C_j'^2 = C_j^2 F_j(q)$$

For the interaction of an electron with polar optical phonons, we have:

$$C_{POL}^2 = 2\pi e^2 \hbar \omega_0 \left\{\frac{1}{\varepsilon_\infty} - \frac{1}{\varepsilon_0}\right\}, \qquad F_{POL} = \frac{N_0^{\pm}}{q^2},$$

$$N_0 = \left[\exp\left(\frac{\hbar\omega_0}{K_B T}\right) - 1\right]^{-1}, \qquad N_0^- = N_0, \; N_0^+ = N_0 + 1$$

For the interaction of an electron with nonpolar optical phonons

$$C_{np}^2 = \frac{\hbar D^2}{2\rho\omega_0\Omega_0}, \qquad F_{np}(q) = N_0^{\pm}$$

For the interaction of electrons with piezo-electric phonons

$$C_{PE}^2 = \frac{e^2 K_B T \beta_p}{2\rho \upsilon_s^2 \Omega_0 \varepsilon^2}, \qquad F_{PE}(q) = \frac{1}{q^2}$$

here β_p is the piezoelectric constant, where ε_∞ and ε_0 are high-frequency and statistic dielectric permittivity of the material, $\omega_q = \omega_0$ is the frequency of a longitudinal optical phonon, dispersion of which is neglected.

In the case of interaction of electrons with acoustic phonons

$$C_{DP}^2 = \frac{E_{ac}^2 K_B T}{2\rho \upsilon_s^2 \Omega_0}, \qquad F_{DP}(q) = 1$$

here E_{ac} is the deformation potential, υ_s is the semiconductor sound speed.

Calculations of spectra of light absorption by 2D electron gas carried out for

GaAs/$In_{0.23}Ga_{0.77}As$ lattice structures. The effective electron mass in $In_{0.23}Ga_{0.77}As$ was chosen equal to $m* = 0.05m_0$, $g = -4.0$, Rashba spin-orbit interaction constant $\alpha = 2.5 \cdot 10^{-11}$ $эВ \cdot м$

Conclusion

Thus, in this work the absorption of light from the so-called rashba plane, of two-dimensional electron gas taking into account spin-orbital interaction in the presence of strong magnetic field normal to the structure plane. It is believed that in the conduction band only the zeroth Landau level is filled., i.e. it is assumed that the factor $2\pi nlH \leq 2$. Splitting is considered to be exactly weak, i.e. supposed that conditions $2m\alpha^2/\omega << 1$, $|g|\ \mu BH/\omega_C = |g|m/2m_0 << 1$. As a result of rather cumbersome transformations of the general formula (5), we manage to highlight pour the contribution of the spin-orbit interaction effects in intraband light magnetoabsorption.

The resulting formula generalizes the expression burning for intraband magnet absorption of light obtained by us earlier for quasi-two-dimensional electron gas in the absence of Rashba plane and Zeeman splittiing [15] .

References

1. *D.D. Aüschalom, D. Loss, N.* Samarth (eds.), Semiconductor Spintronics and Quantum Computation, Springer-Verlag, Berlin (2002) 311.
2. *Y.A. Bychkov, V.I. Melnikov, E.I. Rashba*, Effect of spin-orbit coupling on the 2D electron spectrum in a tilted magnetic field, JETP 98, (1990) 717-726.
3. *X.F. Wang, P.Vasilopoulos*, Magnetotransport in a two-dimensional electron gas in the presence of spin-orbit interaction, Phys.Rev. B 67, (2003) 085313-08518.
4. *M.-C. Chang*, Effect of in-plane magnetic field on the spin Hall effect on a Rashba-Dresselhaus system, Phys.Rev.B 72, (2005) 085315-085319.
5. *M. Zarea, S.E. Ulloa*, Landau level mixing by full spin-orbit interactions, Phys. Rev.B 72, (2005) 085342-085345.
6. *V.Ya. Demikhovskii, A.A. Perov*, Harper-Hofstadter problem for 2D electron gas with k-linear Rashba spin-orbit coupling, Europhys. Lett 76, (2006) 477-480.
7. *E.I. Rashba*, Combined resonance in semiconductors(Combined resonance mechanisms in semiconductors, noting spin-orbital interactions, zone carrier and local center resonances), Advances in Physical Sciences 84, (1964) 557-578.
8. *V.L.Gurevich, D.A. Parshin, K.E. Shtengel*, Optical phonon-assisted light-absorption by free-carriers in quasi two-dimensional systems, Physics of the Solid State 30, (1988) 1466-1475.
9. *G.G. Zegrya, V.E. Perlin*, Interband light absorption in quantum wells at the expense of electron-electron collisions, Semiconductors 32, (1998) 466-471.
10. *G.B. Ibragimov*, Free-carrier absorption in quantum well structures for alloy-disorder scattering, Phys. Stat.Sol(b) 231, (2002) 589-594.
11. *G.B. Ibragimov*, Free-carrier absorption in quantum wires for boundary roughness scattering, Journal of Physics: Condensed Matter 15, (2003) 1427-1435.
12. *G.B. Ibragimov*, Free-carrier absorption in semiconducting quantum well wires for alloy-disorder scattering, Journal of Physics: Condensed Matter 14, (2002) 8145-8152.
13. *N.G. Galkin, V.A. Margulis, A.V. Shorokhov*, Intraband absorption of electromagnetic radiation by quantum nanostructures with parabolic confinement potential, Physics of the Solid State 43, (2001) 530-538.
14. *G.B. Ibragimov*, Journal of Physics: Condensed Matter 15, (2003) 8949-8956.
15. *G.B. Ibragimov*, Free-carrier magneto-absorption in quantum well structures, Ukr. J. Phys. 241, (2004) 1923-1927.
16. Solid State Physics: Semiconductor Heterostructures and Nanotstructures, v.44, ed. by H.Ehrenreich and D. Turnbull, New York, Academic Press, 44 (1991) 294p.

ВНУТРИЗОННОЕ ОПТИЧЕСКОЕ ПОГЛОЩЕНИЕ ИНДУЦИРОВАННОЕ СПИН-ОРБИТАЛЬНОМ ВЗАИМОДЕЙСТВИЕМ РАШБА В ДВУМЕРНОМ ЭЛЕКТРОННОМ ГАЗЕ

Г.Б. Ибрагимов, Р.З. Ибаева

Институт Физики НАНА Азербайджана

Поступила 21.07. 2020

С учетом спин – орбитального взаимодействия Рашбы исследовано поглощение инфракрасного электромагнитного излучения в квазидвумерной системе свободными носителями, рассеяние которых происходит на оптических, пьезоэлектрических и акустических колебаниях решетки.

Ключевые слова: спин-орбитальное взаимодействие Рашба, внутризонное магнитопоглощение, квазидвумерная электронная система

ВОЗМОЖНОСТЬ ВОЗБУЖДЕНИЯ НЕРАВНОВЕСНЫХ ОЖЕ – ПЕРЕХОДОВ КОРПУСКУЛЯРНЫМИ ЗОНДАМИ

Е.Р. Бурмистров[1,a], Л.П. Авакянц[1,b]

[1]*Московский государственный университет имени М.В. Ломоносова, физический факультет, кафедра общей физики, Россия, 119991, Москва, Ленинские горы, д.1, стр.2*

E-mail: [a]*eugeni.conovaloff@yandex.ru,* [b]*avakyants@genphys.phys.msu.su*

Поступила 10.12.2020

Главное содержание исследования составляет теоретическое обоснование возможности возбуждения неравновесного Оже – перехода. Основное внимание в работе авторы акцентируют на возможные механизмы обменного взаимодействия электронных оболочек, которые приводят к возникновению данного квантового эффекта. В качестве исследовательской задачи была определена попытка провести вычисление вероятности неравновесного Оже – перехода и оценить его скорость. Результаты показывают, что данный процесс характеризуется значением $\tau \sim 10^{-14}$ с. Таким образом, в нашей статье впервые обосновывается мысль о возможности возбуждении нового неравновесного Оже – перехода и приводится его вероятностная оценка.

Ключевые слова: вероятность, квантовые переходы, Оже – спектры, эмиссия, зонд.

УДК 13.2
PACS: 12.90.+b
DOI: 10.31145/2224-8412-2020-21-1-37-49

Введение

В настоящее время актуальны методы анализа структуры вещества по энергетическому спектру потока вторичных частиц. Перспективу исследований открывает метод эмиссионной Оже – спектроскопии, позволяющий установить структуру поверхности твердого тела на атомном уровне, а также распределение частиц во вторичном потоке по скоростям и энергиям [1, 2]. Полученное распределение несет информацию, важную для дальнейших исследований, в частности, характеризует прочность химической связи и степень возбуждения атомных структур в твердом теле. Следует отметить, что метод эмиссионной Оже – спектроскопии может быть использован для анализа состава и структуры химических связей в растворах сложных органических смесей, коллоидных соединений, твердых тел и газовых смесей [3, 4].

Распыление частицами корпускулярных зондов многокомпонентных и многофазных поверхностей представляет большой интерес как с точки зрения получения более полной картины состояния атомов приповерхностного слоя, так и для решения ряда прикладных задач [5, 6, 7]. Это стало темой обсуждения нескольких докладов международной конференции «ВИМС-18», «ВИМС-19» [8, 9]. Актуальность проводимых нами исследований обоснована получением дополнительной информации из Оже – спектров в процессе совершенствования оптических устройств, в частности, разрешающей способности анализаторов в оптических установках. Действительно, апертура современных записывающих оптических систем настроена таким образом, что значительная часть информации теряется при регистрации спектров на основе применения метода Оже – спектроскопии [10]. В частности, теряется информация об особенностях возбужденных атомных структур, о наличии которых свидетельствуют дополнительные амплитудные пики в Оже – спектрах, которые не регистрируются оптической системой из-за узкого выбранного диапазона энергий. Практическая значимость исследования обоснована и заключается в том, что по наличию неравновесных Оже – электронов во вторичном потоке можно судить не только о структуре, но и о состоянии вещества. Таким образом, повышение селективности в процессе интерпретации Оже – спектра приводит к открытию новых особенностей в открытии свойств атомных оболочек [11].

Существование неравновесного динамического процесса как нового эффекта с участием электронов в рамках Оже – рекомбинации другими авторами не исследовалось. Данное исследование впервые ставит перед собой задачу теоретически оценить вероятность нового неравновесного Оже – перехода. Показано, что исследуемый эффект характеризуется временем релаксации $1.82 \cdot 10^{-14}$ с и заключается в том, что система стремится снять возбуждение за счет испускания электронов, энергия которых превышает энергию падающих частиц зонда. Обнаружено, что интенсивность неравновесного Оже – перехода увеличивается с ростом порядкового номера и лучше всего наблюдается для элементов Al, Na, Br. Для реализации поставленных задач была проведена численная оценка вероятности неравновесного Оже – перехода и получен Оже – спектр с целью выявления амплитудных пиков, свидетельствующих о существовании исследуемого эффекта.

Методология и общая концепция

Неравновесные Оже – электроны являются индикаторами состояния исследуемой поверхности. По скорости счета неравновесных Оже – электроны, а также по амплитуде электрического сигнала, генерируемого регистрирующей системой, можно судить о степени возбуждения и электронной структуре атомов, находящихся на поверхности исследуемого образца.

Энергетическое распределение вторичных частиц зависит от их структурных модификаций. Это могут быть возбужденные одиночные атомы или сложные молекулярные кластеры. Следовательно, различия в состоянии молекул приводят к трансформации Оже – спектров [12]. Из них (см., например, обзоры [13, 14]) следует, что энергия частиц, вылетающих с поверхности исследуемого образца, как правило, меньше их исходных значений. Этот результат является превалирующим и проявляется при кулоновском рассеянии, нормальном эффекте Комптона, Оже – рекомбинации и комбинационного рассеяния света [15, 16].

В некоторых случаях наблюдается отклонение от общих закономерностей, когда энергетические параметры падающих частиц ниже их значений после взаимодействия с поверхностными атомами [17]. В [18, 19] показано, что этот эффект наблюдается в спектроскопии комбинационного рассеяния света (эффект Рамана), а также в аномальном эффекте Комптона.

Рекомбинация вакансии и электрона может сопровождать новый механизм взаимодействия электронных оболочек, в котором энергия вторичной частицы больше энергии первичной [20]. Реализация такого квантово-механического процесса обусловлена особенностями электронного обмена атомами в поверхностных слоях, в которых вакансия занята носителем заряда, внешним по отношению к электронной оболочке. Для дальнейшей оценки вероятности неравновесного Оже – перехода стоит остановиться на более подробном рассмотрении кинетики процесса.

В случае классической Оже – рекомбинации, когда возбужденный электрон переходит на более низкие энергетические уровни, испускается квант электромагнитной энергии (γ квант при переходе валентных электронов), который передается другому возбужденному носителю заряда. В результате атомная система избавляется от избыточной энергии за счет испускания Оже – электрона в область пространственного континуума. Однако цель данной работы – рассмотреть новый механизм электронного обмена, в условиях которого возможна эмиссия неравновесного Оже – электрона с энергией, превышающей скорость зондирующих частиц.

На рис.1. представлена энергетическая диаграмма исследуемого процесса с участием электронной пары, локализованной на K – слое и одного неспаренного электрона на L – слое.

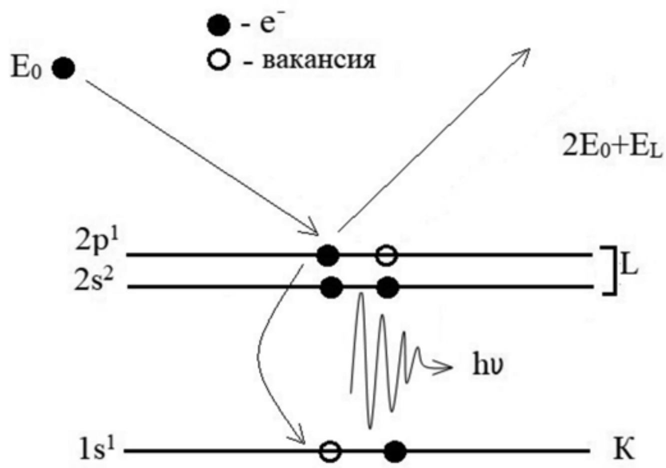


Рис. 1. Энергетическая диаграмма электронного обмена с участием связанных и свободных электронов

Как уже упоминалось выше, поверхность исследуемого образца в рамках эмиссионной Оже – спектроскопии бомбардируется частицами корпускулярного зонда (в данном случае электронными пучками). Скорость распыляемых частиц по порядку величины сравнима с энергией связи $1s^1$-электронов с ядром, что позволяет выбить $1s^1$-электрон и получить вакансию на K-слое. Основная идея данной работы состоит в том, что состояние $2p^1$ занимается падающей частицей зонда, которая не потеряла энергию в приповерхностных слоях исследуемого вещества и не испытывала неупругого рассеяния на атомах кристаллической решетки. В результате перехода электрона из состояния $2p^1$ в состояние $1s^1$ высвобождается квант энергии, который передается частице зонда, занявшей состояние $2p^1$. Развивая эту концепцию, можно сделать вывод, что атомная система избавляется от избыточной энергии за счет испускания неравновесного Оже – электрона с энергией выше, чем у частиц, падающих на поверхность исследуемого образца. В свете вышеизложенного важны следующие оценки и выводы, которые представляют интерес для нашего исследования:

1) Исходя из представлений об электронных конфигурациях атомов наиболее вероятный процесс неравновесного Оже – перехода будет наблюдаться для элементов Al, Na, Br;

2) Влияние примесей и загрязняющих веществ в исследуемом образце сильно повлияет на форму спектра и дальнейшую сложность его интерпретации. Это приводит к необходимости отжига пленки при высоких температурах (около 1000K) в техническом вакууме;

3) Толщина образца должна быть порядка 8 10 атомных слоев. Это связано с экспоненциальным уменьшением вероятности неравновесного Оже – перехода с увеличением глубины проникновения зондирующих частиц в образец;

4) Энергия налетающих электронов должна лежать в диапазоне, соответствующем энергиям связи электронов на K – слое выбранных атомов (порядка 1-15 кВ);

5) Для обеспечения локального воздействия на поверхность образца диаметр зондирующего пучка должен быть узким, что соответствует 1-2 мкм.

Важным моментом для исследований является положение о том, что концентрация электронов в поверхностном слое должна быть высокой, а время перехода между состояниями $2p^1$ и $1s^1$ должно быть достаточно коротким. Действительно, когда пробная частица выбивает электрон из K – слоя, образуется вакансия, которая заполняется за очень короткое время. Необходимо, чтобы $2p^1$-электрон успел его заполнить, а значит, плотность электронного окружения в месте локализации атома должна быть высокой.

Математическая модель

Оценим численно возможность реализации неравновесного Оже – перехода. Вероятность перехода для нормального эффекта Оже в общем случае рассчитывается с помощью формулы [13]

$$W_A = \frac{2\pi}{\hbar}\rho(k)\left|\int\int \varphi_f(r_1)\psi_f(r_2)U\varphi_i(r_1)\psi_i(r_2)\,dV_1\,dV_2\right|^2, \tag{1}$$

где интегрирование производится по всему объёму; $\rho(k) = mVk \cdot d\Theta / 8\pi^3\hbar^2$ – плотность состояний с нормировкой для ящика объёмом V; $d\Theta = \sin\theta d\theta d\phi$ – элемент телесного угла; k – модуль волнового вектора; $\hbar = 1.055 \cdot 10^{-34}$ Дж·с; $e = 1.6 \cdot 10^{-19}$ Кл; $U = e^2/(r_1 - r_2)$ – потенциал взаимодействия. Волновые функции KLL – перехода в водородоподобном атоме имеют вид [14]:

$$\varphi_i(r_1) = \frac{r_1}{a\sqrt{6a^3}}\exp(-r_1/2a)Y_{1m}(\theta_1,\phi_1), \qquad \psi_f(r_2) = \frac{1}{\sqrt{V}}\exp(ikr_2),$$
$$\varphi_f(r_1) = \frac{2}{\sqrt{a^3}}\exp(-r_1/a), \qquad \psi_i(r_2) = \frac{r_2}{a\sqrt{6a^3}}\exp(-r_2/2\,a), \tag{2}$$

где $a = a_0/Z = \hbar^2/me^2$ – первый радиус Бора; $m = 9.1 \cdot 10^{-31}$ кг; Z – атомный номер.

Сферические функции с точностью до нормировочного коэффициента в общем виде выражаются через присоединённые полиномы Лежандра

$$Y_{l,m}(\theta,\phi) = \Phi(\phi)\cdot i^l(-1)^m\sqrt{\frac{(2l+1)(l-m)!}{2(l+m)!}}(1-x^2)^{m/2}\frac{d^m}{d(x)^m}\left[\frac{1}{2^l l!}\cdot\frac{d^l}{d(x)^l}(x^2-1)^l\right], \tag{3}$$

где l, m, – орбитальное и магнитное квантовые числа; $\Phi(\phi) = 1/\sqrt{2\pi}\exp(im\phi)$ – собственные функции оператора проекции момента количества движения на ось; $x = \cos\theta$.

Тот факт, что радиальная составляющая волновой функции s – состояния мала по сравнению с волновой функцией p – состояния, позволяет применить грубое приближение для потенциала взаимодействия, согласно которому

$$e^2/(r_1 - r_2) \approx (e^2/r_2)(1 + (r_1/r_2)\cos\theta_{12}) = e^2(1/r_2 + (r_1/r_2^2)(C_1C_2 + S_1S_2)), \qquad (4)$$

где $C_1 = \cos\theta_1, C_2 = \cos\theta_2, S_1 = \sin\theta_1, S_2 = \sin\theta_2$; $Y_{10}(\theta_2, \phi_2) = \frac{i}{2}\sqrt{\frac{3}{4\pi}}C_2 = A.$

В водородоподобном атоме энергия неравновесного Оже – электрона E_0 связана с энергиями $(E_k)K$ – и $(E_L)L$ – оболочек соотношением [15]

$$\begin{cases} E_O = E_k - 2E_L \\ E_L = E_K/4 \end{cases} \Rightarrow E_O = E_K/2, \qquad (5)$$

следовательно

$$E_k = (eZ)^2/2a_0 \Rightarrow \ ak = (a_0/Z)\sqrt{mE_K/\hbar^2} = 1/\sqrt{2} \Rightarrow \ k = 1/a\sqrt{2}. \qquad (6)$$

Эмиссия неравновесного Оже – электрона обусловлена тем, что вакансию на L – слое занимает первичный электрон. Следовательно, вместо волновой функции – состояния следует использовать плоскую волну де – Бройля. Вероятностные функции принимают вид

$$\varphi_i(x_1) = \frac{1}{\sqrt{a^3}}\exp(ik'r_1C_1), \qquad \psi_f(x_2) = \frac{1}{\sqrt{V}}\exp(ikr_2C_2).$$
$$\varphi_f(x_1) = \frac{2}{\sqrt{a^3}}\exp(-r_1/a), \qquad \psi_i(x_2) = \frac{Ar_2}{a\sqrt{6a^3}}\exp(-r_2/2a), \qquad (7)$$

Соответствующую вероятность неравновесного Оже – перехода можно оценить на основе интеграла

$$W_A = R\left\{\iint \varphi_f(r_1)\varphi_f^*(r_1)\psi_f(r_2)\psi_f^*(r_2)(1/r_2 + (r_1/r_2^2)(C_1C_2 + S_1S_2))^2 \times\right.$$

$$\left.\times\varphi_i(r_1)\varphi_i^*(r_1)\psi_i(r_2)\psi_i^*(r_2)d^3r_1d^3r_2\right\} = \quad (8)$$

$$= R\left\{\iint (2A^2/3a^{12}V)\exp(-2r_1/a)(1/r_2 + (r_1/r_2^2)(C_1C_2 + S_1S_2))^2 r_2^2 \exp(-r_2/a)d^3r_1d^3r_2\right\},$$

где $R = (2e^4mVk/8\pi^2\hbar^3)\cdot d\Theta$. Полагая, что $d\Theta = 4\pi$, $\int\limits_0^{2\pi} d\phi_1 \int\limits_0^{2\pi} d\phi_2 = 4\pi^2$ и переходя к сферическим координатам, матричный элемент можно рассчитывается на основе формулы

$$W_A = R'\left\{\iiiint A^2\exp(-1/a\cdot(2r_1 + r_2))(1/r_2 + (r_1/r_2^2)(C_1C_2 + S_1S_2))^2 r_2^2 r_1^2 r_2^2 S_1S_2 dr_1dr_2d\theta_1d\theta_2\right\} =$$

$$= R'\left[\iiiint A^2\exp(-1/a\cdot(2r_1 + r_2))r_1^2r_2^2S_1S_2dr_1dr_2d\theta_1d\theta_2 + \right.$$

$$+2\iiiint A^2\exp(-1/a\cdot(2r_1 + r_2))r_1(C_1C_2 + S_1S_2)S_1S_2\,dr_1\,dr_2\,d\theta_1\,d\theta_2 + \quad (9)$$

$$\left.+\iiiint A^2r_1^4\exp(-1/a\cdot(2r_1 + r_2))(C_1C_2 + S_1S_2)^2 S_1S_2\,dr_1\,dr_2\,d\theta_1\,d\theta_2\right],$$

где $R' = 4\sqrt{2}e^4m/3a^{12}\hbar^3$.

Расчёт интеграла (9) производился аналитическими методами. Результаты представлены в таблице 1.

Ширина спектральной линии связана со временем жизни соотношением неопределённости Гейзенберга соотношением

$$\Delta E\cdot\Delta t \geq \hbar. \quad (10)$$

Для энергетических уровней атомов $\Delta t = \tau$, где τ – время жизни возбуждённого состояния. С другой стороны, согласно классическому определению соотношению неопределённости

$$\Delta p_x\cdot\Delta x_{per} \geq \hbar, \quad (11)$$

где Δx_{per} – длина свободного пробега вторичной частицы (рис.2).

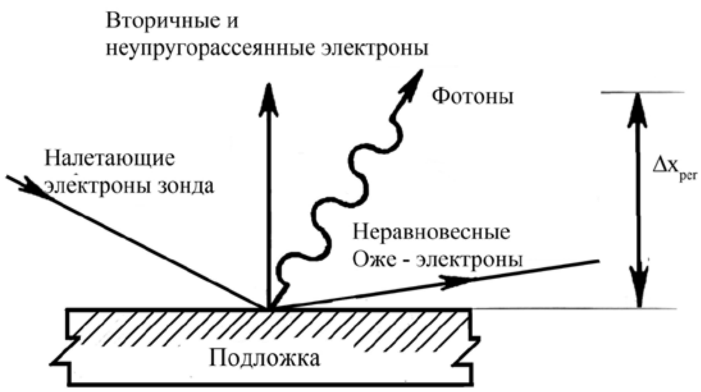


Рис.2. Схематическое изображение процесса бомбардировки поверхности электронами зонда. Показаны падающие электроны, а также частицы во вторичном потоке

Учитывая, что неопределённость в импульсе связана с неопределённостью в энергии соотношением $\Delta p_x = 2m\Delta E$, то

$$\begin{cases}\Delta E \geq \hbar / 2m\Delta x_{per} \\ \Delta E \geq \hbar / \tau\end{cases} \Rightarrow \quad \begin{cases}\Delta E \cong \hbar / 2m\Delta x_{per} \\ \Delta E \cong \hbar / \tau\end{cases} \Rightarrow \begin{cases}\Gamma \cong \hbar / 2m\Delta x_{per} \\ \Gamma \cong \hbar W_A\end{cases} \qquad (12)$$

Два последних соотношения определяют характер зависимости уширения спектральной линии как функции массы налетающей частицы и скорости неравновесного Оже – перехода.

Обсуждение результатов

Результаты расчёта интеграла (9), характеризующего энергетические переходы KLL с образованием неравновесного Оже – электрона для водородоподобного атома показывают, что данный процесс характеризуется высокой вероятностью W_A=0.386·10^{13} с$^{-1}$. В таблице 1 представлены теоретически рассчитанные по формуле (7) значения W_A и времена жизни возбуждённого состояния $\tau = 1/W_A$. Приведены атомные массы химических элементов с учётом изотопического состава m.

Таблица 1

Теоретические значения неравновесных Оже – переходов. Представлены данные энергетического уширения вторичных частиц

Порядковый номер	Элемент	W_A,10^{13} с$^{-1}$ Неравновесный Оже – переход	τ , 10^{-14} с	Γ, мэВ
35	Br	0.78	2.15	4.8
11	Na	0.50	1.98	5.2
12	Mg	0.54	1.82	5.7
13	Al	0.56	1.75	5.9
14	Si	0.61	1.63	6.4

В таблице приняты обозначения, согласно которым Г – ширина спектральной линии, соответствующая энергетическому распределению частиц во вторичном потоке; W_A, τ – вероятность и время неравновесного Оже – перехода в возбуждённой атомной структуре.

Из данных таблицы 1 видно, что с увеличением порядкового номера атома ширина полосы пропускания и вероятность неравновесного Оже – перехода увеличиваются. Стоит отметить, что неравновесные Оже – электроны, вышедшие из более глубоких слоёв, теряют энергию из-за неупругих столкновений и не дают вклад в Оже – линии, однако уширяют их.

Положение максимумов, соответствующих неравновесным Оже – переходам, можно определить с точностью до ±5 эВ, в то время как повышение селективности Оже – анализа с поверхности плёнки или твёрдого тела приводит к неизбежному сужению интенсивности и ширины пропускания энергоанализатора [16, 17, 18]. Следовательно, значительная часть информации о состоянии исследуемого вещества теряется.

Учитывая указанные особенности, методика выращивания образцов выбрана в пользу атомно – слоевого осаждения газовой фазы напыляемого вещества на подложку с выбранными адгезионными свойствами. Формирование исследуемых образцов проводилось на установке «ALD P-1000 Picosun» за счет последовательного нанесения моноатомных слоев, что обеспечило непрерывный контроль напыляемого покрытия. Образцы представляли собой тонкие пленки, полученные осаждением оксида серебра Ag_2O на подложке Si_3N_4 толщиной 2.5 мкм. Дальнейший отжиг при температуре обработки 170ºC производился с целью модификации Ag_2O в его кубическую η-фазу и устранения загрязняющих примесей. Рис. 3, 4 дают изображение экспериментальной кривой распределения вторичных заряженных частиц по энергиям, а также результаты первого и второго дифференцирования.

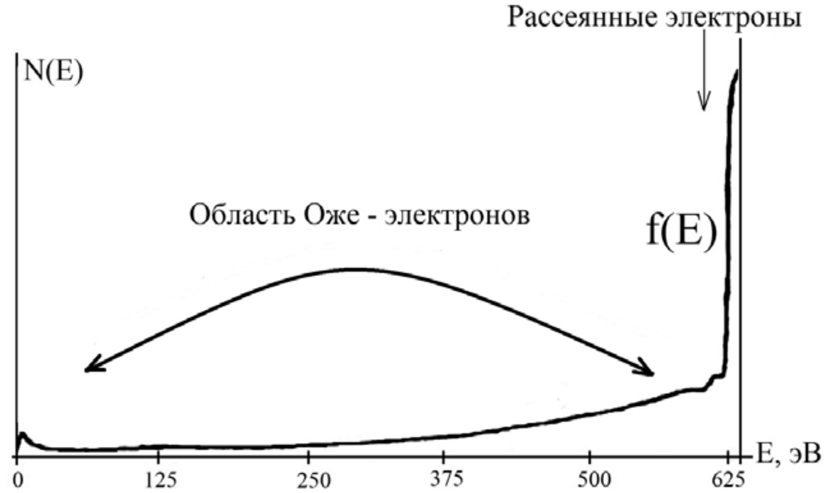


Рис.3. Экспериментальный Оже – спектр

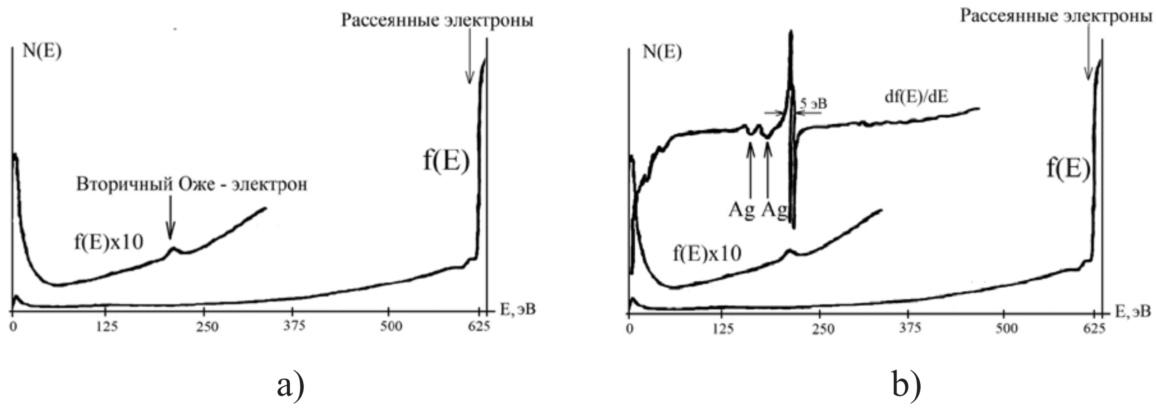


Рис.4. a) Усиленный в 10 раз результат первого дифференцирования экспериментального Оже – спектра; b) Результат второго дифференцирования и выявленные фундаментальные переходы

Неравновесные Оже – электроны являются частью всех вышедших с поверхности образца частиц, имеющих различное происхождение [19, 20]. Следует отметить, что при интерпретации Оже – спектров следует анализировать весь энергетический диапазон, чтобы найти характерные пики, соответствующие неравновесным Оже – переходам.

Операция дифференцирования кривой вольт – амперной характеристики позволяет выявить особенности Оже – процесса при обнаружении максимумов, величина которых соответствует частоте переходов. На рис.4. нижняя кривая $f(E)=dN/dE$, которая является результатом первого дифференцирования экспериментальной кривой вольт – амперной характеристики, даёт распределение вторичных электронов по энергиям. Второе дифференцирование (см. верхнюю кривую $df(E)/dE$) выявляет дополнительные особенности спектра возбуждённой атомной структуры. Важно отметить, что именно неравновесные Оже – электроны дают основной вклад в возрастающий участок производной, отвечающей области истинно вторичных электронов.

Установлено, что спектральная ширина локального максимума, изображенного на рис.4. b) определяется наличием во вторичном потоке именно неравновесных Оже – электронов. Действительно, вышедшие с поверхности неравновесные Оже – электроны не могут усиливать сигнал или давать дискретный набор линий в силу их малой интенсивности во вторичном потоке. Однако они вносят основной вклад в возрастающий участок производной и уширяют локальный максимум, что видно из рис.4.b). Ширина линии спектра, равная 5эВ является нижним пределом в обнаружении неравновесных Оже – спектров, начиная с которого интенсивность их вылета с поверхности исследуемого образца увеличивается с ростом энергии падающих электронов. Это приводит в дальнейшем к уширению локального максимума.

Выводы

Установлено, что одним из возможных каналов релаксации атомных или кластерных структур в поверхностных слоях твердого тела является эмиссия неравновесных Оже – электронов. Эффект неравновесного Оже – перехода, исследуемый в данной работе, заключается в том, что вакантное место заполняется падающим электроном корпускулярного зонда, а энергия перехода высвобождается и передается эмитирующему электрону зонда. Показано, что этот процесс становится конкурирующим с нормальной Оже – рекомбинацией с увеличением порядкового номера химического элемента и с увеличением радиуса атома в приповерхностном слое.

В ходе работы удалось получить математическую модель, описывающую процесс неравновесного Оже – перехода и позволяющую получить численную оценку вероятности этого эффекта. Следует отметить, что математическая модель, полученная в рамках исследования, была основана на общем формализме механических кинетических процессов. Представленный в работе расчет вероятности неравновесного Оже – перехода проводился с учетом кулоновского взаимодействия электрон-электронного взаимодействия и учитывал частичное экранирование атома остальными частицами. На основании полученных данных в данной работе показано, что релаксация атомной структуры, сопровождающаяся эмиссией неравновесного Оже – электрона, характеризуется временем релаксации $\sim 10^{-14}$с. В рамках математической модели удалось получить численную оценку вероятности неравновесного Оже – перехода $\sim 10^{13}$с$^{-1}$.

Проведенный расчет вероятности с учетом кулоновского частичного экранирования атомов позволил изучить кинетику неравновесного процесса. Важно, что по наличию во вторичном потоке неравновесных Оже – электронов можно судить о прочности химической связи атомов и электронной структуре атомных оболочек. Это необходимо учитывать в процессе совершенствования оптических приборов, в частности разрешающей способности анализаторов в оптических установках.

Авторы работы полагают, что этот эффект можно наблюдать не только с помощью Оже – электронной спектроскопии, но и при возбуждении атомов твердых тел ионами средних или высоких энергий и фотонными зондами.

Литература

1. *Зельцер И.А., Моос Е.Н., Кукушкин С.А.* // Письма в ЖТФ. 2008. № 13. С. 56 – 60.
2. *Абгарян В.К., Гидаспов В.Ю., Надирадзе А.Б., и др.* // Письма в ЖТФ. 2019. **45**. № 4. С. 4 – 6.
3. *Русских А.Г., Орешкин В.И., Жигалин А.С., и др.* // Письма в ЖТФ. 2016. **42**. № 5. С. 3 – 7.
4. *Repsilber T., M. Borghesi, J., Gauthier C., et al.* // Appl. Phys. 2005. **80**, N 134, P 905 – 910.
5. *Faenov A. Ya., Pikuz T. A., Fukuda Y., et al.* // Appl. Phys. Lett. 2009, **95**, N 27, P 673 – 680.
6. *Ландау Л.Д., Лифшиц Е.М.* Квантовая механика. Нерелятивистская теория. – М.: Физматлит. Т.3. М., 1963.
7. *Windhorn L., Witte T., Yeston J. S., et al.* // Chem. Phys. Lett. 2002, **37**, N 14, P 120 – 189.
8. *Windhorn L., Yeston J.S., Witte T., et al.* // J. Chem. Phys. 2003, **641**, N 358, P 119 – 127.
9. *Малышев В.И.* // УФН. 1957. **LXII**. С. 320 – 350.
10. *Сущинский М.М., Бажулин П.А.* // УФН. 1957. **LXII**. С. 301 – 320.
11. *Ландсберг Г. С., Мандельштам Л. И.* // Журнал Русского физ.-хим. об-ва. 1928. **60**. С. 55 – 61.
12. Ковальчук М.В. // Приборы и техника эксперимента. 1987. 3. С. 191-195.
13. *Hertel N.* // Phys. Lett. A. 1980, **75**, N 6, P 501 – 502.
14. *Зельцер И.А., Кукушкин С.А., Моос Е.Н.* // Известия РАН. Серия физическая. 2008. **72**. №7. С. 873-877.
15. *Burmistrov E.R., Afanasova M.M.* // SN Applied Sciences. 2020, **2**, N 1532, 10 p.
16. *Райхбаум Я.Д., Костюкова Е.С.* // Изв. АН СССР, сер. Физ. 1954. **18**. №2. С. 289 – 290.
17. *Богачев Г.Г., Ремета Е.Ю.* // Оптика и спектроскопия. 2020. **128**, № 10. 9 с.
18. *Иванова Е.П.* // Оптика и спектросокпия. 2018. **125**. № 8. С. 3-13.
19. *Коробов А.И., Агафонов А.А., Изосимова И.Ю.* // ЖТФ. 2018. **88**. № 3. 8 с.
20. *Хисамов Р.Х., Юмагузин Ю.М., Мулюков Р.Р., Назарок К.С. и др.* // Письма в ЖТФ. 2013. 7. С. 76-83.

POSSIBILITY OF EXCITATION OF NONEQUILIBRIUM AUGER TRANSITIONS BY CORPUSCULAR PROBES

Burmistrov E.R.[1,a], Avakyants L.P.[1,b]

[1]*Department of General Physics, Faculty of Physics,*
M.V. Lomonosov Moscow State University. Moscow 119991, Russia.
E-mail: [a]eugeni. conovaloff@yandex.ru, [b]avakyants@genphys.phys.msu.su

Received 10.12.2020

The main content of the research is the theoretical substantiation of the possibility of excitation of a nonequilibrimn Auger transition. The main attention in the work is focused on the possible mechanisms of the exchange interaction of electron shells which lead to the appearance of this quantum effect. As a research problem, an attempt was made to calculate the probability of a nonequilibrimn Auger transition and to estunate its speed. The results show that this process is characterized by a value of 10^{-14}s. Thus m our article for the first time the idea of the possibility of exciting a new nonequilibrimn Auger transition is substantiated and its probabilistic estimate is given.

Keywords: probability, quantum transitions, Auger – spectra, emission, probe.

НА ПУТИ К ТЕОРЕТИЧЕСКОЙ БИОЛОГИИ. I. ПРОЛЕГОМЕНЫ.

Перевод с английского
С.Г. Васецкого

Под редакцией и с предисловием
акад. Б.Л. Астаурова

Издательство «Мир» Москва, 1970.

Предисловие к английскому изданию

Теоретическая физика представляет собой вполне сложившуюся самостоятельную науку, и во многих университетах ею занимаются специальные лаборатории и кафедры. Более того, наши теории о природе окружающего нас материального мира, безусловно, оказывают глубокое влияние на общефилософские концепции. Что же касается теоретической биологии, то едва ли можно сказать, что такая наука уже существует. Трудно сказать, чем она должна заниматься и по каким путям ей следует развиваться; к тому же очень редко случается, что философы ощущают связь таких биологических проблем, как теория эволюции или восприятие раздражения, с традиционными проблемами философии.

Международный союз биологических наук (МСБН) счел своим долгом, как организация, объединяющая биологов из разных стран, стимулировать создание некоего костяка понятий и методов, на котором могла бы формироваться теоретическая биология. Это совсем не простая задача; поэтому было решено провести три симпозиума на эту тему с годичными интервалами. Эти симпозиумы предполагалось посвятить не обсуждению теоретических основ каких-либо частных биологических процессов, например проницаемости мембран, наследственности, нервной деятельности и т. д., а попыткам выявить и сформулировать основные концепции и логические связи, характеризующие живые системы в отличие от неживых, и рассмотрению вытекающих из них общефилософских представлений.

На меня была возложена обязанность пригласить докладчиков и организовать заседания.

Первый симпозиум проходил с 28 августа по 3 сентября 1966 г. на вилле Сербеллони в Беладжо (озеро Комо). Чтобы создать известную базу для дискуссии и сосредоточить внимание на некоторых проблемах, я разослал участникам симпозиума свои лекции, прочитанные за год до этого в университете Северного Уэльса и нарочито переработанные с тем, чтобы придать им несколько полемический характер. Одновременно были разосланы некоторые комментарии Рене Тома к этим лекциям, а также статья Эрнста Майра.

Заседания на вилле Сербеллони, носившие весьма непринужденный характер и оказавшиеся очень плодотворными, не стенографировались.

В процессе обсуждения внимание было сосредоточено главным образом на проблемах биологической теории, а не на более общих проблемах. Хотя в результате работы симпозиума стали вырисовываться пусть еще не очень четкие, но уже определенные контуры теоретической биологии, было совершенно ясно, что необходимо продолжить обсуждение и обмен мнениями между приверженцами различных точек зрения, прежде чем удастся разработать некое подобие схемы стройной и самостоятельной науки. Поэтому предлагаемая вниманию читателя книга состоит из отдельных статей, написанных после симпозиума в духе проводившегося на нем обсуждения. Они еще не связаны друг с другом в некое единое целое. Именно сознание того, что такого единого целого не существует, что его создание представляет собой длительную и нелегкую задачу, и заставило принять решение провести три симпозиума. Мы надеемся, что на втором симпозиуме будут сделаны дальнейшие шаги на пути к синтезу различных точек зрения. Поэтому этот первый том и получил подзаголовок «Пролегомены».

К. Уоддингтон

ПРИЧИНА И СЛЕДСТВИЕ В БИОЛОГИИ

Э. Майр
(Гарвардский университет)

Продолжение. Начало в предыдущих номерах.

Будучи по профессии биологом, я, разумеется, не могу предпринять такой анализ причины и следствия в биологических явлениях, какой провел бы логик. Вместо этого мне хотелось бы сосредоточить внимание на особых трудностях, связанных с классической концепцией причинности в биологии. Эти трудности начинают терзать исследователей при первых же попытках создать единое понятие причины. Крайне механистическое представление Декарта о жизни и та логическая крайность, до которой его идеи были доведены Гольбахом и Де Л а Меттри, неизбежно вызывают обратную реакцию, приводящую к созданию виталистических теорий, время от времени входящих в моду вплоть до наших дней. Среди наиболее видных представителей этого направления достаточно назвать Дриша (энтелехия), Бергсона (жизненный порыв) и Леконта дю Нуи. Хотя точки зрения этих исследователей в каких-то деталях различны, все они сходятся в тем, что живые существа и жизненные процессы не поддаются причинному объяснению в рамках физических и химических явлений. Наша задача заключается в том, чтобы задать вопрос, оправданно ли это мнение, и в случае отрицательного ответа установить источник заблуждений.

Принято считать, что причинность независимо от ее определения в терминах логики содержит три элемента: 1) объяснение прошедших событий (причинность a posteriori); 2) предсказание грядущих событий; 3) истолкование телеологических, т. е. «целенаправленных», явлений.

Три аспекта причинности (объяснение, предсказание и телеология) должны составлять основные вехи любого обсуждения причинности и совершенно справедливо были выделены в качестве таковых Нагелем [1]. Биология может внести существенный вклад в обсуждение всех трех аспектов причинности. Но прежде чем перейти к детальному разбору этого вклада, я должен сказать несколько слов о биологии как науке.

Биология. Термин *биология* подразумевает некую однородную и единую науку. Однако события последних лет со все большей ясностью показывают, что биология — наука в высшей степени сложная. В сущности «биологией» называют две в значительной мере самостоятельные области науки, различающиеся по методу, проблематике и основным концепциям. Едва зайдя за уровень чисто описательной биолоти, мы находим две сильно различающиеся области, которые можно было бы назвать *функциональной* и *эволюционной* биологией. Разумеется, эти две области во многих точках соприкасаются и перекрываются. Любой биолог, работающий в одной из этих областей, должен иметь представление о другой области, если он хочет избежать ярлыка ограниченного специалиста. Однако его собственные исследования обычно бывают посвящены проблемам в какой-либо одной из этих областей. Мы не можем обсуждать понятия причины и следствия в биологии, не охарактеризовав предварительно эти две области.

Функциональная биология. Специалист в этой области имеет дело прежде всего с действием и взаимодействием структурных элементов — от молекул до органов и целых особей. Он непрерывно задается вопросом: «Как?» — как действует тот или иной элемент, как он функционирует? Этот подход объединяет анатома, изучающего сочленения, с молекулярным биологом, исследующим роль молекулы ДНК в передаче генетической информации. Функциональный биолог стремится выделить тот частный элемент, который он изучает, и в каждом данном исследовании он обычно имеет дело с одной особью, одним органом, одной клеткой или одной частью клетки. Он пытается исключить все переменные или управлять ими и повторяет свои опыты при постоянных или варьирующих условиях до тех пор, пока не убедится, что он выяснил функцию изучаемого им элемента. Эксперимент служит ему основным методом исследования, и его подход к изучаемым явлениям по существу сходен с подходом физика или химика. И в самом деле, вычленив в достаточной мере изучаемое явление из всей совокупности процессов жизни, он может достигнуть идеала — свести свое исследование к чисто физическому или химическому эксперименту. Несмотря на некоторую ограниченность этого метода, нельзя не согласиться, что подобный упрощенный подход абсолютно необходим для достижения тех целей, которые ставит себе такой биолог. Бурное развитие биохимии и биофизики оправдывает этот прямой, хотя и явно упрощенный подход.

Эволюционная биология. Биолог-эволюционист пользуется другими методами, и его интересуют иные проблемы. Он главным образом задает вопрос: «Почему?» Спрашивая: «Почему?», мы всегда должны сознавать двусмысленность этого вопроса. Он может означать: «Как же это происходит?», но его можно понять и в телеологическом смысле: «Для чего?» Ясно, что эволюционист, задавая этот вопрос, имеет в виду первое из этих значений. Любой организм, особь или вид является продуктом длительной истории, истории, насчитывающей более двух миллиардов лет. Как сказал Макс Дельбрюк [2], «зрелого физика, впервые сталкивающегося с проблемами биологии, ставит в тупик то обстоятельство, что в биологии нет «абсолютных явлений». Каждое явление представляется иным в разных местах и в разное время. Любое животное, растение или микроорганизм, которое изучает биолог, — лишь одно звено в эволюционной цепи изменяющихся форм, ни одна из которых не остается сколько-нибудь постоянной». Едва ли можно до конца понять какую-нибудь структуру или функцию в организме, не изучив ее становления в ходе эволюции. Основная задача биолога-эволюциониста заключается в том, чтобы найти причины имеющихся у данного организма признаков, и в частности приспособлений. Его поражает невероятное многообразие органического мира. Он хочет знать причины этого многообразия, а также путь, которым оно было достигнуто. Он изучает силы, вызывающие изменения в фауне и флоре (частично зафиксированные палеонтологией), и этапы развития удивительных приспособлений, столь характерных для органического мира.

Мы можем охарактеризовать эти две области биологии в несколько ином плане, воспользовавшись языком теории информации. Специалист в области функциональной биологии имеет дело со всеми аспектами реализации запрограммированной информации, содержащейся в ДНК-м коде оплодотворенной зиготы, тогда как биолога-эволюциониста интересуют история этих кодов и законы, управляющие их изменениями от поколения к поколению. Иными словами, его интересуют причины этих изменений.

Многие давние спорные вопросы в области философии естествознания можно значительно точнее сформулировать на языке теории информации. Например, как показано И. И. Шмальгаузеном в СССР и независимо от него мною, невозможность наследования приобретенных признаков становится совершенно ясной, если построить модель передачи генетической информации с периферии (от фенотипа) к ДНК половых клеток.

Однако при всем том не следует переоценивать эти генетические коды. Их характерная черта состоит в том, что заключенная в них программа не предопределяет все и вся в организме. Такие явления, как обучение, память, негенетическое структурное изменение и регенерация, показывают, в какой мере «открыты» эти программы. Но даже и эти явления весьма специфичны, если говорить о том, чему может «обучиться» организм, на какой стадии жизненного цикла происходит обучение и как долго сохраняется в памяти выученное. Таким образом, хотя программа может быть в некоторой своей части совершенно неспецифичной, однако диапазон возможных отклонений в ней зафиксирован. Поэтому коды в некоторых отношениях высокоспецифичны, в других же они лишь определяют «нормы реакции» или общие способности и потенции.

Я попытаюсь показать эту двойственность кодов на примере различий в способности к «распознаванию особей своего вида» у двух видов птиц. Птенца желтушника выкармливают приемные родители, например какие-нибудь певчие птицы из воробьиных. Как только птенец перестает нуждаться в помощи приемных родителей, он начинает искать других слетков своего вида, хотя никогда раньше их не видел. Напротив, молодой гусенок, только что вылупившийся из яйца, принимает за родителей первый же попавшийся ему на глаза движущийся (и предпочтительно также издающий звуки) объект, за которым он может следовать и который он может запечатлеть в мозгу. В первом случае запрограммирован определенный «образ», а во втором — просто способность к импринтингу. Подобные различия в специфичности наследуемой программы характерны для всего органического мира.

Вернемся к основной теме нашего исследования и поставим вопрос о том, идентичны ли понятия *причины* в функциональной и эволюционной биологии.

И вновь Макс Дельбрюк [2] напоминает нам, как еще в 1870 г. Гельмгольц постулировал, что «поведение живых клеток должно быть объяснимо в терминах движения молекул, подчиняющегося определенным законам взаимодействия сил». В настоящее же время, как справедливо отмечает Дельбрюк, мы не можем даже объяснить поведение отдельного атома водорода. Он напоминает, что «любая живая клетка несет в себе опыт экспериментирования ее предков на протяжении миллиарда лет».

Трудности, возникающие при рассмотрении понятия причинности в биологии, можно проиллюстрировать на следующем примере. Давайте зададим вопрос: в чем причина перелетов птиц? Или более конкретно: почему славка, живущая в саду моего загородного дома в Нью-Гемпшире, начинает свой перелет на вечером 25 августа?

Я могу перечислить четыре в равной степени законные причины этого перелета:

1. Экологическая причина. Славки, будучи насекомоядными птицами, должны мигрировать на юг, так как, оставшись на зиму в Нью-Гемпшире, они погибнут от голода.

2. Генетическая причина. В процессе эволюции вида славка приобрела генетическую конституцию, в которой заложена способность правильно реагировать на соответствующие воздействия внешней среды. Так, совка, гнездящаяся рядом со славкой, не обладает такой конституцией и не реагирует на эти воздействия. В результате она ведет оседлый образ жизни.

3. Внутренняя физиологическая причина. Славка летит на юг, так как она привыкла к определенной длине дня. Как только число дневных часов уменьшится до определенного уровня, она готова мигрировать.

4. Внешняя физиологическая причина. Славка отправилась в путь 25 августа, потому что в этот день над нашей местностью прошли холодные массы воздуха, принесенные северными ветрами. Резкое падение температуры и связанное с этим ухудшение погоды подействовали на птиц, уже находившихся в состоянии общей физиологической готовности к перелету, и они начали перелет именно в этот день.

Среди названных нами четырех причин перелета этой птицы можно выделить комплекс причин, прямо вызвавших перелет: это физиологическое состояние птиц, связанное с фотопериодичностью, и падение температуры. Их можно назвать *непосредственными* причинами перелета. Две другие причины — недостаток пищи зимой и генетическая конституция вида — можно назвать *основными*, или *первичными*. Именно эти причины возникли и исторически вошли «в плоть и кровь» системы в результате естественного отбора на протяжении многих тысяч поколений. Очевидно, что задача функциональной биологии заключается в анализе непосредственных причин, тогда как эволюционная биология занята изучением первичных причин. Так обстоит дело почти с каждым биологическим явлением, которое мы захотели бы изучать. Всегда есть какие-то непосредственные и первичные причины, причем для того, чтобы до конца понять данное явление, необходимо выявить и объяснить как те, так и другие.

Различия между этими двумя категориями причин можно также выразить, сказав, что непосредственные причины определяют реакции особи (и ее органов) на факторы внешней среды, тогда как первичные причины ответственны за эволюцию того частного ДНК-го кода, которым обладает каждая особь любого вида. Эти различия, по-видимому, не столь существенны с точки зрения логики. Однако биологи знают, что многих ожесточенных споров о «причине» какого-нибудь биологического явления можно было бы избежать, если бы противники поняли, что один из них имеет в виду непосредственные, а другой — первичные причины этого явления. Я могу иллюстрировать сказанное цитатой из Леба [3]: «Ранние исследователи объясняли рост ног у головастика лягушки или жабы приспособлением к жизни на земле. Но благодаря опытам Гудер– нача мы знаем, что рост ног можно вызвать в любое время, даже у только что вылупившегося головастика, который еще не может жить на суше, если давать животному с кормом щитовидную железу».

Вернемся теперь к определению «причины» в формальной логике и посмотрим, насколько оно совпадает с понятиями «причины» в функциональной и эволюционной биологии. Мы можем, например, определить причину как «условие (необходимое, но недостаточное), без которого явление не могло бы произойти», или как «одно из множества в совокупности достаточных условий, без которого явление не могло бы произойти» [4]. Подобные определения вполне адекватно описывают причинные связи в некоторых областях биологии, в частности в тех, которые рассматривают химические и физические основы явлений. В строго формальном смысле такие определения приложимы и к более сложным явлениям. Однако в областях биологии, связанных с изучением сложных явлений, польза от таких определений невелика. Вряд ли кто-нибудь среди ученых сомневается в причинном характере всех биологических явлений, т. е. в том, что любое уже свершившееся явление можно в конечном счете объяснить с точки зрения его причины. И тем не менее такое объяснение во многих случаях неизбежно будет столь неспецифичным и столь формальным, что его ценность, несомненно, окажется сомнительной. При изучении какой-либо сложной системы едва ли можно считать откровением следующее объяснение: «Явление *A* обусловлено сложным рядом взаимодействующих факторов, одним из которых служит *B*». А часто ничего более сказать нельзя. Мы еще вернемся к этой трудности в связи с проблемой предсказания. Однако прежде следует рассмотреть проблему телеологии.

Телеология. Никакое обсуждение причинности не будет полным, если оно не затрагивает проблему телеологии. Эта проблема берет свое начало от «конечных» причин, входивших еще в аристотелеву классификацию причин. Эта категория причин создана в результате размышлений об упорядоченном и целенаправленном развитии особи от яйца до «конечной» стадии взрослого организма и о развитии вселенной от ее начала (хаоса?) до современного упорядоченного состояния. Конечная причина была определена как «причина, ответственная за упорядоченное достижение заранее известной конечной цели». Всякое целенаправленное поведение относили к категории «телеологических», но к этой же категории относили и многие другие явления, которые по своей природе не были обязательно целенаправленными.

Исследователи научного наследия Аристотеля справедливо обращали внимание на то, что Аристотель — по своему образованию и по призванию — был прежде всего биологом и что именно его увлеченность биологическими явлениями оказала решающее влияние на его представления о причинах и привела к тому, что помимо материальной, формальной и эффективной причин он ввел еще и категорию конечной причины. Мыслители от Аристотеля до наших дней ломали голову над явным противоречием между механистическим толкованием естественных процессов и кажущейся целенаправленной последовательностью процессов роста, размножения и поведения животных. Такой разумный ученый, как Бернард [5], выразил этот парадокс в следующих словах:

«Существует, так сказать, заранее установленный план каждого существа и каждого органа, согласно которому любое явление само по себе зависит от общих сил природы, но если его рассматривать вместе с другими явлениями, то создается впечатление, что какая-то невидимая сила направляет это явление на тот путь, по которому оно следует, и приводит его на занимаемое им место.

Мы допускаем, что жизненные процессы связаны с физико-химическими проявлениями, но нельзя не признать, что это не позволяет объяснить самую их суть, ибо случайное сочетание физико-химических явлений не может привести к образованию каждого организма по твердо установленному (заранее) плану и к возникновению удивительной соподчиненности и гармоничной согласованности жизненных процессов... Детерминизм может быть лишь физико-химическим детерминизмом. Жизненная сила и жизнь относятся к миру метафизики».

Что же представляет собой этот *x*, этот очевидно целенаправленный агент, эта «жизненная сила» в органических явлениях? Лишь в наше время были предложены удовлетворительные объяснения этого парадокса.

Многие дуалистические, телеологические и виталистические теории прошлого попросту подменяли неизвестное *x* другим неизвестным, *y* или *z*, так как назвать неизвестный фактор «энтелехией» или «жизненным порывом» еще не значит объяснить его. Я не буду тратить времени на демонстрацию того, сколь ошибочными были эти попытки в своем большинстве. Даже несмотря на то, что некоторые наблюдения, положенные в основу этих общих схем, вполне правильны, из них были сделаны совершенно неверные, противоестественные выводы.

Так в каких же случаях можно говорить о цели и целенаправленности в природе и в каких нельзя этого делать? На этот вопрос мы теперь можем дать четкий и недвусмысленный ответ. Индивидуум, который, выражаясь языком теории информации, «запрограммирован», может действовать целенаправленно. Однако исторические процессы *не могут* действовать целенаправленно. Птица, начинающая свой перелет; насекомое, находящее свое растение-хозяина; животное, спасающееся от хищника; самец, старающийся привлечь внимание самки, — все они действуют целенаправленно, поскольку они соответствующим образом запрограммированы. Когда я говорю о запрограммированном индивидууме, я употребляю это выражение в широком смысле. Запрограммированное счетно-решающее устройство также представляет собой «индивидуум» в этом смысле, равно как и пара птиц, чьи инстинктивные и приобретенные в результате обучения реакции подчиняются, так сказать, единой программе.

Полностью индивидуальная, но при этом также видоспецифичная программа, закодированная в ДНК любой зиготы (оплодотворенной яйцеклетки), которая регулирует развитие центральной и периферической нервной системы, органов чувств, гормонов, физиологии и морфологии, представляет собой *программу* счетно-решающего устройства, определяющего поведение этой особи.

Естественный отбор всемерно благоприятствует выработке кодов, гарантирующих такое поведение, которое повышает приспособленность. Программа поведения, гарантирующая мгновенную правильную реакцию на потенциальный источник пищи, потенциального врага или потенциального партнера для спаривания, безусловно, обеспечит большую приспособленность в дарвиновском смысле, чем программа, не предусматривающая таких реакций. Равным образом программа поведения, допускающая соответствующее обученней совершенствование поведенческих реакций посредством различного рода обратных связей, обеспечивает большую вероятность выживания, чем программа, лишенная этих свойств.

Целенаправленное действие особи, поскольку оно обусловлено свойствами ее генетического кода, целенаправленно не более и не менее, чем действия счетно-решающего устройства, которое реагирует на различные сигналы в соответствии с заложенной в него программой. Это, если можно так сказать, чисто механистическая целенаправленность.

Мы, биологи, уже давно чувствовали, что называть такое запрограммированное целенаправленное поведение *телеологическим* несколько двусмысленно, потому, что это слово используется, кроме того, совсем в другом смысле для обозначения конечной стадии эволюционных приспособительных процессов. Когда Аристотель говорил о конечных причинах, он имел в виду, в частности, изумительную приспособленность, наблюдаемую в растительном и животном царствах. Его интересовало то, что более поздние авторы называли «замыслом» или «планом» в природе. Он считал, что конечными причинами можно объяснить не только мимикрию или симбиоз, но и все другие приспособления животных и растений друг к другу и к физической внешней среде. Представители школы Аристотеля и их последователи спрашивали себя, какой целенаправленный процесс мог привести к возникновению столь упорядоченного плана в природе.

Теперь очевидно, что термины *телеология* и *телеологический применялись* для обозначения явлений двух различных типов: 1) выработки и усовершенствования на протяжении всей истории растительного и животного царств все обновляющейся программы и все совершенствующейся информации, закодированной в ДНК; 2) испытания этих программ и расшифровки этих кодов на протяжении всей жизни каждой особи. Имеются существенные различия, с одной стороны, между целенаправленными поведенческими реакциями или процессами развития особи или системы, которые управляются программой, и с другой — постоянным совершенствованием генетических кодов. Это совершенствование генома представляет собой эволюционное приспособление, регулируемое естественным отбором.

Чтобы избежать путаницы между этими двумя совершенно различными типами целенаправленности, Питтендрай [6] ввел термин *телеономический* для описания всех целенаправленных систем, «не связанных с аристотелевой телеологией». Это негативное определение не только переносит центр тяжести на термин «система», но и не дает возможности четко разграничить две различные аристотелевы телеологии. По-видимому, применение термина *телеономический* следует стоого ограничить, используя его лишь для систем, действующих на основе какой-то программы или закодированной информации. Телеономия в биологии обозначает, по выражению Дж. Гекели [7], «кажущуюся целенаправленность организмов и их признаков».

Столь четкое отделение телеономии, которая имеет поддающуюся анализу физико-химическую основу, от телеологии, имеющей дело со всеобщей гармонией органического мира в более широком плане, весьма полезно, потому что эти два совершенно различных явления часто смешивают друг с другом.

Развитие или поведение особи целенаправленно, а естественный отбор, разумеется, нет. Когда Маклеод [8] писал: «Наиболее вызывающим у Дарвина было то, что он вновь ввел цель в природу», он выбрал неподходящее слово. Слово *цель* совершенно не подходит для эволюционных изменений, которые в конечном счете рассматривал Дарвин. Если организм высокоприспособлен к условиям своей среды, то это обусловлено не какой-то целью, которую ставили себе его предки или какой– либо внешний агент, например «Природа» или «Бог», сотворившие некий верховный замысел или план. Как справедливо сказал Симпсон [9], Дарвин «вымел подобную финалистическую телеологию через парадную дверь».

Подводя итоги нашему обсуждению, можно сказать, что между те– леономией и причинностью нет противоречия, но что научная биология не нашла никаких доводов, которые бы служили подтверждением телеологии в духе, соответствующем разнообразным виталистическим или финалистическим теориям {9—11]. Все так называемые телеологические системы, которые обсуждает Нагель [12], на самом деле служат примерами телеономии.

Проблема предсказания. Третья важная проблема, связанная с причинностью в биологии, — это проблема предсказания. В классической теории причинности пробным камнем для любого причинного объяснения служит возможность использовать его для предсказания будущих событий. Эта точка зрения сохраняется и в современных теориях [13]: «Теория может предсказывать в той же степени, в какой она может описывать или объяснять». Это утверждение ясно свидетельствует о том, что его автор — физик, ибо ни один биолог не отважился бы на такое заявление. Теория естественного отбора позволяет довольно точно описывать и объяснять явления, но она не дает возможности делать надежные предсказания, если не считать таких тривиальных и бессмысленных заявлении, как, например, «более приспособленные особи оставят в среднем более многочисленное потомство». Скрайвен [14] совершенно справедливо подчеркнул, что один из самых существенных вкладов эволюционной теории в философию состоял в том, что она продемонстрировала независимость объяснения от предсказания.

Хотя предсказание и не является неотъемлемой частью причинности, все же каждый ученый стремится к тому, чтобы его причинные объяснения позволили в то же время предсказывать грядущие события с большой точностью. Мы различаем много видов предсказания в биологических объяснениях, так что определить понятие «предсказание» в биологии весьма затруднительно. Сведущий зоогеограф может довольно точно предсказать, какие животные будут найдены на еще не исследованном горном хребте или острове. Точно так же палеонтолог с большой вероятностью предскажет, какие ископаемые можно будет обнаружить во вскрываемом геологическом разрезе. Можно ли считать такое правильное угадывание результатов прошедших событий гениальным предвидением? Это сомнение распространяется и на таксономические предсказания, рассматриваемые ниже. Однако использовать термин *предсказание* для будущих событий представляется совершенно законным. Я приведу четыре примера, чтобы проиллюстрировать диапазон различных категорий предсказания:

1. *Предсказание в классификации*. Если я определил на основе расположения щетинок и относительных размеров головы и глаз, что данная плодовая мушка относится к виду *Drosophila melanogaster*, то я могу «предсказать» многие другие черты строения и поведения, которые я обнаружу при дальнейшем изучении этой мушки. Если я найду новый вид с диагностическими признаками, характерными для рода *Drosophila*, то я сразу же смогу «предсказать» целый ряд его биологических свойств.

2. *Предсказание большей части физико-химических процессов на молекулярном уровне*. Можно весьма точно предсказать большую часть основных биохимических процессов, протекающих в организмах, например пути обмена веществ, а также биофизические явления в простых системах, например физиологическое действие света, тепла и электричества.

Примеры *1* и *2* показывают, что причинные объяснения позволяют делать весьма точные предсказания. Однако в биологии существует множество других обобщений или причинных объяснений, на основании которых нельзя сделать точных прогнозов. Ниже приводятся соответствующие примеры.

3. *Предсказание исхода сложных экологических взаимодействий*. Если сказать, что «заброшенное пастбище на юге Новой Ашлии зарастет тополелистной березой (*Betula populifolia*) и веймутовой сосной (*Pinus strobus*)», то это нередко окажется правильным. Однако еще чаще пастбище зарастает почти одной только веймутовой сосной или же, напротив, сосна отсутствует, а пастбище зарастает вишней (*Prutius*), виргинским можжевельником (*Juniperus virginianus*), кленами, сумахом и некоторыми другими видами.

Другой пример также иллюстрирует невозможность предсказания. Если два вида мучных хрущаков (*Tribolium confusum и T. castaneum*) поместить вместе в какую-либо однородную среду (на просеянную пшеничную муку), то один из этих видов всегда вытеснит другой. При высокой температуре и влажности побеждает *T. castaneum*, а при низкой температуре и влажности — *T. confusum*. При промежуточных условиях исход конкуренции неопределенен и поэтому не может быть предсказан (табл. 1).

Таблица 1

КОНКУРЕНЦИЯ ДВУХ ВИДОВ *Tribolium* [15]

Условия		Число опытов	Число опытов, в которых победил	
температура, °C	влажность, %		*T. confusum*	*T. castaneum*
34	70	30	—	30
29	70	66	11	55
24	70	30	21	9
34,29	30	60	53	7
24	30	20	20	—

4. *Предсказание эволюционных событий*. Возможно, что в биологии труднее всего предсказать дальнейший ход эволюции. Кто бы мог сказать, глядя на рептилий пермского периода, что большая часть наиболее процветающих групп вымрет (причем многие довольно быстро) и что от одной из наименее примечательных ветвей произойдут млекопитающие? Какой исследователь фауны кембрия предсказал бы революционные изменения в жизни морей последующих геологических эр? Нельзя предсказать и исход микроэволюции. Животноводы и эволюционисты вновь и вновь сталкиваются с тем, что независимые и параллельные линии, подвергнутые одному и тому же давлению отбора, реагируют на него с разной скоростью и в форме различных коррелированных реакций.

Как и во многих других областях науки (если не считать некоторые основные химические или физические процессы), предсказание биологических явлений почти всегда носит статистический характер. Мы можем с большой точностью сказать, что среди 1000 новорожденных будет немногим более 500 мальчиков. Но мы не можем предсказать пол данного, еще не родившегося ребенка,

Причины неопределенности в биологии. Не претендуя на исчерпывающее изложение всех возможных причин неопределенности, я могу перечислить четыре класса таких причин. Хотя они до некоторой степени перекрываются, каждый из них можно рассматривать в отдельности.

1. *Случайность события, не связанная с его значением*. Хорошей иллюстрацией этого класса причин неопределенности служит спонтанная мутация, вызванная «ошибкой» в репликации ДНК. Появление данной мутации

никак не связано с эволюционными потребностями отдельного организма или той популяции, к которой он принадлежит. Точные результаты определенного давления отбора непредсказуемы, поскольку реакция на него зависит от мутаций, рекомбинаций и гомеостаза в ходе развития, причем относительная роль каждого из этих факторов варьирует. В определении генетического состава на всех ступенях участвует в качестве одного из основных компонентов этот вид случайности. Сказанное относительно мутационного процесса справедливо также для кроссинговера, распределения хромосом, отбора гамет, выбора партнера и выживания зигот на ранних стадиях развития. Однако ни лежащие в основе этих явлений молекулярные процессы, ни механические перемещения, ответственные за эту случайность, не связаны с их биологическими проявлениями.

2. *Уникальность всех существ на более высоких уровнях биологической интеграции*. Уникальность биологических существ и явлений составляет одно из основных различий между биологией и физическими науками. Физикам и химикам зачастую бывает весьма трудно понять эту подчеркиваемую биологами уникальность, хотя последние достижения в современной физике облегчают понимание этой особенности. Если физик говорит: «Лед плавает в воде», то его утверждение справедливо для любого куска льда и для любого водоема. В подобных случаях отдельные представители данного класса обычно лишены индивидуальности, столь характерной для органического мира, в котором уникальна каждая особь, каждая стадия жизненного цикла, каждая популяция, каждый вид и каждая более высокая таксономическая категория, любые взаимоотношения между особями, все естественные сообщества видов и все этапы эволюции. В приложении к человеку справедливость этих положений очевидна. Однако они в равной мере приложимы ко всем животным и растениям, размножающимся половым путем. Уникальность, разумеется, не может полностью исключить предсказание. Мы можем весьма точно предсказать многие признаки и особенности поведения как человека, так и других организмов. Однако в большинстве случаев эти предсказания (не считая тех, которые относятся к систематике) носят чисто статистический характер. Уникальность особенно присуща эволюционной биологии. В мире уникальных явлений совершенно невозможно выявить общие законы, подобные существующим в классической механике.

3. *Исключительная сложность*. Физик Эльзассер сказал на одном из недавних симпозиумов: «Примечательной чертой всех организмов является их почти неограниченная структурная и динамическая сложность». Это верно. Всякая органическая система обладает столь большим числом обратных связей, гомеостатических механизмов и потенциальных множественных путей обмена веществ, что описать ее полностью невозможно.

Более того, чтобы исследовать такую систему, необходимо разрушить ее, так что анализ станет бесполезным.

4. *Появление новых свойств на высших уровнях интеграции*. Обсуждение щекотливой проблемы «появления» в связи с поднимаемыми вопросами завело бы нас слишком далеко. Мне придется лишь догматически сформулировать общий принцип: «Если два существа объединяются на более высоком уровне интеграции, то не все свойства нового существа можно логически вывести или предсказать на основании свойств этих двух компонентов». Эта трудность присуща не одной только биологии, но она, несомненно, является одним из главных источников неопределенности в биологии. Не следует забывать, что неопределенность означает не отсутствие причины, а лишь непредсказуемость.

Все четыре причины неопределенности, как по отдельности, так и в сочетании друг с другом, снижают точность предсказания.

В связи с этим можно задать вопрос, обусловлена ли предсказуемость в классической механике и непредсказуемость в биологии количественными или качественными различиями. Существует немало доводов за то, что эти различия в основном носят количественный характер. Классическая механика находится, так сказать, на одном конце непрерывного спектра, а биология — на другом. Возьмем классический пример газовых законов. Можно считать, что эти законы носят статистический характер, однако число молекул газа, подчиняющихся этим законам, столь велико, что действия отдельных молекул приводят к некоему предсказуемому — можно сказать «абсолютному» — результату. Если же взять 5 или 20 молекул, то они проявят определенную индивидуальность. Различия в размерах изучаемых «популяций» также, конечно, участвуют в создании различия между физическими науками и биологией.

Выводы. Вернемся к нашему исходному вопросу и попытаемся суммировать некоторые из наших выводов о природе связей между причиной и следствием в биологии.

1. Причинность в биологии сильно отличается от причинности в классической механике.

2. Объяснение всех биологических явлений, кроме самых простых, обычно слагается из ряда причин. Так обстоит дело, в частности, с теми биологическими явлениями, которые можно понять лишь в свете их эволюционной истории. Каждый такой ряд причин напоминает скобки, в которых содержится много неизвестного и много такого, что, вероятно, никогда не удастся исследовать до конца.

3. Вследствие многочисленности альтернативных путей, возможных для большинства биологических процессов (за исключением чисто физико-химических процессов), и случайности многих биологических процессов, в

частности на молекулярном уровне (а также в силу других причин), причинность в биологических системах не дает возможности предсказывать или в лучшем случае позволяет делать предсказания лишь статистического характера.

4. Существование сложных информационных кодов в ДНК зародышевой плазмы допускает телеономическую целенаправленность. Вместе с тем эволюционные исследования не позволили обнаружить никаких доказательств «целенаправленности» эволюционных линий, как это постулируют те сторонники телеологии, которые видят «план и замысел» в природе. Гармония живой вселенной, поскольку она существует, представляет собой, так сказать, «апостериорный» продукт естественного отбора.

Наконец, причинность в биологии не противоречит на самом деле причинности в классической механике. Как показала современная физика, причинность в классической механике представляет собой лишь очень простой, частный случай причинности. Например, предсказуемость нельзя рассматривать как необходимый компонент причинности. Сложность биологической причинности не оправдывает использования ненаучных теории, например виталистических или телеологических; она лишь должна вдохновлять всех тех, кто пытается создать более широкую основу для понятия причинности.

ЛИТЕРАТУРА

1. *Nagel K.,* Lecture presented at the Massachusetts Institute of Technology in the Hayden Lectures Series, 1960—1961.

2. *Delbruck M.*, Trans. Conn. Acad. Arts Sci., 33, 173 (1949).

3. *Loeb J.*, The organism as a whole, Putman, New York, 1916.

4. *Scriven M.*, неопубликованные данные.

5. *Bernard C.*, Lemons sur les phcnomcnes do la vie. vol. 1, 1885.

6. *Pittendrigh C. S.*, in: Behaviour and Evolution, A. Roc and G. G. Simpson, eds., Yale Univ. Press. New Haven, Conn., p. 394, 1958.

7. *Huxley J.,* Zool. Jahrb. Abt. Anal, und Ontog. Tiere, 88, 9 (I960).

8. *Maeleod R. B*., Science, 125, 477 (1957).

9. *Simpson G. G.*, Ibid., 131, 966 (1960).

10. *Simpson G. G*., Sci. Monthly, 71, 262 (1950).

11. *Koch L. F*., Ibid., 85, 245 (1957).

12. *Nagel E.*, The structure of science. New York, 1961.

13. *Bunge M.,* Causally, Harvard Univ. Press. Cambridge, Mass., p. 307, 1959.

14. *Scriven M.,* Science, 130, 477 (1959).

15. *Park T.*, Physiol. Zool., 27, 177 (1954).

КОММЕНТАРИИ К. Х. УОДДИНГТОНА

В связи с очень интересной статьей Майра хочется сделать некоторые замечания. Майр проводит грань между неприемлемой и приемлемой формами телеологического объяснения; первая из них — собственно телеологическая или виталистическая, а вторая — «гелеономическая» или, если использовать введенный мною термин, «квазителеологическая». Но проведя это различие между типами гипотез, Майр затем приходит к утверждению о существовании соответствующих различий между типами явлений. «Развитие или поведение особи, — пишет он, — целенаправленно (в приемлемом смысле. — *К. У.*), а естественный отбор, разумеется, нет». И под этим «нет» он, по-видимому, имеет в виду «не может быть», поскольку в другом месте он пишет: «Однако исторические процессы не *могут* действовать целенаправленно». Уже в течение нескольких лет я настаиваю на том, что в теории эволюции, как и в теории развития [2, 3], следует использовать объяснения квазителеологического типа.

Если запустить какой-либо процесс (например, смешать два реагирующих между собой химических вещества), то он в конечном счете придет к какому-то концу. О «телеологии» начинают говорить, когда при этом в конечном результате возникает что-либо интересное, например нечто сложное и в то же время определенное. При этом возможны три основных типа объяснения: 1) конечный результат сам действует как причина, направляя процесс таким образом, что он завершается этим предетерминированным конечным состоянием, — это аристотелева телеология, которую мы отвергаем, потому что входящее в нее представление о причинности лежит вне принятого нами круга идей; 2) некий нематериальный агент направляет процесс к предетерминированному концу — это «витализм», который мы также отвергаем; 3) конечный результат процесса определяется его исходными свойствами — это «механицизм», а наш приобретенный за последние годы опыт работы с такими механическими системами, как счетно-решающие устройства, наводит на мысль, что «механистическое» объяснение более перспективно, чем это казалось сначала. Мы можем запустить какой-то процесс таким образом, чтобы он приводил к некоему установленному заранее конечному состоянию, вводя в исходное состояние ряд условий, которые действуют как «программа», и возвращая этот процесс на правильный путь посредством отрицательных обратных связей в случае отклонения. И наоборот, если какой-либо процесс характеризуется программой и обратными связями, можно сделать вывод, что он придет к какому-то конечному состоянию, которое в принципе можно установить по характеру программы и обратных связей (степень точности, с которой устанавливается конечное состояние, зависит, разумеется, от особенностей программы и обратных связей).

Майр принимает теорию [4], согласно которой онтогенетическое развитие зависит от какого-то квазителеологического механизма такого рода, причем и программа и обратные связи введены в генотип, созданный естественным отбором. Но в природе таких квазителеологических механизмов нет ничего, что исключало бы предположение о квазителеологи– ческой природе эволюционного процесса. Очевидно, что он характеризуется программой, содержащейся в теореме естественного отбора. Этого уже достаточно, чтобы в какой-то степени определить природу конечного состояния, к которому будет идти эволюция; она должна привести к повышению эффективности биосистемы в целом, проявляющейся в нахождении путей к воспроизведению. Несомненно, существует множество типов обратных связей, которые позволили бы более точно определить конечный результат. На два типа таких обратных связей я уже обращал внимание: 1) связь, благодаря которой поведение организма влияет на характер давлений отбора, действующих на сам организм (грубо говоря, животное выбирает среду, прежде чем среда оказывает на него селективное действие); 2) связь, обусловленная тем, что отбор предшествующих поколений на стабильность или лабильность развития влияет на характер фенотипического эффекта, который может возникнуть в результате новой мутации. Существуют и многие другие типы обратных связей. Например, возрастание фенотипического разнообразия популяции, свяаанное с необходимостью приспособления к различным местам обитания, в конечном счете уравновесится возникновением преград, препятствующих скрещиванию между отдельными разновидностями.

Мне представляется в настоящее время неправомерным говорить, как это делает Майр, что естественный отбор не является целенаправленным процессом. Разумеется, сам по себе он не более целенаправлен, чем процесс образования межатомных химических связей. Но подобно тому, как этот последний процесс лежит в основе белковых синтезов, которые в свою очередь объединяются в квазителеологический механизм зародышевого развития, естественный отбор представляет собой основной механизм другого квазителеологического процесса — эволюции. Сейчас нам следует использовать наши новые представления о природе квазителеологических механизмов для того, чтобы углубить понимание эволюционных процессов.

ЛИТЕРАТУРА

1. *Mayr E.*, Science, 134, 1501 (1961).
2. *Waddington C. H.*, The Strategy of Genes, Allen and Unwin, London, 1957.
3. *Waddington C. H.*, The Nature of Life, Allen and Unwin, London, 1961.
4. *Waddington C. H.*, Organisers and genes, Cambridge Univ. Press, Cambridge, 1940. (К. Уоддингтон, Организаторы и гены, ИЛ, М., 1947.)

ПОПЫТКА СОЗДАНИЯ СХЕМЫ ПЕРВИЧНОГО ОРГАНИЗМА

А. Кэрнс-Смит
(Университет в Глазго)

Пытаясь создать схемы первого организма, Холдейн [1] высказал предположение, что простейшая возможная сгруктура могла бы состоять из одного РНК-го гена, определяющего синтез одного фермента. Он считал, что появление даже такой системы кажется слишком мало вероятным, если не допустить участия какого-то до сих пор не обнаруженного принципа, который придает биологическим системам некую присущую им большую вероятность, которой не было бы в его отсутствие. Обсуждая «самовоспроизведение», фон Нейман [2] также приходит к обескураживающему выводу о высокой нижней границе необходимой сложности для любой «саморепродуцируюгцейся» системы [3].

Я допускаю, что интересующие нас первичные организмы удовлетворяли как общему определению жизни по Мёллеру, так и «самому важному условию» Уоддингтона, т. е. обладали способностью участвовать в длительных процессах эволюции. На мой взгляд, Холдейн хотел подчеркнуть, что принципы организации самых первых организмов отличались от принципов построения современных организмов: как указывает Патти, нет никакой логической необходимости настаивать на том, что универсальные для современных организмов молекулы непременно присутствовали в самых примитивных формах или несли в них функции, аналогичные нынешним. Хорошо известные синтезы «биохимических веществ» в условиях, постулируемых для древней Земли, не дают никаких веских доказательств сходства между современными и примитивными формами жизни: организмы, развивающиеся в среде, которая содержит аминокислоты, пурины и т. п., по всей вероятности, включали эти молекулы в свои структуры на какой-то стадии, но не обязательно в момент своего возникновения.

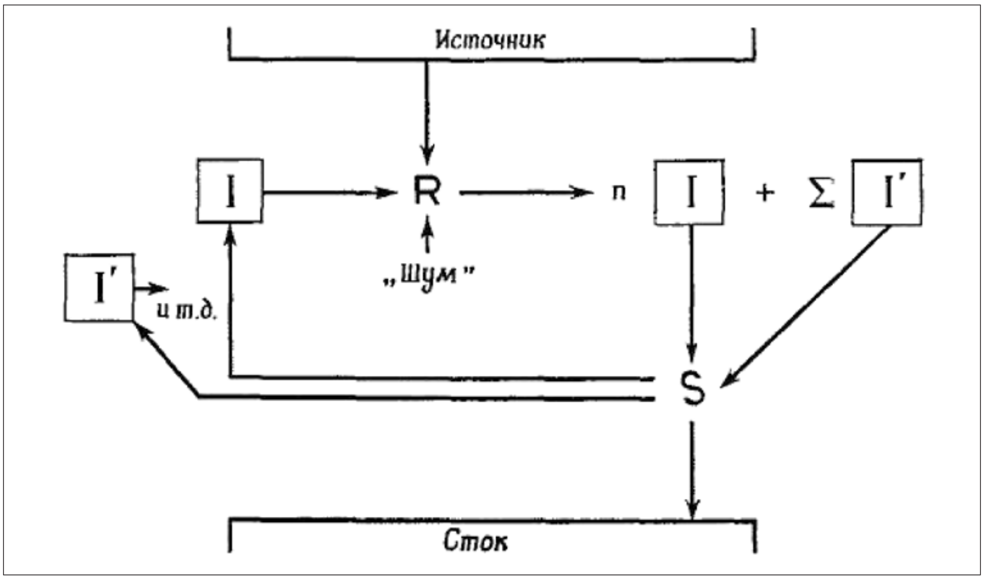
Затруднения, выявленные фон Нейманом, следует рассматривать как указание на то, что самовоспроизведение примитивных организмов в значительной мере или даже полностью было обусловлено их средой, т. е. что процесс самовоспроизведения как таковой не требовал предсуществования в этих организмах каких-либо детально разработанных внутренних инструкций или механизмов. Можно, было бы представить себе репликацию элементарной ячейки кристалла или, еще лучше, репликацию картины дислокаций в процессе роста кристалла. Сама обыденность процессов роста кристалла, которые могут идти неопределенно долго в подходящей среде при соответствующих условиях, служит указанием на то, что для них не требуются тщательно разработанные специальные инструкции. Возможно, как говорит Уоддингтон, что простая дислокация не представляет достаточного интереса, чтобы ее можно было назвать организмом; однако, как мне кажется, нам следует принять, что жизнь – «интересная» жизнь – возникла из систем, которые были «неинтересными» в том смысле, что если они и обладали функциональной специфичностью, то она, вероятно, была весьма ограниченной. Вместо того чтобы рассматривать теоретические модели процессов самовоспроизведения, близко отражающие аналогичные системы в современных организмах, мы должны внимательнее присмотреться к простым процессам репликации, в изобилии встречающимся в мире физикохимических явлений, рассматривая эти процессы не просто как модели, а как возможных предков. Интересующая нас проблема, вероятно, целиком относится к области эволюции, а не происхождения: быть может, она состоит в том, чтобы найти пути эволюции от какого-нибудь хорошо известного кристалла к одному из тех организмов, которые сейчас населяют Землю.

Ген в потоке? Поддержание некоторого непрерывного потока в окружающей среде, разумеется, существенно для жизни: современные организмы в той или иной степени зависят от притока фотонов с солнца. Шеррингтон [4] сравнивал организмы с водоворотами в потоке. Мы можем продолжить эту аналогию и сказать, что сами водовороты представляют собой фенотипы, продуцируемые генотипами, которые состоят из камней и берегов потока, обладающих большим постоянством и определяющих форму этих водоворотов. Для полноты аналогии надо бы иметь камень, способный к репликации, и такую ситуацию, чтобы специфическая форма водоворота, созданного камнем специфической формы, могла бы определять потенциал выживания этого камня в потоке.

Если рассматривать эту аналогию на молекулярном уровне, то она становится еще ближе к действительности и служит возможной моделью примитивного организма. Поток мы понимаем в буквальном смысле, как некий стационарный гидродинамический поток к поверхности или близ поверхности древней Земли. «Камень» – это кристаллит, обладающий рядом дефектов. Поток (иногда) бывает перенасыщен компонентами кристаллита, в результате чего путем роста и деления образуются новые кристаллы с такими же дефектами. «Водовороты» возникают в результате зависящего от структуры взаимодействия между кристаллитом и компонентами потока (т. е. они состоят из изменяющейся популяции молекул, адсорбированных с разной степенью специфичности, которые либо стремятся удержать кристаллит в той части потока, где имеются благоприятные для роста кристаллов условия, либо защищают кристаллит от растворения, если поток время от времени перестает быть насыщенным, либо ускоряют процессы роста и деления кристаллов).

Я уже высказывал предположение [5], что частицы глины могли служить генами первичных организмов; при этом благоприятствующие выживанию инструкции могли быть записаны в форме определенного пространственного набора катионных связей, а фенотипы представляли бы собой адсорбированные, возникшие абиогенным путем органические молекулы. Реальность такого предположения зависит и от того, обладал ли данный набор катионных связей способностью реплицироваться в процессе образования глины из раствора. Мы будем развивать здесь более общую идею о том, что нашим исходным предком был в конечном счете некий «ген в потоке», не указывая, был ли этот ген частицей глины и был ли он кристаллитом вообще.

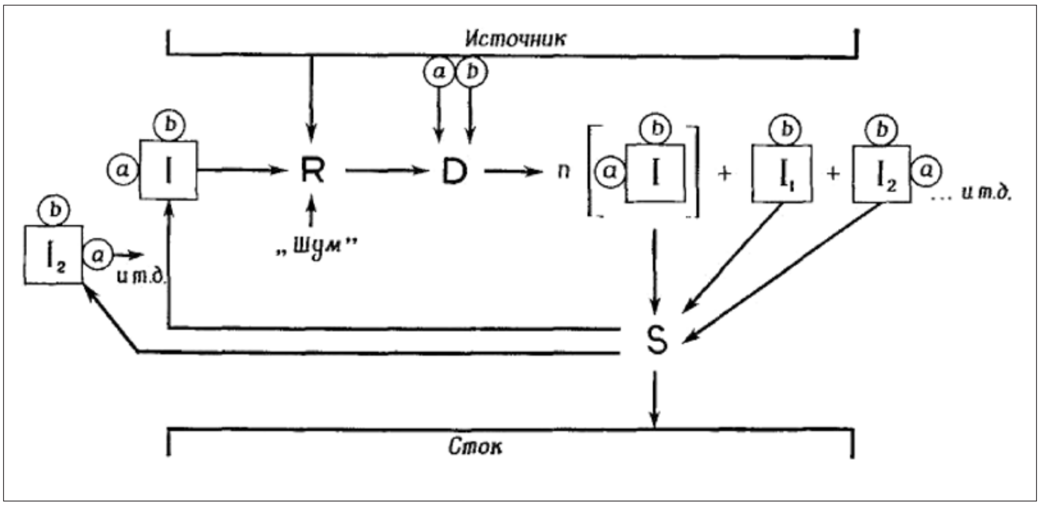
Общие требования, которым должна удовлетворять развивающаяся физико-химическая система. Нам требуется система, которая может неопределенно долго проходить дарвиновский цикл, представленный на фиг. 1. Она будет состоять из популяции маленьких ящичков нескольких разных типов, например типа I, и существующих в окружающем их пб- токе, который воздействует на них двояким образом: 1) реплицирует разные типы (обычно, хотя далеко не всегда точно), синтезируя новые ящички, и 2) отбирает ящички в зависимости от их типа. Некоторые типы ящичков отбираются потому, что они ускоряют самый процесс репликации. Если происходит только такой отбор, например предпочтительное образование в кристалле дислокаций, способствующих росту, то такую систему можно назвать организмом типа I. В организме типа II существует (альтернативно или дополнительно) стадия отбора, не зависящая от процесса репликации, т. е. данный тип специфически повышает шансы на выживание ящичков этого типа. Мы можем представить себе колеблющуюся среду, которая в разное время стимулирует: 1) образование ящичков (условия репликации – *R*) и 2) разрушение ящичков (условия отбора —*S*). Эволюция такой системы привела бы к созданию таких типов ящичков, которые: 1) благоприятствуют образованию и 2) устойчивы к разрушению или, иначе говоря, обеспечивают наиболее эффективное прохождение цикла. Эти типы и соответствуют генетическим инструкциям, связанным со специфической средой.



Фиг. 1. Дарвиновский цикл для организма типа II (см. текст).
Возникающие иногда ошибки («шум») в репликации приводят к изменениям структуры и некоторые из них (например, *I*) могут привести к возникновению конкурентного цикла.

Окружающая среда предстает, таким образом, в виде активной части сложной системы, которая вырабатывает инструкции, записанные в форме ряда устойчивых типов на молекулярном уровне. В результате мы отходим от обычных представлений и рассматриваем среду («поток») как преимущественно динамический, а организм («ген») – как статический компонент системы. Это позволит нам представить себе первичные организмы весьма простыми, однако при этом окажется, что первичная среда должна быть очень сложной! Последнее положение – скорее догадка, однако следовало бы полагать, что простая среда может породить лишь простые, способствующие выживанию инструкции. Но если предположить, что заключенные в ящичке инструкции обеспечивают себе самозащиту путем адсорбции из среды молекул, причем форма и характер адсорбированных молекул зависят от содержащихся в ящичке инструкций (организм типа III, фиг. 2), то при этом с точки зрения ящичка среда изменилась, создав возможности для эволюции новых инструкций, относящихся уже к этой изменившейся среде. Эти новые инструкции вновь изменят непосредственную среду ящичков, создавая возможности для последующих изменений, и т. д.

При этом первичная окружающая среда, которая вырабатывает первоначальные инструкции, не должна быть очень сложной при условии, что эти инструкции, по крайней мере частично, в свою очередь воздействуют на среду и изменяют ее, особенно если они способны вызвать стойкое изменение в своем непосредственном окружении, т. е. образование фенотипа. (По-видимому, фенотип появится независимо от вашей воли, даже при самых строгих исходных генетических взглядах, если поставить непременным условием, что любая система, называемая организмом, должна быть способна к неопределенно долгой эволюции, т. е. если к определению жизни по Мёллеру добавить «самое важное» условие Уоддингтона.)



Фиг. 2. Дарвиновский цикл для организма типа III.
Этот организм в процессе своего развития проходит стадию *D*,
на которой он в соответствии с заключенными в нем инструкциями поглощает
из среды определенные защитные сочетания молекул, например *a* и *b*.

Возможен ли «генетический метаморфоз»? Ящички, представленные на фиг. 1 и 2, изготовлены из некоего «генетического материала». На основании некоторых общих соображений можно думать, что в первичных организмах этим материалом была не ДНК и не какое-нибудь сходное с ней вещество: свойства, необходимые первичному генетическому материалу, который должен был действовать «на ровном месте», практически без всяких вспомогательных механизмов, по-видимому, сильно отличаются от свойств, обеспечивающих наибольшую эффективность генетическому материалу высокоразвитых организмов, в которых в репликации генов и образовании фенотипа участвует сложная система предшествующих механизмов, как, например, в цитоплазме современных клеток. Требования, предъявляемые к «наилучшему из возможных» генетическому материалу на ранних этапах эволюции, были, вероятно, совершенно иными, однако действительная замена этого материала, по-видимому, зависела от наличия необходимого для этого механизма. В процессе обсуждения на этом симпозиуме было высказано предположение, что такой процесс логически невозможен. Если это так, то с момента возникновения жизни роль генетического материала должна была играть ДНК как единственный «компромиссный» материал. Я хотел бы показать, что «генетический метаморфоз» не только возможен, но что в самом строении первичных организмов мог быть заложен готовый механизм, обеспечивающий такой метаморфоз.

Сравнительно небольшие изменения генетического материала, которые можно назвать *генетическими модификациями*, не вызывают никаких серьезных изменений. Например, Холдейн [1] предполагал, что РНК предшествовала ДНК в качестве генетического материала: это нетрудно представить себе, поскольку РНК и ДНК «совместимы». Однако подобное изменение равносильно всего лишь изменению цвета туши, которой вычерчена схема. Нам хотелось бы знать, могла ли эволюция проделать путь от вычерченных схем до магнитной ленты! Вряд ли можно предполагать, что механизм, обеспечивший коренное изменение способа кодирования, сводится к ряду генетических модификаций через образование пар в совместимых системах. Однако теоретически должен существовать некий альтернативный механизм, для которого химическое сходство между различными видами генетического материала не обязательно, поскольку он не связан с транскрипцией инструкций с одного материала на другой.

«Гетерогенетические» организмы. Уоддингтон и Сэджер рассмотрели возможность существования у современных организмов более чем одного процесса передачи биологической информации. Предположение о развитии первичных организмов из сильно *гетерогенетических форм*, т. е. форм, содержащих более одного вида генетического материала, представляет определенный интерес.

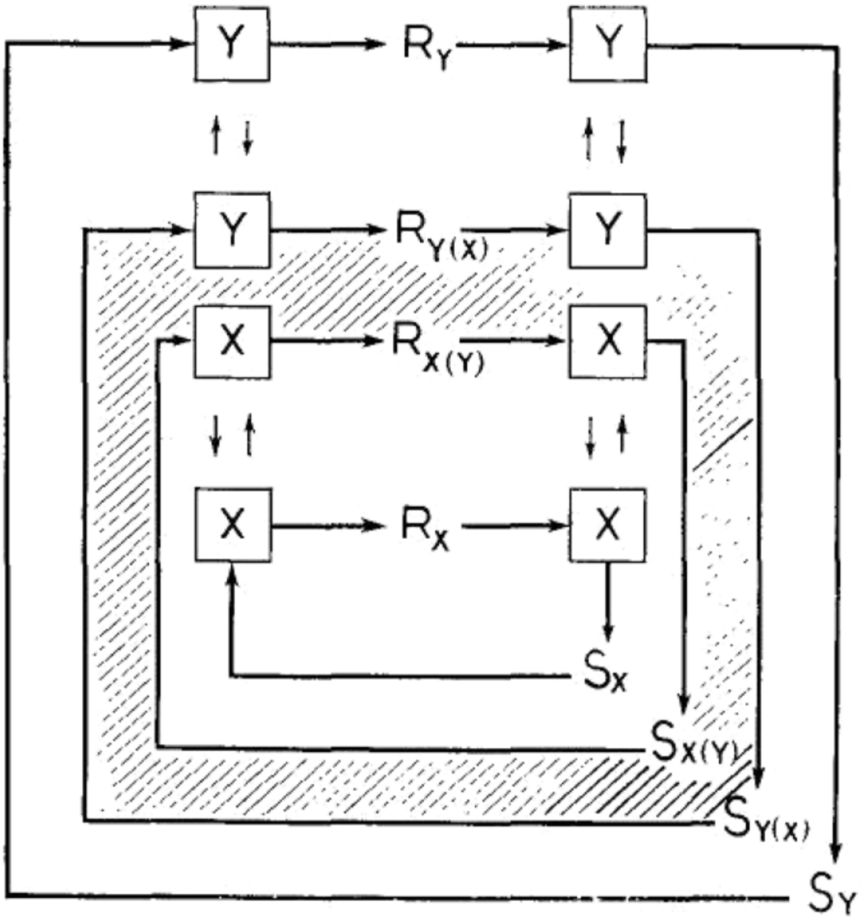
Прежде всего можно задать вопрос, почему современные организмы, по крайней мере большая их часть, *гомогенетичны*? Гомогенетические системы могут быть в принципе весьма простыми, однако современные организмы отличаются довольно сложной «простотой», определяемой большим разнообразием одной-единственной системы; это разнообразие в свою очередь зависит от разделения функций между двумя совершенно различными видами молекул: ДНК, обеспечивающей хранение и передачу генетических инструкций, и белка, реализующего эти инструкции в форме различных структур и функций. Современная «простота» зависит от наличия сложных механизмов трансляции. (Формально было бы проще хранить всю информацию, которую мы используем в повседневной жизни, в одной-единственной форме, скажем на магнитной ленте, но пока у каждого человека не будет соответствующей аппаратуры, удобнее пользоваться книгами, пластинками, фильмами и т. д.)

Пири [6] полагал, что современная биохимическая однородность – использование всеми организмами по существу одной и той же системы передачи информации – возникла в результате длительного процесса эволюции, основанного на отборе из многочисленных химически различных альтернативных систем. Нас же интересуют не столько различия *между* первичными организмами, сколько их *внутреннее* разнообразие.

Механизм генетического метаморфоза. Вполне возможно, что первые организмы были гомогенетическими и обладали весьма простыми, непосредственными и ограниченными механизмами выживания (см. [5]). Можно представить себе два процесса, которые могли бы повысить разнообразие первичной генетической системы: 1) *генетическая экспансия*, связанная с эволюцией все более косвенных способов действия; так, например, гены (см. фиг. 2) не только адсорбируют молекулы из своего непосредственного окружения, чтобы защитить их от разрушения, но и образуют на своей поверхности центры катализа, делая возможным воз- иикновсиие новых и более эффективных молекул. Дальнейшая экспансия могла бы произойти, если бы некоторые из вновь возникших молекул сами представляли собой регулирующие устройства, например катализаторы. Генетическая экспансия могла повысить изменчивость первичного генетического материала, но нет никаких особых причин думать, что такой .материал, будучи хорошей стартовой системой, мог бы успешно развиваться и дальше. Вторым процессом, способствующим увеличению генетического разнообразия, мог быть переход к гетерогенетическому состоянию. По всей вероятности, до появления какой-либо единственной весьма лабильной системы у первичных организмов, хотя, возможно, и не у самых первых, было несколько различных видов генетического материала, обладавших более или менее непосредственным действием, каждый из которых был способен к образованию различных видов биологически полезных структур. Далее, чем сильнее *различались* бы эти генетические материалы в химическом отношении, тем разнообразнее были бы выполняемые ими функции и тем меньше возможность их смешения. Можно предположить, например, что одно вещество могло реплицироваться и одновременно функционировать в качестве мембраны или гелеобразующего волокна, другое—реплицироваться и действовать в качестве катализатора одной определенной реакции, третье — катализировать другую реакцию и т. п. Каждое из этих веществ можно представить себе как организм (или суборганизм), симбиотически сосуществующий с другими, а весь организм – как студневидную массу неопределенных размеров, содержащую большое число каждого из этих суборганизмов. Размножение целого организма могло происходить сначала путем более или менее случайного отделения от этой массы участков, содержавших достаточный набор различных видов генетического материала. Позже могло появиться более упорядоченное размножение с синхронной репликацией генетического материала и его пространственной организацией в клетки, но оно не было необходимо в самом начале. (Группа организмов может существовать совместно к взаимной выгоде, но это не требует синхронизации их процессов репликации или сколько-нибудь сложной пространственной организации: поскольку члены группы зависят друг от друга, в пределах данного участка будет наблюдаться тенденция к установлению определенного равновесия между их численностями.)

Можно было бы ожидать, что на ранних стадиях эволюции преобладали факторы, способствовавшие усилению гетерогенетичности, причем весьма велика вероятность того, что какой-либо иной генетический материал с иным кругом прямых функций мог бы оказаться полезным для организма, вырабатывающего более стабильную генетическую систему. Позже взяли верх противоположные факторы, способствовавшие возврату к гомогенетичности. Главными здесь были процессы генетической экспансии в различных системах, приводящие к перекрыванию функций, выполняемых этими системами, в результате чего между ними начиналась конкуренция. В то время как отбор первичного генетического материала был направлен на усиление непосредственности его действия (без предсуществовавших механизмов), в последующей конкуренции предпочтение отдавалось косвенному эффекту, поскольку он повышал изменчивость системы.

Все вышесказанное можно вкратце изложить следующим образом, Наш первичный предок был гомогенетическим организмом; его генетический материал был простым, устойчивым и обладал прямым ограниченным действием. От него произошли в конечном счете современные организмы, также (в основном) гомогенетические, однако их генетический материал отличается сложностью, хрупкостью и весьма косвенным характером своих разносторонних эффектов. В ходе длительной эволюции произошел «метаморфоз»: первичные организмы стали гетерогенетическими. В конкуренции между различными видами генетического материала внутри одного организма более стойкой оказалась генетическая система, близкая к современной, а не исходная генетическая система.



Фиг. 3. Схема, показывающая, какими путями некий «самостоятельно запускающийся» цикл (в высшей степени упрощенный),оперирующий с генетическим материалом *X*, мог содействовать установлению критического цикла другого материала *Y* (см. текст). $R_{y(x)}$ и $S_{y(x)}$ (дуозиачают соответственно репликацию и отбор генетического материала *Y* с помощью *X* и (или) связанного с иим фенотипа. $R_{y(x)}$ и $S_{y(x)}$ имеют аналогичный смысл. Маленькие вертикальные стрелки указывают на возможность частичных ассоциаций. Штриховкой показан спаренный цикл сдигснстнческого организма).

Таким образом, исходный генетический материал оказался среди «побежденных», а «победитель» был лишь немного похож или вообще не имел никакого химического сходства с «побежденным».

Первичные организмы в роли «стартеров». На фиг. 3 схематически изображен исходный цикл с участием генетического материала *X*, т. е. R_xS_x. Этот цикл представляет собой упрощенный вариант рассмотренных ранее циклов (см. фиг. 1 и 2). Мы можем рассмотреть общий вопрос о том, как эволюция инструкций, записанных в материале *X*, может обусловить эволюцию инструкций, содержащихся в каком-то другом генетическом материале, например У, или способствовать такой эволюции без непосредственной транскрипции с одного из этих материалов на другой. Генетический метаморфоз в том виде, в каком мы его только что рассматривали, представляет собой крайний из трех возможных родственных процессов, посредством которых R_xS_x может действовать как стартер для R_yS_y. Не следует думать, что эти процессы взаимно исключают друг друга.

1. *«Сапрофиты» – образ жизни*. По мере эволюции первичных форм должен был изменяться состав имеющихся на Земле веществ. Это обстоятельство уже само по себе повышало вероятность появления на каком-то этапе альтернативных видов генетического материала. Более того, среди вновь появляющихся веществ, по-видимому, были оптически активные молекулы, которые в большей степени подходили для образования генетических полимеров, чем обычные рацемические смеси, обычно содержавшиеся в исходной среде. Таким образом, R_xS_x мог привести к образованию совсем другого цикла R_yS_y, впервые предоставив подходящий для него субстрат.

2. *«Паразитический» образ жизни*. Если новый вид генетического материала возник в рамках фенотипа какой-то первичной формы, то это могло создать особенно подходящую среду для установления некоторого критического цикла (в ходе которого образуется в среднем достаточно жизнеспособных потомков, чтобы обеспечить его сохранение на неопределенно долгое время), причем не только потому, что эта новая жизненная форма могла использовать новый запас молекул (и, возможно, более высокоорганизованных структур типа трубок и бороздчатых поверхностей), но также и потому, что некоторые стороны жизнедеятельности первичной формы могли содействовать установлению критического цикла. Например, в первичной форме могли возникнуть черты, характерные для «гудвиновского осциллятора», которые уменьшили бы ее зависимость от колебаний условий внешней среды при регуляции своей генетической репликации и таким образом создали бы регулируемую колеблющуюся среду для паразита. Подобная паразитическая ассоциация могла быть постоянной ($R_{y(x)}S_{y(x)}$) или захватывать лишь часть жизненного цикла паразита ($R_yS_{y(x)}$ или $R_{y(x)}S_y$). Следовательно, мы должны учитывать возможность того, что наш исходный предок был «спонтанно размножавшимся облигатным паразитом», который постепенно перешел к свободному образу жизни.

3. *«Симбиотические» отношения*. В нашем представлении истинный генетический метаморфоз проходит через стадию, на которой вторичный (облигатный) суборганнзм слегка повышает возможности первичного организма. В этом случае можно ожидать взаимной эволюции в результате сочетания инструкций, заключенных в различных видах генетического материала, – тем и другим было бы выгодно «держаться вместе» в буквальном смысле. Затем в результате отбора мутаций в каждом из генетических материалов должна была возникнуть сложная система – «дигенетический» организм. Циклы $R_{y(x)}S_{y(x)}$ и $R_{x(y)}S_{x(y)}$ (фиг. 3) можно было бы рассматривать при этом как единый цикл. Если допустить для простоты, что во всем этом не участвует никакой другой генетический материал, можно представить

себе, что при наличии у Y большего эволюционного потенциала (способности к более обширной генетической экспансии), чем у X, произойдет полный генетический метаморфоз, так что генетический материал X будет в конечном счете устранен. В результате должен установиться цикл R_yS_y, характеризующийся сравнительно сложными инструкциями и предшествующими фенотипическими механизмами, отсутствие которых не позволяло Y стать первичным генетическим материалом.

А как же нуклеиновая кислота? Появление нуклеиновой кислоты или какого-то предшественника, способного превратиться в нуклеиновую кислоту просто путем генетического изменения, несомненно, было критической стадией в эволюции. Некоторые исследователи обычно называют этот момент «происхождением жизни», хотя это (ср. [6]) довольно предвзятая точка зрения. Согласно развиваемым нами представлениям, «происхождение жизни» было весьма постепенным процессом. Он мог начаться с появления в рамках фенотипа какой-либо уже образовавшейся первичной формы вещества типа РНК, действовавшего в качестве подсобного генетического материала при выполнении какой-либо весьма незначительной функции, которая (в чем бы она ни состояла), по-видимому, определялась способностью к образованию специфичной, зависящей от последовательности составляющих ее единиц третичной структуры путем внутримолекулярного спаривания оснований. Затем могла произойти генетическая экспансия – между такими РНК-ми органеллами и маленькими молекулами из их окружения, в частности аминокислотами, могли возникать специфичные ассоциации, вносившие более важный и сложный вклад в гетерогенетический организм в целом. Можно думать, что дальнейшая экспансия шла по пути специфичного спаривания абсорбированных аминокислот, приводившего к образованию полипептидов с независимыми функциями. В конечном счете некоторые из этих полипептидов достигали таких размеров, что сами начинали образовывать последовательные третичные структуры, – за этим последовало «открытие» ферментов. И только тогда могло стать ясно (первому молекулярному биологу!), что РНК образовала «выигрышную» систему, способную принять на себя всю генетическую регуляцию всего организма.

Каким бы ни был эволюционный путь становления современного механизма наследственности (изложенные выше представления о «генетической рибосоме» наверняка слишком просты), с момента первого возникновения нуклеиновой кислоты как *одного* из видов генетического материала и до окончательного ее утверждения как *единственного* генетического материала, по-видимому, должно было пройти очень много времени. Возможность создать искусственные условия, которые позволили бы провести этот процесс в обозримое время, представляется весьма проблематичной. С другой стороны, этому творческому, но ограниченному в своих возможностях Слепому Случаю, по-видимому, удалось, по крайней мере однажды, создать исходный организм, тогда как мы смогли синтезировать такое вещество, как молекула инсулина, чего Слепой Случай почти наверняка не смог бы сделать, даже если допустить существование самых идеальных условий на древней Земле (ср. [7, 8]). Следовательно, сделать исходный организм было бы совсем несложно – если бы мы знали, как его делать. Прежде всего мы должны попытаться представить себе какую-либо реальную физико-химическую систему, способную к неопределенно долгой дарвиновской эволюции. И мы не должны настаивать на том, чтобы наш гипотетический организм непременно содержал нуклеиновую кислоту или вообще какой-либо органический полимер.

ЛИТЕРАТУРА

1. *Haldane J. B. S.*, in: The Origins of Prebiological Systems, S. W. Fox, ed., Academic Press, New York, 1965.
2. *J. Neumann* von, Collected Works, vol. 5, Pergamon Press, 1965.
3. *Pattce H. H.*, Advance. Enzymol., **27**, 381 (1965).
4. *Sherrington C.*, Man on his Nature, ch. 3, Univ. Press, Cambridge, 1940.
5. *Cairns-Smith A. G.,* J. Theoret. Biol., **10**, 53 (1966).
6. *Pirie N. W.*, in: The Origin of Life on the Earth, F. Clark and R. L. M. Synge, eds., Pergamon Press, New York, p. 76, 1959.
7. *Dixon M., Webb E. C.*, Enzymes, Academic Press, New Yorft, p. 666, 1958. (М. Диксон, Э. Уэбб, Ферменты, ИЛ, М., 1961.)
8. *Pat tee H. H.*, Biophys., J., **1**, 683 (1961).

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 издателю 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 заверенное 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

- Рекомендуемый шрифт – Times New Roman.
- Строки в пределах абзаца не должны разделяться символом возврата каретки (Enter).
- Нельзя использовать автоматическое создание сносок, 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 редактров 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 Jornal, 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 институт электроники и математики (МИЭМ),
Европейский центр по качеству

Издатель
Европейский центр по качеству
ООО Сенсор Микрон
Журнал входит в перечень ВАК РФ
Статьи рецензируются

ПОДПИСКА НА ЖУРНАЛ НМФМ

Подписной индекс журнала в каталоге агентства «Урал-Пресс» 70017.
Электронный подписной каталог и контакты всех представительств «Урал-Пресс»
на сайте www.ural-press.ru.
Редакция предлагает подписчикам возможность безвозмездно получить подборку прошлых выпусков журнала. Пришлите на электронный адрес nanostructures@hse.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.