

УДК 519.17:547.1:547.25

МОЛЕКУЛЯРНЫЕ ГИПЕРГРАФЫ<sup>\*</sup>)

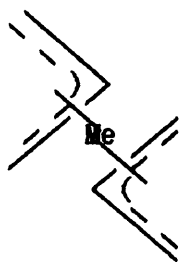
Е.В.Константинова, В.А.Скоробогатов

В в е д е н и е

Теоретико-графовые представления достаточно широко используются в компьютерной химии и теоретических химических исследованиях [1,2]. Молекулярные структуры представляют в виде графов, в которых вершины соответствуют атомам, а ребра - химическим связям. Такой граф, называемый молекулярным [1], является объектом изучения теории обыкновенных графов [3]. Однако обыкновенные графы дают неадекватное описание химических соединений неклассического строения. Как отмечалось в [4] существенным недостатком структурной теории является отсутствие удобного способа представления молекул с делокализованными многоцентровыми связями. К таким структурам относятся, в частности, металлоорганические соединения [5].

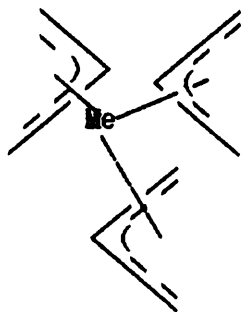
В металлоорганической химии рассматриваются соединения, содержащие по крайней мере одну связь металл-углерод (Me-C). Эта связь может быть обыкновенной ковалентной, например,  $Pb(C_2H_5)_4$ , но может быть и ионной, например,  $Na^+C_2H_5^-$ , или  $\pi$ -связью, как, например, в ферроцене, где железо образует две связи с цикlopентадиенилами, являющимися пяти-электронными лигандами. В прин-

<sup>\*</sup>) Работа выполнена при поддержке Российского Фонда Фундаментальных Исследований по теме № 93-03-18657.



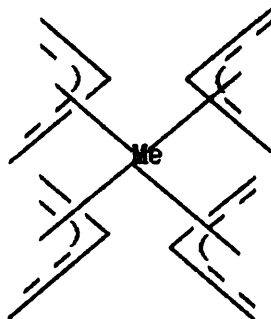
1)  $\text{MeL}_2$

$\text{Me}=\text{Ni}, \text{Pd}, \text{Pt}$



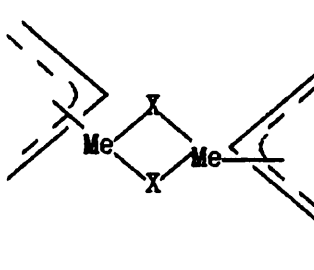
2)  $\text{MeL}_3$

$\text{Me}=\text{V}, \text{Cr}, \text{Fe}, \text{Co}, \text{Rh}$



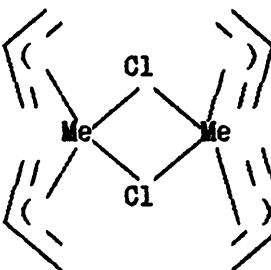
3)  $\text{MeL}_4$

$\text{Me}=\text{Zr}, \text{Nb}, \text{Ta}, \text{Mo}, \text{W}$



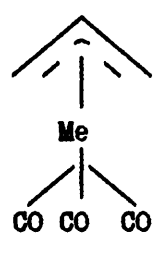
4)  $[\text{LMeX}]_2$

$\text{Me}=\text{Ni}, \text{Pd}$



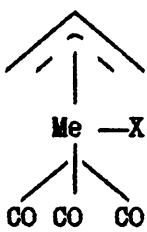
5)  $[\text{L}_2\text{MeX}]_2$

$\text{Me}=\text{Rh}$



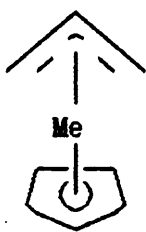
6)

$\text{Me}=\text{Co}, \text{Fe}^{\oplus}$



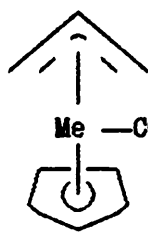
7)

$\text{Me}=\text{Fe}$



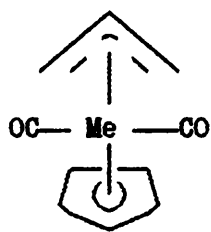
8)

$\text{Me}=\text{Ni}$



9)

$\text{Me}=\text{Fe}, \text{Co}^{\oplus}$



10)

$\text{Me}=\text{Mo}$

Рис.1. Молекулярные структуры соединений с трехэлектронными лигандами

цепе, лигандом может быть любая ненасыщенная или ароматическая органическая молекула или радикал.

Одна и та же молекула может выполнять роль различных лигандов. Типы связей  $\text{Me}-\text{C}$ , а также различные типы лигандов приводят к значительному разнообразию металлоорганических соединений и являются основой их классификации [5]. На рис.1 представлены примеры молекулярных структур соединений с трех-электронными лигандами.

Рассмотрим некоторые графовые модели, используемые для представления таких соединений. В [6,7] приводятся примеры описания структур "сэндвичевого" и олефинового типов в виде несвязных молекулярных графов (см.рис.2а). Очевидно, что такое представление структур является ненаглядным и не позволяет анализировать структуры в целом, так как отсутствуют связи между молекулярными графами, представляющими отдельные фрагменты структуры.

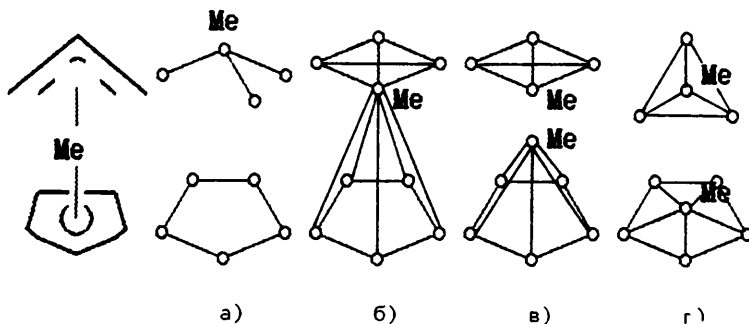


Рис.2

Более наглядными, но также не лишенными недостатков, являются связанные молекулярные графы, в которых все вершины, соответствующие атомам углерода, соединены с вершиной  $\text{Me}$ , соответст-

вующей атому металла (рис.2б). В этом случае степень вершины  $Me$ , равная числу соединенных с ней вершин, не обязательно равна валентности атома  $Me$ . Кроме того, как в первом представлении так и во втором, "стирается" различие между обыкновенными и многоцентровыми связями.

Иногда при изучении стереохимических свойств металлоорганических комплексов для их описания используют плоские графы [8]. В комплексе выделяют фрагменты, представимые в виде какой-либо пирамиды, например, тригональной и пентагональной, как это показано на рис.2в, а затем преобразуют пирамиду в плоский граф (рис.2г). И в этом случае получают описание структуры в виде несвязного плоского графа.

Не только металлоорганические соединения содержат делокализованные связи. Они имеются и в ароматических системах, представление которых, как отмечалось в [9], "в виде систем с фиксированными двойными связями становится неоправданным, когда возможно написание двух и более эквивалентных кекулевских структур". Неоправдано, видимо, и представление кекулевских структур обыкновенными молекулярными графами, поскольку последние не различают обыкновенные связи и  $\pi$ -связи. Для обозначения шести  $\pi$ -электронов введен специальный символ - круг внутри шестиугольника, позволяющий формально обозначать положение ароматического секстета среди колец полициклической системы, как это показано, например, на рис.3 для тетрацена.

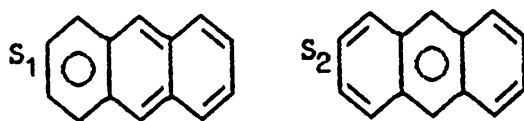


Рис.3

Для описания структур, содержащих делокализованные многоцентровые связи, предлагается использовать их представление в

виде гиперграфов, позволяющих устранить приведенные выше недостатки.

### 1. Гиперграфы молекулярных структур

Приведем некоторые понятия из теории гиперграфов [10,11].

Гиперграф  $H = (V, E)$  состоит из непустого множества вершин  $V = \{v_i \mid i = 1, \dots, p\}$  и семейства  $E = \{E_j \mid j = 1, \dots, q\}$  различных подмножеств множества вершин. Множества  $E_j$  называют ребрами гиперграфа или гиперребрами. На рис.4 приведен пример гиперграфа  $H$ .

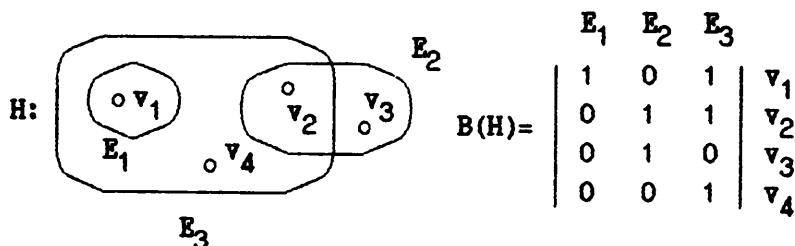


Рис.4

Ребра и вершины гиперграфа называют его элементами. Если  $v_i \in E_j$ , то вершина  $v_i$  и ребро  $E_j$  инцидентны. Множество всех ребер, инцидентных вершине  $v_i$ , называют степенью вершины  $v_i$  и обозначают  $\deg v_i$ . Степенью ребра  $E_j$  гиперграфа называют множество всех инцидентных ему вершин и обозначают  $\deg E_j$ . У гиперграфа  $H$  на рис.4 гиперребра  $E_1, E_2, E_3$  имеют степени один, два и три соответственно; вершина  $v_2$  инцидентна ребрам  $E_2$  и  $E_3$ , т.е.  $\deg v_2 = 2$ , а вершина  $v_4$  инцидентна только ребру  $E_3$ .

Обыкновенный граф является частным случаем гиперграфа, в котором для всех ребер  $E_j$ ,  $j = 1, \dots, q$ ,  $\deg E_j = 2$ .

Конечный гиперграф (множество  $V$  и семейство  $E$  конечны) однозначно задается матрицей инцидентий  $B(H) = \|b_{ij}\|$ , где

$$b_{ij} = \begin{cases} 1, & \text{если } v_i \in E_j, \\ 0, & \text{если } v_i \notin E_j. \end{cases}$$

На рис.4 представлена матрица инцидентий гиперграфа  $H$ .

Два гиперграфа  $H^1$  и  $H^2$  изоморфны, если между их множествами вершин и ребер существует взаимно-однозначное соответствие, сохраняющее инцидентность.

1.1. Молекулярный гиперграф. Гиперграф  $H = (V, E)$  представляющий молекулярную структуру  $F$ , в котором вершины  $v \in V(H)$  соответствуют отдельным атомам, гиперребра  $E \in E(H)$ , степень которых больше двух, соответствуют делокализованным многоцентровым связям и гиперребра  $e \in E(H)$ ,  $\deg e = 2$ , соответствуют обыкновенным ковалентным связям, будем называть молекулярным гиперграфом.

Молекулярный гиперграф  $H = (V, E)$  называется помеченным, если на множестве вершин  $V$  и на множестве ребер  $E$  введены функции помечивания, ставящие в соответствие каждой вершине и каждому ребру некоторую метку. Гиперребра степени два будем изображать в виде обыкновенных ребер, чтобы подчеркнуть различие между обыкновенными и многоцентровыми связями.

Молекулярные гиперграфы, как и обыкновенные молекулярные графы, могут представлять молекулярные структуры с разной степенью детализации. Покажем это на примере четырех молекулярных гиперграфов (рис.5) рассмотренного ранее аллильного комплекса (рис.2). Известно, что аллильный лиганд содержит открытую цепочку трех  $sp^2$ -гибридизованных атомов углерода, каждый из которых имеет  $\pi$ -электрон, свободный для образования связи металл-лиганд [5]. На рис.5а изображен молекулярный гиперграф  $H_1$ , в котором делокализованная связь в аллиле представлена ребром  $E_1$ ,  $\deg E_1 = 3$ , а связь аллила с металлом описывается ребром  $E_2$ ,  $\deg E_2 = 4$ . Аналогично, ребрами  $E_3$  и  $E_4$  описываются  $\pi$ -связь в цикlopentadiенильном лиганде и его связь с металлом.

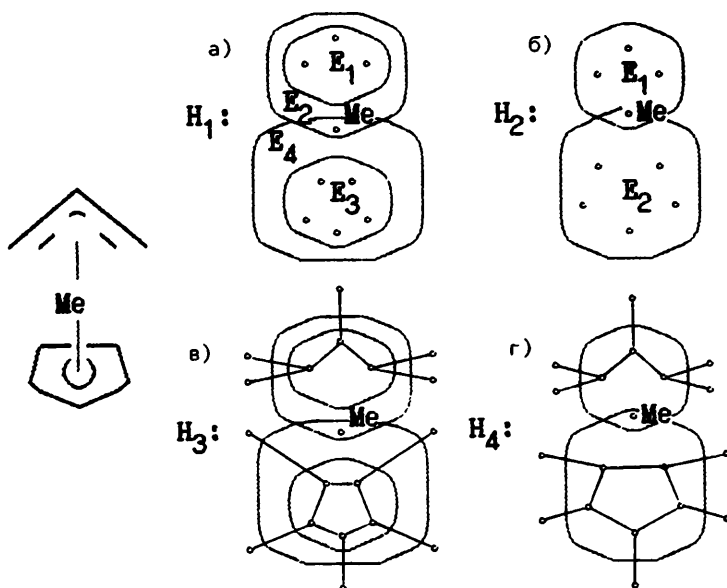


Рис. 5. Молекулярные гиперграфы, представляющие аллильный комплекс с разной степенью детализации

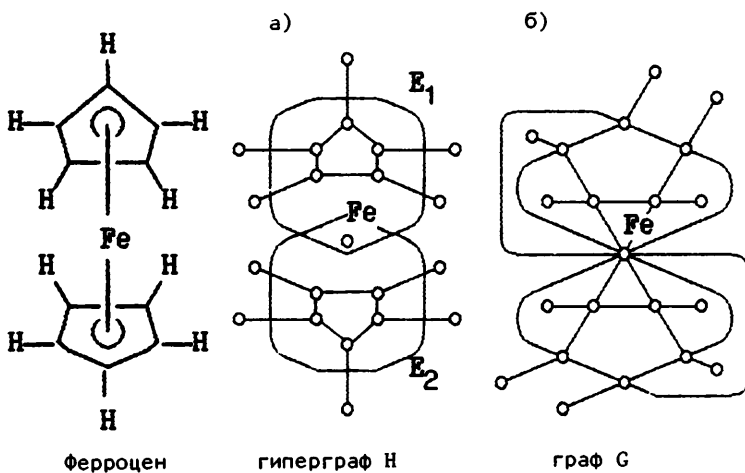


Рис. 6. Графовое и гиперграфовое представление ферроцена

В рассмотренном молекулярном гиперграфе представлены все делокализованные многоцентровые связи аллильного комплекса. Можно также рассматривать молекулярный гиперграф  $H_2$ , в котором присутствуют лишь гиперребра, соответствующие в молекулярной структуре связям атома металла с лигандом (рис.5б), либо молекулярные гиперграфы  $H_3$  и  $H_4$ , в которых представлены как делокализованные, так и ковалентные сигма-связи углерод-углерод и углерод-водород (рис.5в,г).

1.2. Примеры представления молекулярных структур. Рассмотрим примеры, демонстрирующие различия в представлении одних и тех же молекулярных структур графами и гиперграфами.

В представлении ферроцена обыкновенным графом связь металл-лиганд требует десяти ребер, соединяющих вершину Fe с остальными десятью вершинами (рис.6б).

Этот недостаток устранен в представлении ферроцена гиперграфом (рис.6а). Гиперребра  $E_1$  и  $E_2$  описывают  $\pi$ -связи между железом и лигандами, другие ребра соответствуют ковалентным связям. Кроме этого, представление ферроцена в виде гиперграфа позволяет наглядно демонстрировать различие между обыкновенными ковалентными связями углерод-углерод и углерод-водород и  $\pi$ -связями. В молекулярном гиперграфе ферроцена также "сохраняется" валентность атома железа, так как вершина Fe в гиперграфе  $H$  имеет степень 2. В обыкновенном графе  $G$ , описывающем данную молекулярную структуру, степень этой вершины равна 10 (рис.6б). Рассмотрим другой пример сохранения валентности атомов в гиперграфовом представлении. Представленный на рис.5г молекулярный гиперграф  $H_3$  аллильного комплекса имеет вершины, степени которых соответствуют валентностям атомов металла и атомов углерода. В случае обыкновенного графа (рис.2) валентность атома металла не сохраняется.

Гиперграфы оказываются удобными при описании кекулевских структур ароматических систем. В этом случае гиперребра в ги-



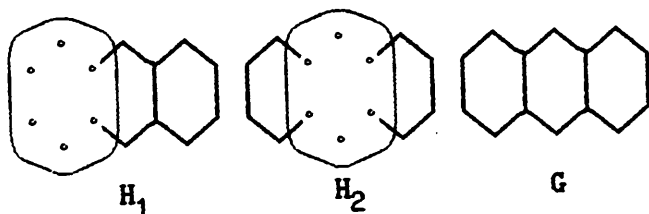


Рис. 7

перграфе соответствуют ароматическим секстетам. На рис.7 изображены гиперграфы  $H_1$ ,  $H_2$  кекулевских структуры  $S_1, S_2$ , представленных на рис.3. Описание этих структур в виде обыкновенных графов дает один и тот же граф  $G$ .

## 2. Идентификация молекулярных структур

При компьютерной обработке молекулярных структур необходимы средства их формального описания и идентификации. Распространенным способом задания обыкновенных молекулярных графов является матрица смежности, определяющая графы с точностью до изоморфизма [1,3]. Применяются также вершинные списки и линейные нотации [12].

В качестве компьютерного представления гиперграфа удобно использовать его матрицу инцидентий, определяющую гиперграф с точностью до изоморфизма. Известно [13], что одним из способов распознавания изоморфизма двух обыкновенных графов является поиск и сравнение их канонических видов.

2.1. Каноническая матрица инцидентий молекулярного гиперграфа. Из двух бинарных матриц одинаковой размерности большей считается та, у которой лексикографически больше строка в первой паре неравных строк. Максимальную из всех возможных матриц называют канонической матрицей.

Канонической  $(p \times q)$ -матрицей инцидентий  $B^*(H)$  гиперграфа  $H$  называется максимальная из всех возможных  $p! \cdot q!$  матриц инцидентий. Определение канонической матрицы инцидентий гиперграфа остается верным и для обыкновенных графов [14,15].

Очевидно, что каноническая матрица инцидентий единственна и может быть использована для получения канонического вида гиперграфа. Под каноническим видом гиперграфа будем понимать гиперграф, представленный канонической матрицей инцидентий. Нумерация вершин и ребер, соответствующая каноническому виду гиперграфа, называется канонической. На рис.8 представлен канонический вид гиперграфа  $H_2$ , рис. 56, аллильного комплекса и каноническая матрица инцидентий.

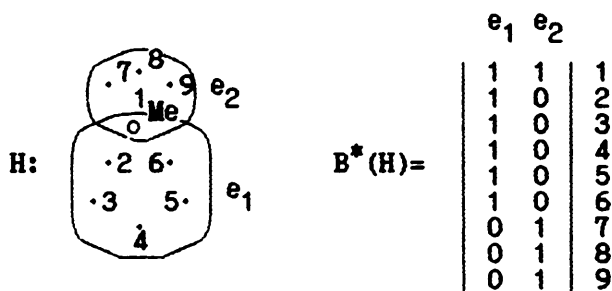


Рис. 8

Каноническим кодом  $C(H)$  гиперграфа  $H$  будем называть его представление в виде  $q$ ,  $q = |E(H)|$ , списков множеств помеченных или непомеченных вершин  $E_j = \{v_{i_j} : j = 1, \dots, q, i_j \in \{1, \dots, p\}\}$ , упорядоченных в соответствии с номерами строк канонической матрицы инцидентий. Для сокращения записи канонического кода условимся, что метки тех помеченных вершин, которые присутствуют в нескольких списках, указываются в  $C(H)$

только в первом по порядку списке. В остальных списках указываются только их номера.

На рис. 9 представлены канонические виды неизоморфных непомеченных гиперграфов  $H_1, H_2$  и графа  $G$ , изображенных на рис. 7. Там же записаны их канонические коды.

Канонические виды помеченных гиперграфов четырех аллильных соединений (см. рис. 1) и их канонические коды представлены на рис. 10. Заметим, что изображенные на рисунке гиперграфы содержат лишь те гиперребра, которые соответствуют связям атома металла с лигандом. Связи в самом лиганде в данном представлении не рассматриваются.

Канонические коды получены на основе простого метода построения канонической матрицы инцидентий гиперграфов.

## 2.2. Алгоритм построения канонической матрицы инцидентий.

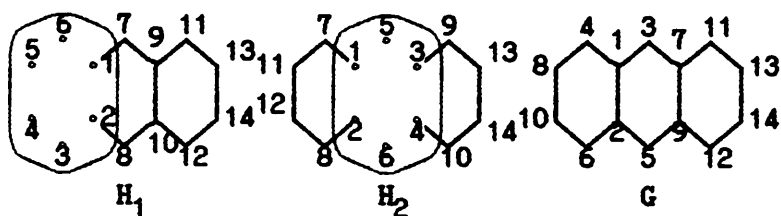
В алгоритме используется представление гиперграфа  $H = (V, E)$  в виде обыкновенного двудольного графа  $K(H) = (V, E, Y)$  (в дальнейшем называемый также кениговым представлением) с множеством вершин  $V \cup E$ ,  $|V \cup E| = p + q$ , и множеством ребер  $Y$ ,  $|Y| =$

$$= \sum_{j=1}^q \deg E_j.$$

Множества  $V$  и  $E$  называют долями графа:  $V$ -доля

и  $E$ -доля. Вершины  $v_i \in V$  и  $E_j \in E$  в  $K(H)$  смежны тогда и только тогда, когда в  $H$  вершина  $v_i$  и ребро  $E_j$  инцидентны.

Любой конечный гиперграф допускает кенигово представление, и наоборот, каждый граф Кенига является представлением некоторого конечного гиперграфа  $H$  и определяет его однозначно с точностью до изоморфизма [11]. Следовательно, изучение свойств гиперграфа можно свести к исследованию аналогичных свойств кенигова представления. В частности, каноническая матрица инцидентий гиперграфа  $H$  может быть получена из канонической матрицы смежности его кенигова представления  $K(H)$ .

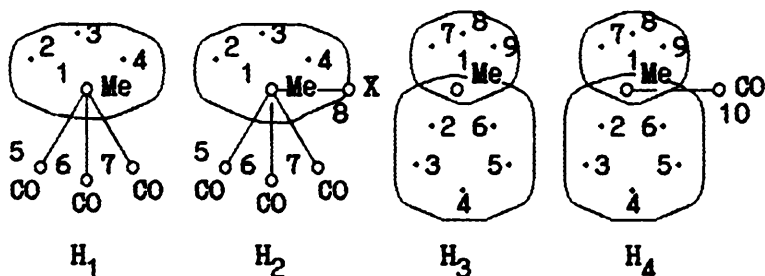


$$C(H_1) = (1,2,3,4,5,6) (1,7) (2,8) (7,9) (8,10) (9,10) (9,11) (10,12) \\ (11,13) (12,14) (13,14)$$

$$C(H_2) = (1,2,3,4,5,6) (1,7) (2,8) (3,9) (4,10) (7,11) (8,12) (9,13) \\ (10,14) (11,12) (13,14)$$

$$C(G) = (1,2) (1,3) (1,4) (2,5) (2,6) (3,7) (4,8) (5,9) (6,10) (7,11) \\ (8,10) (9,12) (11,13) (12,14) (13,14)$$

Рис. 9. Канонические виды и канонические коды графов, представляющих кекулевские структуры тетрацена



$$C(H_1) = (1Me, 2, 3, 4) (1, 5CO) (1, 6CO) (1, 7CO)$$

$$C(H_2) = (1Me, 2, 3, 4) (1, 5CO) (1, 6CO) (1, 7CO) (1, 8X)$$

$$C(H_3) = (1Me, 2, 3, 4, 5, 6) (1, 7, 8, 9)$$

$$C(H_4) = (1Me, 2, 3, 4, 5, 6) (1, 7, 8, 9) (1, 10CO)$$

Рис. 10. Канонические виды и канонические коды гиперграфов аллильных соединений, изображенных на рис. 1

Пусть  $A(K(H)) = \|a_{ij}\|$ ,  $i, j = 1, \dots, p$  матрица смежности графа  $K(H)$ , в которой  $a_{ij} = 1$ , если вершины  $i$  и  $j$  смежны, и  $a_{ij} = 0$  в противном случае. Канонической  $(p \times q)$ -матрицей смежности  $A^*(K(H))$  графа  $K(H)$  называется максимальная из всех возможных  $p!$  матриц смежности.

Кроме обыкновенной матрицы смежности двудольного графа рассматривают [16] также приведенную матрицу смежности  $R(K(H)) = \|r_{ij}\|$ ,  $i = 1, \dots, p$ ,  $j = 1, \dots, q$ , где  $r_{ij} = 1$ , если вершины  $v_i$  и  $E_j$  смежны, и  $r_{ij} = 0$  в противном случае.

Очевидно, что матрицы  $R(K(H))$  и  $B(H)$  определяются одинаково. Тогда, если определить каноническую приведенную матрицу смежности  $R^*(K(H))$  двудольного графа  $K(H)$  аналогично канонической матрице инцидентов  $B^*(H)$  гиперграфа  $H$ , то верно следующее

**УТВЕРЖДЕНИЕ 1.**  $R^*(K(H)) = B^*(H)$ .

Следовательно, поиск канонической матрицы инцидентов  $B^*(H)$  сводится к поиску канонической приведенной матрицы смежности  $R^*(K(H))$ . Последняя может быть получена из матрицы  $A^*(K(H))$  на основе следующих утверждений.

**УТВЕРЖДЕНИЕ 2.** Каноническая матрица смежности кенигова представления гиперграфа  $H$  имеет симметричный блочно-диагональный вид

$$A^*(K(H)) = \begin{array}{c} \begin{array}{cccccc} n_1 & n_2 & n_3 & \dots & n_{t-1} & n_t \\ \begin{array}{|c|c|c|c|c|c|} \hline 0 & A_{12} & & & & \\ \hline A_{21} & 0 & A_{23} & & & \\ \hline & A_{32} & 0 & & & \\ \hline & \dots & & \dots & & \\ \hline & & & & 0 & A_{t-1,t} \\ \hline & & & & A_{t,t-1} & 0 \\ \hline \end{array} & \begin{array}{l} n_1 \\ n_2 \\ n_3 \\ \dots \\ n_{t-1} \\ n_t \end{array} \end{array} \end{array}$$

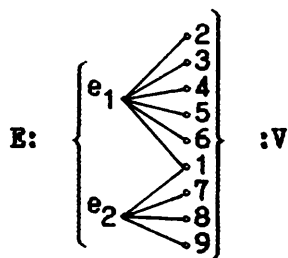
где строки (столбцы) ненулевых матриц  $A_{n_i n_{i+1}}$  ( $A_{n_{i+1} n_i}$ ),

$i = 1, \dots, t-1$ , определяются либо только вершинами из  $V$ -доли, либо только вершинами из  $E$ -доли, причем если  $n_i = |V_i|$ , где  $V_i$  есть подмножество множества вершин,  $V_i \subset V$ , то  $n_{i+1} = |E_{i+1}|$ , где  $E_{i+1}$  есть подмножество множества гиперребер,  $E_{i+1} \subset E$ , и наоборот.

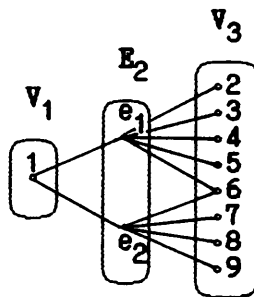
**ДОКАЗАТЕЛЬСТВО.** Пусть известна каноническая нумерация вершин графа  $K(H)$  и пусть первые номера в канонической нумерации имеют вершины множества  $V_1$ , принадлежащие  $V$ -доле. По определению две вершины в графе  $K(H)$  смежны тогда и только тогда, когда они принадлежат разным долям. Следовательно, вершины  $E$ -доли определяют множество вершин  $E_2$ , смежных с вершинами множества  $V_1$  и имеющих следующие за вершинами из  $V_1$  номера в канонической нумерации. Для вершин множества  $E_2$  смежными являются вершины из  $V$ -доли и т.д. Таким образом, граф  $K(H)$  с учетом канонической нумерации может быть представлен в виде  $t$ -дольного графа, в котором смежными являются вершины из соседних долей. Отсюда следует, что ненулевые элементы в матрице  $A^*(K(H))$  определяются только вершинами двух соседних долей  $V_i \subset V$  и  $E_{i+1} \subset E$ , т.е.  $A_{n_i n_{i+1}} \neq 0$  ( $A_{n_{i+1} n_i} \neq 0$ ), где  $n_i = |V_i|$  и  $n_{i+1} = |E_{i+1}|$ . Утверждение доказано.

Иллюстрацией утверждения 2 служит пример на рис.11.

В дальнейшем будем называть  $i$ -строку ( $i$ -столбец) блоков в матрице  $A^*(K(H))$   $i$ -блочной строкой ( $i$ -блочным столбцом). Пусть все  $i$ -блочные строки (столбцы), где  $i = 2k+1$ ,  $k = 0, 1, \dots$ , содержат номера всех вершин, принадлежащих  $V$ -доле. Тогда, по утверждению 2, все  $i$ -блочные строки (столбцы), где  $i = 2k$ ,  $k = 1, 2, \dots$ , содержат номера всех вершин, принадлежащих  $E$ -доле. Удалим из  $A^*(K(H))$  все четные блочные строки и все нечетные блочные столбцы. Полученная приведенная матрица смежности, обо-



Двудольный граф  $K(H)$   
гиперграфа



Граф  $K(H)$  в виде  
3-дольного графа

		$n_1$		$n_2$		$n_3$										
		1	$e_1$	$e_2$	2	3	4	5	6	7	8	9				
$A^* =$		0	1	1											1	
		1			1	1	1	1	1				$e_1$			
		1		0							1	1	$e_2$			
				1	0											2
				1												3
				1												4
				1												5
				1												6
				1												7
																8
																9

Каноническая матрица смежности  
графа  $K(H)$

Рис.11. Кенингово представление  $K(H)$  гиперграфа  $H$  (рис.8)  
и его каноническая матрица смежности

значим ее  $R(A^*(K(H)))$ , размерности  $(p \times q)$  будет иметь следующий вид:

$$A^*(K(H)) = \begin{array}{c|cccccc} & n_2 & n_4 & n_6 & & n_{t-3} & n_{t-1} \\ \hline A_{12} & 0 & & & & & \\ A_{32} & A_{34} & 0 & & & & \\ & A_{54} & A_{56} & \dots & & & \\ \dots & & \dots & \dots & \dots & & \dots \\ & & & & A_{t-2,t-3} & A_{t-2,t-1} & \\ & & & & 0 & A_{t-1,t} & \end{array} \begin{array}{l} n_1 \\ n_3 \\ n_5 \\ \\ n_{t-2} \\ n_t \end{array}$$

УТВЕРЖДЕНИЕ 3.  $R(A^*(K(H))) = R^*(K(H))$ .

ДОКАЗАТЕЛЬСТВО. По определению канонической приведенной матрицей смежности двудольного графа  $K(H)$  является максимальная из всех возможных  $p! \cdot q!$  матриц. Покажем, что матрица  $R(A^*(K(H)))$  удовлетворяет требованию максимальности. Последнее следует из свойств канонической матрицы  $A^*(K(H))$ , являющейся максимальной, и способа построения матрицы  $R(A^*(K(H)))$ . Предположим, что  $R(A^*(K(H)))$  не является максимальной, т.е. существует приведенная  $(p \times q)$ -матрица  $R'(A^*(K(H)))$ , имеющая лексикографически большую строку в первой паре неравных строк. Из способа построения приведенной матрицы смежности следует, что соответствующая ей обыкновенная матрица смежности  $A'(K(H))$  также должна иметь лексикографически большую строку, чем  $A^*(K(H))$ , что противоречит максимальности  $A^*(K(H))$ . Таким образом,  $R(A^*(K(H)))$  является максимальной и  $R(A^*(K(H))) = R^*(K(H))$ . Утверждение доказано.

На рис.12 представлена каноническая приведенная матрица смежности  $R^*(K(H))$ , полученная описанным выше способом из канонической матрицы смежности  $A^*(K(H))$  графа  $K(H)$  (рис.11). Заметим, что  $R^*(K(H))$  совпадает с канонической матрицей инцидентов  $B^*(H)$  гиперграфа  $H$  (см.рис.8).



$$B^*(H) = R^*(K(H)) =$$

$n_2$		
$e_1$	$e_2$	
1	1	1
1		2
1		3
1		4
1		5
1		6
	1	7
	1	8
	1	9

$\left. \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{matrix} \right\} n_1$   
 $\left. \begin{matrix} 7 \\ 8 \\ 9 \end{matrix} \right\} n_3$

Рис. 12

Таким образом, алгоритм поиска канонической матрицы инцидентий гиперграфа включает следующие шаги: построение графа  $K(H)$  гиперграфа  $H$ ; поиск канонической матрицы смежности графа  $K(H)$ ; построение приведенной канонической матрицы смежности графа  $K(H)$ .

Оценка сложности алгоритма определяется оценкой сложности алгоритма поиска канонической матрицы смежности. Заметим, что теоретически канонизация двудольных графов с долями  $N$  и  $M$  на основе их приведенных матриц смежности (по определению совпадающих с матрицами инцидентий) является более эффективной по сравнению с канонизацией их на основе обыкновенных матриц смежности, поскольку для первого подхода в худшем случае требуется  $N! \cdot M!$  операций, а для второго подхода требуется  $(N+M)!$  операций. На практике же реализованы эффективные алгоритмы канонизации именно обыкновенных матриц смежности [13]. Известные авторам алгоритмы канонизации матриц инцидентий [14, 15] являются переборными, оценки сложности данных алгоритмов либо не приводятся [14], либо таковы, что не позволяют оценить быстродействие алгоритма на практике [15].

### 3. Топологические индексы графов и гиперграфов

Различия, имеющиеся в представлении молекулярных структур обыкновенными графами и гиперграфами, отражаются на значениях топологических индексов. Ниже приведены результаты сравнительного анализа некоторых топологических индексов, основанных на расстоянии в графе. Расстоянием  $d(u,v)$  между вершинами  $u,v$  в графе  $G$  (гиперграфе  $H$ ) называется длина кратчайшей простой  $(u,v)$ -цепи. Длина любого гиперребра в  $H$  равна единице, т.е. все вершины, инцидентные одному и тому же гиперребру находятся друг от друга на расстоянии 1.

Пусть  $G = (V,E)$  и  $H = (V,E)$  есть представление молекулярной структуры  $F$  с делокализованными многоцентровыми связями в виде обыкновенного молекулярного графа и молекулярного гиперграфа, причем  $V(G) = V(H)$ . Тогда для  $u,v \in V(G) = V(H)$  верно следующее

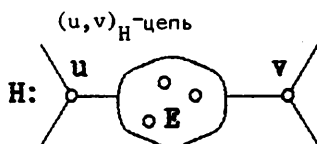
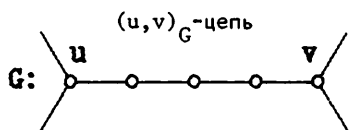
УТВЕРЖДЕНИЕ 4.  $d_G(u,v) \geq d_H(u,v)$ .

ДОКАЗАТЕЛЬСТВО. Возможны случаи.

Случай 1. Простая кратчайшая  $(u,v)$ -цепь в графе  $G$  и гиперграфе  $H$  состоит только из ребер степени 2. Поскольку  $G$  и  $H$  описывают одну и ту же молекулярную структуру  $F$  на одном и том же множестве вершин и поскольку  $(u,v)$ -цепи - кратчайшие, следовательно, данные цепи в обоих графах содержат одни и те же вершины и одно и то же число ребер, т.е. цепи имеют одну и ту же длину. Отсюда  $d_G(u,v) = d_H(u,v)$ .

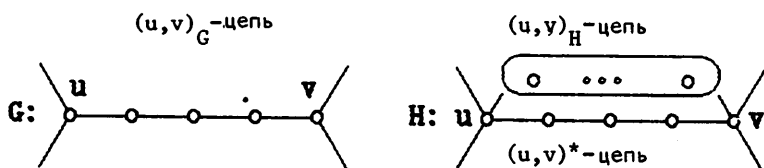
Случай 2. Простая кратчайшая  $(u,v)_G$ -цепь в графе  $G$  состоит только из ребер степени 2, а простая кратчайшая  $(u,v)_H$ -цепь в гиперграфе  $H$  содержит по крайней мере одно ребро  $E$ :  $\deg E > 2$ .

Предположим, что обеим цепям принадлежат одни и те же вершины:



Поскольку  $\deg E > 2$ , следовательно, расстояние между всеми вершинами, принадлежащими ребру, равно 1 и  $d_H(u,v) = d_G(u,v) - \deg E + 2$ , т.е.  $d_H(u,v) < d_G(u,v)$ .

Теперь предположим, что в графах  $G$  и  $H$   $(u,v)$ -цепи содержат разные вершины:



Пусть  $(u,v)^*$ -цепь есть цепь в гиперграфе  $H$ , содержащая те же вершины, что и кратчайшая  $(u,v)_G$ -цепь в графе  $G$ . Тогда в силу рассмотренных выше случаев

$$\text{либо } d_G(u,v) = d_H(u,v)^*, \quad (1)$$

$$\text{либо } d_G(u,v) > d_H(u,v)^*. \quad (2)$$

Но поскольку  $(u,v)_H$ -цепь является кратчайшей в  $H$ , то для нее верно:

$$d_H(u,v) < d_H(u,v)^*. \quad (3)$$

Тогда из (1)-(3) следует, что  $d_H(u,v) < d_G(u,v)$ . Утверждение доказано.

Из утверждения 4 следует, что для любой  $F$ , представленной  $G$  и  $H$ , выполняется неравенство  $W(G) > W(H)$ , где  $W$  есть индекс Винера (см. табл. 1). То есть значение индекса Винера для гиперграфа, представляющего молекулярную структуру  $F$ , меньше значения этого же индекса для обыкновенного графа структуры  $F$ .

Далеко не для всех топологических индексов удастся найти аналогичные закономерности. Также не всегда удастся оценить поведение топологических индексов на всем множестве графов и гиперграфов.

Т а б л и ц а 1

Формулы некоторых топологических индексов

№	Обобщенный индекс	Обобщенная формула
1	Индекс Винера [17]	$W = \frac{1}{2} \sum_{v,u} d(u,v)$
2	Индекс среднеквадратичных расстояний [цит.18]	$D^2 = \left[ \frac{\max \sum_{i=1} d_i \cdot i^2}{\sum_{i=1} d_i} \right]^{\frac{1}{2}}$
3	Индекс расстояний в графе [цит.18]	$VDI = \sum_{i=1}^{\max} (d_i)^2$

$d_i$  - число пар вершин, находящихся на расстоянии  $i$  друг от друга.

Т а б л и ц а 2

Упорядочение молекулярных структур по значениям индексов  $W$  и  $D^2$

№	W(G)	W(H)	№	$D^2(G)$	№	$D^2(H)$
1	32	30	1	1.6036	8	1.5000
6	34	33	8	1.6583	1	1.5119
7	47	46	6	1.6903	9	1.5916
8	57	51	9	1.7127	6	1.6475
9	74	68	2	1.7320	10	1.6569
2	75	72	7	1.7423	2	1.6733
10	93	87	10	1.7529	7	1.7113
4	100	98	3	1.7974	3	1.7541
3	136	132	4	2.4766	4	2.4494
5	300	296	5	2.7386	5	2.7203

Рассмотрим изменение значений топологических индексов (табл.1) на множестве молекулярных структур, состоящем из десяти комплексов рис.1 (по одной структуре каждого типа), представленных графами и гиперграфами. В табл.2 приведены значения индекса Винера  $W$  и индекса среднеквадратичных расстояний  $D^2$  для обыкновенных графов и гиперграфов. Как видно из табл.2 значения и первого, и второго индексов на множестве обыкновенных графов, превышают значения этих же индексов на множестве гиперграфов. Однако, если для индекса Винера упорядочение структур по его значениям на множествах графов и гиперграфов совпадают, то упорядочение структур по значениям индекса  $D^2$  на тех же множествах существенно различаются. Такие различия в общем случае могут оказаться полезными при исследовании корреляций структура-свойство [19].

Молекулярные гиперграфы можно применять для изучения зависимости свойств кекулевских структур от распределения в них ароматических секстетов. В табл.3 даны значения трех топологи-

Т а б л и ц а 3

Значения топологических индексов молекулярных гиперграфов кекулевских структур гептацена

№	VDI	W	$D^2(G)$
H1	17237	2371	6.3779
H2	18413	2251	6.0448
H3	19133	2179	5.8358
H4	19469	2155	5.7645

ческих индексов для молекулярных гиперграфов кекулевских структур гептацена (рис.13). Из таблицы видно, что между изменением значений индексов и расположением ароматического секстета в структурах имеется прямая зависимость: значения индексов либо строго уменьшаются (индексы  $W$  и  $D^2$ ), либо строго увеличиваются (индекс VDI) при "перемещении" секстета от концевой кольца к

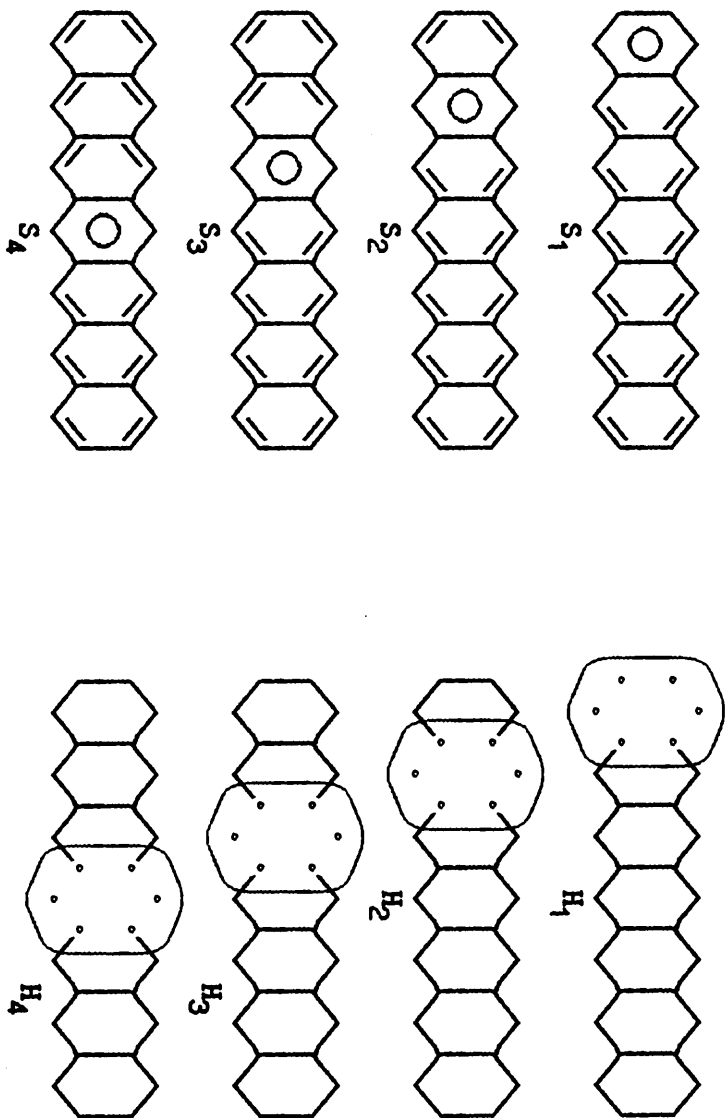


Рис. 13. Молекулярные гиперграфы кекулевских структур гептацена

центральному. В случае обыкновенных графов все структуры представляются одним и тем же графом.

### З а к л ю ч е н и е

В работе предложен новый способ представления молекулярных структур с многоцентровыми делокализованными связями, а также структур Кекуле в виде гиперграфов. Этот способ позволяет устранить недостатки представления таких структур в виде обыкновенных графов и облегчает задачу сравнения обычных молекулярных структур со структурами, содержащими многоцентровые связи [20].

Введено понятие молекулярного гиперграфа и на примерах показаны возможности его применения. Рассмотрен алгоритм построения канонических кодов молекулярных гиперграфов и их идентификации. Каноническое кодирование гиперграфов может быть использовано для получения формального описания неклассических молекулярных структур с целью их регистрации и поиска в базах данных, а также для решения проблем, связанных с созданием их номенклатуры [21].

Проведен сравнительный анализ топологических индексов молекулярных графов и молекулярных гиперграфов одних и тех же соединений. Из анализа следует, что упорядочение структур по значениям некоторых индексов может существенно различаться. Имеющиеся различия могут быть использованы в качестве дополнительной информации при прогнозировании свойств неклассических соединений по их структуре.

Применение гиперграфов для описания молекулярных структур неклассических соединений открывает новую область исследований, обобщающую результаты, полученные в химических приложениях теории обыкновенных графов, и расширяющую сферы использования математических подходов, которыми обладает теория гиперграфов.

## Л и т е р а т у р а

1. TRINAJSTIĆ N. Chemical Graph Theory. - Florida: CRC Press, Boca Raton. - Vol. 1. - 1983. - 147 p.
2. BONCHEV D. Information Theoretic Indices for Characterization of Chemical Structures. - New York: Research Studies Press, 1983.
3. ХАРАРИ Ф. Теория графов. - М.: Мир, 1973. - 300 с.
4. БАБАЕВ Е.В. Пространство изостерных ансамблей как форма естественной системы молекул //Принципы симметрии и системности в химии. - М., 1987. - С. 30-53.
5. HAIDUC I., ZUCKERMAN J.J. Basic Organometallic Chemistry. - Berlin, New York: de Gruyter, 1985. - 488 p.
6. СТАНКЕВИЧ И.В. Графы в структурной химии //Применение графов в химии /Под ред. Зефинова Н.С., Кучанова С.И. - Новосибирск: Наука, 1988. - С. 7-69.
7. ШУЛЬПИН Г.Б. Мир необычных молекул. - М.: Мир, 1986. - 174 с. (Серия "Наука и технический прогресс").
8. БЕРЖ К. Теория графов и ее применения. - М.: ИЛ, 1962. - 319 с.
9. КЛАР Э. Полициклические углеводороды. Т.1. - М.: Химия, 1971. - 442 с.
10. BERGE C. Graphes et hypergraphes.- Paris: Dunod, 1970.
11. ЗЫКОВ А.А. Гиперграфы //Успехи мат.наук. - 1974.-Т.29, вып. 6. - С. 89-154.
12. КОЧЕТОВА А.А., СКОРОБОГАТОВ В.А., ХВОРОСТОВ П.В. Язык описания структурной информации ОГРА-3.0 //Машинные методы обнаружения закономерностей, анализа структур, проектирования. - Новосибирск, 1982. - Вып. 92: Вычислительные системы. - С.70-79.
13. АРЛАЗАРОВ В.А., ЗУЕВ И.И., УСКОВ А.В., ФАРАДЖЕВ И.А.Алгоритм приведения конечных неориентированных графов к каноническому виду //Журн. вычисл.математики и мат.физики. - 1974. - Т.14, № 3. - С. 737-743.
14. БАРАЕВ А.М., ФАРАДЖЕВ И.А. Построение и исследование на ЭВМ однородных и неоднородных двудольных графов //Алгоритми - ческие исследования в комбинаторике. - М., 1978. - С. 25-61.
15. OVERTON M.L., PROSKUROWSKI A. Canonical incidence matrices of graphs //BIT-1979. - Vol. 19, N 2. - P. 271-273.



16. Лекции по теории графов /Емеличев В.А., Мельников О.И., Сарванов В.И., Тышкевич Р.И. - М.: Наука, Гл.изд.Физ.-мат.лит., 1990. - 384 с.

17. WIENER H. Structural determination of paraffin boiling points //J.Am.Chem.Soc. - 1947. - Vol. 69. - P. 17-20.

18. РУВРЭ Д. Следует ли заниматься разработкой топологических индексов? //Химические приложения топологии и теории графов: Пер. с англ. /Под ред. Р.Кинга. - М.: Мир, 1987. - С.183-205.

19. РАНДИЧ М., КРАУС Дж., ДЗОНОВА-ДЖЕРМАН-БЛАЗИЧ Б. Упорядочивание графов как подход к исследованию корреляций структур-активность //Химические приложения топологии и теории графов: Пер. с англ. /Под ред. Р.Кинга. - М., 1987. - С. 222-235.

20. СКОРОБОГАТОВ В.А., КОНСТАНТИНОВА Е.В. Гиперграфы молекул с делокализованными многоцентровыми связями //Тез. докл. межвузовской конференции "Молекулярные графы в химических исследованиях", Калинин, май 1990. - Калинин, 1990. - С.90.

21. КОНСТАНТИНОВА Е.В., СКОРОБОГАТОВ В.А. Компьютерный анализ молекулярных гиперграфов //Тез. докл. 9 Всесоюз.конф. "Химическая информатика", январь, 1992. - Черногловка, 1992. - С. 156.

Поступила в ред.-изд.отд.

22 июня 1994 года