

УДК 546.22/24:621.315.592:004.89

ПРОГНОЗИРОВАНИЕ СУЩЕСТВОВАНИЯ AB_3X_3 (X–S, SE, TE)

© 2009 г. Н. Н. Киселева

*Институт металлургии и материаловедения им. А.А. Байкова
Российской академии наук, Москваe-mail: kis@ultra.imet.ac.ru*

Поступила в редакцию 09.07.2008 г.

Впервые рассмотрена возможность существования при нормальных условиях соединений состава AB_3X_3 в системах $A_2X_3-B_2X$ (A и B – разные элементы; X–S, Se, Te). Для прогнозирования использована специально разработанная информационно-аналитическая система, объединяющая комплекс баз данных по свойствам неорганических веществ и материалов, и программы анализа данных на основе алгоритмов обучения ЭВМ.

ВВЕДЕНИЕ

Большинство соединений состава $A^III B_3^I X_3$ (X–S, Se, Te) относятся к полупроводниковым веществам [1–4]. Кристаллы соединений $AsAg_3S_3$ (прустит), $SbAg_3S_3$ (пираргирит) и $AsPt_3Se_3$ перспективны также для применения в нелинейной оптике и электрооптике [5]. В [5–7] обсуждены возможности использования соединений составов $AsAg_3S_3$ и $AsPt_3Se_3$ в акустооптике.

Цель данной работы – прогнозирование возможности образования новых соединений $A^III B_3^I X_3$ (X–S, Se, Te).

МЕТОДЫ ПРОГНОЗИРОВАНИЯ

В настоящее время существует два основных подхода к прогнозированию новых неорганических соединений: квантово-механические методы и эмпирические критерии. Несмотря на успехи физики в объяснении свойств твердых фаз, попытки расчета свойств сложных неорганических веществ на основе знания только свойств компонентов малоэффективны, что связано с трудностями квантово-механических расчетов многоэлектронных систем. Это и явилось причиной разработки эмпирических критериев образования фаз с определенными свойствами с целью сокращения времени и затрат на экспериментальный поиск новых веществ (примером могут служить известные правила Юм–Розери, Лавеса, Маттиаса, Гольдшмидта, Даркена–Гурри и т.п.).

Обычно такие критерии включают лишь несколько параметров компонентов. Достоинством таких критериев является простота и наглядность. Однако простейшие эмпирические критерии имеют существенный недостаток – неполное разделение классов систем, связанное, как правило, с тем, что не учитывается множество других, возможно не менее существенных, свойств компонентов.

Для поиска классифицирующих правил, включающих широкую совокупность свойств компонентов физико-химических систем, нами предложено использовать методы компьютерного анализа данных, основанные на алгоритмах обучения ЭВМ [8]. Этот подход обеспечивает конструирование сложных многомерных классифицирующих закономерностей на основе обработки больших массивов информации из баз данных (БД). При этом средняя ошибка прогнозирования с использованием сформированных критериев не превышает 80% [8].

Для автоматизации процедуры поиска многомерных критериев и их использования для прогнозирования новых неорганических соединений разработана специальная программная информационно-аналитическая система (ИАС) [9]. Эта система включает в себя комплекс БД по свойствам неорганических соединений [8–10] и подсистему анализа данных, в состав которой входят наиболее эффективные для применения в неорганической химии и материаловедении программы обучения ЭВМ [11, 12].

ПРОГНОЗИРОВАНИЕ НОВЫХ СОЕДИНЕНИЙ СОСТАВА AB_3X_3

Для компьютерного анализа использовали информацию входящей в состав ИАС БД “Фазы” [8–10] о 117 примерах образования соединений состава AB_3X_3 (X–S, Se, Te) и 58 примерах отсутствия соединений этого состава в системах $A_2X_3-B_2X$ при нормальных условиях (обучающая выборка). Для описания соединений в памяти ЭВМ на основе физико-химических представлений о природе веществ этого типа были выбраны две подсистемы свойств:

– свойства элементов A, B и X (табл. 1), значения которых взяты из разработанной нами БД “Элементы” (www.phases.imet-db.ru/elements), входящей в состав ИАС;

– свойства простых халькогенидов A_2X_3 и B_2X (стандартная энтропия и энтальпия, а также кова-

Таблица 1. Свойства элементов, использованные для описания соединений состава AB_3X_3

№	Свойство	№	Свойство
1	Температура плавления	14	3-й потенциал ионизации (I_3)
2	Ковалентный радиус (r_{cov})	15	Теплопроводность
3	Псевдопотенциальный радиус (по Цангеру)	16	Химический потенциал Мидемы (только для элементов А, В)
4	Квантовый номер	17	Энтальпия плавления
5	Расстояние до валентных электронов (по Шуберту)	18	Электроотрицательность (по Полингу)
6	Расстояние до внутренних электронов (по Шуберту)	19	Номер группы (только для элементов А и В)
7	Температура кипения	20	Молярная теплоемкость
8	Энтальпия испарения	21	Энтропия твердого тела
9	Ионный радиус (по Бокию и Белову)	22	Температура Дебая (только для элементов А и В)
10	Энтальпия атомизации	23	$(r_{covA} - r_{covB})/r_{covC}$
11	Регулярный номер (по Менделееву–Петтифору) (N_{MP})	24	$(I_{1A} - I_{1B})/I_{1C}$
12	1-й потенциал ионизации (I_1)	25	$(I_{2A} - I_{2B})/I_{2C}$
13	2-й потенциал ионизации (I_2)	26	$(I_{3A} - I_{3B})/I_{3C}$
		27	$(N_{MPA} - N_{MPB})/N_{MPC}$

лентные радиусы и отношение ковалентного радиуса к металлическому радиусу для элементов А, В и Х).

Для анализа данных использовали несколько алгоритмов обучения ЭВМ, включенных в разработанную ИАС: алгоритм вычисления оценок, метод бинарных решающих деревьев, двумерные линейные разделители, линейный дискриминант Фишера, алгоритм линейной машины, обучение нейронных сетей, метод k -ближайших соседей, генетический алгоритм, метод опорных векторов, голосование по тупиковым тестам, статистически взвешенное голосование, *LoReg* [11] и *ConFor* [12].

Для оценки точности прогнозирования использовали экзаменационное распознавание, которое проводили на материале обучающей выборки в двух режимах: без скользящего контроля и со скользящим контролем. В первом случае после обучения ЭВМ распознавались физико-химические системы из обучающей выборки. Однако такой способ оценки качества обучения и последующего прогнозирования является крайне ненадежным. В связи с этим использовали второй способ оценки качества обучения и прогноза – скользящий контроль. В этом случае заданная часть (f) объектов – описаний физико-химических систем – изымается из обучающей выборки, проводится обучение ЭВМ на оставшихся объектах, а изъятые объекты предъявляются для распознавания. Фиксируется количество ошибок. Затем “контрольные” объекты возвращаются в обучающую выборку, и из нее изымаются другие f объектов. Процедура повторяется до тех пор, пока все объекты обучающей выборки не побывают в роли “контрольных”. Такая процедура позволяет более объективно оценить точность алгоритмов обучения ЭВМ для конкретной обучающей выборки, особенно в

случае методов распознавания образов, для которых наблюдается нежелательный эффект “переобучивания”, когда алгоритм “настраивается” только на обучающую выборку и впоследствии дает плохие результаты для новых объектов. По завершении обучения со скользящим контролем рассчитывается средняя ошибка экзаменационного распознавания, которая далее используется для выбора наилучших алгоритмов для принятия коллективного решения.

Цель применения стратегии коллективных решений – повышение точности прогнозирования. Идея этого подхода заключается в следующем. Сначала для прогнозирования новых соединений независимо применяются разные алгоритмы обучения ЭВМ. Далее результаты работы отдельных алгоритмов специальным образом обрабатываются [11] и формируется окончательное коллективное решение об отнесении соединений к тому или иному из классов. В этом случае ошибки прогнозов с использованием отдельных алгоритмов чаще всего взаимно компенсируются, за счет чего в большинстве случаев удается получить более точный результат.

По результатам экзаменационного распознавания были отобраны следующие алгоритмы обучения ЭВМ:

– для формирования классифицирующих критериев, в состав которых входят свойства элементов (табл. 1), – линейная машина, *LoReg*, k -ближайших соседей [11] и *ConFor* [12], для которых достоверность прогноза с использованием скользящего контроля равна 85.7, 82.5, 85.7 и 95% соответственно;

– для поиска классифицирующих критериев, в состав которых входят свойства простых халькогенидов, – линейная машина, *LoReg*, линейный дискриминант Фишера, метод опорных векторов [11] и

Таблица 2. Прогноз возможности образования соединений состава АВ₃X₃ (X–S, Se, Te)

A	B	Al	P	Sc	Ti	Cr	Fe	Ga	As	Y	Rh	In	Sn	Sb	La	Ce	Pr	Nd	Pm	Sm	Eu	Gd	Tb	Dy	Ho	Er	Tm	Yb	Lu	Bi			
S																																	
Li	#1		1	1	#1	1	1	#2	#1			#2		#1	2	2	2	2														#2	
Na	#1	#1	1	1	1	1	#1	#2	#1	1	#1	#1	1	#1					1	1	1	1	1	1	1	1	1	1	1	1	1	#1	
K	#1	1	1	1	1	1	1	#2	#1	1	#1	1	1	#1	1		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	#2	
Cu	1	#2	#1	#1		#2	#1		#1	#1	1	#2	#1	#1	#2	#2	2	#2			#1	#1	#1	#1	#1	#1	#1	#1	#1	#1	#1	#1	
Rb	#1	1	1	1	1	1	1		1	1	1	#1	1	1	1			1	1	1	1	1	1	1	1	1	1	1	1	1	1	#1	
Ag		#2					#2	#2	#1			#2	#2	#1	#2	#2	#2	#2	2	2	2			2	#2	2						#2	
Cs	#1	1	1	1	1	1	#1	#2			1	1	1	#2	1			1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
Tl	#1		#2	1	1	#1	1	2	#1		1	#1	#1	#1	#2	2	#2	2	2	2	2	2								1	1		
Se																																	
A	B	Al	P	Sc	Ti	Cr	Fe	Ga	As	Y	Rh	In	Sn	Sb	La	Ce	Pr	Nd	Pm	Sm	Eu	Gd	Tb	Dy	Ho	Er	Tm	Yb	Lu	Bi			
Li	1		1	1	1	1	1	2	#2	1	1	2			2	2	2	2					1	1	1	1	1	1	1	1	1		
Na	1	#1	1	1	1	1	#1	#1	#1	1	1	1	1	#2									1	1	1	1	1	1	1	1	1	1	
K	1	#1	1	1	1	1	#1	#1	#1	1	1	1	1	#1		1			1	1	1	1	1	1	1	1	1	1	1	1	1	#1	
Cu	1		1	#1	1	1	1	#2	#1	#1	1	#2	#1	#1	2	2	2	2		#1	#1	#1	#1	#1	#1	#1	#1	#1	#1	#1	#1	#1	#1
Rb	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1			1	1	1	1	1	1	1	1	1	1	1	1	1	1	#1
Ag	1		2			2	2	#2	#1	2		2		#2	2	2	2	2	2	2	2			2	2	2	2	2	2	2	2	2	2
Cs	1	1	1	1	1	1	1	#1	1	1	1	1	1	1	1				1	1	1	1	1	1	1	1	1	1	1	1	1	1	#1
Tl	#1		1	1	1	#1	1	2	#1	1	1	#2	#1	#1	2	2	2	2								1	1	1	1	1	1	1	#2
Te																																	
A	B	Al	P	Sc	Ti	Cr	Fe	Ga	As	Y	Rh	In	Sn	Sb	La	Ce	Pr	Nd	Pm	Sm	Eu	Gd	Tb	Dy	Ho	Er	Tm	Yb	Lu	Bi			
Li	1			1	1		1	2		1	1				2	2	2	2	2						1	1	1	1	1	1	1		
Na	1	#1	1	1	1	1	1		1	1	1	1	1	#1				1	1	1	1	1	1		1	1	1	1	1	1	1	1	
K	1	#1	1	1	1	1	1	#1	1	1	1	1	#1	#1	1				1	1	1	1	1	1	1	1	1	1	1	1	1	1	#1
Cu	1	#2		1	1	1	1	#2	#2	#1	1	#2		#2	2	2	2	2	2	#1	#1	#1	#1	#1	#1	#1	#1	#1	#1	#1	#1	#1	#2
Rb	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1			1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Ag	1		2	2		2	2	#2	2	2		#2	2	#2	2	2	2	2	2	2	2	2	#2	2	2	2	2	2	2	2	2	2	2
Cs	1	#1	1	1	1	1	1	1	1	1	1	1	1	1	1	1			1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Tl	1			1	1		1	2			1	2	#1	#2	#2	2	2	2	2													#2	

ConFor [12], для которых достоверность прогноза с использованием скользящего контроля равна 65.3, 64.5, 64.5, 68.6 и 93.1% соответственно.

Далее результаты работы этих алгоритмов использовали в разных коллективных методах принятия решения [11] (методе Байеса, нахождении областей компетенции и шаблонов принятия решений, динамическом методе Вудса, методе логической коррекции, алгоритме выпуклого стабилизатора, комплексных комитетных методах). Наилучшие результаты экзаменационного распознавания при принятии коллективных решений [11] дали методы, основанные на нахождении шаблонов принятия решений, – поиск критериев, в состав которых входят

свойства элементов (достоверность экзаменационного распознавания 98.7%), и метод Байеса – поиск критериев, в состав которых входят свойства простых халькогенидов (достоверность экзаменационного распознавания 98.3%). Именно их использовали для прогноза еще не полученных соединений состава АВ₃X₃. При этом для прогнозирования необходима информация только о свойствах химических элементов, входящих в состав системы (табл. 1) или простых халькогенидов.

Окончательные прогнозы были получены на основе сравнения результатов расчетов с использованием сформированных критериев, включающих свойства элементов и простых халькогенидов. Если

результаты противоречили друг другу, окончательный прогноз считался неопределенным.

В табл. 2 даны примеры прогнозов соединений состава AB_3X_3 . Приняты следующие обозначения: 1 – прогноз образования соединения AB_3X_3 при обычных условиях; 2 – прогноз отсутствия AB_3X_3 соединения при обычных условиях; знаком # отмечены примеры, информация о которых использована для обучения ЭВМ; пустые клетки – неопределенный прогноз.

ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Вследствие противоречивости информации о существовании соединений состава AB_3X_3 отбор примеров для обучения ЭВМ был самой сложной задачей. Например, крайне противоречивы данные о существовании соединений состава $LnCu_3X_3$ (Ln –Eu–Lu, X–S, Se) [2, 13, 14]. Практически не исследованы халькогенидные системы щелочных металлов с лантанидами. Естественно, что отсутствие достоверных данных значительно снижает точность прогнозирования. Однако при появлении новых данных разработанная нами ИАС [9] позволит легко и быстро перенастроить классифицирующие критерии и уточнить прогнозы.

Следующий этап компьютерного конструирования соединений состава AB_3X_3 (X–S, Se, Te) связан с прогнозированием типа их кристаллической структуры при нормальных условиях и оценкой свойств (температуры плавления, ширины запрещенной зоны и т.д.) с использованием разработанного нами подхода [8, 15].

ЗАКЛЮЧЕНИЕ

С использованием методов обучения ЭВМ впервые получены прогнозы новых соединений состава AB_3X_3 (A и B – разные элементы; X–S, Se, Te). Анализ полученных прогнозов показывает, что существует большой резерв еще не синтезированных халькогенидных соединений состава AB_3X_3 . Наибольший интерес для применений в электронике представляет проверка полученных прогнозов для соединений с медью, серебром и таллием.

Автор благодарит А.В. Столяренко, В.А. Дударева, В.В. Подбельского, В.В. Рязанова и О.В. Сенько за помощь в создании ИАС.

Работа выполнена при частичной поддержке Российского фонда фундаментальных исследований (гранты 06-07-89120, 08-01-90427-Укр_а, 08-07-00437-а и №05-03-39009).

СПИСОК ЛИТЕРАТУРЫ

1. Лазарев В.Б., Куш З.З., Переш Е.Ю., Семрад Е.Е. Сложные халькогениды в системах $A^I-B^{II}-C^{VI}$. М.: Металлургия, 1993. 240 с.
2. Физико-химические свойства полупроводниковых веществ: Справочник. М.: Наука, 1979. 339 с.
3. Рустамов П.Г., Алиев О.М., Эйнуллаев А.В., Алиев И.П. Хальколантанаты редких элементов. М.: Наука, 1989. 284 с.
4. Гавриленко В.И., Грехов А.М., Корбутяк Д.В., Литовченко В.Г. Оптические свойства полупроводников: Справочник. Киев: Наук. думка, 1987. 607 с.
5. Берча Д.М., Ворошилов Ю.В., Сливка В.Ю. Сложные халькогениды и халькогенгалогениды (получение и свойства), Львов: Вища шк., 1983. 181 с.
6. Distanov V.E., Nenashev B.G., Kirdyashkin A.G., Serboulenco M.G. Proustite Single-Crystal Growth by the Bridgman–Stockbarger Method Using ACRT // J. Cryst. Growth. 2002. V. 235. № 1–4. P. 457–464.
7. Singh N.B., Suhre D., Gupta N. et al. Performance of TAS Crystal for AOTF Imaging // J. Cryst. Growth. 2001. V. 225. № 2–4. P. 124–128.
8. Киселева Н.Н. Компьютерное конструирование неорганических соединений. Использование баз данных и методов искусственного интеллекта. М.: Наука, 2005. 288 с.
9. Kiselyova N., Stolyarenko A., Ryzanov V., Podbel'skii V. Information-Analytical System for Design of New Inorganic Compounds // Int. J. "Information Theories & Applications". 2008. V. 2. № 4. P. 345–350.
10. Киселева Н., Мурат Д., Столяренко А. и др. База данных по свойствам тройных неорганических соединений "Фазы" в сети Интернет // Информационные ресурсы России. 2006. № 4. С. 21–23.
11. Журавлев Ю.И., Рязанов В.В., Сенько О.В. "РАСПОЗНАВАНИЕ". Математические методы. Программная система. Практические применения. М.: ФАЗИС, 2006. 176 с.
12. Гладун В.П. Процессы формирования новых знаний. София: СД "Педагог 6", 1995. 192 с.
13. Gulay L.D., Shemet V.Y., Olekseyuk I.D. et al. Investigation of the $R_2S_3-Cu_2S-PbS$ ($R = Y, Dy, Ho$ and Er) Systems // J. Alloys Compd. 2007. V. 431. № 1–2. P. 77–84.
14. Gulay L.D., Wolcyrz M., Pitraszko A., Olekseyuk I.D. Investigation of the $Tm_2Se_3-Cu_2Se-PbSe$ and $Lu_2Se_3-Cu_2Se-PbSe$ Systems at 870 K // Pol. J. Chem. 2006. V. 80. № 10. P. 1703–1714.
15. Kiselyova N.N., Stolyarenko A.V., Gu T., Lu W. Computer-Aided Design of New Wide Bandgap Semiconductors with Chalcopyrite Structure // Перспективные материалы (Спецвыпуск: Сб. тр. IX Российско-Китайского симпозиума. "Новые материалы и технологии"). 2007. С. 351–355.