

ТЕОРЕТИЧЕСКАЯ
НЕОРГАНИЧЕСКАЯ ХИМИЯ

УДК 546.121:004.855.5

ПРОГНОЗИРОВАНИЕ ТЕМПЕРАТУРЫ ПЛАВЛЕНИЯ
ДВОЙНЫХ ГАЛОГЕНИДОВ

© 2023 г. Н. Н. Киселёва^{a, *}, В. А. Дударев^a, А. В. Столяренко^a, А. А. Докукин^{a, b},
О. В. Сенько^b, Ю. О. Кузнецова^a

^aИнститут металлургии и материаловедения им. А.А. Байкова РАН,
Ленинский пр-т, 49, Москва, 119334 Россия

^bФедеральный исследовательский центр “Информатика и управление” РАН,
ул. Вавилова, 40, Москва, 119333 Россия

*e-mail: kis@imet.ac.ru

Поступила в редакцию 25.11.2022 г.

После доработки 06.01.2023 г.

Принята к публикации 13.01.2023 г.

Проведено прогнозирование температуры плавления при атмосферном давлении двойных галогенидов состава $AVHal_3$, $AVHal_4$, A_2VHal_4 , A_2VHal_5 и A_3VHal_6 (A и B – разные элементы, Hal = F, Cl, Br или I). Для расчетов была применена разработанная авторами система, основанная на использовании методов машинного обучения. Поиск компьютерных моделей был проведен на основе анализа информации об уже известных температурах плавления галогенидов. Для прогнозирования неизвестных значений температуры плавления галогенидов использовали только значения свойств элементов A, B и Hal. Было показано, что применение программ, основанных на методологии ансамблей алгоритмов машинного обучения, позволяет получить наиболее точные оценки температур плавления (средние абсолютные ошибки, определенные методом скользящего контроля, в пределах 29–52 К в зависимости от состава галогенидов и выбранного алгоритма). Коэффициент множественной детерминации для моделей, использованных для прогнозирования, не ниже 0.7.

Ключевые слова: галогенид, температура плавления, машинное обучение, прогнозирование

DOI: 10.31857/S0044457X22602012, EDN: SNDUYN

ВВЕДЕНИЕ

Прогнозирование температуры плавления ($T_{пл}$) неорганических соединений на основе знания только свойств элементов, входящих в их состав, относится к одной из важнейших задач как химии (прогнозирование фазовых диаграмм, оценка физических свойств еще не полученных веществ и т.п.), так и материаловедения (поиск новых материалов с заданными свойствами). Еще в 1910 г. Ф.А. Линдеман попытался предсказать температуру плавления кристаллических материалов [1]. Метод был основан на наблюдении, что рост температуры связан с ростом амплитуды тепловых колебаний. При этом плавление начинается, когда среднеквадратическое значение амплитуды колебаний атомов превышает некоторую пороговую величину. К сожалению, точность оценки температуры плавления с помощью формулы, предложенной Линдеманом, была низкой. Многочисленные попытки улучшить эту формулу [2–5] также были безуспешны. Для решения этой задачи использовали и квантово-механические методы, например, с применением методов

молекулярной динамики были рассчитаны [6] значения температуры плавления фаз в давно изученной системе Hf–Ta–C [7]. При всей фундаментальности последнего исследования, его результаты пока не дают надежды на прогнозирование температур плавления еще не полученных соединений.

ТЕОРЕТИЧЕСКИЙ АНАЛИЗ

В последние годы для прогнозирования температур плавления неорганических соединений стали широко использоваться методы машинного обучения. Первые работы в этом направлении были проведены еще в 70-х годах [8]: с применением машинного обучения удалось предсказать температуру плавления простых оксидов состава A_2O_3 , зная только свойства элемента A. Обучение ЭВМ при его использовании в неорганической химии основано на поиске сложных закономерностей в экспериментальных данных, связывающих некоторое заданное свойство соединений со свойствами химических элементов, входящих в их состав. В настоящее время разработано мно-

жество программ машинного обучения, успешно решающих такие задачи, например, пакеты scikit-learn [9] и R [10], разработанная нами [11] система ParIS (Parameters of Inorganic Substances) и т.д. Применение машинного обучения позволило успешно решить множество задач прогнозирования температур плавления неорганических соединений. Например, в работах [12–14] с достаточно высокой точностью были предсказаны значения температур плавления эквиатомных соединений состава АВ. Прогнозы температуры плавления двойных и тройных неорганических соединений разного состава были получены авторами [15–19]. Настоящая работа посвящена прогнозированию температуры плавления двойных галогенидов состава $ABnAl_3$, $ABnAl_4$, A_2BnAl_4 , A_2BnAl_5 и A_3BnAl_6 (А и В – разные элементы, $Hal = F, Cl, Br$ или I), образование и тип кристаллической структуры которых при обычных условиях были предсказаны нами ранее [20, 21].

МЕТОДЫ РАСЧЕТА

Информационной основой для прогнозирования свойств соединений являются базы данных (БД). Репрезентативность и достоверность анализируемой информации в первую очередь определяют точность прогнозирования. Исходная информация для компьютерного анализа представляется в виде матрицы (обучающей выборки), каждая строка которой содержит описание уже известного соединения, для которого экспериментально определено значение оцениваемого параметра этого соединения (например, температуры плавления). В описание соединения входят значения свойств химических элементов (признаков) – компонентов соединения, а также известное значение целевого параметра. В настоящей работе вся эта информация извлекается из разработанных нами баз данных: информация о свойствах элементов – из БД “Elements” [22], информация об известных значениях выбранного свойства галогенидов (температуры плавления) – из БД “Фазы” [23]. Для формирования обучающей выборки пользователю нужно только ввести набор символов элементов, входящих в состав уже изученных соединений, и значение параметра, например температуры плавления соединения, и специальная программа автоматически формирует выборку для анализа. Выборка для прогнозирования также включает подобные описания, состоящие из значений свойств химических элементов, но не содержит значений прогнозируемого параметра. В процессе анализа информации обучающей выборки по специальным программам обучения ЭВМ происходит поиск закономерностей, связывающих свойство соединения со свойствами элементов. Форма этих закономерностей определяется типом применяемой

программы. Это может быть обученная нейронная сеть, система уравнений или неравенств, булево выражение и т.д. Как правило, такие закономерности имеют очень сложный вид, что объясняется сложной зависимостью свойств веществ от свойств химических элементов, поэтому процедура прогнозирования выполняется с помощью специальных программ, что значительно облегчает процедуру предсказания свойств веществ. В разработанной нами системе ParIS [11] для прогнозирования количественных свойств неорганических веществ пользователю достаточно ввести символы химических элементов, входящих в состав прогнозируемого соединения, и он сразу же получает прогноз.

В настоящее время система ParIS включает специально разработанные нами программы [24–26], также 31 программу из свободно распространяемого пакета scikit-learn [9].

Процедура компьютерного анализа и прогнозирования состоит из нескольких этапов.

1. Отбор экспериментально изученных соединений и экспертная оценка исходных данных. На этом этапе из БД “Фазы” [23] по запросу пользователя выводится весь массив хранящейся информации о соединениях заданного состава и температурах их плавления. Поскольку в этой БД хранятся “сырые” данные, извлеченные из публикаций, задача оценки качества информации решается специалистом, готовящим выборку для последующего компьютерного анализа. Эту задачу нельзя полностью формализовать, однако нами предложено использовать для этих целей анализ диаграмм отклонений рассчитанных значений параметров от экспериментальных, процедура которого подробно описана в [27]. Использование таких диаграмм позволяет сократить время на экспертизу, так как специалист сразу получает информацию о составе соединения, экспериментальное и расчетное значения, “кликавая” по точке на диаграмме, наиболее отклоняющейся от экспериментального значения. Следует отметить, что экспертная оценка правильности данных является наиболее сложной и длительной процедурой при применении машинного обучения в химии, а разброс экспериментальных значений, например температуры плавления двойных галогенидов, доходит до ± 100 К по данным разных авторов. Большое значение имеет наличие полных текстов публикаций в разработанных нами БД, что упрощает экспертизу.

2. Отбор свойств химических элементов для включения в описание соединений. Первичный отбор проводится специалистом, обучающим компьютерные программы, на основе физико-химических представлений о природе изучаемых соединений и прогнозируемого свойства. Информация извлекается из БД “Elements” [22]. Пользователи

могут выбрать также различные простые алгебраические функции от исходных свойств элементов для включения в описание, и соответствующие результаты будут автоматически добавлены в обучающую выборку. Можно также вручную ввести более сложные уравнения, и программа добавит результаты вычисления в описание соединений. Более того, изменяя набор свойств элементов и оценивая точность прогнозирования при машинном обучении, можно отобрать наиболее важную совокупность свойств, обеспечивающую более точные прогнозы.

3. Обучение ЭВМ и отбор наиболее точных программ. При прогнозировании новых галогенидных соединений [20, 21] использовали разработанную нами информационно-аналитическую систему [28], включающую БД по свойствам неорганических веществ и программы машинного обучения, предназначенные для прогнозирования категориальных (дискретных) свойств. Для прогноза температур плавления применяли разработанную нами систему ParIS [11], предназначенную для прогнозирования количественных свойств веществ. Точность прогнозирования с использованием разных программ оценивали на основе вычисления средней абсолютной (MAE – Mean Absolute Error) и среднеквадратичной ошибки (MSE – Mean Squared Error) (при скользящем контроле в режиме LOOCV – Leave-One-Out Cross-Validation, который является обычной процедурой при малых выборках для обучения), коэффициента детерминации R^2 и т.д., а также анализа диаграмм отклонений рассчитанных значений параметров от экспериментальных для веществ, информация о которых была использована при обучении ЭВМ. Наиболее “точные” программы в дальнейшем применяли для прогнозирования.

4. Прогнозирование. В настоящей работе прогнозировали температуры плавления галогенидов при обычных условиях. Для прогнозирования использовали только значения свойств элементов, входящих в состав соединений. Система сама подставляет значения свойств элементов из БД “Elements” в закономерность, полученную в результате машинного обучения с использованием самых “точных” программ, и выводит прогнозируемые значения. Пользователи вводят только набор символов элементов, входящих в состав прогнозируемого соединения. Конечный прогноз вычисляется на основе использования либо лучшего алгоритма, либо усреднения полученных результатов.

РАСЧЕТНАЯ ЧАСТЬ

Задачу прогнозирования температуры плавления решали отдельно для каждого из составов двойных галогенидов $ABHal_3$, $ABHal_4$, A_2VHal_4 ,

A_2VHal_5 и A_3VHal_6 . Количество примеров для обучения ЭВМ: $ABHal_3$ – 241, $ABHal_4$ – 229, A_2VHal_4 – 204, A_2VHal_5 – 141 и A_3VHal_6 – 439. Обучающие выборки включали экспериментальные результаты разных исследователей. Одной из самых сложных проблем в этом случае было отсеивание резко отличающихся значений, для чего использовали экспертную оценку, в том числе и полные тексты публикаций, включенные в БД “Фазы”. Следует отметить, что для многих известных соединений не был определен тип кристаллической структуры или значение температуры плавления, поэтому они были добавлены в выборку для прогноза наряду с предсказанными ранее соединениями [20, 21].

На основе экспертной оценки были отобраны свойства химических элементов A, B и Hal для включения в искомые закономерности, а именно: псевдопотенциальный орбитальный радиус (по Цангеру), ионный радиус (по Шеннону), расстояния до внутренних и валентных электронов (по Шуберту), энергии ионизации первого, второго и третьего электронов, номера по Менделееву–Петтифору, квантовый номер, электроотрицательность (по Полингу), химический потенциал Мидемы, температуры плавления и кипения, стандартные энтропия, энтальпия атомизации, теплопроводность, молярная теплоемкость и т.д. [22] – всего 87 значений параметров элементов для каждого соединения.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Сравнение значений параметров MAE, MSE и R^2 наиболее точных алгоритмов системы ParIS при экзаменационном распознавании обучающей выборки в режиме LOOCV приведено в табл. 1. В большинстве случаев наилучшие показатели были получены при использовании алгоритмов Extra Trees Regressor (ETR), Gradient Boosting Regressor (GBR), Hist Gradient Boosting Regressor (HGBR) из пакета scikit-learn и специально разработанных для системы ParIS метода Simple Syndrome Regressor (SAND) и двухуровневого метода Recursive Regressor (RR), который является комбинацией нескольких методов из пакета scikit-learn (Bagging Regressor (BR), GBR или Random Forest Regressor) с методом кластеризации массива параметров элементов по степени корреляционной связи. Более точные результаты, получаемые с помощью алгоритмов ETR, GBR, HGBR, можно объяснить тем, что в их основе лежит принцип создания ансамблей алгоритмов, позволяющих компенсировать недостатки одного алгоритма преимуществами других. Особенно предпочтительно применение таких алгоритмов для решения задач, в которых обучающие выборки имеют сравнительно малый размер.

Таблица 1. Результаты оценки точности прогнозирования температуры плавления двойных галогенидов

Состав	Алгоритм	R^2	MAE, K	MSE, K
ABHal ₄	ETR	0.887	47	7362
	GBR	0.888	51	7276
	HGBR	0.907	49	6021
	SAND	0.897	46	6716
	BR	0.898	49	6615
A ₂ BHal ₄	ETR	0.905	30	2037
	GBR	0.905	34	2030
	HGBR	0.907	34	1992
	RR	0.919	29	1747
A ₂ BHal ₅	KNR	0.727	52	4947
	GBR	0.719	48	5097
	HGBR	0.701	51	5423
A ₃ BHal ₆	ETR	0.910	35	3399
	GBR	0.900	38	3747
	HGBR	0.906	38	3538
	RR	0.911	35	3340
ABHal ₃	ETR	0.898	43	5690
	GBR	0.892	49	5987
	HGBR	0.892	51	5974
	RR	0.909	42	5052

Примечание. KNR – Kneighbors Regressor.

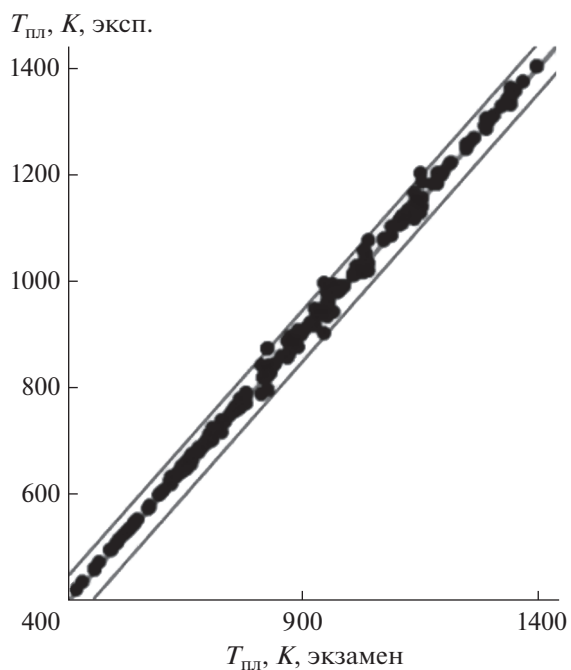


Рис. 1. Диаграмма отклонений прогнозируемых температур плавления соединений состава ABHal₃ от экспериментальных для алгоритма RR.

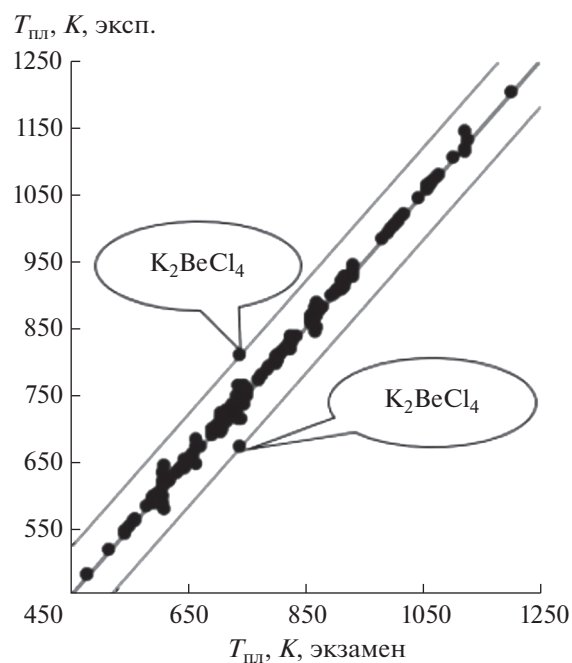


Рис. 2. Диаграмма отклонений прогнозируемых температур плавления соединений состава A₂BHal₄ от экспериментальных для алгоритма ETR.

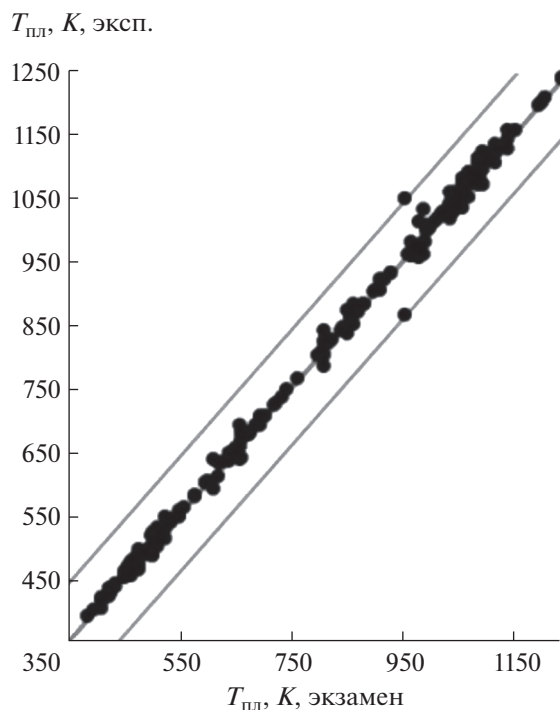


Рис. 3. Диаграмма отклонений прогнозируемых температур плавления соединений состава $ABHal_4$ от экспериментальных для алгоритма ETR.

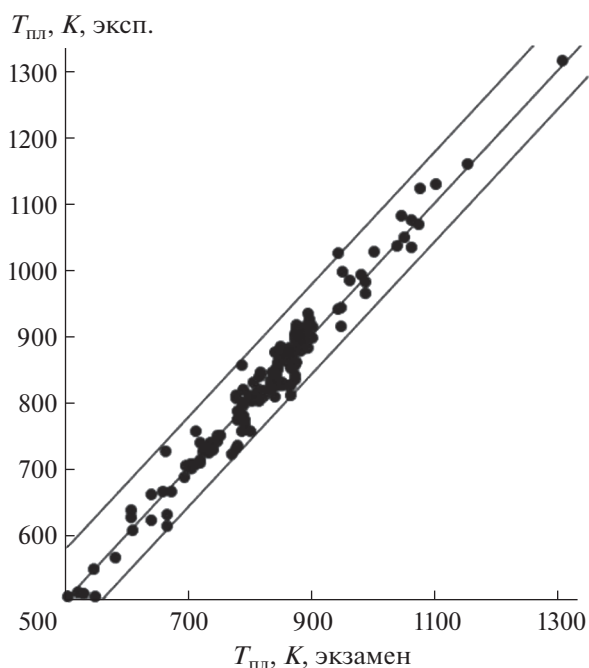


Рис. 4. Диаграмма отклонений прогнозируемых температур плавления соединений состава A_2BHal_5 от экспериментальных для алгоритма GBR.

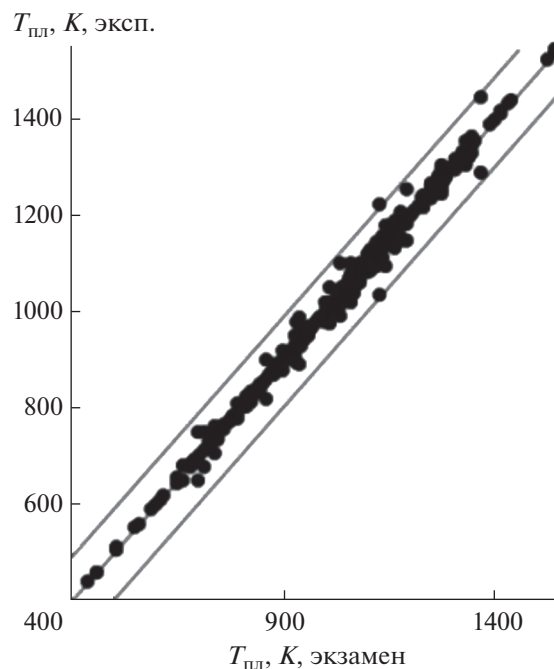


Рис. 5. Диаграмма отклонений прогнозируемых температур плавления соединений состава A_3BHal_6 от экспериментальных для алгоритма ETR.

В табл. S1–S5 даны результаты прогноза температуры плавления для галогенидов разного состава, чей тип кристаллической структуры был предсказан нами ранее [20, 21]. Для состава $ABHal_3$ был использован алгоритм RR, для других составов результат есть среднее значение прогнозов с использованием лучших алгоритмов (табл. 1): A_2BHal_5 – Kneighbors Regressor, GBR и HGBR, A_3BHal_6 – ETR и RR, A_2BHal_4 – ETR, GBR, HGBR и RR; $ABHal_3$ – HGBR, SAND и BR. В тех случаях, когда были известны температуры плавления соединений, но не были определены их кристаллические структуры, даны для сравнения и экспериментальные значения. Точность прогнозирования, определенная методом скользящего контроля в режиме LOOCV, колеблется от ± 29 до ± 52 K в зависимости от состава галогенидов и выбранного алгоритма (табл. 1).

Анализ диаграмм отклонений прогнозируемых значений температур плавления двойных галогенидов от экспериментальных, информация о которых использована для машинного обучения (рис. 1–5), дополнительно позволяет оценить качество прогнозов. Следует отметить, что чем ближе значения температур плавления (точки) к диагонали на таких диаграммах, тем точнее прогнозы. Для соединений состава $ABHal_3$ прогнозируемые и экспериментальные значения (рис. 1) практически полностью совпадают, что свидетельствует о высоком качестве модели, полученной с использовани-

ем алгоритма RR. При прогнозировании температур плавления соединений состава A_2BnA_4 (рис. 2) наиболее отклоняется от известных значений (809 [29] и 671 К [33]) прогноз 740 К для соединения K_2VeCl_4 , что является поводом дополнительной экспериментальной проверки. В целом количество подобных выбросов для соединений других составов равно 1–2 примерам, что свидетельствует о приемлемой точности прогнозирования на основе моделей, полученных в результате машинного обучения. Следует отметить, что предварительный анализ диаграмм отклонений позволил выявить десятки ошибок в публикациях, особенно в справочниках, которые были исправлены при формировании обучающих выборок.

ЗАКЛЮЧЕНИЕ

Впервые прогнозированы температуры плавления при атмосферном давлении двойных галогенидов разного состава, в том числе еще не полученных. Показана эффективность использования для прогнозирования программ, основанных на ансамблях алгоритмов машинного обучения. Соответствие полученных моделей, связывающих температуру плавления галогенидов определенного состава со свойствами образующих их химических элементов, экспериментальным данным, определенное на основе расчета коэффициента множественной детерминации R^2 , не ниже 0.7. Отмечены важность экспертной оценки информации о параметрах уже известных соединений, включаемой в выборку для компьютерного анализа, и влияние ошибочных экспериментальных данных на точность прогнозирования. Полученные прогнозы значений температур плавления двойных галогенидов могут использоваться в металлургической и электронной промышленности, также при поиске новых люминесцентных, магнитных, акустооптических, нелинейно-оптических, электрооптических и сегнетоэлектрических материалов и суперионных проводников.

БЛАГОДАРНОСТЬ

Авторы выражают благодарность проф. П.П. Федорову за ценные замечания при обсуждении статьи.

ФИНАНСИРОВАНИЕ РАБОТЫ

Работа выполнена по государственному заданию № 075-01176-23-00 и при частичной финансовой поддержке РФФИ, проекты 20-01-00609 и 21-51-53019.

КОНФЛИКТ ИНТЕРЕСОВ

Авторы заявляют, что у них нет конфликта интересов.

ДОПОЛНИТЕЛЬНЫЕ МАТЕРИАЛЫ

Дополнительная информация доступна в on-line версии: <https://doi.org/10.31857/S0044457X22602012>.

СПИСОК ЛИТЕРАТУРЫ

1. Lindemann F.A. // Phys. Z. 1910. Bd. 11. S. 609.
2. Ross M. // Phys. Rev. 1969. V. 184. № 1. P. 233. <https://doi.org/10.1103/PhysRev.184.233>
3. Stacey F.D., Irvine R.D. // Aust. J. Phys. 1977. V. 30. № 6. P. 631. <https://doi.org/10.1071/PH770631>
4. Boyer L.L. // Phase Transitions. 1985. V. 5. № 1. P. 1. <https://doi.org/10.1080/01411598508219144>
5. Owens F.J. // Phase Transitions. 2018. V. 91. № 5. P. 503. <https://doi.org/10.1080/01411594.2018.1432052>
6. Hong Q.-J., van de Walle A. // Phys. Rev. B: Condens. Matter. 2015. V. 92. № 2. P. 020104. <https://doi.org/10.1103/PhysRevB.92.020104>
7. Andrievskii R.A., Strel'nikova N.S., Poltoratskii N.I. et al. // Soviet Powder Metallurgy and Metal Ceramics. 1967. V. 6. № 1. P. 65. <https://doi.org/10.1007/BF00773385>
8. Савицкий Е.М., Грибуля В.Б. // Структура и свойства жаропрочных металлических материалов. М.: Наука, 1973. С. 3.
9. Pedregosa F., Varoquaux G., Gramfort A. et al. // J. Machine Learning Res. 2011. V. 12 (Oct.). P. 2825.
10. Сайт проекта R: <https://www.r-project.org/> (visited on 10.11.2022)
11. Dudarev V.A., Kiselyova N.N., Stolyarenko A.V. et al. // CEUR Workshop Proceedings (CEUR-WS.org), v. 2790. Supplementary Proceedings of the XXII International Conference on Data Analytics and Management in Data Intensive Domains (DAMDID/RCDL 2020). P. 89. <http://ceur-ws.org/Vol-2790/paper09.pdf>
12. Saad Y., Gao D., Ngo T. et al. // Phys. Rev. B: Condens. Matter. 2012. V. 85. № 10. P. 104104. <https://doi.org/10.1103/PhysRevB.85.104104>
13. Paliana G., Gubernatis J.E., Lookman T. // Phys. Rev. B: Condens. Matter. 2015. V. 91. № 21. P. 214302. <https://doi.org/10.1103/PhysRevB.91.214302>
14. Zhang Z., Peng R., Chen N. // Mater. Sci. Eng., B. 1998. V. 54. № 3. P. 149. [https://doi.org/10.1016/S0921-5107\(98\)00157-3](https://doi.org/10.1016/S0921-5107(98)00157-3)
15. Seko A., Maekawa T., Tsuda K., Tanaka I. // Phys. Rev. B: Condens. Matter. 2014. V. 89. № 5. P. 054303. <https://doi.org/10.1103/PhysRevB.89.054303>
16. Chen N., Li C., Yao S., Wang X. // J. Alloys Compd. 1996. V. 234. № 1–2. P. 130. [https://doi.org/10.1016/0925-8388\(95\)01963-4](https://doi.org/10.1016/0925-8388(95)01963-4)
17. Yan L.-M., Zhan Q.-B., Qin P., Chen N.-Y. // J. Rare Earths. 1994. V. 12. № 2. P. 102.
18. Seko A., Hayashi H., Nakayama K. et al. // Phys. Rev. B: Condens. Matter. 2017. V. 99. № 14. P. 144110. <https://doi.org/10.1103/PhysRevB.95.144110>
19. Gu T., Lu W., Bao X., Chen N. // Solid State Sci. 2006. V. 8. № 2. P. 129. <https://doi.org/10.1016/j.solidstatesciences.2005.10.01>

20. *Киселева Н.Н.* // Журн. неорган. химии. 2014. Т. 59. № 5. С. 665.
<https://doi.org/10.7868/S0044457X14050110>
21. *Киселева Н.Н., Столяренко А.В., Рязанов В.В. и др.* // Журн. неорган. химии. 2014. Т. 59. № 12. С. 1709.
<https://doi.org/10.7868/S0044457X1412010>
22. БД “Elements”: <https://phase.imet-db.ru/elements> (visited on 10.11.2022).
23. БД “Фазы”: <https://phase.imet-db.ru> (visited on 10.11.2022).
24. *Сенько О.В., Докукин А.А., Киселева Н.Н., Хомутов Н.Ю.* // Доклады Академии наук. 2018. Т. 479. № 1. С. 11.
<https://doi.org/10.7868/S086956521801-0016>
25. *Журавлев Ю.И., Сенько О.В., Докукин А.А. и др.* // Доклады Академии наук. 2021. Т. 499. С. 63.
<https://doi.org/10.31857/S2686954321040172>
26. *Ващенко Е.А., Витушко М.А., Дударев В.А. и др.* // Информационные процессы. 2019. Т. 19. № 4. С. 415.
27. *Киселева Н.Н., Дударев В.А., Столяренко А.В. и др.* // Перспективные материалы. 2021. № 9. С. 1.
<https://doi.org/10.30791/1028-978X-2021-9-5-23>
28. *Kiselyova N.N., Stolyarenko A.V., Ryzanov V.V. et al.* // Pattern Recognition and Image Analysis. 2011. V. 21. № 1. P. 88.
<https://doi.org/10.1134/S1054661811010081>
29. *Коршунов Б.Г., Сафонов В.В., Дробот Д.В.* Фазовые равновесия в галогенидных системах. Справочник. М.: Металлургия, 1979. 182 с.
30. *Коршунов Б.Г., Сафонов В.В.* Галогениды. Диаграммы плавкости. Справочник. М.: Металлургия, 1991. 288 с.
31. *Медведев В.А., Бергман Г.А., Васильев В.П. и др.* Термические константы веществ / Под ред. Глушко В.П. Вып. IX (Be, Mg, Ca, Sr, Ba, Ra). М.: Изд-во АН СССР, 1979. 574 с.
32. *Посыпайко В.И., Алексеева Е.А., Васина Н.А. и др.* Диаграммы плавкости солевых систем. Справочник. М.: Металлургия, 1977. Ч. I. 416 с. Ч. II. 304 с.
33. *Медведев В.А., Бергман Г.А., Васильев В.П. и др.* Термические константы веществ. Вып. X (Li, Na, K, Rb, Cs, Fr). Ч. 2. Таблицы принятых значений: K, Rb, Cs, Fr. М.: Изд-во АН СССР, 1981. Ч. 1. 299 с. Ч. 2. 441 с.