ISSN 0005-1055, Automatic Documentation and Mathematical Linguistics, 2012, Vol. 46, No. 1, pp. 40–49. © Allerton Press, Inc., 2012.
Original Russian Text © T.G. Petrov, A.A. Andriyanets-Buyko, S.V. Moshkin, 2012, published in Nauchno-Tekhnicheskaya Informatsiya, Seriya 2, 2012, No. 2, pp. 15–23.

A Two-Parameter Alphabet for Coding Structural–Chemical Information and its Systematization (Using the Example of Tourmaline)

T. G. Petrov[a], A. A. Andriyanets-Buyko[a], and S. V. Moshkin[b]

[a]*St. Petersburg State University, Universitetskaya nab. 7/9, St. Petersburg, 119164 Russia*
e-mail: tomas_petrov@rambler.ru; a-baa@mail.ru
[b]*OOO INFOCOMM, St. Petersburg, Russia*
e-mail; svmoshkin@rambler.ru

Received November 29, 2011

Abstract—A method for constructing a two-parameter *crystal chemical alphabet* for coding crystal chemical formulas (*CCFs*) for the purpose of their systematization is described. The general principles for coding the crystal chemical formulas are formulated. The two-parameter alphabet is an ordered collection of pairs of symbols in which data on a position and element (PE) that are peculiar to this structural type of mineral are fixed. Stoichiometric coefficients of elements in positions of particular CCFs allow one to construct rank crystal chemical formulas, i.e., sequences of PE pairs in the order of their coefficient decrease. The collection of these rank formulas is sorted on the dictionary principle in accordance with the PE in the proposed alphabet, allowing one to obtain a hierarchical classification of the CCF codes. The construction of rank CCF formulas enables one to calculate the entropy characteristics of the obtained codes in studies of transient processes of the structural–chemical states of substances. The method is described using the example of the mineral tourmaline.

Keywords: coding, crystal chemical alphabet, position–element, rank crystal chemical formula, hierarchical classification, relationship of crystal chemical formula and tourmaline.

DOI: 10.3103/S0005105512010086

INTRODUCTION

The problem of the uniform sorting of chemical compositions of minerals and other geological objects was solved previously [1–6] by creating the *RHA* informational language–method, which was intended for the description of compositions of any nature. Here, the *R*-rank formula is the sequence of symbols of chemical elements as their atomic contents decrease in the object composition; *H* (the measure of the composition complexity) is the negative sum of the products of the contents by their logarithms (Shannon entropy); and *A* (the measure of the composition purity) is the negative sum of the content logarithms (anisotropy). The sorting of the rank formulas as *words*, in which symbols of elements are taken as *letters*, is performed lexicographically (as in a dictionary). The Periodic Table of elements is used as an alphabet. As a result, the hierarchical periodic classification of compositions was obtained [6]. The use of the rank formulas of chemical compositions of minerals allowed us to create the *R-dictionary–catalog of the chemical compositions of minerals* [7]. The entire variety of the chemical compositions of minerals from known mineralogical Internet databases and several other sources are presented in it, as rank formulas. Collections of compositions of several mineral groups with entropy characteristics (*H* and *A*) are presented in [8] (scaploites), [9, 10] (eudialytes), and [11] (tourmalines), and on the Internet at the address: http://geology.spbu.ru/department/scientific/rha-language-method (micas, garnets, and tourmalines).

The standard method of writing the chemical compositions of minerals does not reflect their structural features, the knowledge of which is extremely important for a mineralogist. The crystal chemical formulas of minerals (CCFMs) contain this information. Therefore, the citing of these formulas in parallel with chemical analyses of minerals became standard in the geological literature. However, no systematized collections of CCFMs exist. This makes it difficult to review large data arrays, identify minerals from their compositions, find similar CCFs, comprehend their variability, estimate the prevalence or originality of a CCF, and reveal new varieties and new minerals. There is no possibility to reflect processes of structure–chemical changes of minerals on a plane, as can be done with the processes of the chemical changes by using entropy characteristics. The absence of system-

atized CCF collections shows that they cannot be used for sorting without any change of their image.

The variability of mineral structures does not allow one to solve the problem of developing a single system for describing the crystal chemical formulas of *all* minerals but does not end the possibilities for solving partial problems. These include the systematization of materials with the possible creation of databanks of the crystal chemical formulas of separate groups of isostructural minerals and the solution of genetic problems, which are similar to those already described for chemical compositions [3, 5, 12, 13]. In addition, the visual perception of complex poorly-perceived sequences of signs in the crystal chemical formulas (symbols of chemical elements with indexes, "+", "–" signs, brackets, and coefficients of the latter) should be simplified.

Experience from operation with chemical compositions using the *RHA* method has demonstrated its wide capabilities for systematizing information [7, 14]. This induced us to develope a system of conversion (coding) and sorting collections of the crystal chemical formulas on basis of the *RHA* method. For this coding, the authors consider it as necessary and sufficient to use three parameters: (i) a position symbol, (ii) an element symbol, and (iii) a number of position–element combinations in a particular structure. This means that as a first step it is necessary to develope a method for the unambiguous construction of a complex alphabet that takes both the position and element, for particular groups of minerals, into account. By an alphabet, a sign system with a fixed value of symbols and their unambiguous linear sorting is meant.

At present, there are no stringent, generally accepted rules of recording crystal-chemical formulas of minerals with complex structures even for minerals of a single group. Thus, there are eight versions of the *recalculation* of the CCFs and, accordingly, their image, for tourmaline alone [15]. This poses the separate problem of searching for the optimal calculation and representation method of the crystal chemical formulas with discussion of the issues of separate mineral groups under a separate cover.

The purpose of this article is to propose a method for coding two-parameter descriptions of objects and, in particular, the crystal chemical formulas of minerals, allowing one to create factual information systems, i.e., allowing one to systematize and search for analogs among properties but not from object names, and operate with theoretical and particular CCFs, which belong to separate groups of minerals. As an example, the alphabet version for coding the crystal chemical formulas of the tourmaline group was proposed. Bearing in mind that an algorithm for coding the crystal chemical information for any mineral is actually given, we will not discuss debating provisions, which relates to particular structures and ion distributions in tourmaline, which was selected as a sample, and strive for completeness of the alphabet itself. The latter is a separate task.

PRINCIPLES AND METHOD OF CONSTRUCTION OF AN ALPHABET FOR CODING CRYSTAL CHEMICAL FORMULAS

The term *alphabet* usually implies a generally accepted sequence of speech sound symbols in natural languages. However, ten types of the alphabet are distinguished in information science [16]. In addition, there are eight types of *new* alphabets among the most known [17]. The alphabets are rather widespread means for the fixation, ordered storage, coding, and transmission of information of different types.

Let us formulate the principles for the construction of the *crystal chemical alphabet*, which must be assumed as a basis of the coding system of two-parameter crystal chemical formulas of minerals (CCFMs) and their sorting. It is possible to construct three-parameter (and more) alphabets on the same logical base.

I. Principle of Conformity

The collection of alphabet symbols should correspond to a set of distinguished structural units of a mineral (positions) and their population by chemical elements, e.g., in the *reference* CCF. The term *reference* signifies the type of the crystal chemical formula in the largest dictionaries and reference books [18, 19].

II. Principle of Content

The complex symbol (code) should reflect the fixed (in the CCF) *position* of an ion (element) and the *symbol* of an ion (element) in abbreviated form: PE (position–element).

III. Principle of Order

The order of the symbols of positions in the created alphabet should correspond to the first alphabet, i.e., the most known (widespread in publications) order of symbols of positions in the CCF. PE pairs, which have *identical symbols of positions*, are ordered in accordance with their positioning in the second alphabet, for which the periodic system of chemical elements is taken.

IV. Principle of Simplicity

The CCF coding should contain: (i) the minimal generally accepted information on the CCF; and (ii) minimal information that is newly formed, i.e. a fact that calls for special explanation.

V. Principle of Conventionalities Minimization

When the alphabet is formed, the number of conventions should be minimal. The conventions include replacements of valence 2 by "k", valence 3 by "1", valence 4 by "m", etc.

VI. Principle of Restriction

The CCF code should not reflect the structure of the crystal in an explicit form. (For this purpose, other mechanisms exist). Thus, all brackets are open and discarded. The combinations of such signs as Si_4O_{10} and BO_3 are partitioned.

According to the special features of the software (Petros3), (i) the maximal number of symbols in the code should not exceed four signs, and (ii) digital information is not included in the symbols of the alphabet. Thus, these two limitations are technical.

So, to construct the alphabet, it is necessary and sufficient to have:

(i) an ordered list of positions;

(ii) an ordered list of elements that occupy these positions.

Two versions for constructing an alphabet that encodes the relationships of the chemical element and its position in the structure of mineral are possible:

(i) Position–element, where the symbols of positions are at the first place in the code and the symbols of chemical elements are in the second place.

(ii) Position–element, where the symbols of elements are in the first place in the code and the symbols of positions are in the second place.

These versions are equally possible. However, the degrees of definiteness of an element's existence and a position's existence in the crystal are substantially different. Thus, as a solid body, the crystal is first characterized by a structure, i.e., by some limited set of possible positions for elements. As to the chemical elements and ions that can form this mineral, uncertainties exist in both the length of the list of elements (over 50 were discovered in tourmaline) and the quantitative relationships between them. Therefore, it seems more natural to use just the *position–element* (PE) version. This approach allows one to create alphabets and, accordingly, systematized databases for groups of isostructural and other structurally related substances. When similar alphabets are constructed for other objects, it is natural to adhere to the similar ground of the symbol positioning in the code; i.e., the symbol of a more inert, more important parameter should be placed in the first place.

To create the *ideally complete* alphabet for a group of minerals, it is required to know all positions and all elements that are placed in each position. This stringent requirement is virtually unrealizable due to the absence of complete information on the distributions of the atoms in the structure. It is also unrealizable because it is impossible to foresee the discoveries of new elements in the mineral if the sensitivity of analytical methods is increased and if new varieties are discovered. Therefore, to illustrate the alphabet-constructing method (in accordance with the principle of conformity), let us consider the CCF representation form, which is the most ordinary and is universally accepted with respect to sets of positions and elements.

The crystal chemical alphabet is constructed as follows.

The first member of the alphabet, i.e., position (P) is in a series, without blanks, unified with symbols of elements (E) that are encountered in this position in an order that corresponds to the periodic system of elements. We obtain, e.g., P_1E_1, P_1E_3, and P_1E_8.

When all versions of elements that can be in P_1 are exhausted we start to construct a section of the alphabet that describes the connectives of the second position with elements (that can be in it). Accordingly, we obtain a row, e.g., P_2E_3, P_2E_9, P_2E_{25}, and so on up to the exhaustion of pairs of *position–element* symbols.

Note that the members of the alphabet are complex (compound) symbols, having a certain sense, as symbols of the periodic system of elements.

Further we call them the *letters of the crystal chemical alphabet*, from which the *words of the informational language of the CCFM* description, namely, *rank crystal chemical formulas of minerals (R-CCFM)*, are formed.

CONSTRUCTION OF THE POSITION–ELEMENT ALPHABET FOR THE CRYSTAL CHEMICAL FORMULAS OF TOURMALINE

Works [18, 20, 21] were taken as references. We confine ourselves to these works, assuming that an increase in the number of distinguished symbols and types of minerals would not influence the methods for constructing the alphabet and rank formulas.

Thus, we have a theoretical formula of minerals of the tourmaline group, i.e., $X\,Y_3\,Z_6\,[T_6O_{18}]\,[BO_3]_3\,V_3\,W$. Eight positions that were recognized in the references (X, Y, Z, T, B, O, V, W), are distinguished in this formula. Since the symbols Y, O, V, and W are also the symbols of the chemical elements in the periodic system, one should recognize that it would be unsuccessful to use them for designating positions in the *R*-CCF. So that needless discussions are not generated, we confine ourselves to only stating this fact.

As can be seen, several existing positions of oxygen that are related to Si and with B are not recognized and indicated in the theoretical formula of tourmaline. According to the reference data in [21], oxygen has positions O4, O5, O6, and O7 in a grouping with Si and O2, O8 in a grouping with B.

Table 1. Versions of the alphabet: (I) initial and (II) abbreviated

	I	II		I	II		I	II
1	XNa^{+}	XNa	14	YFe^{3+}	YFel	27	$O_{Si}O6$	OSi
2	XK^{+}	XK	15	YCu^{2+}	YCu	28	$O_{Si}O7$	
3	XCa^{2+}	XCa	16	YZn^{2+}	YZn	29	BB^{3+}	BB
4	X(G)	XG	17	ZMg	ZMg	30	B(G)	BG
5	YLi^{+}	YLi	18	ZAl	ZAl	31	O_BO2	OB
6	YMg^{2+}	YMg	19	ZV^{3+}	ZV	32	O_BO8	
7	YAl^{3+}	YAl	20	ZCr^{3+}	ZCr	33	$V(OH)^{-}$	VOH
8	YTi^{4+}	YTi	21	ZFe^{3+}	ZFel	34	VO^{2-}	VO
9	YV^{3+}	YV	22	$T(B^{3+})$	TB	35	$W(OH)^{-}$	WOH
10	YCr^{3+}	YCr	23	TAl	TAl	36	(OH)	VWOH
11	YMn^{3+}	YMnk	24	TSi	Tsi	37	WO^{2-}	WO
12	YMn^{3+}	YMnl	25	$O_{Si}O4$	OSi	38	WF^{-}	WF
13	YFe^{2+}	YFek	26	$O_{Si}O5$		39		

Note: Codings $O_{Si}O4$–$O_{Si}O7$ and O_BO2, O_BO8 designate positions of oxygen in a grouping with silicon Si_6O_{18} and boron BO_3, respectively. The code VW designates undivided positions, i.e., the absence of information on the division of these two positions.

According to [18], we give a list of ions in the theoretical crystal chemical formulas of tourmaline:

Na^{+}, K^{+}, Ca^{2+}, vacancy (G); Li^{+}, Mg^{2+}, Al^{3+}, Ti^{4+}, V^{3+}, Cr^{3+}, Mn^{2+}, Mn^{3+}, Fe^{2+}, Fe^{3+}, Cu^{2+}, Zn^{2+}; Mg, Al, Si, B^{3+}, O^{2-}, O_{Si}, O_B, $(OH)^{-}$, F^{-}.

We construct the complex alphabet, adding the ions that are peculiar to it to the symbol of each position. We obtain the row:

XNa^{+}, XK^{+}, XCa^{2+}, X(G), YLi^{+}, YMg^{2+}, YAl^{3+}, YTi^{4+}, YV^{3+}, YCr^{3+}, YMn^{2+}, YMn^{3+}, YFe^{2+}, YFe^{3+}, YCu^{2+}, YZn^{2+}, ZMg, ZAl, ZV^{3+}, ZCr^{3+}, ZFe^{3+}, $T(B^{3+})$, TAl, TSi, BB^{3+}, B(G), $V(OH)^{-}$, VO^{2-}, $W(OH)^{-}$, WF, WO^2, $O_{Si}O4$, $O_{Si}O5$, $O_{Si}O6$, $O_{Si}O7$, O_BO2, and O_BO8.

This is the *complex crystal chemical* alphabet for describing the CCF of tourmaline, represented by 37 complex symbols. According to principle *IV*, the stated symbolism of the alphabet should be simplified. In those cases where the ion has a constant charge, the sign of the charge is removed. Valence 2 is replaced by "k", valence 3 is replaced by "1", etc., according to the standard mathematical symbolism for the cases, when the enumeration starts not from the beginning of the natural scale. The symbol VWOH designates ion OH, whose positions V and W are not distinguished in the context of the crystal chemical formula. The symbol OB designates atoms of oxygen, being part of the complex ion BO_3; the symbol OSi designates the atoms of oxygen, being part of the complex Si_6O_{18}. Thus, we obtain the simplified alphabet: XNa, XK, XCa, XG, YLi, YMg, YAl, YTi, YV, YCr, YMnk, YMnl, YFek, YFel, YCu, YZn, ZMg, ZAl, ZV, ZCr, ZFel, TB, TAl, TSi, OSi, BB, BG, OB, VOH, VO, WOH, VWOH, WO, and WF.

There are 34 symbols here. For obviousness, we summarized the obtained alphabet versions in a table (Table 1).

The analysis of oxygen calls for solving the special structural problem. It has been solved in single works [21]. Oxygen always occupies the first places in the crystal chemical rank formula (due to its large content); therefore, it does not clearly influence the distinguishing of tourmaline types. According to the principle of conformity, when oxygen is encoded, only its presence in central ions (silicon and boron) is taken into account.

OBTAINING THE RANK CRYSTAL CHEMICAL FORMULA OF THE MINERAL (R-CCFM)

The construction of the rank formula with using the formed PE alphabet is as follows:

(i) The number of atoms in each position is determined. This can be determined directly from the initial crystal chemical formula.

(ii) All PE pairs placed in the particular CCFM are arranged in accordance with a decrease in their numbers. (These numbers are discarded at this stage of operation). As a result, we obtain the rank crystal chemical formula, i.e., *word* of the crystal chemical

language. This word, *letters* of which are divided by blanks, can be written:

OSi OB ZAl = TSi VWOH YMg = BB XNa

Signs of equality are placed when the coefficients of the PE are equal and when they differ in no more than 15 rel %, i.e., a value of no more than 1.15, when a larger value is divided by a smaller value.

The obtained *R-CCF* images, i.e., obtained *words*, are positioned in the vertical direction, and the symbols of one rank are written one under another. The sorting is performed on the dictionary (alphabetical) principle. The sequence of the PE pairs is used in the obtained crystal chemical alphabet as an alphabet for sorting *words*.

As a result of the alphabetical sorting of the rank CCF formulas, the linear hierarchical periodic *R*-classification of the crystal chemical formulas is automatically obtained. Based on quantitative popularity relationships of position–element combinations, this classification always has the most significant (in the quantitative relation) PE pairs in the first places in the rank formula, i.e., distinguishes the skeleton of the mineral structure. Therefore, the shape of the initial CCF and its image in the form of the rank crystal chemical formula differ.

CODING OF THEORETICAL CRYSTAL CHEMICAL FORMULAS OF THE TOURMALINE GROUP AND CONSTRUCTION OF THEIR R-CCF CLASSIFICATION

As initial materials, theoretical crystal chemical formulas of all the known actual and hypothetical types of tourmaline from the reference book were taken [18]. They are listed in Table 2.

In Table 2, the material is arranged in accordance with its order in the initial edition. Sorting of this type usually means the absence of any relationship between adjacent records. This is related to the fact that the letters of a natural language, in contrast to the symbols of chemical elements, do not carry any information [7]. Therefore, in all dictionaries based on alphabets of natural languages, connections are usually absent between words that begin identically and are located nearby, excluding the similarity of word images (*bed, blood, crocodile, crocus, rabbit, crown, top, cross*).

On the other hand, the totality of these symbols, e.g., OSiAl, clearly orients the intuition of a professional, even one who is not familiar with the informational language *RHA*, to the chemical type of the geological object.

Overcoming the above-shown property of the ordinary method of sorting totalities of mineral names with their CCF is achieved as follows.

For each type of tourmaline, the complex PE symbols are ranked in the line of their coefficient decrease and this forms their rank formulas (*R*-CCF). As an example, we consider the construction of the *R*-CCF of uvite, whose theoretical crystal chemical formula is listed in Table 3 (line 1). The positions are indicated in line 2 and elements corresponding to them are listed in line 3. The PE symbols are listed in line 4. The PE coefficients are placed in line 5. The procedure of their sorting is named in line 6. Line 7 contains the constructed crystal chemical rank formula (*R*-CCF) of uvite and the corresponding coefficients of PE pairs are given below. As it is seen, the entire coding procedure is extremely simple and can be fulfilled manually.

The obtained totality of the *crystal chemical rank formulas* was sorted on the dictionary principle in accordance with the above position–element alphabet. As a result, the linear hierarchical *R*-classification of the crystal chemical formulas of tourmaline was created (Table 4).

DISCUSSION

As can be seen, the rank formulas of all CCFs differ. This means that when the crystal chemical formulas are encoded by the method that was used, their variety (namely, 34 types) did not decrease. This could occur if the principle of conformity (I) were not satisfied, or the principle of simplicity (IV) was violated.

The bottom line of Table 4 gives a number of different *R-CCFs* with different lengths of their rank formula. As is seen, the increase in the variety of rank formulas stops from the 10th rank. This indicates the sufficiency of the distinguished position–element pairs for the complete description of the entire set of currently existing theoretical compositions of tourmaline. It is understandable that the level of detail of the rank PE formulas both for the *real compositions* and for creating new theoretical ones is not limited by a value of 10 ns.

The two first (top) ranks, which encode the oxygen structure skeleton, which is organized and *cemented* by silicon and boron, are common for the entire set of the crystal chemical rank formulas. The third rank distinguishes six classes (types of tourmaline). There are 8 *R*-CCF classes in the fourth rank, 15 in the fifth rank, etc. The hierarchy of the structure of *R* classification of the crystal chemical formulas becomes apparent in this manner. Horizontal separators make this structure easily perceptible. Formally, the hierarchy means that each rank formula with length n is part of the only class with length $n - 1$.

The hierarchical structure of the classification revealed a sharp nonuniformity of partitioning the field of the crystal chemical compositions into varieties. Thus, if the *R*-formula OSi > OB > ZAl has 20 versions of extensions (types of tourmaline), single types of tourmalines correspond to the *R*-formulas OSi > OB > YCr, OSi > OB > ZV, and OSi > OB > ZCr and two belong to the formula OSi > OB > ZFel. The fact that chromdravite, vanadiodravite, fluor-chromdrav-

Table 2. Nomenclature and crystal chemical formulas for tourmaline

1	Buergerite	$NaFe_3Al_6(BO_3)_3(Si_6O_{18})O_3F$
2	Vanadiodravite	$NaMg_3V_6(Si_6O_{18})(BO_3)_3(OH)_4$
3	Hydroxy-liddicoatite_theor	$Ca(Li_2Al)Al_6(BO_3)_3(Si_6O_{18})(OH)_3O$
4	Hydroxy-uvite_theor	$CaMg_3(Al_5Mg)(BO_3)_3(Si_6O_{18})(OH)_3$
5	Dravite	$NaMg_3Al_6(BO3)_3(Si_6O_{18})(OH)_4$
6	Liddicoatite	$Ca(Li_2Al)Al_6(BO_3)_3(Si_6O_{18})(OH)_3F$
7	Magnesio-foitite_theor	$\square(Mg_2Al)Al_6(Si_6O_{18})(BO_3)_3(OH)_4$
8	Oxy-dravite	$Na(Al_2Mg)(Al_5Mg)(Si_6O_{18})(BO_3)_3(OH)_3O$
9	Oxy-liddicoatite_theor	$Ca(Li_{1.5}Al_{1.5})Al_6(Si_6O_{18})(BO_3)_3(OH)_3O$
10	Oxy-rossmanite_theor	$\square(Li_{0.5}Al_{2.5})Al_6(Si_6O_{18})(BO_3)_3O(OH)_3$
11	Oxy-uvite_theor	$Ca(Al_2Mg)(Al_4Mg_2)(Si_6O_{18})(BO_3)_3(OH)_3O$
12	Oxy-ferri-foitite_theor	$\square(Fe^{3+}_2Fe^{2+})(Fe^{3+})_6(Si_6O_{18})(BO_3)_3(OH)_3O$
13	Oxy-feruvite_theor	$Ca(Al_2Fe^{2+})(Al_4Mg_2)(Si_6O_{18})(BO_3)_3(OH)_3O$
14	Oxy-foitite_theor	$\square(Al_2Fe^{2+})Al_6(Si_6O_{18})(BO_3)_3(OH)_3O$
15	Oxy-chromdravite_theor	$Na(Cr_2Mg)(Cr_5Mg)(Si_6O_{18})(BO_3)_3(OH)_3O$
16	Oxy-schorl_theor	$Na(Al_2Fe^{2+})(Al_5Fe^{2+})(Si_6O_{18})(BO_3)_3(OH)_3O$
17	Oxy-elbaite_theor	$Na(Al_2Li)Al_6(Si_6O_{18})(BO_3)_3(OH)_3O$
18	Olenite	$NaAl_3Al_6(BO_3)_3(Si_6O_{18})O_3(OH)$
19	Povondraite	$NaFe_3Fe_6(BO_3)_3(Si_6O_{18})(OH)_4$
20	Rossmanite	$\square(LiAl_2)Al_6(Si_6O_{18})(BO_3)_3(OH)_4$
21	Uvite	$CaMg_3(Al_5Mg)(BO_3)_3(Si_6O_{18})(OH)_3F$
22	Ferri-uvite_theor	$Ca(Fe^{2+})_3(Al_5Mg)(BO_3)_3(Si_6O_{18})_3(OH)_3O$
23	Ferri-feruvite_theor	$Ca(Fe^{3+}_2Fe^{2+})(Fe^{3+}_4Mg_2)(Si_6O_{18})_3(BO3)_3(OH)_3$
24	Foitite	$\square(Fe_2Al)Al_6(BO_3)_3(Si_6O_{18})(OH)_4$
25	Fluor-dravite_theor	$NaMg_3Al_6(BO_3)_3(Si_6O_{18})O_3F$
26	Fluor-olenite	$NaAl_3Al_6(BO_3)_3(Si_6O_{18})O_3F$
27	Fluor-rossmanite_theor	$\square(LiAl_2)Al_6(Si_6O_{18})(BO_3)_3(OH)_3F$
28	Fluor-foitite_theor	$\square(Fe_2Al)Al_6(BO_3)_3(Si_6O_{18})(OH)_3F$
29	Fluor-chromdravite_theor	$NaMg_3Cr_6(BO_3)_3(Si_6O_{18})(OH)_3F$
30	Fluor-schorl_theor	$NaFe_3Al_6(BO_3)_3(Si_6O_{18})(OH)_3F$
31	Fluor-elbaite_theor	$Na(Li_{1.5}Al_{1.5})Al_6(BO_3)_3(Si_6O_{18})(OH)_3F$
32	Chromdravite	$NaMg_3Cr_6(BO_3)_3(Si_6O_{18})(OH)_4$
33	Schorl	$NaFe_3Al_6(BO_3)_3(Si_6O_{18})(OH)_4$
34	Elbaite	$Na(Li_{1.5}Al_{1.5})Al_6(BO_3)_3(Si_6O_{18})(OH)_4$

Note: The abbreviation *theor* relates to types that are distinguished as extreme members of hypothetical isomorphic rows. These minerals have not been found in nature.

ite, oxy-ferri-foitite, and povondravite sharply differ in *crystal chemical compositions* from most of the remaining ones corresponds to this, while the differences between liddicoatite and hydroxy-liddicoatite, as well as those between oxy-uvite and oxy-feruvite, are minimal. Thus, the differences in the degree of relationship between the types are visually apparent. The longer the total part of the rank crystal chemical formula is, the shorter the separators are, and the closer to each other the varieties of tourmaline are.

Table 3. Conversion of the crystal chemical formulas of uvite into the rank crystal chemical formulas of uvite

1	CCF of uvite	$CaMg_3(Al_5Mg)(BO_3)_3(Si_6O_{18})(OH)_3F$									
2	Positions	X	Y	Z	Z	B	OB	T	OSi	V	W
3	Elements	Ca	Mg	Al	Mg	B	O	Si	O	OH	F
4	PE symbols	XCa	YMg	ZAl	ZMg	BB	OB	TSi	OSi	VOH	WF
5	PE coefficients	1	3	5	1	3	9	6	18	3	1
6	PE sorting in the order of their coefficient decrease										
7	R-CCF of uvite (PE coefficients)	OSi OB TSi ZAl YMg = BB = V OH XCa = ZMg = WF (18 9 6 5 3 = 3 = 3 1 = 1 = 1)									

The same relationships will occur between any objects whose properties will be described by using alphabets similar to that presented here.

The degrees of the differences between minerals are not usually revealed in their names and only with difficulty in the crystal chemical formulas. The rank CCF formula shows the structural significance of each element. In this regard, it is interesting to note that the ranks of type-forming elements widely vary. The *group-forming* boron occupies 5–7 ranks and it is exceeded in quantity by many elements. This can be recognized as a manifestation of its large structure-forming role. One should also pay attention to the distinguishing eight classes from the ratio of filling ranks 3 and 4 by elements in positions T and Z (these are distinguished in bold type in Table 4). It was interesting to determine the connection between the differences in the physical properties and the degree of the relationship from the crystal chemical properties of the tourmaline groups that were distinguished by the described method.

We accepted the above listed Principle–positions as intuitively clear and oriented to obtaining the *first* initial result with the specified properties (alphabet). In turn, this was used to obtain the *second* result (for constructing *words*, i.e., rank crystal chemical formulas). In turn, the latter were used for constructing the final result (hierarchical CCF classification). The quality of the alphabet, rank CCFs and the *obviousness*, and *convenience* of their use are judged from the results of using this system. Thus, there is no sense in estimating the simplicity, convenience, and usefulness of such designations as *three*, *3*, I I I, out of the context of their use. In addition, it is necessary to teach users how to operate within this system, but it is always difficult to some extent to master a new system [22, 23].

The designed alphabet takes into account only positions and elements of the crystal chemical formulas that are relating to a limited list of hypothetical and theoretical compositions of tourmaline. However, the method and proposed principles have no any limitations on their use when operating with actual compositions. This calls for expanding the alphabet. For this purpose, it is necessary and sufficient to position new PE symbols into the corresponding places of the alphabet without changing the relative positions of the remaining symbols. It may be necessary to supplement the alphabet if additional structural information on the positions of separate oxygen atoms and the hydroxyl group is gained.

It is known that meaningful systematization substantially facilitates working with data and allows one to better see both advantages and drawbacks of the data under study. When the quantity of data is high, the coding and sorting of the crystal chemical formulas, together with corresponding names of minerals, allow one to unambiguously classify data, visually reflect fullness of data collections, whether objects are unique or common, and to identify types inside mineral groups. The creation of databanks for mineral groups by the proposed method of coding allows one to analyze collections of minerals with a higher specificity level than occurs when only their chemical compositions are taken into account. It seems that this new type of systematization of materials during detection of abnormalities in large databanks promotes refining the positions of elements in crystalline structures and the discovery of new minerals.

It is important to note that the construction of the rank crystal chemical formulas (first compound part of the *RHA* method) opens up the possibility of using entropy and anentropy as the most adequate characteristics of changes of compositions during the mixture and division processes [24]. The latter include crystallization [25]. It is known that when the number of the parameters of comparable systems that are taken into account is larger the systems can be better recognized. The distributions of the PE frequencies are more informative, as compared with the distributions of chemical elements. Therefore, the use of the entropy characteristics of the crystal chemical compositions allows one to increase the distinguishability of objects on *HA* diagrams in aggregates of points, whether they are images of mineral varieties or processes of their formation [12]. As a trial, the entropy characteristics of chemical and crystal chemical compositions of chlorites (for which the alphabet is different) of different zones of the ore deposit were calculated from the

Table 4. Hierarchical *R*-classification of the crystal chemical rank formulas of minerals of the tourmaline group

Rank crystal chemical formulas											Crystal chemical formulas
1	2	3	4	5	6	7	8	9	10	11	
OSi	OB	YCr =	TSi	VWOH	YMg =	BB	XNa				32_$NaMg_3Cr_6(BO_3)_3(Si_6O_{18})(OH)_4$
OSi	OB	ZAl =	TSi	YMg =	BB =	VOH	XNa =	WF			25_$NaMg_3Al_6(BO_3)_3(Si_6O_{18})(OH)_3F$
OSi	OB	ZAl =	TSi	YAl =	BB =	VO	XNa =	WF			26_$NaAl_3Al_6(BO_3)_3(Si_6O_{18})O_3F$
OSi	OB	ZAl =	TSi	YAl =	BB =	VO	XNa =	WOH			18_$NaAl_3Al_6(BO_3)_3(Si_6O_{18})O_3(OH)$
OSi	OB	ZAl =	TSi	YFek =	BB =	VOH	XNa =	WF			30_$NaFe_3Al_6(BO_3)_3(Si_6O_{18})(OH)_3F$
OSi	OB	ZAl =	TSi	YFek =	BB =	VO	XNa =	WF			1_$NaFe_3Al_6(BO_3)_3(Si_6O_{18})O_3F$
OSi	OB	ZAl =	TSi	BB =	VOH	YLi	XCa =	YAl =	WF		6_$Ca(Li_2Al)Al_6(BO_3)_3(Si_6O_{18})(OH)_3F$
OSi	OB	ZAl =	TSi	BB =	VOH	YLi	XCa =	YAl =	WO		3_$Ca(Li_2Al)Al_6(BO_3)_3(Si_6O_{18})(OH)_3O$
OSi	OB	ZAl =	TSi	BB =	VOH	YLi =	YAl	XNa =	WF		31_$Na(Li_{1.5}Al_{1.5})Al_6(BO_3)_3(Si_6O_{18})(OH)_3F$
OSi	OB	ZAl =	TSi	BB =	VOH	YLi =	YAl	XCa =	WO		9_$Ca(Li_{1.5}Al_{1.5})Al_6(Si_6O_{18})(BO_3)_3(OH)_3O$
OSi	OB	ZAl =	TSi	BB =	VOH	YAl	XNa =	YLi =	WO		17_$Na(Al_2Li)Al_6(Si_6O_{18})(BO_3)_3(OH)_3O$
OSi	OB	ZAl =	TSi	BB =	VOH	YAl	XG =	YLi =	WF		27_$\square(LiAl_2)Al_6(Si_6O_{18})(BO_3)_3(OH)_3F$
OSi	OB	ZAl =	TSi	BB =	VOH	YAl	XG =	YFek	WO		14_$\square(Al_2Fe^{2+})Al_6(Si_6O_{18})(BO_3)_3(OH)_3O$
OSi	OB	ZAl =	TSi	BB =	VOH	YFek	XG =	YAl =	WF		28_$\square(Fe_2Al)Al_6(BO_3)_3(Si_6O_{18})(OH)_3F$
OSi	OB	ZAl =	TSi	BB =	WOH	YAl	XG =	VO =	YLi		10_$\square(Li_{0.5}Al_{2.5})Al_6(Si_6O_{18})(BO_3)_3O(OH)_3$
OSi	OB	ZAl =	TSi	VWOH	YMg =	BB	XNa				5_$NaMg_3Al_6(BO_3)_3(Si_6O_{18})(OH)_4$
OSi	OB	ZAl =	TSi	VWOH	YFek =	BB	XNa				33_$NaFe_3Al_6(BO_3)_3(Si_6O_{18})(OH)_4$
OSi	OB	ZAl =	TSi	VWOH	BB	YLi =	YAl	XNa			34_$Na(Li_{1.5}Al_{1.5})Al_6(BO_3)_3(Si_6O_{18})(OH)_4$
OSi	OB	ZAl =	TSi	VWOH	BB	YMg	XG =	YAl			7_$\square(Mg_2Al)Al_6(Si_6O_{18})(BO_3)_3(OH)_4$
OSi	OB	ZAl =	TSi	VWOH	BB	YAl	XG =	YLi			20_$\square(LiAl_2)Al_6(Si_6O_{18})(BO_3)_3(OH)_4$
OSi	OB	ZAl =	TSi	VWOH	BB	YFek	XG =	YAl			24_$\square(Fe_2Al)Al_6(BO_3)_3(Si_6O_{18})(OH)_4$
OSi	OB	ZV =	TSi	VWOH	YMg =	BB	XNa =				2_$NaMg_3V_6(Si_6O_{18})(BO_3)_3(OH)_4$
OSi	OB	ZCr =	TSi	YMg =	BB =	VOH	XNa =	WF			29_$NaMg_3Cr_6(BO_3)_3(Si_6O_{18})(OH)_3F$
OSi	OB	ZFel =	TSi	BB =	VOH	YFek =	XG =	YFel =	WO		12_$\square(Fe_2^{3+}Fe^{2+})(Fe^{3+})_6(Si_6O_{18})(BO_3)_3(OH)_3O$
OSi	OB	ZFel =	TSi	VWOH	YFek =	BB	XNa				19_$NaFe_3Fe_6(BO_3)_3(Si_6O_{18})(OH)_4$
OSi	OB	TSi	ZAl	YMg =	BB =	VOH	XCa	ZMg	WF		21_$CaMg_3(Al_5Mg)(BO_3)_3(Si_6O_{18})(OH)_3F$
OSi	OB	TSi	ZAl	YMg =	BB =	VWOH	XCa =	ZMg			4_$CaMg_3(Al_5Mg)(BO_3)_3(Si_6O_{18})(OH)_3$
OSi	OB	TSi	ZAl	YFek =	BB =	VOH	XCa =	ZMg =	WO		22_$CaFe_3^{2+}(Al_5Mg)(BO_3)_3(Si_6O_{18})_3(OH)_3O$
OSi	OB	TSi	ZAl	BB =	VOH	YAl	XNa =	YMg =	ZMg =	WO	8_$Na(Al_2Mg)(Al_5Mg)(Si_6O_{18})(BO_3)_3(OH)_3O$
OSi	OB	TSi	ZAl	BB =	VOH	YAl	XNa =	YFek	ZFel	WO	16_$Na(Al_2Fe^{2+})(Al_5Fe^{2+})(Si_6O_{18})(BO_3)_3(OH)_3O$
OSi	OB	TSi	ZAl	BB =	VOH	YAl =	ZMg	XCa =	YMg =	WO	11_$Ca(Al_2Mg)(Al_4Mg_2)(Si_6O_{18})(BO_3)_3(OH)_3O$
OSi	OB	TSi	ZAl	BB =	VOH	YAl =	ZMg	XCa =	YFek =	WO	13_$Ca(Al_2Fe^{2+})(Al_4Mg_2)(Si_6O_{18})(BO_3)_3(OH)_3O$
OSi	OB	TSi	ZCr	BB =	VOH	YCr	XNa =	YMg =	ZMg =	WO	15_$Na(Cr_2Mg)(Cr_5Mg)(Si_6O_{18})(BO_3)_3(OH)_3O$
OSi	OB	TSi	ZFel	BB =	VWOH	ZFel	ZMg	XCa =	YFek		23_$Ca(Fe_3^{2+}Fe^{2+})(Fe_4^{3+}Mg_2)(Si_6O_{18})(BO_3)_3(OH)_3$
1	1	6	8	15	18	25	28	32	34	34	

Note: Numbers of the crystal chemical formulas correspond to the description and numbers in Table 2: (1) Buergerite, (2) vanadiodravite, (3) hydroxy-liddicoatite, (4) hydroxy-uvite, (5) dravite, (6) liddicoatite, (7) magnesio-foitite, (8) oxy-dravite, (9) oxy-liddicoatite, (10) oxy-rossmanite, (11) oxy-uvite, (12) oxy-ferri-foitite, (13) oxy-feruvite, (14) oxy-foitite, (15) oxy-chromdravite, (16) oxy-schorl, (17) oxy-elbaite, (18) olenite, (19) povondraite, (20) rossmanite, (21) uvite, (22) ferri-uvite, (23) ferri-feruvite, (24) foitite, (25) fluor-dravite, (26) fluor-olenite, (27) fluor-rossmanite, (28) fluor-foitite, (29) fluor-chromdravite, (30) fluor-schorl, (31) fluor-elbaite, (32) chromdravite, (33) schorl, and (34) elbaite. The theoretical members of the tourmaline group are distinguished by bold type in the numbers of the crystal chemical formulas. In the *R-CCFs*, the bold type distinguishes the elements of three to four ranks, as being important after oxygen, of the tourmaline skeleton, distinguishing eight classes from relationships of elements in positions T and *Z*. The bottom line of the table contains the number of different *R-CCFs* with different lengths of the rank formula (NR).

data given in [26]. It was determined that the fields of *HA* points of the crystal chemical compositions of the deposit zones were divided, while the fields of the *HA* points of the chemical compositions of the same chlorites overlapped.

The authors believe that the principles presented in this article can be used as a basis for creating similar alphabets for describing, not only other groups of minerals, but also complex alphabets to systematize similar analytic materials in other fields of knowledge.

CONCLUSIONS

One should stress that the proposed method for coding crystal chemical formulas is not a replacement for the formulas themselves, just as a code name, title, or description does not replace the subject itself and is not similar to it. The CCF code is not only the CCF image form. Therefore, the advantages and drawbacks of the proposed code can be compared only with other versions of CCF codes but not with the crystal chemical formulas themselves. The main issue is whether the use of the code opens up new possibilities, as compared with the customary use of the initial data. Some of these possibilities are described above.

By using two-parameter alphabets it is possible to create databases with hierarchical periodic structures that unify different types of information on socio–economic or other pairs of properties of complex systems.

For example:

Alphabet1 contain the professions that are studied at universities and Alphabet2 gives the positions that are actually occupied by their graduates.

Alphabet1 contains nationalities and Alphabet2 are professions, or types of crimes, or activity areas abroad.

Alphabet1 consists of countries and Alphabet2 are types of products for export, or activity areas of citizens abroad, or activity areas, for which citizens of this country have received Nobel Prizes.

The symbols of the first parameter (Alph1) are discrete essences, which are unchanged for the time of study (fixation). The symbols of the second parameter (Alph2) are also discrete, but they are parameters whose intensity and frequency of popularity vary.

The rank formulas are constructed with consideration for the frequencies of joint events. Further calculations of informational entropy and anentropy, as measures of the complexity and purity of a system that is used in studies of the evolutionary processes of the composition of a system, are possible.

Thus, two-parameter alphabets are applicable for organizing information that is studied by statistics with the use of correlation coefficients, but in which concrete information is submerged. Here (in the database), it becomes easily visible and useful in various ways, remembering that its representation opens up possibilities for using the *RHA* language–method, which was assumed as a basis for designing the presented alphabet.

As far as we know, the proposed method for coding crystal chemical formulas, the formation of an alphabet of a new type, and the construction of the crystal chemical formulas themselves and their sorting to obtain the *R-CCF* hierarchical classification has no analog in the existing references.

The calculations and sorting of the rank formulas of the crystal chemical composition were performed using the Petros3 program complex, which was created by S.V. Moshkin [27, 28]; this provides high speed and accuracy of material processing.

REFERENCES

1. Petrov, T.G., Ground of Variant of General Classification of Geochemical Systems, *Vestn. Leningr. Gos. Univ.*, 1971, no. 18, pp. 30–38.
2. Petrov, T.G., Information Language for Description of Multicomponent Object Compositions, *Nauchn. Tekhn. Inform., Ser. 2*, 2001, no. 3, pp. 8–18.
3. Petrov, T.G., Range-Entropy Approach to the Description of the Geological Object Compositions and Their Changes (on the Geological Coenology), *Obshch. Prikladn. Tsoenolog.*, 2007, no. 5, pp. 27–33.
4. Krasnova, N.I., Petrov, T.G., and Retyunina, A.V., Practical Aspects of RHA Method Use for Classification of Mica Group Minerals, *Vestn. St. Petersb. Gos. Univ., Ser. 7*, 2008, no. 2, pp. 3–19.
5. Petrov, T.G. and Farafonova, O.I., *Informatsionno-komponentnyi analiz. Metod RHA* (Information-Component Analysis. RHA Method), St. Petersburg: St. Petersb. Gos. Univ., 2005.
6. Petrov, T.G., Hierarchical Periodic System of Chemical Compositions and Its Connection wuth Periodical System of Elements, *Vestn. St. Peterb. Gos. Univ., Ser. 7,* 2009, no. 2, pp. 21–28.
7. Petrov, T.G. and Krasnova, N.I., *R-slovar'-katalog khimicheskikh sostavov mineralov.* - (R-Dictionary-Catalog of Chemical Compositions of Minerals), St. Petersburg: Nauka, 2010.
8. Zolotarev, A.A., Petrov, T.G., and Moshkin, S.V., Peculiarities of Chemical Composition of Scapolite Group Minerals, *Zapiski Vseross. Mineral. Obshch.*, 2003, no. 6, pp. 63–84.
9. Bulakh, A.G. and Petrov, T.G., Chemical Variety of Eudialyte Group Minerals, Range Formula for Their Compositions, Chemical and Chemical-Structural Varieties, *Zapiski Vseross. Mineral. Obshch.*, 2003, no. 4, pp. 1–17.
10. Bulakh, A.G. and Petrov, T.G., Chemical Variability of Eudialyte-Group Minerals and Their Sorting, *Neues Jahrb. Min.*, 2004, no. 3, pp. 127–144.
11. Andriyanets-Buyko, A.A., Krasnova, N.I., and Petrov, T.G., Variety of Tourmaline Composition and Their Chemical RHA Method-Based Classification, *Zapiski Vseross. Mineral. Obshch.*, 2007, no. 1, pp. 25–41.

12. Petrov, T.G., Farafonova, O.I., and Sokolov, P.B., Information Entropy Characteristics of Mineral and Rock as Reflection of Crystallization Process Tension, *Zapiski Vseross. Mineral. Obshch.*, 2003, no. 2, pp. 33–40.
13. Mukhamedzhanov, A.F. and Petrov, T.G., Consideration of Differential Intrusions with the Help of Chemical Information RHA Method, in *Struktura, veshchestvo, istoriya litosfery Timano-Severo-Ural'skogo segmenta* (Structure, Substance, Lithosphere History of Timan-North Ural Segment), Syktyvkar, 2006.
14. Petrov, T.G., RHA Method as Solution of Problem of Systematization of Analytical Data about Substantial Composition of Geological Objects, *Otech. Geologiya*, 2008, no. 4, pp. 98–105.
15. Krivovichev, V.G. and Zolotarev, A.A., *Khimicheskii sostav mineralov i graficheskie sposoby ego izobrazheniya* (Chemical Composition of Minerals and Graphical Method of Its Representation), St. Petersburg, 2002.
16. Pershikov, V.I. and Savinkov, V.M., *Tolkovyi slovar' po informatike* (Explanatory Dictionary on Informatics), Moscow: Finan. Statistics, 1991.
17. *Unikal'naya illyustrirovannaya entsiklopediya v tablitsakh i skhemakh* (Unique Illustrated Encyclopedia in Tables and Schemes), Moscow: AST, 2006.
18. Krivovichev, V.G., *Mineralogicheskii slovar* (Mineralogical Dictionary), St. Petersburg: St. Petersb. Gos. Univ., 2008.
19. *Mineralogicheskaya entsiklopediya* (Mineralogical Encyclopedia), Frey, K., Ed., Leningrad: Nedra, 1985.
20. Hawthorne, F.C. and Henry, D.J., Classification of the Minerals of the Tourmaline Group, *Eur. J. Mineral.,* 1999, vol. 11, pp. 201–215.
21. Rozhdestvenskaya, I.V., Frank-Kamenetskaya, O.V., Zolotarev, A.A., et al., Refinement of the Crystal Structures of Three Fluorine-Bearing Elbaites, *Crystall. Rep.*, 2005, vol. 50, no. 6, pp. 907–913.
22. Salyamon, L.S., About Some Factors Determining of New Word Perception in Science, in *Nauchnoe otkrytie i ego vospriyatie*, (Scientific Discovery and Its Perception), Mikulinskii, S.R. and Yaroshevskii, M.G., Eds., Moscow: Nauka, 1971, pp. 95–114.
23. Ivanov, S.A., *1000 let ozarenii* (1000 Years of Strikenings), Moscow: Slovo, 2002.
24. Shurubor, Yu.V., About One Property of Measure of Complication of Geochemical Systems, *Dokl. Akad. Nauk SSSR*, 1972, vol. 205, no. 2, pp. 453–456.
25. Petrov, T.G., Problems of Separation and Mixing in Inorganic Systems, *Geologiya*, 1995, vol. 2, pp. 181–186.
26. Barsukov, V.L., Volosov, A.G., Kozerenko, S.V., et al., *Nauchnye osnovy geokhimicheskikh metodov poiskov glubokozalegayushchikh rudnykh mestorozhdenii* (Scientific Fundamentals of Geochemical Methods of Searching of Deeply Imbedded Ore Deposits), Irkutsk, 1970.
27. Moshkin, S.V., Shelemotov, A.S., Bogachev, et al., RETROS – New Program Complex for Treatment and Analysis of Petrogeochemical Information, in *Petrografiya na rubezhe XXI veka. Itogi i perspektivy* (Petrography on the Boundary of XXI Century. Results and Perspectives), Syktyvkar, 2000.
28. Petrov, T.G. and Moshkin, S.V., RHA Method and Its Realization in Retros-3 Program Complex, *Vychisl. Geolog.*, 2011, no. 1, pp. 50–53.