

Research Article

Open Access



Composition genes in materials

Shuang Zhang¹, Qing Wang², Chuang Dong^{1,2}

¹School of Materials Science and Engineering, Dalian Jiaotong University, Dalian 116028, Liaoning, China.

²School of Materials Science and Engineering, Dalian University of Technology, Dalian 116024, Liaoning, China.

Correspondence to: Prof. Chuang Dong, School of Materials Science and Engineering, Dalian Jiaotong University, No. 794 Huanghe Road, Dalian 116028, Liaoning, China; School of Materials Science and Engineering, Dalian University of Technology, No. 2 Linggong Road, Dalian 116024, Liaoning, China. E-mail: dong@djtu.edu.cn

How to cite this article: Zhang S, Wang Q, Dong C. Composition genes in materials. *J Mater Inf* 2021;1:8. <http://dx.doi.org/10.20517/jmi.2021.04>

Received: 8 Jul 2021 **First Decision:** 8 Oct 2021 **Revised:** 1 Nov 2021 **Accepted:** 9 Nov 2021 **Published:** 23 Nov 2021

Academic Editor: Xingjun Liu **Copy Editor:** Yue-Yue Zhang **Production Editor:** Yue-Yue Zhang

Abstract

High-performance materials always possess specific chemical compositions. The present work points out that the composition genes, which are the basic structural units that serve as the composition carriers, are actually the molecule-like chemical units. Friedel oscillations, in combination with the cluster-plus-gluce-atom model, are fully presented to show how to uncover the composition genes hidden in chemical short-range orders in any material. Examples are given in three categories of materials, i.e., metallic alloys including solid solutions and metallic glasses, inorganic compounds as well as relevant glasses, and polymers. Furthermore, materials can be classified into single-, dual-, and multi-gene types. The proposition of composition genes facilitates the understanding of prevailing materials and can be a useful tool to guide the exploration of new composition space.

Keywords: Composition genes, chemical units, cluster-plus-gluce-atom model, Friedel oscillations

INTRODUCTION

Solids originate from strong chemical bonding between atoms and make a rich material world^[1]. The commonly used materials are conventionally classified into metallic alloys, inorganic compounds (as well as relevant glasses), and polymers (as well as their hybrids). In terms of chemical compositions, the materials are differently expressed. Metallic alloys, for the convenience of large-scale industrial production, are generally expressed in mass percentages of elements, such as 304 stainless steel 06Cr19Ni10 (the percent is placed after the element) and brass 70Cu-30Zn (the percent is placed before the element),



© The Author(s) 2021. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.



which are summarized into various industrial grades. Inorganic compounds are mostly represented by atomic fractions or chemical formulas, as exemplified by silica SiO_2 and ceramic TiN_x (x represents non-stoichiometry). Polymers are composed of macromolecules by combining several monomers (repeating subunits), such as polyethylene $(\text{C}_2\text{H}_4)_n$ and polyvinyl chloride $(\text{C}_2\text{H}_3\text{Cl})_n$. However, there is a long-standing mystery on the structural origin of the materials' chemistry, which is especially important for metals and inorganic compounds that contain multiple elements, or alloyed materials. Even for polymers, the same question persists: is the monomer the composition unit of the macromolecule? In other words, what are the composition genes of materials? Here, the gene refers to the smallest structural unit that serves as the composition carrier of the material. Different from molecular substances, such as ice composed of H_2O molecules and their connections via relative weak inter-molecular bonding, solid materials are dominated by inter-atomic chemical bonding. In metals and inorganic compounds, inter-molecular bonding is completely missing, and it is impossible to define molecules. The unit cell or atomic motif in crystallography, despite carrying the composition information, cannot be universally accepted as the chemical unit for the lack of structural stability contribution and for being unable to deal with short-range ordering. For example, for a Cu_3Au structure, ordered or disordered alike, the unit cell always contains three Cu and one Au atoms, which is too small in comparison with the short-range ordering that extends to a few atomic shells. In polymers, the basic units within the macromolecules are still to be defined.

The composition genes of materials should represent the local short-range ordering, just as molecules. Specifically, a composition gene should possess typical molecule-like characteristics, such as chemical composition of the entire structure, charge balance (or stable electronic structure), local atomic configuration, and mean atomic density. We developed a new structural model for short-range order structures, the so-called cluster-plus-glue-atom model^[2,3]. This model regards any structure from the viewpoint of a local unit composed of a central atom, its shell (the cluster part), plus a few atoms located at the next-neighbor sites (the glue part). In the following, we show that such local units are exactly the composition genes of materials. Termed chemical units, they are molecule-like and meet nearly all the conditions of molecules.

The method of using nearest-neighbor clusters as the basic structural units to characterize phase structures has a long history. As early as the 1970s, Mackay^[4,5] proposed that nearest-neighbor coordination polyhedral clusters (hereinafter referred to as clusters) can be introduced into the structure description of complex phases, instead of the traditional crystallographic method using space group and atomic position information; thus, the characteristics of the material structures can be more effectively reflected, and even the relationship between the compositions and structures can be established. By the 1990s, it was pointed out that quasicrystals and their corresponding crystalline phases have the same electron concentrations^[6]. Subsequently, a great deal of research on the electron behavior formed the basis for the proposition of the cluster-plus-glue-atom model, mainly destined for amorphous structures^[7-9]. Based on the cluster-plus-glue-atom model, the composition genes of materials are the molecule-like chemical units, expressed as the cluster formula $[\text{cluster}](\text{glue atom})_x$, where x means the number of glue atoms. The cluster part is the local structure with the strongest interactions between atoms, and the existence of glue atoms is to balance the charge neutrality of the structural units and maintain the average atomic density. Therefore, this structural picture contains structural stability information, being different fundamentally from the unit cell concept in conventional crystallography. These composition genes exist in liquids, amorphous solids, solid solutions, and crystalline states. The composition design is made simple once the composition gene is known.

In the following, the molecule-like chemical units defined via the cluster-plus-glue-atom model are first presented, stressing their role as the composition genes in materials. Then, examples of the composition genes in various materials are given. Finally, a new classification method for materials according to the composition genes is proposed.

MOLECULE-LIKE CHEMICAL UNITS FROM SHORT-RANGE ORDERS

In a classical particle system composed of a large number of atoms, the most obvious manifestation of ordering is positional order, which means that the positions of atoms in different places are related^[10]. If there is no correlation at all, the distribution of atomic positions is completely random, that is, the system is in a completely disordered state, such as the ideal gas; if the correlation range of atomic positions is limited to neighboring atoms, the system exhibits short-range order (SRO); and, if the correlation range reaches infinity, the system shows long-range order (LRO)^[11]. Many materials we commonly use today are characterized by locally disordered states, or SROs, such as those based on solid solutions (e.g., steels) or glasses (e.g., amorphous silicates), in which local symmetry is partially destroyed. The traditional crystallography is no longer applicable in dealing with such locally disordered states. There is an urgent need to establish a theoretical model to describe SRO structures and their composition formulas, in order to define the composition genes hidden in them.

To describe the phenomenon of ordering in alloys, various theoretical approaches have been proposed. Bragg and Williams^[12] introduced the LRO parameter, S , and devised a simple theory giving S as a function of temperature. Subsequently, Bethe^[13] proposed the “order of neighbors” σ to express the difference of the probabilities of finding an unequal and an equal neighbor beside a given atom and determined the long-range and nearest-neighbor orders for the case of AB alloys. Later, Peierls^[14] extended this theory to the case of the face-centered cubic (FCC) A_3B alloys such as Cu_3Au , and the results better accord with experiments than did the theory of Bragg and Williams. Similarly, the theory of Kirkwood^[15] describes the order and disorder in solid solutions based on a direct evaluation of the crystalline partition function. Profiting from the advances in diffraction techniques, the quantitative determination of LRO and SRO parameters becomes possible. Cowley^[16] measured the SRO parameters for the first ten shells of neighbors in Cu_3Au using the single-crystal diffuse scattering technique. Furthermore, he proposed the SRO parameters α_i to express the interaction of a given atom in an alloy with the atoms at the i th shell surrounding it, which is still widely used today^[17-19].

The physical root of ordering in classical particle systems lies in the interaction between particles, while this interaction itself is a quantum mechanics problem. With the development of quantum mechanics, people began to pay more attention to the electron fluctuation in the process of studying the structures of metals and alloys. By the 1950s, the advent of Friedel oscillation theory^[20-22] had revolutionized the understanding of the structure of matter. This theory is derived from Friedel's early research on the distribution of electrons around impurities in monovalent metals^[20]. Subsequently, Friedel^[21] extended his research into the electronic structure of primary solid solutions in metals and proposed that it is simply a question of solving a Schrödinger equation. He specifically considered the charge screening phenomenon generated by the introduction of small and strong perturbations in a uniform potential field and described the collective behavior of conduction electrons in metallic alloys^[22]. Due to this collective behavior, when an impurity charge is introduced into the uniform potential field, the electron cloud around the impurity charge is polarized, thereby shielding the disturbance of the impurity charge to the whole system. However, this shielding is incomplete. The disturbance in the short range is not completely shielded, and the electron density around the impurity shows an oscillating distribution. This phenomenon, similar to water waves, is known as Friedel oscillations.

Following Friedel, Langer and Vosko^[23], Heine and Weaire^[24], Harrison^[25], and Ziman^[26] also conducted specific studies on the shielding potential energy generated by the introduction of impurity charges, and the calculation results were mutually verified, that is, the effective pair potential $\phi_{\text{eff}}(r)$ at distance r is proportional to a damped cosine-function: $\phi_{\text{eff}}(r) \propto \cos(2k_F \cdot r + \theta)/r^3$. Here, θ signifies the phase shift angle changing from zero to $\pi/2$ and is related to the scattering amplitude at the Fermi level^[24]. k_F means Fermi vector. For liquid and amorphous states, at short and medium distances, Häussler^[27] experimentally

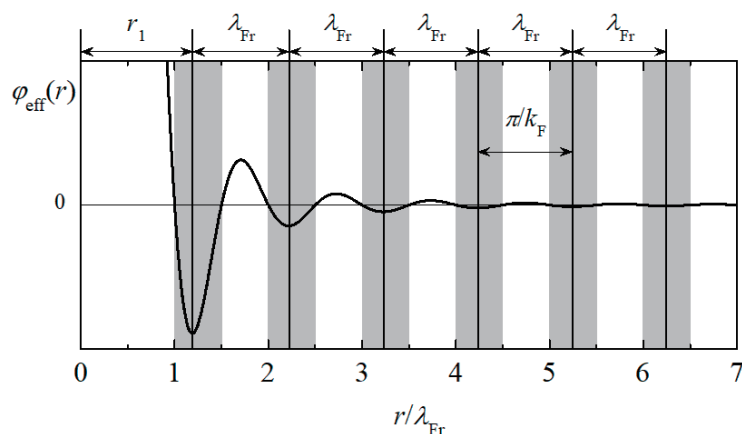


Figure 1. Effective potential of Friedel oscillations $\phi_{\text{eff}}(r) \propto -\sin(2k_{\text{F}}r)/r^3$, with the negative potential zones being shaded, where atoms tend to gather.

verified that the phase shift θ is equal to $\pi/2$, and Kroha *et al.*^[28] also obtained $\theta = \pi/2$ through theoretical calculations. Therefore, the expression of the effective potential of Friedel oscillations is written as $\phi_{\text{eff}}(r) \propto -\sin(2k_{\text{F}}r)/r^3$, as shown in Figure 1^[29]. To reach a stable state, atoms tend to be located near the mid-points of the negative potential zones in Friedel oscillations, at the distances of $r_n = (1/4 + n)\lambda_{\text{Fr}}$, where $\lambda_{\text{Fr}} = \pi/k_{\text{F}}$ is the Friedel wavelength and $n = 1, 2, 3, \dots$ is the shell sequence, so as to minimize the total energy of the system^[29]. This arrangement sequence of atoms with the inter-shell spacing of Friedel wavelength is the so-called spherical-periodic order, which occurs due to the resonance between the electronic and static atomic structures.

The compositions of materials come from chemical SRO, which is related to Friedel oscillations. According to the Friedel-oscillation-based spherical-periodic resonance theory, the molecule-like chemical unit covering only the nearest-neighbor cluster plus a few next-neighbor glue atoms can be rationalized^[3]. Since the function forms of the charge distribution and atomic density distribution are consistent with the effective potential function, it is easy to obtain the charge-neutral and mean-density radial distances by integrating the function $-\sin(2k_{\text{F}}r)/r^3$. The first of such positions falls near $1.76\lambda_{\text{Fr}}$, close to the mid-point $1.75\lambda_{\text{Fr}}$ in the first positive potential zone of Friedel oscillations^[30], defining the size of the corresponding composition gene. Generally, the cluster part can be derived from the homologous crystalline phase structure, and the glue atom part can be further calculated according to the chemical unit as well as the actual composition, from which the composition gene can be ultimately determined. Figure 2 exhibits a typical cluster-plus-glue-atom composition gene relevant to FCC structure, whose cluster configuration is a cuboctahedron, and a few glue atoms are located outside the cluster.

For binary FCC-based solid solution alloys, the coordination number of the cluster is 12, and the glue-atom shell in the next neighborhood contains 1-5 atoms. Thus, the chemical unit of a binary system can be expressed as $[A-B_{12}]A_xB_y$, containing 14-18 atoms, where the integer $x + y$ represents the number of glue atoms, $0 < x + y < 6$. It is assumed that the volume of each element remains unchanged after mixing^[31], so the chemical unit volume is the sum of all atomic volumes, i.e., $(1+x) \cdot (4\pi/3) \cdot R_A^3/0.74 + (12+y) \cdot (4\pi/3) \cdot R_B^3/0.74$, where R 's are atomic radii and 0.74 is the packing efficiency of FCC structure; it is also equal to the spherical volume enclosed by the cut-off distance $1.76\lambda_{\text{Fr}}$, i.e., $(4\pi/3) \cdot (1.76\lambda_{\text{Fr}})^3 = (4\pi/3) \cdot [1.76 \times (R_A + R_B)/1.25]^3$, where $R_A + R_B = 1.25\lambda_{\text{Fr}}$ is the nearest-neighbor distance in the cluster according to the spherical periodicity theory. This equation leads to $(1+x) \cdot R_{A/B}^3 + (12+y) \approx 2 \cdot (R_{A/B} + 1)^3$, with $R_{A/B}$ meaning the ratio of R_A over R_B . Goldschmidt radii of atoms are generally adopted. According to this relationship, for a given binary FCC-based solid solution alloy of known $R_{A/B}$, the chemical unit can be calculated. For instance, when $R_A = R_B$,

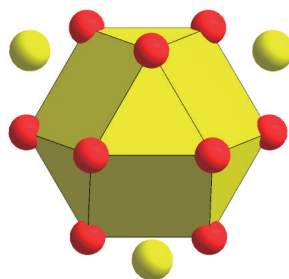


Figure 2. Configuration of the cluster-plus-glue-atom composition gene relevant to FCC structure, with the cuboctahedron representing the nearest-neighbor cluster and the scattered yellow atoms representing the glue atoms located at the next neighbors.

$x + y = 3$, which means that, for an FCC solid solution alloy composed of solute and solvent atoms with equal atomic radii, which is equivalent to a single-element FCC structure or a completely disordered FCC structure composed of one average atom, the composition gene consists of 16 atoms, written as $[A-B_{12}]_3(A,B)_3$ (as exemplified by austenitic stainless steels 304 and 316). When $R_A \neq R_B$, the value of $x + y$ can be calculated accordingly, as exemplified in Cu-Zn alloys. To allow more A in solution of B, the general formula of the composition genes in FCC solid solutions should be $[A-(A,B)_{12}]_x A_x B_y$, where A and B are mixed in the nearest-neighbor shell and can be determined by Cowley's α short-range parameter. The calculation can also be done following similar equations by regarding the mixed A and B atoms in the shell site as an average atom. This method can also be extended to solid solutions based on body-centered cubic and hexagonal close-packed structures.

In addition to solid solution alloys, the composition genes of other materials can also be identified, such as amorphous alloys, inorganic compounds, glasses, polymers, *etc.* First, the clusters are derived from relevant phases of known structures (for amorphous alloys and glasses, they are devitrification phases); then, combined with the cluster-resonance model^[32,33], the electronic factor is introduced to address the structural stability (the electron number per unit e/u being a multiple of 8, conforming to the octet rule), thus to determine the glue atoms. In the calculation process of the composition genes, the chemical compositions, spatial configurations, charge neutrality, and average atomic densities are taken into account.

COMPOSITION GENES IN MATERIALS

In this section, the molecule-like chemical units as the composition genes in various materials, covering metallic alloys, inorganic compounds and relevant glasses, and polymers, are examined via examples.

Metallic alloys

Metallic alloys generally consist of solid-solution-based industrial alloys and metallic glasses.

Solid-solution-based industrial alloys

Austenitic stainless steels are the most widely used type of stainless steels, most of which are derived from 304 grade, or 06Cr19Ni10, whose composition (using weight percentage, wt.%) is $C \leq 0.08$, $Si \leq 1.00$, $Mn \leq 2.00$, $P \leq 0.045$, $S \leq 0.030$, $Cr: 18.0-20.0$, $Ni: 8.0-11.0$ ^[34]. Due to the austenite structure, the cluster is a cuboctahedron with the coordination number of 12 [Figure 2]. The elements in substitutional solution of austenite have approximately equal atomic radii, so the number of glue atoms can be determined as 3 by using the calculation method about composition genes mentioned above. Thus, the composition gene of 304 stainless steel can be determined to be $(Cr,Si)_{3-3.5}-(Ni,Mn)_{1.25-1.75}-Fe_{10.75-11.75}$, which covers the composition zones as specified in different industrial standards. This formula contains at least three Cr atoms, which signifies that Cr_3 is the minimum quantity required to guarantee to passivate the steel. Similarly, the composition gene of 316 stainless steel is $(Cr,Mo,Si)_{3-3.5}-(Ni,Mn)_{1.75-2.25}-Fe_{10.25-11.25}$. The major

difference from that of 304 lies in higher Ni content and hence lower C content, to minimize sensitization (the tendency of Cr-rich carbide precipitation that deteriorates corrosion resistance). The complex chemistries of stainless steels are now made clear and concise. We can also see that these steels have large composition tolerance, as each steel actually corresponds to two cluster formulas, which is a significant advantage of steels. This is in clear contrast to Ni-based superalloys whose formulas are usually unique, which makes their preparation intrinsically difficult, as illustrated next.

Superalloys are used to withstand extreme high-temperature service conditions and are typical multi-element complex industrial alloys. DD406 (DD6) belongs to the second generation of Ni-based superalloys developed in China, which is comparable to PWA1484, Rene N5, and CMSX-4. It is also cost-effective for the low Re content. This alloy has a moderate creep rupture strength to suit for service temperature below 1100 °C. Its chemical composition is quite complex: 0.001-0.040C, 3.80-4.80Cr, bal. Ni, 8.50-9.50Co, 7.00-9.00W, 1.50-2.50Mo, 5.20-6.20Al, Ti ≤ 0.10, Fe ≤ 0.20, 0.60-1.20Nb, 6.00-8.50Ta, 1.60-2.40Re, 0.05-0.15Hf, B ≤ 0.02, Zr ≤ 0.10, Mn ≤ 0.15, Si ≤ 0.20, P ≤ 0.018, S ≤ 0.001, Cu ≤ 0.05 (trace elements neglected). The mid-values of substitutional elements are taken, and the composition becomes Ni-5.7Al-0.9Nb-7.25Ta-4.3Cr-2Mo-8W-2Re-9Co, which is transformed into a 16-atom chemical formula $Al_{2.10}Nb_{0.10}Ta_{0.40}Cr_{0.82}Mo_{0.21}W_{0.43}Re_{0.11}Co_{1.52}Ni_{10.31}$. The elements are grouped into the precipitation-type (Al,Nb,Ta), the solution-type (Cr,Mo,W,Re), and the solvent-type (Co,Ni), i.e., $(Al_{0.81}Nb_{0.04}Ta_{0.15})_{2.60}-(Cr_{0.52}Mo_{0.13}W_{0.30}Re_{0.07})_{1.57}-(Co_{0.13}Ni_{0.87})_{11.8}$, which falls close to $(Al_{0.81}Nb_{0.04}Ta_{0.15})_{2.5}-(Cr_{0.52}Mo_{0.13}W_{0.30}Re_{0.07})_{1.5}-(Co_{0.13}Ni_{0.87})_{12}$. A tentative cluster formula is thus obtained: $[(Al_{0.81}Nb_{0.04}Ta_{0.15})-(Co_{0.13}Ni_{0.87})_{12}](Al_{0.81}Nb_{0.04}Ta_{0.15})_{1.5}(Cr_{0.52}Mo_{0.13}W_{0.30}Re_{0.07})_{1.5}$. Actually, this is the general formula for all single-crystal superalloys. For example, PWA1484 satisfies 2.56-1.67-11.78, Rene N5 2.61-1.94-11.44 (where apparently there is excessive solution content), and CMSX-4 2.56-1.72-11.72. By comparing these formulas, it is straightforward to say that DD6 best satisfies the modeled formula. These composition genes formulated by the cluster formulas simplify the composition analysis of superalloys and can be used as guiding tools for the development of new superalloys.

More examples have been analyzed, for instance, maraging stainless steel Custom-465 is formulated as $[Ni-Fe_{12}]Cr_2(Ni,Mo,Ti,Nb,Al,V)^{[35]}$, high-entropy Al-Cr-Fe-Co-Ni alloys as $[Al-(Cr,Fe,Co,Ni)_4]Al^{[36,37]}$, the most widely used brass Cu-30.0Zn as $[Zn-Cu_{12}]Zn_4^{[38]}$, and the most popular Ti alloy Ti-6Al-4V as a dual-gene formula $12[Al-Ti_{12}]AlTi_2+5[Al-Ti_{14}]V_2Ti \approx Ti-6.05Al-3.94V^{[39]}$.

Metallic glasses

Metallic glasses are characterized by short-range ordering. Good glass-formers conform to well-specified compositions. The steps of composition analysis and alloy design based on the cluster-plus-glue-atom model have been summarized^[40,41], and the resulting cluster formulas are exactly the composition genes. Here, the metallic glasses in Cu-Zr and Ni-(Nb,Ta) systems are taken as examples to illustrate the calculations of composition genes, and the calculated results are listed in [Table 1](#). The three Cu-Zr bulk metallic glass compositions with high glass forming ability are $Cu_{64}Zr_{36}$, $Cu_{56}Zr_{44}$, and $Cu_{50}Zr_{50}$ ^[42]. To interpret the composition of $Cu_{64}Zr_{36}$, the deep eutectic point $Cu_{0.618}Zr_{0.382}$ is selected according to the phase diagram^[43]. Then, the crystalline phase Cu_8Zr_3 is determined to be the devitrification phase due to the structural homology^[44,45], which generates the cluster $[Cu-Cu_7Zr_5]$ with the configuration of icosahedron. Finally, one Cu atom is added as the glue atom, so the cluster formula $[Cu-Cu_7Zr_5]Cu \approx Cu_{64.3}Zr_{35.7}$ is obtained, with $e/u \approx 24$. For the bulk metallic glass $Cu_{56}Zr_{44}$ (exactly the eutectic point), two eutectic phases, $Cu_{10}Zr_7$ and CuZr, are related. Thus, the dual-cluster formulization is introduced to decipher this alloy^[46], and its composition gene is calculated to be $[Zr-Cu_{10}Zr_4+Cu-Zr_8Cu_6]Cu_2Zr_2 \approx Cu_{55.9}Zr_{44.1}$. Half of this dual-cluster formula readily interprets the bulk metallic glass $Cu_{50}Zr_{50}$: $[Cu-Zr_8Cu_6]Cu = Cu_{50}Zr_{50}$. Similarly, the composition genes of the bulk metallic glasses $Ni_{62}Nb_{38}$ and $Ni_{59-62}Ta_{38-41}$ are $[Ni-Nb_4Ni_8]Nb_2Ni = Ni_{62.5}Nb_{37.5}$ and $[(Ni,Ta)-Ni_6Ta_6]Ni_3 = Ni_{56.25-62.5}Ta_{37.5-43.75}$, respectively^[41], where the clusters are derived from eutectic phases and the electron numbers per unit fall close to 24.

Table 1. Composition genes of typical bulk metallic glasses in Cu-Zr system and Ni-(Nb,Ta) systems

| Bulk metallic glasses | Crystalline phases | Cluster formulas of composition genes | Calculated compositions |
|---|--|---|---|
| Cu ₆₄ Zr ₃₆ | Cu ₈ Zr ₃ | [Cu-Cu ₇ Zr ₅]Cu | Cu _{64.3} Zr _{35.7} |
| Cu ₅₆ Zr ₄₄ | Cu ₁₀ Zr ₇ +CuZr | [Zr-Cu ₁₀ Zr ₄ +Cu-Zr ₈ Cu ₆]Cu ₂ Zr ₂ | Cr _{55.9} Zr _{44.1} |
| Cu ₅₀ Zr ₅₀ | CuZr | [Cu-Zr ₈ Cu ₆]Cu | Cu ₅₀ Zr ₅₀ |
| Ni ₆₂ Nb ₃₈ | NbNi ₃ | [Ni-Nb ₄ Ni ₈]Nb ₂ Ni | Ni _{62.5} Nb _{37.5} |
| Ni ₅₉₋₆₂ Ta ₃₈₋₄₁ | NiTa | [(Ni,Ta)-Ni ₆ Ta ₆]Ni ₃ | Ni _{56.25-62.5} Ta _{37.5-43.75} |

The cluster formula approach can also deal with dual-phase glass-crystal alloys. For example, Mg₄₉Cu₄₂Y₉ is composed of an amorphous phase and a nanocrystal phase and can be deciphered into a composition gene [Mg-Cu₄Mg₈Y₃]Cu₂Mg+[Cu-Cu₆Mg₆]Cu ≈ Mg_{48.5}Cu_{42.4}Y_{9.1}^[47]. Binary eutectics can be dissociated into two stable liquids so that their compositions are formulated by two subunits, i.e., dual-genes, as exemplified by [Al-Co₁₄+Al-Co₁₂]Al₄ = Al_{18.75}Co_{81.25}, [Cr-Co₁₂+Co-Cr₁₄]Co₆ ≈ Co_{55.9}Cr_{44.1}, [Hg-Mg₁₂+Hg-Mg₁₁]Hg₃Mg ≈ Mg_{82.8}Hg_{17.2}, and [Sb-Mg₁₂+Sb-Mg₇]Mg₅Sb ≈ Mg_{88.9}Sb_{11.1}^[48]. Besides, the multi-element eutectic-type bulk metallic glasses are always derived from binary eutectic alloys; thus, their composition genes usually stem from multiple phases^[49].

Inorganic compounds and glasses

Inorganic compounds

A compound is a pure substance composed of two or more different elements, usually in simple proportions. Compounds are generally classified as stoichiometric and non-stoichiometric ones. In this part, SiO₂ and TiN_x are chosen as typical examples to unveil the composition genes hidden in inorganic compounds.

Silica SiO₂ is the most popular inorganic material. The outer electron orbital of Si is 3s²3p², so that this atom provides four valence electrons to form chemical bonding. The atom O has an outer electron orbital of 2s²2p⁴, giving six valence electrons. Therefore, one Si atom attracts four O atoms through four covalent sp³-type bonds, forming a tetrahedron with the Si atom as the center and O atoms as the nearest neighbors, [Si-O₄]. However, [Si-O₄] itself does not satisfy charge balance, and four more electrons are needed to reach the octet state. The [Si-O₄] cluster should be linked to one more Si atom as the glue atom to obtain enough electrons, thus maintaining the charge neutrality. Therefore, the molecule-like chemical unit as the composition gene in SiO₂ is determined to be [Si-O₄]Si, with e/u = 32^[50], as shown in Figure 3. Actually, this is also the composition gene of all silicate glasses. Obviously, the composition gene [Si-O₄]Si is different from the chemical formula SiO₂.

The compounds TiN_x are typical for non-stoichiometry. The structure can be viewed as N atoms in interstitial solution in FCC Ti lattice, forming the nearest-neighbor coordination octahedral cluster [N-Ti₆]. The composition gene for the stoichiometric TiN compound is [N-Ti₆]N₅. However, theoretically, it is revealed that the hardest (or the most stable) compound is TiN_{0.8}^[50], which corresponds to a Ni-deficient formula [N-Ti₆]N_{3.8}, with e/u = 48.

Inorganic glasses

The advent of silicate glasses is traced back to Egyptian times, when people accidentally found that silica mixed with soda easily forms glasses. Soda-lime silicate glasses remain the most widely used category for their easy production and low cost, occupying about 90% of the glass market. However, as with industrial metals, their compositions are developed out of tedious trial-and-error attempts. The silicate glass compositions should be hidden in the amorphous structure, of which the modeling is a big challenge for the scientific community even today. The widely recognized random network model of silicate glasses was proposed by Zachariasen^[51] in 1932. The major assumption is that the coordination number centered by a

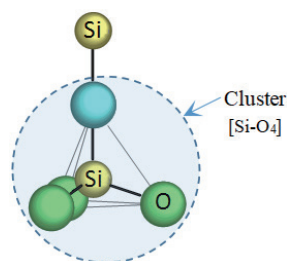


Figure 3. Configuration of the composition gene $[\text{Si-O}_4]\text{Si}$ in SiO_2 , where the cluster part is outlined.

metal cation ranges from 3 to 4, which is equivalent to the average valency of cations. This model does not contain any chemical information so cannot be used for composition interpretation. We here re-interpret this model using our cluster-plus-glue-atom model as follows: a silicate glass is formed by mixing four-valency unit $[\text{Si-O}_4]\text{Si}$ and three-valency unit $[\text{M-O}_3]\text{M}$ (M represents the average cations in glasses) to develop into a three-dimensional random network with an average coordination number ranging from 3 to 4. Since the addition of oxides is for defined purposes, it should be maximized so long as the glass remains stable, i.e., the proportion of $[\text{M-O}_3]\text{M}$ is generally high so that the final average cation valence falls close to three. For instance, the most common composition of soda-lime float glasses for windows is $72\text{SiO}_2-15\text{Na}_2\text{O}-9\text{CaO}$, whose average cation valence is 3.04. Therefore, the composition gene of this glass is formulated basically by the three-valence unit $[\text{M-O}_3]\text{M}$, where M is $\text{Na}_{0.26}\text{Ca}_{0.09}\text{Si}_{0.65}$. It is also noted that the commercial glasses all have cation valences slightly above three; this is to facilitate industrial-scale production control, as any composition with a cation valence below three would suffer from low structural stability and thus from easy crystallization.

Cr-doped amorphous carbon has a diamond-like structure and is widely used as protective coatings. Diamond-like amorphous carbon is mainly composed of sp^2 and sp^3 bondings. Cr-doping is generally adopted to increase sp^2 content and hence the electrical conductivity. However, the optimal Cr-doping content is not known. According to the cluster-plus-glue-atom model, the composition gene of stable Cr-doped diamond-like carbon is $[\text{Cr-C}_6]\text{CrC}$ or $[\text{Cr-C}_6]\text{Cr}_3^{[50]}$, which provides a theoretical direction for the coating industry.

Polymers

Polymers are basically molecular materials so that in principle their compositions are not of major concern for the researchers: the materials' chemistry is identical to that of the macromolecules composing the polymers. For example, polyethylene, the most common plastic in use today, is made of monomer ethylene C_2H_4 connected in chain, so that the composition of polyethylene is also that of the monomer. However, the macromolecules are so huge that there is a need to further identify the basic units, which is similar to the identification of chemical units in well-ordered inorganic compounds. Here, polyethylene is taken as the example to define the composition gene in a long-chained macromolecule. Polymerization of ethylene to polyethylene is described by $n \text{CH}_2 = \text{CH}_2$ (gas) \rightarrow $[-\text{CH}_2-\text{CH}_2-]_n$ (solid). From this equation, one may have the impression that the repeating unit is a pair of methylene groups. However, in terms of the cluster-plus-glue-atom model, the composition gene of this long chain $(\text{C}_2\text{H}_4)_n$ is $[\text{C-H}_2\text{C}_2]\text{H}_4 = \text{C}_3\text{H}_6$, with $e/u = 18$. Likewise, the macromolecules of polypropylene $(\text{C}_3\text{H}_6)_n$ and polyvinyl chloride $(\text{C}_2\text{H}_3\text{Cl})_n$, the world's second- and third-most widely produced synthetic plastic polymers, respectively, are deciphered into the composition genes $[\text{C-H}_1\text{C}_3]\text{H}_7 = \text{C}_4\text{H}_8$ and $[\text{C-H}_1\text{Cl}_1\text{C}_2]\text{H}_4 = \text{C}_3\text{H}_5\text{Cl}_1$, both having $e/u = 24$.

In terms of composition genes, materials can be simply classified into single-, dual-, and multi-gene types. Single-gene materials refer to the materials consisting of only one composition gene, such as single-crystal Ni-based superalloys, most binary bulk metallic glasses, single-phase inorganic compounds, and single-

monomer polymers. Dual-gene materials are the ones composed of two material genes, including eutectic alloys, most industrial alloys, amorphous-crystalline dual-phase alloys, *etc.* Multi-gene materials are made up of three or more composition genes, such as glasses that are mixtures of several units and multi-monomer polymers.

The concept of composition gene proposed here can guide material design, as the genes directly provide the composition formulas, which can greatly improve the efficiency of material research and development. The applications of such an approach in alloy optimization and development has been exemplified in a number of alloys (cf. Ti alloys^[30,39], maraging stainless steels^[35], high-entropy alloys^[36,37], Cu alloys^[38], metallic glasses^[40,41,47], eutectics^[46,48]).

CONCLUSIONS

In summary, molecule-like chemical units as composition genes of materials are proposed based on the cluster-plus-glue-atom model, reflecting the chemistries, characteristic short-range ordering, electronic structure stability, and overall atomic densities of the materials. The compositions of various materials, including metallic alloys, inorganic compounds, and polymers, are analyzed and deciphered into different kinds of genes in terms of cluster formulas, expressed as the nearest-neighbor clusters plus a few next-neighbor glue atoms. Materials are thus classified into single-, dual-, and multi-gene types.

DECLARATIONS

Authors' contributions

Conducted the calculations, analyzed the data, and wrote the paper: Zhang S

Participated in the discussions about the cluster models: Wang Q

Conceived the ideas, proposed the cluster model theory, and revised the manuscript: Dong C

Availability of data and materials

Not applicable.

Financial support and sponsorship

This work was supported by the Key Discipline and Major Project of Dalian Science and Technology Innovation Foundation (No. 2020JJ25CY004) and by the Foundation of Liaoning Province Education Administration (No. JDL2019023).

Conflicts of interest

All authors declared that there are no conflicts of interest.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Copyright

© The Author(s) 2021.

REFERENCES

1. Kittel C. Introduction to solid state physics. 8th ed. Hoboken: John Wiley & Sons; 2005.
2. Dong C, Wang Q, Qiang JB, et al. From clusters to phase diagrams: composition rules of quasicrystals and bulk metallic glasses. *J Phys D: Appl Phys* 2007;40:R273-91.

3. Dong C, Wang Z, Zhang S, Wang Y. Review of structural models for the compositional interpretation of metallic glasses. *Int Mater Rev* 2020;65:286-96.
4. Mackay AL, Finney JL. Structuration. *J Appl Crystallogr* 1973;6:284-9.
5. Mackay AL, Finney JL, Gotoh K. The closest packing of equal spheres on a spherical surface. *Acta Cryst A* 1977;33:98-100.
6. Dong C, Perrot A, Dubois JM, Belin E. Hume-rothery phases with constant e/a value and their related electronic properties in Al-Cu-Fe(-Cr) quasicrystalline systems. *Mater Sci Forum* 1994;150-151:403-16.
7. Qiang J, Wang D, Bao C, et al. Formation rule for Al-based ternary quasi-crystals: Example of Al-Ni-Fe decagonal phase. *J Mater Res* 2001;16:2653-60.
8. Wang YM, Qiang JB, Wong CH, Shek CH, Dong C. Composition rule of bulk metallic glasses and quasicrystals using electron concentration criterion. *J Mater Res* 2003;18:642-8.
9. Wang Y. The e/a factor governing the formation and stability of $(Zr_{76}Ni_{24})_{1-x}Al_x$ bulk metallic glasses. *Scripta Materialia* 2003;48:1525-9.
10. Feng D, Jin G. Condensed state physics. Beijing: High Education Press; 2012.
11. Ziman JM. Models of disorder - the theoretical physics of homogeneously disordered systems. Cambridge: Cambridge University Press; 1979.
12. Bragg WL, Williams EJ. The effect of thermal agitation on atomic arrangement in alloys. *Proc R Soc Lond A* 1934;145:699-730.
13. Bethe HA. Statistical theory of superlattices. *Proc R Soc Lond A* 1935;150:552-75.
14. Peierls R. Statistical theory of superlattices with unequal concentrations of the components. *Proc R Soc Lond A* 1936;154:207-22.
15. Kirkwood JG. Order and disorder in binary solid solutions. *J Chem Phys* 1938;6:70-5.
16. Cowley JM. X-ray measurement of order in single crystals of Cu_3Au . *J Appl Phys* 1950;21:24-30.
17. Cowley JM. An approximate theory of order in alloys. *Phys Rev* 1950;77:669-75.
18. Cowley JM. Short- and long-range order parameters in disordered solid solutions. *Phys Rev* 1960;120:1648-57.
19. Cowley JM. Short-range order and long-range order parameters. *Phys Rev* 1965;138:A1384-9.
20. Friedel J. XIV. The distribution of electrons round impurities in monovalent metals. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 2010;43:153-89.
21. Friedel J. Electronic structure of primary solid solutions in metals. *Adv Phys* 1954;3:446-507.
22. Friedel J. Metallic alloys. *Nuovo Cim* 1958;7:287-311.
23. Langer J, Vosko S. The shielding of a fixed charge in a high-density electron gas. *J Phys Condens Matter* 1960;12:196-205.
24. Heine V, Weaire D. Pseudopotential theory of cohesion and structure. Elsevier; 1970. p. 249-463.
25. Harrison WA. Solid state theory. New York: McGraw-Hill, Inc.; 1970.
26. Ziman JM. Principles of the theory of solids. Cambridge: Cambridge university press; 1972.
27. Häussler P. A new hume-rothery phase with an amorphous structure in noble-metal/simple-metal alloys. *J Phys Colloques* 1985;46:C8-361-5.
28. Kroha J, Huck A, Kopp T. Coulomb interaction and disorder at $q=2kF$: a novel instability of the fermi sea and implications for amorphous alloys. *Phys Rev Lett* 1995;75:4278-81.
29. Häussler P. Interrelations between atomic and electronic structures-Liquid and amorphous metals as model systems. *Physics Reports* 1992;222:65-143.
30. Jiang B, Wang Q, Dong C, Liaw PK. Exploration of phase structure evolution induced by alloying elements in Ti alloys via a chemical-short-range-order cluster model. *Sci Rep* 2019;9:3404.
31. Senkov O, Miracle D, Keppens V, Liaw P. Development and characterization of low-density ca-based bulk metallic glasses: an overview. *Metall Mater Trans A* 2008;39:1888-900.
32. Han G, Qiang J, Li F, et al. The e/a values of ideal metallic glasses in relation to cluster formulae. *Acta Materialia* 2011;59:5917-23.
33. Luo L, Chen H, Wang Y, et al. 24 electron cluster formulas as the 'molecular' units of ideal metallic glasses. *Philos Mag* 2014;94:2520-40.
34. Stainless Steel Branch of China Special Steel Enterprises Association. Stainless steel practical manual. Beijing: China Science and Technology Press; 2003.
35. Li Z, Zhang R, Zha Q, Wang Y, Qiang J, Dong C. Composition design of superhigh strength maraging stainless steels using a cluster model. *Progress in Natural Science: Materials International* 2014;24:35-41.
36. Wang Q, Ma Y, Jiang B, et al. A cuboidal B2 nanoprecipitation-enhanced body-centered-cubic alloy $Al_{10}CoCrFe_2Ni$ with prominent tensile properties. *Scripta Materialia* 2016;120:85-9.
37. Ma Y, Wang Q, Jiang B, et al. Controlled formation of coherent cuboidal nanoprecipitates in body-centered cubic high-entropy alloys based on $Al_2(Ni,Co,Fe,Cr)_{14}$ compositions. *Acta Mater* 2018;147:213-25.
38. Hong HL, Wang Q, Dong C, Liaw PK. Understanding the Cu-Zn brass alloys using a short-range-order cluster model: significance of specific compositions of industrial alloys. *Sci Rep* 2014;4:7065.
39. Liu T, Zhang S, Wang Q, Min X, Dong C. Composition formulas of Ti alloys derived by interpreting Ti-6Al-4V. *Sci China Technol Sci* 2021;64:1732-40.
40. Wang ZR, Qiang JB, Wang YM, et al. Composition design procedures of Ti-based bulk metallic glasses using the cluster-plus-glue-atom model. *Acta Mater* 2016;111:366-76.
41. Zhang S, Dong D, Wang Z, Dong C, Häussler P. Composition formulas of Ni-(Nb, Ta) bulk metallic glasses. *Intermetallics* 2017;85:176-9.

42. Li Y, Guo Q, Kalb JA, Thompson CV. Matching glass-forming ability with the density of the amorphous phase. *Science* 2008;322:1816-9.
43. Massalski TB, Okamoto H, Subramanian PR, et al. Binary alloy phase diagrams. 2nd ed. USA: ASM International; 1990.
44. Zhang S, Dong D, Wang Z, Dong C, Häussler P. Spherical periodicity as structural homology of crystalline and amorphous states. *Sci China Mater* 2018;61:409-16.
45. Zhang S, Dong C, Häussler P. Structural relationship between crystalline and amorphous states in Cu-(Zr, Ti) binary systems. *J Phys Condens Matter* 2021;33:074001.
46. Ma YP, Dong DD, Dong C, et al. Composition formulas of binary eutectics. *Sci Rep* 2015;5:17880.
47. Wang Z, Chen R, Qiang J, Zhang S, Zhao Y. Cluster-based composition interpretation of dual-phase glass-crystal alloys via the example of Mg₄₉Cu₄₂Y₉. *J Non Cryst Solids* 2021;566:120886.
48. Zhang S, Dong C. Dual-cluster interpretation of binary eutectics associated with hexagonal close-packed solid solution phases. *Mater Lett* 2018;233:71-3.
49. Han K, Wang Y, Qiang J, et al. Dual-cluster formulas for eutectic-type bulk metallic glasses and experimental verification in Zr-Al-Fe-Cu system. *Materials & Design* 2019;183:108142.
50. Zhang S, Dong C, Ma Y. et al. Materials characteristics of thin films. *VACUUM* 2020;57:11-8. (in Chinese)
51. Zachariasen WH. The atomic arrangement in glass. *J Am Chem Soc* 1932;54:3841-51.