



Quantifying multipoint ordering in alloys

James M. Goff ^{1,*}, Bryant Y. Li ¹, Susan B. Sinnott,^{1,2,3} and Ismaila Dabo^{1,3,4}

¹*Department of Materials Science and Engineering, The Pennsylvania State University, University Park, Pennsylvania 16802, USA*

²*Department of Chemistry, The Pennsylvania State University, University Park, Pennsylvania 16802, USA*

³*Materials Research Institute, The Pennsylvania State University, University Park, Pennsylvania 16802, USA*

⁴*Penn State Institutes of Energy and the Environment, The Pennsylvania State University, University Park, Pennsylvania 16802, USA*



(Received 10 February 2021; revised 8 June 2021; accepted 6 July 2021; published 9 August 2021)

A central problem in multicomponent lattice systems is to systematically quantify multipoint ordering. Ordering in such systems is often described in terms of pairs, even though this is not sufficient when three-point and higher-order interactions are included in the Hamiltonian. Current models and parameters for multipoint ordering are often only applicable for very specific cases or require approximating a subset of correlated occupational variables on a lattice as being uncorrelated. In this paper, cluster order parameters are introduced to systematically quantify arbitrary multipoint ordering motifs in substitutional systems through direct calculations of normalized cluster probabilities. These parameters can describe multipoint chemical ordering in crystal systems with multiple sublattices, multiple components, and systems with reduced symmetry. These are defined in this paper and applied to quantify four-point chemical ordering motifs in platinum/palladium alloy nanoparticles that are of practical interest to the synthesis of catalytic nanocages. Impacts of chemical ordering on nanocage stability are discussed. It is demonstrated that approximating four-point probabilities from superpositions of lower-order pair probabilities is not sufficient in cases where three- and four-body terms are included in the energy expression. Conclusions about the formation mechanisms of nanocages may change significantly when using common pair approximations.

DOI: [10.1103/PhysRevB.104.054109](https://doi.org/10.1103/PhysRevB.104.054109)

I. INTRODUCTION

Chemical ordering in alloys and multicomponent crystal systems strongly influences the properties of materials such as mechanical strength, durability, and thermodynamic stability. This includes both the long-range periodic arrangement of elements in ordered alloys such as Cu₃Au and the short-range order (SRO) that occurs in solid-solution crystal systems [1]. In alloys, SRO can influence thermodynamic stability as well as mechanical properties; increased SRO in CrCoNi alloys leads to increased hardness [2,3]. In semiconductors, the optical and electronic properties are affected by chemical ordering [4]. At the solid-solution interfaces of alloy catalysts, the adsorption of solution species is correlated with alloy ordering, and this influences the electrochemical response [5]. In the case of platinum-based alloy nanoshells applied in hydrogen fuel cells as oxygen reduction catalysts, the chemical SRO and structure of the surface alloy have a strong influence on catalyst durability [6]. In these catalytic surface alloys and in many cases, such as high-entropy oxides, semiconductor crystals with multiple sublattices, and other catalyst systems, the chemical ordering motifs of interest are composed of multiple points and may span multiple sublattices [5,7,8]. Due to the strong interdependence between SRO and the properties of materials, it is desirable to systematically quantify chemical ordering in substitutional systems.

The Warren-Cowley SRO parameters are among the most commonly used descriptors for pair ordering in alloys, both experimentally and theoretically [9–14]. These parameters can be written for binary systems as

$$\gamma_m^{pq} = 1 - \frac{P(q|p)_m}{c_q}. \quad (1)$$

The parameter is given in terms of the conditional probability that atom p is at a site with atom q in some neighbor shell around it, labeled by m [Fig. 1(a)]. These probabilities, which can be obtained by inversion of pair correlations, are then divided by the concentration c_q . In a random alloy the pair parameter is 0; when $\gamma^{pq} > 0$, there is a tendency of p - q ordering, and when $\gamma^{pq} < 0$, there is a tendency of p - p and q - q pair ordering. While the description of chemical SRO in terms of pair ordering is useful in many cases, pairs alone do not completely describe a substitutional system. For example, in Fig. 1(b), the Warren-Cowley parameters do not describe the ordering motif where a blue atom occupies a site adjacent to a gray-blue pair (a three-point ordering). It also cannot describe a three-point ordering between sublattices in an alloy oxide. While the high-order (three-body and beyond) correlations of this sort are often less significant than pair correlations, all of the n -body correlations are needed to completely describe a system [15]. Many alloys and substitutional crystal systems can be represented with an Ising-like Hamiltonian that depends on chemical occupation variables of sites in the lattice. In such models, it has been proven that correlations up to the order of the interactions in the

*jmg670@psu.edu

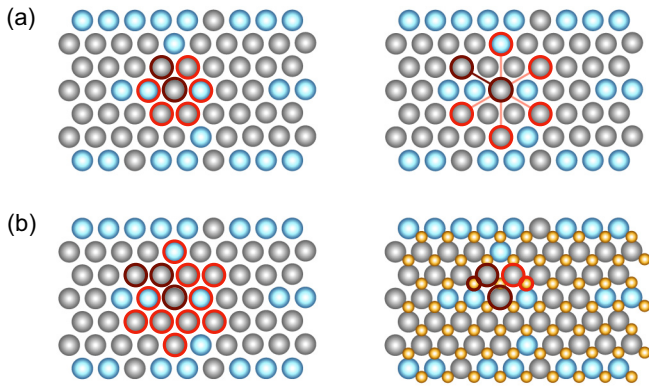


FIG. 1. (a) Chemical ordering is often described in terms of pairs (dark red motifs) and how much the chemistry of a given neighbor shell deviates from its nominal stoichiometry on average in the crystal. (b) Chemical ordering across multiple shells and/or sublattices can be difficult to quantify and often requires approximation.

Hamiltonian (e.g., three-point correlations if three-body interactions are included) are required to completely describe all other correlations in the system [16]. Neglecting many-body correlations, three atomic sites or more, can lead to poor predictions of materials properties, as in the inverse Monte Carlo method [17–19]. Incorporation of high-order correlations and associated SRO into models and analysis of substitutional systems would be beneficial, but it is often dismissed due to challenges in obtaining the multipoint probabilities both experimentally and theoretically.

A number of other developments have been made to extend the Warren-Cowley parameters beyond their typical application to pair ordering in AB alloys, including the extension of the Warren-Cowley parameters to systems with more than two components [20,21]. Work by Clapp and co-workers showed that some multipoint correlations can be obtained from lower-order ones through the Kirkwood superposition (multiplication of pair probabilities) [22–25]. In special cases such as linear binary chains or in equimolar AB alloys, some of the high-order correlations may be exactly expressed in terms of lower-order pair and single-site correlations. In general this is approximate and not suitable for systems with highly correlated lattice occupations or with strong multipoint interactions [16,23]. Similar approximations have been made by Shirley and Wilkins, reconstructing multipoint correlations from pair combinations contained in the motif [26,27]. This method still approximates the occupations of correlated lattice sites contained in the motif as combinations of pairs that occur independently from one another. Its main utility is at the order-disorder transition temperature. Definitions of multipoint order parameters were included by Shirley and Wilkins, but these suffer from the deficiencies associated with the approximated multipoint correlations used to define them. It was proven by Nicholson *et al.* [16], and demonstrated by other authors, that such methods only work in cases where interactions

beyond pairs are negligible [28]. When only pair interactions are significant, obtaining multipoint orderings from pair correlations can be highly successful [29,30]. Some three- and four-point ordering parameters have been defined and used for stochastic generation of two-dimensional substitutional lattices possessing high-order correlations [31,32]. Methods such as the geometrical locus method that quantify the ordering of derivative structures are currently limited to specific crystal systems and motifs [33–35]. Exact quantification of general multipoint orderings is still needed for substitutional lattice systems with multiple components and between sublattices. Approximating these from low-order correlations is desirable for connection to experimental SRO intensities but, as we show in this paper, does not apply well for all systems with many-body interactions above pairs.

The extraction of three-point and higher-order correlations from crystals in x-ray experiments is still an active area of study [36–39]. Impressive strides have been made in energy-resolved scanning tunneling electron microscopy to directly measure SRO domains in alloys, but atomic-level chemical ordering across multiple points in alloy systems is still challenging to quantify [3,40]. Simulation and theory could be used to directly evaluate multipoint chemical ordering to support experimental findings, but it can be challenging to obtain meaningful statistics in substitutional or alloy systems with many degrees of freedom. The notions of statistical efficiency and accuracy need to be addressed as descriptors of chemical ordering, such as the Warren-Cowley parameter, are extended to multipoint motifs. This was partly addressed by the work of de Fontaine when the order parameters were recast as normalized pair probabilities and the number of independent pair parameters was defined for systems with arbitrary numbers of components [20]. We aim to extend the description of normalized probabilities to multipoint ordering in alloys. In this paper, cluster order parameters (ClstOPs) are introduced for systematically quantifying multipoint ordering in multicomponent crystals through direct measurements of normalized cluster probabilities.

II. DEFINING THE CLUSTER ORDER PARAMETERS

A. Order parameters on a single sublattice

We define a set of order parameters to quantify arbitrary multipoint chemical ordering. Like the Warren-Cowley parameters and the current three- and four-point parameters in the literature, the new set of parameters should be 0 for the disordered phase. To begin, an occupation variable representation of a single substitutional lattice is adopted, much like that for Ising or cluster expansion models. The occupations of sites on a lattice are designated by a collection of variables as an occupation or spin vector,

$$\vec{\sigma} = \{\sigma_1, \sigma_2, \sigma_3, \dots, \sigma_N\}. \quad (2)$$

A spin variable σ_i is assigned to each lattice site of an N -site crystal. The spin variables are integers that take on the values

$$\sigma_i = \begin{cases} -m, -m+1, \dots, -1, 1, \dots, m-1, m : & m = d/2 \\ -m, -m+1, \dots, -1, 0, 1, \dots, m-1, m : & m = (d-1)/2; \end{cases} \quad (3)$$

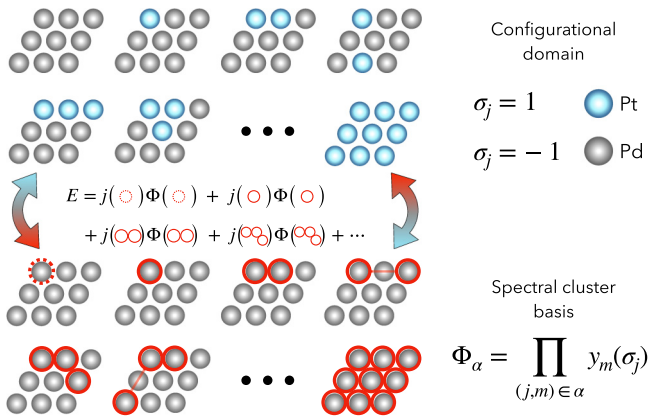


FIG. 2. A spectral expansion of the energy is enabled by defining a complete basis over the entire domain of alloy configurations. The alloy configuration is given by a collection of spin variables that specify the occupation of the lattice sites. The basis functions are defined as all possible products of the spin variables (or appropriate site basis function to maintain orthogonality conditions). For a cluster expansion in this basis, there are terms corresponding to not only single sites and pairs (of different ranges) similar to an Ising model but also all high-order terms such as triplets and beyond.

the first case occurs when the compositional degrees of freedom for a lattice site, d , are even, and the second case occurs when the degrees of freedom are odd. In a binary alloy containing two species A and B, the spin variables can take the values $\sigma_i = \{-1, 1\}$ corresponding to A and B, respectively.

We are interested in developing an order parameter for arbitrary collections of lattice points in a crystal. The complete, orthonormal set of basis functions defined for the discrete variable space in Eq. (2) are a convenient starting point [15]. The cluster basis, spin products, is defined over the entire configurational domain of an alloy crystal (Fig. 2). A scalar extensive property such as the energy may be represented as a linear expansion in this basis, $E = \sum_{\alpha} j_{\alpha} \Phi_{\alpha}$, where the sum runs over clusters α and j_{α} are the expansion coefficients associated with the respective cluster basis function Φ_{α} . The chemical occupations for any combination of lattice sites (e.g., two lattice sites, three lattice sites, or more) are related to the measured value of a corresponding cluster basis function(s) that is defined for the same lattice points. Therefore the chemical ordering of *any* combination of lattice sites may be described within the cluster expansion formalism, because the cluster basis is complete. This formalism is defined below, and the exact relationship is shown.

For a given cluster in the set of cluster basis functions $\{\Phi_{\alpha}(\vec{\sigma}_{\alpha})\}$, the measured value of the basis function for an alloy configuration specified by the spin vector, Eq. (2), depends on the spin variables of the sites contained in the cluster, $\vec{\sigma}_{\alpha}$. Each cluster basis function is given by

$$\Phi_{\alpha} = \prod_{(j,m) \in \alpha} y_m(\sigma_j), \quad (4)$$

where the product of particular site basis functions, indexed by m , is taken over all sites j in the cluster [15]. The combination of site indices, $j \in \alpha$, as well as the site basis function indices, $m \in \alpha$, is implied by the cluster index α . The cluster basis

functions obey the orthogonality condition

$$\langle \Phi_a(\vec{\sigma}) | \Phi_b(\vec{\sigma}) \rangle = \delta_{ab} \quad (5)$$

and the completeness relationship

$$\sum_{\alpha} \Phi_{\alpha}(\vec{\sigma}_1) \Phi_{\alpha}(\vec{\sigma}_2) = \delta_{12}. \quad (6)$$

The choice of the site basis in Eq. (4) is somewhat arbitrary as long as the corresponding cluster functions obey the completeness and orthogonality conditions, Eqs. (5) and (6). In this paper, an appropriate trigonometric basis is used, and the specification of site basis indices, m in Eq. (4), is set by the cluster index α [41]. The set of cluster basis functions includes an empty identity cluster for completeness, single-site, pairs, triplets, quadruplets, and so on, as shown schematically in Fig. 2.

In practice, for periodic crystals, the symmetry of the crystal imposes constraints on the expansion coefficients. For practical cases, it is more convenient to average the cluster basis functions over the crystal. The average cluster basis functions, often referred to as cluster correlation functions, are given as

$$\bar{\Phi}_{\alpha}(\vec{\sigma}) = \frac{1}{m_{\alpha} N} \sum_{\beta \equiv \alpha} \sum_p^{N_p} \Phi_{\beta}[\vec{\sigma}_{\beta}(p)], \quad (7)$$

where the inner sum runs over all distinct locations of a cluster, p , and the outer sum runs over all symmetrically equivalent clusters, $\beta \equiv \alpha$. Per-site correlations are obtained by dividing by the number of lattice sites N and the number of symmetrically equivalent clusters m_{α} . It is noted that the number of distinct cluster locations N_p may differ from the total number of sites N in crystals with reduced symmetry, such as a two-dimensional surface with a set thickness. This is the case in the example given later. A scalar extensive property such as the energy of a substitutional lattice system $E(\vec{\sigma})$ may be represented per site as a linear expansion in the cluster correlations:

$$E(\vec{\sigma}) = \sum_{\alpha} m_{\alpha} J_{\alpha} \bar{\Phi}(\vec{\sigma}_{\alpha}). \quad (8)$$

Here, the sum is performed over all symmetrically distinct clusters, and the coefficients J_{α} , referred to as the effective cluster interactions (ECIs), describe the strength of an interaction averaged over the lattice.

The correlations in Eq. (7) can be written in terms of occupational pair, triplet, quadruplet, and higher-order multiplet probabilities as a weighted average:

$$\bar{\Phi}_{\alpha}(\vec{\sigma}) = \frac{N_p}{N} \sum_{\vec{\sigma}_{\alpha}} \Phi_{\alpha}(\vec{\sigma}_{\alpha}) \bar{P}(\vec{\sigma}_{\alpha}). \quad (9)$$

Here, the sum now runs over all distinct occupations of the cluster multiplied by the respective probability of finding any symmetrically equivalent cluster with that specific occupation in the crystal, $\bar{P}(\vec{\sigma}_{\alpha})$. These probabilities, referred to as cluster probabilities, can be defined as

$$\bar{P}(\vec{\sigma}_{\alpha}) = \frac{1}{m_{\alpha} N_p} \sum_{\beta \equiv \alpha} n(\vec{\sigma}_{\beta}). \quad (10)$$

The total number of clusters with the desired occupation is counted at each distinct location, p , in the crystal to give $n(\vec{\sigma}_\alpha)$. The counts are summed over all symmetrically equivalent clusters and divided by the symmetry multiplicity and total number of occurrences of the cluster in the crystal, N_p . These probabilities sum to 1:

$$\sum_{\vec{\sigma}_\alpha} \bar{P}(\vec{\sigma}_\alpha) = 1. \quad (11)$$

The sum is taken over all possible distinct occupations of the cluster. Using the multipoint probabilities defined in Eq. (10), the ClstOP is defined as

$$\gamma_\alpha(\vec{\sigma}_\alpha^k) = 1 - \frac{\bar{P}(\vec{\sigma}_\alpha^k)}{P_{\text{random}}}, \quad (12)$$

where the order parameter γ_α is given in terms of the average probability of finding a cluster with a desired occupation ($\vec{\sigma}_\alpha^k$) in the crystal that is normalized by the probability of the cluster forming with the desired occupation in a random alloy, P_{random} . In the random alloy the probability of a site being occupied by a specific species, $P(\sigma_i)$, is given by the atomic concentration of that species C_i ; the probability in the denominator is the product of the site probabilities for a given cluster occupation. For example, P_{random} in an AB alloy for a three-point occupation of ($\vec{\sigma}_\alpha^k$) = $[-1, -1, -1]$ corresponding to (AAA) is given by $C_A C_A C_A$. It can be directly shown that by selecting a pair cluster $\alpha \in \{\text{pairs}\}$ the ClstOP reduces to the binary Warren-Cowley SRO parameters.

The analysis of the ClstOPs is similar to that for the Warren-Cowley pair parameters [10]. When the ClstOP is zero, the cluster shape with the specified occupation occurs as frequently as it would in a random alloy. When $\gamma_\alpha > 0$, the cluster with the specified occupation is found less often than in a random alloy of the same composition. Finally, when $\gamma_\alpha < 0$, the cluster with desired occupation occurs more frequently than in a random alloy. The cluster expansion of the energy in Eq. (8) may be written in terms of the ClstOPs. Using Eqs. (9) and (8), the cluster expansion of the energy may be written as

$$E(\vec{\sigma}) = \frac{N_p}{N} \sum_{\alpha} m_\alpha J_\alpha \sum_{\vec{\sigma}_\alpha} \Phi_\alpha(\vec{\sigma}_\alpha) P_{\text{random}} [1 - \gamma_\alpha(\vec{\sigma}_\alpha)], \quad (13)$$

where the outer sum runs over all symmetrically distinct clusters and the inner sum runs over the chemical labelings of that cluster. In this equation, the cluster correlations Φ_α have been rewritten in terms of ClstOPs rather than the cluster probabilities as in Eq. (9). It is noted that the inner sum runs over the same cluster labelings and sites for multicomponent cluster functions belonging to the same orbit; the different site basis functions just give a different value of $\Phi_\alpha(\vec{\sigma}_\alpha)$ [15]. What can be inferred from Eq. (13) is that sums of the chemical ordering parameters scaled by the evaluated basis function for a given cluster labeling, $\Phi_\alpha(\vec{\sigma}_\alpha)$, determine the energy contribution from a given cluster. If the expansion coefficient associated with a cluster correlation Φ_α is large, the influence of the associated chemical orderings will have more of an impact on the system energetics. The relationship between the chemical ordering and the energy is not linear in general; the value of the ClstOP, $\gamma_\alpha(\vec{\sigma}_\alpha^k)$, and all other ClstOPs associated with the

possible occupations of the cluster, $\gamma_\alpha(\vec{\sigma}_\alpha^{k \neq k})$, is constrained dependently by the composition of the system.

One benefit of the completeness of the cluster basis is that the energy contribution from any correlation may be calculated, and through Eq. (13), the energy contributions from any set of associated chemical orderings. These may correspond to long- or short-range correlations. Conversely, a given set of chemical orderings for a system can be used to extract expansion coefficients that may yield said chemically ordered structures (inverse Monte Carlo method) [42]. The inverse Monte Carlo method performed with ClstOPs could be distinguished from traditional inverse Monte Carlo methods that provide interaction energies from pair ordering alone. Limiting values of the ClstOPs may be related to long-range or superstructure orderings depending on the cluster and crystal. The limits of the ClstOPs at large separations could be used to generalize long-range order parameters [43]. Special cases of derivative structure orderings and multipoint motifs can be inferred by the geometrical locus method. Recall that the generalized geometrical locus method provided constraints on the pair parameters spanning derivative polyhedra in certain AB crystals (rocksalt, CsCl structure, and SnS structure) [33,34]. The ordering of octahedra in the rocksalt lattice can be inferred on the basis of composition, whether or not the octahedra are arranged periodically, the composition of the octahedron, and noting that the octahedra span the crystal. With ClstOPs the ordering of octahedra (or other derivative polyhedra) can be directly measured with the polyhedral ‘‘cluster.’’ Defining similar parameters with the cluster probabilities could allow for a generalization of the geometrical locus method to derivative structures beyond polyhedra and to alloys with more than two components [44].

B. Order parameters on multiple sublattices

The ClstOPs can be generalized to crystals with multiple sublattices. This is demonstrated here in a two-sublattice system, as an example. The occupations of the sites in each respective sublattice are designated with spin vectors $\vec{\sigma}$ and $\vec{\delta}$ as in Eq. (2) of the main text. A set of cluster basis functions can be assigned to each sublattice, $\{\Phi_\alpha^1(\vec{\sigma})\}$ and $\{\Phi_\alpha^2(\vec{\delta})\}$. A complete set of composite basis functions can be defined for the multilattice crystal by taking the tensor product of the single-sublattice cluster spaces [45]. Using a notation similar to that given by Tepeš *et al.*, these composite basis functions, indexed by θ , can be denoted as

$$\{\Theta_\theta(\vec{\tau}_\theta)\}, \quad (14)$$

where we have defined a single vector containing the spin variables for the entire crystal:

$$\vec{\tau} = (\sigma_1, \sigma_2, \dots, \sigma_{N_1}, \delta_1, \delta_2, \dots, \delta_{N_2}). \quad (15)$$

The intersublattice and single-sublattice correlations are given as the expectation value of the composite basis functions over the crystal. These are calculated as

$$\bar{\Theta}_\theta(\vec{\tau}) = \frac{1}{m_\theta N} \sum_{\phi=\theta} \sum_c^{N_c} \Theta_\phi[\vec{\tau}_\phi(c)]. \quad (16)$$

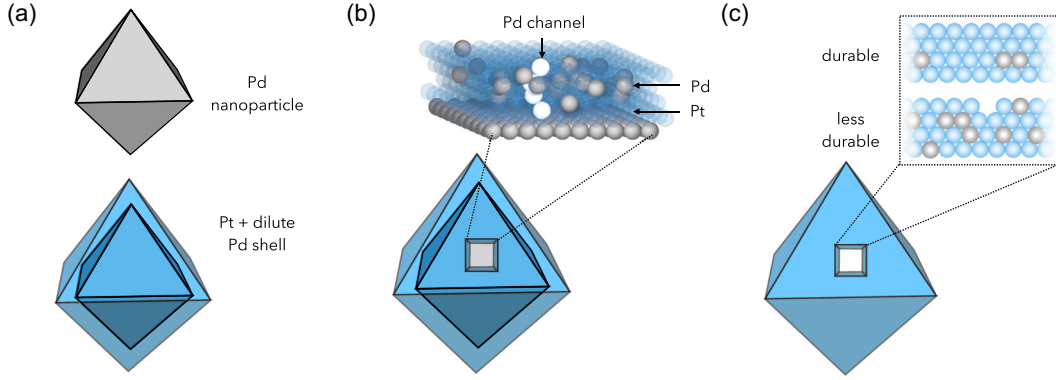


FIG. 3. The palladium (111) face-centered-cubic nanoparticles coated with a few atomic layers of a Pt-Pd alloy (a), prepared in Ref. [6], rely on specific multipoint chemical ordering motifs, palladium channels that span the surface alloy region (b), to generate high-surface-area catalyst cages for the oxygen reduction reaction (c). A sufficient amount of Pd is needed to observe the “channel” order motif [shown in (b)] to allow for the etching of the palladium core but not so much that it diminishes the catalyst durability.

The inner sum runs over all distinct locations of the composite cluster, c , and the outer sum now runs over all symmetrically equivalent composite clusters, $\theta = \phi$. The sum of the evaluated composite cluster functions at all of these points is then divided by the total number of sites in the crystal coming from sublattices 1 and 2, $N = (N_1 + N_2)$, multiplied by the number of symmetrically equivalent composite clusters m_θ .

The correlations in Eq. (16), can also be written as weighted averages of cluster basis functions evaluated for specific *cluster* occupations following a form similar to that for a single sublattice.

$$\bar{\Theta}_\theta(\bar{\tau}) = \frac{N_c}{N} \sum_{\bar{\tau}_\theta} \Theta_\theta(\bar{\tau}_\theta) \bar{P}(\bar{\tau}_\theta), \quad (17)$$

where the sum now runs over all occupations possible in the composite cluster. The multipoint probabilities for the intersublattice correlations are defined as

$$\bar{P}(\bar{\tau}_\theta) = \frac{1}{m_\theta N_c} \sum_{\phi=\theta} n(\bar{\tau}_\phi). \quad (18)$$

The counts of composite clusters with a specific intersublattice occupation, τ_θ , are summed for all symmetrically equivalent composite clusters and divided by the total number of occurrences of the composite cluster in the crystal, N_c , multiplied by the symmetry multiplicity. With this, the intersublattice ClstOP can be defined as

$$\gamma_\theta = 1 - \frac{\bar{P}(\bar{\tau}_\theta^d)}{P_{\text{random}}}. \quad (19)$$

The correlations in Eq. (17) could also be inverted to obtain specific probabilities. In the case of the composite cluster formed by the product of the two single-site basis functions, $\Theta(\sigma_i, \delta_j) = \Phi_{\text{single}}^1(\sigma_i) \Phi_{\text{single}}^2(\delta_j)$, the intersublattice pair probabilities could be extracted. Because correlations beyond pairs can be considered, the ordering of cations about an anion vacancy could be considered in the rocksalt crystal structure, for example [8]. Pair ordering alone would likely show a large tendency of unlike pair ordering between the cation and anion species [46].

III. APPLICATION TO Pt-BASED ALLOY NANOSHELL CATALYST

The need for an exact description of chemical SRO is highlighted in the case of platinum-based nanoshells in Fig. 3 [6]. To make the platinum-based nanoshells, thin layers of dilute Pt-Pd alloy are deposited on palladium nanoparticles. The Pd cores are subsequently leached out leaving a highly active, predominantly Pt shell (9.1% mass Pd). Though the thicknesses of these catalyst shells are as low as four atomic layers, this resembles the surface of many other core-shell and alloy interfaces [47–49]. As evidenced by Zhang *et al.* [6], the formation of Pd channels that span, or nearly span, the deposited surface alloy allows for subsequent etching of the Pd cores. They also found that excess Pd content decreased mechanical stability of the shells. Nanoshell catalysts with increased durability could potentially be produced by decreasing the Pd content while still allowing for Pd channel formation [Fig. 3(b)].

Pair-ordering analysis would provide some insights into the Pd channel content in alloy surfaces, but the pair approximations used to quantify the Pd channel occurrence are somewhat arbitrary. The ClstOPs were used instead to directly quantify Pd channel content in models of this alloy surface and were compared with an approximated channel parameter constructed from pair probabilities. The surfaces of these alloy-coated nanoparticles were modeled using cluster expansions fit with and without a continuum solvent interface above the alloy surface to simulate some effects of the experimental environment [41,50]. In these models, the (111) face-centered-cubic surface was represented with four layers of Pt-Pd alloy on top of four pure palladium layers to describe the palladium core. The occurrence of four-point Pd channels spanning the alloy region determines whether the catalyst can be synthesized and its durability.

A. Model for the alloy surface

1. Density functional theory calculations

A selection of 398 symmetrically unique alloy configurations were used as training data for the cluster expansion.

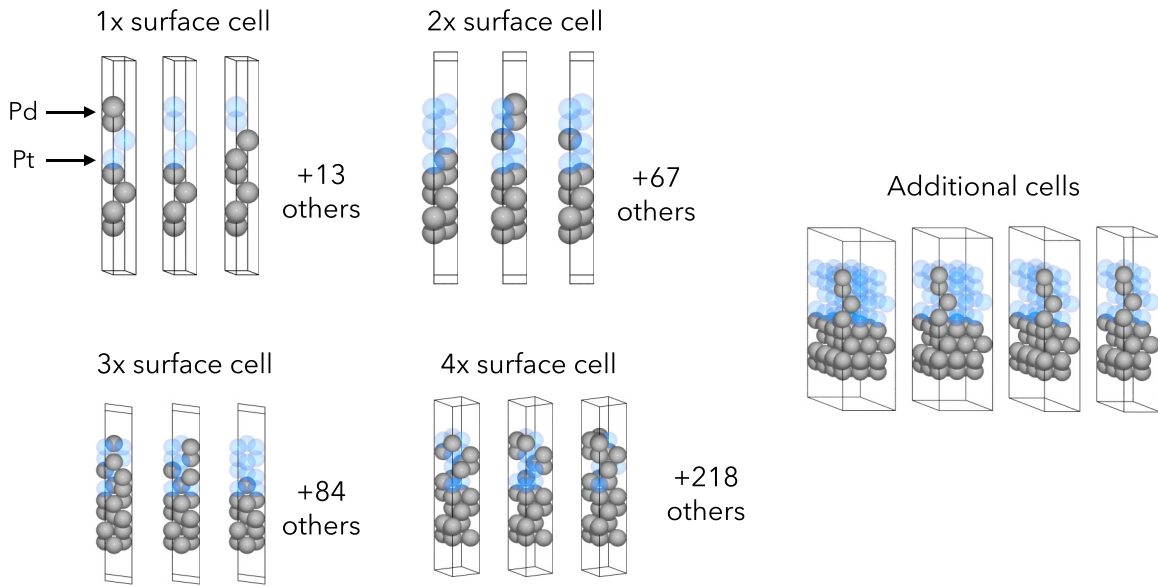


FIG. 4. Examples of surface cells used to fit the cluster expansion coefficients (grouped as multiples of the primitive surface cell). All symmetrically distinct configurations in the $1 \times$ and $2 \times$ cells are used for fitting along with larger randomized surface cells.

Some examples are provided in Fig. 4. The mixing energies of these alloy configurations were calculated using density functional theory (DFT). The surface alloy slabs were composed of eight layers in total with a vacuum height of 6 Å on either side. Four of the layers on the bottom were pure Pd to represent the Pd core of the nanoparticles, and the top four layers were composed of both Pt and Pd with varied concentration. The DFT calculations were performed using the QUANTUM ESPRESSO suite with plane-wave basis sets [51]. The kinetic energy cutoff for the basis sets was 100 Ry. Norm-conserving pseudopotentials from the PSEUDODOJO library were used to represent ion cores [52]. An approximately uniform distribution of reciprocal Bloch vectors (\mathbf{k} points) was used to sample the Brillouin zone across the cells by using a \mathbf{k} -point density of $(11/m \times 11/n \times 1)$ for an $m \times n$ surface cell. Electronic occupations were smoothed with 0.001 Ry of Marzari-Vanderbilt smearing [53]. The \mathbf{k} points, smearing, slab thickness, energy cutoffs, and vacuum height were converged with respect to the Fermi energy of the system, to within 0.05 eV, ensuring that the interfacial dipole was converged. During geometry optimizations of surface alloy slabs the bottom two layers of Pd were fixed at calculated bulk lattice parameters, and total forces were below 25 meV/Å.

To account for average solvent effects on the surface alloy, the DFT energies for the surface alloy configurations were also calculated in the presence of a continuum solvent via the self-consistent continuum solvation (SCCS) method [50]. The shape and onset of the dielectric cavity are defined using minimum and maximum charge density cutoffs ($\rho_{\min} = 0.0013$ a.u. and $\rho_{\max} = 0.010$ 25 a.u.), and the original switching function provided by Andreussi *et al.* [50] was used. The dielectric constant inside the cavity is 1 and switches smoothly to a dielectric constant of 78.3 outside the cavity. Surface tension, pressure, and volume terms were omitted as done by Huang *et al.* [54]. Though the fitting cells in Fig. 4 are not symmetric, the contributions from solvent on the palladium core side should cancel out when calculating the mixing energies.

Models including explicit solvation, adsorption of solution ions, and the etchant will likely show a stronger influence on the surface alloy structure and would help determine solvent conditions suitable for making more durable cages [5,55].

2. Cluster expansions

The cluster expansion model represents a scalar extensive quantity as a linear expansion in the cluster basis functions of Eq. (7) and was obtained using the ICET software package [41]. The mixing enthalpy (per site) for the representative surface alloy was expanded as

$$\Delta H_{\text{mix}}(\vec{\sigma}) = \sum_{\alpha} m_{\alpha} J_{\alpha} \Phi(\vec{\sigma}_{\alpha}), \quad (20)$$

with the sum being taken over all symmetrically distinct clusters up to some maximum size and order. Aside from identity and single-site clusters, all pair and triplet clusters with within three neighbor shells were included. Additionally, some larger quadruplet clusters are included that span the fourth- and fifth-neighbor shells. This is done to include the cluster corresponding to the channel shape in Fig. 3(b) in the energy expression. This resulted in 25 pair, 138 triplet, and 34 quadruplet clusters for a total of 201. The expansion coefficients were trained against DFT fitting data in Fig. 4 using the automatic relevance determination regression (ARDR) method implemented in SCIKIT-LEARN to obtain an optimally sparse set of ECIs and reduce overfitting yielding 68 nonzero ECIs after training [56]. A weighting function was added on the basis of convex hull distances, by $W = (1 + e^{-\frac{D}{kT}})/(1 - e^{-\frac{D}{kT}})$, where D is the distance between the convex hull and the mixing energy of the configuration, k is Boltzmann's constant, and T is the temperature. After optimization of regularization parameters with respect to $k =$ tenfold cross validation, the test error was 5 meV/site. The ECIs as a function of cluster radius and numbers of vertices are given in Fig. 5, with the exception of the identity and single-site clusters with ECIs

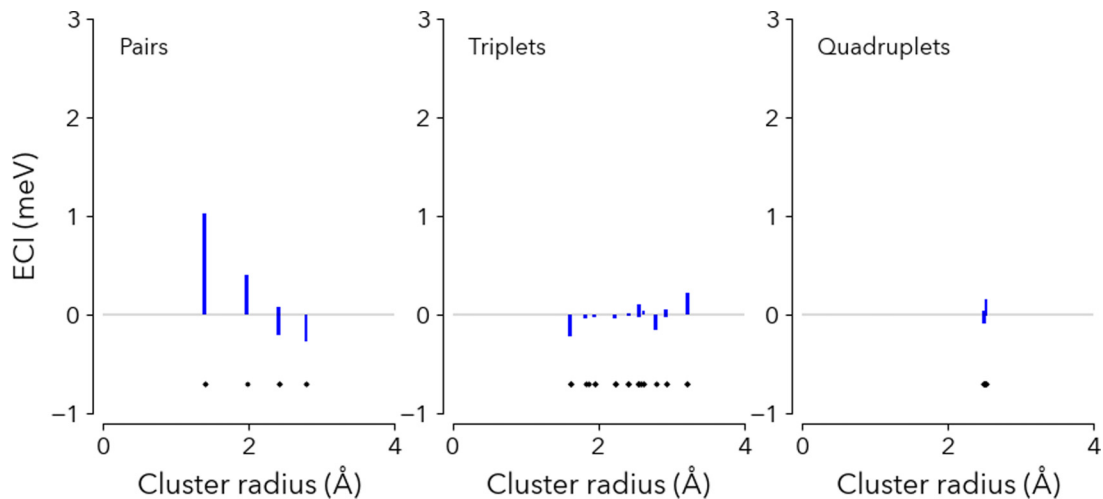


FIG. 5. Effective cluster interactions and expansion coefficients for the surface alloy system in contact with the implicit solvent. The blue bars mark the magnitude and sign of the coefficient for pair, triplet, and quadruplet clusters, order 2, 3, and 4, respectively.

of -15.3 and -2.1 meV, respectively. The relative sizes of the ECIs reflect the relative contribution to the mixing energy for some surface alloy configuration. Relatively large ECIs associated with clusters containing four vertices highlight the importance of interactions beyond pairs in this system.

3. Monte Carlo sampling

Using stochastic algorithms, a representative 10×10 (2.7×2.7 nm) surface cell was sampled in the canonical ensemble at 473 K. Only shells with low concentrations of Pd (1–20% surface content) were considered because of the mechanical destabilization of shells with high palladium content. For each composition, an order parameter was obtained for the chemical ordering motif associated with the the channel shape, a four-point vertically oriented cluster in Fig. 3(b) along with relevant pair orderings. The relatively accurate predictions of the multipoint orderings were facilitated by sampling converged simulations. The convergence of the parallel Monte Carlo simulations was quantified with the potential scale reduction factor from Brooks and Gelman [57], which uses the the effective number of independent measurements, $M_{\text{eff}}(k) = 1/[1 + 2\lambda(k)]$, and ratios of pooled and within-simulation variances to quantify the convergence of the parallel chains [57,58]. Each parallel simulation was run for 1000 passes after a 100-pass burn-in. The resulting potential scale reduction factors are on the order of 1.000 18, and the simulations are considered well converged [57]. The statistical analysis and error measurements are detailed further in the Supplemental Material [59].

The MCHAMMER Monte Carlo software was used through ICET to carry out Markov-chain Monte Carlo samplings of the surface alloy system [41,60]. The ClstOPs were evaluated from the Monte Carlo trajectories using the CLST_ORDER software package developed for this work. The CLST_ORDER PYTHON software calculates the normalized, symmetry-averaged probabilities in Eq. (9) for arbitrary cluster shape and order. The parameters can be calculated in general two- or three-dimensional crystal systems provided that the atoms can be projected onto pristine parent lattice(s). This software

is compatible with the trajectories produced from MCHAMMER but can also be used with other software packages that generate trajectories or structure files compatible with Atomistic Simulation Environment (ASE) such as the Large-Scale Atomic/Molecular Massively Parallel Simulator (LAMMPS) package [61,62].

4. Chemical ordering quantification

Chemical ordering analysis for alloys is commonly given in terms of pair ordering in certain nearest-neighbor shells. Discussions of multipoint ordering motifs generally involve a consideration of all constituent pair orderings contained in the motif [24,34]. In the case of the Pd channel, the Pd pair ordering in the zeroth- to third-nearest-neighbor shells contained in the channel shows some limiting factors for the occurrences of the channels as well as the structure of the alloy overall. Due to the anisotropy of the system along the surface normal, pair orderings in a given neighbor shell are not the same throughout the surface alloy; pair ordering oriented along the surface normal differs from pair ordering parallel to the surface. For this reason, we extract from the Monte Carlo simulations both sets, the set of zeroth- to third-neighbor pair parameters starting from the top of the surface alloy and also from the bottom, of vertically oriented pair order parameters that can be used for a Kirkwood superposition of the Pd channel. The respective anisotropic constituent pair parameters are reported in Fig. S3 of the Supplemental Material to further demonstrate the Kirkwood superposition, while the combined order parameters for the Pd-Pd pairs are reported in Fig. 6 as a function of Pd fraction [59]. The respective combined pair parameters are highlighted in the plot. In crystals possessing three-dimensional periodicity, this same model can be applied with averaging pair probabilities over all equivalent orientations and constant single-site correlations that are given simply by the concentrations.

The parameters in Fig. 6 describe how likely Pd-Pd pairs in the first- to third-neighbor shells of the channel shape are relative to a completely random alloy (along with the single-site Pd ordering in the zeroth-neighbor shell). Recalling that

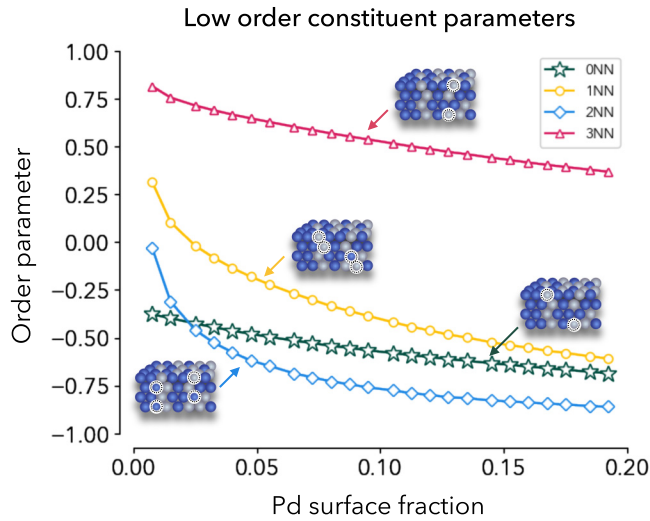


FIG. 6. The pair parameters for all of the nearest-neighbor (NN) shells contained in the four-point channel are provided along with the combined single-site correlations at the surface and bottom of the alloy.

these parameters are zero in a completely random alloy, the lower values for the Pd-Pd pairs in the first-neighbor shell (circle markers) indicate that Pd-Pd pairs are likely to form at the top or bottom of the surface alloy. The likelihood for the occurrence of precursors of the Pd channels is high. In the third shell (triangle markers), the occurrence of Pd-Pd pairs is highly unlikely, and this is one of the key factors that limits the occurrence of Pd channels overall. The Pd tends to reside in the middle of the shell as indicated by the third-neighbor shell parameter as well as the single-site correlation for the top or bottom of the alloy. This supports the experimental findings of Zhang *et al.* [6], because a significant amount of Pd remains after etching of the core. This low-order analysis, which is common in the literature, is useful for determining limiting factors for channel occurrence and alloy structure.

The Pd channels were quantified during parallel Monte Carlo simulations of the surface alloy using ClstOPs, and the results are reported in Fig. 7. For the compositions tested, the channels occur less frequently than in a randomized alloy lattice, because the results show that the four-point Pd channel ClstOP is greater than 0. The channels occur only in small amounts. This supports the need for small amounts of Pd channels without excess that mechanically destabilizes the shell in experiment. From pair analysis alone, one may expect insignificant amounts of Pd channels at experimentally relevant compositions (with an atomic fraction of Pd of $\sim 15\%$) given the limiting formation of third-neighbor shell Pd-Pd pairs (Fig. 6), but there are significant amounts of channels when quantified exactly with ClstOPs. Using the limiting pair parameter alone is not sufficient for quantifying the Pd channel occurrence.

Improved quantification of Pd channels can be made by approximating the four-point probability with combinations of pair probabilities. There are various combination or superposition expressions that could be used to approximate the four-point Pd channel probability, but two key examples are

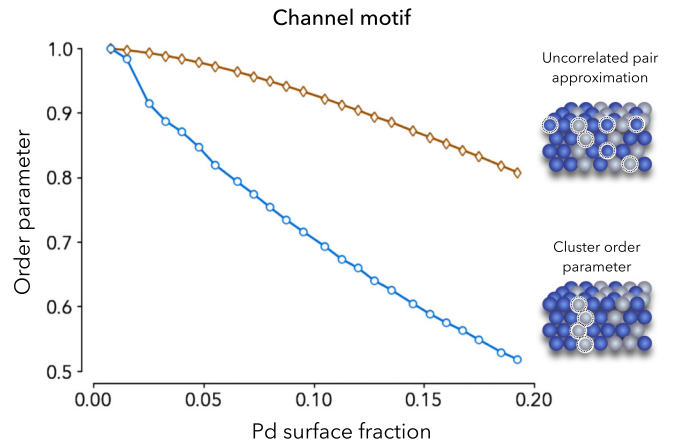


FIG. 7. The exact four-point order parameter corresponding to the channel-shaped cluster from direct calculation during Monte Carlo simulations (blue circles) is compared with that calculated from Clapp's implementation of the Kirkwood superposition (brown diamonds). A visualization of low-order probabilities used to approximate the four-point channel probability is shown for each case.

provided for comparison. Using a slightly modified form of Eq. (7) of Ref. [23], the relative four-point probability can be approximated using Kirkwood's superposition,

$$\bar{P}(\vec{\sigma}_\alpha)/P_{\text{random}} \approx (P_0 \times P_{01} \times P_{02} \times P_{03} + \mathcal{P}_r)/c_{\text{Pd}}^7, \quad (21)$$

where P_{ij} are the Pd probabilities between the i th- and j th-neighbor shells and c_{Pd}^7 is the product of the marginal probabilities for the constituent pairs and single-site correlations. \mathcal{P}_r is the product of the equivalent pair probabilities in reversed order (e.g., starting from the bottom of the surface alloy rather than the top). The corresponding approximation to the ClstOP is given as the curve with diamond markers in Fig. 7. This approximation describes some qualitative trends correctly but deviates from the exact ClstOP value. From a consideration of constituent pair probabilities in this way, the amount of Pd channels at low concentrations may still be misleadingly small. One benefit of this approximation is that it can often be obtained from experimentally determined Warren-Cowley parameters. It is additionally pointed out that the parameters estimated with cluster probabilities constructed via the Kirkwood superposition are known with less certainty. Figure 8 shows the relative standard errors in the measured cluster probabilities using the Kirkwood superposition as well as the ClstOP.

In Fig. 8, it is shown that the relative standard error for the measured quantities for the ClstOP ranges between 0.9 and 580% over the full Pt composition range tested (0.9–22% over the experimentally relevant composition range of $c_{\text{Pd}} = 0.05\text{--}0.2$). The Kirkwood superposition gives a relative standard error range of 1.6–1800% over the full composition range tested (1.6–25% over the experimentally relevant composition range). After propagation of error, the approximation method quantifies the Pd channel occurrence with larger uncertainty than the exact ClstOP due to the multiple measurements needed to construct the Kirkwood superposition. The increased relative error in the low-Pd limit is due to the low Pd concentration, which is related exponentially to

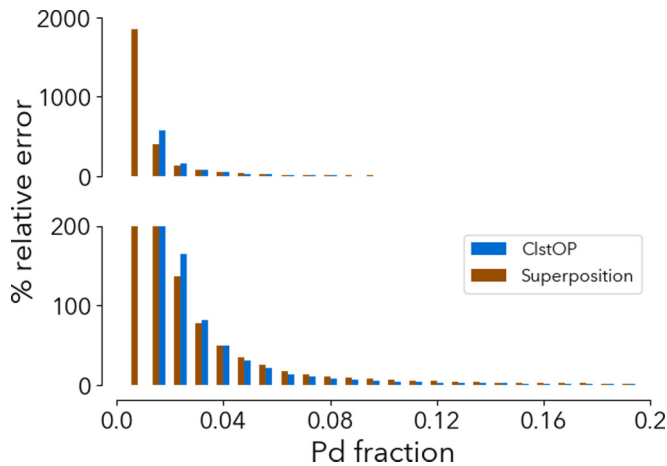


FIG. 8. The relative error in the measured cluster probabilities using the cluster order parameter formalism compared with a constituent pair superposition approximation.

the probability of observing a Pd channel, and possibly the quality of the model in the very dilute Pd limit for the cluster expansion model. Error comparisons and further discussion of the measured cluster probability relative to random alloy fluctuations are contained in Fig. S2 of the Supplemental Material [59]. Connections are provided with original theory for special quasirandom structures [63].

The approximation suggested by Shirley and Wilkins [26,27] may provide a slightly improved prediction of the exact parameter in the surface alloy system. For this case, the four-point probability is approximated as

$$\bar{P}(\vec{\sigma}_\alpha) \approx (P_{01} \times P_{23} + P_{02} \times P_{13} + P_{03} \times P_{12}). \quad (22)$$

This approximation is not given in this paper because it relies on spatially resolved ordering probabilities that are not obtained from typical experimentally derived pair parameters. In many three-dimensional alloy crystals, the cluster probabilities are averaged over the distinct cluster locations needed to construct the four-point probability in Eq. (22).

IV. CONCLUSION

A general cluster order parameter (ClstOP) was introduced to systematically quantify multipoint ordering motifs in alloy crystals via direct calculation of cluster probabilities. This parameter can be used to evaluate chemical ordering in alloys and substitutional systems that cannot be addressed by pair-ordering analysis alone. Though the pair ordering and pair interactions are often most important, there are systems where higher-order correlations are significant and cannot be ignored. The utility of the ClstOPs is that a specific multipoint motif of interest can be quantified directly in simulations or in theoretical applications. In certain special cases where “clusters” are chosen such that they can represent derivative structures of a crystal, these parameters could be used to help generalize the geometrical locus method. Despite the expected low probability of occurrence for multipoint ordering motifs, meaningful predictions of ClstOPs can be made through ef-

ficient sampling during parallel Markov-chain Monte Carlo simulations. The average ClstOPs can be predicted with reasonable certainty with relative standard errors of 22–0.9% within the experimentally relevant composition range of $c_{\text{Pd}} = 0.05$ – 0.20 , respectively. The ClstOP associated with the four-point Pd channel motif quantifies Pd channel occurrence with improved certainty and accuracy over approximate methods such as the Kirkwood superposition. Similar sampling approaches could be used to predict multipoint ClstOPs in many other practical applications such as descriptors for data-driven materials discovery and machine-learning models of alloy systems [64]. With the multilattice generalization, chemical orderings between multiple sublattices can also be described. This generalization is particularly useful for describing ordering between ligand vacancies on one sublattice and alloying metals in another.

The utility of the parameters was demonstrated while modeling representative surfaces of the Pt-Pd nanoparticle alloy system in Fig. 3. Cluster expansion models were generated for a four-layer dilute Pt-Pd alloy on top of a four-layer bulk palladium core region with the (111) surface orientation, and these systems were sampled using Markov-chain Monte Carlo simulations. In this system, Pd channels that span the surface alloy are needed for the synthesis of high-surface-area Pt nanocage catalysts. The calculated ClstOP corresponding to a channel shape in these simulations suggests that Pd channels occur in significant amounts, with 0.12 channels per nanoparticle with experimental facet sizes (octahedra with 19.4 Å edge lengths with a 15% atomic fraction of Pd in the surface). Using pair probabilities alone to quantify Pd channels leads to a significant underestimation of channel ordering, and less than half as many channels occurring with a predicted 0.05 channels per nanoparticle. Though pair ordering provides a wealth of information about the alloy structure overall and the factors that limit the occurrence of a Pd channel, it provides poor estimates of the true number of channels occurring in the alloy. This is because multipoint energetics are significant in this system. Additional results in the Supplemental Material suggest that different solvent environments could aid in the design of nanostructured catalysts that are more durable [59]. Solvation induces a surface enrichment of Pd without completely eliminating the occurrence of Pd channels that span the surface alloy. This was shown by exact quantification of the four-point channel motif where the limiting factors derived from pair ordering alone may lead one to believe that the Pd channels do not occur in significant amounts.

The data are available upon reasonable request from the authors. Software to calculate the order parameters is also available [65].

ACKNOWLEDGMENTS

J.M.G., S.B.S., and I.D. acknowledge financial support from the U.S. Department of Energy, Office of Science, Basic Energy Sciences, CPIMS Program, under Award No. DE-SC0018646. First-principles calculations for this research were performed on the Pennsylvania State University’s ROAR supercomputer. A portion of the Monte Carlo analysis was performed on the CNMS CADES supercomputer at Oak Ridge National Laboratory.

- [1] C. Wolverton, V. Ozolinš, and A. Zunger, First-principles theory of short-range order in size-mismatched metal alloys: Cu-Au, Cu-Ag, and Ni-Au, *Phys. Rev. B* **57**, 4332 (1998).
- [2] J. C. Fisher, On the strength of solid solution alloys, *Acta Metall.* **2**, 9 (1954).
- [3] R. Zhang, S. Zhao, J. Ding, Y. Chong, T. Jia, C. Ophus, M. Asta, R. O. Ritchie, and A. M. Minor, Short-range order and its impact on the CrCoNi medium-entropy alloy, *Nature (London)* **581**, 283 (2020).
- [4] J. D. Joannopoulos and M. L. Cohen, Theory of short-range order and disorder in tetrahedrally bonded semiconductors, in *Solid State Physics*, edited by H. Ehrenreich, F. Seitz, and D. Turnbull (Academic, New York, 1976), Vol. 31, pp. 71–148.
- [5] B. C. Han, A. VanderVen, G. Ceder, and B. J. Hwang, Surface segregation and ordering of alloy surfaces in the presence of adsorbates, *Phys. Rev. B* **72**, 205409 (2005).
- [6] L. Zhang, L. T. Roling, X. Wang, M. Vara, M. Chi, J. Liu, S.-I. Choi, J. Park, J. A. Herron, Z. Xie, M. Mavrikakis, and Y. Xia, Platinum-based nanocages with subnanometer-thick walls and well-defined, controllable facets, *Science* **349**, 412 (2015).
- [7] J. Pan, J. J. Cordell, G. J. Tucker, A. Zakutayev, A. C. Tamboli, and S. Lany, Perfect short-range ordered alloy with line-compound-like properties in the ZnSnN₂:ZnO system, *npj Comput. Mater.* **6**, 63 (2020).
- [8] C. M. Rost, E. Sachet, T. Borman, A. Moballegh, E. C. Dickey, D. Hou, J. L. Jones, S. Curtarolo, and J.-P. Maria, Entropy-stabilized oxides, *Nat. Commun.* **6**, 8485 (2015).
- [9] J. M. Cowley, X-ray measurement of order in single crystals of Cu₃Au, *J. Appl. Phys. (Melville, NY)* **21**, 24 (1950).
- [10] J. M. Cowley, An approximate theory of order in alloys, *Phys. Rev.* **77**, 669 (1950).
- [11] I. Mirebeau, M. Hennion, and G. Parette, First Measurement of Short-Range-Order Inversion as a Function of Concentration in a Transition Alloy, *Phys. Rev. Lett.* **53**, 687 (1984).
- [12] T. Mohri, K. Terakura, S. Takizawa, and J. M. Sanchez, First-principles study of short range order and instabilities in AuCu, AuAg and AuPd alloys, *Acta Metall. Mater.* **39**, 493 (1991).
- [13] B. Schönfeld, H. Roelofs, A. Malik, G. Kostorz, J. Plessing, and H. Neuhäuser, The microstructure of CuAl, *Acta Mater.* **44**, 335 (1996).
- [14] A. Fernández-Caballero, J. S. Wróbel, P. M. Mummery, and D. Nguyen-Manh, Short-range order in high entropy alloys: Theoretical formulation and application to Mo-Nb-Ta-V-W system, *J. Phase Equilib. Diffus.* **38**, 391 (2017).
- [15] J. M. Sanchez, F. Ducastelle, and D. Gratias, Generalized cluster description of multicomponent systems, *Phys. A (Amsterdam)* **128**, 334 (1984).
- [16] D. M. C. Nicholson, R. I. Barabash, G. E. Ice, C. J. Sparks, J. L. Robertson, and C. Wolverton, Relationship between pair and higher-order correlations in solid solutions and other Ising systems, *J. Phys.: Condens. Matter* **18**, 11585 (2006).
- [17] W. Schweika and A. E. Carlsson, Short-range order in Ising-like models with many-body interactions: Description via effective pair interactions, *Phys. Rev. B* **40**, 4990 (1989).
- [18] D. B. Laks, L. G. Ferreira, S. Froyen, and A. Zunger, Efficient cluster expansion for substitutional systems, *Phys. Rev. B* **46**, 12587 (1992).
- [19] C. Wolverton, A. Zunger, and B. Schönfeld, Invertible and non-invertible alloy Ising problems, *Solid State Commun.* **101**, 519 (1997).
- [20] D. de Fontaine, The number of independent pair-correlation functions in multicomponent systems, *J. Appl. Crystallogr.* **4**, 15 (1971).
- [21] A. V. Ceguerra, M. P. Moody, R. C. Powles, T. C. Petersen, R. K. Marceau, and S. P. Ringer, Short-range order in multicomponent materials, *Acta Crystallogr., Sect. A: Found. Crystallogr.* **68**, 5 (2012).
- [22] P. C. Clapp and S. C. Moss, Correlation functions of disordered binary alloys. I, *Phys. Rev.* **142**, 418 (1966).
- [23] P. C. Clapp, Exact relations between triplet probabilities and pair correlations in AB binary alloys and ising spin-1/2 systems, *Phys. Rev.* **164**, 1018 (1967).
- [24] P. C. Clapp, Theoretical determination of *n*-site configuration probabilities from pair correlations in binary lattices, *J. Phys. Chem. Solids* **30**, 2589 (1969).
- [25] J. G. Kirkwood, Statistical mechanics of fluid mixtures, *J. Chem. Phys.* **3**, 300 (1935).
- [26] C. G. Shirley and S. Wilkins, Many-site interactions and correlations in disordered binary alloys, *Phys. Rev. B* **6**, 1252 (1972).
- [27] C. G. Shirley and S. W. Wilkins, Disordered binary alloys. II. Decoupling schemes for many-site correlation functions, *Phys. Rev. B* **16**, 3484 (1977).
- [28] P. C. Clapp, Atomic configurations in binary alloys, *Phys. Rev. B* **4**, 255 (1971).
- [29] D. G. Gratias and P. Cenedese, A C.V.M. approach of multiplet correlation functions in substitutional solid solutions, *J. Phys., Colloq.* **46**, C9-149 (1985).
- [30] R. De Ridder, Iterative method for the calculation of atomic cluster probabilities from pair correlations in binary systems, *Phys. A (Amsterdam)* **79**, 217 (1975).
- [31] T. R. Welberry and D. P. Craig, Solution of crystal growth disorder models by imposition of symmetry, *Proc. R. Soc. London, Ser. A* **353**, 363 (1977).
- [32] T. Welberry, Multi-site correlations and the atomic size effect, *J. Appl. Crystallogr.* **19**, 382 (1986).
- [33] M. Brunel, F. De Bergevin, and M. Gondrand, Determination theorique et domaines d'existence des differentes surstructures dans les composes A³⁺B¹⁺X₂²⁻ de type NaCl, *J. Phys. Chem. Solids* **33**, 1927 (1972).
- [34] M. Sauvage and E. Parthé, Prediction of diffuse intensity surfaces in short-range-ordered ternary derivative structures based on SnS, NaCl, CsCl and other structures, *Acta Crystallogr., Sect. A: Cryst. Phys., Diff., Theor. Gen. Crystallogr.* **30**, 239 (1974).
- [35] D. van Dyck, C. Conde-Amiano, and S. Amelinckx, The diffraction pattern of crystals presenting a digenite type of disorder II. The structure of the digenite-related phases derived by means of the cluster theory, *Phys. Status Solidi A* **58**, 451 (1980).
- [36] F. Lehmkuhler, G. Grübel, and C. Gutt, Detecting orientational order in model systems by X-ray cross-correlation methods, *J. Appl. Crystallogr.* **47**, 1315 (2014).
- [37] B. Pedrini, A. Menzel, M. Guizar-Sicairos, V. A. Guzenko, S. Gorelick, C. David, B. D. Patterson, and R. Abela, Two-dimensional structure from random multiparticle X-ray scattering images using cross-correlations, *Nat. Commun.* **4**, 1647 (2013).

- [38] B. Pedrini, A. Menzel, V. A. Guzenko, C. David, R. Abela, and C. Gutt, Model-independent particle species disentanglement by X-ray cross-correlation scattering, *Sci. Rep.* **7**, 45618 (2017).
- [39] P.-A. Lemieux and D. J. Durian, Investigating non-Gaussian scattering processes by using n th-order intensity correlation functions, *J. Opt. Soc. Am. A* **16**, 1651 (1999).
- [40] R. Zhang, S. Zhao, C. Ophus, Y. Deng, S. J. Vachhani, B. Ozdol, R. Traylor, K. C. Bustillo, J. W. Morris, D. C. Chrzan, M. Asta, and A. M. Minor, Direct imaging of short-range order and its impact on deformation in Ti-6Al, *Sci. Adv.* **5**, eaax2799 (2019).
- [41] M. Ångqvist, W. A. Muñoz, J. M. Rahm, E. Fransson, C. Durniak, P. Rozyczko, T. H. Rod, and P. Erhart, ICET – A Python library for constructing and sampling alloy cluster expansions, *Adv. Theory Simul.* **2**, 1900015 (2019).
- [42] V. Gerold and J. Kern, The determination of atomic interaction energies in solid solutions from short range order coefficients—an inverse monte-carlo method, *Acta Metall.* **35**, 393 (1987).
- [43] J. M. Cowley, Short-range order and long-range order parameters, *Phys. Rev.* **138**, A1384 (1965).
- [44] W. S. Morgan, G. L. Hart, and R. W. Forcade, Generating derivative superstructures for systems with high configurational freedom, *Comput. Mater. Sci.* **136**, 144 (2017).
- [45] P. D. Tapesch, G. D. Garbulsky, and G. Ceder, Model for Configurational Thermodynamics in Ionic Systems, *Phys. Rev. Lett.* **74**, 2272 (1995).
- [46] M. Stana, B. Sepiol, R. Kozubski, and M. Leitner, Chemical ordering beyond the superstructure in long-range ordered systems, *New J. Phys.* **18**, 113051 (2016).
- [47] P. Mani, R. Srivastava, and P. Strasser, Dealloyed Pt-Cu core-shell nanoparticle electrocatalysts for use in PEM fuel cell cathodes, *J. Phys. Chem. C* **112**, 2770 (2008).
- [48] S. Alayoglu, A. U. Nilekar, M. Mavrikakis, and B. Eichhorn, Ru-Pt core-shell nanoparticles for preferential oxidation of carbon monoxide in hydrogen, *Nat. Mater.* **7**, 333 (2008).
- [49] V. R. Stamenkovic, B. Fowler, B. S. Mun, G. Wang, P. N. Ross, C. A. Lucas, and N. M. Markovi, Improved oxygen reduction activity on Pt₃Ni(111) via increased surface site availability, *Science* **315**, 493 (2007).
- [50] O. Andreussi, I. Dabo, and N. Marzari, Revised self-consistent continuum solvation in electronic-structure calculations, *J. Chem. Phys.* **136**, 064102 (2012).
- [51] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, A. D. Corso, S. de Gironcoli, S. Fabris, G. Fratesi, R. Gebauer, U. Gerstmann, C. Gougoussis, A. Kokalj, M. Lazzeri, L. Martin-Samos *et al.*, QUANTUM ESPRESSO: A modular and open-source software project for quantum simulations of materials, *J. Phys.: Condens. Matter* **21**, 395502 (2009).
- [52] M. J. van Setten, M. Giantomassi, E. Bousquet, M. J. Verstraete, D. R. Hamann, X. Gonze, and G. M. Rignanese, The PseudoDojo: Training and grading a 85 element optimized norm-conserving pseudopotential table, *Comput. Phys. Commun.* **226**, 39 (2018).
- [53] N. Marzari, D. Vanderbilt, A. DeVita, and M. C. Payne, Thermal Contraction and Disordering of the Al(110) Surface, *Phys. Rev. Lett.* **82**, 3296 (1999).
- [54] J. Huang, N. Hörmann, E. Oveisi, A. Loidice, G. L. De Gregorio, O. Andreussi, N. Marzari, and R. Buonsanti, Potential-induced nanoclustering of metallic catalysts during electrochemical CO₂ reduction, *Nat. Commun.* **9**, 3117 (2018).
- [55] L. Cao, L. Niu, and T. Mueller, Computationally generated maps of surface structures and catalytic activities for alloy phase diagrams, *Proc. Natl. Acad. Sci. USA* **116**, 22044 (2019).
- [56] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, Scikit-learn: Machine learning in Python, *J. Mach. Learn. Res.* **12**, 2825 (2011).
- [57] S. P. Brooks and A. Gelman, General methods for monitoring convergence of iterative simulations, *J. Comput. Graph. Stat.* **7**, 434 (1998).
- [58] A. Gelman and D. B. Rubin, Inference from iterative simulation using multiple sequences, *Stat. Sci.* **7**, 457 (1992).
- [59] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevB.104.054109> for the statistical analysis and error measurements.
- [60] D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman, emcee: The MCMC Hammer, *Publ. Astron. Soc. Pac.* **125**, 306 (2013).
- [61] A. H. Larsen, J. J. Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dułak, J. Friis, M. N. Groves, B. Hammer, C. Hargus, E. D. Hermes, P. C. Jennings, P. B. Jensen, J. Kermode, J. R. Kitchin, E. L. Kolsbjerg, J. Kubal, K. Kaasbjerg, S. Lysgaard, J. B. Maronsson *et al.*, The atomic simulation environment—a Python library for working with atoms, *J. Phys.: Condens. Matter* **29**, 273002 (2017).
- [62] S. Plimpton, Fast parallel algorithms for short-range molecular dynamics, *J. Comput. Phys.* **117**, 1 (1995).
- [63] A. Zunger, S.-H. Wei, L. G. Ferreira, and J. E. Bernard, Special Quasirandom Structures, *Phys. Rev. Lett.* **65**, 353 (1990).
- [64] P. C. Jennings, S. Lysgaard, J. S. Hummelshøj, T. Vegge, and T. Bligaard, Genetic algorithms for computational materials discovery accelerated by machine learning, *npj Comput. Mater.* **5**, 46 (2019).
- [65] https://github.com/jmgoff/clst_order.