Cluster expansion methods from physical concepts

Paul E. Lammert
Department of Physics, Pennsylvania State University, University Park, PA 16802-6300

Vincent H. Crespi
Department of Physics, Pennsylvania State University, University Park, PA 16802-6300
Department of Chemistry, Pennsylvania State University, University Park, PA 16802-6300 and
Department of Materials Science and Engineering Pennsylvania State University, University Park, PA 16802-6300
(Dated: Aug. 30, 2022)

The cluster expansion formalism used in materials science is reconstructed on an axiomatic basis with the aims of clarifying underlying concepts and improving computational procedures, and without using conventional cluster functions. Instead, cluster components of configuration functions are defined in an intrinsic manner, which can be viewed as Möbius inversion of conditional expectation. The associated method for fitting a model to a configurational sample is grounded entirely in Hilbert space geometry. By constructing models directly from the given data, we avoid an underdetermination problem to which the conventional approach is subject. Tensor observables are treated on an equal footing with scalar observables.

Keywords: cluster expansion; alloy theory; Möbius inversion

CONTENTS

1. Introduction Outline

2. Critical review of conventional formalism
   A. Apparatus
      1. encoding
      2. orthonormal bases of cluster functions
   B. Shortcomings

3. Notation and nomenclature

4. Rough ideas
   A. Effective and reduced entities
   B. A heuristic shortcut

5. Axioms for best-approximation projectors and consequences
   A. Axioms
   B. cluster components and decomposition
   C. Defending the axioms
   D. Probabilistic interpretation
      1. Basic probability
      2. $E$ is an expectation operator
      3. $E$ induces a Hilbert space structure
      4. The $E_I$ are orthogonal (self-adjoint) projections
      5. Marginalization
      6. Internalizing evaluation, or, hats
   E. Many equivalents of independent sites
   F. Why take $IS$ as an axiom?
   G. Proof of Proposition 6.1

7. Taking stock
   A. Assessing the framework
   B. On the rest of the paper

8. Symmetry
   A. Permutation actions
   B. Symmetries of $E$
   C. Tensor-valued observables
   D. Averaging over the group
   E. Cluster-shape subspaces

9. Two points of computational efficiency
   A. Folding and unfolding
   B. Storage and use of cluster observables

10. Sampling, modeling, and Hilbert space geometry
    A. Defining a precise problem
    B. Solving the abstract problem
    C. Application to the sampling context
       1. Loss function
       2. Model parametrization

11. Conclusion

Acknowledgments

References
1. INTRODUCTION

Discovery and elucidation of structure-property relationships is a core concern of materials science\[1\]. For dealing with configurational dependence, cluster expansion methods\[2\,11\] have become a widespread tool, as attested by recent reviews\[12\,14\]. There are now many software packages and libraries\[15\,19\] allowing one to incorporate cluster expansion into multiscale modeling. We suggest that, despite this level of development, the core formalism is not understood as well as it should be. A symptom of this is rigid adherence to specific, pre-defined, orthonormal bases of so-called cluster functions $\Phi_\alpha$, so that one cannot describe what is going on in a basis-free manner. This is like distinguishing $p$-orbitals not by using the concept of angular momentum, but through a fixed $p_x$, $p_y$ and $p_z$ basis. From a practical standpoint, the problem with such a situation is that, internally, one may fail to see opportunities to do computations more efficiently or robustly, while externally, distinguishing cluster expansion from other potential ways of solving the modeling problem (see paragraph after next) or even conceiving possible contours of such a thing are more difficult. For example, imagine some seemingly new of machine learning algorithm. How could we tell whether it was really a sort of cluster expansion in disguise?

The project of this paper is first to unearth and clearly display the logic of cluster expansion methods — what do they mean? Then we show the practical utility of our answer. In addition to its intrinsic interest, this supports the claim that the proposed view of matters is a specially good one. The fundamental issue is to understand whether there is some well-defined and nonarbitrary principle whereby a configuration function is broken down into elementary building blocks like $n$-body interactions. We show that there is such a principle (See Def. 5.1). However, the decomposition is not absolute, but relative to a probability distribution on configurations. A Hilbert space structure (i.e., inner product), an unexplained basic ingredient of the conventional formalism, appear as a natural consequence of this conceptual analysis. Section 6 is then able to explain what makes the conventional cluster functions $\Phi_\alpha$ special — they provide bases of structurally defined cluster subspaces when sites are probabilistically independent.

Return to the modeling problem mentioned earlier. Given total energy on a sampling of configurations, how can we find an approximate decomposition into zero-body, one-body, two-body, $\ldots$ interactions? This problem (not only for energy) makes up the great majority of applications of cluster expansion methods. Our position is that such modeling should explicitly reflect and incorporate the Hilbert space structure of the space of configuration functions. The conventional approach does not do that, but rather fits coefficients $J^\alpha_i$ of a model $\sum J^\alpha_i \Phi^\alpha_i$ as though the $\Phi^\alpha_i$ were generic expansion functions. In contrast, we treat the problem from a perspective based on Hilbert space geometry. A model is expanded on a basis derived directly from the sample data, thereby avoiding an underdetermination problem to which the conventional approach is prone.

Outline

We give now a more detailed guide to the contents. But first, some general comments. It may appear at first glance that the paper builds up a huge and unwieldy apparatus. However, much of it is one-time analysis, with the goal of a clear and flexible formalism. We also pay attention to computational complexity. The mathematics involved is mostly discrete probability and finite-dimensional vector/Hilbert spaces, theory covered in many quantum mechanics courses such as 20 and 21. Sections 8 and 9 involve a little general group representation theory\[22\,23\]. Generally, it is safe to skip things enclosed in “Proof … □”.

Section 2 is a brief review of the conventional cluster expansion formalism. It expands on the motivations already given in this Introduction. Section 3 establishes notation which will be used throughout. Section 4 begins our new development from the ground up, introducing the “best approximation” idea and showing how elementary building blocks (cluster components) can be extracted. At least from a logical perspective, Section 5 is the core of the new formulation. We give informal arguments for a small set of axioms, the idea being to sharpen and codify an initially inchoate concept of best approximation in this context of arbitrary configuration functions. Unpacked, the axioms say we have a probability distribution on configurations. This is not surprising, and possibly even reassuring. Very importantly for what follows, this section sets up the naturally associated Hilbert space.

The conventional formalism requires probabilistic independence of the degrees of freedom on distinct sites. Section 6 explores this restriction, developing a long list of equivalents and consequences. From that point on, we assume that the independence condition holds. Section 7 explains that without that restriction, cluster components have properties which are probably counter to intuitions. Site independence is therefore presented as a secondary axiom. Section 8 recovers the conventional cluster functions in asking for orthonormal bases (ONBs) of the cluster subspaces. At that point, the core formalism is finished. Section 9 recapitulates and does some stage-setting for the more applied part which follows. Section 10 brings both space-group symmetry and tensor-valued observables into the formalism; it is natural to consider these together. Many important material properties are tensorial, and they are incorporated into our framework with no more difficulty than into the conventional one\[1\]. Section 11 shows that eschewing cluster functions is computationally reasonable. Finally, section 12 tackles the modeling problem referred to in the
second paragraph of this introduction. Our approach is exclusively through Hilbert space geometry. We show that it is practical to construct a model directly from the sampled data, and that this even avoids an underdetermination problem which affects the conventional formalism.

2. CRITICAL REVIEW OF CONVENTIONAL FORMALISM

This Section gives a brief overview of the conventional cluster expansion formalism. The purposes are to amplify points made earlier, make the contrast with our approach clearer, and help those familiar with the conventional methods to avoid confusion when we do something different. In principle, none of this is crucial for what follows. The new formalism is developed from the ground up in the following sections.

A. Apparatus

The basic context for cluster expansion methods is a finite collection Site of sites, and a collection Sps of species that may occupy the sites. Typically Sps is something like \{Ni, Zn, Mg, Cu, Co\} and the sites are sites of a (finite or periodic) crystal structure. A configuration \(x\) is an assignment \(i \mapsto x_i\) of species to sites. The set of all configurations is denoted \(\text{Cfg}\). The subject that cluster expansion talks about is real or complex configuration functions. These comprise a vector space \(\mathcal{F}\).

There are two ubiquitous pieces of machinery in the cluster expansion literature: encoding by numbers, and the use of so-called cluster functions as a basis for \(\mathcal{F}\). The former is really a somewhat superficial aspect, but we discuss it anyway due to its ubiquity and the influence it has on the cluster functions.

1. encoding

Encoding is almost universally presented as a matter of replacing Sps with a set of numbers. Typically, with our example Sps, we would have Ni \(\mapsto 1\), \ldots, Co \(\mapsto 5\), or Ni \(\mapsto -2\), \ldots, Co \(\mapsto 2\). Each site \(i\) is then equipped with a site variable, occupation variable, or spin variable (common terms), such that, for example, \(\sigma_i = 1\) means that there is Ni at site \(i\).

We believe that, although formally it changes almost nothing, a better perspective is to consider \(\sigma_i\) not as a variable, but as a fixed function, an element of \(\mathcal{F}\), because of the way the \(\sigma_i\)'s are used. \(\mathcal{F}\) is not only a vector space, it is an algebra, that is, its members can be multiplied together: \((fg)(x) = f(x)g(x)\). The algebra generated by the \(\sigma_i\)'s, that is, the set of polynomials in them collectively, is necessarily a subset of \(\mathcal{F}\). If each \(\sigma_i\) is one-to-one, this algebra is all of \(\mathcal{F}\). The key point is the following lemma. The * appearing in it can be thought of as standing for a generic, unspecified, or abstract site. A proof is given for completeness, but none of the ideas therein are used anywhere else.

**Lemma 2.1.** Suppose \(\sigma_i\) is a one-to-one real (or complex) function on a set Sps with \(N\) elements \(|\text{Sps}| = N\). Then, the powers \(1 = \sigma_i^0, \sigma_i, \sigma_i^2, \ldots, \sigma_i^{\text{Sps}-1}\) are a linear basis for the space \(\mathcal{F}(\ast)\) of functions \(\text{Sps} \rightarrow \mathbb{R}\). For some polynomial \(p\) of degree \(N - 1\), \(\sigma_i^m = p(\sigma_i)\).

Thus, for each site \(i\), the space \(\mathcal{F}(i)\) of functions depending on only the species at site \(i\) is a copy of \(\mathcal{F}(\ast)\), and \(1, \sigma_i, \ldots, \sigma_i^{\text{Sps}-1}\) is a basis of \(\mathcal{F}(i)\). Furthermore, the set of all monomials \(\prod_{i \in \text{Site}} \sigma_i^{m_i}\), with each \(m_i\) in \(\{0, 1, \ldots, |\text{Sps}| - 1\}\), is therefore a linear basis of \(\mathcal{F}\). Since factors \(\sigma_i^0\) make no difference to the product, we can equally describe this basis as consisting of monomials \(\prod_{i \in I} \sigma_i^{m_i}\) for some set of sites \(I\), and with \(1 \leq m_i \leq |\text{Sps}| - 1\).

Any member of \(\mathcal{F}\) can be expressed as a polynomial in the \(\sigma_i\)'s, this representation is highly nonunique. Although this fact has little effect on most applications, it certainly complicates theoretical considerations, nor does it seem to us to have practical advantages over dealing directly with configuration functions.

**Proof of Lemma 2.1.** Let \(\text{Sps} = \{s_1, \ldots, s_N\}\), and \(\delta_j\) be the indicator function of \(s_j\) (\(\delta_i(s_j) = 1\) if \(j = i\), otherwise zero). (1) The dimension of \(\mathcal{F}(\ast)\) is \(N\) because the \(\delta_i\)'s are a basis. (2) Since \(\delta_i\) is a multiple of \([\sigma_s - \sigma_s(s_2)] \cdots [\sigma_s - \sigma_s(s_N)]\), and similarly for the other indicators, polynomials in \(\sigma_s\) span \(\mathcal{F}(\ast)\). Hence, any linearly independent collections of polynomials can be expanded to one of cardinality \(N\), but no larger. (3) Now, let \(M\) be smallest such that \(\sigma_i^M\) is linearly dependent on lower powers, so \(\sigma_i^0, \sigma_i^1, \ldots, \sigma_i^{M-1}\) are linearly independent and for some polynomial \(p\) of degree \(M - 1\), \(\sigma_i^M = p(\sigma_i)\). Repeated use of this equation allows allows to express also every higher power of \(\sigma_i\) as a linear combination of \(\sigma_i^0, \ldots, \sigma_i^{M-1}\). Hence these must form a basis, and \(M = N\). \(\square\)

2. orthonormal bases of cluster functions

The most straightforward way to obtain the standard orthonormal bases (ONBs) of cluster functions, and one common in the literature, is to treat \(\mathcal{F}\) as a tensor product of Hilbert spaces. Give the vector space \(\mathcal{F}(i)\) of functions of the configuration variable at site \(i\) an inner product according to

\[
\langle h | g \rangle_i = \frac{1}{|\text{Sps}|} \sum_{x_i \in \text{Sps}} h(x_i)g(x_i).
\]

This turns \(\mathcal{F}(i)\) into a Hilbert space. Now regarding \(\mathcal{F}\) as the tensor product \(\bigotimes_i \mathcal{F}(i)\), it acquires the inner product
Typically, the notion of tensor product is explained in terms of orthonormal bases (ONBs). Returning to the linear basis $1, \sigma_1, \ldots, \sigma_{|Sps|}$ of the generic single-site space $\mathcal{F}(\ast)$, construct an ONB by Gram-Schmidt orthonormalization. This gives orthonormal functions $\varphi_k(\sigma_i)$ for $k = 0, 1, \ldots, |Sps| - 1$, where $\varphi_k$ is a degree-$k$ polynomial. All possible products $\prod \varphi_{k(i)}(\sigma_t)$ over all sites then constitutes an ONB for $\mathcal{F}$. Factors of $\varphi_0$ can be ignored, since they equal the constant one, so we can also organize this ONB as all products

$$\Phi^\alpha_I = \prod_{i \in I} \varphi_{\alpha(i)}(\sigma_i),$$

where $I$ runs over sets of sites and $\alpha$ assigns a nonzero index to each site in $I$. We will call the set $I$ the support of $\Phi^\alpha_I$. If $I$ is empty, then there are no factors in the product. The functions $\Phi^\alpha_I$ are often called cluster functions, though sometimes simply basis functions. Our notation is close to that of Ref. [3]. Other common notations are $\Phi_\alpha$, where the support is left implicit in $\alpha$, or $\Pi_\alpha$, suggesting a product. We prefer to make the support explicit and leave the range of $\alpha$ implicit.

### B. Shortcomings

A Hilbert space structure, that is an inner product, has been introduced here, along with a supposedly distinguished kind of orthonormal basis of cluster functions. However, it is unclear where the inner product comes from or in precisely what way the $\Phi^\alpha_I$ are distinguished or are related to the intuitive idea of an $n$-body interaction. Even if the structure is judged reasonable, a deeper explanation is desirable for added flexibility and a clearer idea of the possibility of alternatives.

Certainly, the cluster functions have an arbitrariness coming from the choice of the $\varphi_k$, or of $\sigma_i$. It might seem plausible that this arbitrariness is limited to $\alpha$, so that the sum

$$C_I f := \sum_\alpha \langle \Phi^\alpha_I | f \rangle \Phi^\alpha_I,$$

actually depends only on the Hilbert space structure. If so, it has intrinsic meaning and characterization, but what are they? We claim that it should be considered the “$I$-body” part (as in two-body, three-body, . . .) of $f$. This characterization will emerge Section 5.

There is a generalization of the simple form of the conventional formalism which has been considered in the literature [3, 4, 24]. This is to replace the inner product (2.1) by

$$\langle h | g \rangle \equiv \sum_{x \in Sps} p(x_i) h(x_i) g(x_i),$$

where the weights $p(x_i)$ add to one, and therefore can be interpreted as probabilities. This effectively makes the inner product a derived notion based on the probability distribution $p$. We will go even farther and not even start with a probability distribution on configurations.

Any observable, energy $\mathcal{E}$ for instance, has an expansion over the cluster functions as

$$\mathcal{E}(x) = \sum_{I,\alpha} \langle \Phi^\alpha_I | f \rangle \Phi^\alpha_I(x).$$

It is common to use the term cluster expansion to refer to the construction of an approximation

$$\mathcal{E} \approx \sum_{S \in \text{all } I,\alpha} J^\alpha_I \Phi^\alpha_I$$

based on the energies of a limited sample of configurations. In the case of energy, the $J^\alpha_I$’s are called effective cluster interactions (ECIs). It is reasonable to call the right-hand side of either 2.4 or 2.5 a cluster expansion. However, the conventional approach to determining the model parameters (ECIs) takes no particular notice of the fact that the $\Phi^\alpha_I$’s are orthonormal with respect to a presumably relevant inner product. The methods are the same as would be deployed for expansion in a generic set of functions. Thus, the conventional modeling procedures are not hardly part of cluster expansion formalism proper. This paper takes the perspective that the Hilbert space structure of $\mathcal{F}$ has implications for modeling that should be taken seriously. Section 10 shows how. Modeling done that way can be considered a proper part of the cluster expansion formalism.

### 3. NOTATION AND NOMENCLATURE

Before starting the development, we establish some terminology, notations, conventions, many of which are recorded in Table 1 for convenient reference.

Site is a set of sites. Ordinarily, these are atomic sites in a crystal structure, but until Section 8, no spatial aspect is involved. A configuration $x$ assigns to each site $i$ a species $x_i$ from the species set $\text{Sps}(i)$, so distinct sites may have distinct sets of allowed species. These are typically simply chemical species, but we only need $\text{Sps}(i)$ to be a finite set; it may include information such as charge state, for example. $\text{Cfg}$ is the set of all configurations. Properties of the system are real or complex functions, e.g. $f$, of configuration. We will often call these observables for brevity, and to avoid overworking the word function. The set $\mathcal{F}$ of all observables is a vector space. Because we will be very interested in which particular configuration variables an observable depends on, some elaboration on these simple ideas is needed. For a set $I$ of sites, a partial configuration $x \in \text{Cfg}(I)$ assigns species only to sites in $I$; we also say $x$ is supported on $I$. For a complete configuration, $x \in \text{Cfg}$, its restriction to $I$ is denoted $x_I$, generalizing the notation $x_i$. An $x$
or $y$ without adornment is a complete configuration unless context indicates otherwise. Thus, we have maps $\text{Cfg} \rightarrow \text{Cfg}(I)$ given by restriction. Dually, there are natural linear maps $\mathcal{F}(I) \hookrightarrow \mathcal{F}$ going the other way, from the space of functions of partial configurations supported on $I$ (that is, $\mathcal{F}(I)$) into $\mathcal{F}$. Namely, from $f \in \mathcal{F}(I)$, we obtain the function $x \mapsto f(x_I)$ in $\mathcal{F}$. Since this is injective (one-to-one), we might as well consider $\mathcal{F}(I)$ to be a subspace of $\mathcal{F}$. With that identification, $f$ in $\mathcal{F}(I)$ is also in $\mathcal{F}(J)$ for any superset $J$ of $I$. If $f \in \mathcal{F}(I)$, we say $f$ is supported on $I$, and the support of $f$, supp $f$, is the smallest such $I$. That is, supp $f$ is the set of sites whose species $f$ really depends on.

4. ROUGH IDEAS

This Section begins construction of a new cluster expansion formalism based on a simple notion of best-approximation which will be axiomatized in the next section.

A. Effective and reduced entities

Reduced or effective descriptions, formulations, quantities, Hamiltonians, actions, theories, and so forth are a venerable way to deal with complex systems in physical science. Roughly, the idea is to eliminate dependence on some independent variables, in such a way that the resulting simplified picture of the physical situation is as accurate as possible within the resulting limited level of detail. For instance, classical fluid dynamics describes the motion of fluids on the basis of only macroscopic degrees of freedom, atomic physics usually ignores the internal structure of atomic nuclei. Renormalization group methods are among the most sophisticated exemplars of this general philosophy.

We propose to base the cluster expansion method on this effective/reduced idea. Begin with a configuration function $f \in \mathcal{F}$. Potentially, $f(x)$ depends on all components of $x$; we seek an effective/reduced description of $f$, to be denoted $\mathcal{E}_I f$, which has lost its dependence on variables with indices outside $I$. The intuitive content of “effective/reduced description” in this context is perhaps better captured by the term best approximation, i.e., $\mathcal{E}_I f$ is to be the best — in some sense — possible approximation to $f$ which is supported in $I$. This is still extremely vague. To make it precise, in the next Section we shall write down conditions, that is, axioms, that the operators $\mathcal{E}_I : \mathcal{F} \rightarrow \mathcal{F}(I)$ should satisfy, not as a logical inevitability, but as the most plausible and reasonable filling out of our fuzzy notion of best approximation.

B. A heuristic shortcut

A more heuristic analysis may be a helpful warm-up. Return to the basic question: does $\mathcal{E}(x)$, energy as a function of configuration, have a natural decomposition into building blocks, absent a mechanistic theory? Write $\mathcal{C}_I \mathcal{E}$ for the hoped-for building block associated with $I$. This must be a function only of the subconfiguration in $I$, $\mathcal{C}_I \mathcal{E} \in \mathcal{F}(I)$, and the total energy must be the sum of all the building blocks, that is

$$
\mathcal{E}(x) = \sum_I (\mathcal{C}_I \mathcal{E})(x).
$$

Of course, this is nowhere near enough to determine the $\mathcal{C}_I \mathcal{E}$’s. So, we make a stronger demand that

$$
\mathcal{E}_I \mathcal{E} := \text{“energy in } I \text{”} = \sum_{J \subseteq I} \mathcal{C}_J \mathcal{E},
$$

the energy in $I$ is the sum of all the building blocks in $I$. At first sight, this seems a completely bad move. Asking for the energy in $I$ is asking for a mechanistic theory behind the scenes. Even if we had such, we could not expect it to cleanly assign parts of the energy to specific sets of sites. However, we can rephrase that in a more informational/observational kind of way. First, replace “energy in $I$” by “what the subconfiguration in $I$ tells us about the energy”. This still does not make sense, nor is it even a quantity of energy. With another little shift, we reach the idea of interpreting $\mathcal{E}_I \mathcal{E}(x)$ as the conditional expectation of the total energy, given the configuration in $I$. This does make sense. The good news is that, given all the $\mathcal{E}_I \mathcal{E}$, the preceding display uniquely determines the $\mathcal{C}_I \mathcal{E}$. The bad news is that a probability distribution over configurations must be brought in. On reflection, though,
that seems inevitable. \(C_I E\) could hardly be expected to have nothing to do with the environment of \(I\). We can, however, reduce it to knowing what that environment is like on average.

The key point in the above argument is to define the cluster components \(C_I E\) somewhat indirectly, in terms of the \(E_I E\), because these are easier to get a solid grip on. In the following formal development we avoid a direct invocation of probability by starting from very explicit requirements on the \(E_I\) operators. That we wind up in the same place anyway underlines the naturality of the probabilistic setting.

5. AXIOMS FOR BEST-APPROXIMATION PROJECTORS AND CONSEQUENCES

Section 5A gives the system of axioms A–D. Section 5B introduces cluster components as a derived concept. Section 5C argues informally for accepting them as the only plausible precisification of the idea of effective/reduced functions. Section 5D brings out the implicit probability and Hilbert space structures.

A. Axioms

We postulate a system of operators \(E_I : \mathcal{F} \to \mathcal{F}(I)\), one for each set of sites \(I\). However, for reasons that become clear soon, we omit the subscript from \(E_{\emptyset}\), writing it as simply \(E\). These operators are subject to the following axioms. In Section 6 we will take up a secondary axiom of independent sites.

A. \(E_I : \mathcal{F} \to \mathcal{F}(I)\) is a linear projection onto \(\mathcal{F}(I)\). 
B. \(f \geq 0\) and \(f \neq 0 \Rightarrow Ef > 0\). 
C. \(EE_I = E\). 
D. \(g \in \mathcal{F}(I) \Rightarrow E_I(gf) = gE_I f\).

Since we have no notion of orthogonality yet, saying that \(E_I\) is a projection merely means that applying it twice in succession is the same as applying it once: \(E_IE_I = E_I E_I = E_I\). However, we will soon see that they really are orthogonal projectors. Generally, we omit the composition symbol \(\circ\), except in cases of ambiguity.

B. cluster components and decomposition

The idea of the following definition is that we want a single set of building blocks to build all of the \(E_I f\). In the case of energy, these building blocks should be somehow describable as two-body potentials, three-body potentials, and so forth.

Definition 5.1. The cluster operators 
\[
C_I : \mathcal{F} \to \mathcal{F}(I) 
\] (5.1)
are implicitly defined by
\[
E_I = \sum_{J \subseteq I} C_J. 
\] (5.2)

Applied to a function \(f \in \mathcal{F}\), these produce the cluster components \(C_I f \in \mathcal{F}(I)\). Collectively, the cluster components comprise the cluster decomposition of \(f\).

Anticipating that our axioms will imply that \(E_{\text{site}}\) is the identity on \(\mathcal{F}\), notice that taking \(I \equiv \text{site}\) in (5.2) yields \(f = \sum_J C_J f\), shows that the appellation decomposition is appropriate.

Def. 5.1 is a legitimate definition because it determines the cluster operators (equivalently, cluster components of arbitrary \(f\)) recursively: First, \(C_{\emptyset} f\) is determined outright, since \(\emptyset\) has no proper subsets, hence \(\emptyset f\) gives \(C_{\emptyset} f = E_{\emptyset} f\). Now, to recurse on the subset relation, assume that \(C_J f\) is determined for every proper subset \(J \subset I\). Then, simple rearrangement of the putative definition gives
\[
C_I f = E_I f - \sum_{J \subset I} C_J f. 
\] (5.3)

Since everything on the right-hand side is known, \(C_I f\) is therefore determined by the equation. Inductive arguments of this sort will appear several more times in the development.

In fact, one can write down a formula for the solution of this recursion, using the notation
\[
\varepsilon(I) = (-1)^{|I|} = \begin{cases} 
1 & \text{if } I \text{ has even size} \\
-1 & \text{if } I \text{ has odd size}.
\end{cases} 
\] (5.4)

Proposition 5.1 (Möbius inversion).
\[
C_J = \varepsilon(J) \sum_{K \subseteq J} \varepsilon(K) E_K. 
\] (5.5)

This is an example of Möbius inversion. The Proposition follows from (5.5) in Appendix A where a self-contained exposition is given.

C. Defending the axioms

Now we suggest some informal arguments for axioms A–D. This is where we probe our intuition about what to see what are the crucial features of the idea of best approximation. The axioms are, however, the official rules and there is no more sneaking intuition in by the back door. Of course, we may later decide that we need to try a different axiomatization. That will not happen in this paper, although the following section will supplement A–D with a secondary independence axiom.

Begin with A, which is equivalent to: (i) the range of \(E_I\) is in \(\mathcal{F}(I)\), (ii) \(E_I g = g\) for every \(g \in \mathcal{F}(I)\), and (iii) \(E_I\) is a linear operator. We defer (iii). The argument for
(i) and (ii) is as follows. Suppose $f \in \mathcal{F}(I)$. Then, $E_I f$ means to formally introduce additional variables that $f$ does not actually depend on (one might be the position of Pluto), and then forget them again. It would be absurd for that to result in anything except $f$ itself.

Now, (ii) above means that for any constant $c$, $E_I c f = c E_I f$, and $E_I (f + g) = E_I f + E_I g$. Since there are certainly effective/reduced situations where this fails, such as nontrivial renormalization group calculations, this is not as obvious. However, even the scalar multiple relation fails for these situations, while it is pretty evident that we want that. To be concrete, if the energy for every complete configuration is doubled, then every effective/reduced energy should surely double as well.

This takes care of axiom A. Note a consequence for $E$. A implies that if $f$ is a constant $c$, then the effective/reduced $f$ depending on no variables is simply $f$ itself. With linearity, that is the same as $E1 = 1$.

We argue for B by noting that it is equivalent to a continuity condition, namely,

$$|g(x) - f(x)| < \epsilon \text{ for every } x \Rightarrow |E g - E f| < \epsilon. \quad (5.6)$$

It is easy to see that this implies B. Simply take $g$ constant equal to the average of the minimum and maximum values attained by $f$. In the other direction, suppose $g < f + \epsilon$. By B and A, $E g < E f + \epsilon$. Now use $|g - f| = \max(g - f, f - g)$.

Suppose we arrive at an effective/reduced function by eliminating some variables, and then we eliminate some more. We might expect that the result is the same as if we had eliminated all of them at once. This would be $J \subseteq I \Rightarrow E_J E_I = E_J$. C is a weak version of this for just $J = \emptyset$.

Finally, we turn to D, which is the most awkward to argue for. If $f$ is an interesting function of configuration, say energy, then it is not so difficult to form some idea of what A, B, and C are saying. D, however, contains a second function, $g$. To motivate it, we start with a special case. For $A \in Sps$, define $\chi_A(x)$ to be one if $x_i \in A$, zero otherwise. Now, consider $\chi_A f$. The factor $\chi_A$ acts as a switch, turning $f$ on only when $\sigma_i$ is in $A$. When the switch is on, it is just as though the factor $\chi_A$ were absent. For $I$ containing $i$, we take a configuration $\sigma$ with $\sigma_i \in A$, and ask ourselves what $(E_I f)(\sigma)$ should be. It is supposed to be some sort of effective/reduced description of $\chi_A f$ using the limited data $\sigma_i$. But this data contains definitive proof that the switch is on. In such case, we should agree that the presence of the switch is irrelevant, that is, $E_I (\chi_A f) = E_I f = \chi_A E_I f$ at configurations with $\sigma_i \in A$. Accepting that, since $1 - \chi_A$ is also a switch, we also have $E_I ((1 - \chi_A) f) = (1 - \chi_A) E_I f$ when $\sigma_i \notin A$. By linearity (axiom A), $E_I (\chi_A f) = \chi_A E_I f$ without restriction. Finally, since any function in $\mathcal{F}(I)$ can be expressed as a sum of products of switches, D follows.

The system of axioms A–D might appear to have considerable freedom in the $E_I$. That is illusory. This subsection shows that $E$ (that is, $E_\emptyset$) alone determines a Hilbert space structure on $\mathcal{F}$, and the $E_I$ are then forced to be orthogonal projections onto the subspaces $\mathcal{F}(I)$. This provides a more concrete structure to work with in the following and is a step toward recovering the conventional formulation.

### D. Probabilistic interpretation

As we have already telegraphed, the best-approximation idea fixed by A–D turns out to be probability in disguise. This subsection carefully develops this rephrasing:

**Proposition 5.2.** Axioms A–D are equivalent to

1. $E$ is a (full support) probabilistic expectation operator
2. $E_I$ is the corresponding conditional expectation operator given the subalgebra $\mathcal{F}(I)$ of $\mathcal{F}$.

We also take up the Hilbert space structure induced on $\mathcal{F}$ by $E$ in Section 5D and use it heavily from that point on. Probability notions are not necessary for that, but allow us to talk about it in a second language. The probability distributions we obtain here are formal. Partly for this reason, we somewhat downplay explicitly probabilistic language. In a concrete setting one may have already one or more relevant probability distributions, e.g., Gibbs distributions with varying temperature and chemical potentials. Introducing a cluster expansion in such a situation, the corresponding probability distribution may not correspond to anything “real”. However, pragmatically, $E$ is misguided if it does not provide a sensible reference distribution. A systematic development of general probability theory prioritizing expectation, as we do, is given by Whittle[26].

#### 1. Basic probability

We recall some probability concepts in preparation for rephrasing the axiom system according to Prop. 5.2.

Over a discrete set such as $\text{Cfg}$, a probability distribution is defined simply by assigning a $P(x) \geq 0$ to each $x \in \text{Cfg}$ in such a way that that $\sum_{x \in \text{Cfg}} P(x) = 1$. If $P(x) \neq 0$ for every $x$, we say $P$ has full support. This is the case we are interested in. We overload the notation so that $P(A) = \sum_{x \in A} P(x)$ is the probability of a subset $A$ of $\text{Cfg}$, an event in probability parlance. We entend the notation once more to expectatons of functions on the outcome space $\text{Cfg}$ (random variables):

$$P(f) = \sum_{x \in \text{Cfg}} P(x) f(x). \quad (5.7)$$

This is unorthodox notation. A more usual notation for the expectation of $f$ would be something like $E(f)$. We will be able to revert to that notation as soon as the identification promised by Prop. 5.2 is established. The abuse of notation committed in (5.7) is justified by the fact that
the probability of a set or single configuration equals the expectation of its characteristic function. Defining
\[ \delta_x(y) = \begin{cases} 1 & y \equiv x \\ 0 & \text{otherwise} \end{cases} \] (5.8)
we can link the two uses of ‘P’ in a different way:
\[ P(x) = P(\delta_x). \] (5.9)
This simple idea of reading an element of \( \text{Cfg} \) as an element of \( \mathcal{F} \) will be surprisingly useful.

2. \( \mathbb{E} \) is an expectation operator

The first step in demonstrating Prop. 5.2 is now simple. \( \mathbb{E} \) obeying \( \mathbf{A} \) and \( \mathbf{B} \) determines a probability \( P \) on \( \text{Cfg} \) and vice versa. Using the abuse of language already introduced, we write simply, albeit somewhat cryptically,
\[ \mathbb{E} = P. \] (5.10)
That is, given \( \mathbb{E} \), defining \( P(x) = \mathbb{E} \delta_x \) gives a full-support probability distribution: \( P(x) > 0 \) by \( \mathbf{B} \), while \( \sum_x P(x) = \mathbb{E}(\sum \delta_x) = \mathbb{E}1 = 1 \) by \( \mathbf{A} \). Going the other direction, if \( \mathbb{E}f \) is defined via \( (5.7) \), it satisfies \( \mathbf{A} \) because \( P(f) \) is a constant and \( P(1) = 1 \), and \( \mathbb{E} \) satisfies \( \mathbf{B} \) because \( P(x) > 0 \). What \( \mathbf{C} \) and \( \mathbf{D} \) say in the case \( I \equiv \emptyset \) is already subsumed by \( \mathbf{A} \).

3. \( \mathbb{E} \) induces a Hilbert space structure

The probability \( P \) also naturally equips \( \mathcal{F} \) with a Hilbert space structure, through the definition
\[ \langle f \mid g \rangle := \sum_{x \in \text{Cfg}} P(x)f(x)^*g(x) = P(f^*g) \] (5.11)
of a positive-definite inner product. Here, \( f(x)^* \) is the complex conjugate of \( f(x) \). We are allowing complex functions, although that is not necessary. Many probabilistic concepts can be recast in Hilbert space terms through the use of this inner product. The expectation of \( f \) is \( P(f) = \langle 1 \mid f \rangle \), where \( 1 \) denotes the constant function with value 1, while its variance is \( \text{var } f = P(|f|^2) - |P(f)|^2 = ||f||^2 - |\langle 1 \mid f \rangle|^2 \).

Finally, consider a vector subspace \( \mathcal{V} \) of \( \mathcal{F} \) containing the constants, and the orthogonal projection \( \mathcal{V} \) onto \( \mathcal{V} \). Then, \( \mathcal{V}f \) is a best approximation to \( f \) by a random variable from \( \mathcal{V} \) in the following probabilistic sense: the residual \( f - \mathcal{V}f \) has mean zero, and as \( g \) ranges over \( \mathcal{V} \), \( \text{var } (f - g) \) is uniquely minimized at \( g = \mathcal{V}f \). Suppose further that \( \mathcal{V} \) is not only a vector subspace but also an algebra, which means it is closed under multiplication \( (g, h \in \mathcal{V} \Rightarrow gh \in \mathcal{V}) \). With a finite outcome space \( \text{Cfg} \), this implies that \( \mathcal{V} \) contains every function that can be made from its elements by any means (see discussion of \( \sigma \) in Sec. 2.4). In such a case, probabilists call \( \mathcal{V}f \) the conditional expectation of \( f \) given \( \mathcal{V} \).

4. The \( \mathbb{E}_I \) are orthogonal (self-adjoint) projections

Proof of Prop. 5.2 is finished by showing that the orthogonal projection onto \( \mathcal{F}(I) \), and nothing else, can play the role of \( \mathbb{E}_I \) in \( \mathbf{A} - \mathbf{D} \).

Lemma 5.3. Axioms \( \mathbf{A} - \mathbf{D} \) imply that \( \mathbb{E}_I \) is an orthogonal, that is, self-adjoint, projection onto \( \mathcal{F}(I) \).

Proof. Since \( \mathbf{A} \) already says \( \mathbb{E}_I \) is a linear projection, only self-adjointness need be shown. On the one hand,
\[ \langle \mathbb{E}_I g \mid f \rangle \overset{5.11}{=} \mathbb{E}(f \mathbb{E}_I g^*) = \mathbb{E}(f \mathbb{E}_I g^*) = \mathbb{E}(f \mathbb{E}_I g^*). \] (5.12)
On the other hand,
\[ \langle g \mid \mathbb{E}_I f \rangle = \langle \mathbb{E}_I f \mid g \rangle^* \overset{5.12}{=} \mathbb{E}(f \mathbb{E}_I g^*). \]

Going the other way, it needs to be checked that orthogonal projections actually satisfy all the axioms. All but \( \mathbf{D} \) are immediate. Thus, assume \( g \in \mathcal{F}(I) \). Then, since \( \mathcal{F}(I) \) is closed under products (it is an algebra), \( \mathbb{E}_I (gf) \) and \( \mathbb{E}_I f \) are also in \( \mathcal{F}(I) \), and to show they are equal only requires testing their inner products with an arbitrary \( h \in \mathcal{F}(I) \):
\[ \langle h \mid \mathbb{E}_I (gf) \rangle = \langle \mathbb{E}_I h \mid gf \rangle \overset{\text{h \in } \mathcal{F}(I)}{=} \langle h \mid gf \rangle 5.11 \langle g^* h \mid f \rangle \]
\[ = \langle g^* h \mid \mathbb{E}_I f \rangle \overset{\text{5.11}}{=} \langle h \mid g \mathbb{E}_I f \rangle. \]
The first equality following by orthogonality of \( \mathbb{E}_I \).

5. Marginalization

The last main probability concept to bring in is marginal distributions. Consider a system of projections \( \mathbb{E}_I \) satisfying \( \mathbf{A} - \mathbf{D} \). Fix a set \( L \) of sites. Restricted to \( \mathcal{F}(L) \) — understood as functions on \( \text{Cfg}(L) \) — the \( \mathbb{E}_I \) with \( I \subseteq L \) still satisfy the axioms. Therefore, just as in the previous few subsections, there is an induced Hilbert space structure on \( \mathcal{F}(L) \) and corresponding probability distribution \( P_L : \text{Cfg}(L) \to (0, 1) \). This is a marginal distribution. For \( f, g \in \mathcal{F}(L) \),
\[ \mathbb{E}(f^*g) = \langle f \mid g \rangle = \sum_{x \in \text{Cfg}(L)} P_L(x)f(x)^*g(x). \] (5.13)
Evaluate the marginal probability of the partial configuration \( x \in \text{Cfg}(L) \) as
\[ P_L(x) = \mathbb{E} \delta_x = \sum_{y \in \text{Cfg}} P(y) \delta_x(y) = \sum_{y \in \text{Cfg}} P(y). \] (5.14)
Like any other observable in $\mathcal{F}(L)$, $\delta_x$ simply ignores species at the other sites when interpreted in $\mathcal{F}$:

$$\delta_x(y) = \begin{cases} 1 & \forall i \in \text{dom } x, \ y_i = x_i \\ 0 & \text{otherwise} \end{cases} \quad (5.15)$$

Essentially, this identifies the partial configuration $x$ with the set of configurations which match it. We could write

$$P_I(x) = P(x), \quad (5.16)$$

where on the right-hand side, ‘$x$’ denotes that set and $P(x)$ is just its ordinary probability under $P$. Thus, the $I$ subscript above is dispensible. However, in the case where $L$ is a single site, we will do so, and with the special notation

$$p_i \equiv P_{\{i\}}. \quad (5.17)$$

Suppose, for example, that Cu (copper) is in $\text{Sps}(i)$. Then

$$p_i(\text{Cu}) = P_{\{i\}}(\text{Cu at } i) = P(\text{Cu at } i). \quad (5.18)$$

The point here is that “Cu at $i$” is an acceptable partial configuration (equivalently, set of configurations), so $P(\text{Cu at } i)$ makes sense. But $P(\text{Cu})$ does not. Somewhere, it must be recorded that the magnesium ion under discussion is at site $i$. In this special case, we choose to retain it as a subscript on $p_i$ (lower-case, note). Then, we can consider $p_i$ as simply a probability distribution over the set $\text{Sps}(i)$.

### 6. Internalizing evaluation, or, hats

The operators $E_I$ fit very naturally into the Hilbert space structure of $\mathcal{F}$. We can further rephrase even evaluation of an observable at a configuration as an inner product. This will be very useful. In Sec. 10 we represent models of $f$ based on a sampling in terms of the sampled configurations, cast as observables. Also, in the presence of symmetry, one wants to average a configuration over a group action. The most reasonable way to make sense of that is to realize the configuration as a special sort of observable.

Actually, the first step has already been taken in associating the observable $\delta_x$ with configuration $x$. The only problem is that the normalization is not quite right for our purposes. Therefore, we introduce an alternative normalization and notation:

$$\hat{x} := \frac{1}{P(x)}\delta_x. \quad (5.19)$$

Here, $x$ may be a partial configuration. For a complete configuration, we have immediately

$$\langle \hat{x} | f \rangle = \frac{1}{P(x)} \sum_{y \in \text{C} \text{fg}} P(y)\delta_x(y)f(y) = f(y), \quad (5.20)$$

since $\delta_x$ reduces the sum to a single term. The generalization to partial configurations follows by the technique of relativization employed in Sec. 5D5.

### Proposition 5.4.

For a partial configuration $x_I \in \text{C} \text{fg}(I)$,

$$\langle E_I f | x_I \rangle = \langle \hat{x_I} | f \rangle. \quad (5.21)$$

Equivalently, for $x \in \text{C} \text{fg}$,

$$E_I \hat{x} = \hat{x_I}. \quad (5.22)$$

In particular, in the important special case of a complete configuration $x$,

$$f(x) = \langle \hat{x} | f \rangle. \quad (5.23)$$

### 6. INDEPENDENT SITES

With the aid of an additional condition, $\text{IS}$, this section makes contact with the conventional cluster functions, among other things. The conventional formalism corresponds to a probability distribution $P$ under which distinct sites are independent. This notion is explained in Section 6A Proposition 6.1 in Section 6C gives a large number of conditions equivalent to $\text{IS}$. Some of these conditions are interesting because they can be used to argue in the spirit of Section 5C for the naturality of $\text{IS}$. This is done in Section 6D Finally, Section 6E shows that the cluster functions of the conventional formalism give ONBs when $\text{IS}$ holds. Section 6F constructs the cluster components of configurations which can provide a useful and convenient alternative to the cluster functions.

#### A. A secondary axiom

We need to define the relevant concept before stating the axiom.

**Definition 6.1 (Independence).** Subalgebras $A$ and $B$ of $\mathcal{F}$ are independent (more precisely: $E$-independent) if

$$\langle f | g \rangle = \langle f | 1 \rangle \langle 1 | g \rangle \text{ for all } f \in A, g \in B.$$

Here, then, is the site-independence axiom:

**IS.** $\mathcal{F}(I)$ and $\mathcal{F}(J)$ are independent whenever $I$ and $J$ are disjoint.

#### B. Elementary products

Elementary products are a class of observables which are especially important and useful when $\text{IS}$ holds.

**Definition 6.2.** The observable $f \in \mathcal{F}$ is an elementary product if it can be written as a product

$$f = \prod_i f_i, \quad f_i \in \mathcal{F}(i) \quad (6.1)$$

of observables, each depending on only a single configuration component. Equivalently, the set of elementary products is

$$\prod_i \mathcal{F}(i). \quad (6.2)$$

9
It is extremely important that \( \mathcal{F} \) is linearly spanned by the elementary products: \( \mathcal{F} = \text{span} \prod_i \mathcal{F}(i) \). This means that every \( f \in \mathcal{F} \) is expressible as a linear combination of elementary products, but such expression is not always unique. A basis, generally many, can always be extracted from a spanning set. The \( \delta \) observables \([5,8]\), for example, provide an orthogonal basis of elementary products. \( \delta_x \) is an elementary product since it can be written as \( \prod_i \delta_{x(i)} \), or, in a more explicit notation,

\[
\delta_x(y) = \prod_i \delta(x_i, y_i). \tag{6.3}
\]

Since an arbitrary \( f \in \mathcal{F} \) can be written
\[
f = \sum_{x \in \mathcal{C}_f} f(x) \delta_x, \tag{6.4}
\]
the \( \delta \) observables span \( \mathcal{F} \). Finally, the definition \( \langle \delta_x | \delta_y \rangle = \sum_x P(x') \delta_x(x') \delta_y(x') \) shows that they are orthogonal, for any \( P \), and therefore form an orthogonal basis.

An elementary product contains exactly one factor from \( \mathcal{F}(i) \) for each \( i \), however some of these factors could simply be constants. This motivates consideration of the sets (not vector spaces) of observables

\[
\prod_{i \in I} \mathcal{F}^0(i). \tag{6.5}
\]

\( \mathcal{F}^0(i) \) is defined to be the orthogonal complement of constants in \( \mathcal{F}(i) \):
\[
\mathcal{F}(i) = \mathbb{C} \oplus \mathcal{F}^0(i). \tag{6.6}
\]

To see why these are important, consider an elementary product \( f = \prod_i f_i \) and write \( f_i \) as \( E f_i + (f_i - E f_i) \), a sum of a constant and a term in \( \mathcal{F}^0(i) \). Expand the product as

\[
f = \prod_i [E f_i + (f_i - E f_i)] = \sum \left( \prod_{j \not\in I} E f_j \right) \prod_{i \in I} (f_i - E f_i). \tag{6.6}
\]

Thus, any elementary product decomposes into a sum of scalar multiples of observables in \( \prod_{i \in I} \mathcal{F}^0(i) \) for varying \( I \). when \( \text{IS} \) holds, these latter are the cluster components \( C_I f \), as will be shown.

C. Many equivalents of independent sites

Proposition 6.1 below lists a large number of conditions equivalent to the proposed axiom \( \text{IS} \). Proof is deferred to Section 6E. Some of the conditions are simple translations from one mode of expression to another, as for example with (b) and (c). Some are important for practical computations, such as (b), (c), (h), and especially (i). The last of these leads directly to useful formulas and connects to the conventional formulation. Others seem of more theoretical interest, such as (f) and (g). The arguments for axiomatic status of \( \text{IS} \) are based on (j) and (g).

**Proposition 6.1.** The following are equivalent. In these, \( I \) and \( J \) implicitly run over all subsets of Site, and \( x \) over all configurations. \( \prod f_i \) is a generic elementary product.

(a) \( \text{IS} \)

(b) With one-site marginal distribution \( p_i \) as in (5.17),
\[
P(x) = \prod_i p_i(x_i). \tag{6.7}
\]

(c) \( \langle \prod f_i | \prod g_j \rangle = \prod_i \langle f_i | g_i \rangle \)

(d) \( E_I (\prod f_j) = (\prod_{j \not\in I} E f_j) \prod_{i \in I} f_i \)

(e) \( I \cap J = \emptyset \Rightarrow E_I E_J = E \)

(f) \( E_I E_J = E_{I \cap J} \)

(g) \( I \not\subseteq J \Rightarrow E_I C_J = 0 \)

(h) \( C_I C_J = \begin{cases} C_I & I = J \\ 0 & I \not= J \end{cases} \)

(i) \( C_I (\prod f_k) = (\prod_{j \not\in I} E f_j) \prod_{i \in I} (f_i - E f_i) \)

(j) \( C(I) = \text{span} \prod_{i \in I} \mathcal{F}^0(i), \) where \( \mathcal{F}^0(i) \) is the orthogonal complement of the constants in \( \mathcal{F}(i) \).

(k) \( \mathcal{F} = \bigoplus_I C(I), \) this being an ordinary direct sum (that is, \( C(I) \perp C(J) \) is not assumed).

D. Why take IS as an axiom?

If the \( E \)-operator we are dealing with happens to satisfy \( \text{IS} \), computational simplifications ensue, as one would expect from conditions (b) or (c) of the Proposition. This is reason to have a preference for such \( E \)-operators. However, just as we did for axioms A–D in Section 5C, we want to find the best arguments toward the conclusion that our semi-choate notion of approximation structure really requires that the condition \( \text{IS} \) be imposed.

The first argument is that the operators \( E \) should always lead to a simplification, in the sense of reducing the support of functions. That is, \( \text{supp} E I f \subseteq \text{supp} f \), whatever \( f \) and \( I \) may be. This translates, almost immediately to \( E_I E_J \subseteq E_{I \cap J} \), condition (f) of Prop. 6.1. While it has some surface plausibility, this argument is not really very compelling. After all, some effective/reduced descriptions of the sorts invoked as motivation in Sec. 4A introduce degrees of freedom which do not even exist in the original description.
Better arguments are to be found by turning attention to cluster components. We did not try to axiomatize the cluster operators directly because it is too difficult to see a priori what a full set of such axioms should be. Let us return to the idea that the cluster operators $C_I$ should extract the building blocks of an observable. Inherent in that idea is some degree of atomicity, stability, and uniqueness. Hence, it might well be reasonable to demand that if we start with a building block $C_I f$, for some particular $I$ and $f$, and extract its building blocks, we simply get it back and nothing else. That would mean $C_I C_I f = C_I f$, while $C_J C_I f = 0$ for $J \neq I$. In other words, Prop. 6.1(h).

Another argument of similar flavor is based on uniqueness. Def. 6.1 of the cluster operators provide a unique expression of any observable as a sum of building blocks in the cluster subspaces $C(I)$. However, that does not mean that there are not others. Indeed, that is exactly what the direct sum condition (j) of the Proposition imposes. If it does not hold, then there are other expressions (not picked out by the $C_I$ operators) of an observable as such a sum.

E. Proof of Proposition 6.1

The equivalence of (a) – (e) follows by recursion from the following Lemma, elementary proof of which is omitted.

**Lemma 6.2.** For given $I$, $J$, the following conditions are equivalent.

1. $I$ and $J$ are independent
2. $E_I E_J = E$
3. $E_I(f_I f_J) = (E f_J) f_I$ whenever $\text{supp } f_I \subseteq I$ and $\text{supp } f_J \subseteq J$.
4. $P(x_{I \cup J}) = P(x_I) P(x_J)$
5. $\langle f_I g_I f_J g_J \rangle = \langle f_I g_I \rangle \langle f_J g_J \rangle$ whenever $\text{supp } f_I \subseteq I$ and so forth.

(e) $\Leftrightarrow$ (f): Implication right-to-left is immediate. For the reverse, it suffices to show $\text{supp } f \subseteq J \Rightarrow E_I f = E_{I \cap J}$. Since products span, we may even suppose $f = f_{J \cap I} f_{J \setminus I}$, where the factors are supported in $J \cap I$ and $J \setminus I$, respectively. Thus, compute

$$E_I(f_{J \cap I} f_{J \setminus I}) = f_{J \cap I} (E_I f_{J \setminus I}) \overset{(d)}{=} f_{J \cap I} (E f_{J \setminus I}),$$

and note that starting with $E_{I \cap J}$ instead of $E_I$ yields the same result by the same reasoning.

(f) $\Rightarrow$ (g):

$$E_J C_J = E_I \left( E_J - \sum_{K \subset J} C_K \right) \overset{(e)}{=} E_{I \cap J} - \sum_{K \subset J} E_I C_K.$$

Inductively, we assume the conclusion $E_I C_K = 0$ for $K \subset J$, so

$$\sum_{K \subset J} E_I C_K = \sum_{K \subset J} E_I C_K = E_I \sum_{K \subset J} C_K = E_I E_{I \cap J} = E_{I \cap J}.$$

$(g) \Rightarrow (h)$: By Möbius inversion formula 5.5, $J \not\subseteq I \Rightarrow C_I C_J = 0$ follows from (g). We will appeal to this shortly.

Now, The cases $J \not\subseteq I$, $I \subseteq J$, and $J = I$ are mutually exclusive and exhaustive. The first case is just the assumption. If $J \subseteq I$, then $I \not\subseteq J$, so $C_J C_I = 0$. Taking the adjoint of each side of this equation yields $C_J C_I = 0$. For the final case, consider

$$C_I^2 = \left( E_I - \sum_{J \subset I} C_J \right) C_I = E_I C_I - \sum_{J \subset I} C_J C_I$$

The first term in the final expression is $C_I$, while in the sum, each term $C_J C_I = 0$ by the second case.

(h) $\Rightarrow$ (e): Use 6.1(h) to express $E_I$ and $E_J$ in terms of cluster operators. Applying (g), only one term $(E)$ in the expanded product survives.

(d) $\Rightarrow$ (i): Use the definition 5.1 and induction.

$$E_I \prod_{i \in I} f_i = c \prod_{i \in I} f_i$$

$$\prod_{i \in I} f_i = \prod_{i \in I} (f_i - E f_i) + \sum_{J \subset I} \left( \prod_{i \in J} E f_i \right) \prod_{j \in J} (f_j - E f_j)$$

The first uses the expansion in (6.6), and the second line follows by induction hypothesis. Now apply Def. 6.1.

(i) $\Rightarrow$ (j): $C_I$ maps an elementary product into

$$\prod_{i \in I} F^0(i),$$

so $C(I) \subseteq \text{span} \prod_{i \in I} F^0(i)$. On the other hand, (h) also shows that every member of $\prod_{i \in I} F^0(i)$ is fixed by $C_I$, hence is in $C(I)$.

(j) $\Rightarrow$ (k): Define $P^\times(x) = \prod_i p_i(x_i)$. We are not assuming that $P^\times$ is the same as $P$. They just have the same one-site marginals. However, $P^\times$ provides a legitimate inner product $\langle \cdot | \cdot \rangle_{P^\times}$ on $F$, as in Section 5.3, and we may apply the already established equivalence (h) $\Leftrightarrow$ (d), showing that the $C(I)$ are orthogonal under the new inner product. Therefore, they are certainly linearly independent, a statement which does not involve any inner product at all. Thus, $F = \bigoplus C(I)$.

(k) $\Rightarrow$ (h): The tactic is to show that the projection operators onto the components of the assumed direct sum, temporarily denoted $\bar{C}_I$, satisfy the defining conditions of the cluster operators, and therefore are equal to them. This suffices because such a system of projections always obeys the equations of (h).
Now, \( F(I) = \sum_{J \subseteq I} C(J) \), but by assumption the \( C(J) \) subspaces are linearly independent, hence \( F(I) = \bigoplus_{J \subseteq I} C(J) \). Therefore \( E_I f = \tilde{C}_I f + \sum_{J \subseteq I} \tilde{C}_J f \).

By induction hypothesis, \( \tilde{C}_I f = C_I f \) in the sum. Removing those tildes then shows that \( \tilde{C}_I f = C_I f \) by (5.3).

F. Conventional cluster functions

The so-called cluster functions of the conventional formalism (or a minor generalization thereof) provide ONBs of the cluster subspaces \( C(I) \) precisely when IS holds. With Prop. 6.1 in hand, the demonstration is very straightforward. For each site \( i \), let \( \{ \phi^\alpha_i \} \) be an ONB of \( F^0(i) \). For each \( I \), define

\[
\Phi^\alpha_I := \prod_{i \in I} \phi^\alpha_i \tag{6.8}
\]

for every collection of indices \( \alpha = (\alpha_i \mid i \in I) \). This construction is just like in Section 2 except that we are not using the encoding function \( \sigma_\cdot \). Then, the \( \Phi^\alpha_I \) comprise an ONB of \( C(I) \). This claim follows from Prop. 6.1 (c.k). It is also clear from Prop. 6.1 (c.k) that we cannot have such a factorized basis unless site independence holds.

G. Cluster components of a configuration

With IS, tidy formulas are also available for the cluster components of observables (5.3). That such observables provide an orthogonal basis of \( F \) indicates the theoretical interest in such formulas, but are they practical? In fact, as discussed in Section 10 information about non-trivial observables generally comes from sampling. In such circumstances, it is entirely natural and efficient to build models/approximations of such observables by use of hat-functions.

Even without IS, \( \delta_x \) factorizes as \( \prod \delta_{\tilde{x}_i} \). With IS, the probability \( P(x) \) also factorizes, resulting in

\[
\tilde{x} = \prod_i \tilde{x}_i. \tag{6.9}
\]

Here we use the abbreviation

\[
\tilde{x}_i \equiv \tilde{x}_{\{i\}}. \tag{6.10}
\]

\( \{i\} \) being a set of sites with one member. Now, according to Prop. 6.1 (d) and (i)

\[
E_I \tilde{x} = \tilde{x}_I = \prod_{i \in I} \tilde{x}_i \tag{6.11}
\]

\[
C_I \tilde{x} = \prod_{i \in I} (\tilde{x}_i - 1). \tag{6.12}
\]

Formula (6.12) is very useful, because it provides practical means to extract cluster components without need for conventional cluster functions. For a more explicit formula, notice that

\[
\langle \tilde{x}_i - 1 \mid y_i \rangle = \begin{cases} p_i(x_i)^{-1} - 1, & x_i = y_i \\ -1, & \text{otherwise} \end{cases}
\]

and therefore

\[
(C_I \tilde{x})(y) = (-1)^{|I|} \prod_{i \in I, y_i = x_i} \left(1 - \frac{1}{p_i(x_i)}\right). \tag{6.13}
\]

7. TAKING STOCK

This Section recapitulates the framework we have assembled and relates it to what follows.

A. Assessing the framework

We started with the problem of breaking arbitrary observables into natural building blocks. While the conventional cluster expansion offers a sort of building block, the cluster functions \( \Phi^\alpha_I \), they depend on some arbitrary choices, and are not adapted to the observable of interest. To improve on the situation, we sought general principles, and wrote axioms for the best-approximation operators \( E_I \). Cluster operators \( C_I \) are obtained from those in a very natural and simple way (Def. 11). Our axioms turned out to be equivalent to requesting a probability distribution over configurations, with the \( E_I \)'s identified as conditional expectation operators. Thus we succeeded in imposing a reasonable amount — and a familiar sort — of structure. To obtain the sort of Hilbert space structure on \( F \) assumed by the conventional cluster expansion requires an additional condition of independent sites. We suggested arguments that IS should be considered an axiom, though the case is not as clear as for A–D.

With that, we arrive at a Hilbert space structure which is the same as presumed by conventional cluster expansion with a variable probability distribution of species. What, then, has been gained? First, we have characterized cluster subspaces \( C(I) \) and cluster components \( C_I f \) intrinsically. This helps insulate even the conventional cluster formalism from the charge of being meaningless curve-fitting. At the same time, we have identified a generalization, obtained by dropping the condition IS. Should an alternative way of decomposing functions of configuration into building blocks be proposed, the abstract formulation puts us in a better position for determining whether it is essentially the same, and even what the prospects are for the existence of such an alternative.

B. On the rest of the paper

As we move toward more applied concerns in the remainder of the paper, we have a flexible set of tools for
working with $\mathcal{F}$ as a Hilbert space: the critical ortho-

gonal direct sum decomposition $\mathcal{F} = \bigoplus C(I) [\text{Prop. 6.1(i)}]$, and a means \[6.1\] to implement the $C_I$ projectors. This puts us in a position to formulate algorithms in an intrin-
sic manner, and possibly find alternative implementations which are more efficient than the use of cluster-function

ONBs. The formula of Prop. \[6.1\] for cluster components of elementary functions, and especially \[6.12\] for hat-functions, hint at such possibilities.

The major application of cluster expansion ideas in materials science is to construction of models of an un-

iform and simple notation for the natural action on any

class of object, i.e., the specific action is determined by

ter components. For tensor observables, elastic, optical,

piezoelectric properties, and so on, the situation is

more complicated since the values of such observables

also transform nontrivially under the space group (rot-

ations, in particular). So far, we have dealt only with scalar-valued observables, namely those in $\mathcal{F}$. Beginning

in this section, we widen the scope to include tensor-valued observables, treating the former as a special case of the latter.

This section extends the framework from $\mathcal{F}$ to $\mathcal{V} \otimes \mathcal{F}$, where $\mathcal{V}$ is a space of tensor values, and shows how symmetry can be effectively exploited. We have been using the term observable in a very broad sense which would include, for example, the $xx$ component of the dielectric constant, or the density of chains of three adjacent Co ions along the $x$ direction. Assuming the $x$-axis is not invariant under our symmetry group $G$, these kinds of observables are not of interest to us as they involve an explicit breaking of the symmetry. We are here really interested only in $G$-equivariant observables. For scalar-valued observables, this is genuine invariance, whereas for a tensor-valued observable, it means we get the same thing if we rotate the tensor value of the observable or rotate the structure and then evaluate the observable.

Proposition 5.3 shows how the subspace $Av \mathcal{V} \otimes \mathcal{F}$ of $G$

equivariant $\mathcal{V}$-valued observables can be “compressed”. For convenient reference, most of the new notations we introduce here are listed in Table II.

### 8. Symmetry

| $G$ | a symmetry group of $E$ |
| $T, T', \ldots$ | element of $G$, site permutation |
| $\mathcal{V}$ | space of “tensors” |
| $Av$ | group averaging operator |
| $\rho$ | representation of $G$ on $\mathcal{V}$ |
| $\mathcal{C}$ | “is $G$-related to” |
| $GI$ | $G$-orbit of $I$, namely $\{J | J \mathcal{C} I\}$ |
| $G_I$ | stabilizer subgroup of $I$ |
| $\mathcal{CSh}$ | cluster-shapes ($G$-orbits of clusters) |
| $C[T]$ | $\bigoplus(C(J) | J \in T\}$ |
| $C_I$ | projector onto $C[I]$ |

| TABLE II. More notation. The notations for $G$-equivalence, orbit, and stabilizer subgroup may be used with other classes of entities than sets of sites. |

To this point, Site has been merely a set of sites with no

structure. In at least the usual applications, however, the

sites are ionic positions in some crystal structure. Sym-

metry considerations therefore emerge, which are very

important for the application of cluster expansion meth-

ods.

For many observables, such as energy, whenever $I$ and

$J$ are related by a space group operation of the par-

ent crystal structure, cluster component $C_I f$ is the same

function of partial configuration on $I$ as $C_J f$ is of par-

tial configuration on $J$. This greatly reduces the compu-

tational burden of finding approximations of such clus-

ter observables. Hence, for some observables, such as energy, whenever $I$ and $J$ are related by a space group operation of the parent crystal structure, cluster component $C_I f$ is the same function of partial configuration on $I$ as $C_J f$ is of partial configuration on $J$. This greatly reduces the computational burden of finding approximations of such cluster components. For tensor observables, elastic, optical, piezoelectric properties, and so on, the situation is more complicated since the values of such observables also transform nontrivially under the space group (rotations, in particular). So far, we have dealt only with scalar-valued observables, namely those in $\mathcal{F}$. Beginning in this section, we widen the scope to include tensor-valued observables, treating the former as a special case of the latter.

This section extends the framework from $\mathcal{F}$ to $\mathcal{V} \otimes \mathcal{F}$, where $\mathcal{V}$ is a space of tensor values, and shows how symmetry can be effectively exploited. We have been using the term observable in a very broad sense which would include, for example, the $xx$ component of the dielectric constant, or the density of chains of three adjacent Co ions along the $x$ direction. Assuming the $x$-axis is not invariant under our symmetry group $G$, these kinds of observables are not of interest to us as they involve an explicit breaking of the symmetry. We are here really interested only in $G$-equivariant observables. For scalar-valued observables, this is genuine invariance, whereas for a tensor-valued observable, it means we get the same thing if we rotate the tensor value of the observable or rotate the structure and then evaluate the observable. Proposition 5.3 shows how the subspace $Av \mathcal{V} \otimes \mathcal{F}$ of $G$-equivariant $\mathcal{V}$-valued observables can be “compressed”. For convenient reference, most of the new notations we introduce here are listed in Table II.

| A. Permutation actions |

Let $T$ be a site permutation. We consider briefly the natural action of $T$ on various types of entity. Permu-
tations defined by space group operations are most im-
portant. However, that restriction is irrelevant to the prelimi-

nary considerations here. By its nature, $T$ has an action on sites $i \leftrightarrow T_i$ with inverse $T^{-1}$. This extends to clusters by $TI = \{T_j | j \in I\}$ and to configurations as $T \{x\}_i = x_{T^{-1}i}$. As here, we generally use a uniform and simple notation for the natural action on any type of object, i.e., the specific action is determined by context. The pattern for configurations, which are func-
tions on Site, continues to higher types. The action of $T$ on $f \in \mathcal{F} = (C_{fg} \to \mathbb{C})$ is given by $(Tf)(x) = f(T^{-1}x)$. For a map $\mathcal{F} \to \mathcal{F}$, such as $E_I$ or $C_I$, matters are a little different because $T$ has a natural action on the target space as well as the domain: $(TC_I)(f) = T(C_I(T^{-1}f))$, or $(TC_I) = T \circ C_I \circ T^{-1}$. If we do not indicate explicitly on which space each $T$ is acting, composition symbols are sometimes needed to disambiguate.

| B. Symmetries of $E$ |

Permutations are not much use to us unless they are symmetries of $E$. $T$ is a symmetry of $E$ if $T \circ E = E \circ T$. In fact, $T \circ E$ is just $E$ if we are dealing with scalar ob-
sobservables, but the more general form is relevant to tensor observables which will be taken up below. In the following basic result about symmetries of \( E \), the equivalence of (a) and (b) is of interest, but normally we only care about (c) and (d) as consequences of symmetry.

**Proposition 8.1.** For a site permutation \( T \), the following conditions are equivalent:

(a) \( T \) is a symmetry of \( E \)

(b) \( T \) is a unitary operator on \( F \):

\[
\langle T f \mid T g \rangle = \langle f \mid g \rangle
\]

(c) for every \( I \), \( E_{T I} = T \circ E_I \circ T^{-1} \)

(d) for every \( I \), \( C_{T I} = T \circ C_I \circ T^{-1} \)

*Proof.* (c) ⇒ (a): \( E = E_f \).

(a) ⇒ (b): \( \langle T f \mid T g \rangle = \langle (T f) T g \rangle = \langle T (f g) \rangle \)

(b) ⇒ (c): \( T (E_f) \) and \( E_{T I} (T f) \) are both in \( F (T I) \). Show they are equal by showing that they have the same inner product with an arbitrary test function \( g \in F (T I) \):

\[
\langle g \mid T E_I f \rangle = \langle T^{-1} g \mid E_I f \rangle = \langle T^{-1} g \mid f \rangle
\]

(d) ⇔ (c): Definition of \( C_I \), Möbius inversion \([5,5] \).

If site independence holds, then \( T \) is a symmetry of \( E \) only if it permutes sites with identical species distributions. Thus, some space group operations of the parent crystal structure may fail to be symmetries of \( E \) if the disordered composition has a spatial modulation. On the other hand, the vast majority of site permutations are not induced by euclidean operations. However, that does not imply that they are totally devoid of practical interest. They are, for example, relevant to the construction of special quasirandom structures (SQS’s\([27] \)), as well as to regular solution models.

**C. Tensor-valued observables**

To this point, we have considered only \( \mathbb{R} \)- or \( \mathbb{C} \)-valued observables, with energy as the primary example of an important one. Many important observables, however, are not of this sort. Polarization and magnetization are vectors, a dielectric tensor is a rank-2 tensor, and, elasticity needs a rank-4 tensor for complete description. The relevant extension of our formalism is straightforward.

Generally, we consider observables taking values in a finite-dimensional Hilbert space \( V \). Such an observable, belonging to the tensor product \( V \otimes F \), can be written as

\[
\tilde{f} = \sum_i \tilde{e}_i \otimes f^i,
\]

where the \( \tilde{e}_i \) are an ONB of \( V \) and \( f^i \in F \). We put an arrow above \( \tilde{f} \) to indicate that it takes values in some non-trivial space, and use the term *tensor observable* rather loosely for all such observables. The inner product on \( V \otimes F \) has two layers:

\[
\langle \tilde{g} \mid \tilde{f} \rangle_{V \otimes F} = \sum_i \langle g^i \mid f^i \rangle_F = \sum_{i \in \mathcal{G}_g} P(x) g^i(x)^* f^i(x).
\]

We have put subscripts on the inner products to be clear. Generally, we will leave these off, relying on the rule that if both entries are in the same space \( (F \) or \( V \otimes F \), then the inner product is the corresponding one, and if one entry is scalar and one tensor, then the inner product is that for \( F \), as in \( \langle g \mid f \rangle = \sum_i \tilde{e}_i (g \mid f^i) \).

Ordinary euclidean tensors fit naturally into this scheme. If \( V \) is rank-2 tensors, for example, then the inner product of two tensors \( A \) and \( B \) is \( \sum A_{ij} B_{ij} \), i.e., the matrix entries are components relative to an orthonormal basis. Rotations and reflections respect this orthonormality.

Elements of the space group of a parent crystal structure are, or (perhaps a better word) induce site permutations. If \( T \) is such an element, then it acts on a tensor observable as \( T (\sum_i \tilde{e}_i f^i) = \sum_i (T \tilde{e}_i) T f^i \), where \( f^i \mapsto T f^i \) is just the permutation action from Section \( 8 A \). For us, \( T \) is not considered a symmetry operation unless it is also a symmetry of \( E \). This identifies a subgroup of the space group that will be denoted \( G \) henceforth.

We do not give the action of \( T \) on elements of \( F \) a special name, but it is sometimes helpful to write the action of \( G \) on \( V \) as

\[
T \tilde{e}_i = \sum_j \rho_{ji}(T) \tilde{e}_j,
\]

so the \( \rho(T) \) are the representation matrices. This representation is not assumed to be irreducible.

Operators previously introduced on \( F \) are now lifted to \( V \otimes F \) in a natural way: the \( V \) part simply passes through unchanged. For example, if \( v \in V \) and \( f \in F \), \( C_I (v \otimes f) = C_I (v f) = v \otimes (C_I f) \). As just demonstrated, we can also often omit writing the tensor product symbol if its clear what kinds of things the factors (here \( v \) and \( f \)) are.

**D. Averaging over the group**

\( G \) will be a symmetry group for the observables of physical interest. That is, such an observable \( \tilde{f} \) commutes with \( T \in G \), \( \tilde{f} (T x) = T (\tilde{f} (x)) \). This is the same as \( T \tilde{f} = \tilde{f} \), equivariance of \( \tilde{f} \). Thus, averaging over the group \( G \), achieved by the operator

\[
\text{Av}^G := \text{Av} := \frac{1}{|G|} \sum_{T \in G} T
\]

is important. Generally, we omit the superscript when averaging is over \( G \). It will be necessary when averaging over a subgroup such as a stabilizer subgroup (e.g.,
On any vector space on which $G$ is represented unitarily, $\text{Av}$ is an orthogonal projection onto the subspace of $G$-equivariant vectors. Thus, we can move it from one side to the other of an inner product, e.g., $\langle g \mid \text{Av} \hat{f} \rangle = \langle \text{Av} g \mid \hat{f} \rangle$. However, such a switch is illegitimate for $\langle g \mid \text{Av} \hat{f} \rangle$, with $\hat{f} \in \mathcal{V} \otimes \mathcal{F}$, but $g \in \mathcal{F}$. The following Lemma gives a useful formula for such a case.

**Lemma 8.2.** The $G$-average of a separable tensor observable $\vec{e}_i \otimes g = \vec{e}_i g$ is

$$\text{Av}(\vec{e}_i \otimes g) = \frac{1}{|G|} \sum_{T \in G} \rho_{ji}(T)Tg.$$  \hfill (8.5)

From this follows

$$\langle g \mid \text{Av} \hat{f} \rangle = \sum_i \vec{e}_i \langle \vec{e}_i g \mid \text{Av} \hat{f} \rangle = \sum_i \langle \text{Av} \vec{e}_i g \mid \hat{f} \rangle.$$  \hfill (8.6)

**Proof.** Compute

$$\text{Av}(\vec{e}_i g) = \frac{1}{|G|} \sum_{T \in G} T \vec{e}_i \otimes Tg.$$  

Inserting (8.3) and rearranging yields (8.6). For (8.6), insert (8.3) into the final expression of

$$\langle g \mid \text{Av} \hat{f} \rangle = \sum_i \vec{e}_i \langle \vec{e}_i g \mid \text{Av} \hat{f} \rangle = \sum_i \langle \text{Av} \vec{e}_i g \mid \hat{f} \rangle.$$  

\hfill $\Box$

**E. Cluster-shape subspaces**

According to Prop. 8.1(i), $\mathcal{F} = \bigoplus_I \mathcal{C}(I)$ is the orthogonal direct sum of all the cluster subspaces. Tensoring with $\mathcal{V}$ on both sides yields

$$\tilde{\mathcal{F}} = \mathcal{V} \otimes \mathcal{F} = \bigoplus_I \mathcal{V} \otimes \mathcal{C}(I) = \bigoplus_I \mathcal{C}(I).$$

We have here introduced an unobtrusive over-arrow notation to indicate the tensor products, and the point is that this change threads cleanly through the cluster decomposition. Symmetry operations have a more profound effect. According to Prop. 8.1 whenever $T$ is a symmetry of $\mathcal{E}$, $\mathcal{C}(TI) = C_{TI} \tilde{\mathcal{F}} = (T \circ C_I \circ T^{-1}) \tilde{\mathcal{F}} = T(C_I \tilde{\mathcal{F}}) = T \mathcal{C}(I)$. Symmetries permute the cluster subspaces for clusters in the $G$-orbit $GI = \{ J \mid J \overset{\mathcal{G}}{\subseteq} I \}$ of $I$. Here, $J \overset{\mathcal{G}}{\subseteq} I$, read “$J$ is $G$-related to $I$”, means that $J = TI$ for some $T \in G$. We use the term *cluster-shape* for a $G$-orbit of clusters, and denote them by calligraphic letters, e.g. $\mathcal{I}$. So, $I = GI$ for any $J \in \mathcal{I}$.

The natural decomposition $\tilde{\mathcal{F}} = \bigoplus_I \mathcal{C}[I]$ is now into cluster-shape subspaces

$$\mathcal{C}[I] := \bigoplus_{J \in \mathcal{I}} \mathcal{C}(J).$$  \hfill (8.7)

Each cluster-shape subspace $\mathcal{C}[I]$ is collectively $G$-invariant, but individual vectors in it generally are not equivariant. To obtain the subspace of $G$-equivariant vectors, we simply apply the $\text{Av}$ projector:

$$\text{Av} \tilde{\mathcal{F}} = \bigoplus_{I \in \text{ClSh}} \text{Av} \mathcal{C}[I].$$  \hfill (8.8)

Because $G$ acts transitively on the orbit of $I$, a nonzero $G$-equivariant vector $\hat{f}$ in $\mathcal{C}[GI]$ necessarily has nonzero component in each subspace $\mathcal{C}(J)$. However, it can be recovered from just one such component $\mathcal{C}_I \hat{f}$, as shown by the following Lemma. The computational significance of this is that $\mathcal{C}_I \hat{f}$ is a function of fewer variables, namely just the part of the configuration on $I$.

**Proposition 8.3.** The maps in the following diagram

$$\text{Av}^{G_1} \mathcal{C}(I) \xrightarrow{\sqrt{|G|} \text{Av}} \text{Av}^G \mathcal{C}[GI]$$  \hfill (8.9)

are mutually inverse unitary equivalences between the subspace of $G_1$-equivariant vectors in $\mathcal{C}(I)$ and the subspace of $G$-equivariant vectors in $\mathcal{C}[GI]$.

**Proof.** It suffices to show that the linear maps in (8.9) have ranges in the indicated spaces, and are isometric. We drop the superscript from $\text{Av}^G$ now.

Ranges are in the indicated spaces: $T \mathcal{C}(I) \subseteq \mathcal{C}[GI], \forall T \in G$, so $\text{Av} \text{Av}^{G_1} \mathcal{C}(I) \subseteq \mathcal{C}[GI].$

In the other direction, $\text{Av}^{G_1}$ and $\mathcal{C}_I$ commute, so $(\mathcal{C}_I \circ \text{Av})\mathcal{C}[GI] \subseteq (\text{Av}^{G_1} \circ \mathcal{C}_I)\mathcal{C}[GI] = \text{Av}^{G_1} \mathcal{C}(I)$.

Lower arrow in (8.9) is isometric: Take $\hat{f} \in \mathcal{Av}\mathcal{C}[GI]$. Since $J \in GI \Rightarrow \| \mathcal{C}_J \hat{f} \| = \| \mathcal{C}_I \hat{f} \| ,

$$\| \hat{f} \|^2 = \sum_{J \in GI} \| \mathcal{C}_J \hat{f} \|^2 = |GI| \| \mathcal{C}_I \hat{f} \|^2.$$  \hfill (8.10)

Upper arrow in (8.9) is isometric: Take $\hat{f} \in \mathcal{Av}^{G_1} \mathcal{C}(I)$. By (8.10), $\| \text{Av} \hat{f} \|^2 = |GI| \| \mathcal{C}_I (\text{Av} \hat{f}) \|^2$. Combine this with $|GI|^2 \| \mathcal{C}_I (\text{Av} \hat{f}) \|^2 = \| \sum_{T \in GI} T \hat{f} \|^2 = |GI|^2 \| \hat{f} \|^2$.  

\hfill $\Box$

**Lemma 8.3** allows us to effectively replace $\mathcal{Av}\mathcal{C}[I]$, by the much smaller space $\mathcal{Av}^{G_1} \mathcal{C}(I)$ for an arbitrarily selected $I$ in the cluster-shape $\mathcal{I}$. This is advantageous for computation. With reference to the decomposition (8.8), we then have

$$\text{Av} \tilde{\mathcal{F}} \cong \bigoplus_{I \in \text{section}} \text{Av}^{G_1} \mathcal{C}(I).$$  \hfill (8.11)

Here, “section” denotes a selection of one cluster from each cluster-shape.
9. TWO POINTS OF COMPUTATIONAL EFFICIENCY

The following Section will discuss how to construct a model for an observable of interest using pieces belonging to $G$-invariant cluster-shape subspaces $\text{AvC}[GI]$. That will be done without using the conventional cluster functions $\Phi^g_\alpha$. Here, we take a brief detour to demonstrate the computational practicality of our approach. Thus, we suppose that we have a cluster observable, i.e., a vector in $\text{AvC}[I]$ for some cluster-shape $I$. How can we represent and use it efficiently without the cluster functions?

A. Folding and unfolding

A function in $\tilde{C}(I)$ has dim $V$ components and depends on $|I|$ site-occupation variables, while a cluster observable in $\text{AvC}[GI]$ depends on $|GI||I|$ variables. Lemma §3 shows how the latter is reducible to the former in principle. Let us see how that can be done in more detail. We have $\tilde{f} \in \text{AvC}[GI]$ and need to evaluate it on configuration $y$. It is useful to introduce maps $\text{fold}^I_{ij} : \text{Cfg} \to \mathcal{F}(I)$ by

$$\text{fold}^I_{ij} y := \frac{1}{|G_I|} \sum_{T \in G} \rho_{ij}(T) E_I(Ty). \quad (9.1)$$

Here is the main result.

**Lemma 9.1.** For $\tilde{f} \in \text{AvC}[GI],$

$$\tilde{f}(y) = \sum_{ij} c^i \left\langle \text{fold}^I_{ij} \hat{y} \right| C_I f^j \right\rangle$$

$$= \sum_{ij} e^i \rho_{ji}(T) \left| C_I f^j \right|(Ty) \quad (9.2)$$

*Proof.* Apply (8.6) to $\tilde{f}(y) \overset{\text{S.9}}{=} |G_I| \left\langle \hat{y} \right| \text{Av(C}_I f \right\rangle.$

The component of $\tilde{f}$ in $\text{Av}^G I C(I)$ is just $\sum_i c^i C_I f^i$, so the representations in (9.2) are obtained directly from this component. The second sum only involves evaluating a function, $\left| C_I f^j \right|(Ty)$, which is in $\mathcal{F}(I)$, on the on the restriction of configuration $Ty$ to $I$. The first form is interesting because $\left\langle \text{fold}^I_{ij} \hat{y} \right| C_I f^j \right\rangle$ involves a uniformity of representation. Both the observable and configuration are effectively in the small Hilbert space $\mathcal{F}(I)$ here.

B. Storage and use of cluster observables

Lemma §1 involves evaluating $C_I f^j \in \mathcal{F}(I)$ on subconfigurations. How may we represent such a function without using a cluster function basis? Simply as a lookup table. Consider the efficiency of this method. For the sake of simplicity, suppose that all species spaces $\text{Sps}(i)$ have the same size. Then, dimensions of relevant spaces are $\dim \mathcal{F}^0(I) = |\text{Sps}| - 1$, $\dim C(I) = (|\text{Sps}| - 1)|I|$, and $\dim \mathcal{F}(I) = |\text{Sps}|^{|I|}$.

Suppose $g$ is in $C(I)$. Generally, we need to store at least $\dim C(I)$ numbers to represent it. Since the conventional cluster functions $\Phi^g_\alpha$ provide an ONB of $C(I)$, an efficient representation is obtained through the components $\langle \Phi^g_\alpha | g \rangle$. The hat-functions do not immediately provide a basis, of $C(I)$, but they do provide an orthogonal basis of the larger space $\mathcal{F}(I)$. The function $g$ can then be represented by its components $g(x) = \langle \hat{x} | g \rangle$ for $x \in \text{Cfg}(I)$, of which there are $|\text{Sps}|^{|I|}$. Effectively, this is to consider $g$ as an element of $\mathcal{F}(I)$. Compared to the conventional formalism, this is somewhat inefficient, but not horribly so, unless $|I|$ is large.

There are countervailing considerations, however. The whole point of having $g$ is to evaluate it on configurations. With the second computational representation, $g(x)$ is evaluated by simply stripping out $x_I$, the restriction of $x$ to $I$, and looking up the corresponding value of $g$ in memory. With the conventional computational representation, we have

$$g(x) = \sum_\alpha \langle \hat{x} | \Phi^g_\alpha \rangle \langle \Phi^g_\alpha | g \rangle. \quad (9.3)$$

It is necessary to first evaluate the factors $\langle \hat{x} | \Phi^g_\alpha \rangle$. If we need to evaluate $g$ on many configurations, this becomes increasingly costly.

10. SAMPLING, MODELING, AND HILBERT SPACE GEOMETRY

A. Defining a precise problem

This section is concerned with how we can put the preceding formalism to work in practical applications. We assume some $G$-equivariant scalar (e.g., energy) or tensor (e.g., dielectric tensor) observable $\tilde{f}$ of some configurationally disordered material, or class of materials sharing a parent structure. In terms both broad and vague, the problem is

P0: Find the most accurate and transferable model of $\tilde{f}$, using limited sample data.

A mode can be transferable with respect to many system characteristics. System size is a particular one that we have in mind here. The basis for thinking that the cluster expansion formalism might lead to a model transferable in that sense is the idea of having components in $C(J)$ and $C(I)$ which are simply related by a $G$ operation whenever $J$ and $I$ are so. It at least provides a plausible recipe for carrying out the transfer. Of a proposed solution to P0, one could not hope to say it is correct or incorrect. It may even be difficult to judge whether it is better or worse than another. Therefore, while something like $\alpha$ should remain a basic aspiration, more precise, and probably
limited, problems are needed to guide us. We return now to a fixed-system-size framework, and put transferability aside for future analysis. Two more precise versions are

P1: Using sampled data, find the best model of \( \vec{f} \) in a model subspace \( \mathcal{M} \) of \( G \)-equivariant observables.

P2: Using sampled data, find the best approximation of the orthogonal projection \( M\vec{f} \) of \( \vec{f} \) into \( \mathcal{M} \).

In practice, the model space is

\[
\mathcal{M} := \bigoplus_{\mathcal{I} \in \text{Small}} \text{Av} \mathcal{C}[\mathcal{I}], 
\]

(10.1)

where \( \text{Small} \) denotes some collection of cluster-shapes which are small, in number of sites and/or spatial extent. The orthogonal projector \( M \) onto \( \mathcal{M} \) is

\[
M = \sum_{\mathcal{I} \in \text{Small}} \text{Av} \circ \mathcal{C}_\mathcal{I}. 
\]

(10.2)

P1 makes no reference to the Hilbert space structure of \( \mathcal{F} \) at all. P2 does, and in the process partially resolves the adjective best in P1. Namely, \( \vec{f} = M\vec{f} + \vec{M} \perp \vec{f} \), where \( \vec{M} := 1 - M \) is the complementary orthogonal projection to \( M \). According to P2, our target is \( M\vec{f} \). This makes sense. \( M\vec{f} \) provides the unique minimizer of \( \| \vec{f} - \vec{F} \| \) over all \( \vec{F} \in \mathcal{M} \), hence is the natural Hilbert space notion of “best approximation to \( \vec{f} \) which lies in \( \mathcal{M} \).

P2 still contains a best that we shall have to interpret if we are to assess any proposed solution. One venerable option is to cash it out in statistical terms. That entails a sampling distribution according to which our sample is randomly drawn. However, we are free to deliberately choose the sample so as to learn as much as possible about \( \vec{f} \) with as little effort as necessary. For that reason, we do not take the statistical route. The results of this section can help us see how to make those choices more intelligently, so as to do more with a smaller sample.

Backing up a bit, we give our sample a name, \( \text{Smpl} \). Sometimes this will be the collection of configurations paired with the corresponding values of \( \vec{f} \), and sometimes just the configurations (which we will tend to write upper case to distinguish from generic configurations). Now, it would be a waste to have two configurations in \( \text{Smpl} \) which are \( G \)-related, since \( \vec{f}(TX) = T(\vec{f}(X)) \), and we know how \( G \) acts on \( \mathcal{V} \). Thus, we assume that \( \text{Smpl} \) contains more than one representative of any \( G \) orbit. Then, we know the values of \( \vec{f} \) on member of \( G \cdot \text{Smpl} \)

\[
G \cdot \text{Smpl} = \{ TX \mid X \in \text{Smpl}, T \in G \}. 
\]

(10.3)

Or, in Hilbert space geometry terms, we know the orthogonal projection of \( \vec{f} \) into the subspace

\[
\mathcal{S} := \text{span}\{ \mathcal{V} \otimes \vec{X} \mid X \in G \cdot \text{Smpl} \} 
\]

(10.4)

of \( \mathcal{V} \otimes \mathcal{F} \). In keeping with our convention, the orthogonal projector into \( \mathcal{S} \) is denoted \( \hat{S} \). \( \mathcal{S} \) is evidently the maximal subspace on which we know \( \vec{f} \), and is called the shadow subspace. The metaphor is that \( \vec{f} \) casts a shadow in \( \mathcal{S} \), and we are going to look for a vector in the model subspace \( \mathcal{M} \) that casts a shadow as nearly indistinguishable from that of \( \vec{f} \) as possible. Thus, our final version of the problem is

P3: Given the shadow subspace \( \mathcal{S} \) constructed from \( \text{Smpl} \) as described, seek \( \hat{F} \in \mathcal{M} \) such that

\[
1^\circ \| \hat{S} \vec{F} - \hat{S} \vec{f} \| \text{ is minimum.} 
\]

\[
2^\circ \| \hat{F} \| \text{ is minimum, subject to } 1^\circ. 
\]

The idea is that, of \( \vec{f} \), we know only that it is \( G \)-equivariant, and its projection \( \hat{S} \vec{f} \). Therefore, we seek a model in \( \mathcal{M} \) matching it in these respects. However, it may not be possible to match it perfectly, hence \( 1^\circ \), and even then, there may be freedom in the model, so \( 2^\circ \) stipulates that idle structure be stripped out. The given norm on \( \mathcal{F} \) provides throughout our standard of comparisons. We proceed to give the solution to this problem, prove its correctness, and contrast it with the traditional approach to fitting a cluster expansion model.

**B. Solving the abstract problem**

The solution to problem P3 is given in Proposition 10.2 below, which shows that it is realized by a linear operator \( \Lambda_{\text{Smpl}}^\mathcal{F} \) which does not depend on \( \vec{f} \) at all, but only on \( \mathcal{S} \) and \( \mathcal{M} \). In principle, the linear operator \( \Lambda_{\text{Smpl}}^\mathcal{F} \) provides simultaneous models for any number of observables measured on the sample \( \text{Smpl} \). While it is not generally practical to construct \( \Lambda_{\text{Smpl}}^\mathcal{F} \) in its entirety, it can be evaluated on an individual observable by (10.8). The following abbreviations will be used:

\[
\mathcal{S}' := \mathcal{SM}, \quad M_0 := M|_{\mathcal{S}'} \quad \mathcal{M}' := \mathcal{M} \mathcal{S}, \quad S_0 := S|_{\mathcal{M}'}.
\]

(10.5)

The preparatory Lemma 10.1 clarifies the geometry, showing that the orthogonal complement of \( \mathcal{M}' \) in the model space \( \mathcal{M} \) is completely inert for modeling purposes, while the orthogonal complement of \( \mathcal{S}' \) in \( \mathcal{S} \) cannot be effectively modeled from \( \mathcal{M} \).

**Lemma 10.1.** The following orthogonal direct sum decompositions hold:

\[
\mathcal{S} = \mathcal{S}' \oplus \ker M_0|_{\mathcal{S}} \quad \mathcal{M} = \mathcal{M}' \oplus \ker S_0|_{\mathcal{M}}
\]

(10.6a)

(10.6b)

Moreover, the restricted projections

\[
M_\ast : \mathcal{S}' \to \mathcal{M}' \quad \text{and} \quad S_\ast : \mathcal{M}' \to \mathcal{S}'
\]

are bijections, hence have well-defined inverses.
Proof. For \( s \in S \),
\[
s \in \ker M \iff \forall v \in M, \langle s \mid v \rangle = 0 \iff s \perp SM.
\]
This establishes (10.6a). It also shows that \( M \) is injective on \( S' \), and since any \( v \in M' \) is the image of something in \( S' \), bijectivity of \( M_\ast \) follows. The rest follows by symmetry. \( \square \)

**Proposition 10.2.** There is a unique vector \( \Lambda S^M \tilde{f} \in M \) satisfying conditions 1° and 2°. It is given by the linear operator
\[
\Lambda S^M = S^{-1}_S: \mathcal{F} \to M',
\]
where \( S' \) denotes orthogonal projection onto \( S' = SM \). The model \( \Lambda S^M \tilde{f} \) is alternatively characterized by
\[
\Lambda S^M \tilde{f} = \arg \min_{h \in MS} \| Sf - Sh \|. \tag{10.8}
\]

**Proof.** The preceding lemma says that \( SM = SMS \). Hence, \( f \) has the form \( Mg \) for some \( g \in S \) because any component of \( f \) orthogonal to \( SM \) would increase its norm without changing its image under \( S \). Appeal to criterion 1° then establishes (10.8).

Differentiation with respect to \( g \) says that minimizing \( \| Sf - SMg \|^2 \) requires \( M(Sf - SMg) = 0 \). Equivalently, \( SMg \) is the projection \( S'Sf = Sf \) of \( f \) into \( SM \). By definition of \( S_\ast \) (see Lemma), \( Mg = S^{-1}_S S' \), which finishes the proof, given (10.8). \( \square \)

### C. Application to the sampling context

Now we apply the abstract Proposition 10.2 to the context in which we are interested, namely, where \( S \) is determined by a sample \( Smpl \). Concerning \( M \), we only assume that its vectors are \( G \)-equivariant. There are essentially two aspects to this: the concrete form of the loss function, given in Proposition 10.3, and concrete parametrization of candidate models in \( M' \), discussed later.

#### 1. Loss function

**Proposition 10.3.** Assume
- \( S \) is determined by \( Smpl \), no two members of which are \( G \)-related
- \( Av \mathcal{M} = \mathcal{M} \), \( Av \tilde{f} = \tilde{f} \).

Then,
\[
\mathcal{S} = \text{span} \left\{ \mathcal{V} \otimes TX \mid T \in G, X \in Smpl \right\}, \tag{10.9}
\]
\[
\mathcal{M}' = \text{span} \left\{ M(\mathcal{V} \otimes \hat{X}) \mid X \in Smpl \right\}. \tag{10.10}
\]

The model is given by
\[
\Lambda S^M \tilde{f} = \arg \min_{\tilde{f} \in M'} L(\tilde{F})
\]
with the concrete loss function
\[
L(\tilde{F}) = \sum_{Y \in Smpl} P(GY) \left\| \tilde{f}(Y) - \tilde{F}(Y) \right\|_{V'}^2. \tag{10.11}
\]

**Proof.** Except for (10.11), this is a straightforward transcription of Prop. 10.2. As an ONB of \( S \), we take \( \sqrt{P(Y)} e_i \hat{Y} \) where \( e_i \) runs over an ONB of \( \mathcal{V} \), and \( Y \) over \( G \cdot Smpl \). \( \| S(f - \tilde{F}) \|_{V \otimes F} \) is then the sum of terms \( P(Y)|\langle e_i \hat{Y} \mid f - \tilde{F} \rangle|^2 \) over the indicated basis. \( \square \)

(10.11) differs from an ordinary least-squares loss function in that each \( G \)-orbit of \( Smpl \) makes a contribution weighted by its total \( E \)-probability. Also, it is completely determined by the conditions of problem P3.

#### 2. Model parametrization

According to (10.10), \( M' \) may be parameterized as
\[
\tilde{F}_{\lambda} := \sum_{X \in Smpl} M(\hat{\lambda}(X) \otimes \hat{X}), \tag{10.12}
\]
with a \( \mathcal{V} \)-valued variational parameter \( \hat{\lambda}(X) \) for each \( X \in Smpl \). Previous sections have discussed how to handle \( M(\hat{\lambda}(X) \hat{X}) \) computationally. See, in particular, Proposition 6.13 and Lemma 8.2. The parameterization in (10.12) might be redundant. In particular, if \( GX \) is nontrivial, then there are vectors \( 0 \neq e_i \in \mathcal{V} \) for which \( Av(e_i X) = 0 \). However, that is harmless; it merely means that different \( \lambda \) parameterize the unique \( \tilde{F}_{\lambda} \) minimizing \( L(\tilde{F}_{\lambda}) \).

On the other hand, if we parameterize candidate models as
\[
\tilde{F}_{\epsilon} = \sum_{I \in Smll} \epsilon_I \otimes \Phi_I, \tag{10.13}
\]
using conventional cluster functions \( \Phi_I \) (or any other basis of \( M \)) with coefficients \( \epsilon_I \in \mathcal{V} \), then the minimizing function will be nonunique whenever \( M' \) is not all of \( M \). A standard fix for such underdetermination is to add a penalty term involving an adjustable scale parameter to the loss function. Such ad hoc fixes are rendered unnecessary by parameterizing only \( M' \), as in (10.12). Circumstances under which \( Smpl \) might be deliberately engineered so that \( M' \) is a proper subspace of \( M \) are not hard to imagine. For instance, one may be interested in a low density of vacancies, and therefore use only configurations with no more than one vacancy in any cluster-shape in \( Smll \).
11. CONCLUSION

Our primary objective here was to advance understanding of what cluster expansion really means. To that end, we proposed a system of axioms for best-approximation operators and a definition connecting them to basic building blocks we call cluster components. The resulting structure is compatible with the traditional cluster expansion formalism in a special case, without the need for the usual cluster functions. Due to the importance of cluster expansion methods in practical computations, such a foundational advance potentially enjoys a large multiplier effect. The approach presented here to the problem of modeling from a sample supports that possibility, and achieves two specific ends. First, a precise version of the problem is fully integrated with the fundamental formalism, in contrast to the traditional cluster expansion formalism in a special case, where it appears ad hoc. Second, it shows that even for practical computations one can do without conventional cluster functions. Indeed, the alternative parametrization using sample data has an advantage in avoiding a potential underdetermination problem.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation under grant DMR-2011839. We thank Miguel Mercado and Ismaila Dabo for asking questions and pointing out obscurities.

Appendix A: Möbius duality

Notation in this appendix is different from in the main text. For some set Ω (not necessarily finite), we are concerned with functions on the set $\mathcal{P}_{\text{fin}}(\Omega)$, with values in some vector space $V$. Such subsets will be denoted $I$, $J$, $K$, etc., and use a function on them written in the normal way, e.g., $f(I)$. Möbius inversion is a procedure applicable in a much wider range of combinatorial settings than discussed here, but this one is more than adequate for our needs.

The equivalence (A.5) is our target; the path is organized as two lemmas.

**Lemma A.1.**

$$\sum_{J: K \subseteq J \subseteq I} \varepsilon(J) = \begin{cases} \varepsilon(I) & K = I \\ 0 & \text{otherwise} \end{cases} \quad (A.1)$$

**Proof.** The cases $K \not\subseteq I$ and $K = I$ are obvious since the sum is empty in the first case and has only one term in the second.

Now consider $I' = I \cup \{i\}$, where $K \subseteq I$ and $i \not\in I$. Subsets of $I'$ come in pairs of the form $(J, J \cup \{i\})$ with $J \subseteq I$. Each such pair makes a contribution $1 + (-1) = 0$ to the sum in (A.1).

For convenience, define the Möbius transform

$$(Mf)(I) := \sum_{J \subseteq I} f(J), \quad (A.2)$$

and a version of $\varepsilon$ rendered as a transformation

$$(\hat{\varepsilon}f)(I) := \varepsilon(I)f(I). \quad (A.3)$$

Then, the composition $\hat{\varepsilon} \cdot M$ is an involution:

**Lemma A.2.** $(\hat{\varepsilon} \cdot M)^2 = \text{id}.$

**Proof.**

$$M \cdot \hat{\varepsilon} \cdot M f(I) = \sum_{J \subseteq I} \hat{\varepsilon} \cdot M f(J)$$

$$= \sum_{J \subseteq I} \varepsilon(J) \sum_{K \subseteq J} f(K)$$

$$= \sum_{K \subseteq I} f(K) \sum_{J: K \subseteq J \subseteq I} \varepsilon(J)$$

claim $\hat{\varepsilon}f$.

This lemma immediately implies the following Möbius duality relation:

$$g = Mf \iff \hat{\varepsilon}f = M(\hat{\varepsilon}g) \quad (A.5)$$

[1] A. Van De Walle, A complete representation of structure-property relationships in crystals, Nature Materials 7, 455 (2008)

[2] J. Sanchez, F. Ducastelle, and D. Gratias, Generalized cluster description of multicomponent systems, Physica A 128, 334 (1984)

[3] D. de Fontaine, Cluster approach to order-disorder transformations in alloys, in Solid State Physics, Vol. 47, edited by H. Ehrenreich and D. Turnbull (Academic Press, 1994) pp. 33–176.

[4] A. Zunger, First-principles statistical mechanics of semiconductor alloys and intermetallic compounds, in Statics and Dynamics of Alloy Phase Transformations edited by P. E. A. Turchi and A. Gonis (Springer US, Boston, MA, 1994) pp. 361–419.

[5] J. M. Sanchez, Cluster expansion and the configurational theory of alloys, Physical Review B 81, 10.1103/PhysRevB.81.224202 (2010).

[6] J. M. Sanchez, Foundations and practical implementations of the cluster expansion,
C. Wolverton and D. de Fontaine, Cluster expansions of alloy energetics in ternary intermetallics, Physical Review B 49, 8627 (1994).

A. van de Walle, M. Asta, and G. Ceder, The alloy theoretic automated toolkit: A user guide, Calphad 26, 539 (2002).

D. Lerch, O. Wieckhorst, G. L. W. Hart, R. W. Forcade, and S. Mueller, Uncle: a code for constructing cluster expansions for arbitrary lattices with minimal user-input, Modelling and Simulation in Materials Science and Engineering 17, 10.1088/0965-0393/17/5/055003 (2009).

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, Advanced Theory and Simulations 2, 1900015 (2019).

J. H. Chang, D. Kleiven, M. Melander, J. Akola, J. M. Garcia-Lastra, and T. Vegge, Clease: a versatile and user-friendly implementation of cluster expansion method, Journal of Physics-Condensed Matter 31, 10.1088/1361-648X/ab1bb6 (2019).

Q. Wu, B. He, T. Song, J. Gao, and S. Shi, Cluster expansion method and its application in computational materials science, Computational Materials Science 125, 243 (2016).

A. Van der Ven, J. C. Thomas, B. Puchala, and A. R. Natarajan, First-principles statistical mechanics of multicomponent crystals, in Annual Review of Materials Research, V. 48, edited by D. Clarke (2018) pp. 27–55.

S. Kadkhodaei and J. A. Munoz, Cluster expansion of alloy theory: A review of historical development and modern innovations, JOM 73, 3326 (2021).

ICET, https://icet.materialsmodeling.org/.

ATAT, https://www.brown.edu/Departments/Engineering/Labs/avdw/atat/.

CASM, https://prisms-center.github.io/CASMcode_docs/.

CELL, https://sol.physik.hu-berlin.de/cell/.

CLEASE, https://close.readthedocs.io/.

J. H. Chang, D. Kleiven, M. Melander, J. Akola, J. M. Garcia-Lastra, and T. Vegge, Clease: a versatile and user-friendly implementation of cluster expansion method, Journal of Physics-Condensed Matter 31, 10.1088/1361-648X/ab1bb6 (2019).

Q. Wu, B. He, T. Song, J. Gao, and S. Shi, Cluster expansion method and its application in computational materials science, Computational Materials Science 125, 243 (2016).

A. Van der Ven, J. C. Thomas, B. Puchala, and A. R. Natarajan, First-principles statistical mechanics of multicomponent crystals, in Annual Review of Materials Research, V. 48, edited by D. Clarke (2018) pp. 27–55.

S. Kadkhodaei and J. A. Munoz, Cluster expansion of alloy theory: A review of historical development and modern innovations, JOM 73, 3326 (2021).

ICET, https://icet.materialsmodeling.org/.

ATAT, https://www.brown.edu/Departments/Engineering/Labs/avdw/atat/.

CASM, https://prisms-center.github.io/CASMcode_docs/.

CELL, https://sol.physik.hu-berlin.de/cell/.

CLEASE, https://close.readthedocs.io/.

J. H. Chang, D. Kleiven, M. Melander, J. Akola, J. M. Garcia-Lastra, and T. Vegge, Clease: a versatile and user-friendly implementation of cluster expansion method, Journal of Physics-Condensed Matter 31, 10.1088/1361-648X/ab1bb6 (2019).

Q. Wu, B. He, T. Song, J. Gao, and S. Shi, Cluster expansion method and its application in computational materials science, Computational Materials Science 125, 243 (2016).