TITLE: The use of cation-cation and anion-anion bonds to augment the bond-valence model

Revision 2

Matthew C. F. Wander, Barry R. Bickmore, Matthew Davis, W. Joel Johansen, Charles Andros, and Larissa Lind

Department of Geological Sciences, Brigham Young University, Provo, UT 84602, U.S.A. E-mail: mcfwander@gmail.com

Abstract

The bond-valence model has, for several decades, been widely used for creating quantitative structure-activity relationships (QSARs), crystal structure modeling, and verification of proposed structures. Certain limitations of the model, such as the neglect of interactions between cations and between anions, have prevented it from being more broadly applied, however. In this work we use cation-cation and anion-anion bonds to augment the existing bonding model in the systems H-Al-Si-O and K-Al-Si-O. The bond valence-length curves for these interactions employ the same mathematical form as ionic bonds, but make only a small contribution to the overall bonding in ionic materials. In the systems examined here, oxygen-oxygen interactions were much more important than those between cations for producing accurate bond-valence sums. Both anion-anion and cation-cation bonding could prove important, however, for our ultimate goal of producing valence-based force fields for use in molecular dynamics simulations. Rolling these interactions into the bond-valence framework would produce behavior similar to hard-sphere repulsive or van der Waals terms, but would more flexibly account for the complete bonding environment. The overall improvement in valence sums was robust, was maintained outside the calibration set, and was invariant to elemental substitution. We conclude that this minor alteration of the bond-valence approach will significantly improve bond-valence models in conjunction with other recent extensions of the approach.

Keyword: crystal structure, bond valence, ligand, silicate, aluminosilicate

Introduction

F. Albert Cotton once quipped that:

> [T]heories of chemical bonding—neglecting not a few which are entirely valueless—fall into one of two categories: those which are too good to be true and those which are too true to be good. "True" in this context is intended to mean "having physical validity" and "good" to mean "providing useful results, especially quantitative ones, with a relatively small amount of computational effort." The proper, rigorous wave equation for any molecular situation represents a theory of that situation which is too true to be good. (Cotton, 1964).

Five decades later, this observation is still apropos. Computational resources and quantum mechanical methods have developed considerably since 1964, but not enough to obviate the need for simpler models of atomic interaction, such as some of the popular bonding models (Gillespie and Popelier, 2001) and molecular mechanics force fields (Rappé and Casewit, 1997), which are computationally less expensive and promote fluent thinking about molecular structure and reactivity (Brown, 2003). In all cases, these simpler models represent atoms and molecules in somewhat physically unrealistic ways, but their associated mathematical descriptions tend to mimic certain aspects of real systems, at least when empirically calibrated. This allows for qualitative, or even quantitative, predictions about certain phenomena, but not others. It is generally the case that, at some point, attempts to make such models more physically realistic end up complicating their mathematical descriptions to the point that they become unusable for most practical purposes. Therefore, attempts to make "good" models more "true" should not be undertaken lightly.

The bond-valence model (BVM) is certainly one that is too good to be true. Over the past several decades, the BVM has been applied to a large number of ionic and polar-covalent systems, successfully rationalizing and predicting energetically favorable combinations of bond lengths about individual atoms (Brown, 1977; 1981; 2002; 2009). In fact, it is a standard tool for screening proposed crystal structures, and has been employed in both structure prediction (Brown, 2002) and the creation of quantitative structure-activity relationships (QSARs) (Hiemstra et al., 1989; Hiemstra and Van Riemsdijk, 1996; Hiemstra et al., 1996; Bickmore et al., 2004; Bickmore et al., 2006a; Bickmore et al., 2006b). The application of the BVM has been limited, however, because at least in its quantitative form, it is concerned solely with bond lengths, and not with the complete spatial distribution of ligands. One factor affecting this spatial distribution is necessarily ligand-ligand interactions, but the BVM has traditionally been developed within a generally ionic framework, in which bonds only exist between cations and anions. If ligand-ligand interactions are treated at all within a typical BVM-based structural model, it is usually by the introduction of simple repulsive potentials (Brown, 2002) or arguments based on symmetry (Brown, 2006; 2011; Bickmore et al., 2013). Both types are likely required. Therefore, the BVM has often been used to rationalize and predict *certain aspects* of structures, such as combinations of bond lengths, but typically not the full structures.

In this contribution, we show how it is possible to extend the BVM to account for ligand-ligand interactions in an internally consistent manner, by allowing anion-anion and cation-cation bonds. In fact, this has been done before by O'Keeffe and Brese (Brese and O'Keeffe, 1991; O'Keeffe and Brese, 1991; 1992), but some differences in our approach allow us to treat much more subtle interactions. We go on to show how these subtleties

might prove important in a BVM-based model that predicts molecular geometry in a manner comprehensive enough to be implemented in a molecular mechanics force field. To accomplish this, however, it is necessary to relax the ionic framework of the BVM. Throughout this paper, therefore, we present an argument for the proposition that the alterations we suggest would make the model both "truer" (i.e., more consistent with quantum mechanics) and "better" (i.e., producing statistically significant and robust improvement), at least for some purposes.

Theory

In this section we briefly explore the underpinnings of the BVM, showing how it is built into a generally ionic framework, and explaining the changes needed to allow for anion-anion and cation-cation bonding. We go on to explain why inclusion of such bonds in the BVM would be desirable for some purposes, and develop the specific hypothesis we test here.

Bond Valence and the Valence-Sum Rule

The most common form of the BVM posits that a quantity called the bond valence (s_{ij}) between ions *i* and *j* can be represented by Eqn. 1, where *R* is the interatomic distance and both R_0 and *B* are pair-specific, empirically calibrated parameters. The sign of s_{ij} is positive in the direction of the anion and negative in the direction of the cation, while the magnitude is expressed in valence units (v.u.).

$$\left|s_{ij}\right| = e^{(R_0 - R)/B} \quad (1)$$

Bond-valence parameters are generally calibrated (Brown and Altermatt, 1985; Brese and O'Keeffe, 1991; Adams, 2001) on numerous empirically determined crystal structures by assuming the valence sum rule (Eqn. 2), which requires that the valence sum of bonds incident to an ion *i* from counter-ions *j* ($S_i = \sum_j s_{ij}$) is equal to negative the atomic valence (V_i), i.e., the oxidation number, of ion *i*.

$$S_i + V_i = 0 \qquad (2)$$

Eqn. 2 is simply a restatement of Pauling's (1929) Second Rule, which treats the oxidation number as a measure of the total bonding power of an atom. Eqn. 1, however, accounts for differences in the strengths of bonds of different lengths.

Bond Valence and the Ionic Model

In the original ionic bonding model of Kossel (1916), atoms are treated as point charges. The atoms gain or lose valence electrons to obey the octet rule (or at least leave no unpaired valence electrons, *i.e.*, Lewis' "rule of two"), resulting in integral numbers of electron charges on the ions (Gillespie and Popelier, 2001). Thus, the electrostatic attraction between cations and anions holds the structure together, and to keep the point-charge atoms from collapsing in upon each other, an arbitrary repulsive potential is introduced between the cations and anions. Anion-anion and cation-cation repulsion is simply the result of Coulomb forces, but all atoms regardless of charge will have repulsive interactions from core-core overlap effects. This model predicts certain typical behaviors of atoms in ionic crystals, e.g., the oxidation numbers add to zero in each formula unit, counter-ions tend to be nearest neighbors, and the ligands about a central atom are distributed as symmetrically as possible.

The BVM treats atoms as point charges as well, but at least for the cation-anion pairs, both the attractive and repulsive potentials are rolled into Eqns. 1-2. That is, approaching cation-anion pairs attract one another until the bond-valence sum reaches the ideal value, and repel one another when it exceeds the ideal value.

The original ionic bonding model assumes complete transfer of bonding valence electrons from cations to anions, even though this is not the case in real structures. Preiser et al. (1999) accommodated this fact within the BVM by noting that for some purposes the spatial distribution of the bonding electron density does not matter. Bond valence is generally interpreted as the electric flux between a cation and anion, a "bond" occurring where electric flux lines connect two atoms. In this case, the valence sum rule (Eqn. 2) simply becomes a restatement of Gauss's Law, the electric flux through any closed surface is proportional to the electric charge contained within. If bonds have some covalent character, we can imagine the valence electron density involved as extra point charges positioned somewhere between the center of a bond and the anion. And if we define the surfaces of the atoms so that the anions include the bonding electron charges, the total electric flux between each anion and its ligands should still be the same, proportional to the number of bonding valence electrons. Furthermore, shorter bond lengths would necessarily be accompanied by larger fluxes (Preiser et al., 1999; Brown, 2002). While we know that such a transfer of electrons would certainly alter the flux lines, it does not appear to affect bond order. Rather, it appears to shift the bonding from ionic to covalent, while maintaining the fundamental inverse relationship between bond order and bond length.

For this reason, the BVM works well for both ionic and polar covalent bonds. That is, the "partial" (non-integral) charges one might assign to the individual ions by various population analysis schemes can be ignored if one knows the total flux incident to the individual ions, which would be proportional to the oxidation numbers, and the flux assigned to bonds of different lengths is empirically calibrated, assuming the exponential form of the relationship defined by Eqn. 1 (Preiser et al., 1999; Brown, 2002). Empirical calibration of bond-valence parameters likely rolls a number of competing effects related to the exact positioning of the valence electron density into a single function, which is reasonable if the drive to pair valence electrons is dominant.

Bond Valence and Covalent Bonding

If the BVM works well for atom pairs like Cl-O, which form almost completely covalent bonds, why is it not used for fully covalent bonds like Cl-Cl or O-O? This appears to be an artifact of the ionic framework of the model. If the atomic valence (oxidation number) of an atom is zero, for instance, how can that be divided between bonds? Also, in a system with polar covalent bonds like Cl-O, there would effectively be no electric flux lines between O atoms, or between Cl atoms.

That there is no fundamental difference between fully covalent and slightly polar covalent bonds is evidenced by the fact that primarily covalent bonding models, e.g., the Valence Shell Electron Pair Repulsion (VSEPR) model, easily account for both, and their electron density distributions behave very similarly (Gillespie and Hargittai, 1991; Popelier, 2000; Gillespie and Popelier, 2001). The primary constraints in both the BVM and

covalent bonding models like VSEPR, furthermore, are the octet rule and the rule of two, so at least in this respect these models are not incompatible.

One of the main differences is in how these models treat the directionality of bonds, a property strongly dependent on covalent character. VSEPR explains bond directionality in terms of repulsion between localized pairs of bonding and non-bonding (lone-pair) valence electrons, which depends on factors such as the strength of the bonds, the relative electronegativity of an atom and its ligands, and the number of lone pairs on the central atom (Gillespie and Hargittai, 1991). The classical BVM always results in ligands distributed as symmetrically as possible about each atom.

The prediction of symmetrically distributed ligands is clearly false in cases where electronic structure effects cause an asymmetric distribution of bonding and non-bonding valence electrons. This fact can be accommodated by the BVM, however, if we relax the requirement for a simple point-charge representation of atoms. In his description of a recent expansion of the BVM called the core-and-valence-shell model, Brown (2011) posited spherically symmetrical atoms in which weaker bonds tend to allow the symmetrical distribution of bonding and non-bonding (lone-pair) valence electron density, whereas strong bonds tend to make the lone pairs stereoactive, concentrating the lone-pair density to one side of the atom. Bickmore et al. (2013) and Shepherd et al. (in prep) quantified the resulting distortions in the coordination sphere, showing that if bonds are represented as vectors in the direction from cation to anion and magnitude equal to the bond valence, the valence dipole moment (i.e., the vectorial valence sum) and the valence quadrupole moment are predictable functions of the expected types of electronic structure effects and the magnitude of the incident bond valence. Clearly, small departures from the

traditional ionic framework of the BVM can yield large dividends in terms of the ability to model the total structure of a much larger class of compounds.

What should be done, then, about the inclusion of anion-anion and cation-cation bonds in the BVM? Certainly *ad hoc* adjustments can be made in clear-cut cases. For instance, S-S bonds form in persulfide compounds like pyrite and marcasite (FeS_2 polymorphs) so that the anions can obtain closed shells. Although the S atoms require 2 v.u. of bonds, we may treat them as S^- ions linked together in S_2^{2-} dimers. The same can be done for cation-cation bonds formed by some ions such as Hg^+ (e.g., in edgarbaileyite—$Hg_6Si_2O_7$)—the metal is treated as univalent, even though it clearly accepts 2 v.u. of bonds. Another interesting example is arsenopyrite (FeAsS), in which each Fe^{3+} is bonded to three S^- and three As^{2-}. The S^- atoms are bonded to three Fe^{3+} and one As, and the As atoms are bonded to three Fe^{3+} and one S^-. Thus, As receives 3 v.u. of bonds, and S receives 2 v.u., so that both anions obtain a closed-shell configuration. Again, one could treat the As-S pair as an AsS^{3-} dimer for the purpose of bond-valence analysis.

Such *ad hoc* adjustments, like dimerization, come with a cost, because they make it difficult or impossible to apply the recent extensions of the BVM that account for bond directionality (Brown, 2011; Bickmore et al., 2013). Bonded pairs still take up space on the surface of an atom, whether the BVM acknowledges their existence, or not. With the type of adjustment just mentioned, no distinction is made between bonds of different length, *i.e.*, they are all assumed to have integral bond orders. But it is well known that bond order is a function of bond length for any bond type (Gillespie and Popelier, 2001), and nuances in the bond network can be missed by ignoring this. Berry et al. (2006), for instance

212 characterized certain mixed-valence Ni compounds in which there are Ni-Ni bonds with a
213 bond order of 0.5.

214 The foregoing examples reveal that the main obstacle to including anion-anion and
215 cation-cation bonds in the BVM is the interpretation of bond valence as the electric flux
216 between non-overlapping atoms. If we interpret bond valence simply as a measure of the
217 spin-paired electron density resulting as the electron clouds of two atoms overlap and
218 redistribute themselves, with no reference to electric flux, it should be possible to include
219 cation-cation and anion-anion bonds, at least insofar as the bond-valence-bond-length
220 relationships can be approximated by the exponential form of Eqn. 1.

221 One cost of glossing over the ionic framework of the BVM is that it will not always be
222 as easy to determine *a priori* the atomic valence of every atom. For the common cations in
223 minerals, however, it is usually clear that their oxidation numbers will still equal their
224 atomic valences. In addition, the atomic valences of many of the common anions may be
225 simply determined based on the octet rule. Oxygen, for instance, can always accommodate
226 2 v.u. of bonds, whether in the diatomic gas, a peroxide, or an oxide.

227

228 ***Motivation***

229 O'Keeffe and Brese (1992) obtained the first anion-anion and cation-cation bond-
230 valence parameters, but since then these parameters have not been extensively used. For
231 many purposes, such bonds are usually weak enough that they can be safely ignored, but
232 recently there have been a number of attempts to include BVM-based structural
233 descriptors in potential energy models such as molecular mechanics force fields (Lufaso
234 and Woodward, 2001; Adams and Swenson, 2002; Grinberg et al., 2002; Cooper et al.,

2003; Adams et al., 2004; Grinberg et al., 2004; Shin et al., 2005; Adams and Rao, 2009; Liu et al., 2013a; Liu et al., 2013b). For models such as these, even relatively small inaccuracies can be costly, because both the absolute value and the gradients of structural descriptors like bond valence sums become important. Small improvements in bond-valence sums due to weak bonds could become significant, whereas O'Keeffe and Brese (1992) focused on anion-anion and cation-cation bonds stronger than 0.25 v.u.

In the context of molecular modeling, the desired level of accuracy is usually less than ~1 kcal/mol, or ~2-4 kJ/mol, and at least under some circumstances, the BVM might achieve this. Brown (2011) notes that bond-valence methods can typically predict bond lengths to within ~0.02 Å. A bond length change of 0.02 Å corresponds to ~4 kJ/mol from a comparison of experimental bond lengths and dissociation energies, which is within the desired range. This is excellent agreement, similar to the best quantum mechanical calculations and to the accuracy of X-ray diffraction crystal structure determinations. It is widely assumed that crystal structures published in major databases are for practical purposes exact, but this is quite wrong (Jones, 1984). Modern crystallographic methods have an average thermal positioning error on the order of 0.001-0.01 Angstroms for the position of each atom, with many authors claiming the lower end, while in fact being closer to the higher end of that range (Jones, 1984). The error in bond length is, therefore, actually double that figure, or up to 0.02 Å.

Brown (2011) specified, however, that the typical error of ~0.02 Å applies in cases where "the structure experiences no steric or electronic stresses." In fact, such stresses are to be expected where there are significant cation-cation or anion-anion interactions, and where covalent bonding induces lone-pair effects. These are common enough that it is

fairly typical for bond-valence sums to be off by as much as 0.1 v.u., and sometimes significantly more. Assuming the traditional *B* value of 0.37 (see Eqn. 1), 0.1 v.u. amounts to ~0.04 Å, using a Taylor expansion to approximate the error. Typical errors might actually be slightly larger, so a value of 0.05 Å is not an unreasonable estimate. Therefore, improvement in bond-valence estimates by a factor of 2-4 is all that could reasonably be expected, and would allow for excellent chemical accuracy in potential energy models utilizing bond-valence structural descriptors. This will not be done in one step, but rather by using a number of different strategies for improving the overall ability of the model to mimic real chemistry.

The motivation for including anion-anion and cation-cation bonds in bond-valence sums goes beyond improving the accuracy of those sums, however, and is also related to our intention to develop BVM-based force fields for use in molecular dynamics simulations. As noted above, the bond-valence equations (Eqns. 1-2) essentially impose both an attractive and a repulsive potential between bonded atoms. The interaction is attractive while the bond-valence sums on the atoms are smaller than the ideal, and repulsive when the sums are greater than the ideal. In fact, Adams and Rao (2009) showed that if we assume potential energy is proportional to the squared deviation from the ideal valence sum, and isolate that deviation to a single bond, the energy-distance curve becomes mathematically identical to a Morse potential, with the energy minimum at the ideal bond length. The real power in this realization, however, is that the bond-valence sum is a multi-body, rather than pair-wise, structural descriptor that is adaptable to different structural environments. That is, the ideal length for a given bond will necessarily change, depending on the valences of the other bonds incident to the atoms in question. In contrast, a typical

molecular mechanics force field might use a Morse potential, with a single ideal length for bonds of a given type, to describe bond-stretching energies. Even assuming equal lengths for all bonds, the ideal length of, for instance, Al-O bonds would change, depending on the Al coordination number. A typical force field might accommodate this by specifying two different types of Al, one 6- and another 4-coordinated, but this would preclude any changes in coordination number during a simulation. Beyond this, however, any potential energy model must incorporate terms to describe cation-cation and anion-anion interactions, in addition to the bonded interactions. This might be done via some combination of Coulomb and Lennard-Jones potentials, or similar, but these suffer the same defect as pair-wise bond-stretch terms, in that unchanging partial atomic charges or ideal distances are typically assumed. If cation-cation and anion-anion interactions were to be included in bond-valence sums, they would automatically be subject to the same type of structure-dependent attractive and repulsive potential as the cation-anion pairs. This would subsume more types of interactions under the same heading, leading to a simpler potential energy model with fewer adjustable parameters.

Supposing cation-cation and anion-anion interactions are fundamentally different than cation-anion bonds, it still might be possible to include them in a bond-valence framework without making the bond-valence sums much worse, especially if the interactions are comparatively weak. Such a strategy would be consistent, for our purposes, with Linus Pauling's (1960) definition of a bond as "whatever is convenient to the chemist to define as a bond."

It seems unlikely, however, that their inclusion would result in significantly improved bond-valence sums unless relatively weak cation-cation and anion-anion

304 interactions really do constitute "bonds" in a similar sense to cation-anion interactions.
305 That is, improved valence sums likely mean that these weaker interactions really do
306 contribute to filling the valence shells of the bonded atoms. There are reasons to believe
307 this might be the case. For example, an Atoms-In-Molecules electron density analysis
308 would show a bond-critical point for every O-O pair in a typical oxide, indicating a bond-
309 like shape to the local density (Bader, 1991; Popelier, 2000). The actual quantity of atomic
310 overlap (and hence bonding) may be tiny, but it is present and contributes to the pairing of
311 the valence electrons. Furthermore, the Ligand Close Packing (LCP) model addresses a
312 number of cases in which ligands such as O pack more closely together, and less
313 symmetrically, than would be predicted by the VSEPR model (Gillespie, 2000; Gillespie and
314 Popelier, 2001), and these deviations might be explained if weak bonding were allowed
315 between the ligands.

316 We conclude that if the inclusion of cation-cation and anion-anion bonds, even
317 relatively weak ones, can be shown to improve the accuracy and reliability of bond-valence
318 sums, then the break required from the ionic framework of the BVM is likely to improve its
319 long-term prospects for complete-structure modeling.

320

321 ***Hypothesis***

322 In the following sections, we test the hypothesis that by including cation-cation and
323 anion-anion bonds in the bond-valence sums, we do not impair existing bond-valence
324 functionality, and we can improve not only the overall reliability of the fits, but their
325 robustness in transferring to different crystal sets.

326

Methods

We test our hypothesis in two systems, Al-Si-H-O and Al-Si-K-O, where strong cation-cation and anion-anion bonding is not expected, by fitting bond-valence parameter values to very carefully chosen calibration sets of crystal structures, and then testing the fitted parameters against wider sets of structures. The following subsections detail the rationale for choosing the calibration sets and the fitting procedure, as well as our procedures for evaluating the results.

Calibration Sets

Crystal structures in the Al-Si-H-O and Al-Si-K-O systems were taken mainly from the American Mineralogist Crystal Structure Database (Downs and Hall-Wallace, 2003), the Crytallography Open Database (Grazulis et al., 2009), and the Inorganic Crystal Structures Database (Belsky et al., 2002), and are listed with the original references in the online Supplemental Information (Tables S1-S4). Crystal structures with partial occupancies of any of the atoms, or that were determined at far-from-ambient temperatures or pressures were eliminated. For reasons discussed below, we found it necessary to include crystal structures of H_2O (Pisani et al., 1996), O_2 (Cox et al., 1973), and H_2O_2 (Busing and Levy, 1965), but to do so we had to relax the temperature restriction and alter the H_2O and O_2 structures to eliminate partial occupancy and disorder in some of the sites. Structures containing H were only included if H positions were explicitly specified. The H positions for many of these were determined via neutron diffraction, which is the ideal, but others were determined by other means. Given that these other methods frequently included quantum mechanical structure optimizations or bond-valence calculations, calibrated on

structures obtained via neutron diffraction, we considered the set of sufficient quality for the proof-of-concept study reported here.

There were two sets generated for the Al-Si-H-O system. The first tempered set of 15 crystals (see Table S1) was selected to have, as far as possible, equal numbers of Si-O, Al-O, and H-O containing structures, and specifically selected to have a wide range of bonding environments for each metal. The second set, a check set (see Table S2), had 14 structures, and contained more unusual environments. Oxygen was allowed to have different oxidation states (0, -1, and -2), but the ideal valence sum was always constrained to be 2 v.u., because we were including fully covalent bonds in the total.

Likewise, we had two sets for the Al-Si-K-O system. A tempered set (see Table S3) was created by removing all the H-containing compounds from the Al-Si-H-O tempered set, and adding in five K-containing compounds, for a total of 13. The check set (see Table S4) also contained 13 structures, including the hydrogen-free structures from the Al-Si-H-O check set.

Optimization

We used a homegrown MATLAB (Mathworks, Inc.) program to read Crystallographic Information Files and perform bond-valence analyses of crystal structures. (The program is available upon request to BRB.) The optimization procedure minimized the summed squared deviation of the bond valence sums from their ideal values (ΔS^2) per unique atom, per structure. Our procedure for optimizing valence parameters (R_0 and B—see Eqn. 1) for the atom pairs utilized the *fmincon* function—a constrained Newton-Raphson-like minimization algorithm—in the MATLAB Optimization Toolbox. We

373 constrained the optimization so R_0 > 0 Å and B > 0.05 Å, and set the convergence tolerance
374 to 10^{-6}, maximum function evaluations to 50,000 (which was never reached), and all other
375 parameters to their default values. (We note that we used the optimizer in version 14 of
376 MATLAB, which produces slightly different results than version 13.

377 In order to have a proper apples-to-apples comparison, we used standard "hard"
378 (Brown and Altermatt, 1985) and "soft" (Adams, 2001) bond-valence parameter sets, and
379 then re-optimized them with the same assumed cutoff valence (0.01 v.u.) for comparison.
380 ("Hard" bond-valence parameter sets typically assume a near-universal B value of 0.37 Å,
381 while Adams's "SoftBV" set was optimized on both R_0 and B, leading to generally higher
382 values of B and softer bonds.) To some degree this will overestimate the quality of the re-
383 optimized set. Due to the extent to which optimization improved the quality of the fitting,
384 one might assume that the standard sets are quite poor. This is actually somewhat
385 misleading. The quality of any parameter set strongly depends on the choice of structures
386 in the fitting set and the chosen cutoffs. Some qualities that the fitting sets should have
387 were revealed in this analysis, and the issue of fitting is one we will return to in the
388 discussion.

389 There are some differences in our overall fitting procedures from earlier parameter
390 sets (Brown and Altermatt, 1985; Adams, 2001). For example, Brown and Altermatt (1985)
391 fit R_0 values one cation at a time, and held B constant. Adams (2001) fit both R_0 and B
392 values, but also treated each cation independently. While there are advantages in such an
393 approach, we chose instead to fit the entire parameter set to all element pairs at once.
394 While this might induce dependencies between parameter pairs, the overall degree of
395 misfit could be dramatically reduced. This is particularly important as we add covalent

parameters. For the single-ion approach to work, there can only be one kind of bond to a given ion, or in cases where there are two or more, only one parameter can be fit at a time. By fitting all sets at the same time we have the potential to obtain a much more robust set.

In addition to performing bond-valence analyses using the standard "hard" and "soft" parameter sets, we performed several optimization runs to make sure that the new parameter sets were stable under optimization, and to produce models with different degrees of freedom. These are designated Run 1, Run 2, etc., and are described as follows. 1) We re-optimized the R_0 values in the "hard" parameter set. 2) We re-optimized R_0 and B in the "soft" parameter set. 2a) We re-optimized both the R_0 and B values, using the "hard" parameter set as a starting point. (The results are not reported here, because they were identical to Run 2.) 3) We added initial parameter guesses for the R_0 and B values for O-O bonds (R_0 = 1.474 Å and B = 0.35 Å) to the parameter values resulting from Run 2, then re-optimized the new parameter set. 4) We added initial guesses for the R_0 and B values of all cation-cation bonds (R_0 = 0.001 Å and B = 0.05 Å) (see Schema 1) to the parameter set from Run 3, then re-optimized. 5) We added initial guesses for R_0 and B for the following cation-cation pairs to the results from Run 3: Al-Al (R_0 = 1.9620 Å and B = 0.7215 Å), Al-Si (R_0 = 2.1818 Å and B = 0.6812 Å), Al-H (R_0 = 1.2835 Å and B = 0.7626 Å), Si-H (R_0 = 1.4658 Å and B = 0.6015 Å), Si-Si (R_0 = 2.1410 Å and B = 0.6646 Å), H-H (R_0 = 1.0174 Å and B = 0.4981 Å). These initial values were estimated from quantum mechanical calculations, and we optimized all valence parameters.

The rationale for the series of optimization runs just described is as follows. We re-optimized the standard "hard" and "soft" parameter sets because they were initially optimized on other calibration sets with different cutoff values. We next added O-O

bonding because one can predict from the geometries that O-O interactions should be more important than any of the cation-cation interactions. That is, the O ions are frequently in contact (nearest neighbors), whereas the cations are always spaced such that an O is always between them. After the O-O parameters were determined, we tried adding combinations of cation-cation parameters. Many different starting configurations were attempted to ensure that the optimizer was not affecting the results. We found that it was not, so the full suite of optimization runs was only performed on the Al-Si-H-O tempered set, and fewer types of runs were performed on the other three. For the tempered Al-Si-K-O set, we performed optimizations corresponding to Runs 1, 2, 3, and 5. There was a slight modification to the procedure for Run 3 from that applied to the tempered Al-Si-H-O set to the corresponding K set. We used the parameter output from Run 3 for the Al-Si-H-O tempered set as the input to the optimization runs for the other data sets, except that K-O parameters had to be obtained from Run 2 on a data set containing K. This modification allowed us to make a direct statistical comparison between the outputted parameter sets prior to optimization. This comparison was required to test the reliability of the initial Al-Si-O parameters under the different conditions. For run 5 it was the parameters from 3, and the values R_0=2.0Å and B=0.5Å. For the two check sets, we analyzed the structures using the parameter sets obtained from Runs 1, 2, 3, and 5 on the small sets. We did not reoptimize any parameter sets on the check sets. The results of these reliability tests are labeled “Check #”, where “#” is the corresponding Run number. The Check results were obtained by analyzing the data set using the input parameter set of the corresponding optimization Run.

442 ***Evaluation***

443 We applied two different kinds of tests to evaluate our results. The first is a
444 statistical test for reliability, and the second includes two transferability tests for
445 robustness. By adding anion-anion and cation-cation bonding into the model, we are
446 obviously adding more adjustable parameters. Mathematically, any increase in the degrees
447 of freedom must increase the quality of the fit, so to demonstrate the value of an increase in
448 degrees of freedom, one must apply an appropriate statistical test. The extra-sum-of-
449 squares F-test (Anderson, 2011) is designed to statistically distinguish between models
450 with different numbers of degrees of freedom. This F-test is calculated slightly differently
451 than the standard F-test as in Eqn. 3:

452 $$F_{stat} = \frac{(n-p_2-1)(SS_2-SS_1)}{(p_1-p_2)(SS_2)} \quad (3)$$

453 where n is the number of crystal structures, p_1 and p_2 are the number of free parameters
454 for models 1 and 2 respectively, and SS_1 and SS_2 are the sums of squared error for the two
455 model fits. The probability of the two models being significantly different can then be
456 calculated using the standard F-distribution function. In essence, it determines whether
457 those extra degrees of freedom are warranted. The statistical differences between model
458 fits can then be visualized using a cumulative distribution function. We also must show
459 that the model parameters are robust in the context of the large sets of structures with
460 different configurations, rather than just the tempered calibration sets, while preserving
461 increased accuracy. Finally, the parameter sets must be insensitive to element swaps, e.g.,
462 Al-O R_0 and B parameters should be the same if H replaces K. If a model fit fails the
463 reliability test, it is unlikely to pass the tests for robustness. Likewise, if it passes the
464 reliability test, it is likely to also pass the robustness tests.

Results

Tables 1 and 2 show the fitting results for the small Al-Si-H-O set. In Table 1, we first report the standard deviation of ΔS per unique atom, per structure ($\sigma_{\Delta S}$). The lower part of Table 1 shows the mean, standard deviation, skewness, and kurtosis of ΔS for each element. The first two columns are the results for the standard "hard" and "soft" parameter sets without re-optimization (Checks 1 and 2). The subsequent columns represent the results of the different runs outlined in the *Methods—Optimization* section above. As expected, the results improve with increasing degrees of freedom.

The first two data columns in Table 2 show the "hard" and "soft" valence parameters used as initial guesses for the optimization runs. Subsequent columns list the R_0 and B values for each element pair obtained from the optimization runs. Generally, the optimized R_0 and B values are consistent between runs to within a few hundredths of an Ångstrom. The exception is hydrogen, which shows larger variation on the order of 0.1 Å in both R_0 and B, indicating that special care will be required when fitting hydrogen-related parameters for a general use set.

Tables 3 through 6 describe the results for the Al-Si-H-O check set (Table 3), the tempered Al-Si-K-O set (Tables 4-5), and the Al-Si-K-O check set (Tables 6). Table 7 is a selection of strongest O-O bonds used in the tempered sets. Finally, Table 8 shows the results from the extra-sum-of-squares F-tests performed on the small Al-Si-H-O set. We used this test to compare the results of Runs 1, 2, 3, and 5.

The F-test results are conclusive. Adding O-O bonds (Run 3) results in a statistically significant improvement (probability = 100.0% and 99.8%) to the model when compared

to the re-optimized "hard" and "soft" models (Runs 1-2). This is true even though the O_2 and the H_2O_2 crystals were eliminated from the comparison. However, introducing cation-cation bonds (Run 8) did not result in a statistically significant improvement in the model fits (probability = 0.0%) when compared to Run 3. In Figure 1, we plot the cumulative distribution function (CDF) for the same four models (Runs 1, 2, 3, and 5), as well as the standard "hard" and "soft" models. They qualitatively show the same strong differences between models shown by the F-tests, and the same strong similarity between Runs 3 and 8. Specifically, the CDFs show that adding O-O bonding (but not cation-cation bonding) greatly reduces the number of outliers.

How much O-O bonding is there in the oxide crystals? Our analysis indicates that ionic bonding accounts for >90% of the total bond valence in these crystals, and that most of the remaining <10% can be accounted for by O-O bonds. The addition of cation-cation bonds resulted in a contribution of less than 1%. The strongest O-O bonds, aside from those in O_2 and H_2O_2, are those with 6-coordinated Al and Si (stishovite). The structures with 5-coordinated Al had roughly ¾ the O-O bonding of their octahedral counterparts. In all cases, O-O bonding in tetrahedral configurations was negligible. (See Table 7.) Thus, including O-O bonds in the valence sums appears to serve the same purpose as an over-bonding repulsion term from the overlap of two O^{2-} electron clouds, but it also appears to improve the predictability of the valence sums.

As expected, the addition of O-O bonding (Run 3) provided consistent results across all the other structure sets (See Check 3 in Tables 3, 4, and 6). Starting with the large set of hydrated aluminosilicates, essentially the same level of statistical improvement is observed between Check 3 and the ionic "hard" and "soft" model results. Further changes in the O-O

and other new parameters due to re-optimization are relatively small; all changes in R_0 and B values between parameter sets are in the second decimal place. This is an important indication that the improvement is physically real, not an artifact of optimizing to our data set. As an aside, an aluminosilicate set was used that did not contain either an elemental-oxygen crystal or hydrogen peroxide. In this case the statistical improvement was maintained, but the sets could not be re-optimized because the O-O term was no longer sufficiently constrained. We will return to this effect in the Discussion.

Turning to the Al-Si-K-O data sets, reliability was again confirmed (see Check 3, Tables 4 and 6). Unfortunately, these sets had only a few K-containing structures. As a result, in these optimizations the only term involving K that we could fit was the K-O parameters. When we tried to optimize the K-Si, K-Al, and K-K terms (see Run 5, Tables 4 and 6) the optimizer had similar issues to those it had with the Al-Si-H-O cation sets, showing that there was poor dependence of the fitting results on these parameter values.

To summarize, several important points can be made about our results. First, addition of O-O bonds to the Al-Si-H-O optimized on the tempered Al-Si-H-O data set resulted in statistically significant improvement by decreasing the number of outliers, whereas the addition of cation-cation bonds did not. Second, the relative statistical improvement obtained by adding O-O bonds was maintained when applied to the large Al-Si-H-O set, and re-optimization resulted in only very small changes to the bond-valence parameters. Third, the optimized bond-valence parameters for the Al-O, Si-O, and O-O bonds remained quite consistent between the Al-Si-H-O and Al-Si-K-O sets after re-optimization (Tables 2 and 5). We note, however, that the stability of the O-O parameters was somewhat dependent on the inclusion of a range of strong O-O bond valences in O_2,

H_2O_2, KO_2, and K_2O_2. The optimized O-O parameters (R_0 = 1.4096 Å, B = 0.2428 Å for Al-Si-H-O; R_0 = 1.3913 Å, B = 0.2255 Å for Al-Si-K-O) are also quite similar to those obtained by O'Keeffe and Brese (1992) (R_0 = 1.48 Å, B = 0.37 Å), except that they assumed B = 0.37 Å, while our B values were optimized. Furthermore, the fully optimized Al-Si-K-O results (Run 3, Table 4) were almost identical in quality to the results obtained by analyzing the Al-Si-K-O data sets using the Al-O, Si-O, and O-O parameters from the Al-Si-H-O optimized set (Run 3, Table 1), and the K-O parameters obtained from Run 2 on the Al-Si-K-O tempered set. Thus, having a range of primary bonding environments is essential to producing a robust parameter set.

Discussion

It is abundantly clear from our results that O-O interactions are important, even in oxides where these interactions are always fairly weak. In this section, we first address important implications for the BVM if such bonds are included. We then outline some critical issues brought out by our analysis that may guide the way to further improvements in the accuracy of the BVM. We also argue that, at least for some purposes, cation-cation bonds should still be included, even in cases (such as the systems studied here) where these bonds are particularly weak.

Implications for the BVM

Supposing that weak anion-anion bonds, and to a lesser (and possibly negligible) extent weak cation-cation bonds, are present in typical compounds such as the oxides studied here, this would have important implications for the application of the BVM.

First, we indicated above that bond valence is taken to be positive in the direction of the anion and negative in the direction of the cation, but this formulation is inadequate if fully covalent bonds are included. In this case, the bond valence-length equation (Eqn. 1) would need to be modified so that the bond valence is positive in any direction, as in Eqn. 4.

$$s_{ij} = e^{(R_0 - R)/B} \quad (4)$$

Likewise, the valence sum rule (Eqn. 2) would need to be modified so that the atomic valence (V_i) is always positive, and the sum of the bond valence incident to an atom tends to equal, as closely as possible, the atomic valence, as in Eqn. 5.

$$S_i = \sum_j s_{ij} \approx V_i \quad (5)$$

Thus, an O^{2-} ion in an oxide would have V_O = 2 v.u., and should have ~2 v.u. of incident bonds. This situation is essentially no different for O^{2-} in typical oxides, for which the atomic valence would still be equal to the absolute value of the oxidation number, but there are other cases that could now be addressed via the new formulation. Instead of assuming V_O = 1 v.u. for O^- in a peroxide and V_O = 0 v.u. in O_2 gas, for instance, we would assume that V_O = 2 v.u. in all cases. In arsenopyrite (FeAsS), discussed above in the Theory section, V_{Fe} = 3 v.u. for Fe^{3+}, V_S = 2 v.u. for S^-, and V_{As} = 3 v.u. for As^{2-}. A cursory analysis of the structure would lead to the correct conclusion that $s_{Fe\text{-}S} \approx 1/3$ v.u., $s_{Fe\text{-}As} \approx 2/3$ v.u., and $s_{As\text{-}S} \approx 1$ v.u.,

The provision that the valence sum should approach the atomic valence *as closely as possible* is also a departure from the standard BVM. On the one hand, if bond valence is equated with the electric flux between ions, Gauss's Law requires that Eqn. 2 *must* be exactly obeyed (Preiser et al., 1999; Brown, 2002). If, on the other hand, we view bond valence simply as the number of electron pairs participating in a particular bond, and the valence sum rule as an expression of the tendency of atoms to obtain filled outer shells

through bonding, then the valence sum rule becomes less of an absolute requirement. If weak cation-cation and anion-anion bonds are allowed, but the anion-anion bonds tend to have larger bond valences, there will be some situations in which the valence sum rule *cannot* be exactly obeyed. That is, the anions will tend to be slightly over-bonded, and the cations will tend to be slightly under-bonded.

To some, this might seem like a shocking departure from the simplicity of the standard BVM, but there are many instances where apparent deviations from ideal valence sums are used in energy cost functions, so it should not be too problematic to assume that such deviations might be real (Hiemstra et al., 1989; Hiemstra and Van Riemsdijk, 1996; Hiemstra et al., 1996; Lufaso and Woodward, 2001; Adams and Swenson, 2002; Grinberg et al., 2002; Cooper et al., 2003; Adams et al., 2004; Bickmore et al., 2004; Grinberg et al., 2004; Shin et al., 2005; Bickmore et al., 2006b; Adams and Rao, 2009; Liu et al., 2013a; Liu et al., 2013b).

Furthermore, given the proper context, we can show that the altered model opens up the possibility of applying the BVM to a much broader array of chemical scenarios in a more realistic manner. The "proper context" to understand this claim involves the realization that bonds with the same valence do not necessarily have the same bond energy. Consider, for instance, two diatomic molecules. The bond dissociation energy for F_2 is 158.67 kJ/mol, whereas that of Cs_2 is 43.919 kJ/mol. Certainly, both are held together by a single (~1 v.u.) bond, and yet their dissociation energies vary by more than 100 kJ/mol. Next, consider the CsF molecule, which is undoubtedly also held together by a ~1 v.u. bond, but which has a bond dissociation energy 517.1 kJ/mol. There are some complexities due to differences in interatomic distances and lone-pair effects (Sanderson,

602 1983), but it is generally true that the most metallic single bonds (lower average
603 electronegativity) tend to have lower dissociation energies than the most covalent single
604 bonds (higher average electronegativity), and both metallic and covalent bonds tend to
605 have lower dissociation energies than the most ionic bonds (larger electronegativity
606 difference). Although some have tried to apply the BVM to problems involving reaction
607 energies without taking account of differences in bond energy due to bond character
608 (Hiemstra et al., 1996), the fact is that such models cannot have very wide applicability.
609 Furthermore, these differences in bond dissociation energies due to bond character have
610 been common knowledge among chemists for decades, and in fact, such differences
611 between metallic/covalent and ionic bonds formed the basis for Pauling's original
612 definition of electronegativity (Pauling, 1932; 1960).

613 Given this context, it is clear that in a mixture of different atom types the available
614 atomic valence would be dominantly taken up in polar bonds, where possible, followed by
615 anion-anion bonds, and finally cation-cation bonds. This would be the most efficient way to
616 minimize the total bond energy, and explains quite a bit of well-known chemical behavior.
617 Strong cation-cation or anion-anion bonds would only occur where there are not enough
618 anions or cations, respectively, to make enough polar bonds to satisfy the atomic valences.
619 In fact, where some polar bonds are present, the cation-cation and anion-anion bonds
620 would be even less energetically favored, because the partial charges that develop on the
621 atoms would weaken the covalent and metallic bonds.

622 If weaker like-like bonds do form, anion-anion bonds would be most likely due to
623 both the energetic considerations outlined above and the relative sizes of typical cations
624 and anions. But if weak (<<1 v.u.) anion-anion and cation-cation bonds form

asymmetrically, this would cause the valence sums to depart from their ideal values. Why then would such weak bonds ever form? It is likely to occur for steric reasons. The N^{5+} ion in a nitrate (NO_3^-) group, for instance, ideally would form three 1.67 v.u. bonds with the O^{2-} ions, but doing so would force the O^{2-} ions quite close together, resulting in significant overlap of their electron clouds. The overlapping electron clouds would necessarily entail some electron sharing, i.e., covalent bonding. Thus, even if the atomic valence of the N^{5+} were fully satisfied with polar N-O bonds, the O^{2-} ions would necessarily be over-bonded. Therefore, the geometry would equilibrate in such a way as to favor the polar bonds as much as possible, while obeying the valence sum rule (Eqn. 5) as closely as possible for all atoms.

Given the energetic differences between the different types of bonds, it is clear why one can get away with ignoring weak anion-anion and cation-cation bonds for many purposes. It is also clear why, in mixed systems of atoms, the like-like bonds tend to closely approximate integral bond orders (~0 v.u., ~1 v.u., ~2 v.u., ~3 v.u.), although there are some rare exceptions (e.g., KO_2 and K_2O_2 in Table 7). That is, fractional bond orders in these bonds would usually force deviations from ideal valence sums. This argument also provides a simple rationale for why there are only two very rare peroxide minerals, studtite and metastudtite (Burns and Hughes, 2003), whereas there are a number of very common persulfide minerals, e.g., pyrite and marcasite. That is, sulfur is considerably less electronegative than oxygen, so it forms less ionic bonds with metals, allowing S-S bonds to be more competitive with Me-S bonds.

While the argument we have outlined here introduces some extra complexity into the BVM, it provides a framework through which we might eventually use the expanded

BVM to model some very complex chemistry, including redox reactions. This is not a trivial task within a molecular mechanics framework (Comba et al., 2009).

It remains to address the issue of compatibility of our O-O parameters with existing bond-valence parameter sets. Unfortunately, in general, our O-O and other like-ion parameter sets will not be able to be added to previously published cation-anion sets without modification, in part because they were optimized with different cutoff criteria, which are quite important for weak bonds. The one exception to this is the peroxides, where the O-O bonds are relatively strong. In these cases a marked improvement in bond valence sums will be realized even without refitting the other parameter sets. However, in all other cases a complete refitting of relevant parameter pairs is recommended. This will substantially improve overall accuracy in the bond valence sums.

Optimization Issues

Our analysis shows that it is necessary to include high-valence bonds in the calibration set when optimizing like-like valence parameters. The factors affecting the stability of the O-O parameters under optimization are a useful starting point to understand the stability of the other elemental pairs. For the O-O interactions, there are four different types of bonds that can be differentiated by bond valence. These included the O-O double bond in O_2, the single O-O bond in hydrogen peroxide, O-O interactions in stishovite, which has 6-coordinated Si, and O-O interactions in the other aluminosilicates, in which Si is 4-coordinated. If our calibration set were only to contain the last two types, we would be trying to fit the O-O valence parameters to a set with a maximum bond strength of 0.25 v.u. Under these conditions, the R_0 value (which is equivalent to the bond

length of a 1 v.u. bond) would be severely underdetermined, as would the corresponding *B* value. In general, if we try to fit valence parameters to a data set that only contains bonds of 0.25 v.u. or less, no optimization procedure can possibly be adequate.

We can visualize this by looking at Schema 2, which shows hypothetical bond-valence curves fit to different bond length distributions. Only short-range interactions contribute significantly to the shape of the curve, and long-range contributions are not important. The range of different bonding configurations—especially at the high end of the bond-valence curve—is actually more important than a balanced set of different elemental combinations.

It is likely that this issue was most acute for the cation-cation pairs, because only very weak bonds of these types were included in the calibration sets. Including examples of the pure elements did not improve the situation significantly, but this may be due to departures from the ideal exponential curve shape of the bond valence-length relationship (Eqns. 1 and 4).

Bonds Involving H

H-O and H-H bonds present an even broader problem. Even when H positions are experimentally determined by neutron diffraction, rather than estimated, the uncertainty is often much larger than that of the heavier elements (Jones, 1984). This is a factor that has been noticed previously, but clearly as one approaches a quantitatively better fit, the errors in hydrogen become even more apparent. Given the geochemical importance of this element, further consideration of how to better fit valence parameters involving H,

693 beginning with paring down the calibration set to those structures with reliable H positions
694 is warranted.

695 The valence parameter values reported for H-O bonds in standard sets (see
696 http://www.iucr.org/resources/data/datasets/bond-valence-parameters) are quite
697 variable. Our R_0 values of ~0.7 are clearly too short to represent a 1 v.u. bond, but R_0
698 values fit to limited data sets are known to be covariant with B values. H atoms in crystals
699 tend to have one covalent bond with O, and one (or more) hydrogen bonds (O-H···O). In this
700 case, fitting regimes cannot distinguish how strong the covalent bond should be relative to
701 the hydrogen bond, and partitions of 0.9 and 0.1, to 0.3 to 0.7 can all be fit approximately
702 equally well to these systems. This is a large part of the reason that this study is intended
703 only as a preliminary proof-of-concept for the need to include O-O bonds and our fitting
704 process, and not intended to produce a final fitted parameter set for a variety of
705 applications. In order to solve this problem we need to have systems that have a balance
706 between hydrogen bonds and systems with no hydrogen bonds. In other words, we will
707 need to include gas phase molecules in our fitting set. This is far too large an issue to be
708 included in the scope of this work and will be explored independently.

709 Brown (1976; 2002) did careful work on H-O bond valence based on structures
710 obtained by neutron diffraction, in which a large range of H-O bond lengths were
711 represented. He found that if we assume the valence sum rule is always obeyed, the H-O
712 bond valence-length curve takes on a very strange, non-exponential shape, with a hump
713 around 0.5 v.u. This strange shape, and the prevalence of asymmetric bond valence
714 distributions about H, was shown to derive from the fact that two 0.5 v.u. H-O bonds would
715 bring the two O atoms too close together if a normal exponential curve were assumed.

Thus, he recommended using three different sets of H-O bond-valence parameters, covering different bond length ranges. In reality, this strange shape of the valence-length curve might mean that the valence sum rule simply is not obeyed in strained systems with H-O bonds ~0.5 v.u. It also is very likely that anion-anion and cation-cation bonding terms are critical missing pieces.

Why Cation-Cation Bonds?

Even though we have shown that, at least in the systems studied here, cation-cation bonds are probably not important, there may be reasons to include them in bond-valence models. For example, if these parameter sets are used in a force field meant for molecular dynamics simulations, cation-cation interaction terms will have to be included. Typically, these would be implemented with van der Waals and/or Coulomb terms to mimic hard-sphere repulsion, but it should be possible to mimic this using a bond-valence approach (Schema 1), as well. Note that the cation-cation bond valence parameters reported in our data tables should not be used as-is for such purposes, because their inclusion proved statistically insignificant, so it is unlikely that their optimized values are very meaningful.

Implications

Inclusion of anion-anion, and possibly cation-cation, bonds in the BVM would certainly make the model more "true". Such bonds clearly exist, even in weak forms, and from a quantum mechanical standpoint are not fundamentally different than the bonds traditionally addressed by the model. Here we have shown that, at least in the case of O-O bonds, their inclusion is a statistically significant addition that improves overall reliability,

739 transferability and robustness of the fitting set. The improvement of the quality of the fit is
740 such that it is now comparable to the experimental error. Where outliers still exist they are
741 likely to be a result of an elevation of free energy rather than a misfit of the model. While
742 the full covalent set, which also included cation-cation bonds, failed the statistical test, they
743 are still worth considering for application in a molecular dynamics force field. Clear cases
744 of strong cation-cation bonds are known, particularly in cases where metals have unpaired
745 *d*-electrons (Müller, 2007).

746 The addition of O-O bonding to the BVM is likely to provide significant incremental
747 improvement in conjunction with our directionalized bond valence work including valence-
748 dipole and valence-quadrupole moments (Bickmore et al., 2013; Shepherd et al., 2013). It
749 is clear that the BV model, which has shown decades of promise in structure interpretation,
750 is amenable to a rigorous analysis that will likely dramatically improve its predictive
751 capability.

752

753 **Acknowledgements**

754 BRB would like to acknowledge support for this project from the National Science
755 Foundation Geobiology and Low-Temperature Geochemistry program (EAR-1227215),
756 NASA (MFRP- NNX11AH11G), and the BYU Mentoring Environment Grant program. The
757 authors acknowledge stimulating conversations with Prof. I. David Brown and Prof. Frank
758 C. Hawthorne. Prof. Dennis L. Eggett, at the BYU Statistical Consulting Center, pointed us to
759 the extra-sum-of-squares F-test applied here. Prof. I. David Brown and an anonymous
760 reviewer provided helpful comments on the initial manuscript.

761

References

Adams, S. (2001) Relationship between bond valence and bond softness of alkali halides and chalcogenides. Acta Crystallographica, B57, 278-287.

Adams, S. and Swenson, J. (2002) Bond valence analysis of transport pathways in RMC models of fast ion conducting glasses. Physical Chemistry Chemical Physics, 4, 3179-3184.

Adams, S. and Rao, R.P. (2009) Transport pathways for mobile ions in disordered solids from the analysis of energy-scaled bond-valence mismatch landscapes. Physical Chemistry Chemical Physics, 11, 3210-3216.

Adams, S., Moretzki, O., and Canadell, E. (2004) Global instability index optimizations for the localization of mobile protons. Solid State Ionics, 168, 281-290.

Anderson, K.B. (2011) Discussion of multicyclic Hubbert modeling as a method for forecasting future petroleum production. Energy & Fuels, 25, 1578-1584.

Bader, R. (1991) A quantum theory of molecular structure and its applications. Chemical Reviews, 91, 893-928.

Belsky, A., Hellenbrandt, M., Karen, V.L., and Luksch, P. (2002) New developments in the Inorganic Crystal Structure Database (ICSD): Accessibility in support of materials research and design. Acta Crystallographica, B58, 364-369.

Berry, J.F., Bothe, E., Cotton, F.A., Ibragimov, S.A., Murillo, C.A., Villagran, D., and Wang, X. (2006) Metal-metal bonding in mixed valence Ni_2(5+) complexes and spectroscopic evidence for a Ni_2(6+) species. Inorganic Chemistry, 45, 4396-4406.

Bickmore, B.R., Rosso, K.M., and Mitchell, S.C. (2006a) Is there hope for multisite complexation modeling? In J. Lützenkirchen, Ed., Surface Complexation Modelling. Elsevier, Amsterdam.

Bickmore, B.R., Tadanier, C.J., Rosso, K.M., Monn, W.D., and Eggett, D.L. (2004) Bond-valence methods for pK_a prediction: Critical reanalysis and a new approach. Geochimica et Cosmochimica Acta, 68, 2025-2042.

Bickmore, B.R., Rosso, K.M., Tadanier, C.J., Bylaska, E.J., and Doud, D. (2006b) Bond-valence methods for pKa prediction. II. Bond-valence, electrostatic, molecular geometry, and solvation effects. Geochimica et Cosmochimica Acta, 70, 4057-4071.

Bickmore, B.R., Wander, M.C.F., Edwards, J., Maurer, J., Shepherd, K., Meyer, E., Johansen, W.J., Frank, R.A., Andros, C., and Davis, M. (2013) Electronic structure effects in the vectorial bond-valence model. American Mineralogist, 98, 340-349.

Brese, N.E. and O'Keeffe, M. (1991) Bond-valence parameters for solids. Acta Crystallographica, B47, 192-197.

Brown, I.D. (1976) On the geometry of O-H...O bonds. Acta Crystallographica, A32, 24-31.

Brown, I.D. (1977) Predicting bond lengths in inorganic crystals. Acta Crystallographica, B33, 1305-1310.

Brown, I.D. (1981) The bond-valence method: An empirical approach to chemical structure and bonding. In M. O'Keefe and A. Navrotsky, Eds., Structure and Bonding in Crystals. Academic Press, New York.

Brown, I.D. (2002) The Chemical Bond in Inorganic Chemistry: The bond valence model, 278 p. Oxford University Press, New York.

Brown, I.D. (2006) On measuring the size of distortions in coordination polyhedra. Acta Crystallographica, B62, 692-694.

Brown, I.D. (2009) Recent developments in the methods and applications of the bond valence model. Chemical Reviews, 109, 6858-6919.

Brown, I.D. (2011) View of lone electron pairs and their role in structural chemistry. Journal of Physical Chemistry, A115, 12638-12645.

Brown, I.D. and Altermatt, D. (1985) Bond-valence parameters obtained from a systematic analysis of the inorganic crystal structure database. Acta Crystallographica, B41, 244-247.

Brown, T.L. (2003) Making Truth: Metaphor in science, 215 p. University of Illinois Press, Urbana.

Burns, P.C. and Hughes, K.-A. (2003) Studtite, $[UO_2](O_2)(H_2O)_2](H_2O)_2$: The first structure of a peroxide mineral. American Mineralogist, 88, 1165-1168.

Busing, W.R. and Levy, H.A. (1965) Crystal and molecular structure of hydrogen peroxide: A neutron-diffraction study. Journal of Chemical Physics, 42, 3054-3058.

Comba, P., Hambley, T.W., and Martin, B. (2009) Molecular modeling of inorganic compounds, 326 p. Wiley-VCH, Weinheim.

Cooper, V.R., Grinberg, I., and Rappe, A.M. (2003) Extending first principles modeling with crystal chemistry: A bond-valence based classical potential. In P.K. Davies and D.J. Singh, Eds., Fundamental Physics of Ferroelectrics. American Institute of Physics, Melville, New York.

Cotton, F.A. (1964) Ligand field theory. Journal of Chemical Education, 41, 466-476.

Cox, D.E., Samuelsen, E.J., and Ceckurts, K.H. (1973) Neutron-Diffraction determination of the crystal structure and magnetic form
factor of gamma-oxygen. Physical Review, B7, 3102-3111.

Downs, R.T. and Hall-Wallace, M. (2003) The American Mineralogist Crystal Structure Database. American Mineralogist, 88, 247-250.

Gillespie, R.J. (2000) Improving our understanding of molecular geometry and the VSEPR model through the ligand close-packing model and the analysis of electron density distributions. Coordination Chemistry Reviews, 197, 51-69.

Gillespie, R.J. and Hargittai, I. (1991) The VSEPR Model of Molecular Geometry, 248 p. Allyn and Bacon, Boston.

Gillespie, R.J. and Popelier, P.L.A. (2001) Chemical Bonding and Molecular Geometry: From Lewis Structures to Electron Densities, 268 p. Oxford University Press, Oxford.

Grazulis, S., Chateigner, D., Downs, R.T., Yokochi, A.F.T., Quirós, M., Lutterotti, L., Manakova, E., Butkus, J., Moeck, P., and Le Bail, A. (2009) Crystallography Open Database: An open-access collection of crystal structures. Journal of Applied Crystallography, 42, 726-729.

Grinberg, I., Cooper, V.R., and Rappe, A.M. (2002) Relationship between local structure and phase transitions of a disordered solid solution. Nature, 419, 909-911.

Grinberg, I., Cooper, V.R., and Rappe, A.M. (2004) Oxide chemistry and local structure of $PbZr_xTi_{1-x}O_3$ studied by density-functional theory supercell calculations. Physical Review B, 69, 144118.

Hiemstra, T. and Van Riemsdijk, W.H. (1996) A surface structural approach to ion adsorption: The charge distribution (CD) model. Journal of Colloid and Interface Science, 179, 488-508.

852 Hiemstra, T., Van Riemsdijk, W.H., and Bolt, G.H. (1989) Multisite proton adsorption
853 modeling at the solid/solution interface of (hydr)oxides: A new approach. I. Model
854 description and evaluation of intrinsic reaction constants. Journal of Colloid and
855 Interface Science, 133, 91-104.
856 Hiemstra, T., Venema, P., and Van Riemsdijk, W.H. (1996) Intrinsic proton affinity of
857 reactive surface groups of metal (hydr)oxides: The bond valence principle. Journal
858 of Colloid and Interface Science, 184, 680-692.
859 Jones, P.G. (1984) Crystal structure determination: A critical view. Chemical Society
860 Reviews, 13, 157-172.
861 Kossel, W. (1916) Über Molekülbildung als Frage des Atombaus. Annalen der Physik, 49,
862 229-362.
863 Liu, S., Grinberg, I., and Rappé, A.M. (2013a) Development of a bond-valence based
864 interatomic potential for BiFeO3 for accurate molecular dynamics simulations.
865 Journal of Physics: Condensed Matter, 25, 102202.
866 Liu, S., Grinberg, I., Takenaka, H., and Rappé, A.M. (2013b) Reinterpretation of the bond-
867 valence model with bond-order formalism: An improved bond-valence-based
868 interatomic potential for $PbTiO_3$. Physical Review B, 88, 104102.
869 Lufaso, M.W. and Woodward, P.M. (2001) Prediction of the crystal structures of perovskites
870 using the software program SPuDS. Acta Crystallographica, B57, 725-738.
871 Müller, U. (2007) Inorganic Structural Chemistry, 268 p. Wiley, Chichester, UK.
872 O'Keeffe, M. and Brese, N.E. (1991) Atom sizes and bond lengths in molecules and crystals.
873 Journal of the American Chemical Society, 13, 3226-3229.
874 O'Keeffe, M. and Brese, N.E. (1992) Bond-valence parameters for anion-anion bonds in
875 solids. Acta Crystallographica, B48, 152-154.
876 Pauling, L. (1929) The principles determining the structure of complex ionic crystals.
877 Journal of the American Chemical Society, 51, 1010-1026.
878 Pauling, L. (1932) The nature of the chemical bond. IV. The energy of single bonds and the
879 relative electronegativity of atoms. Journal of the American Chemical Society, 54,
880 3570-3582.
881 Pauling, L. (1960) The Nature of the Chemical Bond, 644 p. Cornell University Press, Ithaca.
882 Pisani, C., Casassa, S., and Ugliengo, P. (1996) Proton-ordered ice structures at zero
883 pressure. A quantum-mechanical investigation. Chemical Physics Letters, 253, 201-
884 208.
885 Popelier, P.L.A. (2000) Atoms in Molecules: An Introduction, 164 p. Pearson Education,
886 Essex.
887 Preiser, C., Lösel, J., Brown, I.D., Kunz, M., and Skowron, A. (1999) Long-range Coulomb
888 forces and localized bonds. Acta Crystallographica, B55, 698-711.
889 Rappé, A.K. and Casewit, C.J. (1997) Molecular mechanics across chemistry, 444 p.
890 University Science Books, Sausalito, CA.
891 Sanderson, R.T. (1983) Polar Covalence, 240 p. Academic Press, New York.
892 Shepherd, K., Johansen, W.J., Goodell, T., Bickmore, B.R., Wander, M.C.F., Davis, M., Andros,
893 C., Lind, L., and Robertson, K. (2013) Centrosymmetric distortions in the vectorial
894 bond-valence model: The valence quadrupole moment. In prep.
895 Shin, Y.-H., Cooper, V.R., Grinberg, I., and Rappe, A.M. (2005) Development of a bond-
896 valence molecular-dynamics model for complex oxides. Physical Review B, 71, No.
897 054104.

898
899

900

901

902

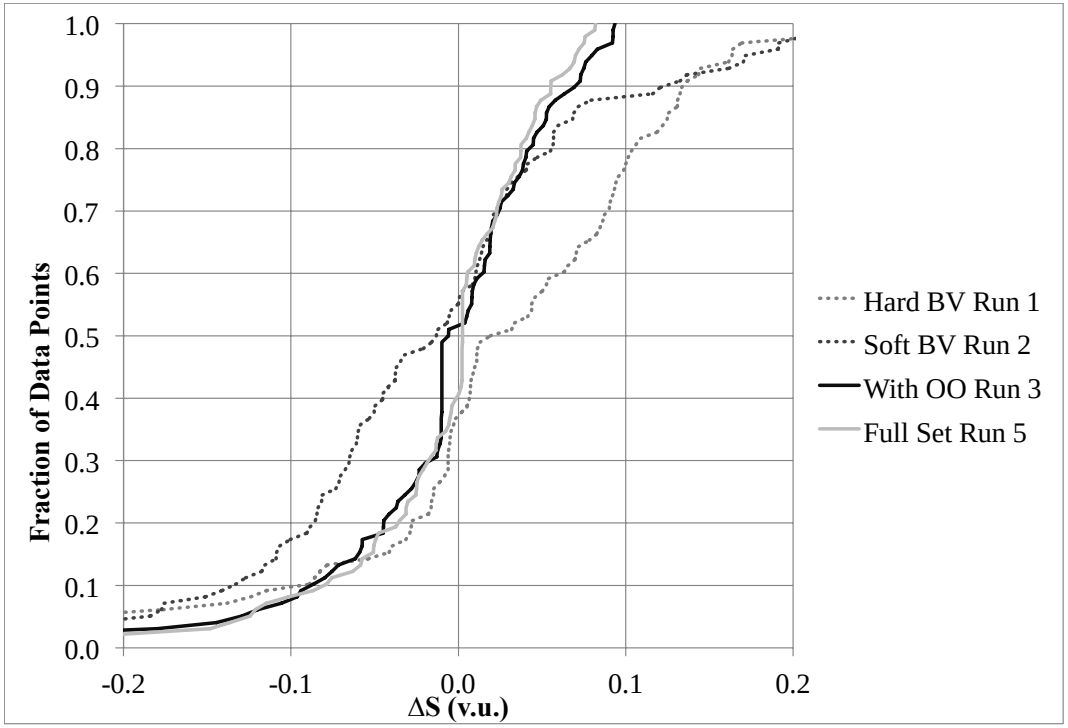


Figure 1. Cumulative distribution functions (CDFs) for the fits to the structurally unique atoms in the Al-Si-H-O set, where Δ*S* is deviation from the ideal valence sums. (Run 1 = no covalent bonds, *B* = 0.37 Å, R_0 optimized. Run 2 = no covalent bonds, R_0 and *B* optimized. Run 3 = O-O bonds allowed, R_0 and *B* optimized. Run 5 = all covalent interactions allowed, R_0 and *B* optimized. The way the largest negative deviations show up at the lower left of the plot and the largest positive deviations at the upper right. The best fits would be hug more closely around the vertical Δ*S* = 0 v.u. line, with smaller tails at the bottom and top of the distributions, indicating fewer outliers. Clearly, the curves from Runs 1 and 2 leave significantly more outliers than the curves that included covalent interactions (Runs 3 and 5), especially at the high end. However, the run that included only cation-anion and O-O bonds (Run 3) was very comparable to the run that included all covalent interactions (Run 5).

914

Table 1. Optimization results for the Al-Si-H-O tempered set. This table summarizes the statistics for each of the runs, and Run numbers correspond to the description in the Methods section. The tabulated numbers correspond to the deviations from the ideal valence sums (Δ*S*) for the atoms in the crystal structures from the calibration set. The first two rows contain the mean Δ*S* values for all the atoms before and after optimization. Subsequent rows have the post-optimization mean Δ*S* values for individual elements, as well as the standard deviation, skewness, and kurtosis.

Atoms	Data	Check1	Check2	Run 1	Run 2	Run 3	Run 4	Run 5
All	$\overline{\Delta S_{init}}$	0.173	0.155	0.173	0.155	0.064	0.073	0.807
	$\overline{\Delta S_{final}}$	N/A	N/A	0.157	0.151	0.025	0.024	0.023
Al	$\overline{\Delta S}$	-0.28	-0.11	-0.04	-0.07	-0.05	-0.05	-0.00
	Std. Dev.	0.08	0.08	0.09	0.09	0.09	0.09	0.06
	Skew	-0.35	-0.44	-0.36	-0.45	-0.40	-0.40	-0.57
	Kurtosis	-1.52	-1.78	-1.55	-1.68	-1.60	-1.70	-0.01
H	$\overline{\Delta S}$	-0.32	-0.04	0.00	0.08	-0.01	-0.01	-0.01
	Std. Dev.	0.11	0.11	0.16	0.10	0.09	0.09	0.09
	Skew	3.44	3.21	3.45	1.09	2.09	2.08	2.34
	Kurtosis	14.15	12.99	14.20	2.61	5.93	5.92	7.00
O (all)	$\overline{\Delta S}$	-0.67	-0.54	-0.52	-0.49	-0.01	-0.02	-0.03
	Std. Dev.	0.81	0.85	0.87	0.88	0.09	0.09	0.09
	Skew	-0.88	-1.11	-1.07	-1.14	-0.21	-0.18	-0.67
	Kurtosis	-0.89	-0.68	-0.68	-0.65	1.49	1.68	1.27
O (red.)‡	$\overline{\Delta S}$	-0.21	-0.05	-0.02	-0.03	*		
	Std. Dev.	0.26	0.12	0.16	0.04			
	Skew	-0.11	1.08	2.62	-0.79			
	Kurtosis	-1.58	2.86	10.26	2.08			
Si	$\overline{\Delta S}$	0.13	0.04	-0.01	-0.03	-0.03	-0.03	-0.01
	Std. Dev.	0.06	0.09	0.05	0.04	0.04	0.05	0.04

	Skew	-0.38	2.53	-0.38	-0.79	-0.85	-0.87	-0.26
	Kurtosis	0.87	7.62	0.82	2.08	2.69	2.55	-0.18

‡The O_2 and H_2O_2 structures were removed to make a proper comparison between models that include O-O bonds and those that do not.
*After this point no recalculations of the oxygen statistics were performed without O_2 and H_2O_2 contributions.

915
916
917

Table 2. Resulting parameter sets from each run for the Al-Si-H-O system. All R_0 and B values are given in Å. Overall, the order of convergence is the order of stability/reliability. The earlier the run number the parameter pair was stabilized on the more reliable it is likely to be.

	"hard" inputs	**"soft" inputs**	**Run 1**	**Run 2**	**Run 3**	**Run 4**	**Run 5**
Al-O (R_0/B)	1.620/ 0.37	1.5990/ 0.4240	1.6502/ 0.37	1.6275/ 0.3950	1.6427/ 0.3795	1.6442/ 0.3750	1.6275/ 0.3797
Si-O (R_0/B)	1.624/ 0.37	1.6082/ 0.4320	1.6106/ 0.37	1.6069/ 0.3911	1.6072/ 0.3877	1.6040/ 0.3793	1.5985/ 0.3589
H-O (R_0/B)	0.790/ 0.37	0.8705/ 0.4570	0.9307/ 0.37	0.7249/ 0.6344	0.7636/ 0.5741	0.7628/ 0.5743	0.7315/ 0.5932
O-O (R_0/B)					1.4096/ 0.2428	1.4147/ 0.2506	1.4322/ 0.2783
Al-Al (R_0/B)						0.3994/ 0.5432	2.4077/ 0.0964
Al-Si (R_0/B)						2.6483/ 0.5289	0.2411/ 0.6508
Si-Si (R_0/B)						2.6483/ 0.5289	2.3199/ 0.5135
H-H (R_0/B)						2.6483/ 0.5289	1.3300/ 0.1203
Al-H (R_0/B)						2.6483/ 0.5289	1.0247/ 0.5301
Si-H (R_0/B)						1.0093/ 0.4918	1.9791/ 0.3043

Table 3. Overall fit and results for each element in the Al-Si-H-O check set. This table summarizes the statistics for each of the runs, and Run numbers correspond to the description in the Methods section. The tabulated numbers correspond to the deviations from the ideal valence sums (ΔS) for the atoms in the crystal structures from the calibration set. The first two rows contain the mean ΔS values for all the atoms before and after optimization. Subsequent rows have the post-optimization mean ΔS values for individual elements, as well as the standard deviation, skewness, and kurtosis.

Atoms	Data	Check Hard	Check Soft	Check 3	Check 5
All	$\overline{\Delta S}$	0.0671	0.0470	0.0467	0.1753
Al	$\overline{\Delta S}$	-0.20	-0.04	0.03	0.61
	Std. Dev.	0.09	0.09	0.09	0.44
	Skew	-0.06	0.22	-0.03	0.53
	Kurtosis	0.01	-0.07	-0.08	0.17
H	$\overline{\Delta S}$	-0.21	0.10	0.12	0.75
	Std. Dev.	0.15	0.14	0.10	1.14
	Skew	1.04	1.04	1.23	2.18
	Kurtosis	0.27	0.22	1.60	4.01
O	$\overline{\Delta S}$	-0.02	0.03	0.04	0.02
	Std. Dev.	0.99	0.92	0.94	1.23
	Skew	0.42	0.43	0.45	0.44
	Kurtosis	-1.85	-1.84	-1.84	-1.22
Si	$\overline{\Delta S}$	0.018	0.09	0.01	0.00
	Std. Dev.	0.27	0.15	0.22	0.25
	Skew	0.09	0.05	0.16	0.02
	Kurtosis	-1.47	-0.79	-1.41	-1.37

924

Table 4. Overall fit and results for each element in the Al-Si-K-O tempered set. This table summarizes the statistics for each of the runs, and Run numbers correspond to the description in the Methods section. The tabulated numbers correspond to the deviations from the ideal valence sums (Δ*S*) for the atoms in the crystal structures from the calibration set. The first two rows contain the mean Δ*S* values for all the atoms before and after optimization. Subsequent rows have the post-optimization mean Δ*S* values for individual elements, as well as the standard deviation, skewness, and kurtosis.

Atoms	Data	Check hard	Check Soft	Check 3	Run 1	Run 2	Run 3	Run 5
All	$\overline{\Delta S_{init}}$	0.201	0.204	0.072	0.201	0.204	0.072	0.108
	$\overline{\Delta S_{final}}$	**N/A**	**N/A**	**N/A**	0.18	0.026	0.028	0.025
Al	$\overline{\Delta S}$	0.066	0.094	0.291	0.101	-0.031	-0.019	0.025
	Std. Dev.	0.365	0.232	0.361	0.369	0.161	0.163	0.062
	Skew	1.199	1.710	1.339	1.200	-1.064	-0.762	1.429
	Kurtosis	N/A*	N/A*	N/A*	N/A*	N/A*	N/A*	N/A*
K	$\overline{\Delta S}$	-0.017	-0.083	0.229	0.260	0.230	0.049	0.078
	Std. Dev.	0.284	0.244	0.378	0.357	0.378	0.321	0.260
	Skew	-0.087	-0.328	0.058	-0.113	0.057	0.036	-0.252
	Kurtosis	-2.612	-1.748	-2.779	-2.528	-2.779	-2.788	-1.937
O (all)	$\overline{\Delta S}$	-0.825	-0.858	0.019	-0.819	-0.831	-0.040	-0.048
	Std. Dev.	0.990	0.967	0.098	0.979	0.967	0.104	0.111
	Skew	-0.305	-0.286	0.166	-0.367	-0.376	-3.518	-3.565
	Kurtosis	-1.914	-1.926	7.001	-1.893	-1.898	19.071	19.500
O (red.) ‡	$\overline{\Delta S}$	0.013	-0.037	**	-0.001	-0.022	**	
	Std. Dev.	0.175	0.187		0.107	0.090		
	Skew	-3.402	-3.903		-1.367	-1.986		
	Kurtosis	13.672	17.238		-1.893	-1.898		
Si	$\overline{\Delta S}$	0.109	0.016	-0.064	-0.025	-0.025	-0.022	-0.003
	Std. Dev.	0.092	0.110	0.079	0.089	0.075	0.076	0.049
	Skew	-0.460	1.096	-0.549	-0.445	-0.745	-0.730	1.051
	Kurtosis	-0.717	4.255	-0.456	-0.749	-0.030	-0.076	-0.054

‡The O_2 and H_2O_2 structures were removed to make a proper comparison between models that include O-O bonds and those that do not.
*With only three Al-containing structures in the tempered set it is impossible to calculate a kurtosis value.
**After this point no recalculations of the oxygen statistics were performed without O_2, peroxide, and superoxide contributions.

925

926

Table 5. Five output sets for Al-Si-K-O tempered set.

Run:	Run 1*	Run 2*	Run 3	Run 8
Al-O (R_0/B)	1.6242/ 0.37	1.5491/ 0.4799	1.5542/ 0.4749	1.5427/ 0.4831
Si-O (R_0/B)	1.6118/ 0.37	1.6099/ 0.3885	1.6103/ 0.3874	1.6078/ 0.3822
K-O (R_0/B)	2.2028/ 0.37	2.2667/ 0.3310	2.2031/ 0.3369	2.1667/ 0.3513
O-O (R_0/B)			1.3913/ 0.2255	1.4034/ 0.2444
Al-Al (R_0/B)				1.2910/ 0.3163
Al-Si (R_0/B)				1.3589/ 0.3427
Si-Si (R_0/B)				2.8553/ 0.0671
K-K (R_0/B)				1.5659/ 0.3705
Al-K (R_0/B)				3.0615/ 0.0553
Si-K (R_0/B)				1.3615/ 0.6302

*Initial K-O "hard" bond-valence parameters were 2.113 & 0.37, and the initial "soft" parameter values were: 1.9548 & 0.4300. The initial Al-Si-O values from Table 3 input.

927
928

Table 6. Overall fit and Individual element results for Al-Si-K-O large set optimizations. This table summarizes the statistics for each of the runs, and Run numbers correspond to the description in the Methods section. The tabulated numbers correspond to the deviations from the ideal valence sums (Δ*S*) for the atoms in the crystal structures from the calibration set. The first two rows contain the mean Δ*S* values for all the atoms before and after optimization. Subsequent rows have the post-optimization mean Δ*S* values for individual elements, as well as the standard deviation, skewness, and kurtosis.

Atoms	Data	Check Hard	Check Soft	Check 3	Check 5
All	$\overline{\Delta S}$	0.0481	0.441	0.0420	0.4648
Al	$\overline{\Delta S}$	-0.22	-0.15	-0.21	-0.15
	Std. Dev.	0.10	0.09	0.13	0.30
	Skew	-0.27	-1.06	-0.57	3.04
	Kurtosis	0.12	3.82	1.03	12.30
K	$\overline{\Delta S}$	-0.05	-0.13	0.02	0.59
	Std. Dev.	0.14	0.11	0.16	3.4
	Skew	1.36	0.80	1.59	5.97
	Kurtosis	3.90	1.49	4.70	35.78
O	$\overline{\Delta S}$	0.01	-0.05	-0.02	-0.03
	Std. Dev.	0.18	0.16	0.14	0.15
	Skew	-0.76	-0.97	-1.34	-1.17
	Kurtosis	1.88	1.48	5.58	4.54
Si	$\overline{\Delta S}$	0.15	0.02	0.01	0.08
	Std. Dev.	0.15	0.17	0.20	0.15
	Skew	0.62	0.37	0.70	0.33
	Kurtosis	-0.83	-0.95	-0.76	0.03

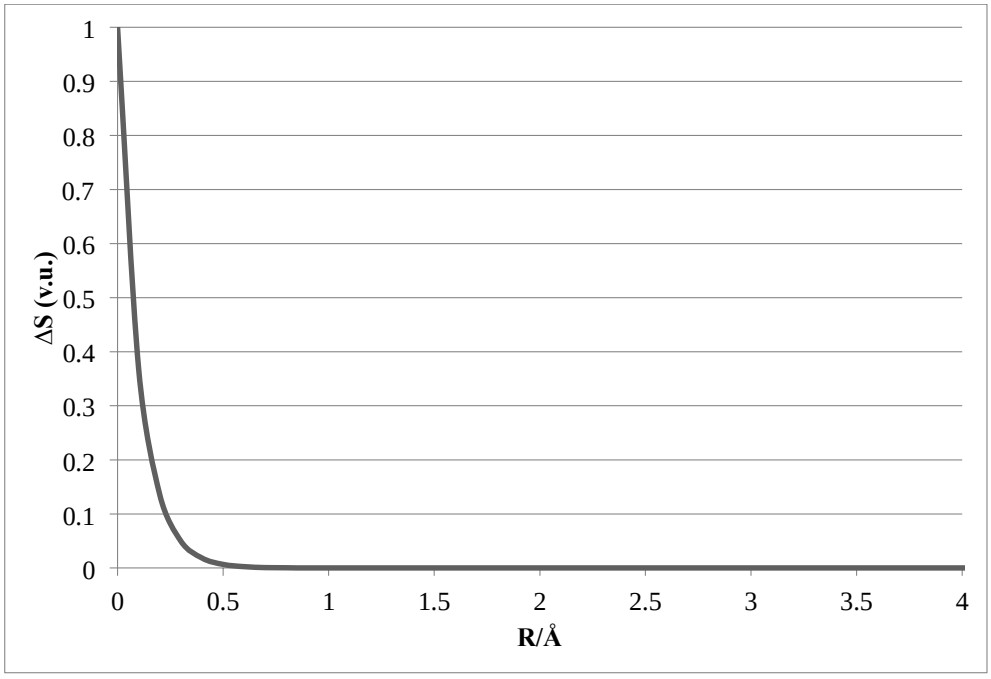
Table 7: Select strongest O-O bonds in our tempered sets crystals.

Crystal	$R_{O\text{-}O}$ (minimum)	$S_{O\text{-}O}$ (v.u.)	Interaction Type
O2	1.24014	2.013	Covalent
KO2	1.30631	1.538	Covalent
H2O2	1.45305	0.847	Covalent
K2O2	1.54058	0.593	Covalent
Bayerite	2.45773	0.014	Repulsive
Bohemite	2.53045	0.011	Repulsive
Corundum	2.52403	0.011	Repulsive
Diaspore	2.45905	0.014	Repulsive
Dickite	2.34906	0.022	Repulsive
Gibbsite	2.4192	0.017	Repulsive
Stishovite	2.52119	0.011	Repulsive

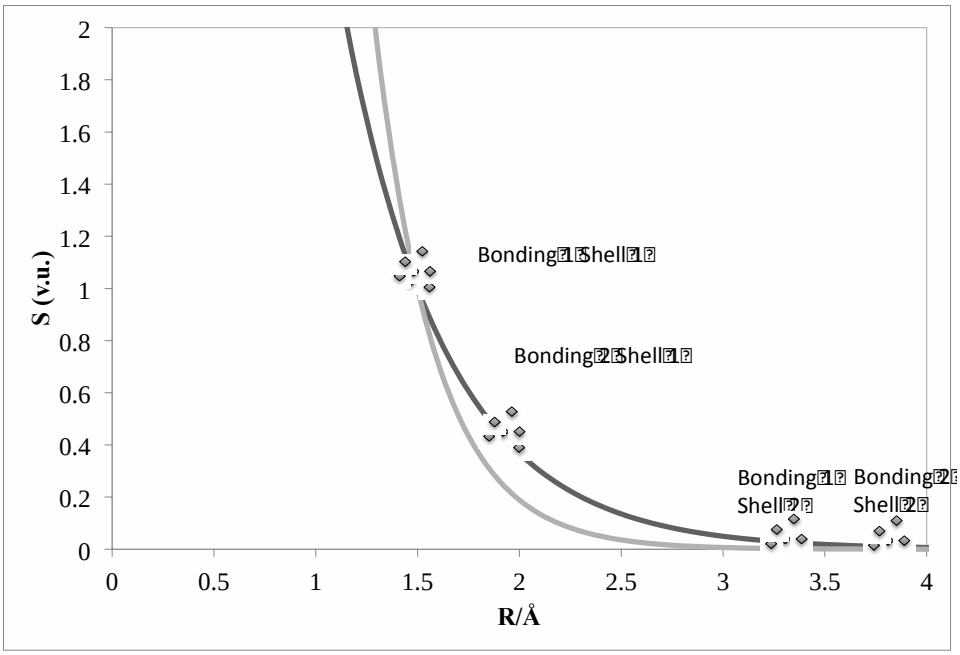
Table 8. F-test results for extra degrees of freedom. Comparing the "hard", "soft", "soft" with O-O bonds, and full covalent models. The sum of squared error (*SS*) and number of free parameters (*p*) is given for each model. In addition, we list the F-statistic (*F-stat*) and probability (Prob.) of a significant difference between models, given the extra free parameters in Model 2, for each model comparison. The tests show that it is justified with >99.8% confidence to fit *b* values in addition to R_0 ("soft" vs. "hard" models), and to add O-O bonding to the "soft" model, but it is not justified to add cation-cation bonding (full covalent).

Model 1	**Model 2**	SS_1	SS_2	p_1	p_2	**F-stat**	**Prob.**
Hard	*Soft*	2.973	1.483	3	6	30.1	99.5%
Hard	*Soft+OO*	2.973	1.187	3	8	26.5	99.6%
Soft	*Soft+OO*	1.483	1.187	6	8	11.0	98.2%
Soft+OO	*Full Cov.*	1.187	1.164	8	20	0.1	0.0%

936



937
938 **Schema 1.** Hard-sphere representation in terms of bond valence. This could also
939 useful as a zero term to check that the addition of covalent bonding does not alter
940 the bonding when its contribution should be zero.
941
942



Schema 2. A hypothetical chart showing the contribution of different types of bonds for defining a valence-length curve. In order to properly fit any "soft" bond-valence curve a range of bonding environments is required. Specifically, different short-range (1st shell) configurations are essential to fitting the softness parameter (*B*).