Factors Governing Oxygen Vacancy Formation in Oxide Perovskites

Robert B. Wexler, Gopalakrishnan Sai Gautam, Ellen B. Stechel, and Emily A. Carter*

ABSTRACT: The control of oxygen vacancy (VO) formation is critical to advancing multiple metal-oxide-perovskite-based technologies. We report the construction of a compact linear model for the neutral VO formation energy in ABO$_3$ perovskites that reproduces, with reasonable fidelity, Hubbard-U-corrected density functional theory calculations based on the state-of-the-art, strongly constrained and appropriately normed exchange-correlation functional. We obtain a mean absolute error of 0.45 eV for perovskites stable at 298 K, an accuracy that holds across a large, electronically diverse set of ABO$_3$ perovskites. Our model considers perovskites containing alkaline-earth metals (Ca, Sr, and Ba) and lanthanides (La and Ce) on the A-site and 3d transition metals (Ti, V, Cr, Mn, Fe, Co, and Ni) on the B-site in six different crystal systems (cubic, tetragonal, orthorhombic, hexagonal, rhombohedral, and monoclinic) common to perovskites. Physically intuitive metrics easily extracted from existing experimental thermochemical data or via inexpensive quantum mechanical calculations, including crystal bond dissociation energies and (solid phase) reduction potentials, are key components of the model. Beyond validation of the model against known experimental trends in materials used in solid oxide fuel cells, the model yields new candidate perovskites not contained in our training data set, such as (Bi,Y)(Fe,Co)O$_3$, which we predict may have favorable thermochemical water-splitting properties.

INTRODUCTION

Metal-oxide perovskites have many applications in fields ranging from energy science to electronics: e.g., two-step thermochemical cycles for water and carbon dioxide splitting (WS and CDS, respectively),$^{1-7}$ thermochemical energy storage,$^{8-11}$ thermochemical O$_2$ storage/pumping,$^{12-14}$ thermochemical air separation,$^{14-18}$ clean electricity production using solid oxide fuel cells (SOFCs),$^{19-23}$ solid oxide electrolyzers that produce hydrogen and carbon monoxide from high-temperature electrochemical WS and CDS,$^{24-26}$ and ferroelectric random-access memory (FRAM).$^{27-29}$ These applications depend sensitively on the concentration of oxygen vacancies (VO$_8$), which are capable of dictating thermodynamic, electronic, and emergent properties. As examples, VO$_8$ reduce H$_2$O to H$_2$ in a two-step thermochemical WS (often referred to as solar thermochemical hydrogen or STCH),$^{2,30,31}$ absorb O$_2$ to maintain low O$_2$ partial pressures in thermochemical O$_2$ storage/pumping and air separation,$^{12-14}$ facilitate O$_2^-$ transport in SOFCs,$^{22,23,32,33}$ and pin ferroelectric domain walls that impede ferroelectric switching in FRAM.$^{27,29,34-37}$ The concentration of VO$_8$ in any given oxide chiefl y depends on their formation energy ($E_f$), which is a well-vetted microscopic descriptor of performance, especially in applications such as STCH$^{38-40}$ and SOFCs.$^{22,23,33,41}$

The ease with which a neutral VO forms in metal (M) oxides, where a neutral VO refers to the removal of an entire oxygen atom (in its neutral, ground-state $^1P_2$ electron configuration) that contributes half of an O$_2$ gas molecule (in its neutral, ground-state $^3Σ^-$), leaving behind two electrons from the O$^2-$ that are donated back to the lattice (typically the cation sublattice), can be described by the following simple physical picture, similar to an extended Born–Haber cycle (see Figure 1).$^{22,23}$ First, forming the VO breaks bonds ($O–M$ bond dissociation). Second, the neutral VO may reduce its neighboring cations (M reduction) or, in materials with aliovalent substitution (e.g., Sr-substituted LaMO$_3$), may fill partial holes in the O sp band (lattice reduction).$^{42-47}$ Third, intrinsic stability/instability can predispose the metal oxide to form fewer/more VO$_8$ (metal oxide stability).$^{48-50}$ Note that stability typically is assessed as the energy above the convex hull ($E_{\text{convex}}$), which is the energy of
Despite their strengths, the reported models are still limited in a number of ways. The first limitation has to do with the level(s) of DFT used for parametrization. A few models\textsuperscript{52,59} have been constructed using the generalized gradient approximation (GGA) exchange-correlation (XC) functional of Perdew, Burke, and Ernzerhof (PBE).\textsuperscript{70} While PBE XC can describe alkaline-earth and La oxides fairly accurately (\textit{vide infra}), this method fails for 3d-transition-metal oxides, where a more accurate treatment of XC is essential.\textsuperscript{71} The most widely used method is PBE+\textit{U}, which pairs the PBE XC functional with Hubbard \textit{U} corrections to reduce the large self-interaction error incurred by highly localized \textit{d} electrons in transition-metal cations.\textsuperscript{71} However, the state-of-the-art XC functional for solids is the strongly constrained and appropriately normed (SCAN) meta-GGA XC functional.\textsuperscript{72,73} We recently showed that an accurate reproduction of the properties of transition-metal oxides using the SCAN functional still requires Hubbard \textit{U} corrections:\textsuperscript{74–76} i.e., a SCAN+\textit{U} approach. Our methodology, in which we fit \textit{U} to relevant oxidation energies,\textsuperscript{77,78} provides superior predictions of bulk thermodynamics, band gaps, and magnetic structures in comparison to PBE, PBE+\textit{U}, and SCAN, thus highlighting the usefulness of a database of SCAN+\textit{U} \textit{V}_0 formation energies that we present herein.

The second major limitation of the existing models is that they, to the best of our knowledge, were trained on homogeneous subsets of perovskite chemical space. Specifically, they lack a diverse sampling of A- and/or B-site metal cations (\# unique A $\times$ \# unique B $\leq$ 10),\textsuperscript{33,57,59–61,64,65} space groups (\# unique = 1),\textsuperscript{33,51,52,59,60,63–65} and electronic structures (i.e., metals, semimetals, and nonmetals).\textsuperscript{66,69} The only exception is the study by Maiti et al.,\textsuperscript{52} which uses the less accurate PBE XC functional for 3d-transition-metal-containing perovskite oxides. The lack of diversity in the training data can significantly limit the predictive power of such models, especially when they are extrapolated to higher-component (e.g., quaternary, quinary, etc.) and structurally diverse (e.g., experimentally accessible metastable polytypes\textsuperscript{79–81}) metal-oxide chemical spaces.

Finally, model-building approaches that use the band center or band gap as descriptors often rely on computationally expensive hybrid functional or GW calculations to obtain accurate predictions, which makes it hard to train a large data set. Moreover, band-center-based descriptors lack the physical intuition with respect to the physical process of \textit{V}_0 formation (i.e., M=O bond dissociation, reduction, and stability), as the band edges normally play a more significant role than the band center. Thus, the need still exists to build a computationally inexpensive yet sufficiently accurate, more physically intuitive, and broadly trained predictive model, especially for the discovery of higher-component or structurally diverse materials.

In this work, we utilize the state-of-the-art SCAN+\textit{U} method to construct a large database of \textit{V}_0 formation energies in ABO$_3$ perovskite oxides for a diverse collection of metal cations, i.e., A-sites (Ca, Sr, Ba, La, and Ce) and B-sites (Ti, V, Cr, Mn, Fe, Co, and Ni), and space groups (cubic, tetragonal, orthorhombic, hexagonal, rhombohedral, and monoclinic), as well as a range of electronic structures, from wide-band-gap insulators and semiconductors to metallic systems. As such, we have set out to provide an accurate (i.e., using SCAN+\textit{U}), extensive (341 unique \textit{V}_0s), and diverse (233 composition–structure combinations) catalog of \textit{V}_0 formation energies. Subsequently, we fit a theoretical model, using machine-learning techniques
with simple and physically intuitive descriptors, such as crystal bond dissociation energies and crystal (i.e., solid phase) reduction potentials, to understand and predict VO formation energies in a diverse set of metal-oxide perovskite frameworks.

The structure of this paper is as follows. First, we detail the computational methods used. Second, we explain the construction of our model and report its performance, key metrics, and trends for various subsets of the data. Third, we introduce our descriptors for VO formation energies and explain how they improve upon the palette of descriptors used in the literature. Fourth, we outline avenues for future research, both in terms of improving our methodology and for applying it to more complex systems (e.g., multinary metal-oxide perovskites and nonoxides). We also discuss the implications of our model with a focus on materials design and candidate identification for various applications. Finally, we conclude by recapitulating our findings and their implications for the understanding and control of VO formation in metal-oxide perovskites.

**COMPUTATIONAL METHODS**

**Quantum Mechanics Calculations.** We performed DFT+U calculations of neutral VO formation energies, crystal reduction potentials, phase diagrams, and band gaps using the Vienna Ab initio Simulation Package (VASP version 5.4.4).\(^1\)\(^2\)\(^3\)\(^4\)\(^5\)\(^6\)\(^7\)\(^8\)\(^9\) We used the SCAN XC density functional,\(^7\)\(^2\)\(^1\)\(^2\)\(^3\)\(^4\)\(^5\)\(^6\)\(^7\)\(^8\)\(^9\) with optimal U values taken from previous work.\(^7\)\(^2\)\(^1\)\(^2\)\(^3\)\(^4\)\(^5\)\(^6\)\(^7\)\(^8\)\(^9\) To apply the Hubbard U correction, we employed the rotationally invariant scheme of Dudarev et al.\(^4\)\(^4\) as implemented in VASP. We find that SCAN+U calculations are only modestly more computationally expensive (∼1.85–2.35×) for each self-consistent-field loop in comparison to those with PBE+U (see Table S1 in the Supporting Information). We utilized the all-electron, frozen-core, projector augmented-wave (PAW) potentials\(^3\)\(^5\)\(^6\) as in our prior work\(^7\)\(^2\)\(^1\)\(^2\)\(^3\)\(^4\)\(^5\)\(^6\)\(^7\)\(^8\)\(^9\) to describe the ion-electron interactions, along with including the nonspherical contributions related to the gradient of the electron density and the kinetic energy density within the PAW spheres for the XC evaluation. We used the Accurate precision mode in VASP to avoid aliasing errors when the fast-Fourier-transform and support grids were set, employed an additional support grid for more accurate evaluation of augmentation charges, and evaluated the projection operators in real space using VASP’s fully automatic optimization scheme. We used a plane-wave kinetic energy cutoff of 520 eV\(^7\)\(^2\)\(^1\)\(^2\)\(^3\)\(^4\)\(^5\)\(^6\)\(^7\)\(^8\)\(^9\) as this threshold converges total energies of oxide perovskites to within 1 meV/atom.

To perform Brillouin-zone integrations, we used a Γ-point-centered Monkhorst-Pack\(^4\)\(^4\) k-mesh with spacing between k-points of \(≤0.025\) Å\(^−1\), as previously optimized,\(^1\) and added a Gaussian smearing function with a width of 0.05 eV. For electronic minimizations, we used the blocked Davidson iteration scheme or the conjugate gradient algorithm,\(^1\)\(^2\)\(^3\)\(^4\)\(^5\)\(^6\)\(^7\)\(^8\)\(^9\) depending on difficulties in convergence. In all cases, we deemed to have reached self-consistent-field convergence when the energy change between steps was \(≤1\times10^{-5}\) eV/atom. We performed collinear, spin-polarized calculations and initialized the atomic magnetic moments in a ferromagnetic configuration.\(^3\)\(^9\)\(^1\)\(^0\) We considered the high-spin states for Ti, V, Cr, Mn, Fe, Co, and Ni, while for Co and Ni, we also considered the intermediate- and low-spin states. We used the conjugate gradient algorithm to perform structure optimization, stopping the process when the norms of all the atomic forces reached \(≤0.03\) eV/Å. For primitive cells, we optimized the atomic positions, cell shape, and cell volume, whereas for supercells, we only optimized the atomic positions to simulate the formation of VO in the dilute limit, in which the bulk lattice structure is not expected to change. Although we preserved symmetry in primitive cell structure optimizations to ensure that there were no symmetry-breaking relaxations (e.g., cubic → tetragonal), we did not preserve the symmetry of ionic positions in supercell calculations, in order to allow for local symmetry breaking upon VO formation.

**ABO₃ Perovskite Structures.** Figure 2 displays the set of elements (Figure 2a) and structures (Figure 2b) that we considered in constructing our VO formation energy data set. On the A-site, we considered alkaline-earth metals (Ca, Sr, and Ba) and lanthanides (La and Ce) that are reasonably abundant and are most likely to adopt the perovskite crystal structure (as opposed to ilmenite-like phases).\(^1\)\(^1\)\(^0\) Of the lanthanides, we focused on La and Ce because they comprise the La₄−₅SrM₀₃ (M = Cr, Mn, Fe, Co) family of perovskite-type materials for SOFC applications and CeO₂, the state-of-the-art metal oxide for STCH, respectively. On the B-site, we considered the redox-active 3d transition metals (i.e., Ti, V, Cr, Mn, Fe, Co, and Ni), with the exception of Cu, since Cu oxides typically will not survive nominal STCH temperatures. SCAN/SCAN+U also inaccurately describes Cu redox chemistry.\(^7\)\(^8\)\(^9\)\(^2\) We did not consider the 4d and 5d transition metals for two reasons, both of which we hope to address in future developments.

---

**Figure 2. ABO₃ perovskite structures.** Elements (panel a, non-gray rectangles) and crystal systems (panel b) considered in this study. A-sites (blue and orange) and B-sites (green) are as follows: Ca (2+), Sr (2+), Ba (2+), La (3+), Ce (3+ or 4+). Ti (3+ or 4+), V (3+ or 4+), Cr (3+ or 4+), Mn (3+ or 4+), Fe (3+ or 4+), Co (2+, 3+, or 4+), and Ni (2+, 3+, or 4+).
work: (1) we have not yet evaluated optimal U values for their oxides within the SCAN+U framework and (2) we wanted to make the construction of our database computationally tractable. For each composition, we considered the following six lattice systems (in Hermann–Mauguin notation) common to perovskites in the Inorganic Crystal Structure Database (ICSD):\textsuperscript{94} cubic (e.g., \textit{Pnma} SrTiO\textsubscript{3}, ICSD \#181652, hexagonal (e.g., \textit{P6\textsubscript{3}/mmc} SrMnO\textsubscript{3}, ICSD \#185417), rhombohedral (e.g., \textit{R3c} LaCrO\textsubscript{3}, ICSD \#167590), tetragonal (e.g., \textit{P4mm} BaTiO\textsubscript{3}, ICSD \#164388), orthorhombic (e.g., \textit{Pnma} CaTiO\textsubscript{3}, ICSD \#165801), and monoclinic (e.g., \textit{P2\textsubscript{1}/b} CeVO\textsubscript{3}, ICSD \#162747). We used the scheme in ref 7 to obtain structures for primitive cells. \textbf{Figure S1} in the Supporting Information has additional details on the composition of the data set.

\textit{Ab Initio Thermodynamics}. In this work, we calculated neutral VO\textsubscript{3−} formation energies (\(E_v\)) in the dilute vacancy limit (see \textbf{Figure S2} for the distribution of \(E_v\) and the data files for tabulated \(E_v\) values). We used the structure matcher\textsuperscript{95} module in pymatgen\textsuperscript{96} to obtain all possible symmetry-distinct neutral VO\textsubscript{3−} configurations. The formation energy of a neutral VO\textsubscript{3−} is given by:

\[
E_v = \frac{p^{\text{SCAN+U}} - p^{\text{pristine}}}{} + \frac{1}{E_{\text{O}_2}}
\]

where \(p^{\text{SCAN+U}}\) is the SCAN+U total energy, \(p^{\text{pristine}}\) is a supercell of the primitive cell, and \(p^{\text{defective}}\) is the same supercell with a neutral \(V\text{O}_{3−}\). \(E_{\text{O}_2}\) is the SCAN total energy of an \(O_2\) gas molecule in its ground-state atomic electron configuration. \(E_v\) is converted to within ±0.1 eV for a 3 \(\times\) 3 \(\times\) 3 supercell of the cubic and tetragonal phases, 2 \(\times\) 2 \(\times\) 2 supercell of the orthorhombic and monoclinic phases, and 2 \(\times\) 2 \(\times\) 1 supercell of the hexagonal and rhombohedral phases (see \textbf{Figure S3} in the Supporting Information for the convergence of \(E_v\) in different crystal systems with respect to supercell size).

For all perovskites considered, we quantified the stability with \(E_{\text{hull}}\) and tested the stability against all elemental phases (i.e., A, B, and O; XC = SCAN) and ordered (i.e., all sites in a given structure have integer atomic occupations), experimentally characterized binary and ternary metal oxides (e.g., AO, B\textsubscript{2}O\textsubscript{3}, AB\textsubscript{2}O\textsubscript{4}, etc.; XC = SCAN+U) containing the elements of the perovskite. For the DFT calculations of elemental solids, we used the SCAN XC functional without Hubbard U corrections, as the latter fails in predicting the properties of metallic systems with delocalized electrons. While DFT-SCAN and SCAN+U total energies are incompatible theoretically, this only influences A-rich and B-rich regions of the phase diagram that are isolated from the stability regions of the perovskite compositions we considered and only exist for the sake of completing the individual ternary phase diagrams. Therefore, the \(E_{\text{hull}}\) values reported here were calculated using a consistent theoretical treatment of all potential structures using SCAN+U. We used the phase diagram module\textsuperscript{97,98} in pymatgen\textsuperscript{96} to generate the 0 K phase diagrams and calculate \(E_{\text{hull}}\) (see \textbf{Figure S4} in the Supporting Information for the distribution of \(E_{\text{hull}}\)).

To calculate crystal bond dissociation energies (\textit{vide infra}), we first calculated cohesive energies from atomic total energies (see \textbf{Figure S5} in the Supporting Information for the periodic trends in experimental cohesive energies). We tested convergence with respect to cell size (12, 16, and 20 \(\AA\)) and used nearly cubic cells with minimal orthorhombic distortions to lift spurious intrashell degeneracies of atoms with partially filled shells (see \textbf{Figure S6} in the Supporting Information for the cell-size convergence of atomic total energies). Our predicted magnetic moments agree with the experimental ground-state atomic electron configurations with the exception of \(V\), which prefers [Ar]4s\textsuperscript{0}3d\textsuperscript{5} over [Ar]4s\textsuperscript{2}3d\textsuperscript{3} by 0.29 eV—this could be explained by inaccuracies of conventional spin-neutral PAW potentials\textsuperscript{99–101}. With that being said, this only leads to a small change of 0.05 eV in the V−O crystal bond dissociation energies. Finally, we used the Huber regressor in scikit-learn\textsuperscript{102} for supervised learning of \(E_v\). The Huber regressor is a regularized linear regression model that is robust to outliers. While nonlinear regression may find better descriptors for many properties (e.g., symbolic regression identified a simple descriptor for the oxygen evolution reaction activity of metal-oxide perovskites\textsuperscript{103}), we find that \(E_v\) can be described by a physically intuitive extended Born–Haber cycle (\textit{vide infra}), justifying the use of a linear model. The coefficients, intercepts, and scores reported here were cross-validated by averaging over 1000 models trained on random permutations of 50% of the data.

\section*{RESULTS}

\textbf{Model for \(V\text{O}_{3−}\) Formation in Metal-Oxide Perovskites.}

First, we report the performance of and key metrics in our best \(E_v\) model for room-temperature-stable (\(E_{\text{hull}} \leq 0.025\) eV/atom \(\approx 298.15\text{K}\)\(k_B\), where \(k_B\) is the Boltzmann constant) perovskites, as well as all perovskites considered. \textbf{Figure 3} shows the agreement between SCAN+U-calculated (vertical axis) and model-predicted (horizontal axis) \(E_v\) values for 142 unique \(V\text{O}_{3−}\)
Crystal Feature Engineering Used in the Model.

Crystal Bond Dissociation Energy. The inspiration for \( \sum E_b \) stems from its molecular analogue, bond dissociation energies (BDEs), for which experimental values are widely available in references such as the CRC Handbook of Chemistry and Physics.\(^{106}\) For transition-metal (M) oxides, however, experimental BDEs are only available for diatomic molecules/ions, e.g., MO(g) and MO\(^+\)(g), which lack the inclusion of effects from the metal’s oxidation state and the crystal field of the solid. To incorporate these effects, we define a crystal O\(^{2-}\)–M\(^{n+}\) bond dissociation energy (\( E_b \)) as:

\[
E_b[O^{2-}\text–M^{n+}] = E_b[MO_{n/2}] / N_b[O^{2-}\text–M^{n+}] \tag{2}
\]

where \( n \) is the oxidation state of M, \( E_b[MO_{n/2}] \) is the cohesive energy of the ground-state polymorph of the binary oxide crystal containing M\(^{n+}\), and \( N_b[O^{2-}\text–M^{n+}] \) is the number of O\(^{2-}\)–M\(^{n+}\) bonds per MO\(_{n/2}\) formula unit. \( E_b \) can be thought of as the cohesive energy per bond in an ionic oxide framework or the energy required to fully break an O\(^{2-}\)–M\(^{n+}\) bond. We use the \( E_b \) values of ground-state binary-oxide polymorphs because they provide a consistent (upper) limit on \( E_b \) and they can be extracted from either experiment or theory.

To calculate \( E_b \), consider the following series of chemical reactions:

\[
egin{align*}
MO_{n/2}(s) &\rightarrow M(s) + (n/2)O_2(g), \quad \Delta E_b[MO_{n/2}] \quad \text{(3a)} \\
M(s) &\rightarrow M(g), \\
(n/4)O_2(g) &\rightarrow (n/2)O(g), \quad (n/4)\text{BDE}[O_2] \quad \text{(3b)} \\
MO_{n/2}(s) &\rightarrow M(g) + (n/2)O(g) \\
E_b[MO_{n/2}] &\rightarrow \Delta E_b[MO_{n/2}] + E_b[M] + (n/4)\text{BDE}[O_2] \quad \text{(3d)}
\end{align*}
\]

where \( \Delta E_b[MO_{n/2}] \) is the binary metal-oxide formation energy/enthalpy, \( E_b[M] \) is the cohesive energy of the pure metal element, and \( \text{BDE}[O_2] \) is the bond dissociation energy of O\(_2\) per atom. The energy change for the net reaction (eq 3d) is \( E_b[MO_{n/2}] \). The experimental \( E_b[MO_{n/2}] \) value can be calculated from the experimental \( \Delta E_b[MO_{n/2}] \),\(^{107-111}\) \( E_b[M] \),\(^{112}\) and \( \text{BDE}[O_2] \) values.\(^{106}\) The experimental \( N_b[O^{2-}\text–M^{n+}] \) value can be obtained from the experimental binary-oxide crystal structures, which can be found in crystallographic databases such as the ICSD.\(^{95}\) In the absence of experimental data, theoretical \( E_b[MO_{n/2}] \) and \( N_b[O^{2-}\text–M^{n+}] \) values can be computed from SCAN+U total energies and optimized crystal structures, which are available in computational databases such as the Materials Project (MP),\(^{113}\) Open Quantum Materials Database (OQMD),\(^{114,115}\) and Automatic-\textsc{flow} for Materials Discovery (\textsc{aflow}).\(^{116}\) Note that because we define \( E_b \) for stable binary metal oxides, namely their ground-state polymorphs, ternary instabilities are captured primarily by \( E_{\text{hal}} \).

As an example, take the \( E_b[O^{2-}\text–Ce^{4+}] \) value for CeO\(_2\), where we are given that M = Ce, \( n = 4 \), \( MO_{n/2} = \text{CeO}_2 \), and its ground-state polymorph adopts the fluorite structure (\( \text{Flm}3m \)). According to the net reaction in eq 3d, the theoretical \( E_b \) value is given by:

\[ E_b[O^{2-}\text–Ce^{4+}] = \]
$E_{\text{c}}[\text{CeO}_2] = E^{\text{SCAN+U}}[\text{Ce}] + 2E^{\text{SCAN}}[\text{O}] - E^{\text{SCAN+U}}[\text{O}]$  \hfill (4)

where $E^{\text{SCAN+U}}[\text{Ce}]$ and $E^{\text{SCAN}}[\text{O}]$ are the SCAN+U and DFT-SCAN total energies of gaseous neutral Ce and O in their ground-state $^1$G$_{4u}$ and $^3$P$_2$ electron configurations, respectively. \cite{118} $E^{\text{SCAN+U}}[\text{CeO}_2]$ is that of fluorite CeO$_2$. Note that using SCAN+U instead of DFT-SCAN to calculate the total energy of an isolated, neutral Ce atom is appropriate, since Ce is metallic only in its bulk solid ground state and the tightly held open-shell $d$ and $f$ electrons require the Hubbard $U$ correction to mitigate the self-interaction error within DFT. That said, the $^1$G$_{4u}$ atomic state is an open-shell singlet state (coupling a $4f$ and a $5d$ electron), requiring two determinants to properly describe its spin coupling. Instead, single-determinant DFT+U theory only describes a spin-polarized, spin-contaminated singlet. As it turns out, the $^3$P$_2$ value is 3.15 eV, which is derived from Co(II,III)$_3$O$_4$ and acts as a proxy for the unstable Co(III)$_2$O$_3$, which can be attributed to the stability (metastability) of Co$_2$O$_4$ (Co$_2$O). Figure S8 in the Supporting Information has additional details on $E_{\text{c}}$ and its molecular analogues.

**Crystal Reduction Potential.** In the same spirit as $\sum E_{\text{b}} V_{\text{r}}$ extends useful concepts from atomic/molecular (electro)chemistry, such as ionization energies ($E_{\text{i}}$) and standard reduction potentials ($V_{\text{r}}$), to the solid state by consistently including crystal field effects and removing any solvation effects in an aqueous (or other liquid) medium. Another motivation for $V_{\text{r}}$ stems from the unavailability of experimental $V_{\text{r}}$ values for the following reduction reactions, relevant in ABO$_3$ perovskites: Fe$^{4+} \rightarrow$ Fe$^{2+}$ and Ni$^{4+} \rightarrow$ Ni$^{2+}$. \cite{106}

We define a crystal $M^{n+} \rightarrow M^{m+}$ (where $n > m$) reduction potential ($V_{\text{r}}$) vs O$_2$ (per 1/4 mol of O$_2$ gas) as:

$$V_{\text{r}}[M^{n+} \rightarrow M^{m+}] = -E_{\text{i}}[M^{n+} \rightarrow M^{m+}]/(n - m)F$$  \hfill (5)

where $n$ and $m$ are the oxidation states of the oxidized and reduced metals in the ground-state polymorphs of their binary metal-oxide crystals MO$_{n/2}$ and MO$_{m/2}$ respectively, $F$ is the Faraday constant, and:

$$E_{\text{i}}[M^{n+} \rightarrow M^{m+}] = E^{\text{SCAN+U}}[\text{MO}_{n/2}] + (n - m)E^{\text{SCAN}}[\text{O}_2]/4 - E^{\text{SCAN+U}}[\text{MO}_{m/2}]$$  \hfill (6)

Note that $V_{\text{r}}$ in eq 5 is strictly defined for the Gibbs energy of reduction, which we approximate here as the internal energy (i.e., $G \approx E$), ignoring $p$-$V$ and entropic effects, in order to apply our scheme readily to large data sets.

$V_{\text{r}}$ can be interpreted as the reduction potential of a metal cation in a crystal. For example, consider $V_{\text{r}}[\text{Ce}^{4+} \rightarrow \text{Ce}^{3+}]$, where $M = \text{Ce}$, $n = 4$, $m = 3$, $n - m = 1$, MO$_{n/2}$ = CeO$_2$ (ground-state polymorph = fluorite structure, $Fm\overline{3}m$), and MO$_{m/2}$ = CeO$_{3/2}$ = 1/2Ce$_2$O$_3$ (ground-state polymorph = hexagonal lattice, $P\bar{3}m1$). Following from eq 6, the theoretical $E_{\text{i}}$ value is:

$$E_{\text{i}}[\text{Ce}^{4+} \rightarrow \text{Ce}^{3+}] = \frac{1}{2}E^{\text{SCAN+U}}[\text{Ce}_2\text{O}_3] + E^{\text{SCAN}}[\text{O}_2]/4 - E^{\text{SCAN+U}}[\text{Ce}_2\text{O}_3] = 1.82 \text{ eV}$$  \hfill (7)

On the other hand, the experimental $E_{\text{i}} = \Delta E[\text{Ce}_2\text{O}_3] = 1.86 \text{ eV}$. \cite{118} Inserting $E_{\text{i}}$, $n - m = 1$, and $F$ into eq 5 yields theoretical and experimental $V_{\text{r}}$ values of $-1.82$ and $-1.86$ V per 1/4 mol of O$_2$, respectively ($AE = 0.04 \text{ V per 1/4 mol of O}_2$). Thus, reducing Ce$^{4+}$ to Ce$^{3+}$ in its bulk oxide state...
should theoretically (experimentally) require a reduction potential of 1.82 V (1.86 V) per 1/4 mol of oxygen gas.

Figure 5 shows that the experimental (blue bars) and theoretical (orange bars) \( V_r \) values are in good agreement, with MAE = 0.10 V vs O\(_2\), implying that, like \( E_b \), experimental values can be used to make accurate predictions in the absence of theoretical data (and vice versa). The aqueous \( V^\circ \) values (green bars), however, poorly capture the qualitative trends in \( V_r \). Note that we plot \( V^\circ \) with respect to the standard hydrogen electrode, whereas both theoretical and experimental \( V_r \) values are plotted with respect to 1/4 mol of O\(_2\)(g). While the \( V^\circ \)s of the transition-metal cations (Figure 5b–d) capture most of the trends in \( V_r \) (but have the wrong sign for V\(^{4+}\), Cr\(^{3+}\), Mn\(^{4+}\), Mn\(^{3+}\), Fe\(^{3+}\), and Co\(^{3+}\)), those of the alkaline-earth-metal and lanthanide cations (Figure 5a) do not. Hence, it is critical to use \( V_r \) and not \( V^\circ \) values because the latter lack the consistent inclusion of crystal field effects required for a quantitatively accurate description of solid-state cation reduction. For specific trends in \( V_r \), please see Figure S9 in the Supporting Information.

Model Intuition. Having presented \( E_b \) and \( V_r \) as descriptors for \( E_v \) and clarified how they relate to fundamental chemical concepts and principles, we now inspect the model’s construction and offer intuitive rationalizations for the signs and magnitudes of its coefficients (c). Our model consists of four terms (Figure 3), \( \sum E_b \) representing the energy penalty of breaking bonds and eventually creating an O\(_2\) molecule from an O\(^{-}\) ion (O\(^{-}\)-M bond dissociation in Figure 1), \( V_r \) quantifying the energy required to reduce “adjacent” cations due to the electrons left behind by the removed O\(^{-}\) (M reduction in Figure 1), \( E_g \) signifying the band structure cost of placing additional electrons in an oxygen-deficient perovskite in its conduction band (not shown in Figure 1 but affects M reduction), and \( E_{\text{hull}} \) highlighting the impact of the underlying (in)stability of the structure in controlling VO formation (metal oxide stability in Figure 1). Our model, with energy terms related to atomization (\( \sum E_b \)), ionization (\( V_r \)), and polarity of chemical bonds (\( E_g \)), can be considered to be an extended Born–Haber cycle.

The first term, \( \sum E_b \), is the sum of the \( E_b \) values for the crystal bonds formed between a specific O\(^{2-}\) that forms the vacancy and its nearest cation neighbors. For perovskites:

\[
\sum E_b = 4E_{b_b}[O^{2-}-A^{6+}] + 2E_{b_b}[O^{2-}-B^{m+}]
\]  

Figure 5. Comparison of experimental (blue) and theoretical (orange) \( V_r \) values with respect to available experimental aqueous standard reduction potentials (green),\(^{106}\) showing that the qualitative trends are obeyed. Experimental \( V^\circ \) values are not available for Fe\(^{4+}/3+\), Ni\(^{4+}/3+\), and Ni\(^{3+}/2+\).

Figure 6. Relationship between SCAN+U-predicted \( E_v \) and \( \sum E_b \) (blue), \( V_r \) (orange), \( E_g \) (green), and \( E_{\text{hull}} \) (red). The numerical ranges of the first three descriptors (i.e., the difference between the largest and smallest horizontal axis values) are approximately equal (i.e., \( \sim 5 \) eV or V). \( V_r \) captures most of the contribution to \( E_v \).
because there are four O$^{2-}$–A$^{n+}$ and two O$^{2-}$–B$^{m+}$ crystal bonds per O$^{2-}$. Note that our definition of $\sum E_b$ is quite general and easily extends to any desired metal-oxide material. The sign of the $E_b$ coefficient, $c_b > 0$, is consistent with the intuition that energy is required to break crystal bonds ($R = 0.77$, blue markers in Figure 6). The magnitude of $c_b$ is smaller ($0.1$–$0.2$) than those of the other coefficients ($0.3$–$5.5$), $c_a < c_b$, suggesting that $\sum E_b$ serves primarily as a correction factor to other terms in the model. Differences between the bond orders in binary and ternary oxides also can affect the magnitude of $c_b$; quantifying such effects will be the focus of future work.

The second term, $V_{\alpha}$, is the maximum $V_{\alpha}$ value among the nearest cation neighbors of a specific O$^{2-}$ and is the most dominant factor determining $E_{\alpha}$ in our model. Indeed, $E_{\alpha}$ anticorrelates quite strongly with $V_{\alpha}$ (orange markers in Figure 6), with $R = -0.88$. Choosing the maximum $V_{\alpha}$ we assume that the most reducible nearest cation neighbor undergoes reduction and hence governs the $V_{\alpha}$ contribution to $E_{\alpha}$. For perovskites:

$$V_{\alpha} = \max(V_{\alpha}[A^{n+} \rightarrow A^{+}], V_{\alpha}[B^{m+} \rightarrow B^{+}])$$

where $m > n$ and $r > s$. The sign ($< 0$) and magnitude ($1.5$–$1.6$) of the $V_{\alpha}$ coefficient ($c_\alpha$) resembles the energy for an $N$-electron reduction reaction, $E_{\alpha} = -N\epsilon_\alpha$ where $N = 2$, $N < 2$, and $N < 2$ for O$_2$ formation in ionic, ionic + covalent, and metallic oxides, respectively. Therefore, $-\epsilon_\alpha$ can be interpreted physically as the number of electrons donated by a specific O$^{2-}$ to its most reducible nearest cation neighbors upon O$_2$ formation in an ionic + covalent metal oxide.

Electrons from the departing oxygen that are left behind occupy the conduction band rather than the valence band. Hence the band gap qualitatively describes the cost of adding electrons to the conduction band upon reduction, which results in the coefficient of the third term, $E_{\text{v}}$, $c_{\text{v}} > 0$ ($R = 0.34$, green markers in Figure 6). Note that this gap is the one at the $\Gamma$-point obtained from a fully self-consistent SCAN+U calculation. Typical band structure and density of states calculations involve a non-self-consistent calculation after the completion of a fully self-consistent calculation. However, to minimize computational costs, we have not performed a non-self-consistent calculation, which is sufficient for our purposes here. Since the numerical ranges of the first three descriptors are approximately equal, the magnitude of $c_{\text{v}}$ ($0.3$–$0.4$) signifies that $E_{\text{v}}$ has a $15$–$20\%$ weight to the overall $E_{\alpha}$.

$\triangle$

Finally, the $E_{\text{hull}}$ coefficient, $c_{\text{hull}} < 0$ ($R = -0.18$, red markers in Figure 6), reflects the intuition that instability favors decomposition: e.g., via facile V$_2$O$_5$ formation. Note that the $c_{\text{hull}}$ coefficients do vary significantly ($5.5$–$5.8$) between our models trained on room-temperature-stable and all perovskites in our data set, likely offsetting the actual $E_{\text{hull}}$ values that are used in the room-temperature-stable ($E_{\text{hull}}$ $\leq$ $25$ meV/atom) and the full (no upper limit on $E_{\text{hull}}$) perovskite sets. In both of our theoretical models, $c_{\text{hull}}$ exhibits the highest magnitude of the coefficients; however, the typical values of $E_{\text{hull}}$ for crystals are in the range of $0$–$100$ meV/atom, which is at least one order of magnitude lower than $E_{\alpha}$, $V_{\alpha}$ or $E_{\text{v}}$ resulting in a significantly smaller contribution for those structures that are stable ($E_{\text{hull}} = 0$) and mildly unstable ($E_{\text{hull}}$ $\approx 50$ meV/atom). Importantly, ours is the first modeling work to include the effects of thermodynamic (in)stability on V$_2$O$_5$ formation via $E_{\text{hull}}$, which could be a source of the inaccuracy and/or nonintuitiveness of previous models.

B-Site Trends for Materials Design and Candidate Identification. In this section, we discuss the specific trends in the SCAN+U-calculated and model-predicted $E_{\alpha}$ with respect to changes in B and its oxidation state (n), focusing on materials design and candidate identification for various applications. Figure 7 shows the SCAN+U-calculated $E_{\alpha}$ distribution (DFT+$U$, circles), average SCAN+U-calculated $E_{\alpha}$ values ([DFT+$U$], filled diamonds), and average model-predicted $E_{\alpha}$ values ([Model], unfilled diamonds, vertical axis) as a function of B (horizontal axis) and n (panels). Overall, the average SCAN+U-calculated and model-predicted $E_{\alpha}$ values are in good agreement (MAE = $0.45$ eV). For $n = 4$ (i.e., B$^{4+}$), $E_{\alpha}$ decreases in the order Ti $>$ V $>$ Mn $\approx$ Cr $>$ Ni $> \approx$ Co $\approx$ Fe (Figure 7a), which largely coincides with increasing $V_{\alpha}[B^{4+} \rightarrow B^{3+}]$ ($R = -0.98$): namely, Ti ($-1.91$ V vs O$_2$) $< V$ ($-0.92$ V vs O$_2$) $<$ Cr ($-0.42$ V vs O$_2$) $\approx$ Mn ($-0.40$ V vs O$_2$) $<$ Ni ($0.40$ V vs O$_2$) $<$ Co ($0.55$ V vs O$_2$) $<$ Fe ($0.97$ V vs O$_2$).
Table 1. List of Promising Perovskites for STCH

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>$\sum E_g$</th>
<th>$V_r$</th>
<th>$E_g$</th>
<th>space group</th>
<th>model $E_v$ (eV)</th>
<th>$T_{90}$ (K)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y$^{3+}$</td>
<td>Fe$^{4+}$</td>
<td>16.23</td>
<td>−1.45</td>
<td>2.45$^{135}$ (exptl)</td>
<td>Pnma</td>
<td>4.76</td>
<td>1993$^{132}$</td>
</tr>
<tr>
<td>Y$^{3+}$</td>
<td>Co$^{4+}$</td>
<td>16.23</td>
<td>−1.02</td>
<td>1.46$^{135}$ (theor)</td>
<td>Pnma</td>
<td>4.34</td>
<td>n/a</td>
</tr>
<tr>
<td>Bi$^{3+}$</td>
<td>Fe$^{3+}$</td>
<td>10.71</td>
<td>−0.99</td>
<td>2.50$^{105}$ (exptl)</td>
<td>R3c</td>
<td>3.99</td>
<td>1234−1336$^{135}$</td>
</tr>
<tr>
<td>Bi$^{3+}$</td>
<td>Co$^{4+}$</td>
<td>10.71</td>
<td>−0.99</td>
<td>1.70$^{105}$ (exptl)</td>
<td>P4mm</td>
<td>3.80</td>
<td>729$^{106}$</td>
</tr>
</tbody>
</table>

*Compounds in boldface type might exhibit simultaneous reduction of the A- and B-site cations. The model $E_v$ is calculated from the experimental $\sum E_g$ value (using eq 8), experimental $V_r$ value (using eq 9), experimental (exptl) or theoretical (theor) $E_g$ value, and MP PBE+U-calculated $E_{\text{bulk}}$ value (not listed because all $E_{\text{bulk}} \approx 0$). We include the measured melting temperatures ($T_{90}$) of the candidates because thermal stability is also a significant screening criterion for STCH.

Similarly, for $n = 3$ (i.e., B$^{3+}$), the $\langle E_v \rangle$ trend (Figure 7b) nearly exactly anticorrelates with that of $V_r[B^{3+} \rightarrow B^{2+}]$ (R = −1.00): Cr (−2.87 V vs O$_2$) < Ti (−2.75 V vs O$_2$) < V (−2.01 V vs O$_2$) < Fe (−1.44 V vs O$_2$) < Mn (−1.02 V vs O$_2$) < Co (−0.90 V vs O$_2$) < Ni (0.05 V vs O$_2$). Finally, for Co and Ni, Figure 7a−c shows that $E_v$ increases with decreasing $n$, which is consistent with the relative instability of their higher oxidation states.

For simplicity in the discussion of materials design strategies, we bin $E_v$ into three discrete ranges: low $E_v$ ($E_v \leq 2.5$ eV), intermediate $E_v$ (2.5 < $E_v$ < 5 eV), and high $E_v$ ($E_v > 5$ eV). Low-$E_v$ oxides are more ideal for SOFCs, thermochemical O$_2$ storage and pumping/separation, thermochemical energy storage, and thermochemical air separation, with optimal B-site cations being Fe$^{3+}$, Co$^{3+}$, and Ni$^{4+}$. While perovskites containing these cations are consistent with state-of-the-art (La,Sr)CoFeO$_3$ SOFC cathodes, this reprediction of their optimality highlights the model’s accuracy and potential transferability to superternary perovskite metal oxides.

Intermediate-$E_v$ metal oxides are more suitable for STCH (WS and/or CDS) because the interactions between O$^{2-}$ and its nearest cation neighbors should be neither too strong for extensive thermal reduction nor too weak for spontaneous water splitting. The best B-site cations for STCH are V$^{4+}$, Mn$^{3+}$, Fe$^{3+}$, Co$^{3+}$, and Ni$^{4+}$. Mn$^{3+}$ appears in Ca-, Ce-, and Mn-based ABO$_3$ perovskites, which we recently identified to be promising candidates that potentially exhibit simultaneous reduction of both Ce$^{3+}$ (A-site) and Mn$^{3+}$ (B-site). While the model-predicted $E_v$ for Mn$^{3+}$ seems like it is too low for STCH, BaCe$_{1/4}$Mn$_{3/4}$O$_3$ is one of the best water splitters to date.$^{119}$ This underestimation is likely due to differences between its $E_g$ value and that of BaMnO$_3$ (since Mn$^{3+}$ is the redox-active cation in both stable materials and therefore their $\sum E_g$, $V_r$, and $E_{\text{bulk}}$ values should be similar), thus making band gap prediction for superternary perovskite metal oxides, an important avenue for improving the model.

On the basis of these B-site cations, there are three ternary metal oxides, an important avenue for improving the model. On the basis of these B-site cations, there are three ternary metal oxides, an important avenue for improving the model.

For prototype 2 or 3, we used the following scheme for enumerating promising candidates, outside of our training data. First, we screened for elements known to exist most commonly in the 3+/4+ oxidation state.$^{123}$ Second, we selected only those 3+/4+ cations whose ionic radii ($r_\text{A}$) produce a Goldschmidt tolerance factor$^{91}$ ($t$) greater than 0.71 (i.e., the stability limit of the perovskite structure), where:

$$t = \frac{r_A + r_B}{\sqrt{2(r_B + r_O)}}$$

$r_\text{B}$ is the ionic radius of the six-fold-coordinated B$^{3+/2+}$, and $r_O$ is the ionic radius of the six-fold-coordinated O$^{2-}$ (1.40 Å). Note that $t$ has a positive rate of 74%, so even if the tolerance criterion predicts a perovskite structure to form, the composition still may not actually form. Recently, Bartel et al. developed a more accurate tolerance factor (positivity rate of 92%).$^{124}$ However, we use eq 10 because its solution is analytical and therefore easier to solve for the relevant range of $r_B$ for which $t > 0.71$. That said, a numerical solution of the new tolerance factor$^{124}$ for $r_B$ is not difficult to obtain and can be a strategy for subsequent studies. We selected only those compositions whose MP-PBE+U-calculated $E_{\text{bulk}} \leq 0.025$ eV/atom, i.e., those that are predicted to be stable at room temperature. Table 1 gives the particularly promising candidates. (La,Co)$_{3+}$ (Mn,Fe,Co)$_{3+}$O$_3$ have been screened theoretically as candidates before by Emery et al.,$^{19}$ whereas those in Table 1 are new materials with known synthesis procedures. For example, Bi(Fe,Co)O$_3$ has a model $E_v$ value in the target range of 3.4−3.9 (±0.1) eV for improved STCH materials.$^{7}$ Of these, BiFeO$_3$ is especially interesting because it is a promising multiferroic material with several synthesis routes already proven.$^{125}$ In addition, the absolute differences between the Bi$^{3+}$ and Fe$^{3+}$ $V_r$ values ($\{V_r[B^{3+} \rightarrow B^{2+}] - V_r[Fe^{3+} \rightarrow Fe^{2+}]\}$ = 0.46 V vs O$_2$) and the Bi$^{3+}$ and Co$^{3/4+}$ $V_r$ values ($\{V_r[B^{3+} \rightarrow B^{2+}] - V_r[Co^{3/4+} \rightarrow Co^{3+/2+}]\}$ = 0.03 V vs O$_2$) are small in comparison to that of the simultaneously redox-active Ce$^{3+}$ and Mn$^{3+}$ cations in Ca-, Ce-, and Mn-based ABO$_3$ perovskites ($\{V_r[Ce^{3+} \rightarrow Ce^{2+}] - V_r[Mn^{3+} \rightarrow Mn^{2+/3+}]\}$ = 1.00 V vs O$_2$). Consequently, Bi(Fe,Co)O$_3$ might exhibit higher simultaneous redox activity and therefore a larger entropy of reduction. On the negative side, BiFeO$_3$ melts at a temperature ($T_{90}$) of 1234−1336 K, which is lower than a STCH cycle demands (≈2000 K). Encouragingly, the $T_{90}$ values can be increased via Co doping,$^{126}$ so the quaternary species could prove sufficiently robust. Note that we list YFeO$_3$ as promising—despite its model $E_v$ (4.76 eV) lying outside the target range (≈3.4−3.9 eV)—because its $T_{90}$ value (1993 K) is greater than that of the $E_v$-ideal Bi(Fe,Co)O$_3$ (1234−1336 K) and therefore Y-doped Bi(Fe,Co)O$_3$ might be a promising candidate for STCH.

Finally, high-$E_v$ metal oxides are optimal for semiconductors and ferroelectrics because their performance can be limited by the formation of deleterious defects. For example, VO$_5$ induces ferroelectric domain wall pinning in BiFeO$_3$$^{127}$ (intermediate $E_v$) but not in A$^{2+}$Ti$^{4+}$O$_3$.$^{36}$ (high $E_v$), which restricts polarization switching in the former. The most favorable B-site cations for high $E_v$ include Ti$^{4+}$, Ti$^{3+}$, V$^{3+}$, and Cr$^{3+}$, which is consistent with the common ferroelectric crystals...
BaTiO$_3$ and PbTiO$_3$,$^{128}$ and those less explored, such as CeCrO$_3$,$^{129}$ and ScCrO$_3$.$^{130}$

**Model Simplification.** When quantum mechanics calculations are unavailable, $E_b$ and $V_r$ can be calculated from experimental$^{107-110}$ formation energies/enthalpies and element cohesive energies.$^{112}$ Moreover, $E_b$ can be obtained from the experimental literature or computational databases$^{113,137-140}$ or by performing a simpler single-step self-consistent SCAN+$U$ calculation. In cases where $E_b$ is unknown or unreliable, we constructed simplified models for room-temperature-stable perovskites:

$$E_v = -1.8V_r - 66.0E_{\text{hull}} + 2.7$$  \hspace{0.5cm} (11)

and all perovskites:

$$E_v = 0.1 \sum E_b - 1.7V_r - 6.5E_{\text{hull}} + 0.5$$  \hspace{0.5cm} (12)

with MAEs of 0.50 ± 0.05 and 0.74 ± 0.05 eV, respectively. Note that in eq 11 the estimated coefficient for $\sum E_b$ is −0.03 ± 0.15 eV and therefore fails to reject the null hypothesis. Figure S10 in the Supporting Information shows the performance of our simplified $E_v$ models (eqs 11 and 12).

In scenarios where $E_{\text{hull}}$ is undetermined or untrustworthy (e.g., for hypothetical and/or strongly correlated phases that are not present in computational materials databases such as MP, OQMD, and AFLOW), we suggest the following strategy when it is impractical to compute the phase diagram. First, one should use the model for stable perovskites in either Figure 3a (with $E_b$) or eq 11 (without $E_b$). If $E_{\text{hull}} = 0.025$ eV/atom, then $E_{\text{hull}} = -1.39 \pm 0.40$ and $-1.65 \pm 0.29$ eV for the model with and without $E_b$, respectively. Therefore, if the perovskite is room-temperature-stable, then the maximum decrease in $E_v$ is 1.39–1.65 ± 0.40 eV. Second, since stability often is desired, if the application-specific ideal- $E_v$ range encompasses the model $E_v$ value and the model $E_v = -1.39$ value (with $E_b$) or $E_v = -1.65$ value (without $E_b$), then the material is worth considering.

### DISCUSSION

**Comparison with a Prevailing Model.** In the first part of the discussion, we compare our model to that of Deml et al.$^{66}$ (the state-of-the-art), which was trained on 45 insulating metal oxides, including binary and ternary main-group and transition-metal oxides in the antifluorite, corundum, rock salt, rutile, perovskite, and spinel structures. Note that both Deml et al. and we considered unstable metal oxides in our databases—58% (42%) of the metal oxides they (we) consider are room-temperature stable. Their model is:

$$E_v = 0.72 \left[ \sum \Delta H_f + 0.6 \left( E_{\text{PBE}} + \frac{c_F}{2} E_g + 2.60 \langle \Delta \chi \rangle \right) \right]$$  \hspace{0.5cm} (13)

where $\Delta H_f$ is the metal oxide formation enthalpy in eV/atom, $E_{\text{PBE}}$ is the O 2p band center in eV, $\langle \Delta \chi \rangle$ is the average Pauling electronegativity difference between O$^{2-}$ and its nearest cation neighbors, and $c_F = 1.00$ (1.50) for the GW-calculated (PBE+U-calculated) $E_g$ in eV. Equation 13 has training and testing MAEs of 0.19–0.21 eV (the first value is for GW and the second is for PBE+U gaps) and 0.39 eV, respectively, where the latter is slightly better than our cross-validated MAE for room-temperature-stable materials (0.45 ± 0.04 eV). When it is normalized by the range of $E_b$, however, their testing MAE (9%) is slightly worse than ours (6–7%).

Despite the differences between some of the terms in our models, there is an underlying relationship between $\Delta H_f$ and $V_r$. To this end, consider the decomposition of binary M$^{n+}$ oxides:

$$\text{MO}_{n/2} \rightarrow M + (n/4)O_2$$  \hspace{0.5cm} (14)

where the energy release upon decomposition is:

$$\Delta E_f = E_f[M^{n+} \rightarrow M^0] = -\Delta H_f$$  \hspace{0.5cm} (15)

Inserting $E_f[M^{n+} \rightarrow M^0]$, $n = 4$, $m = 0$, and $F$ into eq 5 yields:

$$V_f[M^{n+} \rightarrow M^0] = \frac{\Delta H_f}{nF}$$  \hspace{0.5cm} (16)

Despite the direct mathematical relationship between $\Delta H_f$ and $V_r$, two important effects are better accounted for with the crystal $V_r$ value used in our model than with $\Delta H_f$. First, a neutral $V_r$ reduces its neighboring cations; hence, the formation energy per reducing electron ($V_r$) provides a more consistent comparison across metal oxides with varied oxidation states. Second, $\Delta H_f$ neglects the presence of other stable nonzero oxidation states for the reduced metal, while the $V_r$ value defined in our work better accounts for oxidation state changes. For example, Ti$^{4+} \rightarrow$ Ti$^{3+}$ is more sensible than Ti$^{2+} \rightarrow$ Ti$^{0}$ when CaTiO$_3$ is undergoing reduction because the Ca-induced two-electron reduction typically involves two or more cation acceptors rather than reduction of one cation all the way to its metallic state.

To provide a more quantitative comparison between our model and that of Deml et al.$^{66}$ we fit a model (using training methods similar to those described in the Computational Methods) to their PBE+U-calculated $E_v$ values with our descriptors. Note that HgTiO$_3$ was excluded from the reported model of Deml et al. and its inclusion increases the MAE from the reported 0.21 to 0.27 eV. Our model for their data, i.e.,

$$E_v = -1.2V_r + 0.3E_g + 1.87 \text{ eV}$$  \hspace{0.5cm} (17)

highlights the important role of $V_r$ in governing $E_v$ trends, as it is one of only two metrics required to achieve MAE = 0.25 eV (see Figure 8). Notably, the coefficient we obtain for the $E_g$ term (0.3) is similar to their model (0.22 or 0.32 depending on whether the PBE+U- or GW-based model is used), while the contribution of other descriptors used in their model drops down to a single $V_r$ in ours. Additionally, our model is able to reproduce their PBE+U calculated $E_v$ with lower errors and fewer parameters, highlighting that our descriptors are both more accurate and yield better physical insight.

Revisiting HgTiO$_3$, the authors in ref 66 hypothesize that its unpredictability (see the deviation between the orange marker with black outline and the dashed black line) “arises from the inadequate treatment of the filled Hg d orbitals in DFT”. Our results show, however, that the inclusion of $V_r$ in the model enables a better prediction (AE = 0.45 eV vs 2.10 eV for our model vs their model, respectively) of the PBE+U-calculated $E_v$ value for HgTiO$_3$ (3.05 eV). This improvement can be ascribed to the greater reducibility and therefore control over the $E_v$ value of Hg$^{2+}$ (MP-PBE+U-calculated $V_f[\text{Hg}^{2+} \rightarrow \text{Hg}^0] = -0.65$ V vs O$_2$) relative to Ti$^{4+}$ (MP-PBE+U-calculated $V_f[\text{Ti}^{4+} \rightarrow \text{Ti}^{3+}] = -2.25$ V vs O$_2$), which is captured by $V_r$ but not by a combination of $\Delta H_f$, $E_{\text{PBE}}$, $E_g$, and $\Delta \chi$.

**Avenues for Improving and Extending the Model.** In the second part of the discussion, we first analyze the six outliers in Figure 3b. On the basis of our low-, intermediate-,
and high-$E_b$ bins, we classify these outliers as either quantitative (where the SCAN+U-calculated and model-predicted $E_v$ values are in the same bin) or qualitative (where these $E_v$ values are in different bins). The quantitative outliers are monoclinic (m-)LaCrO$_3$ (P2$_1$/c) and hexagonal (h-)CaTiO$_3$ (P6$_3$/mmc), whereas the qualitative outliers are rhombohedral (r-)BaCrO$_3$ (R3c), tetragonal (t-)LaCoO$_3$ (P4mm), cubic (c-)LaCoO$_3$ (Pm$ar{3}$m, ICSD #28921), and c-LaTiO$_3$ (Pm$ar{3}$m, ICSD #28908). In general, multiple possible reasons exist for deviations in the SCAN+U-calculated and model-predicted $E_v$ values. First, the intrinsic instability ($E_{\text{bulk}} > 0.042$ eV/atom for all six outliers) may catalyze nonlocal changes in the structure that models with local features cannot capture. That said, given the small number of outliers, $E_{\text{bulk}}$ captures most of the stability-related contributions to $E_v$. Second, for simplicity in the screening process, we aligned ferromagnetically the magnetic moments of all cations. Despite this approximation leading to a good description of many basic properties of standard ferromagnetic materials, such as permanent magnetic moments and phase diagrams, antiferromagnetism, relevant for the Cr-containing outliers, may lead to inconsistencies between the initial and final spin configurations. Finally, although we considered the high-$E_b$, intermediate-$E_b$, and low-spin states for the Co-containing outliers, a more comprehensive analysis of their magnetic structures is left for future work.

Next, we outline avenues for improving the model. One limitation of the model is using $E_b$ from binary systems, as the actual bond lengths in ternary perovskites (and beyond) structures may be significantly different from the binary oxide bond lengths. So, to account for this, one could either do more calculations to see how $E_b$ changes with bond length for individual systems or fit a simple function (e.g., an exponent or a Gaussian centered around the binary metal oxide bond length) as a weighting factor. In any case, modifying $E_b$ should lead to small improvements in the MAE of models since $c_b$ is quite small.

With respect to the dependence of $E_v$ on $V_n$, our ideas for refining the model are motivated by the fact that the cross-validated $c_v$ value can be understood as the number of electrons ($N_e$) donated by an O$_2^-$ to its most reducible cation neighbors upon V$_O$ formation. Moreover, $N_e$ being approximately equal to 1.5 (see Figure 3a) and not 2 is related to how the cation-reduction energy cost for creating V$_O$ gets split. $V_n$ contributes about 75–80% of the cost, whereas ~15% of the cost is accounted for by $E_b$ (which captures crystal field effects and any delocalization of electrons upon V$_O$ formation); this combination is why $c_v$ splits across 1 to 1.9. We hypothesize that the 10% remaining contribution comes from stability and bond-dissociation effects (where ions that do not undergo reduction can contribute). To explore further the role of electron delocalization in determining the $c_v$ splitting, it is important to consider that this approach be extended to other defect types such as oxygen vacancy. For electronically diverse compounds whose anions neighbor O in the periodic table, such as metallic transition-metal phosphides for electrocatalysis, semiconducting chalcogenides (sulfides and selenides for solar energy conversion), and hybrid organic-inorganic metal halides (e.g., for optoelectronics),

**CONCLUSIONS**

Metal oxide perovskites have received intense scrutiny for clean energy and computer memory applications due to their flexible structure, stability, wide composition space, and mixed ionic electronic conductivity. While exciting progress continues, a simple phenomenological explanation for the trends in oxygen vacancy ($V_O$) formation energies ($E_v$), which play a critical role in determining the performance of perovskite metal-oxide-based devices, has remained elusive. Here, we have introduced a machine-learned linear model that is based on $E_v$ values of ABO$_3$ perovskites, where $A = \text{Ca, Sr, Ba, La, Ce}$ and $B = \text{Ti, V, Cr, Mn, Fe, Co, Ni}$, in six lattice systems (monoclinic, orthorhombic, tetragonal, rhombohedral, hexagonal, and cubic). The model takes as inputs crystal bond dissociation energies, crystal reduction potentials, band gaps, and energies above the convex hull, which one can obtain from theoretical or experimental databases. Notably, we have devised crystal bond dissociation energies and crystal reduction potentials as the solid-state variants of their molecular and aqueous electrochemistry analogues, respectively, to systematically improve the description of bond dissociation and reduction of multivalent cations in solids. Additionally, we demonstrate that the model can be simplified, with acceptable losses in accuracy, such that only crystal bond dissociation energies and crystal reduction potentials are needed in scenarios where band gaps and energies above the convex hull are difficult to
estimate. Also, we predict BiFeO$_3$ and BiCoO$_3$ or a mixture might be viable candidates for solar thermochemical water and/or carbon dioxide splitting applications (intermediate $E_v$, if the melting point can be pushed high enough: e.g., via Y-doping), with CeCrO$_3$ and ScCrO$_3$ as possible V$_0$-resistant ferroelectrics (high $E_v$), indicating the effect that our model can make in materials discovery. Adding further credibility to our model is our rediscovery of Co-based perovskites as robust solid oxide fuel cell cathodes (low $E_v$). Finally, we present our perspectives on how to improve and extend the model, which already provides both accurate and efficient predictions for high-throughput screening and an intuitive and modular phenomenology for applications to metal oxide perovskites and beyond.

**ASSOCIATED CONTENT**

Supporting Information
The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jacs.1c05570.

Further details about the distribution of ABO$_3$ perovskite structures, distribution of neutral oxygen vacancy formation energies, convergence of oxygen vacancy formation energies in different crystal systems with respect to supercell size, distribution of energies above the convex hull, periodic trends in experimental cohesive energies, cell-size convergence of total energies for neutral atoms in their electronic ground states, theoretical and experimental cohesive energies across all metal oxides considered, correlation of experimental bond dissociation energies for neutral diatomic molecules and diatomic cations with respect to SCAN+$U$-calculated crystal bond dissociation energies, specific trends in the SCAN+$U$-calculated and experimental crystal reduction potentials, and SCAN+$U$-calculated vs simplified-model-predicted oxygen vacancy formation energies (PDF)

Comma-separated-value files containing crystal structures, total energies, magnetic moments, charges, band gaps, and crystal features for all systems considered (ZIP)

**AUTHOR INFORMATION**

Corresponding Author
Emily A. Carter — Department of Mechanical and Aerospace Engineering, Princeton University, Princeton, New Jersey 08544-5263, United States; Office of the Chancellor and Department of Chemical and Biomolecular Engineering, University of California, Los Angeles, Los Angeles, California 90095, United States; orcid.org/0000-0001-7330-7554; Email: eac@princeton.edu, eac@ucla.edu

Authors
Robert B. Wexler — Department of Mechanical and Aerospace Engineering, Princeton University, Princeton, New Jersey 08544-5263, United States; orcid.org/0000-0002-6861-6421

Gopalakrishnan Sai Gautam — Department of Mechanical and Aerospace Engineering, Princeton University, Princeton, New Jersey 08544-5263, United States; Present Address: Department of Materials Engineering, Indian Institute of Science, Bengaluru, Karnataka 560012, India; orcid.org/0000-0002-1303-0976

Ellen B. Stechel — ASU LightWorks and the School of Molecular Sciences, Arizona State University, Tempe, Arizona 85287-5402, United States; orcid.org/0000-0002-5379-2908

Complete contact information is available at: https://pubs.acs.org/doi/10.1021/jacs.1c05570

Notes
The authors declare no competing financial interest.

**ACKNOWLEDGMENTS**

The authors gratefully acknowledge research support from the HydroGEN Advanced Water Splitting Materials Consortium, established as part of the Energy Materials Network under the U.S. Department of Energy, Office of Energy Efficiency and Renewable Energy, Fuel Cell Technologies Office, under Award Number DE-EE0008090. The authors also acknowledge the computational resources sponsored by the Department of Energy’s Office of Energy Efficiency and Renewable Energy located at the National Renewable Energy Laboratory. The authors also thank Princeton University for computing resources. The views and opinions of the authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights.

**REFERENCES**


