

---

# Steerable Wavelet Scattering for 3D Atomic Systems with Application to Li-Si Energy Prediction

---

**Xavier Brumwell\***  
Department of CMSE<sup>†</sup>  
Michigan State University  
brumwel1@msu.edu

**Paul Sinz\***  
Department of CMSE<sup>†</sup>  
Michigan State University  
sinzpaul@msu.edu

**Kwang Jin Kim**  
Department of ChEMS<sup>‡</sup>  
Michigan State University  
kimkwa12@msu.edu

**Yue Qi<sup>§</sup>**  
Department of ChEMS<sup>‡</sup>  
Michigan State University  
yueqi@egr.msu.edu

**Matthew Hirn<sup>¶</sup>**  
Department of CMSE<sup>†</sup>  
Department of Mathematics  
East Lansing, MI 48864  
mhirn@msu.edu

## Abstract

A general machine learning architecture is introduced that uses wavelet scattering coefficients of an inputted three dimensional signal as features. Solid harmonic wavelet scattering transforms of three dimensional signals were previously introduced in a machine learning framework for the regression of properties of small organic molecules. Here this approach is extended for general steerable wavelets which are equivariant to translations and rotations, resulting in a sparse model of the target function. The scattering coefficients inherit from the wavelets invariance to translations and rotations. As an illustration of this approach a linear regression model is learned for the formation energy of amorphous lithium-silicon material states trained over a database generated using plane-wave Density Functional Theory methods. State-of-the-art results are produced as compared to other machine learning approaches over similarly generated databases.

## 1 Introduction

Modeling the behaviors of complex multiscale materials has become a domain of ever increasing research interest bringing together experts from a variety of fields ranging from chemistry, biology, and physics, to experts in numerical methods, both experimental and theoretical. The multiscale nature of such materials implies a dependence of material properties upon multiple length and/or time scales. This leads to several fundamental challenges such as the curse of dimensionality, in which the discretization of a material grows exponentially with system size. The coupling of scales is frequently nonlinear, meaning that it is not sufficient to simply separate length scales in computations and recombine the results in a linear fashion. The challenge of multiscale materials modeling is therefore to create models of phenomena of interest which include the influence and coupling of multiple scales in length and time while maintaining acceptably low computational costs. In the computation of energies of atomic systems, for example, plane-wave density functional theory (DFT) provides

---

\*Contributed equally

<sup>†</sup>Computational Mathematics, Science & Engineering

<sup>‡</sup>Chemical Engineering and Materials Science

<sup>§</sup><https://researchgroups.msu.edu/msce>

<sup>¶</sup>Corresponding author; <https://matthewhirn.com/>

state-of-the-art prediction capabilities for computing the ground state energy of a periodic atomic system; however, the computational costs of DFT restrict its practical usage to periodic systems on the order of a few hundred atoms [1]. This has led to the development of novel numerical methods with the chief objectives being computational speedups and accuracies that are competitive with the state-of-the-art methods (e.g., DFT). Machine learning models trained on datasets constructed with trusted methods have seen growing interest in recent years [2, 3, 4, 5, 6, 7, 8, 9, 10]. We introduce one such model as the focus of this work.

We present a general machine learning architecture for the modeling of multiscale properties of complex materials. Any such machine learning algorithm must address the fundamental challenges in materials modeling, including the multiscale nature of many materials in addition to properties of invariance and stability. The machine learning architecture proposed here consists of a linear regression over a new type of three dimensional wavelet scattering coefficients [11], which encode localized fluctuations of signals that represent the atomistic state of the system. The scattering transform has an architecture similar to that of convolutional neural networks (CNNs), but replaces learned, fixed scale filters with a set of designed, multiscale wavelets. The resulting transform thus not only separates the scales of the atomistic system, but also extracts complex coefficients that incorporate geometric patterns at multiple scales. In order to obtain a rotationally invariant representation of the state of the atomistic system, this new scattering transform is constructed out of a general class of three dimensional steerable wavelets, which can be considered as generalizations of two dimensional steerable filters [12]. Within this framework, we additionally introduce a specific family of wavelets, which we refer to as atomic orbital wavelets, that generalizes the family of solid harmonic wavelets introduced in [13].

Using the wavelet scattering coefficients, we train a linear regression model for predicting the formation energy of amorphous lithium-silicon (Li-Si) systems. The resulting model has an MAE of  $0.78 \pm 0.01$ meV/atom and an RMSE of  $1.24 \pm 0.05$ meV/atom. Formation energies are the fundamental driving force for Li storage in Li-Si systems, which is of special interest for next generation high energy Li-ion batteries. Lithiation and delithiation of Si have been widely studied using reactive force field based MD simulations [14, 15]. However, constructing interatomic potentials (or force fields) is a repetitive task of training on extensive first-principles calculations, generally resulting in several months of laborious iterative fitting. Furthermore, computed force fields do not generalize well due to the complex interdependence of the parameters. A paradigm shift is taking place in which machine learning potentials are rising as an alternative to harness first-principles accuracy and force field efficiency [16, 17, 18, 19, 20, 21, 22, 23, 24].

The remainder of this paper is organized as follows. In section 2 we introduce a general class of three dimensional, steerable wavelets and the resulting equivariant scattering transform. Invariant scattering coefficients are derived in section 3. The Li-Si database is described in section 4, and in section 5 numerical regression results on the formation energies of these systems, in addition to further analysis, are presented. Section 6 contains some concluding remarks.

## 2 Equivariant scattering networks with steerable wavelet filters

### 2.1 Periodic steerable wavelets

We introduce a general class of 3-dimensional steerable wavelet filters for which the resulting wavelet transform is equivariant with respect to translations and rotations when combined with an appropriately chosen nonlinearity. The equivariance of the transform leads to the associated invariance of the scattering coefficients introduced in section 3. These wavelet filters have the form

$$\Psi_{\gamma l}^m(u) = Q_{\gamma}(r)Y_l^m(\phi, \theta), \quad (r, \phi, \theta) \in \mathbb{R}^3, \quad (1)$$

for an arbitrary radial function  $Q_{\gamma}$  and the well-known spherical harmonic functions  $Y_l^m$ , indexed by parameters  $l = 0, 1, 2, \dots$  and  $-l \leq m \leq l$ . Here we use spherical coordinates with  $r$  the radial component,  $\phi \in [0, 2\pi]$  the azimuthal angle, and  $\theta \in [0, \pi]$  the polar angle. While equivariance can be obtained for any  $Q_{\gamma}$  (see section 2.2), we constrain  $Q_{\gamma}$  so that the resulting function  $\Psi_{\gamma l}^m$  is either a wavelet or low pass filter. There is therefore great freedom of choice in the radial component. Guidance in choosing  $Q_{\gamma}$  may come from knowledge of the application or the types of signals to be analyzed; it can also be learned from data in a task driven fashion. In example 1 below, as well as the numerical experiments reported in section 5, we specify a class of radial functions that are useful for

modeling atomic systems. In this case it is also beneficial (for computational efficiency) to choose  $Q_\gamma$  so that the Fourier transform of the wavelet,  $\widehat{\Psi}_{\gamma l}^m$ , can be computed analytically.

Wavelets  $\Psi_{\gamma l}^m \in L^2(\mathbb{R}^3)$  are filters with zero average,  $\int_{\mathbb{R}^3} \Psi_{\gamma l}^m = 0$ , and which are localized (they have fast decay) in space and frequency. A multiscale wavelet transform of a signal  $\rho \in L^2(\mathbb{R}^3)$  is given by the convolution of the signal with dilations of the wavelets  $\Psi_{\gamma l j}^m(u) = 2^{-3j} \Psi_{\gamma l}^m(2^{-j}u)$ . For periodic signals  $\rho \in L^2(\mathbb{T}^3)$  (where  $\mathbb{T}$  is the torus), which are used to model materials, we periodize the wavelets and compute circular convolutions. Periodized wavelets with period  $P$  are defined as:

$$\psi_{\gamma l j}^m(u) = \sum_{k \in \mathbb{Z}} \Psi_{\gamma l j}^m(u - kP)$$

and circular convolutions over  $\mathbb{T}$  are denoted by  $\circledast$ . A periodized wavelet can be obtained efficiently by appropriately sub-sampling its Fourier transform. The resulting wavelet coefficients are:

$$W\rho = \{\rho \circledast \psi_{\gamma l j}^m(u) : \gamma \in \Gamma, j \in \mathbb{Z}, -l \leq m \leq l, 0 \leq l \leq L\}.$$

Choosing wavelets with simple Fourier transforms allows for  $W\rho$  to be computed in frequency and brought back to space via the Fast Fourier Transform rather than computing the convolution directly, allowing for considerable time savings.

The wavelet transform can be viewed as an alternative to the Fourier transform. Whereas the Fourier transform extracts responses against complex sinusoidal waveforms with a continuum of frequencies along the entire domain of a signal, the wavelet transform uses wavelets to filter out behavior localized in space which is captured by the various dilations of the mother wavelets.

For the present study we illustrate the flexibility of (1) by introducing wavelets where the radial component depends on  $\gamma = (n, l)$ , where  $n$  is an extra parameter. In this way we may create multiple mother wavelets that independently vary in radial and angular directions as specified by the parameters  $n$ ,  $l$ , and  $m$ . Example 1 below introduces a specific choice of  $Q_{nl}$  that is designed for the regression of energies of atomic systems. We refer to these wavelets as atomic orbital wavelets. These wavelets are used in all scattering regressions, which are described in section 5.

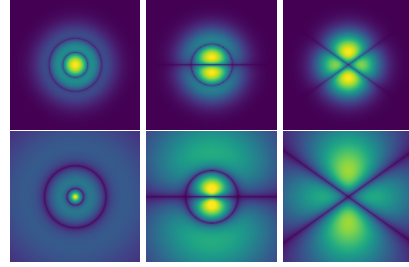


Figure 1: Top row left to right: Density plot cross sections in the  $xz$ -plane of atomic orbital wavelets for  $(n, l, m) = (3, 0, 0)$ ,  $(3, 1, 0)$ , and  $(3, 2, 0)$ . Bottom Row: Corresponding plots for the hydrogen atom orbitals ( $3s$ ,  $3p$ , and  $3d$  orbitals, respectively). All images are rescaled for visualization. Note that the exponential radial decay  $e^{-r}$  of the hydrogen orbitals is replaced by a Gaussian decay  $e^{-r^2}$  in the wavelets accounting for the greater localization of the wavelets.

**Example 1** *The atomic orbital wavelets have the form of (1) with radial component*

$$Q_{nl}(r) = K_{nl}(\alpha) r^l L_{n-l-1}^{l+1/2}(r^2/(2\alpha^2)) e^{-r^2/(2\alpha^2)}, \quad K_{nl\alpha} = \frac{2^{n-l-1} (n-l-1)!}{(\sqrt{2\pi})^3 \alpha^{2n+1}} \sqrt{\frac{4\pi}{2l+1}}$$

for normalizing factor  $K_{nl\alpha}$ , the usual quantum numbers  $n$ ,  $l$ ,  $m$ , and parameter  $\alpha$  corresponding to the standard deviation of the Gaussian which determines the locality of the mother wavelet. The Fourier transform has a simple form and is given by

$$\widehat{\Psi}_{nl}^m(\eta, \xi, \zeta) = (-i)^l \eta^{2(n-1)-l} e^{-\alpha^2 \eta^2 / 2} \sqrt{\frac{4\pi}{2l+1}} Y_l^m(\xi, \zeta).$$

Here  $L_k^\nu$  are the associated Laguerre polynomials defined by  $L_k^\nu(x) = (x^\nu k!)^{-1} (\frac{d}{dx} - 1)^k x^{k+\nu}$ . The frequency coordinates are written in spherical coordinates with  $\eta$  the radius,  $\xi \in [0, 2\pi]$  the azimuthal angle, and  $\zeta \in [0, \pi]$  the polar angle. These wavelets include the solid harmonic wavelets which are recovered for  $l = n - 1$  since  $L_0^\nu = 1$  for any  $\nu$ . The solid harmonic wavelets are used for molecular energy regression in [13, 25].

The choice of  $Q_{nl}$  in example 1 is made to mimic the geometry of the atomic orbitals of the hydrogen atom; these orbitals are closed form solutions to the time-independent Schrödinger equation. We note

that the solid harmonic wavelets of [13, 25] mimic the  $1s, 2p, 3d, 4f, \dots$ , orbitals of the hydrogen atom. This leaves out the majority of the orbitals which, however, the atomic orbital wavelets recover by mimicking the full set of orbitals:  $1s, 2s, 2p, 3s, 3p, 3d, \dots$ . The atomic orbital wavelets are a modification of the spherical Gaussian type orbitals used in [26], and references therein, where they are used for producing simple formulas for multi-centered molecular integrals.

Cross sections of three atomic orbital wavelets are illustrated in figure 1 along with plots of the corresponding orbitals associated to the hydrogen atom. Unlike atomic orbitals which are at a fixed scale, we utilize several dilations of the atomic orbital wavelets for use in the multiscale wavelet transform. In the next section we see that a nonlinear operator, similar to the modulus, applied to the wavelet coefficients is used to recover equivariance to rotations and reflections.

## 2.2 Nonlinear equivariance

Translation equivariance is automatically guaranteed for any convolution operator, that is, for  $t \in \mathbb{R}^3$ ,  $f \otimes g(u-t) = f_t \otimes g(u) = f \otimes g_t(u)$  where  $f_t(u) = f(u-t)$ . Rotations of the spherical harmonics are distributed across all values of the parameter  $m$  for a fixed  $l$  as

$$Y_l^m(Ru) = \sum_{m'=-l}^l [D_{mm'}^{(l)}(R)]^* Y_l^{m'}(u)$$

where  $[D_{mm'}^{(l)}(R)]^*$  is the complex conjugate of the  $(m, m')$  entry of the Wigner D-matrix  $D^{(l)}(R)$  associated to the rotation  $R \in SO(3)$ . It follows that the wavelet filters  $\Psi_{\gamma l}^m$  of (1) are steerable.

To recover rotationally equivariant coefficients we apply a nonlinear operator  $\sigma$  (which we shall refer to as the modulus), first introduced in [13], which is defined by

$$\sigma(\rho \otimes \psi_{\gamma l j})(u) = \left( \sum_{m=-l}^l |\rho \otimes \psi_{\gamma l j}^m(u)|^2 \right)^{1/2} \quad (2)$$

with a slight abuse of notation by omitting the summing variable  $m$  on the left hand side. The resulting set of coefficients

$$\sigma(W\rho) = \{ \sigma(\rho \otimes \psi_{\gamma l j}^m) : \gamma \in \Gamma, j \in \mathbb{Z}, 0 \leq l \leq L \}$$

are equivariant to rotations, as follows from Unsöld's Theorem and by noting the Wigner D-matrix is unitary; thus  $\sigma(\rho \otimes \psi_{\gamma l j})(Ru) = \sigma(\rho_R \otimes \psi_{\gamma l j})(u)$  for any rotation  $R \in SO(3)$  where  $\rho_R(u) = \rho(Ru)$  is a rotation of the input signal. The rotational equivariance and steerability of this class of wavelets is entirely dependent upon the spherical harmonics; this modulus operator guarantees rotational equivariance regardless of the choice of  $Q_\gamma$  in (1). Furthermore, the modulus operator guarantees invariance to reflections as  $Q_\gamma(-r) = Q_\gamma(r)$  and  $Y_l^m(-\mathbf{r}) = (-1)^l Y_l^m(\mathbf{r})$ .

## 2.3 Scale coupling through multiple layers

The strength of (1) and (2) lies not only in providing equivariance to rigid motions, but also in that the radial component can be tailor-made for a specific application. For multiscale problems choosing an appropriate radial component will allow for the separation of length scales of the input signal  $\rho$ . For example, the electronic density  $\rho$  of an atomic system depends upon interactions of electrons with other nearby electrons and nuclei as well as interactions with groupings of atoms at longer length scales. Separating scales is advantageous for computational purposes; however, to adequately model a multiscale system, the inherent coupling between scales must be recovered.

A second wavelet modulus may be computed to recover interactions between length scales. For example, if a wavelet coefficient shows a response to a certain geometric feature of the input at a given scale  $2^{j_1}$ , then the cross-scale interactions of this feature will become visible to wavelet modulus coefficients at a larger scale  $2^{j_2}$ . Therefore we compute second order wavelet modulus coefficients

$$\sigma(\sigma(\rho \otimes \psi_{\gamma_1 l_1 j_1}) \otimes \psi_{\gamma_2 l_2 j_2}) = \left( \sum_{m=-l_2}^{l_2} |\sigma(\rho \otimes \psi_{\gamma_1 l_1 j_1}) \otimes \psi_{\gamma_2 l_2 j_2}^m(u)|^2 \right)^{1/2}, \quad j_1 < j_2, \quad (3)$$

where we apply the modulus operator  $\sigma$  at each level to maintain the equivariance established for the first order coefficients. First order wavelet modulus coefficients encode interference within a signal  $\rho$  at the scale  $2^{j_1}$ . Applying a second order wavelet transform with the modulus operator then encodes the interference within the first order interference at the scale  $2^{j_2}$ , coupling these two scales. We may therefore compute a cascade of wavelet transforms, applying the nonlinear modulus operator at each level, and capture higher order interactions at each level

$$\sigma(\sigma(\cdots\sigma(\rho \otimes \psi_{\gamma_1 l_1 j_1}) \otimes \psi_{\gamma_2 l_2 j_2}) \cdots \otimes \psi_{\gamma_k l_k j_k}), \quad j_1 < \cdots < j_k. \quad (4)$$

The resulting set of coefficients at order  $k$  is a subset of

$$\underbrace{\sigma(W\sigma(W\cdots\sigma(W\rho)))}_{k \text{ times}} = \left\{ \sigma(\sigma(\cdots\sigma(\rho \otimes \psi_{\gamma_1 l_1 j_1}) \otimes \psi_{\gamma_2 l_2 j_2}) \cdots \otimes \psi_{\gamma_k l_k j_k}) \right\}_{\gamma_k, j_k, l_k}$$

which mimics the alternating cascade of linear and nonlinear transformations found in CNNs. Since each wavelet modulus operator  $\sigma \circ W$  is equivariant to rotations, the entire cascade is equivariant. The transform (4) is thus similar to an equivariant three dimensional CNN constructed out of steerable filters; related studies on equivariant neural networks have previously been undertaken for 2D CNNs [27, 28], groups and graphs [29, 30], and 2D scattering transforms [31].

In this work we restrict computations to first and second order cascades of wavelet modulus transforms. Figure 2 shows a cross-section of a three dimensional signal  $\rho$  that represents an atomic state, and the same cross-sectional view of the first and second order wavelet modulus coefficients. The wavelets used are those of example 1 and the input signal is a fictional electronic density given by a sum of normalized Gaussians scaled by electronic charges. The multiscale interactions are clearly visible. The second image from the left shows interactions between the five visible atoms in the cross-section along with interactions into and out of the page. The third image from the left contains second order coefficients of the prior image and shows similar but more complicated interference patterns. The fourth image shows perimeters of the atoms as well as connections between neighboring atoms and the final image is an associated second order coefficient which shows long length scale interference. We see here that different wavelets as specified by the parameters  $(n, l, j)$  capture very different scale interactions, from fine scales which show the atomic locations clearly, to large length scales that show larger interactions.

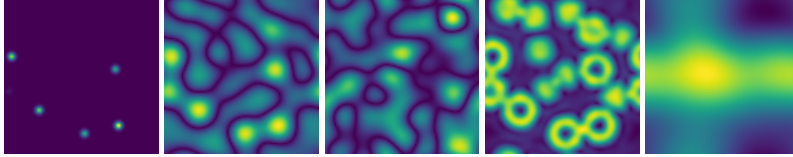


Figure 2: Cross-sectional views of a 3-dimensional signal  $\rho$  and its wavelet modulus coefficients specified by parameters  $(n, l, j)$  and  $(n_1, l_1, j_1, n_2, l_2, j_2)$  for  $\sigma(\rho \otimes \psi_{nlj})$  and  $\sigma(\sigma(\rho \otimes \psi_{n_1 l_1 j_1}) \otimes \psi_{n_2 l_2 j_2})$ . From left to right they are  $\rho$ ,  $(2, 0, 2)$ ,  $(2, 0, 2, 5, 0, 2.5)$ ,  $(5, 2, 1.5)$ , and  $(5, 2, 1.5, 5, 4, 2)$ . Values range from dark purple at zero to yellow at large values.

### 3 Invariant scattering coefficients

#### 3.1 Translation and rotation invariance

We seek features that are *invariant* to translations and rotations; however, the wavelet modulus coefficients (2), (3), and (4) are only *equivariant* to these transformations. Invariant coefficients are recovered by summing the wavelet modulus coefficients giving the first and second order wavelet scattering coefficients as

$$\|\sigma(\rho \otimes \psi_{\gamma l j})\|_q^q = \int_{\mathbb{R}^3} |\sigma(\rho \otimes \psi_{\gamma l j})(u)|^q du, \quad (5)$$

$$\|\sigma(\sigma(\rho \otimes \psi_{\gamma_1 l_1 j_1}) \otimes \psi_{\gamma_2 l_2 j_2})\|_q^q = \int_{\mathbb{R}^3} |\sigma(\sigma(\rho \otimes \psi_{\gamma_1 l_1 j_1}) \otimes \psi_{\gamma_2 l_2 j_2})(u)|^q du. \quad (6)$$

Here the wavelet modulus coefficients for a fixed  $(\gamma, l, j, q)$  or  $(\gamma_1, l_1, j_1, \gamma_2, l_2, j_2, q)$  are aggregated into a single number and are chosen as the features for prediction of properties of input signals

$\rho \in L^2(\mathbb{T}^3)$ . These coefficients are invariant to translations and rotations of  $\rho$  and encode information at a fixed scale of  $2^{j_1}$  in the first order and cross-scale features between  $2^{j_1}$  and  $2^{j_2}$  in the second order coefficients. Multiple integration factors  $q$  do not effect the invariance of the coefficients and are utilized in order to emphasize different responses of the input.

The specific integration factors may be chosen to correspond to interpretable phenomena. In the numerical experiments described in section 5 we select  $q = 1, 4/3, 5/3, 2$ . The power  $q = 1$  scales with the number of particles in the system, whereas it is shown in [32] that first order scattering coefficients for  $q = 2$  encode pairwise interactions, which capture the electrostatic Coulomb interactions between electrons. Furthermore, one can show that  $\|\rho\|_{5/3}^{5/3}$  combined with the collection  $\|\sigma(\sqrt{\rho} \otimes \psi_{n(n-1)j})\|_2^2$  capture the kinetic energy of the system, as approximated by the Thomas-Fermi-Dirac-von Weizsäcker (TFDW) model. Finally, a multiple of the  $L^{4/3}$  integral of  $\rho$  is a first order approximation of the exchange energy; see again the TFDW model [33].

While this analysis maps the scattering coefficients nicely onto the TFDW model, it is only an approximation of the true ground state energy of an atomistic state. Furthermore, there is information loss in the summing of the wavelet modulus coefficients. These considerations provide further motivation for including higher order scattering coefficients (6) as features in our regression model.

### 3.2 System size invariance

We seek to build a machine learning model to predict the formation energy per  $\text{Li}_\beta\text{Si}_{1-\beta}$  formula unit of an infinitely large periodic atomic state; see section 4 below. The formation energy is independent of system size (number of atoms), and therefore our features must recover this additional invariant. Normalizing the input signals to the wavelet transform gives signals  $\rho/\|\rho\|_q$  and  $\sigma(\rho \otimes \psi_{\gamma l j})/\|\sigma(\rho \otimes \psi_{\gamma l j})\|_q$ , which results in the following features that are invariant to system size:

$$\frac{\|\sigma(\rho \otimes \psi_{\gamma l j})\|_q^q}{\|\rho\|_q^q}, \quad \frac{\|\sigma(\sigma(\rho \otimes \psi_{\gamma_1 l_1 j_1}) \otimes \psi_{\gamma_2 l_2 j_2})\|_q^q}{\|\sigma(\rho \otimes \psi_{\gamma_1 l_1 j_1})\|_q^q}. \quad (7)$$

These features are normalized versions of the features (5) and (6).

Infinitely large amorphous systems are inherently disordered and, as such, can be modeled as a 3D compound Poisson process in space where the location of an atom is an arrival event of the process. In practice, plane-wave DFT simulations utilize a finite unit cell with periodic boundary that creates order within the system; however, this phenomena is mitigated by taking a large enough unit cell. In [34], general wavelets are used to compute scattering coefficients using a similar normalization to (7). There it is shown that normalized first and second order scattering coefficients characterize regular Poisson processes in 1D. We can thus interpret these scattering coefficients (7) as statistical moments of  $\rho$ . We train a second order model using these normalized features on a database specified in section 4 and discuss the results of our numerical experiments in section 5.

## 4 Database of Li-Si systems

Let  $x = \{(z_k, r_k)\}_{k=1}^N$  denote the state of an atomistic system, in which  $z_k$  denotes the total protonic charge of the  $k^{\text{th}}$  atom and  $r_k \in \mathbb{R}^3$  its position. In this paper we consider amorphous Li-Si systems  $\text{a-Li}_{\beta(x)}\text{Si}_{1-\beta(x)}$  for  $0 \leq \beta(x) \leq 1$  the atomic fraction of Li for state  $x$ . For notational clarity we write  $\beta$  rather than  $\beta(x)$  with the dependence on the state implied. The formation energy per formula unit of an atomic system is defined relative to the total energy per formula unit of the system,  $E(x)$ , the per atom energy of crystalline Si,  $E_{\text{Si}}$ , and the per atom energy of body-centered cubic Li,  $E_{\text{Li}}$ . The formation energy of an  $\text{a-Li}_{\beta}\text{Si}_{1-\beta}$  state  $x$  is computed as

$$E_f(x) = E(x) - \beta E_{\text{Li}} - (1 - \beta) E_{\text{Si}}. \quad (8)$$

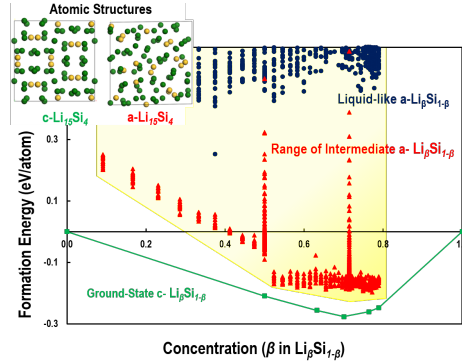


Figure 3: Formation energies of  $\text{a-Li}_\beta\text{Si}_{1-\beta}$ . Green crystalline states are not used but shown here for reference. In the atomic structures Li atoms are green and Si are yellow.

To generate an extensive training set that includes all local structural motifs and therefore covers the entire configuration space of  $a\text{-Li}_\beta\text{Si}_{1-\beta}$ , a wide range of  $a\text{-Li}_\beta\text{Si}_{1-\beta}$  structures with  $\beta$  ranging from 0 to approximately 0.79 were prepared. Figure 3 contains a scatterplot showing the range of energies computed and two example Li-Si structures. Liquid-like  $a\text{-Li}_\beta\text{Si}_{1-\beta}$  structures were generated for each of 37 concentrations ( $\beta$ ) using ReaxFF-MD at a temperature of 2500K for 10ps. Five random states were selected along each of the 37 trajectories (blue dots in figure 3). Then each state was relaxed with DFT-GGA computations at fixed volumes producing strings of intermediate energy states (red dots in figure 3). For clarity only two strings are shown in the figure, at  $\beta = 0.5, 0.71$ , the rest of the red dots show the final states of each string. The yellow region shows the configuration space that the intermediate states cover. The database is comprised of 44,596 states, and we note that the precision of the Li-Si training set is based on a DFT precision of 1 meV/atom. For a complete account of this database see [35].

## 5 Numerics

### 5.1 Model and Training Method

Recall that each state is specified by  $x = \{(z_k, r_k)\}_{k=1}^N$  for  $z_k$  the total protonic charge of the atom at  $r_k \in \mathbb{R}^3$ . We define a naive electronic density  $\rho_x$  as  $\rho_x(u) = \sum_k z_k g(u - r_k)$  with centered and normalized Gaussians  $g(u) = (\sqrt{2\pi}\alpha)^{-3} e^{-|u|^2/2\alpha^2}$ . Therefore  $\|\rho_x\|_1 = \sum_k z_k$  is the total protonic charge in the unit cell (periodic box) of the state  $x$ . Rather than use  $\rho_x$  as a single channel input to the scattering transform, we instead partition the state information into five density channels corresponding to the core electrons of the system, the valence electrons of the system, the Li atoms only, the Si atoms only, and the square root of the full system (to capture the kinetic energy). We denote these channels by  $\rho_x^\delta(u)$ , where  $\delta \in \{\text{core, valence, Li, Si, atomic sqrt}\}$ .

In this work our focus is the predictive capabilities of the wavelet scattering coefficients, and so we perform a simple linear regression over the scattering coefficients as

$$\tilde{E}_f(x) = w_0 + \sum_{\delta, \lambda_1, q} w_{\delta\lambda_1 q} \frac{\|\sigma(\rho_x^\delta \otimes \psi_{\lambda_1})\|_q^q}{\|\rho_x^\delta\|_q^q} + \sum_{\delta, \lambda_1, \lambda_2, q} w_{\delta\lambda_1\lambda_2 q} \frac{\|\sigma(\sigma(\rho_x^\delta \otimes \psi_{\lambda_1}) \otimes \psi_{\lambda_2})\|_q^q}{\|\sigma(\rho_x^\delta \otimes \psi_{\lambda_1})\|_q^q}, \quad (9)$$

subject to  $\|w\|_0 \leq M$ , with wavelet parameters  $\lambda = (n, l, j)$  (using the atomic orbital wavelets of example 1) and where we select  $M$  non-zero weights  $w = \{w_0, w_{\delta\lambda_1 q}, w_{\delta\lambda_1\lambda_2 q}\}$  using greedy orthogonal least squares regression [32, Sec. 3.2]. For small  $M$  we thus obtain a sparse regression, which can lead to an interpretable model.

### 5.2 Numerical Results and Analysis

While there is potential for using the scattering coefficients to predict many properties, the aim for this paper is to predict the formation energy to high accuracy. Motivated by the precision of DFT, we seek to obtain a regression accuracy of 1.0 meV/atom relative to the DFT computed energies.

To robustly test our model against all points in the data set, we used five-fold cross validation. Folds were selected randomly. In numerical computations we used the five density channels described in section 5.1, six wavelet scales increasing by factors of  $\sqrt{2}$ , atomic orbital wavelets for  $1 \leq n \leq 5$  and all valid choices of  $l$ , and  $q = 1, 4/3, 5/3, 2$ . Second order scattering coefficients were restricted so that  $j_2 > j_1$  and  $(n_2, l_2) > (n_1, l_1)$  (by dictionary ordering). The hyperparameter  $M$ , indicating the number of scattering coefficients to use in the linear model (9), was selected using cross validation on the training set. Bagging was used to learn multiple models for each test set, and the final predicted formation energy was taken as the average of the predicted energies of each model.

Table 1: Formation energy regression errors on Li-Si test data measured in meV/atom. Note that while training and test data sets are qualitatively similar across the methods, they are not the same and thus the average errors presented here are not directly comparable.

|      | Artrith [16] | Onat [17] | First order scattering | First & second order scattering |
|------|--------------|-----------|------------------------|---------------------------------|
| MAE  | 5.9          | –         | $2.13 \pm 0.02$        | $0.78 \pm 0.01$                 |
| RMSE | 7.7          | 4.5       | $2.95 \pm 0.04$        | $1.24 \pm 0.05$                 |

Utilizing only first order scattering coefficients in the linear model (9), we obtain a mean absolute error (MAE) of 2.1 meV/atom on the testing folds, which surpasses the results of neural network models trained and tested on similar Li-Si databases as reported in [16, 17]. We note, however, that the data in [16, 17] is not currently available, and thus this is not a precise comparison. In order to reduce the regression error, we also trained the linear model (9) using both first and second order scattering coefficients. With these additional coefficients, the linear regression model (9) achieved an MAE of 0.78 meV/atom with a standard deviation across the five folds of 0.01 meV/atom. Table 1 summarizes these results.

Our results display several expected machine learning phenomena that indicate the scattering coefficients are capturing the structural variability in the Li-Si database. The top of figure 4 plots the RMSE as a function of  $M$ , the number of scattering coefficients used in the linear model (9). It shows the error decays steeply at first before flattening, which indicates the first few selected features are highly correlated with the formation energy; one can thus obtain a very sparse (i.e., simple) model that still obtains good accuracy (e.g.,  $M = 100$  scattering coefficients obtains an RMSE of 5.8 meV/atom). Also of importance, the validation curves maintain close proximity to the training curves, which suggests that we are learning more than the noise of the training set. Of particular note in figure 4 is the slope of the error curve remains negative, meaning there is a potential to extend the model using additional scattering coefficients and improve the accuracy further.

The greedy selection process enables us to extract an ordering of the scattering coefficients which were most critical in predicting the energy. Figure 4 (bottom) plots the corresponding weights of the scattering coefficients according to their selection index; we see a rapid decay, again indicating a sparse model. We also compute statistics of the selected features to get a broader picture of the process. The two plots on the left of figure 6 plot the usage of scattering coefficients in the model (9) by channel  $\delta$  and power  $q$ , respectively, according to weights  $w$ . From these we see that coefficients derived from the Li-only channel were used most by the regression, and similarly for the power  $q = 2$ . Interestingly, the power  $q = 4/3$  was used the least; these terms correspond to the exchange energy, which constitutes a small fraction of the total energy. Additionally, the valence channel is favored over the core channel, matching the larger role of valence electrons in bond formation.

The rightmost plot of figure 6 shows similar information for scales. The empty first space represents the bias term  $w_0$ . The blue portions represent first order features and all other colors are second order scattering coefficients. As indicated by the orange, green, and red bands in the second, third, and fourth columns, the model prefers to couple wavelets of similar scales. Figure 5 plots the sum of the weights corresponding to the wavelets, where the diagonal represents first order wavelet coefficients. We can get an idea of which wavelets are used most frequently, but no clear pattern emerges.

## 6 Conclusion

We have presented a general family of wavelet scattering coefficients as features for use in machine learning tasks on three dimensional signals. The features are specially suited for modeling properties of complex multiscale materials; they have desirable invariant properties, including invariance to rigid translations, rotations, and reflections. Additional invariants, such as to system size, are recovered through normalization of the scattering coefficients. These features may be tailored for the application at hand through the selection of the radial component of the wavelets (1). General considerations for selecting the radial components and an example selection were presented in example 1. These wavelets give state-of-the-art results for modeling formation energies of amorphous Li-Si systems.

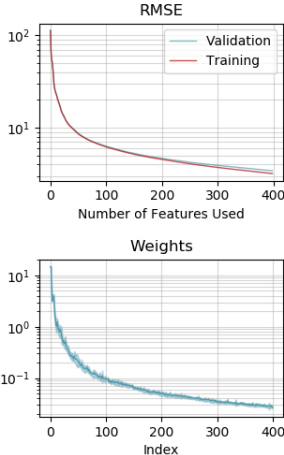


Figure 4: Top: Regression RMSE vs. model size. Bottom: weights vs. number of training points.

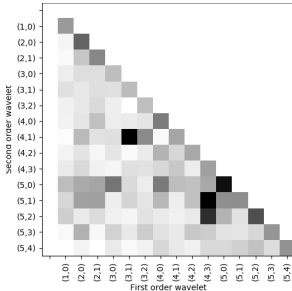


Figure 5: Sum of model weights. Darker squares indicate larger weights.



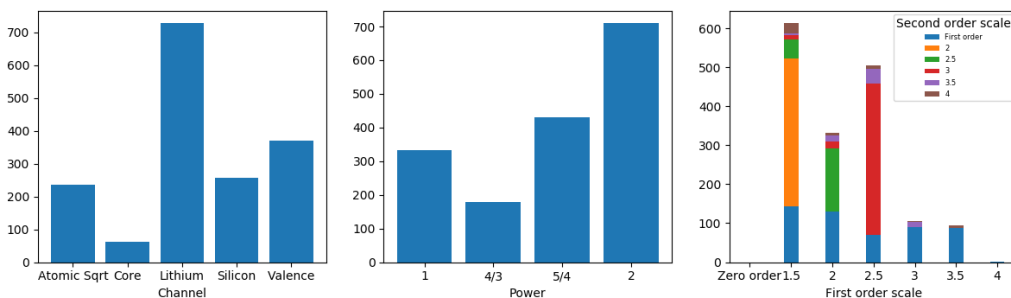


Figure 6: Statistics of selected wavelet coefficients. From left to right: histogram by channel  $\delta$ ; histogram by power  $q$ ; histogram by scales  $(j_1, j_2)$ .

The numerics show that an interpretable (sparse) model is recovered which appears to be improvable with the inclusion of additional features. The computed model performed admirably when compared with two other methods trained on similar Li-Si databases. Future work will include investigations into alternative choices of radial functions for the proposed wavelets as well as the effectiveness of these features in modeling other material properties and signals.

### Acknowledgments

We would like to acknowledge Mr. Jialin Liu’s contribution to automate the training set generation. X.B., P.S., K.J.K. and M.H. were (partially) supported by DARPA YFA #D16AP00117. M.H. is also supported by Alfred P. Sloan Fellowship #FG-2016-6607 and NSF grant #1620216. Y.Q. also appreciates the financial support from Michigan State Research Foundation (SPG grant).

### References

- [1] Yihan Shao et al. Advances in methods and algorithms in a modern quantum chemistry program package. *Phys. Chem. Chem. Phys.*, 8:3172–3191, 2006.
- [2] Rafael Gómez-Bombarelli et al. Automatic chemical design using a data-driven continuous representation of molecules. *ACS Cent. Sci.*, 4(2):268–276, 2018.
- [3] Sandip De, Albert P. Bartok, Gabor Csanyi, and Michele Ceriotti. Comparing molecules and solids across structural and alchemical space. *Phys. Chem. Chem. Phys.*, 18(20):13754–13769, 2016.
- [4] Kristof T. Schütt, Pieter-Jan Kindermans, Huziel E. Sauceda, Stefan Chmiela, Alexandre Tkatchenko, and Klaus-Robert Müller. Schnet: A continuous-filter convolutional neural network for modeling quantum interactions. In *NIPS 2017*, pages 991–1001, 2017.
- [5] Justin Gilmer, Samuel S. Schoenholz, Patrick F. Riley, Oriol Vinyals, and George E. Dahl. Neural message passing for quantum chemistry. In *ICML 2017*, 2017.
- [6] Grégoire Montavon, Matthias Rupp, Vivekanand Gobre, Alvaro Vazquez-Mayagoitia, Katja Hansen, Alexandre Tkatchenko, Klaus-Robert Müller, and O. Anatole von Lilienfeld. Machine learning of molecular electronic properties in chemical compound space. *New J. Phys.*, 15:095003 (16pp), 2013.
- [7] Felix Brockherde, Leslie Vogt, Li Li, Mark E Tuckerman, Kieron Burke, and Klaus-Robert Müller. By-passing the Kohn-Sham equations with machine learning. *Nat. Commun.*, 8, 2017.
- [8] Zhenwei Li, James R. Kermode, and Alessandro De Vita. Molecular dynamics with on-the-fly machine learning of quantum-mechanical forces. *Phys. Rev. Lett.*, 114:096405, Mar 2015.
- [9] Timothy L Fletcher, Stuart J Davie, and Paul LA Popelier. Prediction of intramolecular polarization of aromatic amino acids using kriging machine learning. *J. Chem. Theory Comput.*, 10(9):3708–3719, 2014.
- [10] Matthias Rupp, Alexandre Tkatchenko, Klaus-Robert Müller, and O Anatole Von Lilienfeld. Fast and accurate modeling of molecular atomization energies with machine learning. *Phys. Rev. Lett.*, 108(5):058301, 2012.
- [11] Stéphane Mallat. Group invariant scattering. *Comm. Pure Appl. Math.*, 65(10):1331–1398, October 2012.

- [12] William T. Freeman and Edward H. Adelson. The design and use of steerable filters. *IEEE Trans. Pattern Anal.*, 13(9):891–906, 1991.
- [13] Michael Eickenberg, Georgios Exarchakis, Matthew Hirn, and Stéphane Mallat. Solid harmonic wavelet scattering: Predicting quantum molecular energy from invariant descriptors of 3D electronic densities. In *NIPS 2017*, pages 6540–6549, 2017.
- [14] Kwang Jin Kim and Yue Qi. Vacancies in Si can improve the concentration-dependent lithiation rate: Molecular dynamics studies of lithiation dynamics of Si electrodes. *J. Phys. Chem. C*, 119(43):24265–24275, 2015.
- [15] Kwang Jin Kim, James Wortman, Sung-Yup Kim, and Yue Qi. Atomistic simulation derived insight on the irreversible structural changes of Si electrode during fast and slow delithiation. *Nano Lett.*, 17(7):4330–4338, 2017. PMID: 28632390.
- [16] Nongnuch Artrith, Alexander Urban, and Gerbrand Ceder. Constructing first-principles phase diagrams of amorphous  $\text{Li}_x\text{Si}$  using machine-learning-assisted sampling with an evolutionary algorithm. *J. Chem. Phys.*, 148:241711, 2018.
- [17] Berk Onat, Ekin D. Cubuk, Brad D. Malone, and Efthimios Kaxiras. Implanted neural network potentials: Application to Li-Si alloys. *Phy. Rev. B*, 97:094106, 2018.
- [18] Atsuto Seko, Atsushi Togo, and Isao Tanaka. Descriptors for machine learning of materials data. In *Nanoinformatics*, pages 3–23. Springer, 2018.
- [19] Stefan Chmiela, Alexandre Tkatchenko, Huziel E. Sauceda, Igor Poltavsky, Kristof T. Schütt, and Klaus-Robert Müller. Machine learning of accurate energy-conserving molecular force fields. *Sci. Adv.*, 3(5):e1603015, 2017.
- [20] Jörg Behler. Perspective: Machine learning potentials for atomistic simulations. *J. Chem. Phys.*, 145(17):170901, 2016.
- [21] J. S. Smith, O. Isayev, and A. E. Roitberg. ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost. *Chem. Sci.*, 8:3192–3203, 2017.
- [22] Katja Hansen et al. Machine learning predictions of molecular properties: Accurate many-body potentials and nonlocality in chemical space. *J. Phys. Chem. Lett.*, 6:2326–2331, 2015.
- [23] Albert P. Bartók, Mike C. Payne, Risi Kondor, and Gábor Csányi. Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons. *Phys. Rev. Lett.*, 104(13):136403(4), 2010.
- [24] Jörg Behler and Michele Parrinello. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.*, 98(14):146401(4), 2007.
- [25] Michael Eickenberg, Georgios Exarchakis, Matthew Hirn, Stéphane Mallat, and Louis Thiry. Solid harmonic wavelet scattering for predictions of molecule properties. *J. Chem. Phys.*, 148:241732, 2018.
- [26] Jiyun Kuang and C D Lin. Molecular integrals over spherical gaussian-type orbitals: I. *J. Phys. B*, 30(11):2529, 1997.
- [27] Taco Cohen and Max Welling. Group equivariant convolutional networks. In *ICML 2016*, volume 48, pages 2990–2999, 2016.
- [28] Taco Cohen and Max Welling. Steerable CNNs. arXiv:1612.08498, 2016.
- [29] Risi Kondor, Hy Truong Son, Horace Pan, Brandon Anderson, and Shubhendu Trivedi. Covariant compositional networks for learning graphs. arXiv:1801.02144, 2018.
- [30] Risi Kondor and Shubhendu Trivedi. On the generalization of equivariance and convolution in neural networks to the action of compact groups. arXiv:1802.03690, 2018.
- [31] Laurent Sifre and Stéphane Mallat. Rotation, scaling and deformation invariant scattering for texture discrimination. In *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, June 2013.
- [32] Matthew Hirn, Stéphane Mallat, and Nicolas Poilvert. Wavelet scattering regression of quantum chemical energies. *Multiscale Model. Sim.*, 15(2):827–863, 2017.
- [33] Eric Cancès, Mireille Defranceschi, Werner Kutzelnigg, Claude Le Bris, and Yvon Maday. Computational quantum chemistry: A primer. *Handb. Numer. Anal.*, 10:3–270, 2003.

- [34] Joan Bruna, Stéphane Mallat, Emmanuel Bacry, and Jean-Francois Muzy. Intermittent process analysis with scattering moments. *Ann. Stat.*, 43(1):323 – 351, 2015.
- [35] Kwang Jin Kim. *Reactive Force Field Based Atomistic Simulations of Silicon Anode Upon Lithiation and Delithiation in Lithium-Ion Batteries*. PhD thesis, Michigan State University, 2018.