Atomic-scale analysis of disorder by similarity learning from tunneling spectroscopy

Petro Maksymovych,[1,*] Jiaqiang Yan,[2] Brian Sales,[2] and Jun Wang[1]
[1]*Center for Nanophase Materials Sciences, Oak Ridge National Laboratory (ORNL), Oak Ridge, Tennessee 37831, USA*
[2]*Materials Science and Technology Division, ORNL, Oak Ridge, Tennessee 37831, USA*

(Received 25 June 2021; revised 3 April 2022; accepted 11 May 2022; published 19 July 2022)

Rapid proliferation of hyperspectral imaging in scanning probe microscopy creates unique opportunities to systematically capture and categorize higher dimensional datasets, toward insights into electronic, mechanical, and chemical properties of materials with nano- and atomic-scale resolution. Effective hyperspectral imaging requires a consistent framework for data analysis that would be broadly applicable, reproducible, and transferrable, conceptually resembling the success of integral transforms in image analysis. Here, we demonstrate application of similarity learning for resolving the structure of tunneling spectroscopy data, characterizing a superconducting material with sparse density of defects. Popular methods for unsupervised learning and discrete representation of the data in terms of clusters of characteristic behaviors were found to produce inconsistencies with respect to capturing the location and tunneling characteristics of defect sites. The underlying reason for their ambiguity was traced to continuous variation of the electronic properties across the surface and therefore the absence of clear structural boundaries in the low-dimensional latent spaces of the data. We supported this hypothesis by direct analysis of the distributions of Euclidean distances within the dataset. We further proposed distance rescaling with probabilistic description as a possible approach to mitigate the detrimental effect of the long tails of the distributions on the performance of clustering methods. Subsequently, we applied a more general, nonlinear similarity learning, where dimension reduction was explicitly trained to amplify similarities and dissimilarities among individual spectra. This approach was found to outperform several widely used methods for dimensionality reduction and produce a clear categorization of tunneling spectra. Significant spectral weight transfer associated with the electronic reconstruction by the vacancy sites was systematically captured, as was the spatial extent of the vacancy region. Given that a great variety of electronic materials will exhibit similarly smooth variation of spectral response due to random or engineered inhomogeneities, we believe our approach will be useful for systematic analysis of hyperspectral imaging with minimal prior knowledge as well as prospective comparison of experimental measurements with theoretical calculations with explicit consideration of disorder.

DOI: 10.1103/PhysRevResearch.4.033058

I. INTRODUCTION

Electrical and mechanical measurements on the nanoscale underpin most widely used scanning probe microscopy (SPM) methods as well as many future electronic devices. An outstanding challenge in such studies is a systematic and quantitative interpretation of observations, often in the absence of well-defined analytical models. On the one hand, localizing the probed volume to near-atomic scale almost inevitably introduces systematic and random effects of electronic, mechanical, and chemical properties of the contacts, resulting in persistent and often time-varying uncertainty in the transfer function of the measurement. On the other hand, exponential amplification of physical properties in the measured observables is often the basis for the enhanced sensitivity and spatial resolution achieved by imaging methods [1–3]. For example, the exponential distance dependence of tunneling probability enabled scanning tunneling microscopy (STM) [4], which subsequently revealed a broad variety of atomic structures and dynamics with atomic resolution [5–8]. However, the joint effects of uncertain and nonlinear transfer functions [9,10] generally complicate quantitative interpretation of data in SPM and related nanoscale device measurements. Moreover, consistent theoretical calculations that account for the exact shape of the measurement probe, the nonequilibrium properties of electron tunneling, and the electronically excited states require extensive numerical simulation [9,10] that may be computationally too prohibitive to complement extensive experimentation. Therefore, a task of increasing importance is the analysis of hyperspectral data with minimum prior information [11]. The hyperspectral data volume can yield rich information even in the absence of a specific model, for example, by identifying variability in the dataset or specific regions of interest.

In the past several years, approaches utilizing dimensionality reduction and classification by clustering were

*maksymovychp@ornl.gov

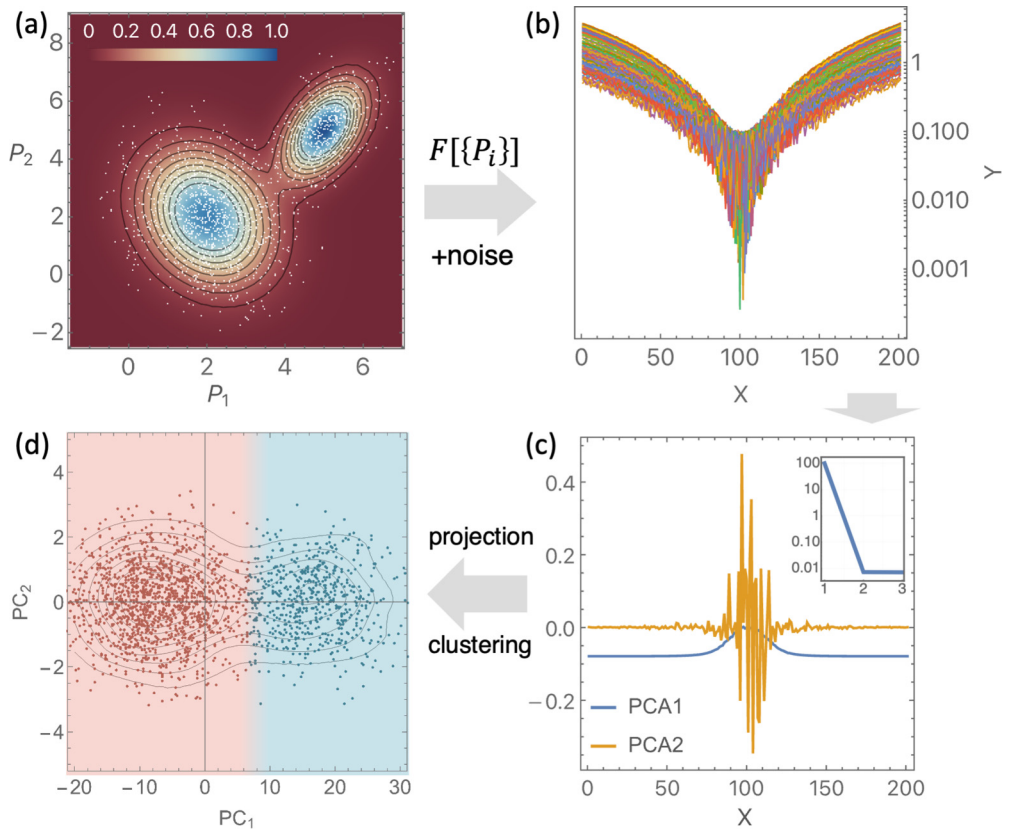

FIG. 1. The conceptual goal of hyperspectral analysis of disorder is to infer the existence and structure of low-dimensional latent space that effectively approximates the true parameter space. (a) Synthetic two-dimensional parameter space, represented by multinormal distribution with two well-separated centroids. (b) A family of polynomials parametrized by randomly sampled parameters from the synthetic space (a) [white dots in (a)]. The polynomial function $Y = 0.1kX^2 + C$ was used, where k was sampled from the distribution in (a), while C was sampled from a uniform distribution between 0 and 0.1. (c) First two principal components of the synthetic dataset (b) and the corresponding scree plot (inset), which show that projection on the first two principal components already captures >99% of variability in the dataset. (d) Latent space of hyperspectral data projected onto first and second principal components (c), revealing structure very closely resembling the original parameter space. Two clusters (red and blue) are obtained by k-means clustering.

successfully applied to hyperspectral data spatially resolved tunneling spectroscopy [12,13], force-distance spectroscopy [14,15], and switching spectroscopy [16,17]. The essential assumption of these approaches is the existence of a low-dimensional parameter space from which measured spectra arise through a convolution with the instrument transfer function. The analysis then aims to construct a low-dimensional representation of the hyperspectral data, for example, in the form of several clusters of characteristic behaviors. Figure 1 applies this concept to synthetic data, wherein random samples $\{P_i\}$ from the a chosen two-dimensional (2D) parameter space [Fig. 1(a)] are used to construct a series of polynomial curves [Fig. 1(b)], representing the effect of the instrument function $F[\{P_i\}]$ and added noise. Subsequently, principal component analysis (PCA) projects the family of curves (standardized to have zero mean and unit sample variance) onto the first two principal components (the eigenvectors of the covariance matrix sorted in order of decreasing eigenvalues). The first two components already capture >99% of variability in the synthetic dataset [Fig. 1(c)]. The latent space of projections onto principal components [Fig. 1(d)] can then be considered an effective representation of the data. Indeed, Fig. 1(d) appears qualitatively very similar to Fig. 1(a) and can become nearly identical by affine transformation that rotates and scales the parameter space. Applying clustering is also very effective in this case, drawing clear decision boundaries that partition the latent space into two parts, reminiscent of the two centers of the original distribution. Therefore, this analysis will work well if the true parameter space and therefore the derived latent space are both well structured. However, disorder in electronic materials can manifest in a broad range of behaviors, many of which will be perturbative rather than drastically changing [18–20]. Moreover, in the case of a sparse density of impurities, the number of data points originating from defect sites may be limited (it appears as an outlier), which may also confuse popular methods, such as k-means clustering [21]. The validity of classification approaches can also be questioned, wherein partitioning the dataset into an integer number of clusters inevitably introduces ambiguity near decision boundaries. As a result, good parameter separation in the latent space may be more of an exception for atomically resolved microscopy data, and alternatives, such as probabilistic description or nonlinear methods that can separate the latent space, are needed for the likely more common case of smooth variation of properties across the dataset.

Here, we demonstrate similarity learning for the hyperspectral data comprising spectra of energy-resolved density of states of the unconventional superconductor FeSe, acquired over a spatial grid with near-atomic-scale resolution. We identified the existence and electronic signature of Fe vacancy defects as well as their spatial locations with minimal prior knowledge. Despite the apparent simplicity of the problem with just one type of chemical vacancy, we show that conventional clustering methods may produce ambiguous if not conflicting results. We traced the origins of the ambiguity to long-tailed structure of the distribution of Euclidian pairwise distances in the dataset. This conclusion was reinforced by the efficiency of *ad hoc* rescaling of pairwise distances to identify the vacancy sites. This analysis can be effective for low-defect scenarios, even though it obscures well-defined geometric meaning of the latent space. To generalize, we applied a weakly supervised similarity learning that explicitly learns nonlinear pairwise distance functions [23] and enables effective analysis of tunneling datasets, even when compared with many popular techniques for nonlinear dimensionality reduction. We believe that similarity learning [24] can play an important role in future analysis of a broad spectrum of problems in materials physics, such as detecting and analyzing effects of disorder as well as classical and quantum phase transitions.

II. RESULTS AND DISCUSSION

FeSe is an unconventional superconductor with a transition temperature of ∼8 K [25]. For this paper, FeSe single crystals were grown out of KCl-$AlCl_3$ flux by a liquid transport technique [26]. The surfaces were cleaved in ultrahigh vacuum at a temperature of ∼150 K before STM measurements. Owing to lack of dangling bonds, FeSe surfaces contain only a small number of well-defined defects [27–29]. Fe vacancies, for example, are readily detectable due their low surface density and characteristic dumbbell shape [Fig. 2(a)]. However, despite their chemical simplicity, the vacancies in FeSe can have a profound impact on the electronic properties. For example, Fe vacancies locally modify magnetic ordering [30]. Se vacancies have been predicted to generate effective hole doping through a combined effect of electron scattering, lattice strains, and charge doping [27,30,31], contrary to intuitive expectation of electron doping. The FeSe monolayer system can achieve ∼10 times higher T_c due to the combined effect of doping and strain at the interface [32]. Meanwhile, the interest in the FeSe/Te system has been vigorously renewed in lieu of predictions for the topologically protected superconducting state in proximity of the 50/50 Se/Te ratio [33], where intrinsic disorder of solid solutions also emerges [34].

To characterize spectral signatures of the Fe vacancies, we acquired tunneling spectra on a 100 × 100 point grid in a nearly pristine surface area with a few vacancies [Fig. 2(a)]. The energy-resolved histogram of all the spectra clearly reveals the superconducting gap [Fig. 2(b)] as well as a rather broad distribution of density of states (about fourfold variation) because of the vacancies, electron scattering (resulting in Friedel oscillations), noise, and other possible factors. Fe vacancies are already known to develop in-gap impurity states and therefore increase the observed density of states inside the

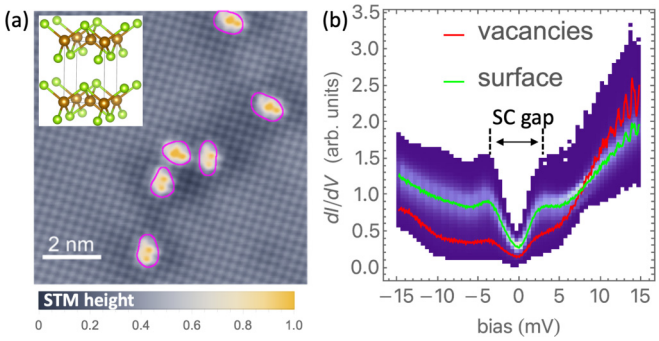

FIG. 2. Tunneling microscopy and spectroscopy of as-cleaved FeSe surface, with a small number of Fe vacancies. (a) Constant-current scanning tunneling microscopy (STM) image showing the atomic lattice of the top surface layer and Fe vacancy sites which appear as clear double-protrusions (this is primarily an electronic effect). Inset shows crystallographic unit cell of tetragonal FeSe after Ref. [22]. (b) Tunneling spectroscopy near the Fermi level, revealing the ∼4 meV wide superconducting gap and a broad distribution of the tunneling conductance due to combined effects of measurement noise and vacancy sites. Individual spectra were acquired on a 100 × 100 pixel spatial grid, with 512 sampling points per energy interval from −15 to +15 mV. Purple in (b) is an energy-resolved histogram of all the spectra. Green and red mark the average spectra from the defect-free and defect regions of the surface, correspondingly. The vacancy sites were identified by segmentation analysis of an isoenergy slice of the hyperspectral volume at ∼−1 mV, and their approximate outlines are shown in purple in (a).

superconducting gap [35]. This effect enables partitioning of the spectra in the dataset into two groups—with and without vacancies—for example, by segmentation of the isoenergy slice of the hyperspectral cube at an energy of ∼−1 mV. The corresponding outlines of the vacancy sites obtained by this segmentation are shown in Fig. 2(a) in purple. Quite notably, Fe vacancies significantly reconstruct the whole spectrum, not just the proximity of the superconducting gap [compare the red and green curves in Fig. 2(b)].

Although detecting vacancies in FeSe by visual inspection is remarkably easy, the multivariate workflow presented in Fig. 1 fails to do so reliably in the reduced dimensionality representation. This is the focus of this paper. Almost all methods presented below provide some level of separation in real space. However, in a very general case, e.g., a sparsely sampled dataset, disordered samples, or multiple kinds of defects, real-space information may not be readily available. We would like to find reliable approaches to separate characteristic behaviors in the spectral space alone. Eventually, both real-space and spectral-space information should be utilized to maximize efficiency of analysis.

Figure 3(a) shows the projection of the hyperspectral dataset onto its first two principal components (that account for > 90% of the observed variance in the data, with three components accounting for >99%). Locating the projections of the spectra due to vacancies (red) and pristine (green) surface sites shows that the two are separated in the latent space, albeit without a clear break into distinct groups. In other words, the dataset is not linearly separable.

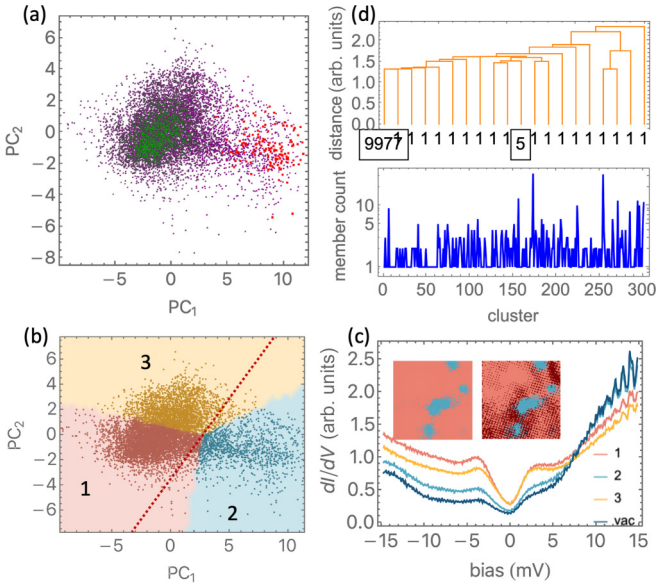

FIG. 3. Dimensionality reduction and clustering applied to tunneling spectroscopy dataset in Fig. 2. (a) Principal component analysis (PCA): projection of the individual tunneling spectra into the latent space of two principal components (PC_1 and PC_2) whose linear combination accounts for >90% of variability in the dataset. Green and red highlight projections due to surface and vacancy regions (determined by segmentation in Fig. 2), respectively. Purple are all the projections. (b) Partitioning of the PCA latent space by k-means clustering with 2 and 3 clusters. The decision boundary for two clusters is shown by the red dashed line. (c) Cluster centers (average IV curves within each cluster) for 3-means analysis in (b). Insets show spatial maps of clusters (left = 2, right = 3). The averaged IV curve for the vacancy region (*vac* in the legend) from Fig. 2 is shown for reference. (d) Results of agglomerative (hierarchical) clustering of the hyperspectral dataset, shown as a dendrogram. The y axis shows linkage distance between individual clusters, while the numbers below show the number of individual IV curves in each shown branch of the dendrogram. Bottom panel shows the number of samples vs the hierarchical order of the specific cluster.

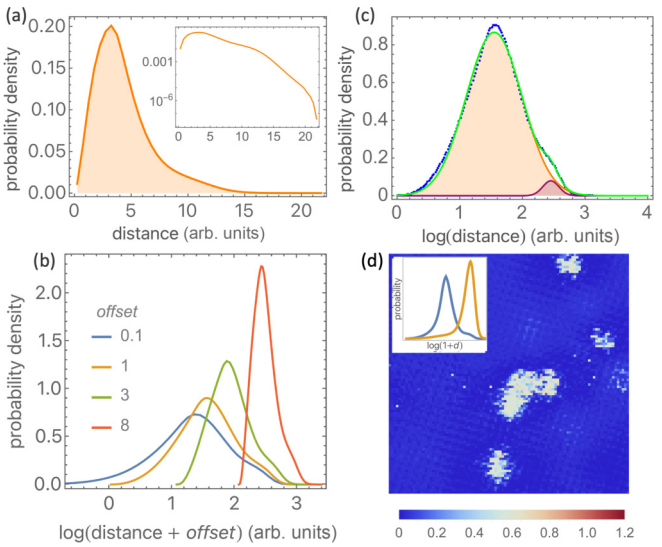

FIG. 4. Statistics of pairwise Euclidean distances between single tunneling spectra in the FeSe dataset. (a) Distribution of Euclidean distances within the dataset from Fig. 2(b) exhibiting a long tail (inset shows distribution on the log scale). (b) Distributions of rescaled pairwise distances, with rescaling as $\log(d_{ij} + \text{offset})$. (c) Fitting rescaled distribution of the log-Euclidean distance by two normal distributions. (d) The abundance map of minority spectra in the dataset, obtained as value for the distance histogram of each spectrum at rescaled distance value of $\log(d_{ij} + 1) = 2.5$. The inset clearly shows typical probability distributions for a majority (blue) and minority (orange) spectrum, with minority exhibiting a peak at ∼2.5.

The assertion is best reflected in the confusing results of clustering methods applied to the latent space: k-means points to few clusters [Fig. 3(b)], while agglomerative clustering points to the opposite—dozens if not more distinct behaviors [Fig. 3(d)]. k-means—a top-down algorithm—identifies k (where k is given) centroids and iteratively refines their position until the net variances of the point-cluster distances are minimized [21,36]. Agglomerative (hierarchical) clustering builds a bottom-up cluster hierarchy [37], beginning with n clusters (where n is the number of spectra in the dataset) and iteratively merging distinct clusters with the same linkage distance. At some intermediate value of linkage distance, a small number of clusters may be revealed. In our case, k-means is effective at approximate identification of the vacancy sites for two-cluster partitioning, which is to be expected [see mask in the inset of Fig. 4(c)]. Adding a third cluster into partitioning, however, does not refine the location of the vacancy, instead partitioning the latent space region of the surface sites and assigning a few spurious locations to the vacancies [Fig. 3(c) inset]. Meanwhile, hierarchical clustering partitions the spectra into essentially arbitrarily many clusters, with dozens and even hundreds of clusters having approximately equal weighting. This is seen in Fig. 3(d), where at no point up to even 300 clusters does the number of members in each cluster exceed 10 spectra [and most often just one, Fig. 3(d) bottom panel]. In an effective agglomerative clustering scenario, we would expect to have much more significant weighting for clusters already at the top of the hierarchy (e.g., for three- to four-cluster partitioning). In other words, according to agglomerative clustering, either there are hundreds of observable behaviors—which is unlikely given the origin of our data—or there are no significant clusters, i.e., vacancies and pristine surface are not statistically distinct, which contradicts direct observations.

Given the extensive amount of literature on clustering methods, such ambiguities are not uncommon [36]. It is rather clear that the Euclidian distance (ED) between individual spectra or between their projections in the latent space plays a key role in the subsequent performance of the clustering methods. We directly examined the distribution of EDs across the whole hyperspectral dataset [Fig. 4(a)]. It is immediately apparent that the distribution is long tailed, which at least partially is caused by a relatively small number of vacancy sites. The sparsity of the vacancies in the dataset creates an immediate problem for the k-means clustering method, which is most efficient when the number of samples from distinct behaviors is comparable in the dataset [21,36]. The long tail also confuses agglomerative clustering, essentially due too

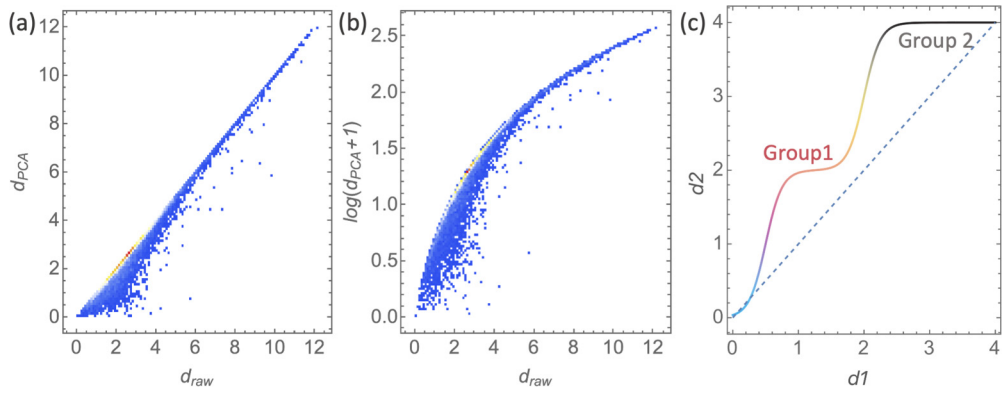

FIG. 5. Distance correlation plot that compares distances between individual points within the dataset (or distance from the mean) among two latent spaces. (a) Comparison of the original data space before dimensionality reduction with the distance distribution in the latent space defined by first two principal components [Fig. 3(a)]. (b) Similar comparison after logarithmic rescaling of the latent space. (c) Schematic of distance correlation plot for an idealized scenario of nonlinear dimensionality reduction, where two (or more) groups of spectra would be clearly separated from each other by the appropriate distortion of the latent space.

slow of a pace of agglomeration, so that inferred clusters end up containing very few spectra up to a very high level in the hierarchy [Fig. 3(d)].

Our reasoning can be further supported by an *ad hoc* rescaling of the Euclidean distances between spectra. Figure 4(b) plots the distributions of $\log(d_{ij}+n)$ with n from 0.1 to 8 (where d_{ij} is the Euclidean distance between a pair of spectra). In these cases, logarithmic rescaling symmetrizes the distance distribution by contracting the long tail of large relative distances while expanding smaller values. For n between 1 and 5, the distribution becomes not only much more symmetric, but it also clearly reveals a shoulder on the side of the main peak [Fig. 4(b)]. A weak shoulder is also visible in Fig. 4(a). In this form, the modified distance distribution can be readily fit by two Gaussian distributions [Fig. 4(c)]. The bimodality of the distribution then points to existence of at least two distinct groups of spectra, which are similar within the group [therefore yielding small relative Euclidean distance captured by the blue peak in Fig. 4(c)] and are dissimilar across the groups [with larger relative distance under the purple peak in Fig. 4(c)]. Extending this logic, we can now take advantage of the sparsity of the vacancies on the surface to label the individual spectra as belonging to majority vs minority species. To this end, we first calculate distributions of log-Euclidean distance for each spectrum and then effectively label each spectrum with its distance histogram. For majority (surface sites), the histograms are dominated by the first peak, while for vacancies, the opposite is true [Fig. 4(d), inset]. Visual inspection of the abundance maps of individual histograms in real space for the minority species in Fig. 4(d) indeed confirms our notions as well as the assignment of the assumed bimodal distribution in Fig. 4(c).

The rescaling described above, which amounts to nonlinear geometric transformation of the data space, is effective, but it is also arbitrary and is likely not generalizable to more complex cases. We therefore sought nonlinear dimensionality reduction methodologies that can achieve nonlinear and effective separation between tunneling spectra algorithmically. Nonlinear dimensionality reduction encompasses a large range of algorithms, such as kernel PCA (kPCA) [38], t-distributed stochastic neighbor embedding [39], neural network autoencoders, among others, each with its own approach to construction of the latent space.

To aid in evaluation of the effectiveness of a specific technique, we tracked the relevant geometry of the projected space by plotting the Euclidean distance of each spectrum from the average in the dataset spectrum vs a similar metric in the latent space (Euclidean distances of coordinates of these spectra in a given latent space from the coordinates of the average spectrum). Examples of such a distance correlation plot, in the form of the density histogram, are shown in Fig. 5(a) for PCA and our logarithmic rescaling of the distances discussed above in Fig. 5(b). PCA is a linear dimensionality reduction technique, and this is very well reflected by the distance correlation plot with a clear linear scaling of distances between original and reduced dimensionality representations [Fig. 5(a)]. Deviation from linearity in this case is the effect of partial loss of information due to dimensionality reduction. Meanwhile, the logarithmic rescaling appears as a log function, by design [Fig. 5(b)]. The transition from the long-tailed to the more symmetric distribution of the relative distances is also apparent by following the redistribution of the point density upon log rescaling. We can further hypothesize what the desired geometric distortion may look like, as shown in Fig. 5(c). Essentially, parts of the original dataset will be grouped into small and well-separated regions of the projected space, creating a large local density of projections. This specific shape is valid under the assumption that one group [group 2 in Fig. 5(c)] has a lot fewer data points than the other (e.g., group 1), which is typically the case for low density of vacancies and defects.

Figure 6 shows the projections and the distance correlation plots for the FeSe dataset generated by several popular methods for nonlinear dimensionality reduction: kPCA [38,40] [Fig. 6(a)], nonnegative matrix factorization (NMF) [41], uniform manifold approximation and projection (UMAP) [42] [Fig. 6(c)], and deep autoencoder [Fig. 6(c)]. All the methods introduce various nonlinearities as seen from the correlation plots: kPCA squeezes the latent space as a whole [Fig. 6(a)], somewhat like the log rescaling in Fig. 5. NMF also squeezes

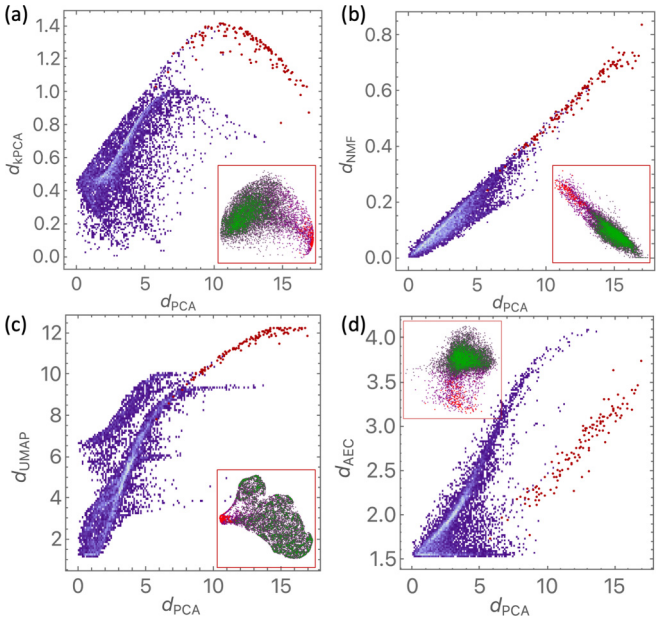

FIG. 6. Distance correlation plots for nonlinear dimensionality reduction methods applied to the spectroscopy dataset: (a) kernel principal component analysis (kPCA), (b) nonnegative matrix factorization (NMF), with implementation in SCIKIT-LEARN [42], (c) uniform manifold approximation and projection (UMAP), and (d) deep autoencoder. The autoencoder was constructed as a sequence of three fully connected layers, with dimensions of 100, 2, and 100. The center bottleneck layer thus represents compressed and generative representation of the dataset. Blue and red accents in the main figure highlight the positions of the surface and vacancy sites correspondingly. The insets show two-dimensional (2D) latent spaces obtained by each method. The following parameters for the nonlinear dimensionality reduction algorithms, implemented in SCIKIT-LEARN library, have been used: kPCA - "rbf" kernel with Γ of 0.04; NMF–2 components with maximum of 1000 iterations; UMAP-2 components with 1000 nearest neighbors. The autoencoder was trained with the ADAM optimizer, batch size of 64, 5376 training examples and 640 validation examples, L2 regularization set to 0.002, and 100 training rounds. All other parameters have been set to their default implementation.

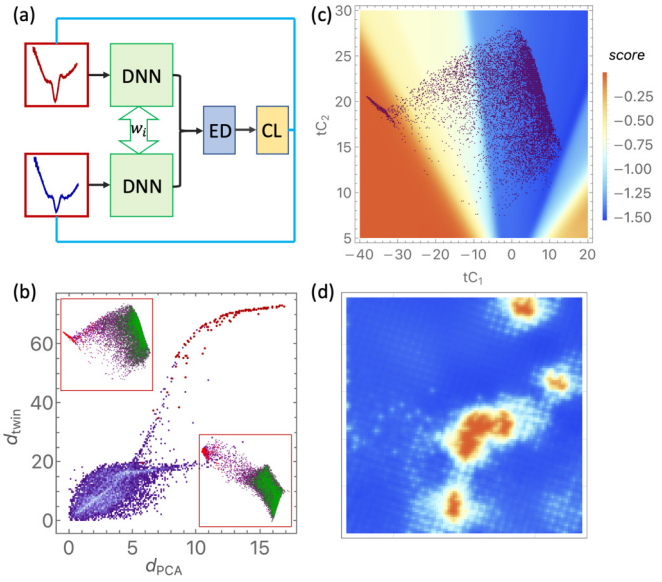

FIG. 7. Application of similarity learning by twin (Siamese) network architecture for the tunneling spectroscopy dataset. (a) Schematic structure of the basic twin architecture (DNN = deep neural network, w_i = shared weights, ED = Euclidean distance, CL = contrastive loss layer). (b) Distance correlation plot for a typical projection of the trained Siamese network applied to the dataset (insets show two different projections, obtained by slight variation of the parameters of the training set, such as the threshold Euclidean distance beyond which the spectra are distinct). (c) The result of applying two Siamese networks to the dataset in sequence. First network to project the dataset onto two components (tC1 and tC2), the projection is the same as the distance correlation plot in Fig. 7(b). Second, simplified network to create one-dimensional (1D) scores for each of the projections, based on their proximity to each other in the projected space. The field of scores is superimposed on top of the projection as a colormap, while (d) shows the abundance map of the scores in real space. Very clearly, the scores act as unique labels for individual spectra, clearly separating vacancy, surface, and intermediate sites. The image has been slightly smoothed by Gaussian convolution for presentation purposes.

the latent space but otherwise does not introduce significant distortion. UMAP partitions the surface spectra but does not clearly separate out the vacancies [Fig. 6(c)]. In none of these cases is there a clear separation between surface (blue) and vacancy (maroon) sites [Figs. 6(a)–6(c)]. Finally, the deep autoencoder does reconstruct the latent space, splitting it into ∼3 regions, two of which [both surface sites, Fig. 6(d)] have comparable density. Thus, although nonlinearities introduced by these methods have not yielded the sought-after result of clearly partitioning the data space, as is also evident from direct inspection of the latent spaces for all the methods (see insets). We note that each of these methods has a set of hyperparameters and a pronounced sensitivity to specific choices of these parameters. However, we are seeking a more general solution. Given the relatively unsatisfactory performance of the unsupervised nonlinear dimensionality reduction techniques, we next applied algorithms where some prior information about the properties of the spectra is passed to the algorithm.

One such architecture is a twin (Siamese) neural network architecture, which was proposed by Koch [43] to achieve better classification of image datasets. The goal of the network is to learn nonlinear embedding of the original data that would maximize Euclidean distance (ED) between distinct spectra and minimize it between similar spectra in the projected space. The schematic of such a network is shown in Fig. 7(a). The network is being trained by propagating a pair of spectra (in our case, selected directly from the raw data in Fig. 2) through a pair of identical networks. The network pair has both identical topology and shared trainable weights. We chose a deep network consisting of three fully connected layers with dimensionality of $20 \rightarrow 5 \rightarrow 2$ separated by nonlinear activation functions. The network therefore maps the spectra into 2D projected space. Subsequently, ED is calculated between the outputs of the two networks [Fig. 7(a)]. The key to twin network performance is a contrastive loss [CL in Fig. 7(a)], which computes a loss based on the distance across the twin network and a target that specifies whether the distance should be minimized or maximized. One added

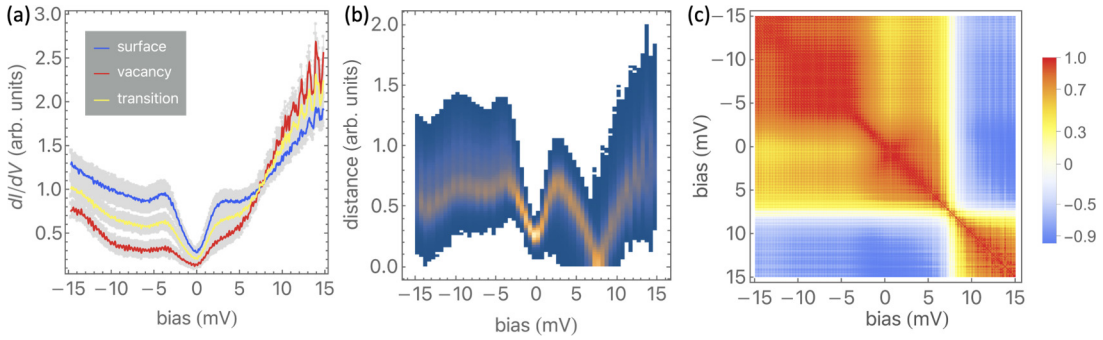

FIG. 8. Spectral weight transfer due to Fe vacancies in FeSe as seen by tunneling spectroscopy. (a) Average spectra due to surface, vacancy, and transition regions (red, blue, and yellow) identified by the appropriate scores in Fig. 7(d). (b) Energy-resolved Euclidean distances between the collection of surface and vacancy spectra, plotted as a density histogram (yellow corresponds to high abundance). (c) Correlation plot for all spectra in the dataset, revealing two distinct anticorrelated energy windows for spectral changes (from −15 to −5 mV and from 7 to 15 mV), as well as window of intermediate correlation (between ∼0 and 7 mV). This plot statistically confirms that Fe vacancies reduce the density of filled states and increase the density of empty states compared with pristine surface.

advantage of this architecture is that the number of pairs scales as N^2 with the N size of the dataset, so that twin networks work well with smaller-sized datasets characteristic of tunneling spectroscopy. In our case, the networks were trained by randomly selecting 5000 pairs from a total possible pool of $\frac{1}{2} \times 10^8$ spectra. Given the relatively small number of vacancy sites in our dataset, we ascertained that the random selection contains a small fraction of vacancy sites.

The input into the twin network is a binary choice of whether the two spectra are of the same overall group. In our case, we simply chose a particular threshold for the ED in the original hyperspectral dataset to partition the pairs within the dataset. As seen in Fig. 7(b), vacancies (red) become well separated from surface sites in the projected space of the twin network (green). This is also clearly seen in the distance correlation, which bears the closest similarity to the scenario in Fig. 5(c). The training is also sensitive to hyperparameters, although producing qualitatively similar results. Figure 7(b) shows two qualitatively similar projected spaces obtained in different initialization and training runs (with randomly generated training sets in each case).

The twin architecture then enables categorization of the projected and latent spaces, as shown in Fig. 7(c), wherein another twin network with smaller and even simpler (10 → 5 → 1) topology was applied to the projected space in the upper corner of Fig. 7(b). The network, trained on random samples from the point cloud in Fig. 7(c), outputs just one number, providing a continuously varying score for each spectrum based on its proximity to other spectra in the projected space. As seen in Fig. 7(c), twin networks very effectively categorize the latent space, drawing clear decision boundaries for the surface (blue) and vacancy (red) sites and rather remarkably outlining the transition region between the two (yellow). Because this type of classification avoids discretization of the clustering procedure, the real-space map of the scores both confirms its assignment and separates the vacancies, surface sites, and the transition region in both projected and real spaces [Fig. 7(d)].

The key prospective advantage of being able to identify distinct tunneling spectra as well as transitions between them is the analysis of the spectral weight transfer associated with disorder in any doped, and particularly highly correlated electronic materials [19,20,44]. Figure 8 shows the tunneling spectra of the vacancies, surface sites, and transition region averaged within their characteristic score values, as determined in Fig. 7(c). For added clarity, we also show the EDs between the spectra in Fig. 8(b) and a correlation plot for the whole dataset in Fig. 8(c). The spectral intensity is redistributed in the whole probed energy window ±15 meV, not just around the superconducting gap (±3 meV). The correlation plot [Fig. 8(c)] reveals that the changes from −15 to ∼+7 mV are anticorrelated with those from +7 to 15 mV, wherein suppression of the density of states in one region causes enhancement in the other. The redistribution of the spectral weight may signal significant electronic reconstruction associated with the vacancy site that goes beyond charge doping expected of impurities and intentional dopants. In fact, Se vacancies were predicted to cause significant reconstruction within as much as a ±2 eV window around the Fermi level, causing substantially weaker doping of the Fe d bands than expected based on stochiometric arguments [27]. The detailed understanding and theoretical analysis of this phenomenon for the Fe vacancies is left for future work. However, we note that the analysis presented here may also conceptually translate to analysis of the first-principles data, providing a path to compare theory with experiment through the effects of disorder on the electronic structure.

III. CONCLUSIONS

To conclude, we applied several dimensionality reduction techniques toward effective measurement of effects of disorder on the electronic properties of FeSe superconductor using tunneling spectroscopy. It is shown that the tunneling spectra are not linearly separable, owing to a likely combination of continuous variation of the electronic density of states, expected for atomic and nanoscale defects and impurities, combined with relatively low defect density. Meanwhile, several common nonlinear dimensionality reduction algorithms were not effective in separating of the vacancy sites in the latent space, which could potentially improve the follow-on clustering. Instead, twin (Siamese) neural network architecture, which learns the similarity directly from the data and

amplifies the similarities and dissimilarities during training, was shown to be effective at both creating well-separated projected space of the tunneling spectroscopy dataset as well as subsequently categorizing tunneling with continuous single-valued scores. The immediate advantage of this approach is the ability not only to detect and describe the properties of defects and impurities but also those of the transition region between distinct behaviors in the dataset. This is arguably a key advantage over discretization of the hyperspectral dataset with clustering methods, wherein neither the uniform state within a cluster nor the abrupt cluster boundaries are physically sound. Moreover, twin networks are well adapted to work with smaller-sized datasets characteristic of tunneling spectroscopy, due to N^2 scaling of the training set size with the N size of the dataset. In the specific case of FeSe, we have found that Fe impurities modify the density of states in a broad window around the Fermi level, well beyond the superconducting gap. This may be an indication of an even more dramatic electronic reconstruction, qualitatively resembling earlier predictions of the effect of Se vacancies [27]. Ultimately, these methods will serve to better understand the effects of disorder in complex electronic materials, toward tailored fundamental and applied properties for advancement of quantum electronics.

IV. METHODS

Tunneling spectroscopy and microscopy were carried out using a Specs Joule-Thomson STM microscope, operating at 4.5 K base temperature. Freshly cleaved surfaces were obtained by delaminating a small sample of FeSe in low vacuum of $\sim 10^{-6}$ Torr, followed by rapid sample transfer to the cryogenic chamber. Data analysis presented above was carried out using primarily Wolfram Language, with additional use of Python libraries for nonlinear dimensionality reduction. All the data and codes are available upon reasonable request.

ACKNOWLEDGMENTS

We thank Rama Vasudevan for discussion and improvement of the manuscript. Research sponsored by Division of Materials Science and Engineering, Basic Energy Sciences, Office of Science, U.S. Department of Energy (DOE). Experiments were carried out as part of a user project at the Center for Nanophase Materials Sciences, ORNL, a U.S. DOE of Science User Facility. This paper used resources of the Compute and Data Environment for Science (CADES) at the ORNL, which is supported by the Office of Science of the U.S. DOE under Contract No. DE-AC05-00OR22725. ORNL is managed by UT-Battelle, LLC, for the U.S. DOE. This paper is a contribution of the U.S. Government, not subject to U.S. copyright. The DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan [45].

P.M. guided research, helped to acquire data, carried out machine learning and interpretation. J.Y. and B.S. grew FeSe crystals. J.W. acquired the data and carried out initial data analysis. All authors contributed to writing of the manuscript.

[1] F. J. Giessibl, Advances in atomic force microscopy, Rev. Mod. Phys. **75**, 949 (2003).

[2] L. Olesen, M. Brandbyge, M. R. Sørensen, K. W. Jacobsen, E. Lægsgaard, I. Stensgaard, and F. Besenbacher, Apparent Barrier Height in Scanning Tunneling Microscopy Revisited, Phys. Rev. Lett. **76**, 1485 (1996).

[3] P. Maksymovych, M. Pan, P. Yu, R. Ramesh, A. P. Baddorf, and S. V. Kalinin, Scaling and disorder analysis of local I-V curves from ferroelectric thin films of lead zirconate titanate, Nanotechnology **22**, 254031 (2011).

[4] G. Binnig and H. Rohrer, Scanning tunneling microscopy—from birth to adolescence, Rev. Mod. Phys. **59**, 615 (1987).

[5] L. Grill, Functionalized molecules studied by STM: Motion, switching and reactivity, J. Phys.: Condens. Matter **20**, 053001 (2008).

[6] P. Maksymovych, Excitation and mechanisms of single molecule reactions in scanning tunneling microscopy, in *Scanning Probe Microscopy of Functional Materials*, edited by S. V. Kalinin and A. Gruverman (Springer, New York, 2010), pp. 3–37.

[7] Ø. Fischer, M. Kugler, I. Maggio-Aprile, C. Berthod, and C. Renner, Scanning tunneling spectroscopy of high-temperature superconductors, Rev. Mod. Phys. **79**, 353 (2007).

[8] J. E. Hoffman, Spectroscopic scanning tunneling microscopy insights into Fe-based superconductors, Rep. Prog. Phys. **74**, 124513 (2011).

[9] H. Lin, J. M. C. Rauba, K. S. Thygesen, K. W. Jacobsen, M. Y. Simmons, and W. A. Hofer, First-principles modelling of scanning tunneling microscopy using non-equilibrium Green's functions, Front. Phys. China **5**, 369 (2010).

[10] J. M. Blanco, F. Flores, and R. Pérez, STM-theory: Image potential, chemistry and surface relaxation, Prog. Surf. Sci. **81**, 403 (2006).

[11] S. V. Kalinin, E. Strelcov, A. Belianinov, S. Somnath, R. K. Vasudevan, E. J. Lingerfelt, R. K. Archibald, C. Chen, R. Proksch, N. Laanait, and S. Jesse, Big, deep, and smart data in scanning probe microscopy, ACS Nano **10**, 9068 (2016).

[12] M. Ziatdinov, A. Maksov, L. Li, A. S. Sefat, P. Maksymovych, and S. V. Kalinin, Deep data mining in a real Space: Separation of intertwined electronic responses in a lightly doped $BaFe_2As_2$, Nanotechnology **27**, 475706 (2016).

[13] A. Belianinov, P. Ganesh, W. Lin, B. C. Sales, A. S. Sefat, S. Jesse, M. Pan, and S. V. Kalinin, Research Update: Spatially resolved mapping of electronic structure on atomic level by multivariate statistical analysis, APL Mater. **2**, 120701 (2014).

[14] B. J. Albers, T. C. Schwendemann, M. Z. Baykara, N. Pilet, M. Liebmann, E. I. Altman, and U. D. Schwarz, Three-dimensional imaging of short-range chemical forces with picometre resolution, Nat. Nano. **4**, 307 (2009).

[15] E. Minelli, G. Ciasca, T. E. Sassun, M. Antonelli, V. Palmieri, M. Papi, G. Maulucci, A. Santoro, F. Giangaspero, R. Delfini, G. Campi, and M. De Spirito, A fully-automated neural network analysis of AFM force-distance curves for cancer tissue diagnosis, Appl. Phys. Lett. **111**, 143701 (2017).

[16] J. C. Agar, Y. Cao, B. Naul, S. Pandya, S. van der Walt, A. I. Luo, J. T. Maher, N. Balke, S. Jesse, S. V. Kalinin, R. K. Vasudevan, and L. W. Martin, Machine detection of enhanced electromechanical energy conversion in $PbZr_{0.2}Ti_{0.8}O_3$ thin films, Adv. Mater. **30**, 1800701 (2018).

[17] S. M. Neumayer, S. Jesse, G. Velarde, A. L. Kholkin, I. Kravchenko, L. W. Martin, N. Balke, and P. Maksymovych, To switch or not to switch—a machine learning approach for ferroelectricity, Nanoscale Adv. **2**, 2063 (2020).

[18] F. Ming, S. Johnston, D. Mulugeta, T. S. Smith, P. Vilmercati, G. Lee, T. A. Maier, P. C. Snijders, and H. H. Weitering, Realization of a Hole-Doped Mott Insulator on a Triangular Silicon Lattice, Phys. Rev. Lett. **119**, 266802 (2017).

[19] B. W. Hoogenboom, C. Berthod, M. Peter, Ø. Fischer, and A. A. Kordyuk, Modeling scanning tunneling spectra of $Bi_2Sr_2CaCu_2O_{8+\delta}$, Phys. Rev. B **67**, 224502 (2003).

[20] M. J. Rozenberg, G. Kotliar, and H. Kajueter, Transfer of spectral weight in spectroscopies of correlated electron systems, Phys. Rev. B **54**, 8452 (1996).

[21] A. K. Jain, Data clustering: 50 years beyond K-means, Pattern Recognit. Lett. **31**, 651 (2010).

[22] G. Haegg and A. L. Kindstroem, Röntgenuntersuchungen am System eisen—Selen, Z. Phys. Chem. **22B**, 453 (1933).

[23] S. Chopra, R. Hadsell, and Y. LeCun, Learning a similarity metric discriminatively, with application to face verification, in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (CVPR'05) (IEEE, San Diego, 2005), Vol. 1, pp. 539–546.

[24] D. Chicco, Siamese neural networks: An overview, in *Artificial Neural Networks*, edited by H. Cartwright, Methods in Molecular Biology (Humana, New York, 2021), Vol. 2190, pp. 73–94.

[25] A. Kreisel, P. J. Hirschfeld, and B. M. Andersen, On the remarkable superconductivity of FeSe and its close cousins, Symmetry **12**, 1402 (2020).

[26] J.-Q. Yan, B. C. Sales, M. A. Susner, and M. A. McGuire, Flux growth in a horizontal configuration: An analog to vapor transport growth, Phys. Rev. Materials **1**, 023402 (2017).

[27] T. Berlijn, H.-P. Cheng, P. J. Hirschfeld, and W. Ku, Doping effects of Se vacancies in monolayer FeSe, Phys. Rev. B **89**, 020501(R) (2014).

[28] X. Liu, L. Zhao, S. He, J. He, D. Liu, D. Mou, B. Shen, Y. Hu, J. Huang, and X. J. Zhou, Electronic structure and superconductivity of FeSe-related superconductors, J. Phys.: Cond. Matter **27**, 183201 (2015).

[29] D. Huang, T. A. Webb, C.-L. Song, C.-Z. Chang, J. S. Moodera, E. Kaxiras, and J. E. Hoffman, Dumbbell defects in FeSe films: a scanning tunneling microscopy and first-principles investigation, Nano Lett. **16**, 4224 (2016).

[30] K.-W. Lee, V. Pardo, and W. E. Pickett, Magnetism driven by anion vacancies in superconducting $\alpha-FeSe_{1-x}$, Phys. Rev. B **78**, 174502 (2008).

[31] T. Berlijn, P. J. Hirschfeld, and W. Ku, Effective Doping and Suppression of Fermi Surface Reconstruction via Fe Vacancy Disorder in $\mathbf{K}_xFe_{2-y}Se_2$, Phys. Rev. Lett. **109**, 147003 (2012).

[32] D. Huang and J. E. Hoffman, Monolayer FeSe on $SrTiO_3$, Annu. Rev. Condens. Matter Phys. **8**, 311 (2017).

[33] Z. Wang, P. Zhang, G. Xu, L. K. Zeng, H. Miao, X. Xu, T. Qian, H. Weng, P. Richard, A. V. Fedorov, H. Ding, X. Dai, and Z. Fang, Topological nature of the $FeSe_{0.5}Te_{0.5}$ superconductor, Phys. Rev. B **92**, 115119 (2015).

[34] L. Vlcek, A. Maksov, M. Pan, R. K. Vasudevan, and S. V. Kalinin, Knowledge extraction from atomically resolved images, ACS Nano **11**, 10313 (2017).

[35] L. Jiao, S. Rößler, C. Koz, U. Schwarz, D. Kasinathan, U. K. Rößler, and S. Wirth, Impurity-induced bound states inside the superconducting gap of FeSe, Phys. Rev. B **96**, 094504 (2017).

[36] Y. P. Raykov, A. Boukouvalas, F. Baig, and M. A. Little, What to do when *K*-means clustering fails: A simple yet principled alternative algorithm, PLoS One **11**, e0162259 (2016).

[37] F. Murtagh and P. Legendre, Ward's hierarchical agglomerative clustering method: Which algorithms implement Ward's criterion? J. Classif. **31**, 274 (2014).

[38] H. Hoffmann, Kernel PCA for novelty detection, Pattern Recognit. **40**, 863 (2007).

[39] L. van der Maaten and G. Hinton, Visualizing data using t-SNE, J. Mach. Learn. Res. **9**, 2579 (2008).

[40] B. Schölkopf, A. Smola, and K. R. Müller, Kernel principal component analysis, in *Artificial Neural Networks—ICANN'97. ICANN 1997*, edited by W. Gerstner, A. Germond, M. Hasler, and J. D. Nicoud, Lecture Notes in Computer Science Vol. 1327 (Springer, Berlin, Heidelberg, 1997), pp. 583–588, https://doi.org/10.1007/BFb0020217.

[41] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, SCIKIT-LEARN: Machine learning in Python, J. Mach. Learn. Res. **12**, 2825 (2011).

[42] L. McInnes, J. Healy, and J. Melville, UMAP: Uniform manifold approximation and projection for dimension reduction, arXiv:1802.03426.

[43] G. Koch, Siamese Neural Networks for One-Shot Image Recognition, Master's thesis, University of Toronto, 2015.

[44] T. C. Koethe, Z. Hu, M. W. Haverkort, C. Schüßler-Langeheine, F. Venturini, N. B. Brookes, O. Tjernberg, W. Reichelt, H. H. Hsieh, H.-J. Lin, C. T. Chen, and L. H. Tjeng, Transfer of Spectral Weight and Symmetry across the Metal-Insulator Transition in VO_2, Phys. Rev. Lett. **97**, 116402 (2006).

[45] http://energy.gov/downloads/doe-public-access-plan.