Unsupervised Machine Learning for Phase Identification and Characterization of High-Resolution STEM EELS in Novel Battery Materials

Zhen Yuan Yeo [1] **Wenhui Lai** [2] **Jong Hak Lee** [3] **Deepan Balakrishnan** [4] **Barbaros Özyilmaz** [2 3 5 6] **N. Duane Loh** [6]

[1] *Institute for Digital Molecular Analytics and Science, Nanyang Technological University, Singapore 636921, Singapore* [2] *Department of Material Science and Engineering, National University of Singapore, Singapore 117575, Singapore* [3] *Center for Advanced 2D Materials, National University of Singapore, Singapore 117546, Singapore* [4] *Department of Biological Sciences, National University of Singapore, Singapore 117543, Singapore* [5] *Institute for Functional Intelligent Materials, National University of Singapore, Singapore 117544, Singapore* [6] *Department of Physics, National University of Singapore, Singapore 117551, Singapore*. Correspondence to: barbaros@nus.edu.sg, duaneloh@nus.edu.sg.

1. Introduction

High-resolution Scanning Transmission Electron Microscopy-Electron Energy Loss Spectroscopy (STEM EELS) provides rich chemical and electronic information crucial for materials discovery [1, 2, 3]. However, when investigating novel battery materials, standard reference spectra may be unavailable or inapplicable due to differences in sample preparation and experimental conditions [4]. These challenges are compounded by high noise levels at the necessary resolutions, which can obscure critical spectral features.

Conventional approaches rely heavily on reference databases that often do not match the new material or experimental setup [4]. The proposed unsupervised machine learning workflow addresses this gap by revealing latent, chemically meaningful features without requiring labeled training data. By reducing spectral dimensionality and grouping similar spectral signatures, it becomes possible to identify previously unknown phases or compositions. This application of dimensionality reduction and information-theoretic measures (such as mutual information) for direct chemical bonding insights is a significant advancement in the field.

As a proof of concept, we apply this machine learning workflow to characterize a novel silicon–carbon composite anode in a high-performance battery. While graphite is the conventional anode material, silicon has drawn attention for its approximately tenfold larger specific capacity [5, 6]. Despite its advantages, the widespread adoption of silicon has been hampered by significant volumetric expansion during charge–discharge cycles [7, 8]. To mitigate this issue, silicon nanoparticles are embedded in nanoporous amorphous carbon using advanced synthesis methods [9]. However, the hierarchical composition of these materials and the low signal-to-noise ratio at individual raster pixels make phase determination particularly challenging.

2. Methodology

Data Acquisition: STEM EELS spectra on the Carbon K-edge, Oxygen K-edge, and Silicon L-edge were collected from Silicon nanoparticles and amorphous Carbon composite. The thickness contrast via low loss EELS is shown in Figure 1a. **Preprocessing**: Each elemental edge was individually background-subtracted to remove thickness-dependent noise from low-loss regions, shown in Figure 1b. A further principal component analysis (PCA) was applied for noise reduction. **Dimensionality Reduction**: Uniform Manifold Approximation and Projection (UMAP) [10] was applied to each elemental map to group pixels with similar spectral signatures, shown in Figure 1c. K-Means clustering was applied to identify classes of spectra in this latent representation. The spatial relationship of these classes are visualized by labelling the pixels in the STEM EELS image. **Correlated Composition Identification**: Mutual information analysis was used to detect regions with strong co-occurrences of chemical signals, suggesting the presence of specific bonds. **Validation**: Average spectra from each cluster were compared to reference EELS data to confirm compositional assignments.

3. Discussion

Cluster Analysis: The UMAP clustering revealed distinct compositional classes in the high-resolution data, even under noisy conditions. **Identification of Chemical Bonds**: Clusters exhibiting high mutual information suggested the coexistence of specific bonding environments, such as Si–C, Si–O, and elemental Si. **Reference Comparison**: Averaged cluster spectra (vector-quantized by class) aligned well with known EELS database references [11, 12, 13], confirming the presence of SiC, SiO_2, and pure Si. **Local Chemical Heterogeneity**: The spatial distribution of these classes revealed significant chemical heterogeneity in the battery anode, which would be difficult to ascertain using conventional, supervised analysis.

4. Conclusion and implications

This work demonstrates that unsupervised machine learning methods, particularly dimensionality reduction and mutual information analysis, can uncover previously elusive compositional details in novel battery materials. By bypassing the need for comprehensive reference databases and effectively mitigating noise inherent to high-resolution imaging, this approach enables rapid identification of new phases and bonding environments. The impli-

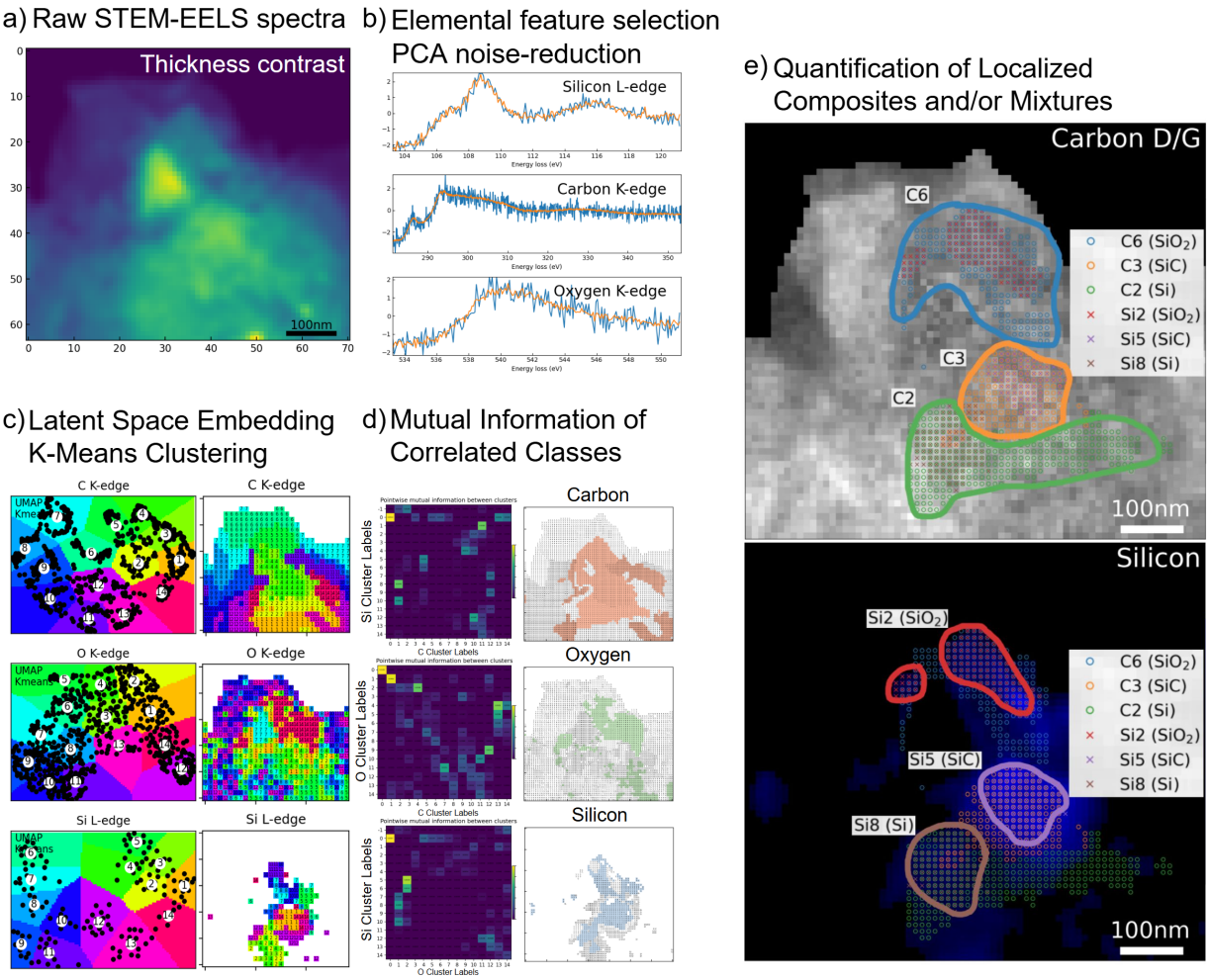

Fig. 1: Unsupervised analysis pipeline for STEM-EELS data in a Silicon–Carbon Composite. (a) Raw STEM-EELS spectral map, showing thickness contrast across the silicon–carbon composite anode region. (b) Example spectra from the silicon L-edge, carbon K-edge, and oxygen K-edge after elemental feature selection and principal component analysis (PCA) for noise reduction. (c) Latent-space embeddings (via UMAP) of each edge, with k-means clustering, to reveal distinct spectral signatures that corresponds to different local chemical environments. (d) Mutual information analysis quantifies correlations between clusters, confirming which chemical phases co-occur (e.g., Si–C, Si–O). (e) Spatial mapping of identified clusters highlights localized compositional regions (e.g., SiC, SiO_2, elemental Si), illustrating how the unsupervised approach distinguishes heterogeneous chemical domains within the material. Scale bars are 100 nm.

cations extend beyond battery research to fields requiring high-precision chemical characterization of materials for which robust reference spectra are unavailable or incomplete, thereby accelerating materials discovery and optimization.

Acknowledgments

The authors gratefully acknowledge the National University of Singapore (NUS), NUS Department of Physics Ph.D. scholarship, Institute for Digital Molecular Analytics and Science (IDMxS), the National Research Foundation, Prime Minister's Office, Singapore, under its Competitive Research Programme (CRP award number NRF-CRP22-2019-008) and Medium-Sized Centre Programme (CA2DM), by Ministry of Education of Singapore, under its Research Centre of Excellence award to the Institute for Functional Intelligent Materials (I-FIM, Project No. EDUNC-33-18-279-V12), and by EDB Singapore, under its Space Technology Development Programme (S22-19013-STDP), and by Agency for Science, Technology and Research (A*STAR), under its LCER Phase 2 Programme, HET FI (Grant No. U2311D4008).

References

[1] Sergei V Kalinin, Debangshu Mukherjee, Kevin Roccapriore, Benjamin J Blaiszik, Ayana Ghosh, Maxim A Ziatdinov, Anees Al-Najjar, Christina Doty, Sarah Akers, Nageswara S Rao, Joshua C Agar, and Steven R Spurgeon. Machine learning for automated experimentation in scanning transmission electron microscopy. *Npj Comput. Mater.*, 9(1):1–16, December 2023.

[2] Yanbin Ning, Feng Yang, Yan Zhang, Zhuomin Qiang, Geping Yin, Jiajun Wang, and Shuaifeng

Lou. Bridging multimodal data and battery science with machine learning. *Matter*, 7(6):2011–2032, June 2024.

[3] Sergei V Kalinin, Kevin M Roccapriore, Shin Hum Cho, Delia J Milliron, Rama Vasudevan, Maxim Ziatdinov, and Jordan A Hachtel. Separating physically distinct mechanisms in complex infrared plasmonic nanostructures via machine learning enhanced electron energy loss spectroscopy. *Adv. Opt. Mater.*, 9(13):2001808, July 2021.

[4] Steven R Spurgeon, Colin Ophus, Lewys Jones, Amanda Petford-Long, Sergei V Kalinin, Matthew J Olszta, Rafal E Dunin-Borkowski, Norman Salmon, Khalid Hattar, Wei-Chang D Yang, Renu Sharma, Yingge Du, Ann Chiaramonti, Haimei Zheng, Edgar C Buck, Libor Kovarik, R Lee Penn, Dongsheng Li, Xin Zhang, Mitsuhiro Murayama, and Mitra L Taheri. Towards data-driven next-generation transmission electron microscopy. *Nat. Mater.*, 20(3):274–279, March 2021.

[5] Jingxing Wu, Yinliang Cao, Haimin Zhao, Jianfeng Mao, and Zaiping Guo. The critical role of carbon in marrying silicon and graphite anodes for high-energy lithium-ion batteries. *Carbon Energy*, 1(1):57–76, September 2019.

[6] Xinghao Zhang, Debin Kong, Xianglong Li, and Linjie Zhi. Dimensionally designed carbon-silicon hybrids for lithium storage. *Adv. Funct. Mater.*, 29(2):1806061, January 2019.

[7] Zhifan Hu, Ran Zhao, Jingjing Yang, Chuan Wu, and Ying Bai. Binders for si based electrodes: Current status, modification strategies and perspective. *Energy Storage Mater.*, 59(102776):102776, May 2023.

[8] Ebrahim Feyzi, Anil Kumar, Xia Li, Sixu Deng, Jagjit Nanda, and Karim Zaghib. A comprehensive review of silicon anodes for high-energy lithium-ion batteries: Challenges, latest developments, and perspectives. *Next Energy*, 5(100176):100176, October 2024.

[9] Jong Hak Lee, N Duane Loh, Zhen Yuan Yeo, Yong Kang Ong, Deepan Balakrishnan, Carlos Maria Alava Limpo, Abhik Datta, Cagdas Cetin, Shoucong Ning, Clarissa Wong, Jian Shi, Fuchen Hou, Junhao Lin, Tadahiro Minamikawa, Tomonori Ito, Hiroyuki Kamisuki, Stephen Pennycook, Paul Matsudaira, and Barbaros Özyilmaz. Engineering a hierarchy of disorder: A new route to synthesize high-performance 3D nanoporous all-carbon materials. *Adv. Mater.*, page e2402628, April 2024.

[10] Leland McInnes, John Healy, Nathaniel Saul, and Lukas Großberger. UMAP: Uniform manifold approximation and projection. *J. Open Source Softw.*, 3(29):861, September 2018.

[11] W M Skiff, R W Carpenter, and S H Lin. Si*L* core edge fine structure in an oxidation series of silicon compounds: A comparison of microelectron energy loss spectra with theory. *J. Appl. Phys.*, 58(9):3463–3469, November 1985.

[12] Philip Ewels, Thierry Sikora, Virginie Serin, Chris P Ewels, and Luc Lajaunie. A complete overhaul of the electron energy-loss spectroscopy and X-ray absorption spectroscopy database: Eelsdb.eu. *Microsc. Microanal.*, 22(3):717–724, June 2016.

[13] Philip Edward Batson. Current trends for EELS studies in physics. *Microscopy Microanalysis Microstructures*, 2(2-3):395–402, 1991.