AI-Guided Design and Discovery of Silicon-Based Anode Materials for Lithium-Ion Batteries

David García Valcarce
HI iberia
Madrid, Spain
dgarcia@hi-iberia.es

César Alonso
HI iberia
Madrid, Spain

Eduardo Abenza
HI iberia
Madrid, Spain

Roberto Gómez-Espinosa
HI iberia
Madrid, Spain

María Carmen Asensio
Instituto de Ciencia de Materiales de Madrid (ICMM/CSIC)
MATINÉE: CSIC Associated Unit Between ICMUV and ICMM
Madrid, Spain

Abstract

The development of high-capacity anodes remains a central challenge for advancing lithium-ion battery (LIB) technology. Silicon (Si) offers an exceptional theoretical capacity but suffers from severe volume expansion, structural degradation, and limited Li-ion mobility after cycling. Here, we present an AI-driven framework for the analysis, generation, and optimization of silicon-based anode materials with enhanced Li-ion transport and controlled volume variation during lithiation and delithiation. Our approach integrates large-scale materials databases with state-of-the-art machine learning to: (i) analyze Li-ion migration pathways in known compounds, (ii) generate lithiation states from charged configurations, and (iii) design novel Si-based materials with enlarged Li-ion migration channels while tracking their volume changes across lithiation levels. By combining graph-based generative models with efficient property predictors, we accelerate the discovery of candidate anodes through database-guided screening. The results highlight the potential of AI to identify next-generation silicon-based anode materials with improved stability and performance, providing valuable guidance for experimental development.

1 Introduction

1.1 Li-ion batteries, Si anode, pros and cons

Lithium-ion batteries (LIBs) are the dominant rechargeable power source for modern electric vehicles and portable electronics due to their high energy density and long cycle life Tarascon and Armand [2001]. In current LIB technology, the anode is typically a graphitic carbon material (LiC_6), which is favored for its stable intercalation chemistry and good cyclability. However, graphite's theoretical specific capacity is only about $372 \mathrm{mAhg}^{-1}$ Tarascon and Armand [2001], and even alternative commercial anodes like spinel lithium titanate ($Li_4Ti_5O_{12}$, LTO) offer just $\sim 175 \mathrm{mAhg}^{-1}$ Zaghib et al. [2010]. These capacity limitations of present anode materials impose a fundamental cap on the energy density of LIBs, motivating extensive research into higher-capacity replacements Armand and Tarascon [2008].

Among the various proposed next-generation anodes, silicon (Si) has emerged as a particularly promising candidate. Silicon can alloy with lithium to form phases such as $Li_{15}Si_4$, endowing it with an exceptionally high theoretical capacity of approximately 3500–$3600 \mathrm{mAhg}^{-1}$ (nearly an order

39th Conference on Neural Information Processing Systems (NeurIPS 2025).

of magnitude greater than graphite) Larcher and Tarascon [2015], Mi et al. [2013]. In addition, Si is an abundant and low-cost element (the second most plentiful in the Earth's crust, accounting for about 25% of its mass) Mi et al. [2013], Toki and Obrovac [2024]. Combined with a relatively low average lithiation potential (0.4V vs Li/Li$^+$) Larcher and Tarascon [2015], these attributes mean that Si anodes could dramatically increase the energy density of LIB cells. Indeed, the incorporation of silicon is widely seen as a key pathway towards next-generation high-energy batteries Larcher and Tarascon [2015], Toki and Obrovac [2024].

Despite its attractive advantages, silicon anode technology still faces critical challenges that have hindered its large-scale commercialization. The most severe issue is the dramatic volume expansion of silicon during lithiation: fully lithiated Si can swell by 2̃80–300%, generating intense mechanical stress Obrovac and Christensen [2004]. This repeated expansion and contraction pulverize Si particles and break electrical contact within the electrode, leading to rapid capacity fading. In addition, the large volume swings destabilize the solid electrolyte interphase (SEI); continual cracking and reformation of the SEI consume both cyclable lithium and electrolyte, resulting in low coulombic efficiency and short cycle life Toki and Obrovac [2024]. Silicon's intrinsically low electrical conductivity further limits rate capability compared to graphite unless properly engineered Larcher and Tarascon [2015]. A wide range of strategies have been explored to mitigate these issues, including nanostructured Si architectures (nanoparticles, nanowires, porous Si) that accommodate strain, Si–carbon composites that enhance conductivity and buffer volume changes, and advanced binders and electrolyte additives that stabilize the SEI Zuo et al. [2017], Toki and Obrovac [2024]. While these approaches have delivered notable improvements in laboratory cells, achieving Si-based anodes that simultaneously provide high capacity and long-term stability remains elusive. This persistent challenge highlights the urgent need for innovative materials, architectures, and interdisciplinary approaches to fully unlock the potential of silicon anodes in next-generation LIBs.

1.2 Material Databases

The *Materials Genome Initiative* (MGI), launched in 2011, aims to accelerate materials discovery through the integration of high-throughput computation, experimental tools, and digital data infrastructure National Science and Technology Council [2011]. In response, several large-scale inorganic materials databases have emerged, forming the foundation for AI-assisted materials design, particularly in lithium-ion battery (LIB) research. This section reviews key resources—including the Materials Project, OQMD, NOMAD, SPSE, COD, and GNoME—emphasizing their data types, reliability, and relevance to Si-based anode discovery.

The **Materials Project** (MP) Jain et al. [2013] and **Open Quantum Materials Database** (OQMD) Saal et al. [2013], Kirklin et al. [2015] are two of the most established repositories. Both perform standardized DFT calculations (typically with VASP) to determine structural, electronic, and thermodynamic properties for thousands of compounds. MP offers curated, consistent data such as formation energies, relaxed structures, and band gaps, while OQMD complements it by systematically generating hypothetical compounds through prototype substitutions, expanding chemical coverage and enabling phase stability analyses.

These datasets have been extensively used to study Li–Si compounds and to train or validate ML models predicting material stability and electrochemical performance. The **NOMAD** (Novel Materials Discovery) repository Draxl and Scheffler [2019], by contrast, aggregates raw computational data from diverse sources and codes. Its heterogeneity poses challenges for curation but provides an unparalleled scale and diversity beneficial for transfer learning and testing model robustness.

The **Screening Platform for Solid Electrolytes** (SPSE) He et al. [2020] focuses on lithium-ion transport properties—data rarely available in general-purpose databases—while the **Crystallography Open Database** (COD) Gražulis et al. [2012] supplies experimentally determined crystal structures useful for model validation or DFT input generation. Finally, **GNoME** Merchant et al. [2023], developed by Google DeepMind, exemplifies AI-driven materials discovery: through an active-learning loop, it generated over two million crystal structures, hundreds of thousands of which are predicted to be stable, including many new Li–Si compounds.

Together, these databases constitute a rich ecosystem for data-driven materials discovery. MP and OQMD provide curated DFT data for model training; NOMAD offers massive, heterogeneous datasets; SPSE and COD address specialized data gaps; and GNoME showcases the potential of

AI-guided materials generation—collectively empowering machine learning approaches to accelerate next-generation silicon anode design.

1.3 Artificial Intelligence in Materials Science

The integration of artificial intelligence (AI) with materials science has opened new avenues for accelerating the design, discovery, and deployment of advanced materials Butler et al. [2018], Lookman et al. [2019], Schmidt et al. [2019]. Machine learning (ML) models can extract complex structure–property relationships from high-dimensional data, while generative approaches can propose entirely new compounds with targeted properties. These techniques are now playing a key role in reducing the time and cost associated with traditional trial-and-error experimental workflows.

Property Prediction via Supervised Learning Supervised learning models, including random forests, kernel ridge regression, and neural networks, are widely used to predict materials properties such as formation energies Ward et al. [2016], band gaps Jha et al. [2018], elastic moduli De Jong et al. [2016], and ionic conductivity Sendek et al. [2018]. These models are typically trained on large datasets from first-principles calculations (e.g., Materials Project, OQMD) and rely on either hand-crafted descriptors (e.g., composition, structure) or learned features. Graph neural networks such as CGCNN Xie and Grossman [2018], MEGNet Chen et al. [2019], and M3GNet Chen and Ong [2022] have become state-of-the-art for structure-based prediction of material properties, including band gaps, formation energies, and elastic constants.

Machine Learning Interatomic Potentials and Structural Relaxation Machine-learned interatomic potentials (ML-IPs) allow for fast, accurate simulations of atomic-scale interactions, enabling structural relaxation and molecular dynamics at near-DFT accuracy but much lower cost. Recent models such as NequIP Batzner et al. [2022] and M3GNet Chen and Ong [2022] exploit equivariant architectures and large-scale datasets to deliver transferable and data-efficient interatomic potentials. These ML-IPs are increasingly used in high-throughput screening, structural refinement, and molecular simulations.

Generative Models and Inverse Design Generative models like variational autoencoders (VAEs) Noh et al. [2020], generative adversarial networks (GANs) Kim et al. [2020], and diffusion models Xie et al. [2022] are being applied to create novel crystal structures and compositions. These models can be conditioned on desired target properties, enabling inverse design of materials. Recent work has demonstrated the generation of stable and synthesizable structures with targeted physical or electronic properties.

Active Learning and Reinforcement Learning Active learning has emerged as a key strategy in closed-loop discovery, where models identify the most informative experiments or simulations to perform next Lookman et al. [2019]. Reinforcement learning (RL) has also been explored for materials optimization and inverse design Karpov et al. [2020]. These approaches reduce the cost of exploration in vast materials spaces, enabling data-efficient discovery cycles.

Applications to Battery Materials AI methods have been successfully applied to various aspects of battery materials research, including cathode and anode discovery Sendek et al. [2018], Kirklin et al. [2013]. In particular, graph-based models and generative approaches have been used to predict and generate promising lithium-ion battery (LIB) anode materials beyond graphite, including silicon-based and alloy-type candidates.

2 Methodology

Our methodology addresses the challenge of designing silicon-based anodes for lithium-ion batteries (LIBs) that retain high capacity while maintaining structural integrity during charge/discharge cycles. We propose a three-stage approach that prioritizes high capacity Si-based alloys exhibiting minimal volume change upon lithiation:

1. **Database mining:** Identify candidate Si-alloy anodes in which Li-ion channel sizes remain relatively unchanged across different lithiation degrees.

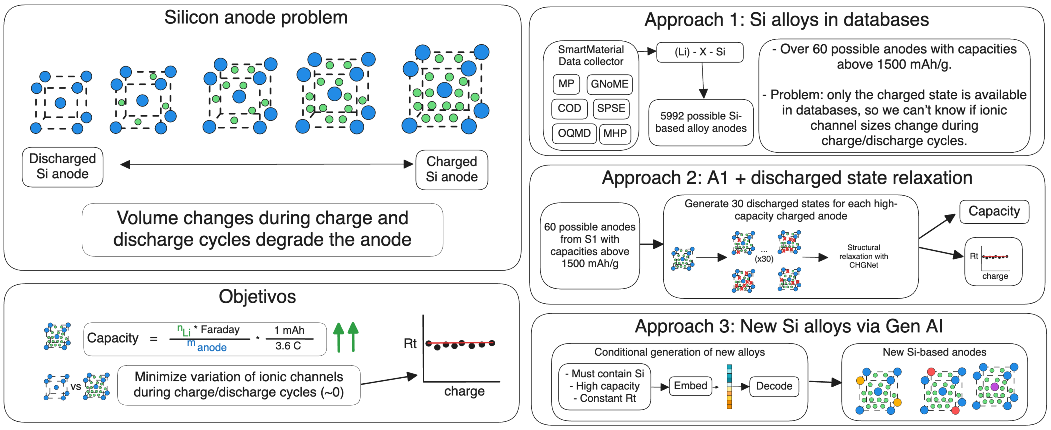

Figure 1: Overview of the proposed methodology. We address degradation in Si anodes due to volume changes by (1) mining Si-alloy compounds from databases, (2) relaxing hypothetical discharged states with CHGNet to evaluate ion channel stability, and (3) generating new stable anodes via generative AI models.

2. **Structural relaxation:** Relax discharged states using machine-learning interatomic potentials to efficiently evaluate stability and ion transport, and calculate the ion channel size for each lithiation degree **?**.
3. **Generative AI design:** Propose novel, stable Si-based anode materials with optimized Li-ion migration pathways and controlled volume variation.

The overall workflow is summarized in Figure 1.

2.1 Problem: Volume Change in Si Anodes

Silicon undergoes large volumetric expansion ($\sim$300%) during lithiation, which leads to pulverization of the anode, unstable SEI formation, and capacity fading. We aim to identify or generate Si-based anodes with high capacity and minimal variation in lithium-ion migration channels across charge/discharge cycles. Two main objectives guide our search: (1) high gravimetric capacity (above 1500 mAh/g), and (2) stable ion migration paths (constant R_t) over the full lithiation range.

2.2 Stage 1: Database Mining of Si-X Alloys

We compiled and harmonized data from multiple materials databases—including the Materials Project (MP), GNoME, OQMD, COD, SPSE, and the Marques–Halle Perovskite Dataset 166 (MHP)—to construct a unified dataset. We filtered for charged $\mathrm{Li}_x\mathrm{Si}_y\mathrm{X}_z$ compounds (where X = metallic or metalloid element) with computed capacities above 1500 mAh/g.

The initial screening yielded 5,992 prospective silicon anode candidates. Yet, existing databases report only charge states, lacking information on structural evolution during delithiation. This is a critical gap, as silicon—despite its high theoretical capacity (3,579 mAh/g) for $\mathrm{Li}_{15}\mathrm{Si}_4$)—suffers from severe volume changes (300%) upon cycling, leading to particle fracture, unstable SEI formation, and rapid capacity fading.

Our analysis shows that variations in lithium-ion migration channel dimensions across charge–discharge cycles are a key structural descriptor directly linked to volumetric instability. Unlike conventional mitigation strategies such as coatings, nanostructuring, or binders, this lattice-level descriptor captures the intrinsic resilience of silicon alloys. To bridge this gap, we systematically quantified Li^+ migration channel sizes in all candidate alloys while assessing their specific capacity. Only those with invariant channel dimensions across lithiation–delithiation can be considered structurally resilient, capable of suppressing expansion and degradation.

This descriptor-driven approach identifies a subset of robust silicon alloys that couple high energy density with improved durability. More broadly, it establishes lattice invariance of migration channels as a predictive design principle for alloy-type anodes, advancing the rational development of practical silicon-based electrodes for LIBs.

2.3 Stage 2: Generation and Relaxation of Discharged States

To overcome the lack of discharged-state structures, we generated a set of hypothetical discharged structures for each high-capacity candidate identified in Stage 1. For each charged anode, 30 discharged configurations were created by probabilistically removing Li atoms.

The generated discharged structures were relaxed using **CHGNet**, a recent graph neural network-based surrogate model for structural relaxation Deng et al. [2023]. CHGNet allows fast and approximate relaxation of large batches of crystal structures, providing an efficient alternative to expensive DFT calculations. Its efficiency against Si-Li-X materials can be studied through the published Matbench Discovery results, showing a MAE for predicted energy of 0.4117 against DFT and a RMSD (Root Mean Squared Displacement) of atomic positions of just over 0.01.Riebesell et al. [2023]

For each relaxed discharged structure, we recomputed the capacity and extracted the channel radius (R_t) for lithium-ion migration using the CAVD package He et al. [2020]. We retained only those candidate materials where R_t remains nearly constant between charged and discharged states, indicating robustness of ion transport paths across cycles.

2.4 Stage 3: Generative Design of New Si-Based Anodes

To explore materials beyond the existing database entries, we developed a generative AI model for conditional design of new Si-based anode materials.

The model takes as input the desired characteristics:

- Presence of silicon in the composition
- Target gravimetric capacity (e.g., $>$1500 mAh/g)
- Target channel radius R_t (and optionally, minimal variation across cycling)

We trained a conditional variational autoencoder (CVAE) model using crystal graph embeddings of existing anodes and their associated properties. The latent space was then sampled and decoded to generate novel crystal structures that meet the target constraints.

Generated candidates were evaluated using the same screening criteria from Stages 1 and 2: capacity and robustness of ion channels during cycling. High-performing candidates are further considered for DFT validation and experimental synthesis.

3 Results and Discussion

3.1 Li$^+$ Migration Channels in Si and Si–X Anodes

Silicon anodes suffer from severe mechanical degradation due to large volume changes during lithiation and delithiation, which disrupt continuous Li$^+$ diffusion channels and cause rapid capacity fade Ashuri et al. [2016], Li et al. [2020]. In particular, pure crystalline silicon expands by over 300% upon full lithiation, leading to fracture, particle pulverization, and unstable solid-electrolyte interphase (SEI) formation Ashuri et al. [2016].

In this context, Figure 2 quantifies this volume variation as a function of the size change of Li$^+$ migration channels in pure Si during charge/discharge cycles. The slope of the channel radius with respect to state of charge is 0.85 ± 0.23 Å/100%, indicating substantial structural evolution in the diffusion pathways as Li is inserted and removed. This variation suggests that Si loses its percolating ion transport networks during delithiation, contributing to mechanical failure and capacity degradation.

To address this limitation, in the next section we explore Si–X alloys with the goal of identifying compositions that retain stable the volume of Li$^+$ migration channels throughout cycling.

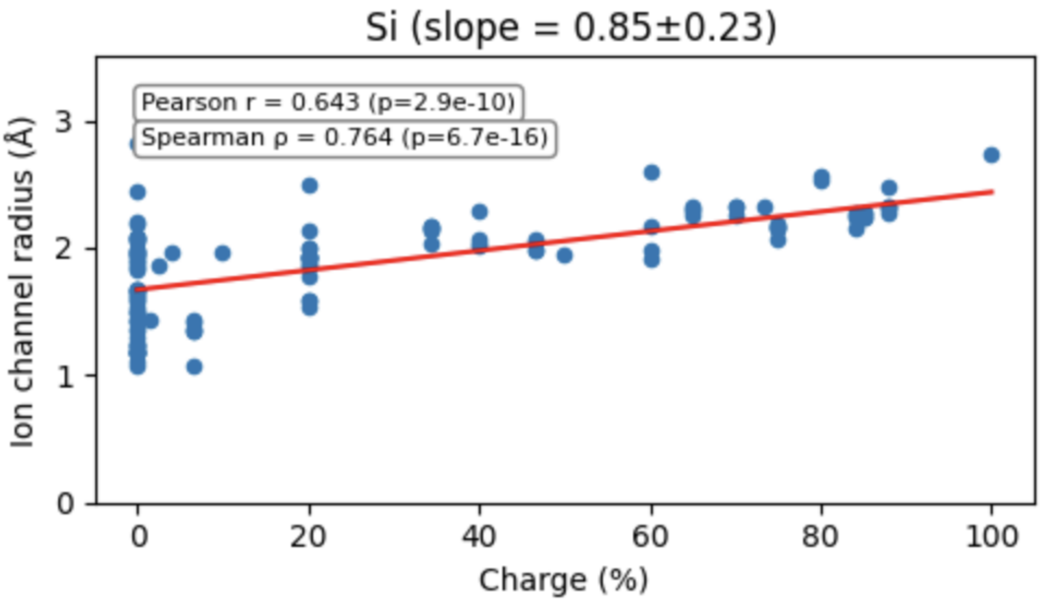

Figure 2: Evolution of Li^+ ion channel radius during delithiation of a pure Si anode. Each point corresponds to a database structure with decreasing Li content, and the red line shows a linear regression fit. The slope of 0.85 ± 0.23 indicates strong variation in channel geometry across charge states. The correlation between channel radius and Li content is strong, with Pearson $r = 0.634$ ($p = 2.9 \times 10^{-10}$) and Spearman $\rho = 0.764$ ($p = 6.7 \times 10^{-16}$), highlighting a robust monotonic relationship.

3.2 High-Throughput Screening: Capacity vs. Channel Size

To identify inherently stable Si-based anodes, we performed a high-throughput search of known crystalline compounds containing Si in fully lithiated form (the charged anode candidates). We aggregated ∼6000 such structures from multiple materials databases (e.g. Materials Project, OQMD, COD), each providing a theoretical specific capacity based on its Li content. For each candidate, we also computed the radius of the largest continuous Li^+ migration channel in the charged-state structure using geometric analysis. Figure 3 plots the screened compounds in terms of gravimetric capacity (mAh g^{-1}) vs. the logarithm of the Li-channel radius (log(Å)). Many Li–Si–X compounds lie in a regime of either high capacity with narrow channels or vice versa. Notably, only a small subset of about 60 candidates exhibit capacities beyond 1500 mAh g^{-1} (far exceeding graphite) while also having moderately large channels (∼1.5–2.5 Å). These high-capacity points tend to correspond to Si-alloy frameworks or intermetallic phases that incorporate Li at high stoichiometry. No obvious linear correlation emerges between capacity and channel size across all data – i.e., increasing Li content (capacity) does not guarantee open channels. Nonetheless, the spread of points in Figure 3 helped pinpoint promising outliers: a few compositions combine relatively spacious Li channels with excellent capacity. We hypothesized that these rare candidates might better tolerate lithiation strain by providing free volume for Li-ion transport. However, a key question remained: do the attractive channels observed in the charged (lithiated) structures persist once the Li is removed? Since database structures are typically reported in their lithiated form only, we next investigated how these materials behave upon delithiation.

3.3 AI-Relaxed Discharged Structures and Channel Stability

To assess the structural resilience of promising Si–X anode candidates, we investigated how their Li^+ ion migration channels sizes evolve across charge/discharge cycles. Starting from the fully lithiated structures found in databases (i.e., charged state), we simulated several partially-discharged states by randomly removing Li atoms and structurally relaxing each resulting configuration using CHGNet Deng et al. [2023], a machine-learning interatomic potential that captures charge effects

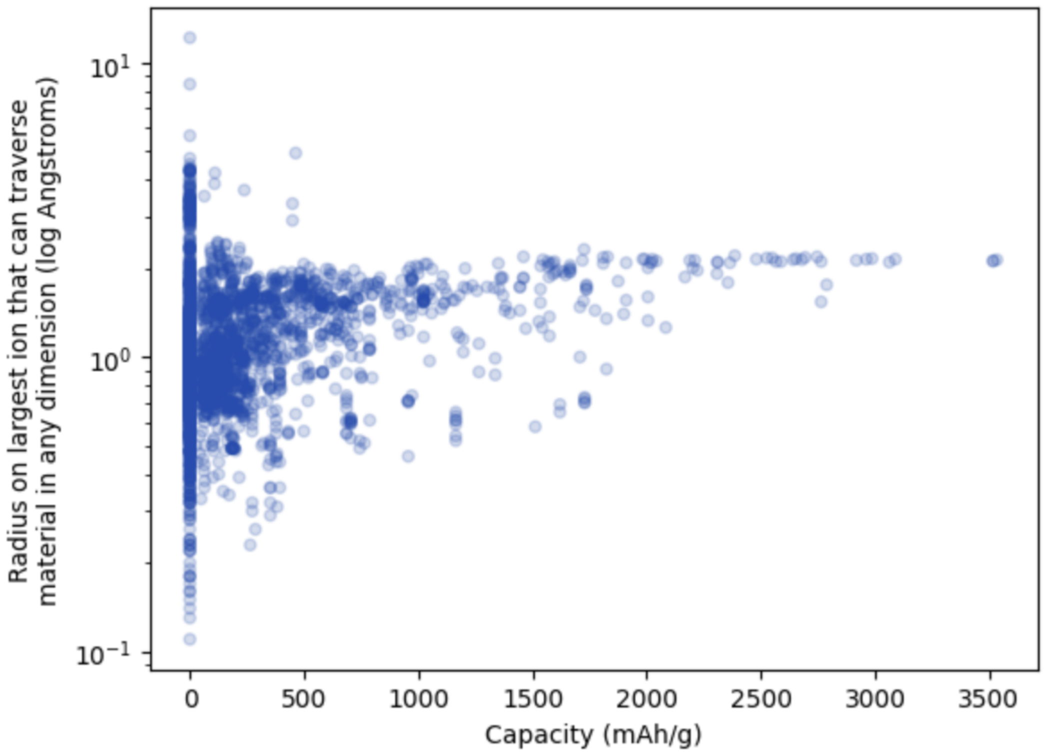

Figure 3: High-throughput screening of ∼6000 fully lithiated Si-containing compounds collected from multiple databases (Materials Project, OQMD, COD). Each point represents a candidate structure plotted by its theoretical gravimetric capacity (mAh g^{-1}) versus the radius of its largest continuous Li^+ migration channel (log Å, from geometric analysis).

and lattice dynamics at low cost. The authors acknowledge that more sophisticated approaches exist to model a realistic delithiation cycle; however, even with this simplified strategy, we successfully identified known alloy phases exhibiting lower variation in channel size. For each candidate, 30 distinct Li-vacancy configurations were generated to probe the variability in ion channel geometry under different de-lithiation paths.

Figure 4 shows two examples of silicon-based anodes with minimal variation in ion channel diameter during delithiation: Si7Ge1 and Zn1Si8, both of which have been previously experimentally studied in the literature as promising LIB silicon alloy anodes. Their slopes—0.44 and 0.39 respectively—are significantly lower than that of pure Si (0.85, see Figure 2), indicating that their frameworks accommodate Li extraction with limited tunnel collapse.

While these known materials validate our screening approach, our method also identified several novel Si–X compounds with similarly low slopes and stable channel behavior. These have not yet been reported in the literature, and due to their potential commercial value, we refrain from disclosing them in this manuscript. Still, the general insight holds: a small subset of Si–X alloys preserve open Li^+ migration pathways across cycling, likely due to supportive sublattice architectures or alloying elements that reinforce the Si framework. These structures are strong candidates for anodes with enhanced cycling stability.

3.4 Conditional Generative Design of Novel Si-Based alloy Anodes

Given the limited number of known materials that inherently satisfy both high capacity and stable-channel criteria, we turned to generative machine learning to explore new Si-based alloy anode compositions. Using a modified version of the conditional crystal generation model developed by Ye et al. [2024], with a stronger emphasis on controlling the presence or absence of specific elements in the generated crystals, we trained on the distribution of Li–Si–X structures and then

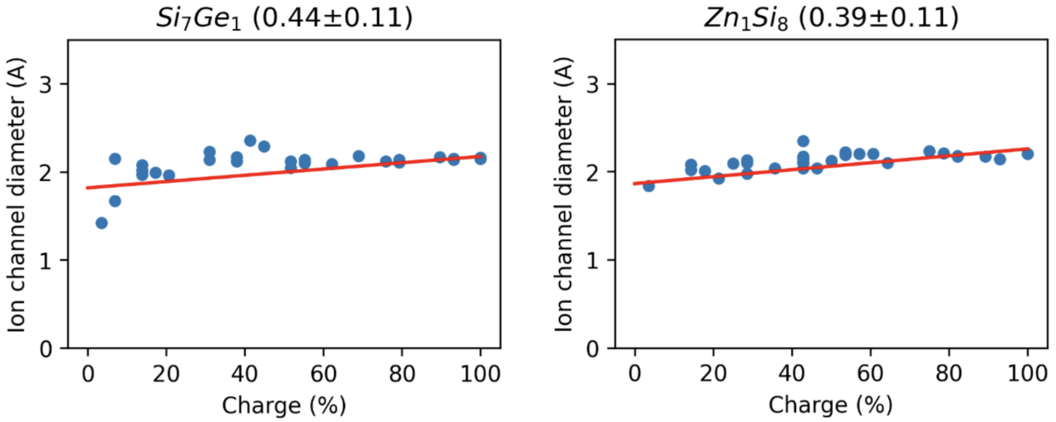

Figure 4: Evolution of Li^+ ion channel diameter during delithiation of two Si–X anodes: Si7Ge1 (left) and Zn1Si8 (right). Each point represents a CHGNet-relaxed structure with a different random Li-vacancy configuration. The fitted slopes (< 0.45) indicate much smaller channel variation than in pure Si (see Figure 2). These two examples, well documented in the literature, serve as validation of our method; other low-slope alloys found are withheld for IP reasons.

directed the generator toward candidates with specified target properties. In our case, we conditioned on (i) the inclusion of Si and Li as constituents, (ii) a capacity above 1500 mAh g^{-1}, and (iii) a large Li-channel radius that remains approximately constant upon (de)lithiation. The generative model learned a continuous latent representation of crystal structures and was able to sample novel compositions that meet these conditions, akin to recent advances in property-driven materials design Xie et al. [2022], Pakornchote et al. [2024]. Figure 5 showcases examples of AI-generated Si-based alloy anodes, as well as an overview of the property-based generation accuracy of the model. The model often reproduces structural motifs favorable for Li diffusion: for instance, several generated structures feature interconnected cages or tunnels that persist after removal of Li. In one example, the algorithm suggested a Si–Cu binary alloy forming a porous clathrate-like lattice – a material not present in our original database. This hypothetical anode achieves an estimated capacity of $\sim$1700 mAh g^{-1} with a continuous channel of $\sim$2.1 Å that remains open post-delithiation, according to our CHGNet relaxations. Another generated candidate is a ternary Li–Si–Sn compound in which Sn atoms occupy positions that stabilize a 3D network of Si polyhedra, yielding minimal volume change during cycling. These qualitative patterns align with our earlier observations: the best designs incorporate a supportive matrix (metal or other element) that holds open the Li diffusion channels. All AI-generated structures were verified for stability before further consideration. To achieve this, each structure underwent CHGNet relaxation following the procedure described in Section 2.3. The mean variation in energy per atom was found to be below 4%. While experimental synthesis and testing of these predictions are ongoing, the ability to tailor new anode materials via conditional generation is a powerful outcome. It demonstrates that machine learning can not only screen existing materials but also inversely design novel compounds that optimize multiple criteria simultaneously Xie et al. [2022], Pakornchote et al. [2024]. Ultimately, such AI-designed Si-based alloy anodes – with their combination of high capacity and robust Li^+ channel architectures – hold promise for breaking the trade-off between energy density and cycle life in next-generation Li-ion batteries.

4 Conclusion

In this work, we present an AI-driven framework for the discovery, evaluation, and design of high-performance silicon-based alloy anode materials for lithium-ion batteries. To address the critical challenge of structural degradation during cycling, we identify and evaluate a physically meaningful descriptor closely linked to the volume changes occurring during lithiation and delithiation. Building on this insight, we propose a three-stage methodology that integrates materials database mining, surrogate-assisted structural relaxation, and conditional generative modeling.

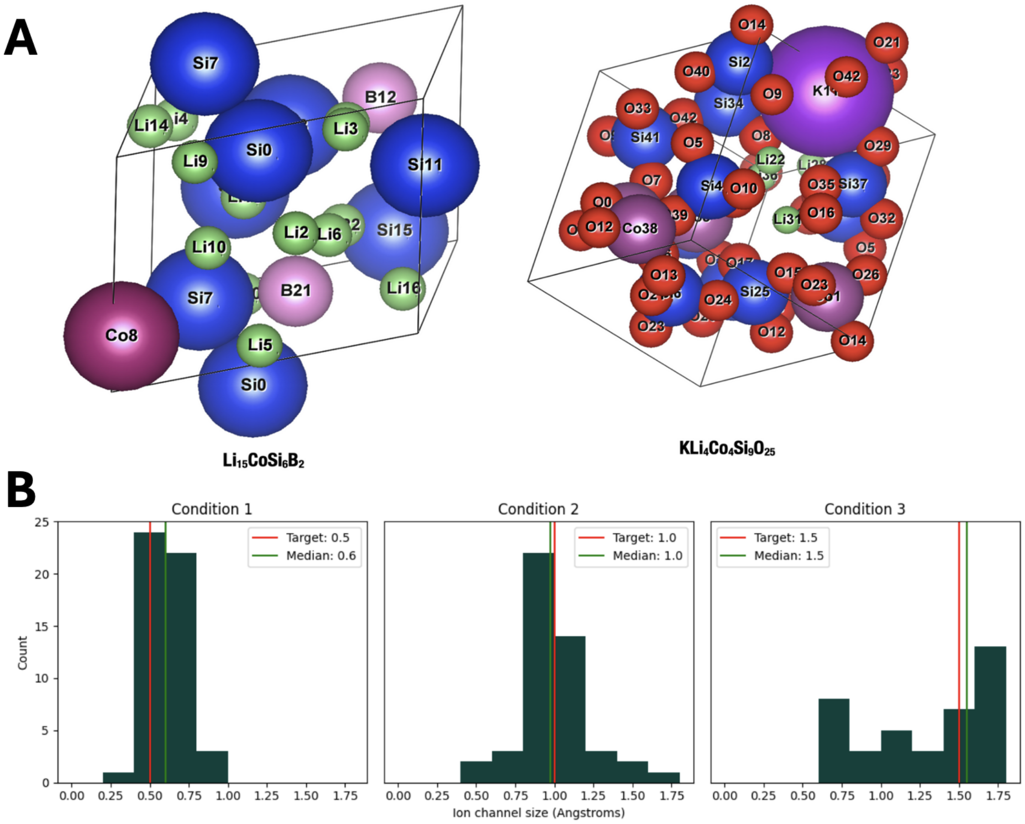

Figure 5: A) Visualization of two novel Si-based alloy anode structures generated by the conditional AI model. The model proposes varying geometries, element counts, and atom numbers without restriction. B) Results from generative processes using ion channel size as the target property. The model received only the target value, and the resulting structures were computationally analyzed to determine their actual channel sizes. Larger channel sizes—less common in the training dataset—proved more challenging, leading to a wider distribution while still keeping the median value close to the target.

Our high-throughput screening of known Si–X compounds revealed that while many exhibit high theoretical capacity, only a small fraction retain stable constant size of the Li^+ migration channels after delithiation. Using CHGNet, a pretrained machine learning interatomic potential, we relaxed hypothetical discharged states and identified robust anode candidates capable of preserving channel geometry, keeping the channel size constant, across cycles.

To expand beyond known chemistries, we implemented a generative model conditioned on desired elements, capacity and channel size properties. The model successfully generated new crystal structures with high capacity and structural resilience, several of which represent previously unexplored compositions with favorable Li-ion transport pathways.

These results demonstrate the potential of AI not only to accelerate the screening of existing materials, but also to generate novel candidates with tailored performance attributes. The integration of database knowledge, fast relaxation tools, and inverse design techniques provides a scalable path toward the discovery of next-generation battery materials. Future work will involve DFT validation and experimental synthesis of top candidates, as well as extension of this framework to other components of electrochemical energy systems.

References

Michel Armand and Jean-Marie Tarascon. Building better batteries. *Nature*, 451:652–657, 2008. doi: 10.1038/451652a.

Maziar Ashuri, Qianran He, and Leon L. Shaw. Silicon as a potential anode material for Li-ion batteries: where size, geometry and structure matter. *Nanoscale*, 8(1):74–103, 2016. doi: 10.1039/C5NR05116A.

Simon Batzner, Albert Musaelian, Lixin Sun, Mario Geiger, Jonathan Mailoa, Mordechai Kornbluth, Nicola Molinari, Tess Smidt, and Boris Kozinsky. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nature Communications*, 13:2453, 2022. doi: 10.1038/s41467-022-29939-5.

Keith T. Butler, Daniel W. Davies, Hugh Cartwright, Olexandr Isayev, and Aron Walsh. Machine learning for molecular and materials science. *Nature*, 559:547–555, 2018. doi: 10.1038/s41586-018-0337-2.

Chi Chen and Shyue Ping Ong. A universal graph deep learning interatomic potential for the periodic table. *Nature Computational Science*, 2:718–728, 2022. doi: 10.1038/s43588-022-00349-3.

Chi Chen, Weike Ye, Yunxing Zuo, Chen Zheng, and Shyue Ping Ong. Graph networks as a universal machine learning framework for molecules and crystals. *Chem. Mater.*, 31(9):3564–3572, 2019. doi: 10.1021/acs.chemmater.9b01294.

Maarten De Jong, Wei Chen, Thomas Angsten, Anubhav Jain, Randy Notestine, Anthony Gamst, Marcel Sluiter, C. Krishna Ande, S. Van der Zwaag, Jose Plata, Cormac Toher, Stefano Curtarolo, Gerbrand Ceder, Kristin Persson, and Mark Asta. Charting the complete elastic properties of inorganic crystalline compounds. *Scientific Data*, 2:150009, 2016. doi: 10.1038/sdata.2015.9.

Bowen Deng, Peichen Zhong, KyuJung Jun, Janosh Riebesell, Christopher J. Bartel, and Gerbrand Ceder. CHGNet: Pretrained universal neural network potential for charge-informed atomistic modeling. *Nature Machine Intelligence*, 5:1031–1041, 2023. doi: 10.1038/s42256-023-00716-3.

Claudia Draxl and Matthias Scheffler. The nomad laboratory: from data sharing to artificial intelligence. *J. Phys. Mater.*, 2(3):036001, 2019. doi: 10.1088/2515-7639/ab13bb.

Saulius Gražulis, Artūras Daškevič, Andrius Merkys, Daniel Chateigner, Luca Lutterotti, Miguel Quiros, Natalia R. Serebryanaya, Peter Moeck, Robert T. Downs, and Armel Le Bail. Crystallography open database (cod): an open-access collection of crystal structures and platform for world-wide collaboration. *Nucleic Acids Research*, 40(D1):D420–D427, 2012. doi: 10.1093/nar/gkr900.

Bing He, Shuting Chi, Anjiang Ye, Penghui Mi, Liwen Zhang, Bowei Pu, Zheyi Zou, Yunbing Ran, Qian Zhao, Da Wang, Wenqing Zhang, Jingtai Zhao, Stefan Adams, Maxim Avdeev, and Siqi Shi. High-throughput screening platform for solid electrolytes combining hierarchical ion-transport prediction algorithms. *Scientific Data*, 7(1):151, 2020. doi: 10.1038/s41597-020-0474-y.

Anubhav Jain, Shyue Ping Ong, Geoffroy Hautier, Wei Chen, William D. Richards, Stephen Dacek, Shreyas Cholia, Daniel Gunter, David Skinner, Gerbrand Ceder, and Kristin A. Persson. Commentary: The materials project: A materials genome approach to accelerating materials innovation. *APL Materials*, 1(1):011002, 2013. doi: 10.1063/1.4812323.

Dipendra Jha, Logan Ward, Abhinav Paul, Wei-keng Liao, Alok Choudhary, Chris Wolverton, and Ankit Agrawal. Elemnet: Deep learning the chemistry of materials from only elemental composition. *Scientific Reports*, 8:17593, 2018. doi: 10.1038/s41598-018-35934-y.

Pavel Karpov, Evgeny V Podryabinkin, and Alexander V Shapeev. A reinforcement learning approach to the inverse problem of design of inorganic crystals. *npj Computational Materials*, 6:119, 2020. doi: 10.1038/s41524-020-00394-2.

Sungwon Kim, Juhwan Noh, Geun Ho Gu, Alán Aspuru-Guzik, and Yousung Jung. Generative adversarial networks for crystal structure prediction. *ACS Central Science*, 6(8):1412–1420, 2020. doi: 10.1021/acscentsci.0c00268.

Scott Kirklin, Bryce Meredig, and Chris Wolverton. High-throughput computational screening of new li-ion battery anode materials. *Advanced Energy Materials*, 3(2):252–262, 2013. doi: 10.1002/aenm.201200593.

Scott Kirklin, James E. Saal, Bryce Meredig, Alexa Thompson, Jeff W. Doak, Muratahan Aykol, Stephan Rühl, and Chris Wolverton. The open quantum materials database (oqmd): assessing the accuracy of dft formation energies. *npj Computational Materials*, 1:15010, 2015. doi: 10.1038/npjcompumats.2015.10.

Dominique Larcher and Jean-Marie Tarascon. Towards greener and more sustainable batteries for electrical energy storage. *Nature Chemistry*, 7:19–29, 2015. doi: 10.1038/nchem.2085.

Hongyi Li, Takitaro Yamaguchi, Shingo Matsumoto, Hiroaki Hoshikawa, Toshiaki Kumagai, Norihiko L. Okamoto, and Tetsu Ichitsubo. Circumventing huge volume strain in alloy anodes of lithium batteries. *Nature Communications*, 11:1584, 2020. doi: 10.1038/s41467-020-15452-0.

Turab Lookman, Prasanna V. Balachandran, Dezhen Xue, and Ruihao Yuan. Active learning in materials science with emphasis on adaptive sampling using uncertainties for targeted design. *npj Computational Materials*, 5:21, 2019. doi: 10.1038/s41524-019-0153-8.

Amil Merchant, Simon Batzner, Samuel S. Schoenholz, Muratahan Aykol, Ekin D. Cubuk, et al. Scaling deep learning for materials discovery. *Nature*, 624:80–85, 2023. doi: 10.1038/s41586-023-06735-9.

Yanfei Mi, Jiajie Liang, Xiaolong Zhou, Linlin Hou, Aiguo Pan, Zehui Jiang, Miaomiao Xu, and Gaoshao Cao. Recent advances in silicon-based anode materials for lithium-ion batteries. *Materials Chemistry Frontiers*, 1(8):1691–1708, 2013. doi: 10.1039/C7QM00156K.

National Science and Technology Council. Materials genome initiative for global competitiveness. Technical report, Executive Office of the President, Washington, DC, 2011.

Junwoong Noh, Seongok Kim, Junghwan Yoon, Hyesu Lyu, Young Kim, Seunghwa Seo, and Jinwoo Kim. Inverse design of solid-state materials via a continuous representation. *Matter*, 2(4):865–877, 2020. doi: 10.1016/j.matt.2020.01.027.

Mark N. Obrovac and Linda Christensen. Structural changes in silicon anodes during lithium insertion/extraction. *Electrochemical and Solid-State Letters*, 7(5):A93–A96, 2004. doi: 10.1149/1.1649993.

Teerachote Pakornchote, Natthaphon Choomphon-anomakhun, Sorrjit Arrerut, Chayanon Atthapak, Sakarn Khamkaeo, and Thiti Bovornratanaraks. Diffusion probabilistic models enhance variational autoencoder for crystal structure generative modeling. *Scientific Reports*, 14:1275, 2024. doi: 10.1038/s41598-024-51400-4.

Janosh Riebesell, Rhys E. A. Goodall, Philipp Benner, Yuan Chiang, Bowen Deng, Mark Asta, Gerbrand Ceder, Alpha A. Lee, Anubhav Jain, and Kristin A. Persson. Matbench discovery — a framework to evaluate machine learning crystal stability predictions. *arXiv*, August 2023. doi: 10.1038/s42256-025-01055-1.

James E. Saal, Scott Kirklin, Muratahan Aykol, Bryce Meredig, and Chris Wolverton. Materials design and discovery with high-throughput density functional theory: The open quantum materials database (oqmd). *JOM*, 65(11):1501–1509, 2013. doi: 10.1007/s11837-013-0755-4.

Jonathan Schmidt, Miguel R. G. Marques, Silvana Botti, and Malte Marques. Recent advances and applications of machine learning in solid-state materials science. *npj Computational Materials*, 5:83, 2019. doi: 10.1038/s41524-019-0221-0.

Austin D. Sendek, Qianxun Yang, Ekin D. Cubuk, Ksenia-Amina N. Duerloo, Yi Cui, and Evan J. Reed. Machine learning-assisted discovery of solid li-ion conducting materials. *Chem. Mater.*, 30(15):5755–5762, 2018. doi: 10.1021/acs.chemmater.8b01963.

Jean-Marie Tarascon and Michel Armand. Issues and challenges facing rechargeable lithium batteries. *Nature*, 414:359–367, 2001. doi: 10.1038/35104644.

Susumu Toki and Mark N. Obrovac. Design strategies for high-performance silicon anodes in lithium-ion batteries. *Advanced Energy Materials*, 14(4):2302056, 2024. doi: 10.1002/aenm.202302056.

Logan Ward, Ankit Agrawal, Alok Choudhary, and Chris Wolverton. A general-purpose machine learning framework for predicting properties of inorganic materials. *npj Computational Materials*, 2:16028, 2016. doi: 10.1038/npjcompumats.2016.28.

Tian Xie and Jeffrey C. Grossman. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Phys. Rev. Lett.*, 120:145301, 2018. doi: 10.1103/PhysRevLett.120.145301.

Tian Xie, Xiang Fu, Octavian Ganea, Regina Barzilay, and Tommi Jaakkola. Crystal diffusion variational autoencoder for periodic material generation. In *International Conference on Learning Representations (ICLR)*, 2022.

Cai-Yuan Ye, Hong-Ming Weng, and Quan-Sheng Wu. Con-cdvae: A method for the conditional generation of crystal structures. *Computational Materials Today*, 1:100003, May 2024. ISSN 2950-4635. doi: 10.1016/j.commt.2024.100003. URL `http://dx.doi.org/10.1016/j.commt.2024.100003`.

Karim Zaghib, Alain Mauger, and Claude M. Julien. Advanced electrode materials for lithium-ion batteries. *Journal of Power Sources*, 196(8):3949–3954, 2010. doi: 10.1016/j.jpowsour.2010.11.118.

Xuejiao Zuo, Jia Zhu, Peter Müller-Buschbaum, and Yingsheng Cheng. Silicon based lithium-ion battery anodes: A chronicle perspective review. *Nano Energy*, 31:113–143, 2017. doi: 10.1016/j.nanoen.2016.11.013.