



Cite this: DOI: 10.1039/d6sc00030d

All publication charges for this article have been paid for by the Royal Society of Chemistry

# Density-dependent sodium-storage mechanisms in hard carbon materials

Alexis Front, \*<sup>ab</sup> Tapio Ala-Nissilä <sup>bc</sup> and Miguel A. Caro <sup>a</sup>

Understanding the sodium-storage mechanism in hard carbon (HC) anodes is crucial for advancing sodium-ion battery (SIB) technology. However, the intrinsic complexity of HC microstructures and their interactions with sodium remain not fully elucidated. We present a multiscale methodology that integrates grand-canonical Monte Carlo (GCMC) simulations with a machine-learning interatomic potential based on the Gaussian approximation potential (GAP) framework to investigate sodium insertion mechanisms in hard carbons with different levels of porosity, achieved by simulating structural models with densities ranging from 0.7 to 1.9 g cm<sup>-3</sup>. Structural and thermodynamic analyses reveal the interplay between pore size and accessibility and the relative contributions of adsorption, intercalation, and pore filling to the overall storage capacity. Low-density carbons favor pore-filling, achieving extremely high capacities at near-zero voltages, whereas high-density carbons primarily store sodium through adsorption and intercalation, leading to lower but more stable capacities. Intermediate-density carbons (1.3–1.6 g cm<sup>-3</sup>) provide the most balanced performance, combining moderate capacity (480 and 310 mAh g<sup>-1</sup>), safe operating voltages, and minimal volume expansion (<10%). These findings establish a direct correlation between carbon density and electrochemical behavior, providing atomic-scale insight into how hard carbon morphology governs sodium storage. The proposed framework offers a rational design principle for optimizing HC-based SIB anodes toward high energy density and long-term cycling stability.

Received 3rd January 2026  
Accepted 17th March 2026

DOI: 10.1039/d6sc00030d

rsc.li/chemical-science

## 1. Introduction

The growing demand for renewable energy-storage solutions has exposed the limitations of conventional lithium-ion batteries (LIBs), particularly due to the scarcity of lithium in the Earth's crust<sup>1</sup> and the challenges associated with its recycling. These constraints hinder the large-scale deployment of LIBs, motivating the search for alternative technologies. Sodium-ion batteries have emerged<sup>2</sup> as a promising candidate in this regard, owing to the natural abundance and low cost of sodium, as well as their favorable cycling performance.<sup>3</sup> Such features make sodium-ion batteries (SIBs) particularly suitable for grid-scale<sup>4</sup> and large-scale energy-storage applications.<sup>5</sup>

Graphitic carbon, the benchmark anode material for LIBs, demonstrates good compatibility with lithium ions and limited volume changes.<sup>6</sup> However, its performance in SIBs is limited by the larger ionic radius of sodium, which induces significant stress and thermodynamic instability during cycling.<sup>5,7</sup> Among

alternative anode materials, hard carbon (HC) has attracted considerable attention due to its intrinsic structural stability and favorable electrochemical properties.<sup>8</sup> Despite its widespread application, the sodium-storage mechanism in HC remains not fully understood. This complexity arises from the heterogeneous nature of HC, which comprises defects, nanoporous domains, and turbostratic graphene layers.<sup>9</sup> These structural features strongly influence sodium-ion storage and ultimately determine the electrochemical performance of SIBs.

Sodium storage in HC proceeds through three primary mechanisms – adsorption, intercalation, and pore filling – which manifest as distinct features in the charge voltage profile.<sup>10</sup> In the high-voltage region (typically above 0.2–0.3 V vs. Na<sup>+</sup>/Na), sodium ions are predominantly adsorbed at defect sites or surface functional groups within the HC framework.<sup>11</sup> This adsorption process leads to a rapid decrease in voltage as the limited number of high-energy sites become occupied. As the sodium content increases, the system transitions to the intercalation stage, in which sodium ions insert into the voids between graphene-like layers.<sup>12</sup> This mechanism contributes to the sloping region of the voltage curve and is highly sensitive to interlayer spacing and the degree of structural disorder. At higher levels of sodium insertion, the pore-filling mechanism dominates. Sodium ions aggregate within larger nanopores and voids, forming clusters that give rise to the characteristic low-

<sup>a</sup>Department of Chemistry and Materials Science, Aalto University, Kemistintie 1, 02150 Espoo, Finland. E-mail: alexis.front@aalto.fi

<sup>b</sup>MSP Group, Department of Applied Physics, Aalto University, FI-00076 Aalto, Espoo, Finland

<sup>c</sup>Interdisciplinary Centre for Mathematical Modelling and Department of Mathematical Sciences, Loughborough University, Loughborough, Leicestershire LE11 3TU, UK



voltage plateau near zero voltage.<sup>13</sup> This stage accounts for a significant portion of the overall capacity, although it may introduce structural stress and influence long-term stability. Recently, Jian *et al.*<sup>14</sup> proposed a new interpretation of the voltage plateau origin by tailoring the pore architecture in HCs. Their study demonstrated a linear correlation between nanopore volume and the experimentally measured plateau capacity, highlighting the critical role of porosity in governing sodium-storage behavior.

Early theoretical studies of sodium storage in hard carbon primarily emphasized adsorption mechanisms, proposing that sodium ions preferentially bind to surface sites or structural defects.<sup>15</sup> Subsequent investigations highlighted the role of enlarged interlayer spacing and structural disorder in facilitating sodium intercalation by weakening interlayer van der Waals interactions.<sup>16</sup> At low potentials, a pore-filling mechanism has been proposed, in which sodium ions aggregate into quasi-metallic clusters within nanopores.<sup>17</sup> These studies form the basis of the widely accepted three-stage sodium-storage model in HC. However, most computational investigations of these mechanisms have relied on empirical interatomic potentials,<sup>11,18</sup> which provide a limited and often inaccurate description of Na–C interactions. The limited fidelity of such potentials hampers their ability to capture the diverse chemical environments present in hard carbon,<sup>19</sup> particularly under conditions of high sodium loading. Furthermore, these models typically employ simplified structural representations that neglect essential microstructural features, including variations in mass density, pore-size distribution, defect concentration, and the presence of turbostratic graphene domains. As a result, fundamental questions remain unresolved, particularly regarding the precise role of microstructural characteristics in governing adsorption, intercalation, and filling processes, as well as the interplay between these mechanisms. Addressing these limitations requires simulation methodologies capable of accurately describing Na–C interactions while explicitly accounting for the inherent structural heterogeneity of HC.

In this work, we develop a machine-learning-based computational framework with density functional theory (DFT) accuracy to investigate sodium storage in hard carbon. Grand-canonical Monte Carlo (GCMC) simulations are performed using a machine-learning interatomic potential constructed within the Gaussian approximation potential (GAP) framework. This approach enables an accurate and efficient description of Na–C interactions across heterogeneous structural environments. Using this framework, we systematically examine the influence of HC porosity on sodium adsorption, intercalation, and pore filling. In particular, we identify density regimes that enable favorable sodium storage while limiting structural deformation and volume expansion. By establishing quantitative links between hard-carbon microstructure and sodium-storage behavior, this work provides methodological insights relevant to the rational design of HC anodes for sodium-ion batteries and, more broadly, to the modeling of ion storage in disordered electrode materials.

## 2. Methodology

To overcome the limitations of both density functional theory (DFT) and empirical interatomic potentials, there has been a recent paradigm shift to construct computationally efficient potentials by machine learning (ML) from DFT reference data. These ML methods perform high-dimensional fits to the DFT potential energy surface (PES) using a limited set of carefully selected configurations, and subsequently interpolate energies and forces for unseen structures. Unlike empirical potentials, ML models do not typically assume a fixed functional form; instead, they adapt flexibly to the training data. This reduces bias during construction but requires careful fitting and validation to prevent unphysical extrapolations. Among the different ML interatomic potential (MLIP) frameworks, GAP<sup>20</sup> is a kernel-based approach founded on Gaussian process regression. In this framework, the PES is regressed using fitting coefficients obtained during training in combination with kernel functions, which are evaluated on the fly. A GAP prediction is made by comparing the atomic descriptor of a current structure to a subset of structures in the database. Each comparison yields a kernel, a measure of similarity, bounded between 0 (two structures are completely different) and 1 (identical). Different descriptors of the atomic structure can be used to describe the atomic environments. The predicted local energy of an atom  $i$  is then expressed for a combination of two-body (2b) and many-body (mb) descriptors as

$$\varepsilon_i = e_0 + (\delta^{(2b)})^2 \sum_s \alpha_s^{(2b)} k^{(2b)}(i, s) + (\delta^{(mb)})^2 \sum_s \alpha_s^{(mb)} k^{(mb)}(i, s) + \text{core} \quad (1)$$

where  $k(i, s)$  is the kernel between the atomic environment  $i$  and the different atomic environment in the sparse set (a subset of structures in the training database), the  $\alpha_s$  are fitting coefficients obtained during training,  $e_0$  is a constant energy per atom,  $\delta$  gives the energy scale of the model, and core refers to a “core potential”.

The choice of descriptors is central to an accurate and data-efficient PES representation. We employ a combination of 2b and mb atomic descriptors, in addition to a tabulated splined “core potential” to describe the interaction between two atoms when they are nearby ( $\leq 1$  Å). The 2b descriptors, constructed with a 5.5 Å cutoff for C–Na and Na–Na interactions, improve the numerical stability of the GAP. To reproduce the underlying DFT PES with higher fidelity, we include many-body descriptors based on the smooth overlap of atomic positions (SOAP) formalism.<sup>21</sup> In this work, we use a more data-efficient implementation of the SOAP descriptor.<sup>22</sup>

### 2.1. Machine-learning interatomic potentials

**2.1.1  $\Delta$ -Learning.** Modeling the potential energy surface of the Na–C system is best done by splitting its representation into two components. A high-quality GAP model for carbon is already available,<sup>23</sup> and it accurately captures the strong (short-range) covalent interactions and weak (long-range) van der



Waals interactions in pure carbon as separate terms in the potential. By comparison, interactions between Na and C are simultaneously short-range and weak. Therefore, rather than constructing a full C–Na binary potential from scratch, we adopt a  $\Delta$ -learning approach:<sup>24</sup> we fit a machine-learning model only to the *energy differences* induced by Na insertion into carbon matrices. The intercalation energy is defined as

$$\Delta E_{\text{ref}}(\text{CNa}) = E_{\text{ref}}(\text{CNa}) - E_{\text{ref}}(\text{C}) \quad (2)$$

where  $E_{\text{ref}}(\text{CNa})$  is the energy of the whole structure (sodium atoms in a carbon structure) and  $E_{\text{ref}}(\text{C})$  is the energy of the same structure after removing the sodium atoms. The subscript “ref” refers to the fact that these are the reference energies used to train the GAP ML model. The total system energy is then expressed as a sum of the reference carbon GAP and the CNa  $\Delta$ -GAP as:

$$E_{\text{GAP}}(\text{CNa}) = E_{\text{GAP}}(\text{C}) + \Delta E_{\text{GAP}}(\text{CNa}) \quad (3)$$

Note that, because of the considerations discussed above, fitting  $E_{\text{GAP}}(\text{C})$  and  $\Delta E_{\text{GAP}}(\text{CNa})$  separately from  $E_{\text{ref}}(\text{C})$  and  $\Delta E_{\text{ref}}(\text{CNa})$ , respectively, and then adding them up is a better approximation to  $E_{\text{ref}}(\text{CNa})$  than fitting  $E_{\text{GAP}}(\text{CNa})$  from  $E_{\text{ref}}(\text{CNa})$  directly. The  $\Delta$  term is also able to accurately model metallic Na, which becomes relevant when intercalation takes place within nanopores larger than  $\approx 1$  nm in diameter.

To improve accuracy in the short-range repulsive regime (interatomic distances below 2 Å), tabulated pairwise core potentials for C–Na and Na–Na interactions are included. This explicit inclusion significantly improves the stability and accuracy of the GAP fit. In practice, the tabulated per-pair core interaction is removed from the DFT reference values prior to training, thereby smoothing the PES and facilitating the GAP fit. At prediction time, the tabulated interactions are reintroduced, ensuring both physical stability and fidelity to the DFT reference.

**2.1.2 Training database and model fitting.** Initial training data were generated by randomly placing Na atoms (up to 10) in all possible  $\text{C}_{60}$  isomers,<sup>25</sup> defected graphite and nanoporous structures. The latter were generated by a melt-graphitization-quench protocol and using the GAP model.<sup>23</sup> This approach ensures sampling of all the relevant atomic environments that Na ions might encounter in graphitic carbon. For the extended systems, we used a simulation box of 216 atoms to recreate a vast variety of possible defects in graphite and to obtain bended graphene layers and small nanopores. The whole database was computed at the DFT level of theory with the PBE exchange-correlation functional<sup>26</sup> using VASP.<sup>27,28</sup> Single-point DFT energies were obtained using a cutoff energy of 650 eV and one  $k$  point. Four valence electrons were explicitly treated for C ( $2s^2 2p^2$ ) and seven for Na ( $2p^6 3s^1$ ). All the calculations used exactly the same convergence parameters to avoid introducing additional noise in the data.

To refine the model accuracy, we carried out an iterative-training approach. Starting from an initial GAP, we used short MD runs to generate 100 structures at each iteration: 25

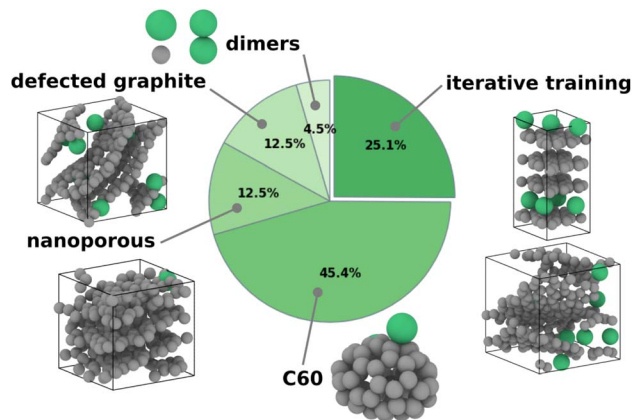


Fig. 1 Overview of the structures in the database. The initial database contains dimers, defected graphite, nanoporous carbon and  $\text{C}_{60}$  structures. At each iteration, graphite and nanoporous carbon structures are added to the database to improve the accuracy of the model. Carbon and sodium atoms are in grey and green, respectively.

graphite, 25 nanoporous carbon (NP-C) with a density of  $1.275 \text{ g cm}^{-3}$ , 25 NP-C with a density of  $1.645 \text{ g cm}^{-3}$ , and 25 amorphous carbon (a-C) with a density of  $2.48 \text{ g cm}^{-3}$ , with 1–25 sodium insertions per structure. We then ran PBE-DFT single-point calculations and added these structures to the growing training database. This procedure was repeated several times to improve the accuracy of the  $\Delta$ -GAP model, reaching optimal performance after the tenth cycle, with a root mean square deviation of  $5 \text{ meV at}^{-1}$  (Fig. S1). The GAP is freely available from Zenodo for further reuse.<sup>29</sup> The composition of the final training dataset, detailing the proportion of each structure type, is summarized in Fig. 1.

All the fits were carried out with the `gap_fit` program,<sup>30</sup> part of the QUIP software package.<sup>31</sup> Atomic structure generation, manipulation, visualization and MD simulations were done with the Atomic Simulation Environment<sup>32</sup> (ASE), different in-house codes, `ovito`<sup>33</sup> and the TurboGAP program.<sup>34</sup>

## 2.2. Hard carbon generation

We used the carbon GAP of Muhli *et al.*,<sup>23</sup> which integrates van der Waals corrections, to perform molecular dynamics (MD) using the TurboGAP code.<sup>34</sup> First, the system was equilibrated in a randomized liquid state at 9000 K for 10 ps. Next, the liquid was quenched with a linear temperature profile over 6 ps down to 3500 K. The graphitization process consist of annealing stages at 3500 K for 400 ps. Finally, the HC structures were quenched down to 300 K over 20 ps. Temperature control was achieved using the velocity-rescaling Bussi thermostat<sup>38</sup> with a time constant of 100 fs. All MD simulations were performed under periodic boundary conditions with a box containing 1728 atoms, using a Verlet integration time step of 1 fs. To explore the effect of density on structural and electrochemical properties, we generated five target densities: 0.7, 1.0, 1.3, 1.6 and  $1.9 \text{ g cm}^{-3}$  corresponding to cubic box lengths of 36.63, 32.52, 29.82, 27.81, 26.28 Å respectively. In this work, the structural density is defined as the total mass of carbon atoms divided by



the simulation cell volume after structural relaxation. Recognizing the stochastic nature of the melt-quench protocol and the inherent heterogeneity of HC, we generated ten independent structures per density using different random seeds. Structural properties reported in this work correspond to the mean values across these ensembles, with standard deviations capturing variability. These structures are available on Zenodo<sup>39</sup> and some examples are shown in the top row of Fig. 2.

**2.2.1 Structural characterization.** The morphology of the quenched HC samples reveals that most of the carbon forms curved graphene-like fragments, predominantly with  $sp^2$  bonding, as quantified in Table S1. These fragments assemble into three-dimensional networks, where edges containing  $sp$  motifs and basal planes can be interlinked *via*  $sp^3$ -hybridized carbon atoms. To further analyze the structural topology of the graphitic planes, we computed ring-size and bond-angle distributions.<sup>40</sup> Here, ring size denotes the number of carbon atoms forming a closed covalent loop within the  $sp^2$  network. These quantities serve as sensitive metrics for local order and curvature in disordered carbon networks. The analysis was performed using matscopy<sup>41</sup> for all five HC densities. As shown in Fig. S2a, the graphene-like layers in all samples are primarily composed of 5-, 6-, and 7-membered carbon rings. All densities display similar ring-size distributions characterized by a dominant peak at 6-membered rings, indicative of well-ordered graphitic domains, along with approximately equal populations of 5- and 7-membered rings. We also note that the small standard deviation confirms the robustness of the simulation protocol. Complementary information is provided by the bond-angle distributions. All samples feature a sharp peak at  $120^\circ$ , a hallmark of regular hexagonal carbon rings in graphitic structures along with an inflection near  $108^\circ$ . These pentagonal motifs, as heptagons, play an important role in curving

graphene fragments in HC and glassy carbons.<sup>42</sup> Neither the ring-size nor the bond-angle distributions show a significant dependence on density. To compare the simulated HC structures with experimental data, we computed diffraction patterns (Fig. 2a) and reduced pair distribution functions (PDF) (Fig. 2b), which serve as metrics for assessing structural similarity. We chose experimental glassy sample HTW2500 reported in ref. 36 and a HC carbon sample reported in ref. 37 as references. The reciprocal-space diffraction scattering was calculated using the Debye scattering equation as implemented in the Debyer code.<sup>43</sup> The reduced PDF commonly used in experiment is related to the PDF  $g(r)$  most often reported in simulation work by  $G(r) = 4\pi\rho(g(r) - 1)$ , where  $\rho$  is the atomic density. The HTW2500 sample exhibits three main diffraction peaks at  $2\theta \approx 24, 44$  and  $80^\circ$ , corresponding to the graphite  $\{002\}$ ,  $\{100\}$ , and  $\{110\}$  reflections. For the lower-density model structures, the  $\{002\}$  reflection is absent, owing to reduced and irregular layer stacking resulting from a more open and diffuse arrangement of graphene sheets. The interlayer spacing can be estimated using Bragg's law. The systematic shift toward smaller scattering angles in Fig. 2a with decreasing density clearly indicates an increase in interlayer spacing. Among the simulated structures, only the HC sample with a density of  $1.9 \text{ g cm}^{-3}$  exhibits a well-defined  $\{002\}$  reflection, although the onset of this peak is already apparent for the  $1.6 \text{ g cm}^{-3}$  model. From the position of this reflection, an average interlayer spacing of approximately  $3.8 \text{ \AA}$  can be inferred. In contrast, no significant shift is observed in the in-plane  $\{100\}$  reflection, indicating an in-plane lattice parameter  $a$  of around  $2.43 \text{ \AA}$ .

**2.2.2 Pore-size distributions.** Porosity is a key structural descriptor of hard carbon, as it governs sodium adsorption, interlayer insertion, and pore-filling storage. Therefore, characterizing porosity is a central focus of this work. The geometric

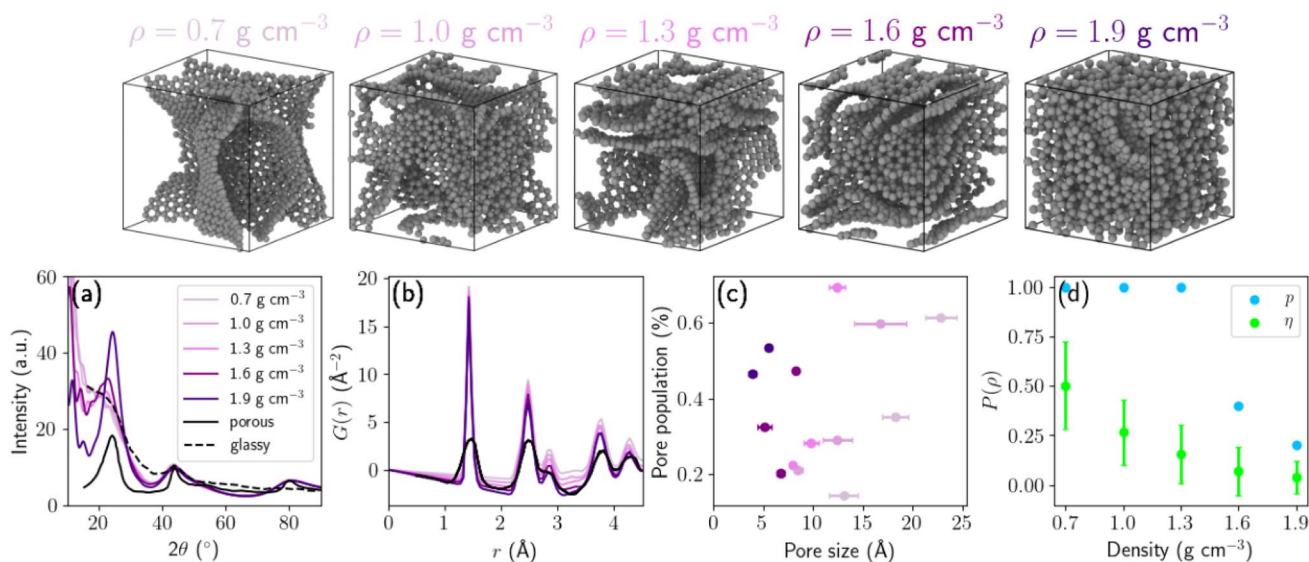


Fig. 2 3D snapshots of five hard carbon samples of different densities. (a) Diffraction patterns, (b) reduced partial distribution functions  $G(r)$ , (c) pore-size populations, and (d) percolation probability (in blue) and geometric transport (in green). Standard deviations are represented with horizontal (c) and vertical (d) bars. Experimental results are extracted from ref. 35 and 36 for the HTW2500 sample (glassy) and from ref. 37 for a HC sample (porous).



porosity of a structure is defined as the fraction of the simulation cell volume not occupied by carbon atoms. Because sodium storage is governed by ion accessibility rather than purely geometric void space, we report here the Na<sup>+</sup>-accessible porosity, defined as the accessible volume fraction  $\epsilon_{\text{access}} = V_{\text{access}}/V_{\text{cell}}$ . A three-dimensional Voronoi network<sup>44</sup> was constructed to identify the void space within the structures, based on Voronoi decomposition as implemented in the Voro++ library.<sup>45</sup> This approach enables a geometric partitioning of the simulation cell into accessible and inaccessible regions based on atomic coordinates. Monte Carlo sampling was performed to determine the Na<sup>+</sup>-accessible pore volume  $V_{\text{access}}$  using a probe radius of 1.16 Å (ionic radius of Na<sup>+</sup>). A total of 130 000 sampling points were employed to ensure statistical convergence of the accessible volume estimate. The pore-size distribution (PSD), calculated using the zeo++ code,<sup>46</sup> is obtained from the distribution of largest included sphere diameters (LCD, denoted in zeo++ as  $D_i$ ), which represent the largest sphere that can be inscribed without overlapping carbon atoms. For consistency, radii are reported as  $R_{\text{LCD}} = D_i/2$ . From these distributions, we extracted the mean pore radius and pore population for each density. The pore population at a given diameter is defined as the normalized frequency of LCD values within that size interval. For densities below 1.6 g cm<sup>-3</sup>, all samples exhibit trimodal or bimodal PSDs, reflecting the coexistence of small interlayer voids and larger pore cavities. In contrast, at a density of 1.9 g cm<sup>-3</sup>, only one third of the samples display a bimodal distribution, while the remaining structures exhibit negligible Na<sup>+</sup>-accessible porosity. Fig. S2a presents the PSDs for each structure used in the sodium insertion simulations. Fig. 2c describes the distribution of pore diameters  $D$  and the corresponding pore populations. With decreasing density, the PSD progressively shifts toward larger pore diameters, and larger pores constitute the dominant population at each density. Conversely, increasing density leads to narrower PSDs with reduced standard deviation, reflecting structural homogenization and collapse of larger cavities. This behavior is consistent with decreasing interplanar spacing between graphitic sheets during densification. As a result, low-density HC structures contain a higher fraction of large cavities capable of accommodating multiple sodium ions, whereas high-density structures are increasingly dominated by small voids and surface adsorption environments. Because PSD analysis alone does not explicitly account for percolation or connectivity, some voids identified geometrically may be isolated. Therefore, PSDs characterize structural pore statistics rather than sodium-accessible transport pathways.

To explicitly evaluate sodium transport accessibility, channel analysis was performed using zeo++ with a Na<sup>+</sup>-sized probe. The pore-limiting diameter (PLD, denoted in zeo++ as  $D_f$ ) corresponds to the diameter of the largest sphere that can traverse the structure without overlap, and therefore defines the narrowest bottleneck along percolating pathways. Radii are reported as  $R_{\text{PLD}} = D_f/2$ . For all structures up to 1.3 g cm<sup>-3</sup>, zeo++ channel analysis indicates continuous percolating Na<sup>+</sup>-accessible networks (one-, two- or three-dimensional connectivity), demonstrating that macroscopic ion transport pathways

are preserved across the low-to-intermediate density range. To quantify the combined effects of accessible porosity and bottleneck restriction, we define a geometric transport factor

$$\eta = \epsilon_{\text{access}} \left( \frac{R_{\text{PLD}}}{R_{\text{LCD}}} \right)^2 \quad (4)$$

where  $R_{\text{LCD}}$  characterizes cavity size and  $R_{\text{PLD}}$  characterizes transport constrictions. The ratio  $R_{\text{PLD}}/R_{\text{LCD}}$  provides a dimensionless measure of bottleneck severity, and the squared term approximates the scaling of transport cross-sectional area. Thus,  $\eta$  represents the fraction of pore volume that is both geometrically accessible and not strongly limited by constrictions. Fig. 2d displays percolation probability and the geometric transport factor. As density increases,  $\epsilon_{\text{access}}$  decreases and  $R_{\text{PLD}}/R_{\text{LCD}}$  progressively declines, resulting in a monotonic reduction of  $\eta$ . Importantly, because percolation is preserved up to 1.3 g cm<sup>-3</sup>, this decrease reflects structural densification and bottleneck narrowing. Because Na<sup>+</sup>-accessible pore networks remain percolating up to 1.3 g cm<sup>-3</sup> and pore-limiting diameters exceed the Na<sup>+</sup> ionic radius, the geometric analysis indicates that long-range ion transport is not connectivity-limited within this density range. A more thorough investigation would require detailed diffusion kinetics, which is beyond the scope of the present work.

### 2.3. Hybrid Monte Carlo/molecular dynamics atomistic simulations

To investigate sodium-storage mechanisms in HC, we performed hybrid MC/MD simulations in the grand-canonical ensemble. In this framework, the system is coupled to a heat bath at temperature  $T$  and can exchange particles with an infinite reservoir at chemical potential  $\mu$ . Within the Metropolis algorithm, starting from an initial pure HC structure, trial configurations are generated by one of the following moves: (i) random displacement of an atom, (ii) insertion or removal of a sodium atom at a random position, (iii) molecular dynamics steps. The inclusion of MD moves accelerates equilibration by allowing both atomic relaxations and volume adjustments. Trial configurations are then accepted or rejected according to the Metropolis acceptance probabilities for particle displacement, insertion, or removal:

$$P_{\text{acc}}^{\text{dis}} = \min[1, \exp(-\beta\Delta E)] \quad (5)$$

for atomic displacements, and

$$P_{\text{acc}}^{\text{ins}} = \min \left[ 1, \frac{V}{(N+1)\lambda^3} \exp(-\beta(\Delta E - \mu)) \right] \quad (6)$$

and  $P_{\text{acc}}^{\text{rem}} = \min \left[ 1, \frac{N\lambda^3}{V} \exp(-\beta(\Delta E + \mu)) \right]$

for sodium insertion and removal, respectively.  $\Delta E$  is the energy difference between the trial and current configurations,  $\beta = 1/(k_{\text{B}}T)$ ,  $V$  is the system's volume,  $N$  is the number of sodium atoms before the move, and  $\lambda$  is the de Broglie thermal wavelength. A trial configuration is accepted if the corresponding acceptance probability is greater than a random number  $r \in [0, 1]$ . Sodium insertion was simulated across the full volume of



each HC structure, without explicit consideration of percolation pathways. Consequently, some sodium ions may occupy isolated voids that would be inaccessible in a real electrochemical environment. Additionally, the GCMC method focuses on thermodynamically accessible configurations and does not incorporate kinetic limitations, electrolyte transport constraints, or surface passivation effects—factors that can (and do) influence real electrode performance. As a result, the absolute storage capacities reported here may represent an upper bound of experimentally achievable values. However, because all densities are evaluated under identical thermodynamic conditions, the relative trends and mechanistic transitions identified as a function of structural density remain robust.

We ran MC/MD simulations for 100 000 steps for each structure for chemical potentials of  $-1.00$ ,  $-1.24$ ,  $-1.50$  eV at 300 K. The chemical potential of  $-1.24$  eV corresponds to the bcc bulk sodium cohesive energy. Fig. 3 shows the evolution of sodium concentration and dimensional change, defined as the relative change in the volume of the hard carbon volume after inserting sodium ions. All simulations reached the equilibrium after 50 000 MC steps, except for the low density structure at the highest chemical potential (*i.e.*,  $-1.00$  eV). The Monte Carlo simulations demonstrate a strong dependence of sodium storage on the structural density of nanoporous carbon. At low density ( $0.7$  g cm $^{-3}$ ), the framework accommodates the highest sodium concentration, reflecting the abundance of adsorption sites and the ease of pore filling. As the density increases to  $1.6$  g cm $^{-3}$ , the equilibrium concentration decreases significantly, suggesting that pore accessibility and connectivity become restricted. At the highest density investigated ( $1.9$  g cm $^{-3}$ ), sodium uptake is limited to 15%, which points to intercalation into a reduced number of energetically favorable sites rather than extended adsorption. This density-dependent trend is pronounced at  $\mu = -1.00$  and  $\mu = -1.24$  eV. In contrast, for  $\mu = -1.50$  eV, sodium uptake is only weakly affected by density, indicating that in this regime insertion is

governed primarily by the strong chemical driving force rather than by structural accessibility.

The dimensional change of the carbon framework follows a similar trend. At low density ( $\rho = 0.7$  g cm $^{-3}$ ) and high chemical potential, sodium insertion induces extreme swelling (up to 150%), directly correlated with the large Na concentration. In this case, drastic structural rearrangements occur, including the breaking of interstitial sp $^3$  bonds to accommodate volume expansion. At  $\mu = -1.24$  eV, the dimensional change is much lower (13%), suggesting a less disruptive insertion mechanism. In the intermediate density range ( $1.0$ – $1.6$  g cm $^{-3}$ ) expansion remains moderate (9 and 6% for  $-1.00$  and  $-1.24$  eV respectively), pointing to a mixed regime where adsorption still contributes but is increasingly restricted by pore confinement. At high density ( $1.9$  g cm $^{-3}$ ), dimensional changes strongly increase (10–28%), which can be attributed to the dominance of graphitic domains that undergo interlayer separation upon intercalation.<sup>47</sup> This resonates with the familiar result that Na ions do not intercalate in graphite and graphite-like materials that are able to accommodate the smaller Li ions. Consistent with the Na uptake results, the dimensional change at  $\mu = -1.50$  eV remains largely independent of density for structures containing pores, underscoring the overriding influence of the chemical potential in this regime, in good agreement with an experimental dimensional change around 2%<sup>48,49</sup> for a similar sodium concentration. The final sodiated structures of each simulation are available from Zenodo.<sup>39</sup>

### 3. Results and discussion

Formation energy and voltage curves are essential descriptors for understanding sodium-storage mechanisms in HC anodes. The formation energy provides direct insight into the thermodynamics of sodium insertion. By quantifying the energy change associated with placing sodium atoms into different sites of the carbon framework, it reveals the relative stability of adsorption at defects, intercalation between graphene layers,

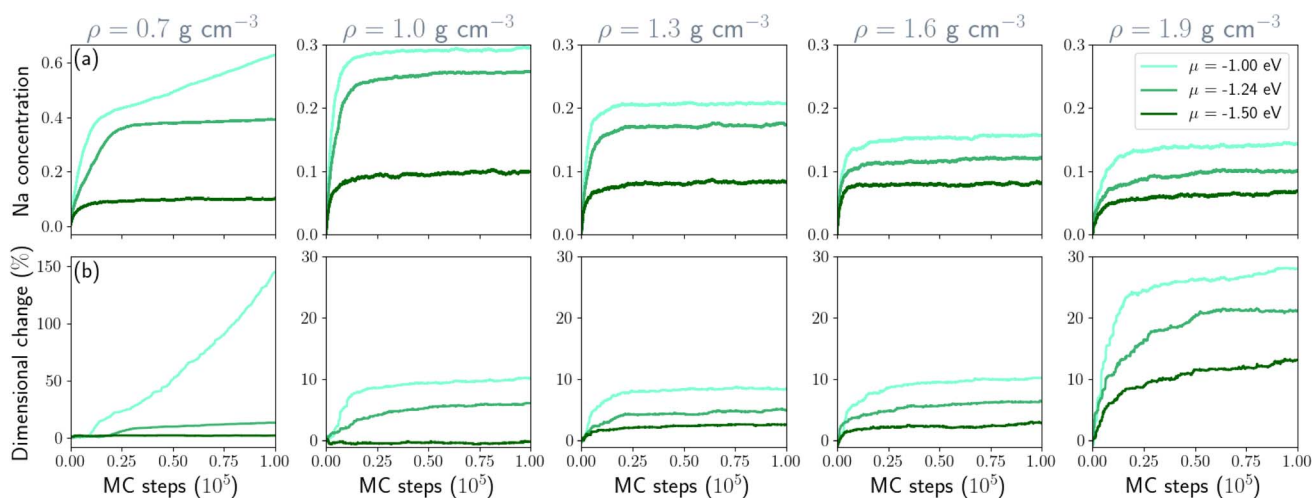


Fig. 3 (a) Sodium concentration and (b) dimensional change as a function of MC steps for the five hard carbon samples at 300 K. Each green line corresponds to a given chemical potential. Note the different vertical scales of the  $\rho = 0.7$  g cm $^{-3}$  sample as compared to the other ones.



and filling of nanopores. These values help identify which storage mechanisms are energetically favored at different sodium concentrations. In contrast, the voltage curve reflects the electrochemical response of the material during sodium insertion and extraction. Derived from the slope of the formation energy with respect to sodium content, it translates atomistic energetics into an experimentally measurable quantity: the electrode potential *versus*  $\text{Na}^+/\text{Na}$ , expressed as

$$V = -\frac{1}{e} \frac{E_f(x_j) - E_f(x_i)}{x_j - x_i} \quad (7)$$

where  $e$  is the elementary charge of an electron,  $x_i$  and  $x_j$  are the sodium ions number of a state  $j$  and  $i$  with  $x_j > x_i$ , and  $E_f(x_j)$  and  $E_f(x_i)$  are the corresponding formation energies of these two states. Sharp drops or step-like features in the voltage curve typically correspond to transitions between different storage mechanisms, while extended plateaus indicate stable two-phase regions or clustering processes. Formation energy calculations explain why sodium ions adopt particular storage configurations, while voltage curves illustrate how these processes appear during charge–discharge cycling. This combined approach bridges atomic-level simulations with electrochemical observables, offering a comprehensive framework for evaluating and optimizing hard carbon as an anode material for sodium-ion batteries. Fig. 4 shows the voltage curves from 0.7 to 1.9  $\text{g cm}^{-3}$  at three chemical potentials.

At  $\mu = -1.00$  eV, the structures exhibit high storage capacities approaching  $4000 \text{ mAh g}^{-1}$  in the extreme limit (which is unphysical). However, most of this uptake occurs at voltage approaching zero, which is unfavorable for practical operation due to the risk of sodium plating.<sup>50</sup> As the chemical potential decreases, the accessible capacity is progressively reduced, reaching  $\approx 1400 \text{ mAh g}^{-1}$  at  $\mu = -1.24$  eV and  $\approx 250 \text{ mAh g}^{-1}$  at  $\mu = -1.50$  eV. This trend illustrates the delicate balance between the driving force for Na insertion and the stability of the host framework. The role of density is also apparent across all values of  $\mu$ . Higher-density structures consistently yield slightly higher voltages at low degrees of sodiation, indicating stronger stabilization of sodium within compact host environments. In contrast, lower-density systems can accommodate significantly more sodium but at lower insertion voltages because, the larger the filled pore, the more the Na within resembles bulk metallic Na. All voltage curves exhibit a sharp initial drop within the first  $50 \text{ mAh g}^{-1}$  or so, associated with the filling of the most favorable adsorption sites, followed by a gradual decay as insertion proceeds into weaker binding environments.

To rationalize these trends, Fig. 5 decomposes the cumulative storage capacity into three distinct contributions: adsorption, intercalation, and pore filling, and illustrates the three corresponding classes of storage sites. To quantify the relative contributions, each sodium atom was assigned to a storage environment based on a hierarchical geometric classification scheme derived from local structural descriptors, including coordination environments, local confinement geometry, and sodium–sodium clustering. Pore-filling sites correspond to

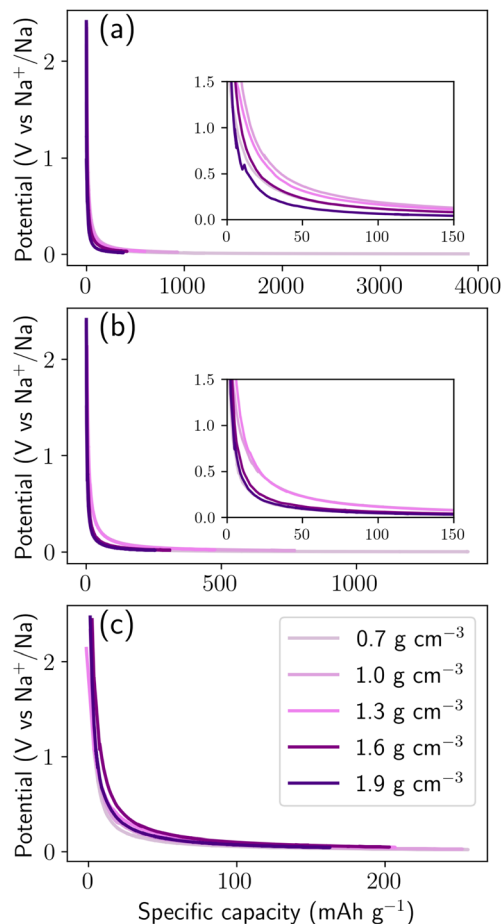


Fig. 4 Potential as a function of specific capacity at a given chemical potential: (a)  $-1.00$  eV, (b)  $-1.24$  eV, (c)  $-1.50$  eV. Each line represents a different density. Insets show details of the voltage inflection.

sodium atoms aggregated within nanopores or voids where Na–Na interactions dominate over Na–C bonding (Fig. 5e). Sodium atoms were considered connected when separated by distances between  $3.0$  and  $4.0 \text{ \AA}$  corresponding to the first coordination shell observed in the Na–Na radial distribution function in Fig. S4c. Clusters containing at least five sodium atoms and located at least  $2.6 \text{ \AA}$  from the nearest carbon atom (corresponding to the first minimum of the Na–C radial distribution function, Fig. S4b) were classified as pore-filling sodium, representing sodium confined within internal nanopore cavities where multi-ion aggregation occurs. Intercalation sites correspond to Na atoms residing in interlayer regions sandwiched between two graphitic sheets (Fig. 5d), analogous to Li intercalation in graphite.<sup>51</sup> For each sodium atom, neighboring carbon atoms within a  $5 \text{ \AA}$  radius were used to construct a local best-fit plane using principal component analysis. When the surrounding carbon atoms formed two approximately parallel layers, identified *via* clustering of their projections along the local plane normal, the interlayer separation was computed. Sodium atoms positioned between layers separated by  $2.8$ – $3.6 \text{ \AA}$  were classified as intercalated sodium, corresponding to insertion within expanded graphitic domains. Adsorption sites refer



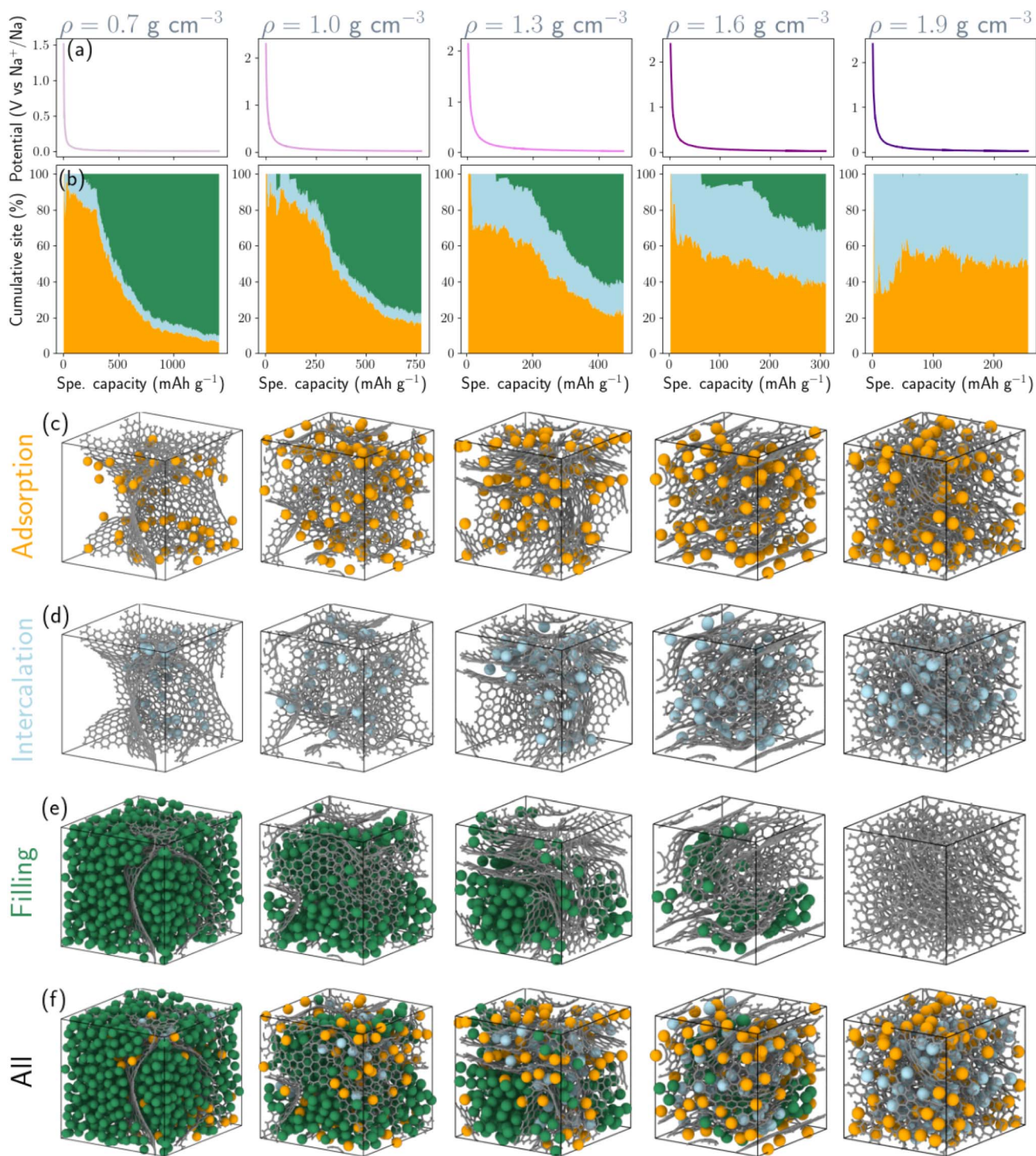


Fig. 5 Contributions of adsorption, intercalation, and pore filling to sodium storage in hard carbon across a density range of 0.7–1.9  $\text{g cm}^{-3}$ . (a) Voltage curves showing potential versus specific capacity. (b) Cumulative storage capacity of adsorption, intercalation, and pore filling as a function of specific capacity. Snapshots in (c–e) illustrate adsorption, intercalation, and pore-filling sites, respectively. The lower panel (f) shows all three classes of storage sites together. Simulations were performed at 300 K with a chemical potential of  $\mu = -1.24 \text{ eV}$ . All snapshots correspond to configurations at the maximum specific capacity. Color coding: orange = adsorption, blue = intercalation, green = pore-filling, and grey = carbon.

to Na-binding sites located on the carbon surface, on pore surfaces, edge sites, or defect regions without extended clustering or layered confinement. These surface binding sites

primarily consist of six-membered rings and edge motifs, with additional contributions from five- and seven-membered rings (Fig. 5c). The classification was applied hierarchically (pore



filling  $\rightarrow$  intercalation  $\rightarrow$  adsorption) to avoid misidentification of clustered pore sodium as adsorption states. Sensitivity analysis performed by varying the geometric cutoff parameters by  $\pm 0.2 \text{ \AA}$  resulted in variations of less than 5% in the relative populations of the three storage categories, confirming the robustness of the classification scheme. In addition, we computed the local energy distribution of the three classes (Fig. S3). Pore filling class exhibit distinct local energy distributions (Fig. S4), whereas adsorption and intercalation classes have a similar mean local energy. Therefore, energetic classification is complement to geometric classification but not enough to distinguish between adsorption and intercalation site. This procedure provides a quantitative and reproducible framework for decomposing sodium storage mechanisms across the density range investigated.

In low-density structures ( $0.7\text{--}1.0 \text{ g cm}^{-3}$ ), sodium uptake is initially dominated by adsorption at defect and surface sites. With increasing capacity, however, pore filling becomes the prevailing storage mechanism, accounting for more than 80% of the total Na uptake. Radial distribution function (RDF) analysis (Fig. S4c) confirms the formation of multi-ion Na aggregates within nanopores, with the first two peaks closely approaching those characteristic of metallic sodium. These confined clusters exhibit formation energies comparable to bulk Na (Fig. S3), resulting in insertion voltages approaching 0 V vs. Na/Na<sup>+</sup> and giving rise to the extended low-voltage plateau. Importantly, these aggregates remain spatially confined within nanopores, forming metallic-like cluster states. At intermediate densities ( $1.3\text{--}1.6 \text{ g cm}^{-3}$ ), adsorption continues to dominate the initial stages of sodiation, while intercalation into graphene-like domains contributes substantially throughout the capacity range. Concurrently, pore filling is progressively constrained as nanopore size and connectivity decrease. The diminished probability of Na–Na aggregation shifts the energetic landscape toward stronger Na–C interactions, leading to moderated insertion voltages and a reduced plateau contribution. In dense hard carbon ( $1.9 \text{ g cm}^{-3}$ ), nanopore volume becomes insufficient to sustain extended Na clustering. Sodium storage is therefore governed primarily by adsorption at defect and surface sites, together with intercalation into graphene-like domains, resulting in a predominantly sloping voltage profile and a reduced capacity ( $\approx 250 \text{ mAh g}^{-1}$ ). We emphasize that this adsorption-dominated behavior arises from a quantitative geometric and energetic decomposition within the simulations, rather than direct experimental site assignment. This mechanistic interpretation aligns with experimental observations of highly densified hard carbons, which typically exhibit suppressed low-voltage plateau capacity and predominantly sloping voltage profiles. Within the established adsorption–intercalation framework, the sloping region is associated with adsorption (with intercalation), while the near-zero-voltage plateau corresponds to nanopore filling and Na clustering. Our simulations provide an atomistic explanation for this density-dependent evolution, demonstrating that increasing density systematically shifts the dominant storage mode from pore filling toward adsorption and intercalation. Fig. 5 illustrates these density-dependent mechanistic contributions at  $\mu =$

$-1.24 \text{ eV}$ . Equivalent analyses at  $\mu = -1.00 \text{ eV}$  (Fig. S5a and S6) and  $-1.50 \text{ eV}$  (Fig. S5c and S7) confirm the robustness of this mechanistic evolution across the relevant chemical potential window. Collectively, these results establish a clear mechanistic linkage between structural density, sodium storage configuration, and electrochemical response. Because the voltage profiles are derived directly from insertion formation energies (eqn (7)), changes in the dominant storage mechanism translate quantitatively into voltage evolution. The predominance of confined Na clustering in low-density carbons explains the experimentally observed near-zero-voltage plateau, whereas adsorption and intercalation dominate the higher-voltage sloping region. Density therefore emerges as a governing structural parameter: low-density frameworks maximize capacity but operate near the sodium plating potential; high-density carbons enhance voltage stability at the expense of storage; and intermediate-density materials achieve the most favorable balance, combining moderated voltage, high reversibility, controlled suppression of metallic-like Na clustering, and limited structural deformation.

Experimental hard carbons synthesized *via* controlled pyrolysis and precursor engineering commonly exhibit skeletal (intrinsic) densities spanning  $1.1\text{--}1.7 \text{ g cm}^{-3}$ , with electrochemical behavior strongly modulated by density-dependent microstructure. Low-density carbons ( $1.0\text{--}1.2 \text{ g cm}^{-3}$ ) generally display pronounced low-voltage plateau capacities, commonly attributed to extensive nanopore filling and Na clustering.<sup>14</sup> In contrast, highly densified carbons ( $\approx 1.7 \text{ g cm}^{-3}$ ) display reduced capacity ( $\approx 286 \text{ mAh g}^{-1}$ )<sup>52</sup> but improved voltage stability and diminished plateau contribution, consistent with suppressed pore filling and limited Na–Na aggregation. Notably, intermediate-density hard carbons ( $1.3\text{--}1.6 \text{ g cm}^{-3}$ ) reported in recent studies achieve a more balanced electrochemical response. For example, carbons with skeletal densities of  $1.6 \text{ g cm}^{-3}$  and  $1.43 \text{ g cm}^{-3}$  deliver reversible capacities of  $390 \text{ mAh g}^{-1}$  (ref. 52) and  $410 \text{ mAh g}^{-1}$ ,<sup>10</sup> respectively, while exhibiting moderate plateau contributions, improved initial coulombic efficiency, and enhanced cycling stability. Similarly, the closed-pore-dominated carbon reported in 53 (true density  $1.64 \text{ g cm}^{-3}$ ) delivers a reversible capacity of  $387 \text{ mAh g}^{-1}$  with a substantial plateau contribution of  $278 \text{ mAh g}^{-1}$ , whereas more turbostratic carbons with higher true densities ( $1.95 \text{ g cm}^{-3}$ ) show markedly lower capacities ( $160 \text{ mAh g}^{-1}$ ). These materials avoid extended near-zero-voltage plateaus yet retain substantial reversible storage. Although the present work employs bulk density as the governing parameter, the experimentally reported intrinsic densities exhibit the same monotonic trend: increasing density suppresses excessive pore filling and metallic-like Na clustering while reducing total capacity beyond an optimal point. These trends are fully consistent with our simulations, which identify the same intermediate density range as an optimal structural window where adsorption and intercalation remain significant while excessive pore filling is constrained. This comparison strengthens the relevance of the predicted density–mechanism relationship for experimentally realizable hard-carbon electrodes.



## 4. Summary and conclusions

We have developed a comprehensive methodology to investigate sodium insertion mechanisms in hard carbon. Using machine-learning potentials coupled to grand-canonical Monte Carlo simulations, we performed realistic atomistic simulations from structure generation to the evaluation of electrochemical properties across a density range from 0.7 to 1.9 g cm<sup>-3</sup>. This approach provides direct insight into the relationship between HC porosity, sodium storage mechanism, and electrochemical performance. Our results reveal that sodium uptake in HC proceeds through three distinct contributions: adsorption, intercalation, and pore filling, whose relative importance strongly depends on mass density. Low-density HC (0.7–1.0 g cm<sup>-3</sup>) achieves high capacities dominated by pore filling but operates near the potential of metallic sodium, introducing a risk of Na plating. High-density HC (1.9 g cm<sup>-3</sup>) confines Na storage to adsorption and intercalation sites, resulting in safer operating voltages but limited capacity. Intermediate-density HC (1.3–1.6 g cm<sup>-3</sup>) achieves the best trade-off, combining balanced capacity, moderate voltage, and minimal structural deformation. These findings provide a microscopic explanation for the experimentally observed sloping and plateau regions in voltage profiles and clarify the density-driven origin of the capacity–voltage trade-off in HC anodes. The insights gained here establish density engineering as a key design strategy for next generation sodium-ion batteries, enabling the rational optimization of pore architecture, mechanical stability, and electrochemical performance.

## Author contributions

A. F. generated the database, trained MLIPs, generated hard carbon structures and performed GCMC simulations. A.F. performed data analysis and interpretation. M.A.C. and T.A-N helped in data analysis and discussion. M.A.C. conceived the idea and obtained funding. A.F. wrote the manuscript. All authors contributed to the discussion and commenting on the paper.

## Conflicts of interest

The authors have no conflicts of interest to declare.

## Data availability

Most of the data that support the findings of this study are available from public online repositories as listed in the references. Any data not publicly shared in this way can be obtained from the corresponding author upon reasonable request.

Supplementary information (SI) is available. See DOI: <https://doi.org/10.1039/d6sc00030d>.

## Acknowledgements

M.A.C. and A.F. acknowledge financial support from the European Union's M-ERA.NET program (NACAB project under grant agreement No. 958174). M.A.C. also acknowledges personal

support from the Research Council of Finland (project number 330488). T.A-N. and A.F. have been supported in part by the Academy of Finland through its QTF Center of Excellence program (project no. 312298) and European Union – NextGenerationEU instrument grant no. 353298. T.A-N. has been also supported by the Academy of Finland grants nos. 370057 and 373647. We acknowledge computational resources from CSC (the Finnish IT Center for Science) and Aalto University's Science-IT Project.

## References

- 1 J.-M. Tarascon, Na-ion versus Li-ion batteries: Complementarity rather than competitiveness, *Joule*, 2020, **4**(8), 1616–1620.
- 2 K. Chayambuka, G. Mulder, D. L. Danilov and P. H. L. Notten, From Li-ion batteries toward Na-ion chemistries: Challenges and opportunities, *Adv. Energy Mater.*, 2020, **10**(38), 2001310.
- 3 J.-Y. Hwang, S.-T. Myung and Y.-K. Sun, Sodium-ion batteries: present and future, *Chem. Soc. Rev.*, 2017, **46**, 3529–3614.
- 4 H. S. Hirsh, Y. Li, D. H. S. Tan, M. Zhang, E. Zhao and Y. Shirley Meng, Sodium-ion batteries paving the way for grid energy storage, *Adv. Energy Mater.*, 2020, **10**(32), 2001274.
- 5 M. D. Slater, D. Kim, E. Lee and C. S. Johnson, Sodium-ion batteries, *Adv. Funct. Mater.*, 2013, **23**(8), 947–958.
- 6 S. Schweidler, L. de Biasi, A. Schiele, P. Hartmann, T. Brezesinski and J. Janek, Volume changes of graphite anodes revisited: A combined operando X-ray diffraction and in situ pressure analysis study, *J. Phys. Chem. C*, 2018, **122**(16), 8829–8835.
- 7 N. Yabuuchi, K. Kubota, M. Dahbi and S. Komaba, Research development on sodium-ion batteries, *Chem. Rev.*, 2014, **114**(23), 11636–11682.
- 8 Y. Chu, J. Zhang, Y. Zhang, L. Qi, Y. Jia, X. Dong, J. Xiao, Y. Tao and Q.-H. Yang, Reconfiguring hard carbons with emerging sodium-ion batteries: A perspective, *Adv. Mater.*, 2023, **35**(31), 2212186.
- 9 X. Dou, I. Hasa, D. Saurel, C. Vaalma, L. Wu, D. Buchholz, D. Bresser, S. Komaba and S. Passerini, Hard carbons for sodium-ion batteries: Structure, analysis, sustainability, and electrochemistry, *Mater. Today*, 2019, **23**, 87–104.
- 10 X. Chen, C. Liu, Y. Fang, X. Ai, F. Zhong, H. Yang and Y. Cao, Understanding of the sodium storage mechanism in hard carbon anodes, *Carbon Energy*, 2022, **4**(6), 1133–1150.
- 11 T. Wesley Surta, E. Koh, Z. Li, D. B. Fast, X. Ji, P. Alex Greaney and M. R. Dolgos, Combining experimental and theoretical techniques to gain an atomic level understanding of the defect binding mechanism in hard carbon anodes for sodium ion batteries, *Adv. Energy Mater.*, 2022, **12**(25), 2200647.
- 12 Y. Huang, X. Zhong, X. Hu, Y. Li, K. Wang, H. Tu, W. Deng, G. Zou, H. Hou and X. Ji, Rationally designing closed pore structure by carbon dots to evoke sodium storage sites of



- hard carbon in low-potential region, *Adv. Funct. Mater.*, 2024, **34**(4), 2308392.
- 13 D. A. Stevens and J. R. Dahn, An in situ small-angle X-ray scattering study of sodium insertion into a nanoporous carbon anode material within an operating electrochemical cell, *J. Electrochem. Soc.*, 2000, **147**(12), 4428.
- 14 W. Jian, X. Qiu, H. Chen, J. Yin, Y. Wen, H. N. Alshareef and W. Zhang, Elucidation of the sodium-ion storage behaviors in hard carbon anodes through pore architecture engineering, *ACS Nano*, 2025, **19**(24), 22201–22216.
- 15 D. Datta, J. Li and V. B. Shenoy, Defective graphene as a high-capacity anode material for Na- and Ca-ion batteries, *ACS Appl. Mater. Interfaces*, 2014, **6**(3), 1788–1795.
- 16 P. Tsai, S.-C. Chung, S. Lin and A. Yamada, Ab initio study of sodium intercalation into disordered carbon, *J. Mater. Chem. A*, 2015, **3**, 9763–9768.
- 17 Y. Youn, B. Gao, A. Kamiyama, K. Kubota, S. Komaba and Y. Tateyama, Nanometer-size Na cluster formation in micropore of hard carbon as origin of higher-capacity Na-ion battery, *npj Comput. Mater.*, 2021, **7**(1), 48.
- 18 J. Li, C. Peng, J. Li, J. Wang and H. Zhang, Insight into sodium storage behaviors in hard carbon by reaxff molecular dynamics simulation, *Energy Fuels*, 2022, **36**(11), 5937–5952.
- 19 M. A. Caro, G. Csányi, T. Laurila and V. L. Deringer, Machine learning driven simulated deposition of carbon films: From low-density to diamondlike amorphous carbon, *Phys. Rev. B*, 2020, **102**, 174201.
- 20 A. P. Bartók, M. C. Payne, R. Kondor and G. Csányi, Gaussian Approximation Potentials: The Accuracy of Quantum Mechanics, without the Electrons, *Phys. Rev. Lett.*, 2010, **104**(13), 136403.
- 21 A. P. Bartók, R. Kondor and G. Csányi, On representing chemical environments, *Phys. Rev. B*, 2013, **87**, 184115.
- 22 M. A. Caro, Optimizing many-body atomic descriptors for enhanced computational performance of machine learning based interatomic potentials, *Phys. Rev. B*, 2019, **100**(2), 024112.
- 23 H. Muhli, X. Chen, A. P. Bartók, P. Hernández-León, G. Csányi, T. Ala-Nissila and M. A. Caro, Machine learning force fields based on local parametrization of dispersion interactions: Application to the phase diagram of  $C_{60}$ , *Phys. Rev. B*, 2021, **104**, 054106.
- 24 So Fujikake, V. L. Deringer, T. H. Lee, M. Krynski, S. R. Elliott and G. Csányi, Gaussian approximation potential modeling of lithium intercalation in carbon nanostructures, *J. Chem. Phys.*, 2018, **148**(24), 241714.
- 25 R. Sure, A. Hansen, P. Schwerdtfeger and S. Grimme, Comprehensive theoretical study of all 1812  $C_{60}$  isomers, *Phys. Chem. Chem. Phys.*, 2017, **19**, 14296–14305.
- 26 J. P. Perdew, K. Burke and M. Ernzerhof, Generalized gradient approximation made simple, *Phys. Rev. Lett.*, 1996, **77**, 3865–3868.
- 27 G. Kresse and J. Furthmüller, Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set, *Phys. Rev. B*, 1996, **54**, 11169–11186.
- 28 G. Kresse and D. Joubert, From ultrasoft pseudopotentials to the projector augmented-wave method, *Phys. Rev. B*, 1999, **59**, 1758–1775.
- 29 A. Front and M. A. Caro, GAP interatomic potential for CNa, *Zenodo*, 2025, DOI: [10.5281/zenodo.17159854](https://doi.org/10.5281/zenodo.17159854).
- 30 S. Klawohn, J. R. Kermode and A. P. Bartók, Massively parallel fitting of gaussian approximation potentials, *Mach. Learn. Sci. Technol.*, 2023, **4**(1), 015020.
- 31 *Quip*, <https://libatoms.github.io/QUIP>, accessed Jan 20, 2024.
- 32 A. Hjorth Larsen, J. J. Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dułak, J. Friis, M. N. Groves, B. Hammer, C. Hargus, E. D. Hermes, P. C. Jennings, P. B. Jensen, J. Kermode, J. R. Kitchin, E. L. Kolsbjerg, J. Kubal, K. Kaasbjerg, S. Lysgaard, J. B. Maronsson, T. Maxson, T. Olsen, L. Pastewka, A. Peterson, C. Rostgaard, J. Schiøtz, O. Schütt, M. Strange, K. S. Thygesen, T. Vegge, L. Villhelsen, M. Walter, Z. Zeng and K. W. Jacobsen, The atomic simulation environment—a python library for working with atoms, *J. Phys.: Condens. Matter*, 2017, **29**(27), 273002.
- 33 S. Alexander, Visualization and analysis of atomistic simulation data with ovito—the open visualization tool, *Model. Simulat. Mater. Sci. Eng.*, 2009, **18**(1), 015012.
- 34 *TurboGAP: Data-Driven Atomistic Simulations*, <http://turbogap.fi>, accessed Jan 20, 2024.
- 35 T. B. Shiell, S. Wong, W. Yang, C. A. Tanner, B. Haberl, R. G. Elliman, D. R. McKenzie, D. G. McCulloch and J. E. Bradby, The composition, structure and properties of four different glassy carbons, *J. Non-Cryst. Solids*, 2019, **522**, 119561.
- 36 T. B. Shiell, D. G. McCulloch, J. E. Bradby, B. Haberl and D. R. McKenzie, Neutron diffraction discriminates between models for the nanoarchitecture of graphene sheets in glassy carbon, *J. Non-Cryst. Solids*, 2021, **554**, 120610.
- 37 W. Huang, H. Zhang, Y. Huang, W. Wang and S. Wei, Hierarchical porous carbon obtained from animal bone and evaluation in electric double-layer capacitors, *Carbon*, 2011, **49**(3), 838–843.
- 38 G. Bussi, D. Donadio and M. Parrinello, Canonical sampling through velocity rescaling, *J. Chem. Phys.*, 2007, **126**(1), 014101.
- 39 A. Front and M. A. Caro, Hard carbon and sodiated hard carbon structures from grand-canonical Monte Carlo simulations, *Zenodo*, 2026, DOI: [10.5281/zenodo.18161390](https://doi.org/10.5281/zenodo.18161390).
- 40 Y. Wang, Z. Fan, P. Qian, T. Ala-Nissila and M. A. Caro, Structure and pore size distribution in nanoporous carbon, *Chem. Mater.*, 2022, **34**(2), 617–628.
- 41 P. Grigorev, L. Frérot, F. Birks, A. Gola, J. Golebiowski, J. Griefser, J. L. Hörmann, A. Klemenz, G. Moras, W. G. Nöhring, J. A. Oldenstaedt, P. Patel, T. Reichenbach, T. Rocke, L. Shenoy, M. Walter, S. Wengert, L. Zhang, J. R. Kermode and L. Pastewka, matscipy: materials science at the atomic scale with Python, *J. Open Source Softw.*, 2024, **9**(93), 5668.



- 42 J. W. Martin, C. de Tomas, I. Suarez-Martinez, M. Kraft and N. A. Marks, Topology of disordered 3d graphene networks, *Phys. Rev. Lett.*, 2019, **123**, 116105.
- 43 Debyer. <https://github.com/wojdyr/debyer>, accessed Dec 12, 2021.
- 44 T. F. Willems, C. H. Rycroft, M. Kazi, J. C. Meza and M. Haranczyk, Algorithms and tools for high-throughput geometry-based analysis of crystalline porous materials, *Microporous Mesoporous Mater.*, 2012, **149**(1), 134–141.
- 45 H. Chris, Rycroft. Voronoi++: A three-dimensional voronoi cell library in c++, *Chaos*, 2009, **19**(4), 041111.
- 46 M. Pinheiro, R. L. Martin, C. H. Rycroft, A. Jones, E. Iglesia and M. Haranczyk, Characterization and comparison of pore landscapes in crystalline porous materials, *J. Mol. Graph. Model.*, 2013, **44**, 208–219.
- 47 G. Wang, M. Yu and X. Feng, Carbon materials for ion-intercalation involved rechargeable battery technologies, *Chem. Soc. Rev.*, 2021, **50**, 2388–2443.
- 48 H. Alptekin, H. Au, A. C. S. Jensen, E. Olsson, M. Goktas, T. F. Headen, P. Adelhelm, Q. Cai, A. J. Drew and M.-M. Titirici, Sodium storage mechanism investigations through structural changes in hard carbons, *ACS Appl. Energy Mater.*, 2020, **3**(10), 9918–9927.
- 49 I. Escher, G. A. Ferrero, M. Goktas and P. Adelhelm, In situ (operando) electrochemical dilatometry as a method to distinguish charge storage mechanisms and metal plating processes for sodium and lithium ions in hard carbon battery electrodes, *Adv. Mater. Interfaces*, 2022, **9**(8), 2100596.
- 50 C. Bommier, T. W. Surta, M. Dolgos and X. Ji, New mechanistic insights on Na-ion storage in nongraphitizable carbon, *Nano Lett.*, 2015, **15**(9), 5888–5892.
- 51 D. A. Stevens and J. R. Dahn, The mechanisms of lithium and sodium insertion in carbon materials, *J. Electrochem. Soc.*, 2001, **148**(8), A803.
- 52 T. Zheng, R. Zhang, H. Wang, S. Zhou, Z. Pan, Y. Huang, D. Sun, Y. Tang, X. Ji, K. Amine and M. Shao, Revealing the closed pore formation of waste wood-derived hard carbon for advanced sodium-ion battery, *Nat. Commun.*, 2023, **14**(1), 6024.
- 53 Y. Xue, Y. Chen, Y. Liang, L. Shi, R. Ma, X. Qiu, Y. Li, N. Guo, Q. Zhuang, B. Xi, Z. Ju and S. Xiong, Substitution index-prediction rules for low-potential plateau of hard carbon anodes in sodium-ion batteries, *Adv. Mater.*, 2025, **37**, 2417886.

