Lawrence Berkeley National Laboratory

Title

A Continuous Action Space Tree search for INverse desiGn (CASTING) framework for materials discovery

Authors

Banik, Suvo
Loefller, Troy
Manna, Sukriti
et al.

Peer reviewed

ARTICLE   OPEN

A Continuous Action Space Tree search for INverse desiGn (CASTING) framework for materials discovery

Suvo Banik[1,2], Troy Loefller[1,2], Sukriti Manna [ID][1,2], Henry Chan [ID][1,2], Srilok Srinivasan [ID][1], Pierre Darancet[1], Alexander Hexemer [ID][3] and Subramanian K. R. S. Sankaranarayanan [ID][1,2 ✉]

Material properties share an intrinsic relationship with their structural attributes, making inverse design approaches crucial for discovering new materials with desired functionalities. Reinforcement Learning (RL) approaches are emerging as powerful inverse design tools, often functioning in discrete action spaces. This constrains their application in materials design problems, which involve continuous search spaces. Here, we introduce an RL-based framework CASTING (Continuous Action Space Tree Search for inverse design), that employs a decision tree-based Monte Carlo Tree Search (MCTS) algorithm with continuous space adaptation through modified policies and sampling. Using representative examples like Silver (Ag) for metals, Carbon (C) for covalent systems, and multicomponent systems such as graphane, boron nitride, and complex correlated oxides, we showcase its accuracy, convergence speed, and scalability in materials discovery and design. Furthermore, with the inverse design of super-hard Carbon phases, we demonstrate CASTING's utility in discovering metastable phases tailored to user-defined target properties and preferences.

INTRODUCTION

The properties of a material, such as chemical, physical, thermal, optical, and mechanical properties, are intimately tied to its crystal structure, topology and/or microstructure. Design, discovery, and structure-property relationships of structurally-distinct metastable crystalline polymorphs have been a long-standing challenge in materials science[1,2]. Crystal Structure Prediction (CSP)[1,3–11] involves navigating through a vast configurational and compositional space with high permutational variability, which makes it a challenging search problem. Global optimization techniques have been traditionally employed in such search problems to predict optimal materials for inverse design applications[6,9,10,12–15]. Alternate approaches were intuition-based and relied on empirical schemes[16]. This not only limits the tractability of the problem but is also very restrictive in terms of exploration.

In the past few decades, significant advancements in algorithmic development[4] and implementation, particularly, in CSP, have unraveled a new paradigm for predicting new materials that display exotic properties[2,4,5,17]. Data-driven approachs[3,7], simulated annealing[6,13], minima hopping[18], and meta dynamics[19,20] have been used with some success. For systems with smaller sizes, even random sampling followed by atomistic relaxation produces structures with stable configurations[21,22]. Metaheuristic techniques such as evolutionary algorithm[5,9,12,23], particle swarm[10,14,15], and basin hopping[24,25], have subsequently been developed and applied to a multifarious class of materials. This allowed a search for the ground state structures based on the chemical composition and synthesis conditions. Not only have the crystal structure prediction methods predicated new materials but many of these theoretically predicted configurations have been experimentally synthesized, bridging theory and experiment in design and discovery[26–29]. More recently, artificial intelligence (AI) and Machine Learning (ML) techniques have emerged as efficient tools in mapping quantitative structure to property relationship[30–36].

The overall success of any crystal structure or topology prediction methodology is widely dependent on the exploratory nature and the convergence strength of the search algorithm. As the dimensionality of the search space increases with either an increase in crystal size or composition, navigating efficiently through the search space with multiple local minima becomes very challenging[14]. Also note that more accurate methods like density functional theory (DFT)[37,38] are computationally expensive and the utilization of these methods in CSP necessitates an algorithm with fast convergence. In this regard, ML has again led to advances in the development of cheaper surrogate models to represent the underlying materials' physics and chemistry[37,38]. Furthermore, we note that inverse design involves searching across energy surface and configurational space that are continuous in nature. Each local minima on the energy surface represents a crystal structure (metastable state), as illustrated in Fig. 1a, and can exhibit exotic properties. Our target solution represents one such minima as highlighted by point 'C' in Fig. 1a. The hills on the energy (and corresponding configurational) surface are barriers that can be overcome with suitable thermodynamic conditions such as for example, temperature, pressure, composition, or their combinations. While we are interested in reaching the target solution, it is also desirable to explore the local minima or metastable states as well. This requires our search algorithm to establish a balance between exploration and exploitation. All sampling techniques[12,15], especially in high dimensional search space suffer from poor solution quality and/ or sluggish convergence owing to an exponentially increasing volume and the number of local minima[1]. A possible solution to this problem is learning from the explored part of the search space and utilizing the knowledge for further exploration. This can greatly improve the efficiency of the optimizer algorithm being

[1]Center for Nanoscale Materials, Argonne National Laboratory, Lemont, IL 60439, USA. [2]Department of Mechanical and Industrial Engineering, University of Illinois, Chicago, IL 60607, USA. [3]Advanced Light Source (ALS) Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. ✉email: skrssank@uic.edu
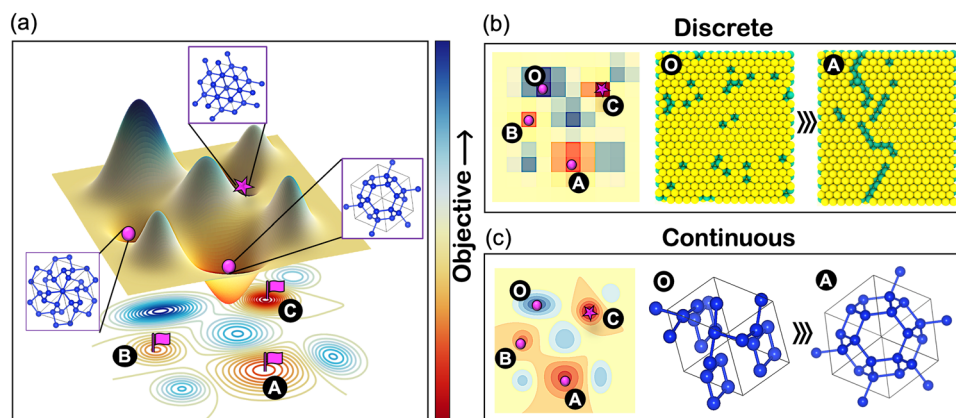
Fig. 1 Schematic illustration of the nature of the search space (discrete vs. continuous) in materials applications. a depicts the typical potential energy landscape of materials with different metastable polymorphs at local minima. **b** depicts a discrete action space with defects movement as actions[32]. Moving from a high energy defect configuration situated at O to a relatively stable polymorph at A consists of finite movement in the defect configurational space where navigating the PES involves finite steps with a discrete jump in energy. **c** Crystal structure optimization represents a continuous action or search space problem with infinite possibilities of moving an atom giving a large number of pathways to navigate from high energy polymorph at O to local minima at A.

used. Reinforcement learning (RL) with the ability to learn on the fly from the current state of the system, and make decisions not only based on the current state of the system, but, also, on the history, can greatly aid in overcoming the so-called 'curse of dimensionality'.

RL-based approaches have achieved remarkable success in solving problems with seemingly large intractable search space, such as board games Chess, Shogi, and Go[39,40], and more recently in materials applications such as chemical synthesis planning[41] or drug discovery[42–44]. Most of the materials applications to this day have been limited to discrete action spaces[45] as shown in Fig. 1b, including, for example, optimization of the geometry of lattice defect[31,32] described as a set of discrete positions on a finite lattice. However, many real-world problems including several grand challenges in materials discovery and design involve decision-making and search in a rather continuous action space[46] (Fig. 1c), which makes the optimization task harder. For example, in the discrete action space as shown in Fig. 1b, moving from defective configurations O to A can be attained via swap moves on a discrete atomistic lattice to navigate via a finite number of paths and reach the global minima at C. On the other hand, for the same task in continuous action space, as shown in Fig. 1c, there are infinite possible intermediate states and transition pathways possible between any two states (crystal or configurations), such as between O and A.

In this work, we introduce a scalable RL approach for structure & topology prediction, design, and optimization. This framework, entitled 'Continuous Action Space Tree search for INverse desiGn' (CASTING), employs a decision tree-based RL algorithm, i.e., Monte Carlo Tree Search (MCTS)[31,39,41]. MCTS efficiently explores a high-dimensional search landscape with multiple objectives by semi-stochastically sampling (playouts) in the proximity of a node, evaluating and learning its quality in a given search tree. It then takes policy-based decisions to explore the regimes of the search space (i.e., part of a tree) while striking a balance between exploration and exploitation to efficiently reach the target objective i.e., a configuration that maps to our desired material properties. We demonstrate the accuracy, speed of convergence, scalability, and applicability of our CASTING framework across a spectrum of problems (from bulk to low-dimensional, single to multiple components, and search space varying from unit to several large supercells) in the domain of CSP and Design. To assess scalability and speed of convergence, we begin with a metal example such as Silver (Ag), with fewer polymorphs and a smaller number of known local minima in its energy landscape.

For this system, we also conduct a performance analysis of our framework, varying different hyperparameters. We then extend our approach to predict the covalent system Carbon, which exhibits a diverse range of metastable states and polymorphs. All previously mentioned applications pertain to bulk (periodic) systems. Our exploration then extends beyond bulk systems as we investigate dimensionality effects on our workflow. Primarily, we explore two different classes of systems: a 0D (cluster) single-component system, such as gold (Au) for representative sizes, and 2D binary systems such as C-H (Graphane) and Boron Nitride (h-BN) to obtain their global minima. To explicitly explore compositional variance-induced metastability, we employ CASTING to explore the compositional space of doped Neodymium Nickel Oxide (NNO), focusing their impact on representative electronic property such as bandgap. Finally, by employing CASTING, we predict super-hard phases of carbon, highlighting its applicability in inverse design.

RESULTS

Crystal structure optimization

To perform a crystal structure optimization, we represent the configuration or the crystal as either periodic (bulk) or a low dimensional crystal by specifying a set of lattice parameters, basis atoms, and/or atomic compositions of its species. We treat the above-described problem as optimization of the lattice parameters (a, b, c, α, β, γ), the number of basis atoms (n), its positions, and atomic compositions of its species. Thus, any crystal structure is represented as a vector with six lattice parameters, and three times the number of atom coordinates (x, y, z) with chemical species belonging to each coordinate. MCTS spawns a tree with each node containing a point in the parameter space being searched for and obtains a score indicating the potential to find a promising structure nearby. The root node is initially assigned to random points in the parameter space or seeded with previously known configurations as shown in Fig. 2a. To sample a node nearby by perturbing the configurations, we implement different perturbation moves. Mainly four types of perturbation (Fig. 2b) moves are used (a) 'Add atom' (retaining the composition), (b) 'Remove atom' (retaining the composition), (c) 'Mutate lattice' (mutation of lattice parameters) and (d) 'Mutate atom' (mutation of atomic coordinates). Note that for the mutation of lattice parameters and coordinates we employ a hypersphere perturbation scheme (refer to methods section). The radius of the hypersphere is gradually reduced using a gaussian 'Depth scaling'
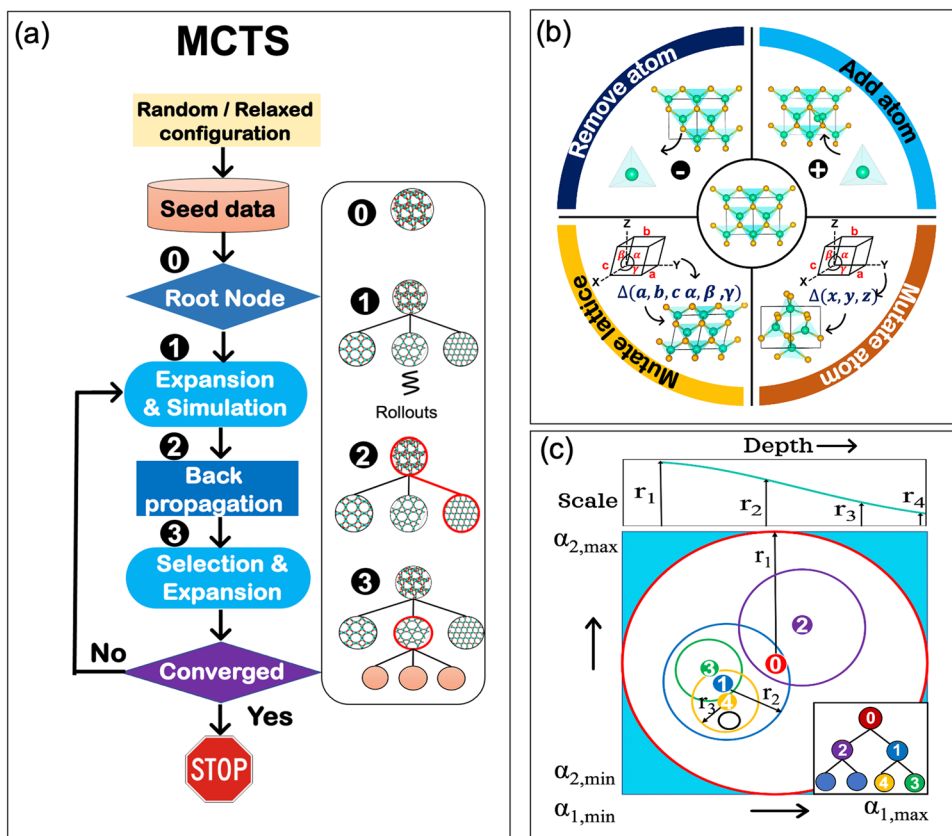
Fig. 2 MCTS working as crystal structure optimizer. a Workflow showing the various stages of MCTS deployed as a crystal structure optimizer constructing a tree search starting from a random or a relaxed configuration as a single node. **b** Four different types of perturbation move imparted on a crystal structure in a node as an offspring crystal is created from a parent. **c** 'Depth Scaling' scheme, implemented as decreasing radius of hypersphere as the depth of the search tree increases.

function (refer to methods section & supplementary Fig. 1b). Also note that the moves that change dimensionality (i.e., size of the system) such as 'Add atom' or 'Remove atom' are done for only one composition unit. For instance, in Graphane (with a C:H ratio of 1:1), a supercell with 10 atoms (5C atoms and 5 H atoms), performing an 'Add atom' move would entail adding one C and one H atom, while performing a 'Remove atom' move would involve eliminating one C and one H atom to maintain the C:H composition during the search. This helps maintain a parent-child correspondence for a given node (some degree of similarity between the parent and child). Initially, the probabilities of selecting each move are assigned an equal value. However, it should be noted that these probabilities may need to be biased for specific applications. For example, for fixed atomic systems such as non-periodic clusters, mutation moves are given higher priority over moves that add or remove atoms. The target objective such as cohesive energies per atom (although any target property computed using Molecular Dynamics (MD) and/or Density Functional Theory (DFT) can be used) of the structures are computed after local atomistic relaxation with the LAMMPS[47] package and the electronic properties such as band-gap were computed using the VASP[48] package.

The optimization with MCTS primarily involves four stages starting from a point in parameter space (root Node) and branching out by sampling new parameter sets (crystal configurations) as shown in Fig. 2a. The first stage involves expanding a node ('Expansion') by sampling new offspring nodes from it by using perturbations (Add atom, Remove atom, Mutate, etc.). Then it is the 'Simulation', where the search learns a qualitative score for selected offspring nodes by carrying out random playouts. A playout is basically random exploration near a parent node in the

search space by spawning new offspring from it, that are not radically different from the parent but inherits some of its traits instead (refer to method section). From the overall quality of these offsprings, a measure of a qualitative score of a parent node is obtained. Learnings are then backpropagated ('Backpropagation') to the root node for updating the score of the tree. And a "Selection" and further "Expansion" are carried out thereafter. Note that modified MCTS follows a UCB (Upper Confidence Bounds) (Eq. 2) policy for the selection of a node (refer to method section). The search is conducted till the termination criterion is reached. All the sampled configurations are then mapped according to their stability and potentially good samples are selected based on filtering descriptors[30,49].

The CASTING framework

Figure 3a, b provides an overview of the CASTING framework developed in this work. It has 6 modules that require input from the user. These include (1) The definition of the optimizer (2) selection of target properties to be predicted (3) objective definition or scoring function (4) definition of the crystal system including types of species and number of components (5) simulator or evaluator for the target property (MD or Ab-initio packages) and (6) output options for data analysis and information extraction. An additional 'Outputs & Monitor' module provides visualization options for the end user (Fig. 3). The first section requires the user to select the optimizer of choice (RL approach such as MCTS or evolutionary such as GA) and set corresponding hyperparameters that are required with it. In this study, we focus on MCTS as our primary optimizer although we make some limited comparisons to a genetic algorithm-based search in
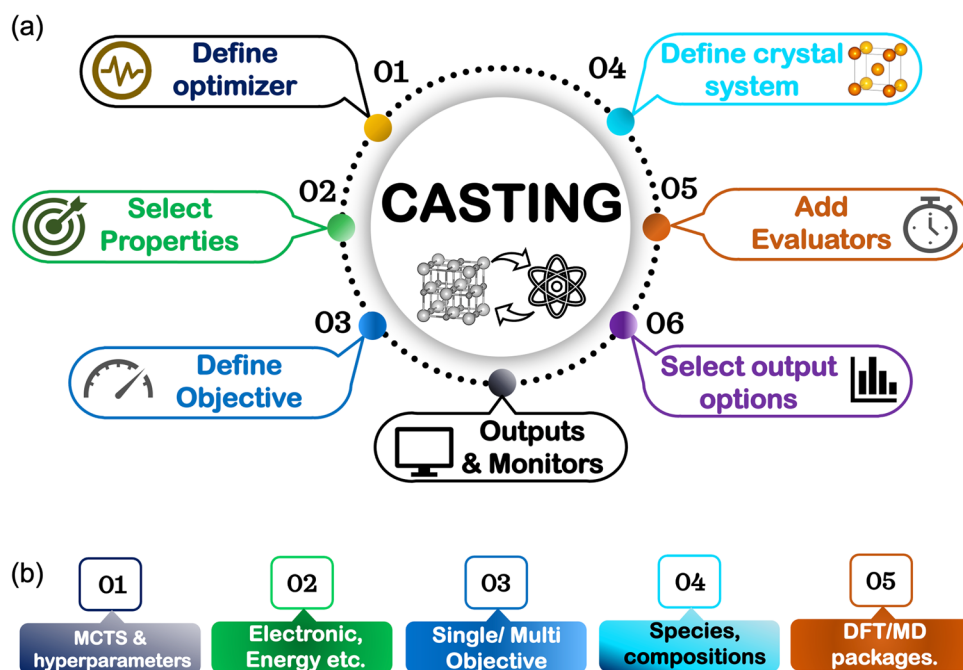
Fig. 3 Schematic depicting the workflow of the CASTING framework for performing inverse design. a User interface for specifying various IO settings leading to a different set of operations at the front-end of CASTING. These include (01) Defining the type of optimizer (02) Selection of properties to be predicted (03) Objective definition (04) Definition of the crystal system or configuration, (05) Evaluators (MD or Ab-initio packages) for computing the rewards or score, and (06) Output options. An additional 'Outputs & Monitor' module is available for visualization. **b** Additional input options associated with each of the operations specified at the front end in (**a**)—this includes (01) MCTS search and associated hyperparameters (02) target properties to be computed (03) single or multiple objectives (04) single or multicomponent or types of species (05) classical MD or electronic structure simulator to evaluate properties.

selected cases. The tree hyperparameters that require explicit input from the user are the number of 'Head expansion', the number of 'Playouts', 'Exploration constant', a 'Depth Scaling' parameter, and the maximum depth of the tree (refer to methods section for details). The target properties that need to be optimized are specified next. The properties can be energetics-based (potential energy, enthalpy, free energy), mechanical (elastic, phonon), electronic (band structure, density of states), and/or thermal (thermal conductivity) to name a few. In this work, we primarily use energy (and elastic moduli) as our target property. Selection of objective function is a crucial step and is entirely dependent on the choice of the optimizer. With MCTS, we use the Upper Confidence Bound (UCB) (Eq. (2)) as the objective function (refer to methods section). The 'UCB' itself requires the 'exploit' or the 'reward' (e.g., configurational energy) to be defined. Additionally, the weights on each 'exploit' may be required in the case of multi-objective optimization. Next, the crystal parameters are to be specified. This includes a range for the number of atoms in the simulation cell, lattice bounds range, lattice angle range, chemical species and compositions, and minimum allowed interatomic distance. These parameters define the search space, size, and dimensionality of the optimization. In cases where the bounds are not known upfront, it is advisable to set large initial bounds for the search, allowing it to explore configurations that meet other constraints, such as minimum interatomic distance criteria (refer to supplementary note 1 for additional details). After the target properties, crystal system, and objective function are defined, the user needs to provide corresponding packages for atomistic and electronic calculations (e.g., LAMMPS & VASP package for MD and DFT respectively, are used in this study). This part also contains the simulation settings and parameter flags associated with these property evaluation packages. Finally, the 'Output options' is for the post-processing section. The user defines the additional outputs such as data formats, visualization monitors, termination criteria, and other metrics that can be used

for a quantitative understanding of the quality of a search. There is an additional 'Outputs & Monitor' section which provides the user with the flexibility to monitor on the fly, search attributes such as current objective status, tree size, node content, sampled configuration, etc.

Applications of CASTING

The application of the CASTING framework involves a collection of pertinent and challenging problems within the realm of CSP and design. Among the various problems we have explored, we also conduct a comparison of the speed of convergence, accuracy of the best solution, and sampling quality achieved using our RL approach against traditional structure prediction methods, such as genetic algorithm (GA)[9,12] basin hopping[24,25], and random search[22]. It's important to note that different runs with the CASTING framework involved varying sets of hyperparameters. A typical strategy for obtaining these hyperparameters is discussed in Supplementary note 2, along with the hyperparameters used for different searches, which are included in Supplementary Table 1.

Exploring the scalability of CASTING framework using an example of metal polymorphs

Silver (Ag) is a well-studied metal and is known to have only a few metastable polymorphs (e.g., hcp, fcc, etc.) with the fcc as the most stable or ground state in its bulk form. We utilize Ag as a representative test case to evaluate the scalability of our framework. Any structural search performed with a decision tree such as MCTS primarily depends on the two aspects of the search parameters. (a) specifications of the crystal parameters (size, lattice parameters), and (b) hyperparameters that control the construction of the tree.

We first explore the impact of the crystal input parameters on the performance of our RL approach. Given that the solution is
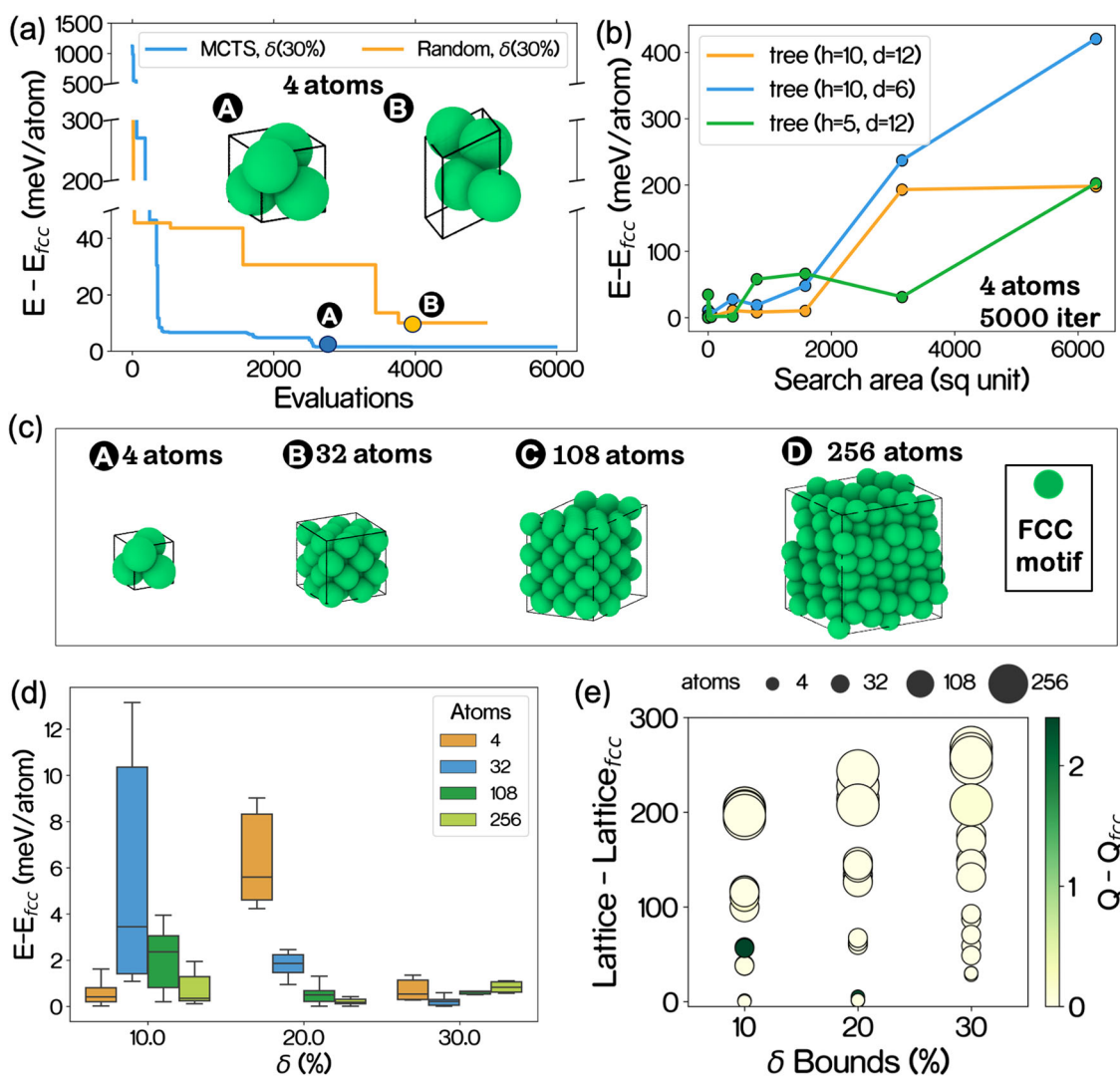
Fig. 4 Exploring the performance and scalability of CASTING framework using an example metal polymorph. a Comparison of the speed of convergence and difference in energy from the best available solution (Ag_{fcc}) between random sampling and MCTS optimizer for four atom system of Ag. **b** Performance of the MCTS optimizer (for different sizes of tree) for the problem in (**a**) as the area of the search space changes. **c** Effect of dimensionality on the predicted crystal structure for different system sizes. **d** Distribution energy difference (from fcc) (meV/atom) of the best solution obtained (in 20,000 iterations) for six independent trials on different sizes of the system with increasing lattice parameter bounds (δ) from a relaxed orthogonal supercell Ag (fcc). **e** Structural variation for the different minima obtained from the independent trials (as in (**d**)) in terms of changes in lattice parameters (from a relaxed orthogonal fcc supercell) and atomic stacking (difference from a pure fcc) for different sizes and lattice parameter bounds (δ).

known (i.e., the lattice petameters and atomic coordinates of ground state fcc structure), we set the search bounds of the lattice parameter in terms of percentage deviation (δ) from its stable counterpart. For example, a deviation in the bounds by 30% means a lattice vector range of $[0.7*l, 1.3*l]$, where **l** is the lattice vector of the pure fcc for a given size of supercell. We first start with a 4-atom search to test the typical convergence profile of the MCTS optimizer and compare it with a purely random search with local minimizations of the configurations to get an idea of the qualitative threshold (Fig. 4a). We use an EAM type empirical potential[50] and set the lattice parameters bounds deviation(δ) to be 30% (Refer to supplementary Table 1 for the hyperparameters). A LAMMPS simulation package was used for the evaluation of the structural property (energy). We find that allowing atoms to approach closer during the search (i.e., specifying a lower value for allowed minimum inter-atomic distance criteria) allows the RL to explore the search space more exhaustively (through high energy

regimes and overcome energy barriers) and helps in overall convergence.

Figure 4a shows that our MCTS search reaches the optimal solution in fewer evaluations compared to the random sampling —the solution quality with MCTS is also better i.e., lower in configurational energy. The stacking of the final predicted structure corresponds to an fcc fingerprint. The energy difference of the final solution from MCTS to that of the pure fcc is negligible ($\ll 1$ meV). Since we are growing a tree of finite size while exploring search space, it is expected that a significant change in the search space size (area) might affect the performance of the search (Fig. 4b). We define a search area to be the magnitude of vector cross product between the upper and lower bound of the lattice parameters vectors. To test this dependence, we spawn 3 trees using the same root node with different head expansions (h) and depth (d) (Fig. 4b). For a tree with less width (head expansion) ($h = 5$, $d = 12$), with the increase in the search area, the performance drops rapidly since the size of the tree is not

adequate to cover the entire search space. As the width of the tree increases ($h = 10$, $d = 12$) the performance becomes much better for lower value areas of the search space. However, we do notice a general decline in the performance, with an increase in the search space area. This is because, in a continuous actions space, an increment in the search space area introduces innumerable configurational possibilities in the energy landscape. While it also increases the possibility of finding a better solution, a greater number of iterations are required to explore it. At the same time, it is also obvious that a shallow tree (less depth) ($h = 10$, $d = 6$) also results in poor performance. As the tree depth increases, the search mostly exploits branches with promising nodes in the tree. A shallow tree restricts the search from exploitation, resulting in delayed or no convergence at all.

We next test the scalability of the CASTING workflow by testing the convergence speed and the energy per atom difference for convergence towards a unit cell of fcc (4 atoms), a supercell of 2*2*2(32 atoms), a supercell of 3*3*3 (108 atoms), and a supercell of 4*4*4 (256 atoms). The width and the depth of the search tree are kept fixed ($h = 10$, $d = 12$). We also select a wide range of the search bounds deviation (δ) from 10 to 30% deviation for testing. We perform six independent trial searches (initializing the root node of the tree at different points in search space) for each of the cases with the maximum number of iterations kept at 20,000. For the best solution from each of these trials, the distribution of energy difference from its fcc supercell counterpart, and the corresponding difference of the structure in terms of lattice parameters and stacking have been shown in Fig. 4d, e. To determine the similarity of the atoms to that of an fcc stacked lattice we used bond order-based parameters based descriptor (Q_2, Q_4, Q_6)[51] (cutoff 3 Å) and coordination number (CN) while the difference in lattice parameters are calculated using 'l_1' norm of the scaled lattice parameter vector ([a, b, c, α, β, γ]) with respect to the lattice parameter vector of the reference fcc structure. One can note that the fcc motif (displayed in green color, Fig. 4) is determined using CNA[51] (Common Neighbor Analysis) method.

It can be observed that for each of these sizes, there is an optimal bounds deviation (δ), for which the search gives the best performance (less variation in final energies and very close to the target) (Fig. 4d). Also note that as we move higher either in size of the system or the bounds deviation (δ), there is a tendency to achieve solutions that have vastly different lattices from the orthogonal supercell, but atoms are stacked in an fcc motif (Fig. 4e) with energies extremely close to the target solution. The effect is more prominent with changes in bounds deviation (δ). These primarily are two contributing factors for MCTS obtaining these degenerate solutions, (1) With an increase either in size or dimension(size) of bounds deviation (δ), the search constraints get lighter allowing atoms to arrange themselves in fcc motif while not having an orthogonal lattice (2) With an increase in the bounds, the corresponding area of the search space also increases, which allows MCTS to explore higher energy regimes of the search space (refer to Supplementary Fig. 2c) causing it to find these energetically close degenerate solution while severely delaying the final stages of the convergence (reaching to the exact orthogonal structure). There is also a dependency on the size of the tree as discussed earlier. For example, with 4 atoms at $\delta = 10\%$, the atom can only arrange themselves in an orthogonal fcc unicell, thus the best solution is obtained. With $\delta = 20\%$, the atoms do not have the flexibility of getting degenerate solutions, and also the size of the tree relatively is large for a given search space area. Hence the search could not get to solutions within fixed iterations (20,000) and the energy distribution is wide (Fig. 4d). For $\delta = 30\%$, the degeneracy can be seen, thus the energy distribution becomes much better owing to these solutions. Similar nonmonotonicity in performance can be observed for the other sizes too. The overall performance, for the given size of the tree ($h = 10$, $d = 12$) is optimum at $\delta = 30\%$, for all the

dimensionalities (system sizes). Note that with the increased dimensionality (Fig. 4d), the best solution obtained by MCTS for each case has a range of energy difference <0.15 meV, indicating the ability of the MCTS optimizer to scale to the dimensionality as high as 774 (256 atoms * 3 cartesian coordinates + 6 lattice parameters) while maintaining a considerable solution accuracy. While for a random search, the performance deteriorates considerably (refer to Supplementary Fig. 2b).

Next, we explicitly explore the different tree hyperparameters and analyze their effect on the convergence and overall sampling quality as shown in Fig. 5. The maximum number of iterations was kept at 2000 and the starting point (root node) of the search was kept the same for all cases. The number of atoms range was fixed at 4 atoms and a bounds deviation (δ) of 30% was maintained. In Fig. 5a, we show the effect of the increasing head expansions for the tree construction on the overall sampling and convergence of the search. The head expansion of the MCTS is somewhat comparable to generating an initial population in the evolutionary approaches. To start with, one would want to have minimal sampled points that cover the search space uniformly. Further branching out from those points helps the search to converge faster. Too many head expansions will generate redundant points in the same regions of search space causing the MCTS to explore unnecessarily more before reaching a converged solution resulting in an energy distribution with a high mean (Fig. 5a) and a typical slower convergence. The converse is true for a very smaller number of head expansions which might cause the search to get stuck in a certain region of the search space and may completely obstruct its convergence. However, with a very large number of evaluations, all the searches, irrespective of the head expansion value, are eventually expected to converge (refer to supplementary Fig. 2d. We next look at the effect of playouts (Fig. 5b). Playouts are basically random perturbations on a node to get a quantitative idea of how likely a node is to yield a good offspring upon further exploration. From the perspective of sampling, it is evident that there is an optimum for the number of playouts required. Too much of a playout will unnecessarily increase the number of iterations thus resulting in a slower convergence and too less of a playout might result in incomplete knowledge regarding any given node leading it to converge at a slower pace as well.

The exploration constant is another crucial parameter for the UCB (refer to methods section—Eq. (2) setting as well as an important parameter that controls the exploration of the tree. For too small of an exploration constant, the tree will greedily pick the nodes with good objective value only making the search confined to a certain region of the search space (Greedy Search). This can have an adverse effect on overall convergence. On, the other hand, selecting a too large constant will make the search to be effectively random. So, a proper selection of exploration constants can help the search to converge efficiently in relatively few numbers of expensive objective function evaluations (Fig. 5c). The final hyperparameter that we explored the effect of is the 'depth scaling'. For any MCTS search, as the depth of the tree increases the parameters at the nodes are expected to be closer to the converged solution than that of a node residing at a higher depth. This is also indicating that the search is moving towards an exploitative phase and thus a scaling of the sampling window is necessary. Otherwise, it might deviate the search from moving towards convergence. We use a gaussian type depth scaling scheme (refer to methods section, Supplementary Fig. 1). From Fig. 5d, we refer to that there is a slightly slower convergence for both higher and low values of 'a'. A low value of 'a' causes the search to become too much exploitive at a shallow depth of the tree. Since it samples only degenerate solutions in a small region of the search space while a high value of 'a' prevents it from being exploitive in a tree with high depth when it is required to do so.
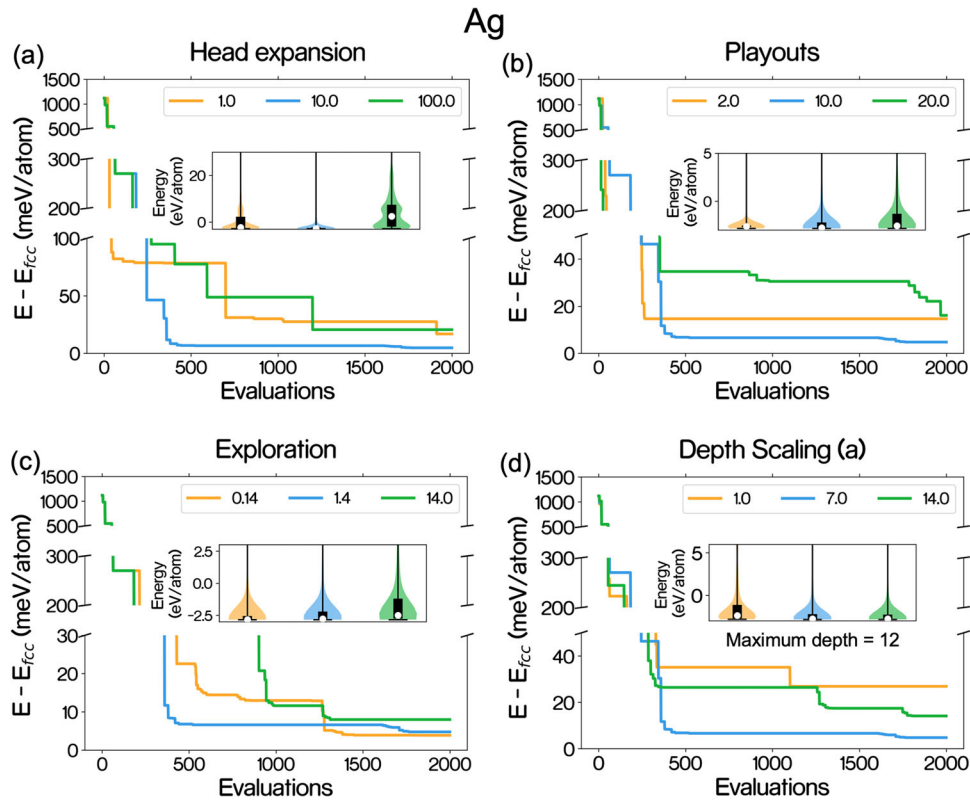
Fig. 5 Effect of tree hyperparameter on the sampling, convergence, and solution quality of Ag polymorphs. a Shows the convergence and energy distributions for different head expansions. **b** Shows the convergence and energy distributions for different playouts used. **c** Shows the convergence and energy distributions for different head expansions used. **c** Shows the convergence and energy distributions for different head exploration constants used. **d** Shows the convergence and energy distributions for different depth scaling factors '*a*' used.
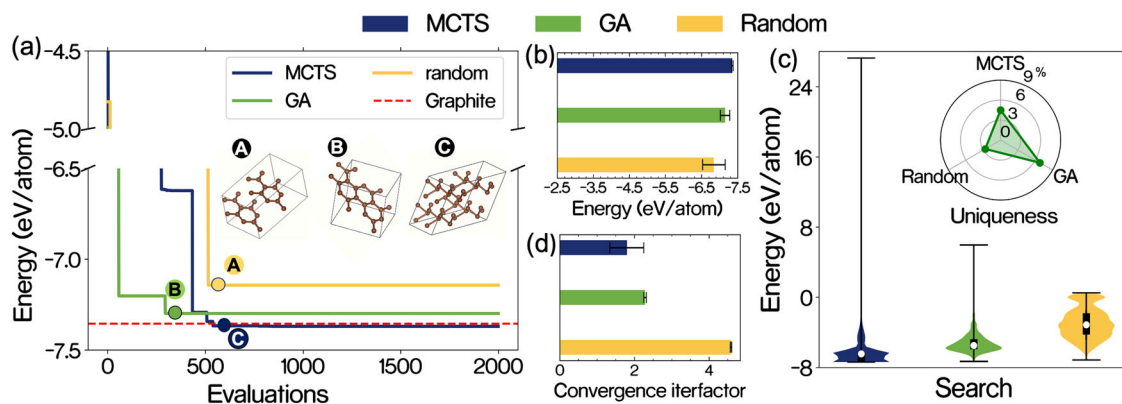
Fig. 6 Comparison of structure prediction for carbon polymorphs with an empirical potential model[52]. a Best Convergence of MCTS, GA, and random sampling out of four independent trials. **b** Mean the best solution obtained for MCTS, GA and random sampling. **c** Typical energy distribution of the sampled configuration during an independent run for MCTS and GA optimizer and their overall uniqueness. **d** Average iteration factor for convergence for different optimizer algorithms used.

Exploring the diverse metastable states and polymorphs of carbon using CASTING

We next explore another system which has a high degree of metastability i.e., has many local minima in its energy surface. Carbon is known to have a diverse range of allotropes, in terms of size, property, and structural diversity. This makes it a suitable test system for benchmarking the sampling quality, accuracy, and speed of convergence of the CASTING framework. Since it is already known that graphite and diamond (at high pressure) are the two most stable allotropes, we set them as our target solution. We start with 3 different search cases (a) CASTING (b) genetic

algorithm (GA)[9] (c) random search with local minimization of the structures (Fig. 6a)—the atom number is in the range [2,10], lattice vector range [2 Å, 8 Å], and lattice angle range [60⁰,120⁰]. The Tree hyperparameter settings are given in the Supplementary Table 1. The empirical LCBOP[52] potential along with the LAMMPS simulation package for local minimization of the configurations and calculation of energy.

From the results of three independent trials (Fig. 6b, d) and the best solution for each case (Fig. 6a), it is very clear that the MCTS optimizer in the CASTING framework not only converges faster to the solution Fig. 6d (The 'convergence iterfactor' is the normalized number of iterations taken for the convergence of the search), but
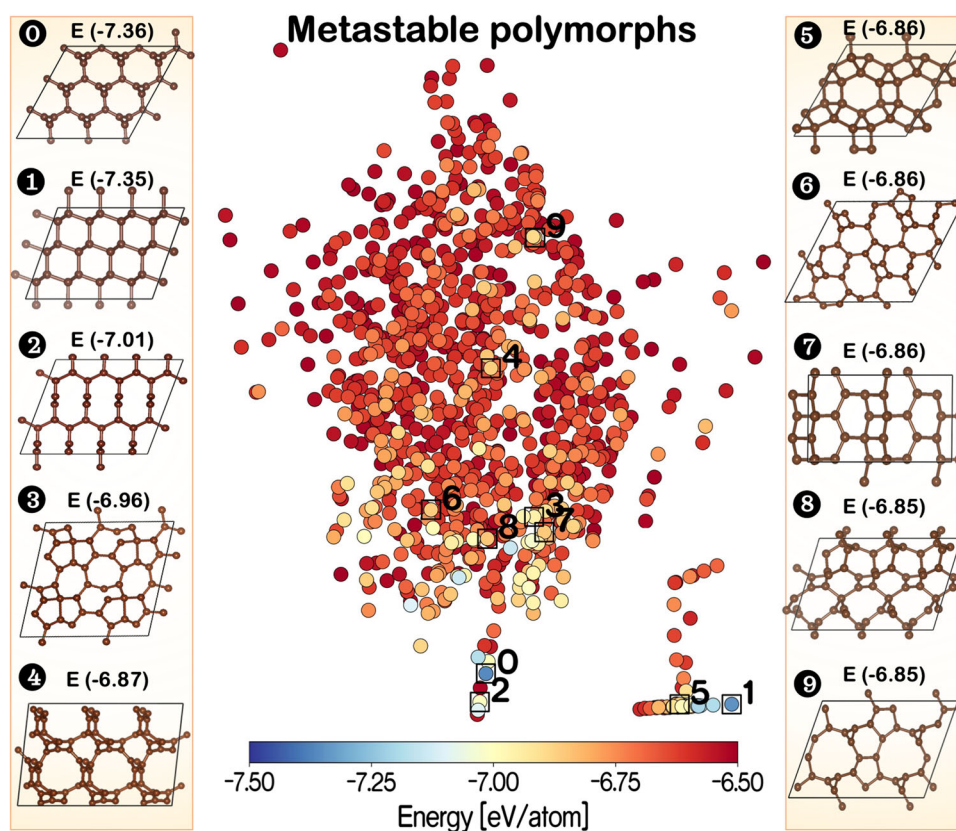
Fig. 7 Structural diversity of sampled Carbon(C) polymorphs using CASTING. ISOMAP representation of Order Parameters (Q_2, Q_4, Q_6) and Coordination Number) Based feature vectors of bulk metastable polymorphs of Carbon(C) sampled using CASTING framework with LCBOP[52] interatomic potential across a range of external stress spanning from 0 to 120 GPa.

the quality (the energy per atom) is also better (Fig. 6b). We also compare the property (energy per atom) distribution of the configurations sampled using MCTS and GA optimizers (Fig. 6c). Clearly, MCTS tends to sample more configurations in the lower energy range as compared to GA, but the overall uniqueness of the sampled configurations is less as compared to the GA (Fig. 6c). This is indicative of the fact the MCTS tends to sample more similar polymorphs near the global minima to reach the absolute best solution (exploitive) since most of the PES of empirical[52] potentials have degenerate solution of the same structure (Graphite in our case) with a very minute difference in energy. Which sometimes hinders more exploratory type search algorithms such as GA to reach the absolute solution. On the other hand, the GA has a slight upper hand in terms of sampling more diverse polymorphs because of its exploratory nature.

Note that the MCTS can also be made exploratory in nature by incrementing the exploration constant 'C' in the UCB (Eq. (2) in method section). By implementing the same for the Carbon polymorphs, we search with our CASTING framework for metastable phases of Carbon polymorphs at different external pressure ranging from 0 to 120 GPa. To find out the unique ones amongst the multiple different structures sampled with MCTS, we adopt a two-step method. Our solution contained a lot of variants of graphite polymorphs. Therefore, we first apply a graph neural network-based characterization workflow[30] to isolate the 2D layered polymorphs from bulk structures. Next, we filter out the unique ones from the bulk configurations using order parameters (Q_2, Q_4, Q_6)[51] + CN feature representation of the bulk configurations and an unsupervised agglomerative clustering[53] technique (refer to supplementary note 3). From the ISOMAP representation[54] feature vectors of the unique bulk polymorphs (Fig. 7), the MCTS optimizer not only sampled a large number of (~1.2 K)

diverse metastable polymorphs but also across a wide energy window (~1 eV). Also, note that MCTS managed to sample the diamond structure (Fig. 7 configuration 1) that exists at higher energy value as compared to the global minima graphite. In the phase diagram of carbon[55], the graphite polymorph is stable at regular thermodynamic conditions whereas the diamond polymorph exists under extreme pressures, which, makes the diamond polymorph metastable at regular thermodynamic conditions. Since there are exponentially many local minima introduced as the overall energy window of the search increases[1], thus discovering diamond becomes difficult. In general, the GA-based structural search converges for bulk[56] systems but typically requires more evaluations to converge compared to the MCTS. It is also worth mentioning that, in terms of computational time associated with the searches (GA and MCTS), the bottleneck lies in the method used for property evaluations (e.g., DFT or MD). Having searches with costly estimators necessitates the search to converge with fewer evaluations to save computational time (refer to the supplementary note 4 and Supplementary Fig. 3 for computational time comparison).

Beyond bulk or periodic systems—exploring dimensionality effects on CASTING's search performance

Low dimensional materials with their high surface to volume ratios present a unique opportunity to tap into properties that cannot be attained in the bulk form[17,57]. As the dimensionality of the atomic particles enters the regime of non-periodicity, the additional abundant surface (nanoclusters, layered materials), weak van der Waals interaction between the layers (2D) leads to electronic changes[58], that begins to play a dominant role in displaying exotic electronic and optical properties having potential in a multitude of applications such as semiconductor electronics[57,59–61], transport[62],

biotechnology[63], medicinal applications[64], systems with mechanical responses[65], etc. Like any atomistic system, the complexity of predicting low dimensional metastable phases increases with the size of the system. For zero-dimensional clusters, the variability in atomic packing[25], a wide energy landscape (from gaseous phase to condensed), and presence of isomorphism makes it challenging to efficiently explore the search space. While for layered two-dimensional materials, a weak van der Waals interaction, variability in coordination environment makes it possible to form numerous local minima that are hard to explore but with potential of having desired properties of interest. It is also worth mentioning that the existing knowledge of the nanoparticle structures does not reach the atom-level resolution from experiments[66]. Traditional diffraction techniques are more suitable for periodic crystalline structures. Thus, for decent accuracy of prediction, comparison of data from multifarious techniques is required[66]. In this regard, CASTING can provide a unified platform for predicting global and local minima of these low dimensional systems to bridge this gap between theory and experiments.

0D—exploring the size dependent diversity in Gold (Au) nanoclusters

We start by employing CASTING for the search of global minima of gold (Au) nanoclusters. Au nanoclusters due to their versatile applicability, have drawn significant attention over the years[61,67–69]. Computationally, most of the global minima of these nanoclusters of different sizes have been extensively explored[70]. Yet due to their relevance in modern-day material science[61], the optimization of these nanoclusters is of great interest. In this work, we use Sutton-Chen (10-8)[71] interatomic potential to recover the known global minima of gold (Au) nanoclusters[70] having 13 Atoms, 20 atoms, and 40 atoms respectively. Figure 8a shows the convergence of the MCTS optimizer for the 3 representative sizes of clusters used in this study. The tree hyperparameters for all the cases of search (refer to Supplementary Table 1) kept being the same. A LAMMPS package was used for the local minimization of the atomic configurations. For the 13-atom cluster which is known

to have an icosahedron structure as the global minima, the MCTS optimizer takes ~150 evaluations to converge to the solution which is considerably less. As their dimensionality increases with an increase in size, the iteration taken by the MCTS optimizer to converge to the global optimum also increases expectedly. For the system with 20 atoms, it takes ~300 iterations to reach the global minimum, while for the system with 40 atoms, MCTS requires ~20,000 evaluations to achieve the optimal solution. We also compare the global minima obtained by MCTS (Fig. 8b) with the known global minimas[70] for each of the sizes in terms of their structural features (pairwise interatomic distances in Fig. 8c). From Fig. 8c it can clearly be seen that, apart from being similar in terms of energy, MCTS optimized structures obtain identical structural similarity to their known counterparts. The overall results display the efficacy of the CASTING workflow in successfully scaling down from the bulk system to 0D systems while acknowledging the fact that the cluster systems are more difficult to optimize than their bulk counterpart because of additional degrees of freedom.

2D—exploring the global minima of two-dimensional Boron Nitride (h-BN) and Graphane (CH)

We next test the performance of the CASTING workflow in sampling two-dimensional (monolayer) systems with a rich compositional degree of freedom such as hexagonal Boron Nitride (h-BN) and Graphane (C-H). h-BN is an exceptional insulator with a direct wide bandgap of ~5–6 eV[72–74]. Being insulating and transparent, it has the potential of becoming an exceptional substrate for the synthesis of Graphene[74,75] and is thus also referred to as 'white graphene'[72]. The h-BN has covalently bonded Boron (B) and Nitrogen (N) atoms that crystalizes in a hexagonal P6₃/mmc space group. On the other hand, Graphane is the hydrogenated version of conductive semi-metal Graphene[76]. It is a fully saturated sp³ hydrocarbon with a 1:1 stoichiometry of C:H. Unlike Graphene, Graphane lacks the Dirac cone and also has an indirect bandgap of ~5.4 eV[77], hence behaving like an insulator. Still, its discovery, and structural attributes[78] have unraveled new paradigms in design of new semiconducting counterparts[77] with

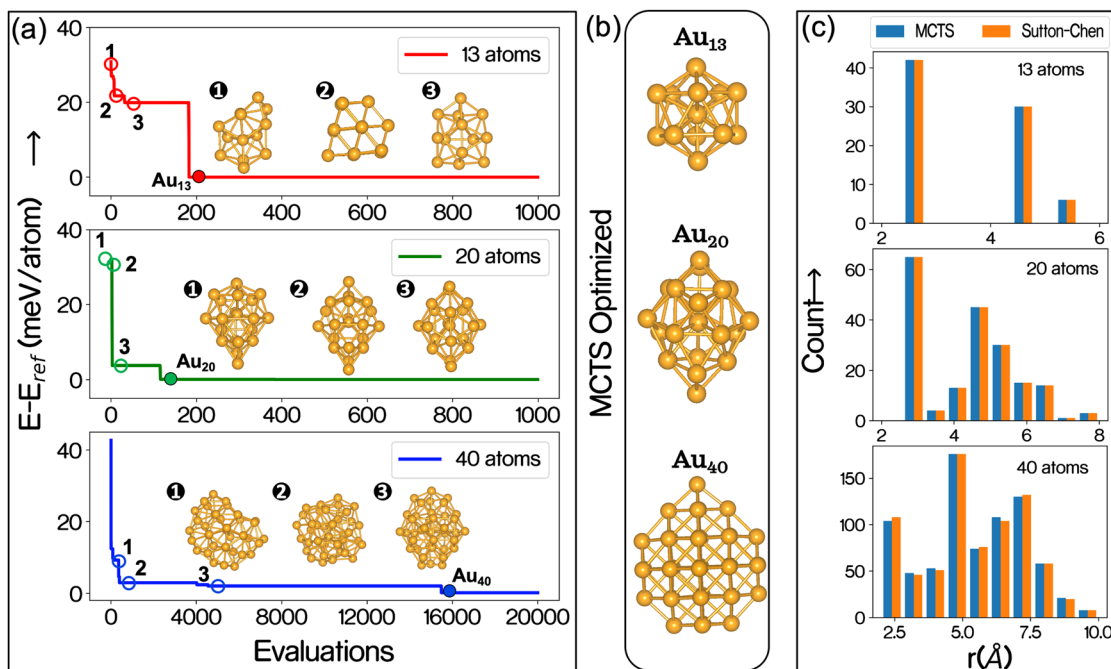**Fig. 8 Convergence with size-dependent diversity in nanoclusters of Gold (Au). a** The convergence of MCTS optimizer for the sampling of gold (Au) nanoclusters of different sizes (13, 20, and 40 atoms). **b** It shows the global minima obtained by MCTS for each case. **c** Comparison of the global minima obtained by MCTS to that of known Sutton-Chen global minima[70] in terms of pairwise distance between the atoms for 13, 20, and 40 atoms respectively.
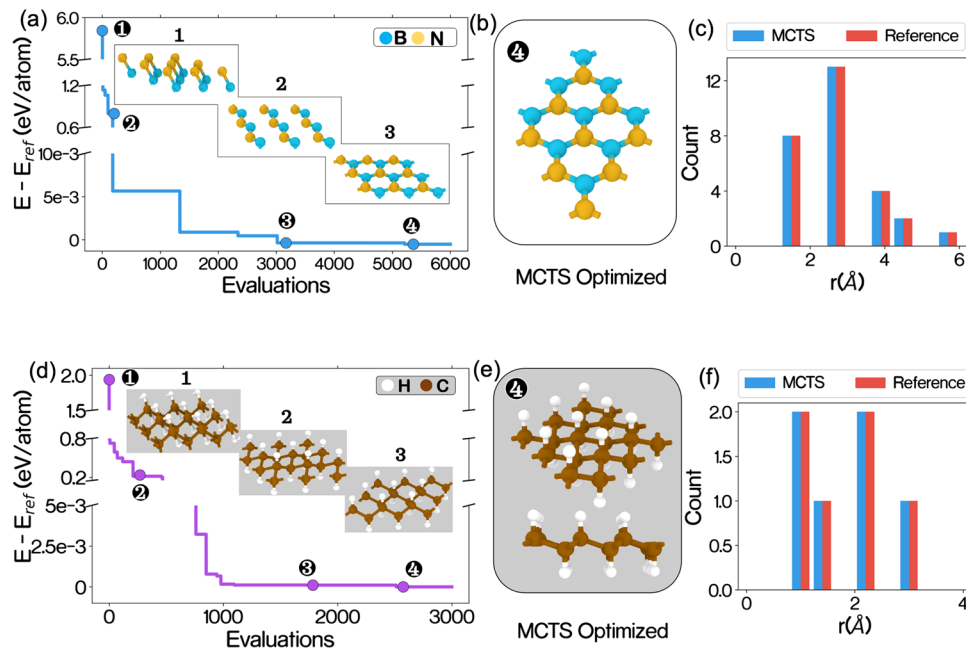
Fig. 9 Exploring 2D polymorphs with CASTING. a The convergence of MCTS optimizer for the sampling of Hexagonal Boron Nitride (h-BN). **b** It shows the global minima of h-BN obtained by MCTS. **c** Comparison of the global minima obtained by MCTS to that of known global minima of h-BN in terms of pairwise distance between the atoms. **d** The convergence of MCTS optimizer for the sampling of Graphane (CH). **e** This shows the global minima of Graphane obtained by MCTS. **f** Comparison of the global minima of Graphane obtained by MCTS to that of known global minima in terms of pairwise distance between the atoms.

exotic electronic properties. From the structural perspective, Graphane has two known conformations are of Chair and Boat type[78] with Chair (P3m1) being the global minima. For both systems, the knowledge of the various intermediate and key transition states can not only unravel crucial aspects of metastability but may also prove insights into the synthesis of these and other similar systems[77]—we deploy CASTING to explore the search for the monolayer h-BN and the chair polymorph of Graphane.

We use an extended Tersoff[74] type potential to model the interactions between B & N atoms and the popular AIREBO potential[79] for C & H atoms. We search for the unit cell of both h-BN and chair polymorph of Graphane. We conduct our search for the same number of atoms in the unit cell and keep the deviation in the lattice parameter (δ) from the known global minima to be 20%. Figure 9a–d shows the convergence of the MCTS optimizer with the number of relaxation evaluations of LAMMPS for h-BN and Graphane respectively. The total number of evaluations taken by the MCTS optimizer to converge to the global minima for h-BN is around ~3 K while that for Graphane is ~2.5 K. The target solution reaches the exact energy per atom value compared to its reference counterpart for both cases (Fig. 9a, d). We also compare the global minima of h-BN obtained by MCTS (Fig. 9c) to the known global minima[80] in terms of pairwise interatomic distances. The excellent match indicates that CASTING has achieved perfect accuracy structurally. A similar observation is seen in the case of the Graphane search (Fig. 9e, f). In addition to having more degrees of freedom, the compositional space adds more challenges to the optimization problem due to the inclusion of multiple species. Note that the increased diversity in the 2D conformations and a richer polymorphism also make it harder to reach the global minima for any search algorithm. This presence of multiple local minima in low dimensional systems translates into a higher number of iterations for MCTS to converge to the solution when compared to the bulk systems.

In Fig. 10, we further compare the performance of the CASTING workflow in predicting 2D polymorphs, such as Hexagonal Boron

Nitride (h-BN), and the global minima of Au nanoclusters (13 atoms) with commonly used optimizers for crystal structure prediction, namely GA, Basin Hopping[24,25], and Random search. Starting with the performance comparison for the prediction of global minima of h-BN, as shown in Fig. 10a, MCTS exhibits faster convergence to a global minimum compared to GA and Random search. The MCTS optimizer demonstrates improved convergence speed and solution accuracy. Similarly, for cluster optimization, the original methodology utilized for obtaining the global minima of Au nanoclusters is Basin Hopping. Thus, we compare the performance of MCTS with Basin Hopping and Random Search (Fig. 10b). Although all searches typically converge to a solution, given the small dimension of the search space, the error magnitude is in the range of ~10^{-8}. However, MCTS outperforms both Basin Hopping and Random search in terms of the final solution quality, as their performance saturates beyond ~2000 evaluations.

Exploring the compositional space of doped neodymium nickelate (NNO) using CASTING—elucidating the correlation between metastability and resistance states

We next deploy CASTING to explore an even more complex compositional landscape of a multi-component system, i.e., perovskite nickelates doped with hydrogen, and elucidate the relationship between metastability in doped NNO and their resistance states. Perovskite nickelate systems such as Neodymium Nickel oxide (NNO) can exhibit electronic properties that have immense potential in a multitude of applications[81,82]. The ground state NNO ($NdNiO_3$) is an orthorhombic perovskite structure with Ni atom bonded to O atom forming a corner-sharing NiO_6 octahedra[80]. A strongly correlated system NNO, however a metal at room temperature (refer to supplementary Fig. 4(a)), the addition of electron donors (H) in the lattice changes electrical conductivity extensively[82]. This makes it an exceptional candidate for being applicable in brained inspired computing[82,83]. Additional donated protons from H interstitials to the Ni not only impact its resistivity severely but also induces a complex potential
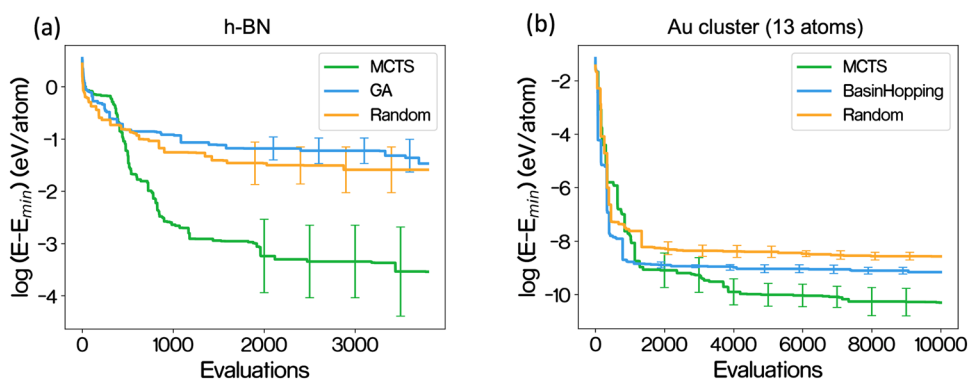
Fig. 10 Comparison of the performance of CASTING with commonly used optimizers in crystal structure prediction. a Average performance and standard deviation comparison (based on ten independent trials) between the MCTS optimizer, GA, and Random search for predicting Hexagonal Boron Nitride (h-BN). **b** Average performance and standard deviation comparison (based on ten independent trials) between the MCTS optimizer, Basin hopping, and Random search for predicting a nanocluster of Au (13 atoms).
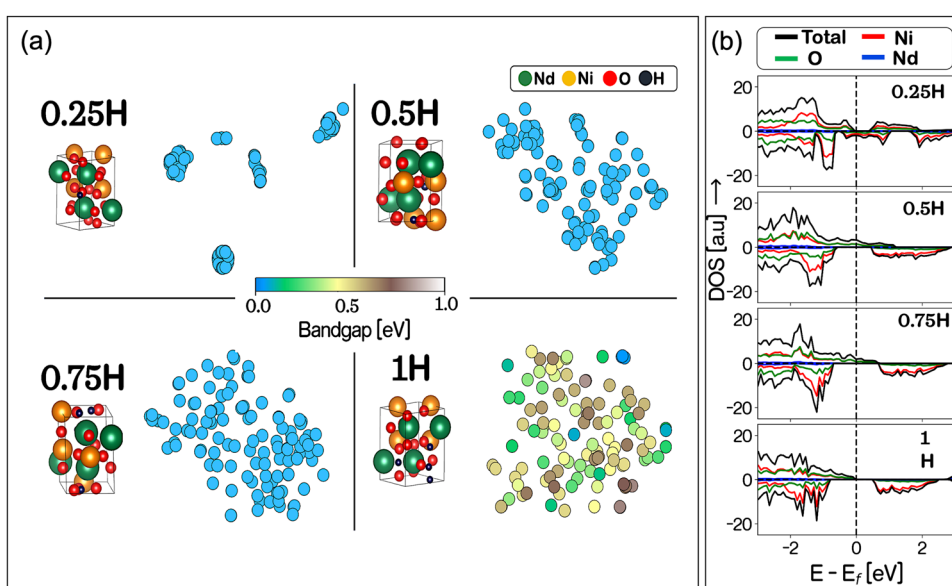
Fig. 11 Exploration of the configurational space of hydrogen doped Neodymium Nickel Oxide (NNO) with CASTING framework. a Shows the t-SNE (t distributed stochastic neighbor embedding) plot for SOAP feature representation of the sampled metastable polymorph at different concentration of hydrogen doping and their corresponding band gap magnitudes. **b** the typical density of states of sampled configurations at doping concentrations of 0.25H, 0.5H, and 0.75H, respectively.

energy surface with a plethora of local minima (metastable states). Additionally, there are two inequivalent O sites in the NNO lattice[80] providing permutational variability towards the location of H atoms. This makes it hard to locate the optimal position of the hydrogen (dopant) atoms in the lattice in search of favorable metastability for resistive switching. The task tends to become more challenging with an increasing concentration of dopants as the number of possible metastable states tends to grow exponentially.

To begin with, we select four concentrations of hydrogen doping 0.25H, 0.5H, 0.75H, and 1H per Ni atom respectively (Fig. 11a). We assume that there will be distortions in the NNO lattice upon insertion of H in it, the symmetry of the fundamental NNO lattice does not get broken even after ionic relaxation in VASP. So, during the sampling, we do not apply any external perturbation to the NNO lattice instead we move the H atoms through the lattice by perturbing its location. This allows us also to find possible locations or H sites in the lattice that alters the electronic structure by creating new eigenstates (Fig. 11b). A VASP package was used for structure relaxation and electronic calculation (refer to

supplementary note 5 for details). It is intuitive that with the increase in the concentration of doping the possibility of having unique metastable states increase drastically. This can also be observed in Fig. 11a. From the t-SNE (t distributed stochastic neighbor embedding) plot of SOAP[49] feature vector representation of the structures having a doping concentration of 0.25H (Fig. 11a), the distinction of the polymorphs in the feature space is not very conspicuous. As the doping concentration increase, the number of discrete and diverse polymorphs tends to grow. It is also very interesting, that the polymorphs having a doping concentration less than 1H, tend to show similar metallic behavior. As the doping concentration reaches 1H, the energy eigenstates vanish near Fermi energy (Fig. 11b) indicating a semiconducting behavior of the polymorphs. The trend persists for almost all the polymorphs sampled at this concentration. This application demonstrates the flexibility of our CASTING towards accurately performing tasks that go beyond simple crystal structure prediction while targeting specific properties of interest in complex material science problems.
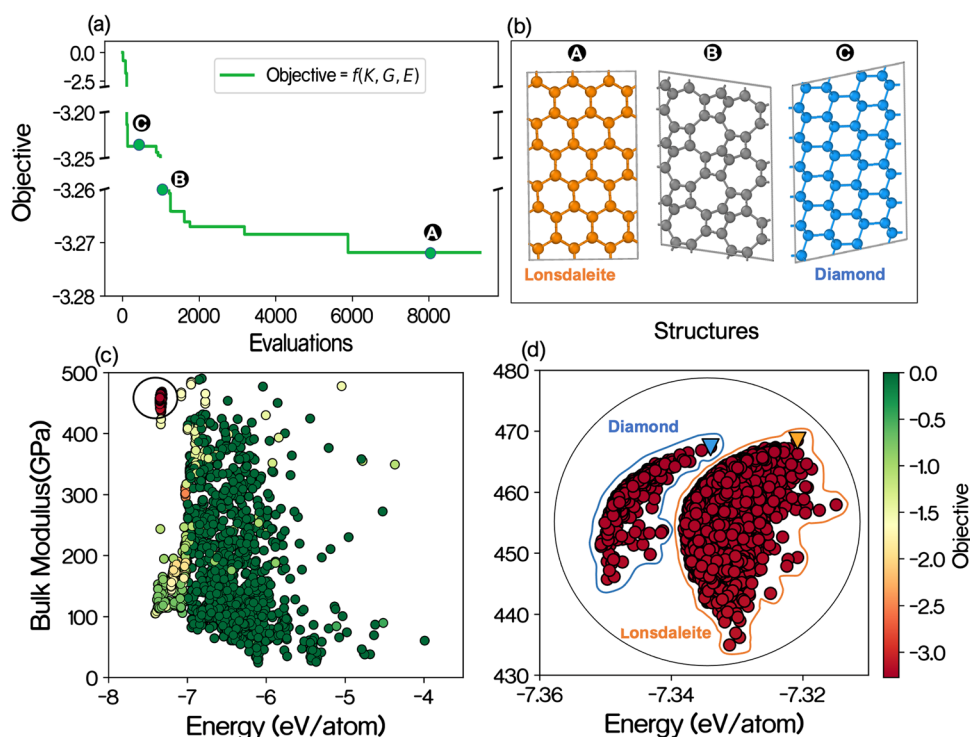
Fig. 12 Inverse design of super hard phases of Carbon (C). a Evolution of the overall objective function with the number of evaluations for the inverse design task. **b** Three phases of Carbon: Lonsdaleite, diamond, cubic diamond, and a metastable phase, which were sampled during the optimization stages depicted in (**a**). **c** Shows the bulk modulus of the sampled configurations and their corresponding energy. **d** This section provides a zoomed-in view of (**c**) and displays the two best configurations: cubic diamond (blue) and Lonsdaleite (orange), in terms of the objective, and similar variants of the configurations forming clusters in the representation space.

Inverse design of super hard phases of carbon through multi-objective optimization with a surrogate evaluator

Super hard materials play a crucial role in a wide range of applications[29,84–86]. Carbon can form two of the hardest known materials: cubic diamond and lonsdaleite[87,88]. Traditionally, diamond has been widely assumed to possess the highest hardness among Carbon polymorphs. However, theoretical studies have revealed that lonsdaleite, also referred to as hexagonal diamond, can exhibit even higher hardness than diamond. We employed CASTING and recovered the global minima of hexagonal diamond, using an objective function comprises of the bulk modulus (K), shear modulus (G) (evaluated using a graph neural network (GNN) model called CGCNN[34]), and cohesive energy (evaluated using empirical potential[52]). The CGCNN model was trained on materials project data[80]. Based on the materials project[80] data, lonsdaleite was found to have a higher elastic modulus compared to cubic diamond. Therefore, according to the trained surrogate model CGCNN, the global minimum for hardness is the hexagonal diamond (lonsdaleite).

CASTING converged to the global minima, the hexagonal diamond, within ~6000 evaluations (Fig. 12a). During the search, it sampled configurations with a wide range of bulk and shear modulus (Supplementary Fig. 5a). It should be noted that although the objective was not solely based on the energy, CASTING sampled polymorphs that are energetically stable (<−7 eV/atom). However, in Fig. 12c (and Supplementary Fig. 5b), for some of the energetically favorable configurations, the bulk and shear modulus are lower, while there is also a region (marked in a circle) with high bulk and shear modulus. Upon zooming into the region (Fig. 12d), it can be observed that two of the polymorphs having the highest objective value, the hexagonal diamond and the cubic diamond, are present and form clusters with similar variants. One noteworthy point here is that even though the cubic diamond polymorph with the best objective

values in Fig. 12d has a slightly lower bulk modulus than the hexagonal diamond counterpart, it is energetically more favorable. This holds true for all the other cubic and hexagonal diamond polymorphs. This further demonstrates the ability of the CASTING framework to find configurations that are not stable (e.g., metastable hexagonal diamond) while targeting specific properties (elastic modulus), making it suitable for inverse design.

DISCUSSION

In summary, we introduce CASTING which is a workflow that implements a continuous action space tree-based RL search algorithm for CSP in a high-dimensional search space. We discuss the important algorithmic modifications that are needed in the MCTS to successfully apply it to continuous search space inverse problems associated with structure and topology predictions. To showcase the efficacy of the CASTING framework, we apply CASTING to a wide range of representative systems—single-component metallic systems such as Ag and Au, covalent systems such as C, binary systems such as h-BN and C-H, and multi-component perovskite systems such as doped NNO. Additionally, we perform the inverse design of super-hard carbon phases using multi-objective optimization. We demonstrate the scalability, accuracy of sampling, and speed of convergence of CASTING on complex material science problems. We discuss the impact of the various RL hyperparameters on search performance. CASTING is also deployed to sample stable and metastable polymorphs across systems with dimensionality ranging from 3D (bulk) to low dimensional systems such as 0D (clusters) and 2D (sheets). Comparisons to other metaheuristic search algorithms such as genetic algorithms, basin hopping, and random sampling are also shown—the MCTS is demonstrated to have a superior performance in terms of the solution quality and the speed of convergence. We expect MCTS to perform well, especially for

complex search landscape with multiple objectives, multiple species, and multi-dimensional systems. Overall, we successfully demonstrate the development and application of an RL technique such as MCTS for inverse materials design and discovery problems related to structure and topology predictions.

METHODS

Monte Carlo Tree Search (MCTS) in continuous action space
Traditional vanilla Monte Carlo Tree Search (MCTS) has been applied to many materials' science problems[32,89,90] involving discreet spaces. But the continuous actions space adaptation for crystal structure prediction requires additional modifications. We have introduced the following to the MCTS to enable its application for continuous search space problems. These include:

Enhanced exploration and degeneracy protection
When performing a search of a very large phase space there can be a multitude of problems that arise which if not accounted for will result in the optimizer spending iterations on unnecessary solutions. In the case of crystal structure searches, there are two problems that can arise owing to the degeneracy of the search results. First, the optimizer can have two branches that initially start at two different positions in the phase space, yet they will converge into the same search location. This is effectively the algorithm retracing its steps repeatedly. The second problem which is more common in structural searches is that the natural entropy of the atomic positions can create many degenerate minima. For example, if one takes all the atoms in a structure and simply translates it a few angstroms in one direction the energy of the system has not changed (Translational invariance). As a result, when performing these searches, one may find a different parameter combination that results in an identical crystal structure. This degeneracy translates into MCTS spending computational cycles on solutions it has already seen before. We define a uniqueness function on the exploration side of the node selection rule to avoid degeneracies in the search space. For situations where we simply wish to limit two branches from approaching the same minima, we found a simple definition as outlined below should suffice:

$$f(\vec{r_i}) = \frac{1.5}{1 + \sum_{j \neq i}^{N_{points}} \delta(|r_i - r_j|)} \tag{1}$$

$$\delta(|r_i - r_j|) = \begin{cases} 1 & |r_i - r_j| < r_{max} \\ 0 & |r_i - r_j| \geq r_{max} \end{cases}$$

where r_{max} is the same r_{max} in the window depth scaling and $|r_i - r_j|$ is the distance between sample points i and j in the reduced parameter space. N_{points} is a count of the number of points generated by other nodes in the tree which also fall into the same area currently being searched by this node. This is a measure of the number of points that 'overlap' into another node's search area. The goal of this is to deprioritize nodes that are searching in a space that has already been searched by another node to prevent duplicate searches. The final node selection rule used is very similar to the classic UCT or UCB with a few key modifications which is called the Upper Confidence Bound for Parameters or UCP[91]. Equation (1) thus defines a uniqueness function on the exploration side of the node selection rule to avoid degeneracies in the search space—in a tree search operating in a continuous search space such as configurational search, there is often a possibility of the different branches converging to the same location in the search space which makes the overall search algorithm sluggish. To avoid this, Eq. (1) is effectively counting the number of points found within an area and scales the uniqueness with the number of points found within the same window. Since previously sampled points do not change their position, one only must keep a running tally of the number of points that have been sampled in the same area as a given node. This means that one only must update this function by comparing existing points to the newly added points which in practice is a very fast operation. Note that the design of this function is to scale the exploration side down toward 0 if the solutions are degenerate with what has already been discovered by the tree. In addition, when a node has a solution that is unique or located in a region that is under-explored, the function will scale to a higher value which promotes searches in these regions.

In reinforcement learning, the UCB (Upper Confidence Bound) technique balances exploration and exploitation by selecting the action with the highest estimated value and confidence bound. It helps find a trade-off between exploring new actions and exploiting known ones. Typical UCP is given

$$UCP(\theta_i) = -\min(p_1, p_2, \dots . p_{n_i}) + C \cdot f(\vec{r_i}) \cdot \sqrt{\frac{\log N_i}{n_i}} \tag{2}$$

Where θ_i represents node i in the MCTS structure, p is the reward for a given playout (calculated using Evaluators as in Fig. 3), C is the exploration constant, $f(\vec{r_i})$ is the uniqueness criteria value for this node, n_i is the number of playout samples taken by this node and all of its child nodes, and N_i is a similar value as n_i except it is the parent node's playout count instead of this node's. Note that $f(\vec{r_i})$ is the uniqueness function specifically introduced in our recent work and is equal to 1 in traditional MCTS settings.) Eq. (2) essentially tries to balance the search between those nodes in the tree which have either returned the maximum reward (left term) or have not been explored enough (right term). In contrast, the playout policy selects random actions (from a node) until the simulated episode is over. The reward is given as the best playout reward discovered as opposed to the average since the algorithm tries to find the best solution instead of the highest probability of winning like in many other MCTS formalisms. One can note that the choice of 'min' in the UCP indicates that the target property is being minimized. It can be 'max' (maximum of the node score) otherwise if the intention is to maximize the score or property (e.g., hardness).

Adaptive sampling in playouts
In discrete space searches such as board games, playouts are performed by randomly moving pieces to evaluate game scenarios ending in a victory or a loss. In a continuous action space, there is not a distinct 'win' scenario. Rather, playouts are viewed as a request for additional random sampling around a given point. When a node is selected for a playout, we perform random vector displacements from the parameter set contained in the node. This is akin to a random walk through the phase space that is guided by the MCTS algorithm. To allow the reinforcement learning to properly determine what path to take next, it is important to ensure that the generated sample points are high in quality. There are a great many stochastical traps that one can fall into depending on the sampling method. One such problem is when generating a vector that corresponds to a perturbation of the parameter space to create a new playout. If one were to use simple distributions such as an N-dimensional uniform, gaussian, etc., where each direction is generated from its own distribution, independent of all other variables, the probability of generating a large displacement increases with the number of parameters. The probability of generating a value between $(-3\sigma, 3\sigma)$ for a 1-dimensional gaussian is ~99%. For a 100-dimensional gaussian the probability of all values being found within 3σ is 0.99^{100} which is simply around 30%. This means the vast majority of vectors generated will have one or more extreme values. This problem becomes even more extreme as a larger number of parameters are introduced. As such better generation schemes are needed when creating points in a high-dimensional space. A simple and effective way to circumvent this is

to generate a vector uniformly on the surface of an N-Sphere of radius 1 and then uniformly pick the vector length. Since we pick within a distance, R, which is a collective variable, one can show that it is actually a biased distribution.

$$\int_0^{r_{max}} \boldsymbol{dr} = \int_0^{r_{max}} J(r)\rho(r)\boldsymbol{dr}$$

Where $J(r)$ is the radial component of the Jacobian for the polar coordinates and $\rho(r)$ is the probability density function. For visual simplicity, the normalization constant is neglected in this equation. This of course assumes that the angular components have already been fixed and thus integrated out. To have a distribution that is uniform on r, the product of the probability density function and the Jacobian must equal a constant. This of course implies

$$\rho(r) = \frac{1}{J(r)}$$

If we examine the radial component of the Jacobian for an N-Sphere we find it is simply given by

$$J(r) = r^{N-1}$$

As such the probability density function regardless of the number of dimensions must equal

$$\rho(r) = \frac{1}{J(r)} = \frac{1}{r^{N-1}}$$

This implies the probability distribution in Cartesian space is given by

$$\int_0^{r=r_{max}} \frac{1}{\left(\sum_{i=1}^N x_i^2\right)^{(N-1)/2}} dx_1 dx_2 \ldots dx_N$$

Thus, regardless of the number of dimensions, there will always be a reasonable probability of picking both large and small displacement vectors. This allows the reinforcement learning algorithm to determine the size of the vector needed to find a better reward function.

Exploitation in continuous action space

To facilitate exploration in a continuous search space, we must allow the algorithm to narrow in on a solution and eventually converge. Using a constant maximum vector length is seen to find a decent solution but remains highly inefficient. Too large a step size is no better than a random search whereas too small requires several node expansions to find a good solution. Additionally, within the tree, there was little correlation between the information stored in a node and the information stored inside its parent node. In a board game MCTS algorithm, each node contains a 'game state' i.e., the game piece's positions on the board. A child node is related to its parent by the fact that you can obtain the child's position by moving a single piece from the parent's position. Restoring this correlation is paramount to have the MCTS algorithm formalism make any logical sense in addition to ensuring that its results are consistent.

We introduce a window scaling scheme (Fig. 2c). Initially, the search space starts has bounds $[a_{1,min}, a_{1,max}]$ and $[a_{2,min}, a_{2,max}]$ respectively. And the largest vector distance r_{max}, corresponding to the sampling radius of the hypersphere that can be generated is given as, r_1. This radius is assigned to smaller and smaller values with the increasing depth of the corresponding node in the MCTS tree (Fig. 2c)). The reduction is done following a gaussian curve using the equation

$$r = \begin{cases} r_{max} * \exp\left(-a * \left(\frac{depth}{maxdepth}\right)^2\right), & depth \leq maxdepth \\ 0, & depth \geq maxdepth \end{cases} \quad (3)$$

'a' is the tunable parameter. The telescoping window scaling approach ensures that the algorithm is incrementally refining the phase space. This allows the algorithm to initially make larger scans of the phase space and as it finds interesting regions it is allowed to zoom in on those regions and begin exploring in more detail. Restoring the correlation between the parent and child node in that a child node is a zoomed-in region around the parent node, it gives the algorithm some direction such that the algorithm is not simply performing a purely random walk, and it also allows it to converge sufficiently close to an optimal solution since it is making smaller and smaller adjustments as it expands the tree depth.

REFERENCES
1. Oganov, A. R. Crystal structure prediction: reflections on present status and challenges. *Faraday Discuss* **211**, 643–660 (2018).
2. Oganov, A. R., Pickard, C. J., Zhu, Q. & Needs, R. J. Structure prediction drives materials discovery. *Nat. Rev. Mater.* **4**, 331–348 (2019).
3. Fischer, C. C., Tibbetts, K. J., Morgan, D. & Ceder, G. Predicting crystal structure by merging data mining with quantum mechanics. *Nat. Mater.* **5**, 641–646 (2006).
4. Oganov, A. R. Modern Methods of Crystal Structure Prediction. (John Wiley & Sons, 2011).
5. Oganov, A. R., Ma, Y. M., Lyakhov, A. O., Valle, M. & Gatti, C. Evolutionary crystal structure prediction as a method for the discovery of minerals and materials. *Rev. Miner. Geochem.* **71**, 271–298 (2010).
6. Pannetier, J., Bassasalsina, J., Rodriguezcarvajal, J. & Caignaert, V. Prediction of crystal-structures from crystal-chemistry rules by simulated annealing. *Nature* **346**, 343–345 (1990).
7. Ryan, K., Lengyel, J. & Shatruk, M. Crystal structure prediction via deep learning. *J. Am. Chem. Soc.* **140**, 10158–10168 (2018).
8. Wang, Y. C., Lv, J., Zhu, L. & Ma, Y. M. CALYPSO: a method for crystal structure prediction. *Comput. Phys. Commun.* **183**, 2063–2070 (2012).
9. Revard, B. C., Tipton, W. W. & Hennig, R. G. Genetic Algorithm for Structure and Phase Prediction. GitHub repository (2018).
10. Wang, Y. C., Lv, J. A., Zhu, L. & Ma, Y. M. Crystal structure prediction via particle-swarm optimization. *Phys. Rev. B* **82**, 094116 (2010).
11. Woodley, S. M. & Catlow, R. Crystal structure prediction from first principles. *Nat. Mater.* **7**, 937–946 (2008).
12. Whitley, D. A genetic algorithm tutorial. *Stat. Comput.* **4**, 65–85 (1994).
13. Van Laarhoven, P. J. M. & Aarts, E. H. L. In: Simulated Annealing: Theory and Applications 7–15 (Springer, 1987).
14. Chen, S., Montgomery, J. & Bolufe-Rohler, A. Measuring the curse of dimensionality and its effects on particle swarm optimization and differential evolution. *Appl. Intell.* **42**, 514–526 (2015).
15. Kennedy, J. & Eberhart, R. Particle swarm optimization. In: *Proc. International Conference on Neural NETWorks (ICNN'95)*. 1942–1948 (IEEE, 1995).
16. Urusov, V. S., Dubrovinskaya, N. A. & Dubrovinsky, L. S. *Generation of likely crystal structures of minerals. Moscow State University Press, Moscow. Valle M.(2005). STM3: a chemistry visualization platform. Z. Krist.* **220**, 585–588 (1990).
17. Wilcoxon, J. P. & Abrams, B. L. Synthesis, structure and properties of metal nanoclusters. *Chem. Soc. Rev.* **35**, 1162–1194 (2006).
18. Goedecker, S. Minima hopping: an efficient search method for the global minimum of the potential energy surface of complex molecular systems. *J. Chem. Phys.* **120**, 9911–9917 (2004).

19. Martonak, R., Donadio, D., Oganov, A. R. & Parrinello, M. Crystal structure transformations in SiO_2 from classical and ab initio metadynamics. *Nat. Mater.* **5**, 623–626 (2006).

20. Martoňák, R. et al. Simulation of structural phase transitions by metadynamics. *Z. Kristallogr. Crystal. Mater.* **220**, 489–498 (2005).

21. Pickard, C. J. & Needs, R. J. High-pressure phases of silane. *Phys. Rev. Lett.* **97**, 045504 (2006).

22. Pickard, C. J. & Needs, R. J. Ab initio random structure searching. *J. Phys. Condens. Matter* **23**, 053201 (2011).

23. Manna, S. et al. A database of low-energy atomically precise nanoclusters. *Sci. Data* **10**, 308 (2023).

24. Wales, D. J. & Doye, J. P. K. Global optimization by basin-hopping and the lowest energy structures of Lennard-Jones clusters containing up to 110 atoms. *J. Phys. Chem. A* **101**, 5111–5116 (1997).

25. Wales, D. J. & Scheraga, H. A. Global optimization of clusters, crystals, and biomolecules. *Science* **285**, 1368–1372 (1999).

26. Lonie, D. C. & Zurek, E. XtalOpt: an open-source evolutionary algorithm for crystal structure prediction. *Comput. Phys. Commun.* **182**, 372–387 (2011).

27. Castillo, R. et al. Germanium dumbbells in a new superconducting modification of BaGe3. *Inorg. Chem.* **55**, 4498–4503 (2016).

28. Gou, H. et al. Discovery of a superhard iron tetraboride superconductor. *Phys. Rev. Lett.* **111**, 157002 (2013).

29. Dwivedi, N. et al. Unusual high hardness and load-dependent mechanical characteristics of hydrogenated carbon-nitrogen hybrid films. *ACS Appl. Mater. Interfaces* **14**, 20220–20229 (2022).

30. Banik, S. et al. CEGANN: Crystal Edge Graph Attention Neural Network for multiscale classification of materials environment. *Npj Comput. Mater.* **9**, 23 (2023).

31. Loeffler, T. D., Banik, S., Patra, T. K., Sternberg, M. & Sankaranarayanan, S. K. R. S. Reinforcement learning in discrete action space applied to inverse defect design. *J. Phys. Commun.* **5**, 031001 (2021).

32. Banik, S. et al. Learning with delayed rewards-a case study on inverse defect design in 2D materials. *ACS Appl. Mater. Interfaces* **13**, 36455–36464 (2021).

33. Le, T., Epa, V. C., Burden, F. R. & Winkler, D. A. Quantitative structure-property relationship modeling of diverse materials properties. *Chem. Rev.* **112**, 2889–2919 (2012).

34. Xie, T. & Grossman, J. C. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Phys. Rev. Lett.* **120**, 145301 (2018).

35. Manna, S., Wang, M. Y., Barbu, A. & Ciobanu, C. V. Machine-learning of piezoelectric coefficients for wurtzite crystals. *Mater. Manuf. Process.* **38**, 1–12 (2023).

36. Banik, S., Balasubramanian, K., Manna, S., Derrible, S. & Sankaranarayananan, S. Evaluating generalized feature importance via performance assessment of machine learning models for predicting elastic properties of materials. ChemRxiv. Cambridge (2023).

37. Cohen, A. J., Mori-Sanchez, P. & Yang, W. Challenges for density functional theory. *Chem. Rev.* **112**, 289–320 (2012).

38. Chen, Y. et al. Pressure-induced phase transformation in β-eucryptite: an X-ray diffraction and density functional theory study. *Scr. Mater.* **122**, 64–67 (2016).

39. Silver, D. et al. Mastering the game of Go with deep neural networks and tree search. *Nature* **529**, 484–489 (2016).

40. Silver, D. et al. A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. *Science* **362**, 1140–1144 (2018).

41. Wang, X. et al. Towards efficient discovery of green synthetic pathways with Monte Carlo tree search and reinforcement learning. *Chem. Sci.* **11**, 10959–10972 (2020).

42. Popova, M., Isayev, O. & Tropsha, A. Deep reinforcement learning for de novo drug design. *Sci. Adv.* **4**, eaap7885 (2018).

43. Srinivasan, S. et al. Artificial intelligence-guided de novo molecular design targeting COVID-19. *ACS Omega* **6**, 12557–12566 (2021).

44. Batra, R. et al. Machine learning overcomes human bias in the discovery of selfassembling peptides. *Nat. Chem.* **14**, 1427–1435 (2022).

45. Dulac-Arnold, G. et al. Deep reinforcement learning in large discrete action spaces. *arXiv preprint arXiv:1512.07679* (2015).

46. Manna, S. et al. Learning in continuous action space for developing high dimensional potential energy models. *Nat. Commun.* **13**, 368 (2022).

47. Plimpton, S. Fast parallel algorithms for short-range molecular-dynamics. *J. Comput. Phys.* **117**, 1–19 (1995).

48. Kresse, G. & Hafner, J. Ab initio molecular dynamics for liquid metals. *Phys. Rev. B Condens. Matter* **47**, 558–561 (1993).

49. Bartók, A. P., Kondor, R. & Csányi, G. On representing chemical environments. *Phys. Rev. B* **87**, 184115 (2013).

50. Williams, P. L., Mishin, Y. & Hamilton, J. C. An embedded-atom potential for the Cu-Ag system. *Modell. Simul. Mater. Sci. Eng.* **14**, 817–833 (2006).

51. Stukowski, A. Structure identification methods for atomistic simulations of crystalline materials. *Modell. Simul. Mater. Sci. Eng.* **20**, 045021 (2012).

52. Los, J. H., Ghiringhelli, L. M., Meijer, E. J. & Fasolino, A. Improved long-range reactive bond-order potential for carbon. *Constr. Phys. Rev. B* **72**, 214102 (2005).

53. Murtagh, F. & Legendre, P. Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion? *J. Classif.* **31**, 274–295 (2014).

54. Balasubramanian, M. & Schwartz, E. L. The isomap algorithm and topological stability. *Science* **295**, 7 (2002).

55. Ghiringhelli, L. M., Los, J. H., Meijer, E. J., Fasolino, A. & Frenkel, D. Modeling the phase diagram of carbon. *Phys. Rev. Lett.* **94**, 145701 (2005).

56. Srinivasan, S. et al. Machine learning the metastable phase diagram of covalently bonded carbon. *Nat. Commun.* **13**, 3251 (2022).

57. Wang, Q. H., Kalantar-Zadeh, K., Kis, A., Coleman, J. N. & Strano, M. S. Electronics and optoelectronics of two-dimensional transition metal dichalcogenides. *Nat. Nanotechnol.* **7**, 699–712 (2012).

58. Geim, A. K. & Grigorieva, I. V. Van der Waals heterostructures. *Nature* **499**, 419–425 (2013).

59. Iannaccone, G., Bonaccorso, F., Colombo, L. & Fiori, G. Quantum engineering of transistors based on 2D materials heterostructures. *Nat. Nanotechnol.* **13**, 183–191 (2018).

60. Schmid, G. et al. Current and future applications of nanoclusters. *Chem. Soc. Rev.* **28**, 179–185 (1999).

61. Chakraborty, I. & Pradeep, T. Atomically precise clusters of noble metals: emerging link between atoms and nanoparticles. *Chem. Rev.* **117**, 8208–8271 (2017).

62. Shao, C., Yu, X. X., Yang, N., Yue, Y. A. & Bao, H. A review of thermal transport in low-dimensional materials under external perturbation: effect of strain, substrate, and clustering. *Nanoscale Microscale Thermophys. Eng.* **21**, 201–236 (2017).

63. Luo, Z., Zheng, K. & Xie, J. Engineering ultrasmall water-soluble gold and silver nanoclusters for biomedical applications. *Chem. Commun.* **50**, 5143–5155 (2014).

64. Zhang, X. D. et al. Ultrasmall Au(10-12)(SG)(10-12) nanomolecules for high tumor specificity and cancer radiotherapy. *Adv. Mater.* **26**, 4565–4568 (2014).

65. Dai, Z., Liu, L. & Zhang, Z. Strain engineering of 2D materials: issues and opportunities at the interface. *Adv. Mater.* **31**, e1805417 (2019).

66. Malola, S. et al. A method for structure prediction of metal-ligand interfaces of hybrid nanoparticles. *Nat. Commun.* **10**, 3973 (2019).

67. Arvizo, R., Bhattacharya, R. & Mukherjee, P. Gold nanoparticles: opportunities and challenges in nanomedicine. *Expert Opin. Drug Deliv.* **7**, 753–763 (2010).

68. Zeng, S. W. et al. A review on functionalized gold nanoparticles for biosensing applications. *Plasmonics* **6**, 491–506 (2011).

69. Sperling, R. A., Rivera Gil, P., Zhang, F., Zanella, M. & Parak, W. J. Biological applications of gold nanoparticles. *Chem. Soc. Rev.* **37**, 1896–1908 (2008).

70. Doye, J. P. K. & Wales, D. J. Global minima for transition metal clusters described by Sutton-Chen potentials. *N. J. Chem.* **22**, 733–744 (1998).

71. Todd, B. D. & Lyndenbell, R. M. Surface and bulk properties of metals modeled with sutton-chen potentials. *Surf. Sci.* **281**, 191–206 (1993).

72. Zeng, M., Xiao, Y., Liu, J., Yang, K. & Fu, L. Exploring two-dimensional materials toward the next-generation circuits: from monomer design to assembly control. *Chem. Rev.* **118**, 6236–6296 (2018).

73. Watanabe, K., Taniguchi, T. & Kanda, H. Direct-bandgap properties and evidence for ultraviolet lasing of hexagonal boron nitride single crystal. *Nat. Mater.* **3**, 404–409 (2004).

74. Los, J. H. et al. Extended Tersoff potential for boron nitride: Energetics and elastic properties of pristine and defective h-BN. *Phys. Rev. B* **96**, 184108 (2017).

75. Zastrow, M. Meet the crystal growers who sparked a revolution in graphene electronics. *Nature* **572**, 429–432 (2019).

76. Elias, D. C. et al. Control of graphene's properties by reversible hydrogenation: evidence for graphane. *Science* **323**, 610–613 (2009).

77. Koski, K. J. & Cui, Y. The new skinny in two-dimensional nanomaterials. *ACS Nano* **7**, 3739–3743 (2013).

78. Sofo, J. O., Chaudhari, A. S. & Barber, G. D. Graphane: a two-dimensional hydrocarbon. *Phys. Rev. B* **75**, 153401 (2007).

79. Stuart, S. J., Tutein, A. B. & Harrison, J. A. A reactive potential for hydrocarbons with intermolecular interactions. *J. Chem. Phys.* **112**, 6472–6486 (2000).

80. Jain, A. et al. Commentary: The Materials Project: a materials genome approach to accelerating materials innovation. *Appl. Mater.* **1**, 011002 (2013).

81. Chen, J. K. et al. Pressure induced unstable electronic states upon correlated nickelates metastable perovskites as batch synthesized via heterogeneous nucleation. *Adv. Funct. Mater.* **30**, 2000987 (2020).

82. Zhang, H. T. et al. Reconfigurable perovskite nickelate electronics for artificial intelligence. *Science* **375**, 533–539 (2022).

83. Park, T. J. et al. Complex oxides for brain-inspired computing: a review. *Adv. Mater.* **385**, 2203352 (2022).

84. Vepřek, S. The search for novel, superhard materials. *J. Vac. Sci. Technol. A Vac. Surf. Films* **17**, 2401–2420 (1999).

85. Mazhnik, E. & Oganov, A. R. Application of machine learning methods for predicting new superhard materials. *J. Appl. Phys.* **128**, 075102 (2020).

86. Veprek, S. & Veprek-Heijman, M. J. G. Industrial applications of superhard nanocomposite coatings. *Surf. Coat. Technol.* **202**, 5063–5073 (2008).

87. Frondel, C. & Marvin, U. B. Lonsdaleite, a hexagonal polymorph of diamond. *Nature* **214**, 587–589 (1967).
88. Qingkun, L., Yi, S., Zhiyuan, L. & Yu, Z. Lonsdaleite–a material stronger and stiffer than diamond. *Scr. Mater.* **65**, 229–232 (2011).
89. Patra, T. K., Loeffler, T. D. & Sankaranarayanan, S. Accelerating copolymer inverse design using monte carlo tree search. *Nanoscale* **12**, 23653–23662 (2020).
90. Kiyohara, S. & Mizoguchi, T. Searching the stable segregation configuration at the grain boundary by a Monte Carlo tree search. *J. Chem. Phys.* **148**, 241741 (2018).
91. Browne, C. B. et al. A survey of Monte Carlo tree search methods. *IEEE Trans. Comput. Intell. AI Games* **4**, 1–43 (2012).

ACKNOWLEDGEMENTS

AUTHOR CONTRIBUTIONS

S.B. and S.K.R.S. conceived the project. S.B. developed the CASTING workflow with input from T.L. S.K.R.S., S.S. and H.C. provided feedback on the CASTING workflow and the crystal structure search/validation. S.B. and S.M. incorporated the use of CASTING in conjunction with the external software package for DFT (VASP). S.B. performed all the calculations. S.B. evaluated the performance of the workflow for different representative systems and analyzed the results with input from S.K.R.S. and P.D. S.B. and S.K.R.S. wrote the manuscript with input from all co-authors. S.M. and H.C. provided feedback on the data analysis and the target systems. All authors participated in discussing the results and provided comments on the various sections of the manuscript. S.K.R.S. supervised and directed the overall work.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41524-023-01128-y.

Correspondence and requests for materials should be addressed to Subramanian K. R. S. Sankaranarayanan.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.