Taylor & Francis
Taylor & Francis Group

Materials Screening Through GPU Accelerated Topological Mapping

Branden B. Kappes and Cristian V. Ciobanu

Department of Mechanical Engineering, Colorado School of Mines, Golden, CO, USA

Selecting materials with properties tailored to a specific application may be accelerated through materials informatics, but kinetic properties whose calculations are too computationally intensive to be incorporated into materials screening must be replaced with appropriate descriptors. Here we present a highly optimized method for general processing on graphics processing hardware through which we map the interstitial subspace of atomic structures that are used as a qualitative predictor for diffusivity. Additionally, analytical methods that determine the largest channel diameter and identify the optimal path through a material are proposed to characterize this topology. Analysis of the interstitial subspace, along with the theoretical capacities for Li ions, has lead us to select high-capacity lithium ion battery (LIB) materials that display both promising capacities and migration pathways able to support lithium insertion and removal. The result is the identification of Li_2MgSi, a LIB anode material with a theoretical capacity of 1023 Ah/kg, from an unfiltered set of 1,754 structures.

Keywords Accessible volume; CUDA; High throughput; GPU; Lithium ion batteries; Materials informatics; Pore characterization; Single-instruction multiple-data (SIMD).

INTRODUCTION

Screening for materials with specific properties using high-throughput thermodynamic modeling is becoming increasingly widespread [1–6], but kinetic properties must also be considered if viable candidate materials are to be found. In many applications, the diffusivity of a guest species in a host structure is of paramount importance. However, identifying the migration pathways available for diffusion by visualizing model structures is neither rigorous nor amenable to high-throughput screening.

Determining migration pathways is a matter of tracing a low-energy route that a guest atom may take through a host structure, such as a crystal lattice. This is particularly effective for systems where the migration of a guest species through the host lattice is predominantly steric, and for that reason, the zeolite community has used analysis of the empty space, e.g., the pore or interstitial subspace, to select metal-organic frameworks for gas reforming, gas separation, and hydrocarbon activation [7–15]. These have focused specifically on properties such as adsorption area, available pore volume, and largest channel diameter (LCD).

Several techniques have been developed to calculate these properties, including Voronoi-Dirichlet polyhedra tessellations (whose analysis is assisted by the Delaunay circumspheres) [16–18] and the collection of polyhedral primitives (i.e., chemical hieroglyphs) [19]. The most conceptually simple approach partitions the unit cell of the crystal into a regular grid, as described by Haldoupis [10], and a closely related variant that identifies continuously accessible pore space through a moving front algorithm [11]. When carried out in serial codes, the latter approaches require orders of magnitude more time than the tessellation methods, but benefit from an intrinsic parallel efficiency that allows them to make use of both multicore CPU and general processing GPU (GPGPU) resources, which brings their total computational cost below that of the tessellation models.

Furthermore, grid structures are amenable to complex cost functions that produce scalar-valued, or even vector-valued, fields. The spatial distribution of the interstitial subspace can be enhanced through the inclusion of crystallochemical information, such as the potential energy of a included guest atom, force mapping, or mapping of electrostatic interactions. With the increased computational cost of these functions comes a reduced ability to rapidly screen through all potential chemistries and crystal structures. Nonetheless, grids permit such screening under a common programmatic framework, which opens the door to hierarchical screening workflows.

Identifying candidate electrode materials for lithium ion battery (LIB) applications is a challenge that is particularly well-suited to high-throughput materials screening. Genetic algorithms, while they represent an excellent and efficient way to determine optimal crystal structures under particular conditions [20–23] or to solve certain engineering problems that involve a large parameter space [24], they are by no means a high-throughput technique for screening as most of them rely on computing directly an expensive cost function, usually energy. Many ongoing efforts focus on the voltage and specific capacity [5, 25, 26], which are set by the thermodynamics of the reversible lithiation reaction, but the power density, and ultimately the viability of a new electrode depends on the kinetics of lithium insertion and removal. Intercalation electrodes, such as $LiCoO_2$ and graphite, are layered structures that

Address correspondence to Cristian V. Ciobanu, Department of Mechanical Engineering, Colorado School of Mines, 1500 Illinois St., Golden, CO 80401, USA; E-mail: cciobanu@mines.edu

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/lmmp.

host lithium between those layers, but otherwise their morphology remains largely unchanged between the lithiated and delithiated states [27]. In contrast, alloying electrodes form a new lithium-containing phase, for example, the lithiation of silicon to form $Li_{15}Si_4$, and promise considerably higher capacities than intercalation electrodes [28, 29], but at the cost of substantial change to the host crystal structure.

The specific capacity, in Ah/kg, of a lithium ion cell is determined by the capacity of the anode, that of the cathode, and is reduced by the mass of the supporting structures, such as the casing, current collectors, electrolyte, and separators according to

$$\frac{1}{C_{cell}} = \frac{1}{C_{anode}} + \frac{1}{C_{cathode}} + \frac{1}{\gamma},$$

where $1/\gamma = 1/C_{cell}^{ref} - 1/C_{anode}^{ref} - 1/C_{cathode}^{ref}$ for a reference cell of identical form factor. While a ten-fold increase in the anode capacity will increase the cell capacity by over 20%, all else unchanged, half of this benefit can be realized by increasing the anode capacity by a factor <3 to 1000 Ah/kg. Beyond this increase, the specific capacity of the cell becomes limited by the cathode capacity and the supporting structures (Fig. 1). Silicon, tin, transition metal oxides and halides, and silicon oxides have been suggested as potential replacements for graphite, but many of these, though their capacity makes them thermodynamically attractive, are kinetically unsuitable for practical applications [30].

Advancement in high-throughput materials screening depends on the rapid assimilation of data produced by different research groups working in the field. This flexibility is best handled through module development in an extensible programming language, which has driven the success of, for example, pymatgen [32] and
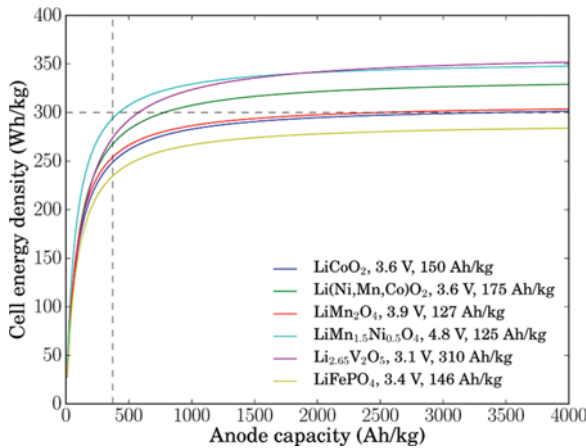

FIGURE 1.—The energy density of the cell shows two distinct regimes. Below an anode capacity of ∼500 Ah/kg, the energy density of the cell can be significantly improved by increasing the capacity of the anode. The vertical dashed line at 372 Ah/kg is the specific capacity of the most commonly used anode material, graphite. The effect of the supporting materials are based on an 18650 cell geometry, which sets the upper bound capacity at 195.74 Ah/kg [31].

the atomistic simulation environment [33]. In order to utilize the growing toolset available in these and other materials informatics frameworks, we have implemented the present approach to be accessible through python.

MATERIALS AND METHODS

We have used a three-step process to screen candidate electrode materials for LIB applications. First, a set of materials is examined by topological mapping of the interstitial subspace. Subsequently, the LCD of each structure is calculated and compared to materials known to perform reasonably well as LIB electrodes. Finally, materials with an insufficient LCD or insufficient capacity are disqualified from the set.

Topological mapping provides a method for identifying possible migration pathways through either crystalline or noncrystalline materials. Only pathways that are continuous across the simulation cell, and are therefore accessible to long range diffusion, are practical for mass transport. Therefore, analysis of topological maps must include methods for determining both the size and the continuity of pores that span a certain structure. Possible kinetic pathways are identified through calculation of a low-cost path between points in the cell.

In the following analyses, the structure is assumed to be a cubic supercell. This is not a limitation, because for any triclinic, periodic supercell with column edge vectors $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$ that define $\mathbf{X} \equiv \mathbf{abc}$, there exists an affine transformation $\mathbf{S} = \mathbf{X}^{-1}$ that transforms the triclinic supercell into a unit cube supercell. All atom positions, $\mathbf{x}_i$ transform to fractional (scaled) coordinates $\mathbf{s}_i = \mathbf{S}\mathbf{x}_i$, and distance vectors in fractional coordinates $\mathbf{s}_{ij} = \mathbf{s}_j - \mathbf{s}_i$ return to cartesian coordinates $\mathbf{r}_{ij} = \mathbf{X}\mathbf{s}_{ij} = \mathbf{x}_j - \mathbf{x}_i$ for measuring real-space distances.

The use of scaled coordinates also greatly simplifies calculations across periodic boundary conditions. The shortest real-space distance between two points in the supercell can be computed from the scaled coordinates representation:

$$\min |\mathbf{r}_{ij}| = \left| \mathbf{X} \cdot \left(\lfloor \mathbf{s}_{ij} + \mathbf{0.5} \rfloor - \mathbf{0.5} \right) \right|,$$

where $\mathbf{0.5}$ is a vector with each component equal to 0.5.

The interstitial subspace is mapped as a scalar field to a regular spatial grid, as shown schematically in Fig. 2. Each volume element in the grid (voxel) stores the Euclidean distance to the nearest atom in the host lattice.

Without requiring an *a priori* knowledge of the guest species, this approach provides the spacing for a guest species size equal to the voxel size. The computed distance at every point in the grid accounts for the finite radius of the host atoms, which can be set at runtime to their atomic, covalent, van der Waals, or ionic radii. For spherically symmetric guest species, the resulting grid may be processed for a different guest without incurring the cost of an additional calculation.

The computational expense of the distance calculation is mitigated through the use of a KD-tree to optimize atom-voxel nearest-neighbor searches, and is parallelized
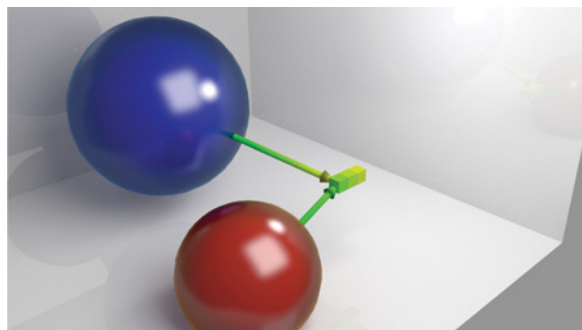
FIGURE 2.—Schematic showing how the interstitial subspace is mapped as a scalar field to a regular spatial grid. The distance to each atom in the cell is calculated at each volume element (voxel), shown as the green/yellow vector extending from the blue and red spheres. The color along the vector becomes increasingly yellow in proportion to the distance from the atom surface. Each voxel is assigned the distance to the nearest atom, e.g., the voxel has the same color as the shortest vector.

over voxels using the python *multiprocessing* package. Array operations are vectorized through the use of *numpy*, vectorized functions, and manually unrolled loops. This reduces the search for the atom nearest a voxel to $O(\log N_{atoms})$. The reduced computational efficiency of python, as compared to C++, is very nearly recovered by the increased search time for an $O(N)$ brute force neighbor search, as used by Martin et al. [17] In python implementations of both the brute force and optimized approaches, the time to map the interstitial subspace on a 1-million voxel grid was reduced from ~1800 s (serial) to 27.42 s, respectively.

The GPGPU kernel optimizes memory access through shared scaled atom positions and flattened (one-dimensional) representations of all multidimensional arrays. Atomic operations are eliminated by grid initialization and declaring *volatile* the global grid data. Atom–voxel distance calculations are fully asynchronous. This algorithm was exposed in python using Klöckner's PyCUDA module [34]. Precompiling the CUDA kernel eliminates compilation overhead.

Underutilization of the GPU cores is due to the limited asynchronicity of the kernel, the main limitations being the parallel efficiency of reduction algorithms (to find the minimum distance) and the fixed number of threads per warp. The simplicity of the distance calculation, the only fully asynchronous method, means that more time will be spent on, e.g., distance calculations and minimizations. Threads, the most basic computational unit, may be divided across either atoms or dimensions, and are grouped into 32-thread warps – threads whose execution occurs synchronously in a single instruction, multiple thread (SIMT) fashion. Collections of warps are further subdivided into blocks. Both dimension-per-thread and atom-per-thread schemes are tested. In both cases, voxels are divided over blocks, one voxel per block.

Earlier implementations refer to pores as those points where the guest-host distance is greater than zero [17]. However, predicting likely diffusion pathways through a compact crystal, e.g., diamond silicon or cubic close-packed

iron, will likely include deformation of the host lattice. That is, in its undeformed state there may not exist a channel diameter that can accommodate guest species diffusion. Assuming an isotropically elastic material, a path is sought through a constricted channel that minimizes the deformation of the lattice during guest migration. To identify these constricted channels, overlap between the guest species and the atoms in the host lattice is treated as a negative distance. Extending the concept of the interstitial subspace to include negative distances has an impact on the two main properties of interest in analyzing their topological maps, the LCD and the optimal migration path.

In this approach, we redefine the LCD as the largest (most positive) guest-host separation distance across the smallest volume that continuously spans the supercell. We use a binary search algorithm to select voxels in the regular grid that match a specific criteria, e.g., those above a threshold distance to any atom. To avoid unnecessary memory usage, which is a reported concern of other implementations [17], this Boolean state is stored as a bit mask matching the dimensions of the grid. Two such masks store the upper and lower bounds for a binary search algorithm. The lower bound selects all voxels greater than some arbitrary separation and creates a virtual isovolume whose contiguous spatial limits are connected and span the entirety of the unit cell [refer to schematic in Fig. 3(b)]. The upper bound produces an analogous isovolume that does *not* span the limits of the unit cell. The basics of the binary search algorithm are provided in the following pseudocode:

```
1  SELECT ABOVE threshold ← returns the set of voxels whose
       value lies above threshold
2  set SPANS axis ← TRUE if the set of voxels spans the unit
       cell along axis
3  set CONNECTS ALONG axis ← TRUE if the set of voxels is
       connected across the periodic boundary along axis
4  epsilon ← stopping criterion, equal to 1/2 × finest grid
       resolution
5  upper ← upper bound in the binary search
6  lower ← lower bound in the binary search

7  lcd ← min(grid)
8  for axis in X, Y, Z
9      upper ← max(grid)
10     lower ← min(grid)
11     while (upper-lower) > epsilon
12         mid ← (upper+lower)/2
13         if (SELECT ABOVE mid SPANS axis) and
14            (SELECT ABOVE mid CONNECTS ALONG axis) then
15             lower ← mid
16         else
17             upper ← mid
18         end if
19     end while
20     if (upper+lower)/2 > lcd then
21         lcd ← (upper+lower)/2
22     end if
23 end for
24 RETURN lcd
```

Figure 3 illustrates isovolumes that span the periodic cell along the *X*-axis, but do not span *Y*. Determining the periodic continuity of isovolumes, that is, whether a pore is open or closed, begins by identifying

contiguous volumes within the supercell without considering periodicity. The two important issues are whether an isovolume spans an axis and, if so, if that isovolume connects to itself periodically.

The result of SELECT ABOVE *mid* is a bit mask equal in size to the grid, storing 0 if the value of the corresponding voxel is less than or equal to *mid* or 1 otherwise. Those voxels corresponding to ones in the bit mask constitute the isovolume. An integer array, equal in size to this isovolume, is used to group the voxels. Contiguous void space is identified using a 26-way flood fill algorithm that "colors" all voxels that can be reached by taking successive, single steps up, down, left, right, forward, backward, or diagonally from the first voxel in the isovolume. Once the first pass of the flood fill algorithm finishes, the elements in the integer array corresponding to the colored voxels are assigned the value 1. In the second pass of the flood fill algorithm, the first uncolored voxel, if any remain, is selected as the starting point. The elements in the integer array corresponding to voxels colored during this second pass are assigned the value 2. Additional passes are made until no voxels remain uncolored. Since periodicity is not taken into account during this process, the results – the connected segments – are labeled as in Fig. 3(a).

The CONNECTS ALONG method returns True if and only if any isovolume both spans the axis and matches across the periodic boundary, as shown in Fig. 3(b).[1]

The diffusivity, embodied in the diffusion constant, D depends on the energy barrier associated with the diffusion path through the Arrhenius relationship. We therefore surmise that the migration paths through a material are those that minimize

$$E[\Gamma] = -\oint \nabla U \cdot d\mathbf{s}, \tag{1}$$

i.e., the energy to follow the path $\Gamma = \Gamma(\mathbf{s})$. The purpose of this approach is not to calculate the diffusivity itself, but to identify the path by which diffusion will occur.

Atomic interactions affecting diffusive migration pathways include electrostatic interactions, van der Waals interactions, and several others. For the sake of simplicity and in order to illustrate proof-of-concept, our analysis below assumes a pairwise potential that describes the interaction of the guest species i (location given by the position vector $\mathbf{r}_i \equiv \mathbf{r}$ with the components x, y, z) with each of the atoms j (located at positions $\mathbf{R}_j$, with components X_j, Y_i, Z_j) of the host lattice. The pairwise potential can readily describe the Coulomb interaction in ionic materials, or some covalent short-ranged bonds; the derivation below is applicable in both cases since it does not assume short-range or long-range for the pairwise interaction. More complex
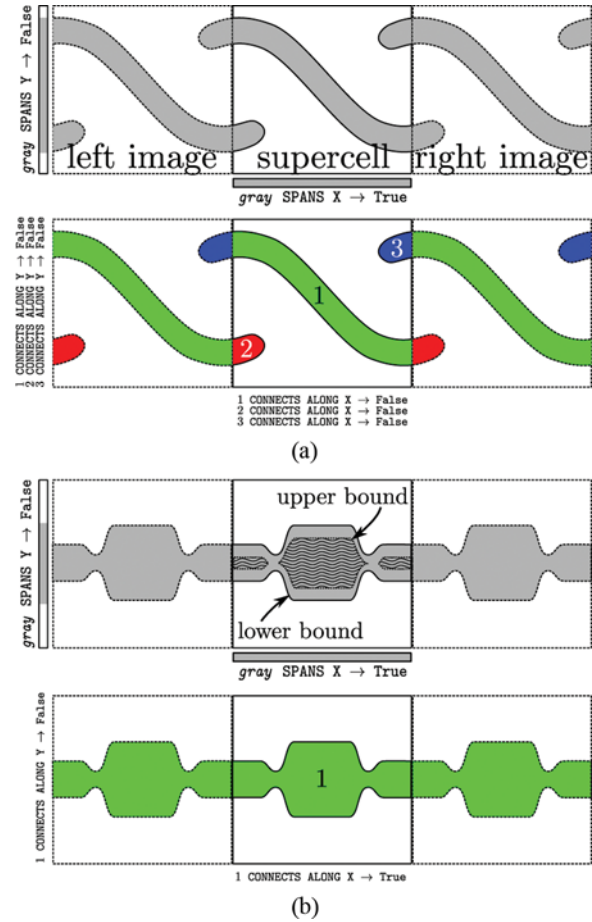

FIGURE 3.—Schematic illustrating cross-sections through isovolumes that span the supercell along the X-axis, but not long Y. Neither volume spans the Y-axis, and so it follows that neither volume can connect along Y. (a) Though isovolume 1 spans the X-axis, it does not connect along X, i.e., isovolume 1 connects to isovolumes 2 and 3, but not 1 (itself). This volume, a closed pore, is not accessible for diffusion and mass transport. (b) Isovolume 1 both spans and connects along the X-axis. An open pore, this volume may be useful for mass transport.

interactions can be also accounted for, with increasing accuracy and computational cost through higher levels of theory, from simple nearest neighbor lattice model models to hybrid density functional theory.

The potential energy landscape experienced by the guest atom i is

$$U = \sum_j u_{ij}(r_{ij})$$

depends only on the distances $|\mathbf{r}_{ij}| = r_{ij}$ between the host atoms j and atom i $r_{ij} = [(X_j - x)^2 + (Y_j - y)^2 + (Z_j - z)^2]^{1/2}$. The force experienced by atom i is given by

$$-\nabla U = -\sum_j \frac{du_{ij}(r_{ij})}{dr_{ij}} \nabla r_{ij}, \tag{2}$$

---

[1]The conditions for CONNECTS ALONG to return true are only satisfied if SPANS is also true, therefore, the test on the 13th line of pseudocode is not strictly necessary, but is included for clarity.

where $\nabla r_{ij} = \mathbf{r}_{ij}/r_{ij} = (\mathbf{R}_j - \mathbf{r})/r_{ij}$ because the positions of the lattice atoms j are assumed fixed. Substituting Eq. (2) into Eq. (1) yields the energy cost to follow the path Γ:

$$E[\Gamma] = - \oint \sum_j \frac{1}{r_{ij}} \frac{du_{ij}(r_{ij})}{dr_{ij}} \mathbf{r}_{ij} \cdot d\mathbf{s}. \qquad (3)$$

Progress along the path Γ is accomplished via the infinitesimal step $d\mathbf{s}$ made by the guess species along the path, so $d\mathbf{s} = d\mathbf{r}$. Equation (3) becomes

$$E = \oint \sum_j F_{ij} \frac{\mathbf{R}_j - \mathbf{r}}{r_{ij}} \cdot d\mathbf{r},$$

where $F_{ij} = -du_{ij}/dr_{ij}$ is the value of the force arising from the interatomic potential u_{ij}. The above equation can be further cast in the form:

$$E = \oint ds |\nabla r| \sum_j \alpha_{ij} F_{ij}, \qquad (4)$$

where α_{ij} derived by straightforward comparison with the previous equation, captures the effect that ∇r has on $\mathbf{r}$ along the path, and lies in the range $[-1, 1]$. If $\alpha_{ij} > 0$, then ∇r increases the distance between the jth atom in the host lattice and the guest atom; conversely, if $\alpha_{ij} < 0$, r decreases. Table 1 summarizes the effect each of these terms have on the cumulative energy.

Intuitively, if the force acting on the guest is repulsive, then the energy can be reduced by elongating the "bond" and the energy cost of traversing a path is minimized by maximizing the distance between guest and host. In a dense system, where insertion of a guest species is done against a persistent (though not constant) repulsive force, the path that maximizes r will minimize the energy. This is true where the magnitude of the energy cost to further compress a bond by an amount Δ is greater than that to decompress the bond by Δ, i.e., where the potential energy curve, $U(r)$ has a positive concavity.

If a region of the host lattice is not fully dense, i.e., the guest atom is smaller than an interstitial site, then the correlation between maximizing distance and the minimum energy is broken. However, if migration is restricted to move through a narrow channel, one that lies within the repulsive range of the potential, then the path that maximizes the interatomic distance is still likely to take the guest species over, or near, the saddle point separating adjacent local energy minima.

Although data that directly provides crystallographically resolved lithium diffusivity is sparse, Fig. 4 consolidates such data and shows the activation energy for

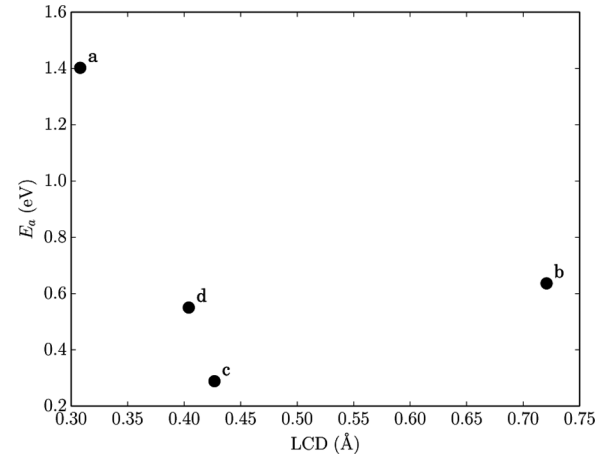

FIGURE 4.—Inverse relationship between the activation energy for (a) LiNbO$_3$ [35], (b) LiFePO$_4$ [36], (c) TiS$_2$ [37], (d) Li$_{12}$Si$_7$ [38] and the largest channel diameter, calculated following the topological mapping analysis described in this work.

diffusion and the LCD, calculated as part of this work. Although this data is not conclusive, it supports the inverse correlation between activation energy and LCD suggested from the preceding derivation.

The resulting geometry of the pore may lend itself to multiple potential diffusion paths, and while direct visualization can provide a route to identifying likely paths, such a manual approach is not acceptable for high-throughput examination. Automatic identification of likely migration pathways is accomplished through the integration of an Eikonal equation:

$$|\nabla U| = C(\mathbf{r}),$$

whose traversal minimizes the cost function, $C(\mathbf{r})$, in the interstitial subspace. It is clear from this equation that $C(\mathbf{r})$ must be everywhere positive, and therefore, the lowest-cost path can be identified using a 26-way Dijkstra algorithm [39].

As demonstrated above, the optimal path through the void subspace may be approximated by maximizing the total separation between the guest species and the host lattice as the former migrates through the crystal; an assertion that is presented graphically in Fig. 5. Since no potential pathways avoid overlap, the interaction between the guest atoms and atoms in the host lattice fall within the red shaded area. Therefore, the energy decreases monotonically with increasing distance between the guest species and atoms in the host lattice.

Overlap, i.e., negative distances, presents a problem for traversal of the Eikonal equation, which requires $C(\mathbf{r})$ to be everywhere positive. Rather than maximizing the distance (minimizing the negative of the distance), $C(\mathbf{r})$ is redefined to favor traversal through non-overlapping, or minimally overlapping, pathways, e.g.,
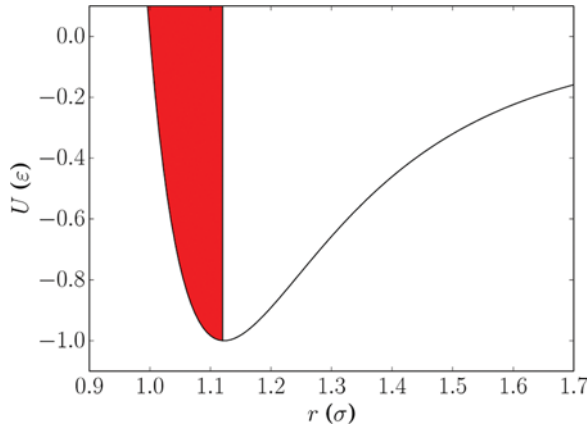
$$C(\mathbf{r}) = e^{-d(\mathbf{r})},$$

TABLE 1.—Effect that each variable in Eq. 4 has on the cumulative energy. (−) implies the energy is reduced; (+), the energy increases.

	$F_{ij} < 0$ (attractive)	$F_{ij} > 0$ (repulsive)
$\alpha_{ij} < 0$ (contraction)	−	+
$\alpha_{ij} > 0$ (elongation)	+	−

FIGURE 5.—In a compact system (see text), no guest species can be inserted with a spacing (r_{ij}) greater than the guest-host atom equilibrium spacing. Using a Lennard-Jones pairwise potential as an example, all guest-host distances fall within the shaded region such that the energy of the system will decrease monotonically with increasing r_{ij}.

where $d(\mathbf{r})$ is the minimum distance between the voxel at $\mathbf{r}$ and any atom in the host lattice. Once the guest specie no longer overlaps any host atom, further increasing the pore diameter only marginally reduces the cost function. The guest atom is effectively free to move about the pore. Physically, increasing overlap increases the cost function exponentially, as in, for example, a Buckingham-style pairwise potential.

RESULTS AND DISCUSSION

Comparison of the scaling efficiency of the optimized CPU and GPGPU algorithms are shown in Fig. 6. The former is run on a six-core Intel® Xeon® E5-2620 2.0 GHz processor, where a best-fit to the 1754 sample

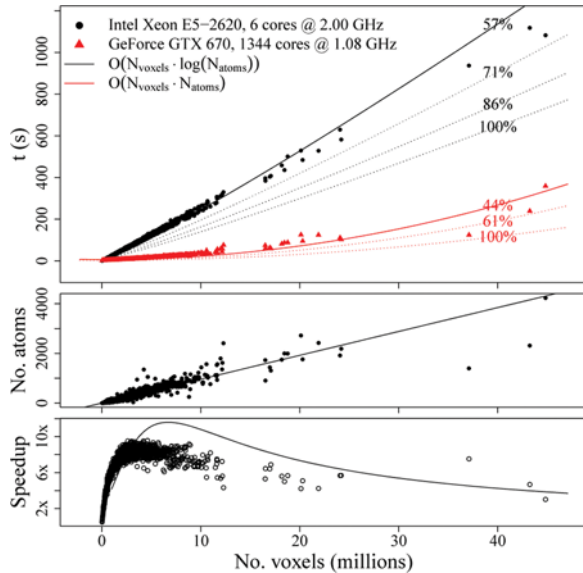FIGURE 6.—Comparison of the scaling efficiency of the optimized CPU and GPGPU algorithms.

timings reveals a 57% parallel efficiency ε, where $\varepsilon = N t_{\text{serial}}/t_{\text{parallel}}$. N is the number of parallel cores; six, in this instance. The GPGPU algorithm maintains a 44% utilization based on best-fit timings. Improved efficiency was realized by switching from a dimension-per-thread parallelization scheme (28% utilization) to an atom-per-thread scheme (44% utilization).

With a memory capacity of 2 GB shared across seven (7) multiprocessors, the Nvidia® GeForce® GTX670 accelerator used in this work is capable of storing a 73 million (73 M) voxel grid of 4-byte (single precision) floats, though practical utilization is less. The block-level shared memory limit of 48 kB precludes block-level sharing of grid data. However, since each voxel will be accessed only ~10 times, the memory latency in storing the grid in global memory is not expected to impact performance significantly.

The problem of topological mapping is partitioned so that each block considers a single voxel, and each thread within that block calculates the distance between one atom and that voxel. The 48 kB of block-level shared memory sets an upper limit of 4096 atoms, for 3×4-byte (single-precision) floats per atom. This limit is superseded, however, by the 1024 threads/block limit on this accelerator; therefore, each block can calculate the distance to only $N = 1024$ atoms. Furthermore since the threads are grouped into n-thread warps ($n = 32$, presently), only systems with $N \bmod n = 0$ have the potential to achieve 100% thread utilization.

For a system of N atoms, each voxel is handled by $\lceil N/1024 \rceil$ blocks. Since each block performs an independent reduction, this establishes a race condition. Consider a system of 2048 atoms. The minimum atom-voxel distance for each voxel will be distributed across two blocks: block A will calculate the distance between the voxel and atoms 1-1024, while block B will calculate the distance between that same voxel and atoms 1025-2048. The following enumeration describes the possible race condition at execution times $t_i < t_{i+1}$.

t_0 Block A calculates the distances between the voxel and atoms 1-1024, then performs a reduction to find the minimum distance, d_A

t_1 Similarly, block B completes its prescribed distance calculations and reduction, d_B

t_2 Block A requests the value at the voxel, v stored in the global memory space.

t_3 Block B similarly requests the value at v.

t_4 Block A performs the comparison $d_A < v$ which evaluates *true* in this scenario, so block A stores $v \leftarrow d_A$ in the global memory space.

t_5 Similarly, block B compares $d_B < v$ where v is the cached value of v. (Recall, the request for v preceded the events of t_4.) If found *true* block B stores $v \leftarrow d_B$.

Following this progression, v ultimately contains d_B. The comparison $d_A < d_B$ is never evaluated, and if $d_A < d_B = $ true, then v contains the wrong value. Therefore, the comparison-assignment operation must be done atomically to ensure algorithmic accuracy.

The minimum distance reduction is the slowest method in this algorithm. With an atom-per-thread parallelization scheme, and assuming a full block, i.e., 1024 atoms, only 512 threads (50% utilization) participate in the first step; 256 in the second; i.e., $N/2^i$ threads: $1 \leq i \leq \lceil \log_2 N \rceil$. Furthermore, to prevent a race condition, similar to the one detailed above for the minimum distance comparisons, each dimidiation step must be done synchronously.

It is clear that these limitations result in accelerator-specific ceilings that must be considered to optimize algorithm performance.

As a demonstration of its utility, this approach is applied to a prototypical application within the Materials Genome Initiative [40], a search for a set of high performance LIB electrode materials. In order to reduce this set, the specific capacity and LCD are plotted for 1,754 lithium-containing compounds found in the Crystallographic Open Database [41–43].

A characteristic example of the results of the topological mapping algorithm is shown in Fig. 7. Since this approach allows for negative distances (overlapping atoms), all pores are open pores at some overlap, and so, in the search for likely migration pathways, the connectivity of the pore volume is identified by determining the LCD as the smallest (most positive) distance that forms a periodically continuous volume.

Assuming all lithium is mobile and that lithium insertion and removal are completely reversible – that is, upon delithiation (discharge) all lithium is removed, and upon lithiation (charging) the initial stoichiometry is recovered – leads to a specific capacity, $\mathcal{C}$, that can be approximated by the stoichiometry of the electrode:
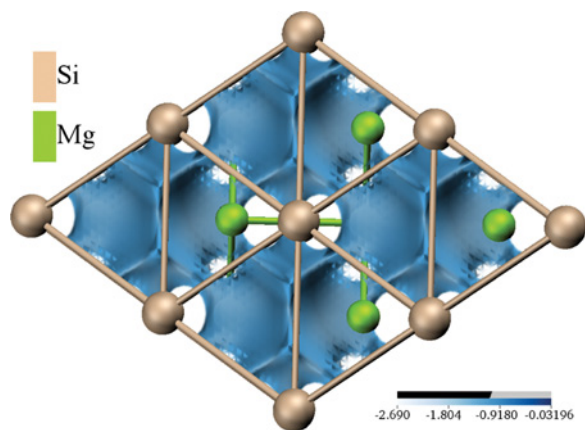
$$\mathcal{C} = \frac{3.6 n F}{M},$$

FIGURE 7.—Topological map of the interstitial subspace at an isovolume level that connects adjacent interstitial sites. The primitive cell is replicated $2 \times 2 \times 2 = 8$ times to show the connection across the periodic boundaries, and the likely migration path taken by the guest species through the host lattice. The negative values in the scalar field correspond to an overlap between the lithium (guest) atom and the nearest host atom, indicating that this lattice is compact (see text).
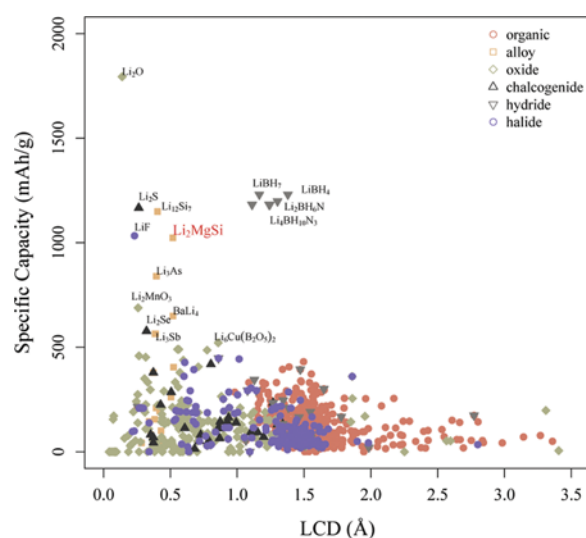

FIGURE 8.—Specific capacity versus largest channel diameter (LCD) of 1754 compounds, categorized by group, reveals chemistries with promising thermodynamics (large specific capacity) and kinetics (large LCD).

here n is the number of electrons involved, i.e., the number of electrons per cation (1 for lithium) times the number of mobile cations, F is Faraday's constant, and M is the formula mass of the electrode.

Figure 8 plots the specific capacity against the LCD, further categorizing each based on their general chemistry, i.e., organics, alloys, oxides, chalcogenides, hydrides, and halides. Many, such as the borohydrides, oxides, and silicides, have been studied extensively [44, 45], and each present challenges for use in secondary (rechargeable) lithium ion batteries. One, Li_2MgSi, has a promising capacity, and an LCD greater than $Li_{12}Si_7$, which itself is known to be a functional alloying anode, and so suggests that Li_2MgSi itself may have sufficient kinetics to function as an electrode.

CONCLUSIONS

A highly parallelized topological mapping algorithm, optimized for acceleration on graphics (GPU) hardware, is presented and characterized for use in high-throughput materials screening.

By dividing the supercell into a regular grid and treating the atom positions as a point cloud, this method makes no assumptions about the crystallinity of the structure, and therefore, is applicable to both crystalline, non-crystalline, organic, and inorganic materials. Furthermore, the relationship between each point in the grid and each atom can be defined in any arbitrary manner. In the present treatment, this relationship is simply the Euclidean distance, but may be readily extended to other scalar-valued property fields, such as potential energy, and to vector fields, such as force.

Analysis of the resulting topological map is enabled by two algorithms. First, identification of the LCD is accomplished through a memory-efficient binary search. Used as an indicator of diffusivity, the LCD is the

largest channel through which a guest species may pass in order to traverse the host structure. Second, the path that minimizes the energy required for the guest atom to traverse the host structure is found through integration of an Eikonal equation by means of a Dijkstra algorithm.

The efficacy of this approach is demonstrated through a test case: identification of the candidate LIB electrode material Li_2MgSi by down-selection from an unfiltered list of 1754 lithium-containing crystal structures. This is done by combining crystallochemical information, available through existing materials informatics platforms, with the above characterization of the interstitial subspace.

References

1. Nørskov, J.K.; Bligaard, T.; Rossmeisl, J.; Christensen, C.H. Towards the computational design of solid catalysts. *Nature Chemistry* **2009**, *1* (1), 37–46.

2. Munter, T.R.; Landis, D.D.; Abild-Pedersen, F. Jones, G.; Wang, S.; Bligaard, T. Virtual materials design using databases of calculated materials properties. *Computational Science & Discovery* **2009**, *2* (1), 015006.

3. Sauter, N.K.; Hattne, J.; Grosse-Kunstleve, R.W.; Nathaniel Echols. New Python-based methods for data processing. *Acta Crystallographica. Section D, Biological Crystallography* **2013**, *69* (Pt 7), 1274–1282.

4. Jain, A.; Hautier, G.; Moore, C.J.; Ong, S.P.; Fischer, C.C.; Mueller, T.; Persson, K.A.; Ceder, G. A high-throughput infrastructure for density functional theory calculations. *Computational Materials Science* **2011**, *50* (8), 2295–2310.

5. Curtarolo, S.; Hart, G.L.W.; Buongiorno-Nardelli, M.; Mingo, N.; Sanvito, S.; Levy, O. The high-throughput highway to computational materials design. *Nature Materials* **2013**, *12* (3), 191–201.

6. Hafner, J.; Wolverton, C.; Ceder, G. Toward computational materials design: The impact of density functional on materials research. *MRS Bulletin* **2006**, *31* (September), 659–668.

7. Barbour, L.J. Crystal porosity and the burden of proof. *Chemical Communications (Cambridge, England)* **2006**, (11), 1163–1168.

8. Blatov, V.A.; Ilyushin, G.D.; Blatova, O.A.; Anurova, N.A.; Ivanov-Schits, A.K.; Dem'yanets, L.N. Analysis of migration paths in fast-ion conductors with Voronoi-Dirichlet partition. *Acta Crystallographica. Section B, Structural science* **2006**, *62* (Pt 6), 1010–1018.

9. First, E.L.; Floudas, C.A. MOFomics: Computational pore characterization of metal–organic frameworks. *Microporous and Mesoporous Materials* **2013**, *165*, 32–39.

10. Haldoupis, E.; Nair, S.; Sholl, D.S. Efficient calculation of diffusion limitations in metal organic framework materials: a tool for identifying materials for kinetic separations. *Journal of the American Chemical Society* **2010**, *132* (21), 7528–7539.

11. Haranczyk, M.; Sethian, J.A. Navigating molecular worms inside chemical labyrinths. *Proceedings of the National Academy of Science* **2009**, *106* (51), 21472–21477.

12. Haranczyk, M.; Sethian, J.A. Automatic structure analysis in high-throughput characterization of porous materials. *Journal of Chemical Theory and Computation* **2010**, *6* (11), 3472–3480.

13. Kim, J.; Martin, R.L. High-throughput Characterization of Porous Materials Using Graphics Processing Units. *Journal of Chemical Theory and Computation* **2012**, *8*, 1684–1693.

14. Pinheiro, M.; Martin, R.L.; Rycroft, C.H.; Jones, A.; Iglesia, E.; Haranczyk, M. Characterization and comparison of pore landscapes in crystalline porous materials. *Journal of Molecular Graphics and Modelling* **2013**, *44*, 208–219.

15. Watanabe, T.; Sholl, D.S. Accelerating applications of metal-organic frameworks for gas adsorption and separation by computational screening of materials. *Langmuir* **2012**, *28* (40), 14114–28.

16. Willems, T.F.; Rycroft, C.H.; Kazi, M.; Meza, J.C.; Haranczyk, M. Algorithms and tools for high-throughput geometry-based analysis of crystalline porous materials. *Microporous and Mesoporous Materials* **2012**, *149* (1), 134–141.

17. Martin, R.L.; Donofrio, D.D.; Sethian, J.A.; Haranczyk, M. Accelerating analysis of void space in porous materials on multicore and GPU platforms. *International Journal of High Performance Computing Applications* **2012**, *26* (4), 347–357.

18. Foster, M.D.; Rivin, I.; Treacy, M.M.J.; Delgado Friedrichs, O. A geometric solution to the largest-free-sphere problem in zeolite frameworks. *Microporous and Mesoporous Materials* **2006**, *90* (1–3), 32–38.

19. Theisen, K.; Smit, B.; Haranczyk, M. Chemical hieroglyphs: abstract depiction of complex void space topology of nanoporous materials. *Journal of Chemical Information and Modeling* **2010**, *50* (4), 461–469.

20. Oganov, A.R., Ed. Modern Methods of Crystal Structure Prediction. Wiley-VCH: Weinheim, Germany, 2010.

21. Ciobanu, C.V.; Wang, C.-Z.; Ho, K.-M. Atomic Structure Prediction of Nanostructures, Clusters, and Surfaces. Wiley-VCH: Weinheim, Germany, 2013.

22. Ciobanu, C.V.; Wang, C.-Z.; Ho, K.-M. Global optimization of 2-dimensional nanoscale structures: A brief review. *Materials and Manufacturing Processes* **2009**, *24*, 109.

23. Davies, T.E.B.; Mehta, D.P.; Rodriguez-Lopez, J.-L. Gilmer, G.H.; Ciobanu, C.V. A variable-number genetic algorithm for growth of 1-D nanostructures into their global minimum configuration under radial confinement. *Materials and Manufacturing Processes* **2009** *24*, 265.

24. Paszkowicz, W. Genetic algorithms, a nature-inspired tool: A survey of applications in materials science and related fields: Part II. *Materials and Manufacturing Processes* **2013**, *28*, 708.

25. Kang, B.; Ceder, G. Battery materials for ultrafast charging and discharging. *Nature* **2009**, *458*, 190–193.

26. Nishimura, S.; Kobayashi, G.; Ohoyama, K.; Kanno, R.; Yashima, M.; Yamada, A. Experimental visualization of lithium diffusion in Li_xFePO_4. *Nature Materials* **2008**, *7* (9), 707–711.

27. Whittingham, M.S. Lithium batteries and cathode materials. *Chemical reviews* **2004**, *104* (10), 4271–4301.

28. Aricò, A.S. Bruce, P.; Scrosati, B.; Tarascon, J.-M.; van Schalkwijk, W. Nanostructured materials for advanced energy conversion and storage devices. *Nature materials* **2005**, *4* (5), 366–377.

29. Li, H.; Wang, Z.; Chen, L.; Huang, X. Research on advanced materials for Li-ion batteries. *Advanced Materials* **2009**, *21* (45), 4593–4607.

30. Ban, C.; Kappes, B.B.; Xu, Q.; Engtrakul, C.; Ciobanu, C.V.; Dillon, A.C.; Zhao, Y. Lithiation of silica through partial reduction. *Applied Physics Letters* **2012**, *100* (24), 243905.

31. Panasonic. Rechargeable Lithium Ion OEM Batteries - Panasonic, 2014. http://www.panasonic.com/industrial/batteries-oem/oem/lithium-ion.aspx (accessed May 20, 2014).

32. Ong, S.P.; Richards, W.D.; Jain, A.; Hautier, G.; Kocher, M.; Cholia, S.; Gunter, D.; Chevrier, V.L.; Persson, K.A.; Ceder, G. Python Materials Genomics (pymatgen), A robust, open-source python library for materials analysis. *Computational Materials Science* **2013**, *68*, 314–319.

33. Bahn, S.R.; Jacobsen, K.W. An object-oriented scripting interface to a legacy electronic structure code. *Computing in Science & Engineering* **2002**, *4* (3), 56–66.

34. Klöckner, A.; Pinto, N.; Lee, Y.; Catanzaro, B.; Ivanov, P.; Fasih, A. PyCUDA and PyOpenCL: A scripting-based approach to GPU run-time code generation. *Parallel Computing* **2012**, *38* (3), 157–174.

35. Hüser, E.; Rahn, J.; Stahn, J.; Geue, T.; Heitjanscd, P.; Schmidt, H. Lithium diffusion in congruent $LiNbO_3$ single crystals at low temperatures probed by neutron reflectometry. *Physical Chemistry Chemical Physics* **2014**, *16*, 3670–3674.

36. Li, J.; Yao, W.; Martin, S.; Vaknin, D. Lithium ion conductivity in single crystal $LiFePO_4$. *Solid State Ionics* **2008**, *179*, 2016–2019.

37. Kanehori, K.; Kirino, F.; Kudo, T.; Miyauchi, K. Chemical diffusion coefficient of lithium in titanium disulfide single crystals. *Journal of the Electrochemical Society* **1991**, *138* (8), 2216–2219.

38. Kuhn, A.; Sreeraj, P.; Pöttgen, R.; Wiemhöfed, H.-D.; Wilkening, M.; Hietjans, P. Li ion diffusion in the anode material $li_{12} si_7$: Ultrafast quasi-1d diffusion and two distinct fast 3d jump processes separately revealed by ^{7}Li NMR relaxometry. *Journal of the American Chemical Society* **2011**, *133*, 11018–11021.

39. Dijkstra, E.W. A note on two problems in connexion with graphs. *Numerische Mathematik* **1959**, *1* (1), 269–271.

40. The White House. Materials Genome Initiative, 2014. http://www.whitehouse.gov/mgi (accessed July 1, 2014).

41. Downs, R.T.; Hall-Wallace, M. The American Mineralogist Crystal Structure Database. *American Mineralogist* **2003**, *88*, 247–250.

42. Gražulis, S.; Chateigner, D. Downs, R.T.; Yokochi, A.F.T.; M. Quirós, Lutterotti, L.; Manakova, E.; Butkus, J.; Moeck, P.; Le Bail, A. Crystallography Open Database – an open-access collection of crystal structures. *Journal of Applied Crystallography* **2009**, *42* (4), 726–729.

43. Gražulis, S.; Daškevic, A.; Merkys, A.; Chateigner, D.; Lutterotti, L.; Quirós, M.; Serebryanaya, N.R.; Moeck, P.; Downs, R.T.; Le Bail, A. Crystallography Open Database (COD), an open-access collection of crystal structures and platform for world-wide collaboration. *Nucleic Acids Research* **2012**, *40* (D1), D420–D427.

44. Wachtler, M.; Winter, M.; Besenhard, J.O. Anodic materials for rechargeable Li-batteries. *Journal of Power Sources* **2002**, *105* (2), 151–160.

45. Zhou, W.; Upreti, S.; Stanley Whittingham, M. High performance Si/MgO/graphite composite as the anode for lithium-ion batteries. *Electrochemistry Communications* **2011**, *13* (10), 1102–1104.